On robustness of Linear Prediction based blind identification

Luc Deneire * and Dirk T.M. Slock

Institut EURECOM, 2229 route des Crêtes, B.P. 193,
F-06904 Sophia Antipolis Cedex, FRANCE
{deneire, slock}@eurecom.fr

Abstract

Linear prediction based algorithms have been applied to the multi-channel FIR identification problem. In [11], it was shown that oversampled and/or multiple antenna received signals may be modeled as well as low rank MA processes as low rank AR processes. Indeed, taking FIR nature and the singularity of the MA process into account (due to the fact that the number of channels is bigger than the number of sources) leads to a finite order prediction filter (i.e. AR($L < \infty$) modeling), which is automatically identified by, e.g., a singular multichannel Levinson algorithm, and can be shown to be robust to AR order overestimation. On the other hand, K.A. Meraim and A. Gorokhov derive other robustness properties based on the equations $\mathbf{P}(z)\mathbf{H}(z) = \boldsymbol{h}(0)$, where $\mathbf{P}(z)$ is the prediction filter $\mathbf{H}(z)$ is the channel and $\boldsymbol{h}(0)$ its first coefficient. Although using $\mathbf{P}(z)$ of overestimated order, clever use of the previous equations leads to robustness of the estimation of $\mathbf{H}(z)$ to channel length overestimation. This paper investigates these robustness issues, comparing both methods (and derived methods) to order estimation algorithms for, e.g., subspace-fitting methods. An important point developed hereunder is the implicit order estimation schemes present in linear prediction based methods and their influence on identification performance. Furthermore, we develop a new order estimation method, of low computational cost and giving the channel estimate as a by-product.

1 Introduction

Lots of batch multichannel identification algorithms based on Second Order Statistics have been developed recently [3, 1]. Among these, the Linear Prediction (LP) based algorithms have the following advantages:

- They are robust to order overestimation. We will study the mechanisms which lead to this robustness and to what extent this robustness holds.
- They are computationally efficient, as will be demonstrated hereunder.
- The weighted LP can be shown to be asymptotically statistically equivalent to Weighted Noise Subspace Fitting with proