International Telecommunications Symposium – ITS2002, Natal, Brazil

Robust Blind Channel Identification through Regularized Linear Prediction

Cláudio José Bordin Jr. and Luiz Antonio Baccalá
Escola Politécnica da Universidade de São Paulo

Abstract—**This work reviews the blind linear prediction (LP) approach to combat the loss of digital communications system performance due to intersymbol interference (ISI) and discusses that its alleged robustness to model order overestimation is restricted to the theoretical conditions of exactly known received signal statistics. To circumvent these limitations, alternative solution criteria to the least-squares optimization underlying the LP approach are examined together with simulations that point to regularization as the best candidate that balances variance and bias.**

I. Introduction

INTERSYMBOL interference (ISI) associated with signal dispersion due to the transmission channel can seriously impair system performance at the high signaling rates required for recently proposed systems such as those considered for implementation in the new 3G wireless networks. Traditional approaches to ISI reduction involve channel characterization via the transmission of training sequences in an often inefficient process as constant retraining is necessary, vis-à-vis time-varying channels, that thereby take a heavy toll on the overall information throughput. Because of this, the goal of "blind" identification (i.e. without training) of the transmission channel parameters has become a kind of holly grail.

Blind identification algorithms can be classified in two groups: they are either based exclusively on second order statistics (SOS) of the received signal or on higher order statistics (HOS). Despite their many merits, HOS algorithms in general converge slowly, making them unsuitable for handling fast changing channels. Furthermore, their performance degrades markedly when the statistics of the transmitted signals approach those of a gaussian distribution. These shortcomings stimulated the development of SOS blind identification algorithms, which were, however, long thought impossible because the output signal SOS statistics (autocorrelation) of a channel driven by stationary inputs is uninformative with regard to channel phase. This picture changed following Gardner's [1] observation that channel phase information is preserved if the transmitted signal is cyclostationary, as is true for analog QAM signals sampled at integer multiples of the signaling rate. Since then, intense research resulted in showing that cyclostationarity is equivalently described by multichannel models which serve as the starting point for most of today's algorithms.

Though generally of faster convergence compared to HOS, SOS algorithms have limitations of their own, namely (a) they require the satisfaction of some identifiability conditions and more importantly (b) they are sensitive to errors in the channel model orders. This lack of robustness to errors in the determination of the channel model order remains a major hindrance to their practical use, since most of today's order determination methods (AIC, MDL, etc) tend to be strongly biased towards order overdetermination when using estimated statistics [7].

Among the various approaches to SOS identification, algorithms based on linear prediction (LP) as proposed originally by Slock [2], stand out as they are robust to order overestimation at least when exact statistics are employed. In this work, we discuss the behaviour of certain batch implementations of LP algorithms when estimated statistics are involved. In Section II, we present a general formulation of the channel identification problem, followed in Sections III and IV by a description of the LP approach. Actual blind identification considerations and simulation results are presented in Section V, followed by the discussion of alternative solution estimation criteria in Section VI.

II. Problem formulation

Let x_k be an i.i.d. unit variance symbol sequence transmitted over a linear digital communication system. Assuming correctly acquired carrier synchronism, the continuous-time received signal can be written as:

$$y(t) = \sum_k x_k \; h(t-kT) \quad + \quad b(t) \; , \tag{1}$$

where $h(t)$ is the combined impulse response of the channel and the transmit and receive filters; $b(t)$ is additive white zero mean noise.

The sequence y_n, obtained by uniformly sampling the received signal $y(t)$ with period $\tau = \frac{T}{L}$, $L \in \mathbb{N}$ can then be expressed as:

$$\begin{array}{rcl} y_n = y(n\tau) & = & \sum_k x_k \; h(n\tau - kT) + b(n\tau) \\ & = & \sum_k x_k \; h\left(\left(\frac{n}{L} - k\right) T\right) + \\ & & b\left(\left(\frac{n}{L} - k\right) T\right) \; . \end{array} \tag{2}$$

For $L > 1$, y_n is no longer the result of a linear time invariant filtering operation. Yet, certain subsequences obtained through the polyphase decomposition of y_n do exhibit this property as is immediately apparent from defining the following subsequences:

$$\begin{array}{rcl} y_n^{(i)} & = & y\left(\left(n + \frac{i}{L}\right) T\right) \\ h_n^{(i)} & = & h\left(\left(n + \frac{i}{L}\right) T\right) \\ b_n^{(i)} & = & b\left(\left(n + \frac{i}{L}\right) T\right) \; . \end{array} \tag{3}$$

Under the additional assumption that $h(t) \neq 0$ only for $t \in [0, (M+1)T]$, (2) can be rewritten as:

$$y_n^{(i)} = \sum_{k=0}^{M} x_{n-k} \; h_k^{(i)} \quad + \quad b_n^{(i)} \; . \tag{4}$$

which characterizes the channel as a SIMO system.

Cláudio José Bordin Jr. and Luiz Antonio Baccalá are with the Communications and Signals Laboratory (LCS), Telecommunications and Controls Department, Escola Politécnica da Universidade de São Paulo, Av. Prof. Luciano Gualberto, trav. 3, n. 158, 05508-900, São Paulo - SP, Brazil, Phone: +55 11 3091 5508, Fax: +55 11 3091 5128, E-mail: {bordin,baccala}@lcs.poli.usp.br

III. The Linear Prediction Based Algorithm (LP)

Linear prediction techniques for blind identification were first introduced by Slock, in [2]. The basic idea is that in the *absence* of additive noise, the received signal y_n admits either a finite order moving average representation or a finite order autoregressive representation. Among other consequences, this allows determining the innovations process $\tilde{y}_n$ associated with y_n through finite order predictors. Moreover, the innovations process $\tilde{y}_n$ at the $n-$th instant is totally defined by the $n-$th transmitted symbol x_n leading to the conclusion that the predictors are equivalent to zero-forcing equalizers in this case.

A. Determination of the prediction error

Before determining the innovations process associated with the received signal y_n , it is convenient to introduce the following notation:

$$\begin{array}{lclcl} H_i & \triangleq & [\; h_i^{(0)}\; h_i^{(1)}\; \ldots\; h_i^{(L-1)}\;]^T & \in \mathbb{C}^{L\times 1} \\ H & \triangleq & [\; H_0^T\; H_1^T\; \ldots\; H_M^T\;]^T & \in \mathbb{C}^{L(M+1)\times 1} \\ Y_i & \triangleq & [\; y_i^{(0)}\; y_i^{(1)}\; \ldots\; y_i^{(L-1)}\;]^T & \in \mathbb{C}^{L\times 1} \\ B_i & \triangleq & [\; b_i^{(0)}\; b_i^{(1)}\; \ldots\; b_i^{(L-1)}\;]^T & \in \mathbb{C}^{L\times 1} . \end{array} \tag{5}$$

Under this notation, the vector innovations process $\tilde{Y}_n$ associated to the vector received signal Y_n can be defined as:

$$\tilde{Y}_n \triangleq Y_n - P(Y_n \mid Y_{n-1},\ldots,Y_0) \ , \tag{6}$$

where $P(Y_n \mid Y_{n-1},\ldots,Y_0)$ is the linear least squares estimator of Y_n given $[Y_{n-1},\ldots,Y_0]$.

As $E\; Y_n Y_m^H = \underline{0}_{L\times L}\ ,\quad m \geq n+M$, the optimal estimator given the whole past of Y_n equals to the optimal estimator given the M last samples of the received signal [4]:

$$P(Y_n \mid Y_{n-1},\ldots,Y_0) = P(Y_n \mid Y_{n-1},\ldots,Y_{n-M}). \tag{7}$$

For $N \geq M$, rewriting (6) leads to:

$$\tilde{Y}_n = F_N \ \left[\, Y_n^T \ \ldots\ Y_{n-N}^T \,\right]^T \ , \tag{8}$$

where $F_N \triangleq [\; I_L \ \ -A_N \;] \ \in \mathbb{C}^{L\times L(N+1)}$. By definition, the filter F_N must be chosen to minimize the prediction error variance:

$$F_N = \arg\min_{F_N} \mathrm{tr}\ \ F_N\, R_N\, F_N^H \ , \tag{9}$$

where $R_N \triangleq E\; [Y_n^H \ldots Y_{n-N}^H]^H [Y_n^H \ldots Y_{n-N}^H]$.

The underlying optimization problem in F_N can be solved by imposing the orthogonality principle [3], i. e.

$$E \begin{bmatrix} Y_{n-1} \\ \vdots \\ Y_{n-N} \end{bmatrix} \tilde{Y}_n^H = \underline{0}_{LN\times L}. \tag{10}$$

Hence

$$\mathcal{H}_N\; E \left\{ \begin{bmatrix} x_{n-1} \\ \vdots \\ x_{n-M-N} \end{bmatrix} [x_n \ \ldots\ x_{n-M-N}]^* \right\}_{N+1}^{H} F_N^H = \underline{0}_{LN\times L}\ , \tag{11}$$

where $\mathcal{H}_N \ \in\ \mathbb{C}^{LN\times M+N}$ is the Sylvester matrix of order N associated with the channel H, i.e.

$$\mathcal{H}_N \triangleq \begin{bmatrix} H_0 & \cdots & H_M & & \\ & \ddots & & \ddots & \\ & & H_0 & \cdots & H_M \end{bmatrix} . \tag{12}$$

Pre-multiplying (11) by $\mathcal{H}_N^{\#}$ (pseudo-inverse) and assuming that the transmitted signal is unit variance i.i.d yields:

$$\left[\underline{0}_{M+N\times 1}\ I_{M+N}\right] \mathcal{H}_{N+1}^H\ F_N^H = \underline{0}_{M+N\times L}\ . \tag{13}$$

The latter multiplication has the effect [8] of selecting the last $M+N$ rows of the matrix $\mathcal{H}_{N+1}^H$ which after adequate partition results in:

$$\mathcal{H}_{N+1}^H = \begin{bmatrix} \mathcal{H}^t \\ \mathcal{H}^b \end{bmatrix} \Rightarrow \mathcal{H}^b \begin{bmatrix} I_L \\ -A_N^H \end{bmatrix} = \underline{0}_{M+N\times L}\ . \tag{14}$$

As show in [8] $\mathcal{H}^t$ is of the form:

$$\mathcal{H}^t = [\; H_0^H \quad 0 \quad \ldots \quad 0 \;] \ \in \mathbb{C}^{1\times L(N+1)}. \tag{15}$$

Substituting (15) into (13) yields:

$$\mathcal{H}_{N+1}^H \begin{bmatrix} I_L \\ -A_N^H \end{bmatrix} = [\, H_0\; 0\; \ldots\; 0\,]^H\ . \tag{16}$$

Substituting (16) into (8) yields the minimum least squares prediction error $\tilde{Y}_n$:

$$\begin{array}{rcl} \tilde{Y}_n & = & F_N\ \left[\, Y_n^H\ \ldots\ Y_{n-N}^H \,\right]^H \\ & = & [\, I_L\ \ -A_N]\, \mathcal{H}_{N+1}\, [\, x_n\ \ldots\ x_{n-N-M}]^H \\ & = & [\, H_0\; 0\; \ldots\; 0\,]\, [\, x_n\ \ldots\ x_{n-N-M}]^H \\ & = & H_0\, x_n\ . \end{array} \tag{17}$$

In other words, by (17) the optimal predictor F_N is effectively a zero-forcing equalizer as $\tilde{Y}_n$ depends only on the signal transmitted at the $n-$th instant. Moreover, it is also possible to show that the estimate of the innovations process remains consistent even if the model order M is overestimated, since the former result holds for any $N \geq M+1$.

B. Determination of the optimal predictors

In the previous section we computed the minimum prediction error $\tilde{Y}_n$, by implicitly solving the optimization problem represented by (9). In this section, we derive a closed expression for the optimal predictors F_N. Again, by the orthogonality principle:

$$E \begin{bmatrix} Y_{n-1} \\ \vdots \\ Y_{n-N} \end{bmatrix} \tilde{Y}_n^H = \underline{0}_{LN\times L}\ . \tag{18}$$

Substituting (18) into (8) yields:

$$\begin{bmatrix} Y_{n-1} \\ \vdots \\ Y_{n-N} \end{bmatrix} \left[\, Y_n^H\ \ldots\ Y_{n-N}^H \,\right] \begin{bmatrix} I \\ -A_N^H \end{bmatrix} = \underline{0}_{LN\times L}\ . \tag{19}$$

Let $r_i \triangleq E\, Y_n Y_{n+i}^H$ and the matrices:

$$\begin{array}{lll} p_N & \triangleq & \left[\, r_1 \;\dots\; r_{N+1} \,\right] \quad \in \mathbb{C}^{L\times L(N+1)} \\ R_N & \triangleq & \begin{bmatrix} r_0 & \dots & r_N \\ \vdots & \ddots & \vdots \\ r_N^H & \dots & r_0 \end{bmatrix} \quad \in \mathbb{C}^{L(N+1)\times L(N+1)} \end{array}, \tag{20}$$

so that:

$$\left[\, p_{N-1}^H \; R_{N-1} \,\right] \begin{bmatrix} I \\ -A_N^H \end{bmatrix} = \underline{0}_{LN\times L} \,. \tag{21}$$

Using (21) we obtain:

$$A_N R_{N-1} = p_{N-1} \,. \tag{22}$$

However, since $\mathcal{N}(p_{N-1}) = \mathcal{N}(\mathcal{H}_{N-1}^H) = \mathcal{R}^{\perp}(R_{N-1})$ [4], we have:

$$A_N = p_{N-1} R_{N-1}^{\#} \,, \tag{23}$$

which finally yields:

$$F_N = \left[\, I_L \quad p_{N-1} R_{N-1}^{\#} \,\right] \,. \tag{24}$$

IV. Blind Identification Algorithms based on Linear Prediction

The LP blind identification algorithm due to Abed-Mehrain *et al.* [4] is based on the particular form of the innovations process $\tilde{Y}_n$ in the additive noise free case:

$$\tilde{Y}_n \triangleq F_N \left[\, Y_n^H \;\dots\; Y_{n-N}^H \,\right]^H = H_0 x_n \,. \tag{25}$$

Because $\tilde{Y}_n$ depends only on the signal transmitted at the $n-$th instant, any $i-$th line of F_N is a null delay zero-forcing equalizer (up to a complex scale factor given by $h_0^{(i)}$). More importantly, F_N is completely determined by the covariance matrix R_N and can be estimated from the received signal.

Though F_N could be used to produce an estimate of the channel parameters, [4] uses an additional step where the rows of F_N are linearly optimally combined through a vector $K_{opt} \in \mathbb{C}^{L\times 1}$, whose function is to minimize the estimation error of x_n from $\tilde{Y}_n$:

$$\hat{x}_n = K_{opt}^H \tilde{Y}_n \,. \tag{26}$$

By definition, K_{opt} must satisfy the Wiener-Hopf Equation:

$$K_{opt}^H \;\; E\, \tilde{Y}_n \tilde{Y}_n^H = E\, x_n \tilde{Y}_n^H \,. \tag{27}$$

but it is easy to see that:

$$\begin{array}{rcl} E\, \tilde{Y}_n \tilde{Y}_n^H & = & H_0 H_0^H \\ E\, x_n \tilde{Y}_n^H & = & H_0^H \,. \end{array} \tag{28}$$

where from $E\;\tilde{Y}_n \tilde{Y}_n^H$'s rank, follows that (28) has an infinite number of solutions and that the one with minimal norm is:

$$K_{opt} = \frac{H_0}{\|H_0\|^2} \,. \tag{29}$$

H_0 is unknown *a priori* but it can be estimated through the most dominant eigenvector of the prediction error covariance matrix, i. e. let

$$D \triangleq E\, \tilde{Y}_n \tilde{Y}_n^H = F_N R_N F_N^H = r_0 + p_{N-1} A_N^H \tag{30}$$

which has a single non-zero eigenvalue $\lambda = \|H_0\|^2$, associated to the eigenvector $v = H_0\; \|H_0\|^{-1}$. Thus, the zero delay ZF equalizer $g_N \in \mathbb{C}^{LN\times 1}$ may be estimated as:

$$g_N = F_N^H \frac{v}{\sqrt{\lambda}} \,. \tag{31}$$

A. Identification

The knowledge of a ZF equalizer g_N and of some received signal second order statistics allows the transmission channel H to be identified up to a unit norm complex constant because

$$E\, Y_n x_{n-l}^* = \sum_{k=0}^{M} H_k\; E\, x_{n-k} x_{n-l}^* = \sum_{k=0}^{M} H_k \delta_{k-l} = H_l. \tag{32}$$

As a direct consequence:

$$\left[\, H_0^T \cdots\; H_M^T \; \underline{0}_{L\times 1}^T \cdots \underline{0}_{L\times 1}^T \,\right]^T = E\, Y_n \left[x_n^* \;\cdots\; x_{n-N+1}^* \right]$$
$$= \begin{bmatrix} r_0 & r_1 & \dots & r_{N-1} \\ r_1 & r_2 & \dots & r_N \\ \vdots & \vdots & & \vdots \\ r_{N-1} & r_{N-2} & \dots & r_{2N-2} \end{bmatrix} g_N \tag{33}$$

B. Taking $\sigma_B^2 \neq 0$ into account

The discussion so far was restricted to exact statistics and no additive noise. To account for the latter, consider B_N represents white, zero mean noise, independent of the transmitted signal and of variance σ_B^2. Its sole effect is to add the term $I\sigma_B^2$ to R_N:

$$\begin{array}{l} \hat{Y}_N = \mathcal{H}_N \;\left[\, x_n \;\dots\; x_{n-M-N+1} \,\right]^H + B_N \;\Rightarrow \\ \quad \hat{R}_N = \mathcal{H}_N \mathcal{H}_N^H + I_N \sigma_B^2 \,. \end{array} \tag{34}$$

In other words, this means that additive noise introduces bias into the LP algorithm as g_N is no longer a ZF equalizer:

$$\begin{array}{l} \hat{g}_N \;=\; g_N + \\ \quad \left[\, I_L - ((R_{N-1} + I_{L(N-1)}\sigma_B^2)^{\#} - R_{N-1}^{\#}) p_{N-1}^H \right]^H H_0 \|H_0\|^{-2} \end{array} \tag{35}$$

A simple, although certainly sub-optimal [4] procedure for obtaining unbiased estimates from the LP algorithm is to estimate and subtract the additive noise contribution $\hat{\sigma}_B^2$:

$$R_N = \hat{R}_N - I\hat{\sigma}_B^2 \,. \tag{36}$$

Note that an estimate $\hat{\sigma}_B^2$ can be obtained from the $LN-N-M$ less dominant eigenvalues of the R_Y. For the sake of robustness, however, it is convenient that the noise power estimates do not depend on the exact model order.

V. Performance of the LP algorithm using estimated statistics

In the previous sections, it was implicitly shown that the LP algorithm is robust to channel order overdetermination, at least under exact statistics. In practice, the performance of the LP algorithm, however, degrades significantly if the channel order is overdetermined.

To illustrate this, we perfomed Monte Carlo experiments over 1,000 independent realizations, using independent symbols extracted from a QPSK constellation. The adopted figure-of-merit was the mean-square estimation error (MSE), defined as:

$$MSE = 10\ \log_{10}\ \frac{1}{\|H\|^2\ N_r} \sum_t \|\hat{H} - H\|^2\ , \qquad (37)$$

where $\hat{H}$ are the estimates corrected by the scale factor that minimizes the error norm and N_r is the number of experiments.

The effects of the additive noise were quantified in terms of the noise-to-signal ratio (NSR):

$$NSR = 10\ \log_{10}\ \frac{\sigma_B^2}{\|H\|^2}\ . \qquad (38)$$

Figure 1 shows LP algorithm performance employing the channel model ($M = 3$) described in (39), using 1,000 symbols for the estimation of R_N, whereas Fig. 2 uses 10,000 symbols to estimate R_N.

$$\begin{aligned} H_0 &= [\ \ 1.62 + 2.30i \quad\ 1.76 - 1.37i \quad\ 1.05 + 0.38i\ ]^T \\ H_1 &= [-1.36 + 2.20i \quad\ 1.46 - 0.06i \quad -0.85 - 1.22i\ ]^T \\ H_2 &= [\ \ 0.05 - 0.47i \quad\ 0.11 + 0.24i \quad\ 1.00 + 0.20i\ ]^T \\ H_3 &= [\ \ 0.05 - 0.15i \quad -0.02 - 0.03i \quad\ 0.09 + 0.01i\ ]^T \end{aligned} \qquad (39)$$

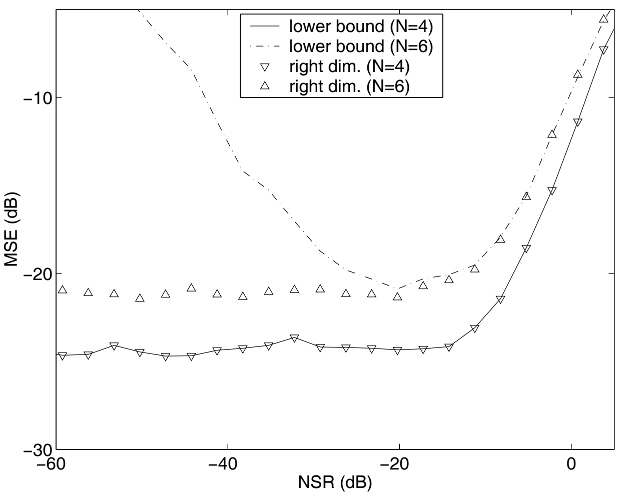

Fig. 1. Average LP algorithm performance as a function of the NSR using 1,000 received signal samples for the channel model in (39). The curves show the performance obtained for correct ($N = 4$) and overestimated ($N = 6$) channel orders, using correct and lower bound noise subspace dimensions.

The results shown in Fig. 1 and 2 illustrate the paramount importance of the exact knowledge of the channel model order. As expected, the performance of the LP algorithm degrades slightly as N increases if the correct noise subspace dimension of R_{N-1} is considered during the determination of $R^{\#}_{N-1}$. In fact, adapting results from [9], it can be shown that:

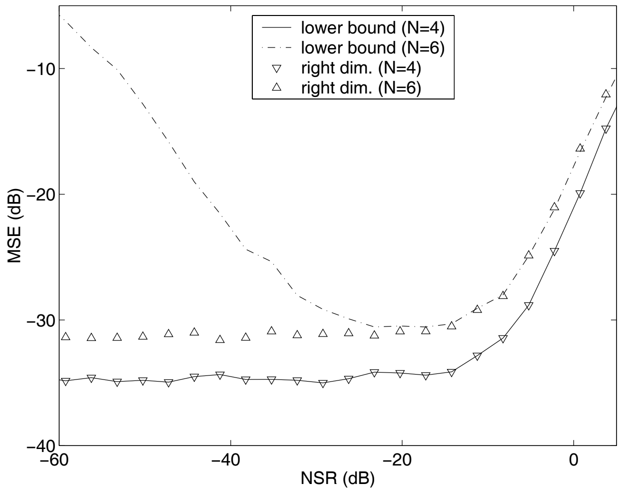

Fig. 2. Average LP algorithm performance as a function of the NSR using 10,000 received signal samples for the channel model in (39). The curves show the performance obtained for correct ($N = 4$) and overestimated ($N = 6$) channel orders, using the correct and the lower bound noise subspace dimensions.

$$\dim \mathcal{N}(R_{N-1}) = \dim \mathcal{N}(\mathcal{H}_{N-1}) = (L-1)(N-1)-M \quad (40)$$

On the other hand, if M is unknown, so is $\dim \mathcal{N}(R_{N-1})$. Despite this, a lower bound to this dimension can be easily obtained as

$$\min_{M \le N-1} \dim\ \mathcal{N}(R_{N-1}) = (L-2)(N-1)\ . \qquad (41)$$

Also from the figures, it can be seen that the underestimation of the R_{N-1} noise subspace dimension produces a strong performance degradation, causing even non-monotonic behaviour of MSE with NSR. It is commonly argued in the literature that the misclassification of some noise subspace eigenvectors as signal eigenvectors would not cause serious trouble to the estimation of A_N, since $\mathcal{N}(p_{N-1}) \subseteq \mathcal{R}(R_{N-1})^{\perp}$ [4], which is approximately verified for the sample estimates. However, as shown in Fig. 3, the eigenvalue spread of R_{N-1} strongly increases as the NSR diminishes, overriding improvements in covariance matrices estimates.

VI. Alternative criteria for the estimation of the predictors

When estimated statistics are involved, it is easy to show that (23) reduces to an estimate of the predictors A_N in the least squares sense. Despite its popularity, the least squares criterion is known to perform poorly in the presence of data model mismatches [11], which is always the case with the LP algorithm, since R_{N-1} is unknown *a priori*. In the next sections, we investigate the use of alternative estimation criteria for the estimation of the A_N predictors.

It is easy to see that (22) is a special case of the general problem $Ax \approx b$, $A \in \mathbb{C}^{m\times n}$ and $b \in \mathbb{C}^{m\times p}$ for which we next consider the Total Least Squares (TLS) and the Least Squares regularization (LSr) approaches.

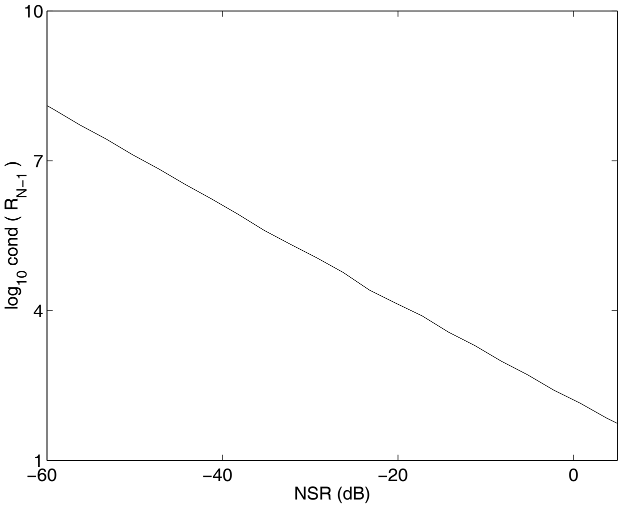

Fig. 3. Average condition number (with respect to the eigenvalues) of R_{N-1} as a function of the NSR for the channel model in (39) and $N = 6$. Each realization was estimated using 10,000 received signal samples.

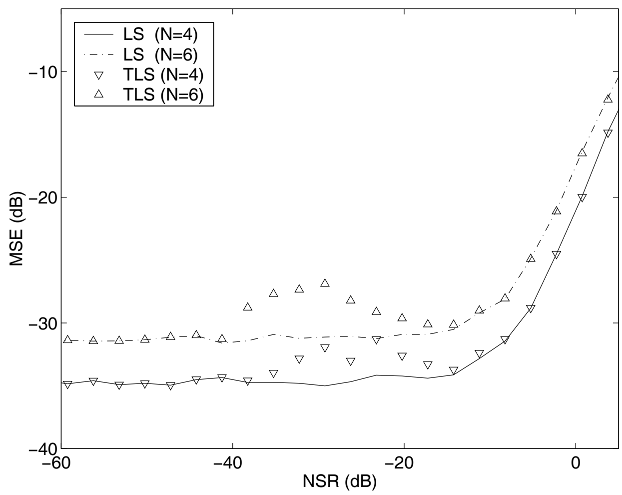

Fig. 4. Average LP algorithm performance for the LS and TLS criteria as a function of the NSR for the channel model in (39) using 10,000 received signal samples. Correct noise subspace dimensions were used.

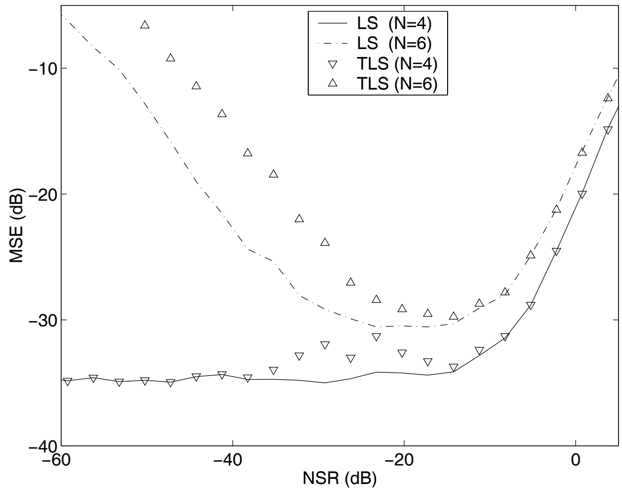

Fig. 5. Average LP algorithm performance for the LS and TLS criteria as a function of the NSR for the channel model in (39) using 10,000 received signal samples. The noise subspace dimension lower bound was used.

A. Estimation by Total Least Squares

The total least squares [6] technique initially seeks:

$$[\ \hat{A}\ \hat{B}\ ] = \underset{[\ \hat{A}\ \hat{B}\ ]}{\arg\min} \|[\ \hat{A}\ \hat{B}\ ] - [\ A\ B\ ]\|_F \quad \text{subject to } \mathcal{R}(\hat{B}) \in \mathcal{R}(\hat{A}) \tag{42}$$

Once a minimizing $[\ \hat{A}\ \hat{B}\ ]$ is found, then any x_{TLS} satisfing:

$$\hat{A}\, x_{TLS} = \hat{B} \tag{43}$$

is called a total least squares solution.

The application of the TLS criterion to (23) is somewhat involved due to the reduced rank of R_{N-1}, which leads the solution to be non-unique. Let $U\Sigma V^H$ be the SVD of $[R_{N-1}^H\ p_{N-1}^H]$. Then, from [6], p.62., one can show that:

$$A_N^{TLS} = -V_{12}V_{22}^{\#}\ , \tag{44}$$

where V_{12} and V_{22} are respectively the blocks formed by the first $L(N-1)$ and the last N rows of the last $q+N$ columns of the matrix V, with $q = dim(\mathcal{N}(R_{N-1}))$.

Figures 4 and 5 compare average LP algorithm performance under the TLS and LS criteria, for both the correct dimension of the noise subspace and its lower bound. As readily apparent, the more elaborate computational nature of the TLS criterion provides no performance benefits.

B. Estimation by Regularized Least Squares (LSr)

Another approach to solve $Ax \approx b$ is the regularized least squares (LSr) [10] technique, whose solution is

$$x_{LSr} = \arg\min_x \left(\delta\|x\|_F^2 + \|Ax - b\|_F^2\right)\ , \tag{45}$$

where δ is a positive constant. Equivalently, the LSr problem solution may be recast as:

$$x_{LSr} = \left(A^H A + I\delta\right)^{-1} A^H b\ .$$

Applying this to find A_N yields:

$$A_N^{LSr} = p_{N-1}R_{N-1}^H\ \left(R_{N-1}R_{N-1}^H + I\delta\right)^{-1}\ . \tag{46}$$

Fig. 6 shows the LP algorithm performance under the LS and LSr ($\delta = 10^{-4}$) solution criteria, using the lower bound noise subspace dimensions. As one can notice, the model model order mismatches have little influence on LSr performance.

However, this result is obtained at the expense of introducing bias into A_N as larger δ values are used. In fact, let $R_{N-1}^+ \triangleq R_{N-1}^H(R_{N-1}R_{N-1}^H + I\delta)^{-1}$ and $\lambda_i,\ 0 \le i \le M+N-1$, be the non-zero eigenvalues of R_{N-1}. Then, it follows that [12]:

$$\begin{aligned} \|A_N^{LSr} - A_N\|_2\ &\le \|p_{N-1}\|_2\ \ \|R_{N-1}^+ - R_{N-1}^{\#}\|_2 \\ &\le \|p_{N-1}\|_2\ \ \max_{\lambda_i}\left\{\tfrac{\delta}{\lambda_i(\lambda_i^*\lambda_i+\delta)}, \tfrac{1}{\delta}\right\}\ . \end{aligned} \tag{47}$$

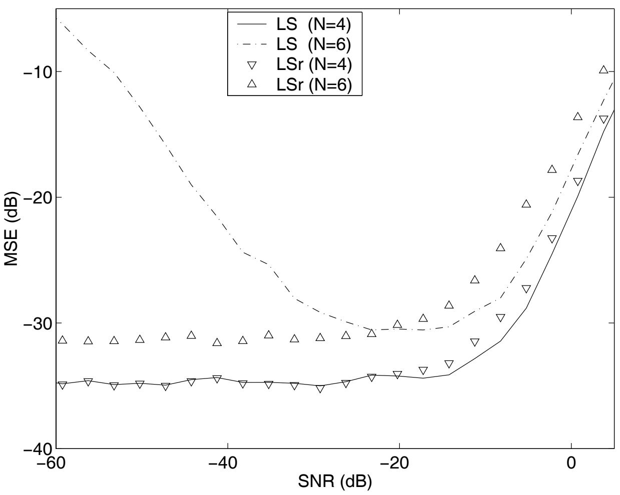

Fig. 6. Average LP algorithm performance for the LS and LSr ($\delta = 10^{-4}$) criteria as a function of NSR for the channel model in (39) using 10,000 received signal samples.

This illustrated in Fig. 7 by depicting average LP algorithm behaviour for a channel containing maximum-phase zeros ($H_1(z) = 1 - 1.5z^{-1}$ and $H_2(z) = 1 - \lambda z^{-1}$) that violate the SOS identifiability condition at $\lambda = 1.5$.

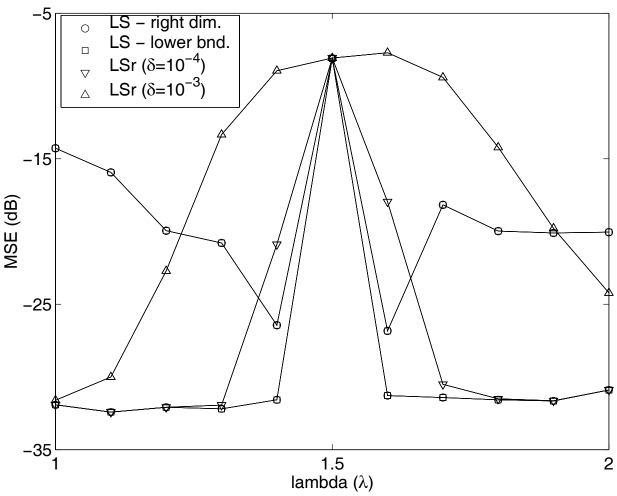

Fig. 7. Average LP algorithm performance as a function of NSR using different optimization criteria for estimating predictors employing 10,000 received signal samples. The channel model has common maximum phase zeros at $\lambda = 1.5$.

In this case, LP algorithm performance under LSr criterion degraded faster as $\lambda \to 1.5$ for larger values of δ. Thus, the value of δ establishes a compromise between variance and bias: the larger its value, the more insensitive the algorithm becomes to model mismatches, at the expense of introducing bias for ill-conditioned channels.

VII. CONCLUSION

In this work, we examined the use of the TLS and LSr optimization criteria to the estimation of the optimal predictors of the LP algorithm. We verified through numerical simulations that the TLS criterion provides neither performance nor robustness improvement. The LSr criterion, however, comes forward as a compromise solution allowing a balance between robustness to model mismatches and bias through an adequate choice of the δ regularization parameter.

REFERENCES

[1] W.A. Gardner, "A New Method of Channel Identification," *IEEE Transactions on Communications*, vol. 39, n. 6, p. 813-817, Juny 1991.

[2] D. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter-banks and multichannel linear prediction," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing 1994* vol. 4, p. 585-588, 1994.

[3] T. Kailath, A.H. Sayed, B. Hassibi, *Linear Estimation*, Prentice Hall, Upper Saddle River, N.J., 2000.

[4] K. Abed-Mehraim, E. Moulines, P. Loubaton, "Prediction Error Method for Second-Order Blind Identification," *IEEE Transactions on Signal Processing*, vol. 45, n. 3, p. 694-705, March 1997.

[5] D. Gesbert, P. Duhamel, "Unbiased Blind Adaptive Channel Identification and Equalization," *IEEE Transactions on Signal Processing*, vol. 48, n. 1, p. 148-158, January 2000.

[6] S. van Huffel, J. Vandewalle *The Total Least Squares Problem: Computational Aspects and Analysis* Siam, Philadelphia, 1991.

[7] A.P. Liavas, P.A. Regalia "On the behavior of information theoretic criteria for model order selection," *IEEE Transactions on Signal Processing*, vol. 49, n. 8, p. 1689-1695, August 2001.

[8] J. Mannerkoski, V. Koivunen "Error Analysis of a Multi-step Prediction Based Blind Equalizer," *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems - ISCAS 99*, vol. 3, p. 86-89, 1999.

[9] E. Moulines, P. Duhamel, J.F. Cardoso, S. Mayrargue "Subspace Methods for Blind Identification of Multichannel FIR Filters," *IEEE Transactions on Signal Processing*, vol. 43, n.2, p. 516-525, February 1995.

[10] A.N. Tikhonov "Regularization of Incorrectly Posed Problems,´´ *Soviet Mathematics*, vol. 4, p. 1624-1627, 1963.

[11] A.H. Sayed, V. Nascimento, S. Chandrasekaran "Estimation and Control with Bounded Data Uncertainties,´´ *Linear Algebra and its Applications*, vol. 284, p. 259-306, November 1998.

[12] R.A. Horn, C.R. Johnson *Matrix Analysis*, Cambridge University Press, 1996.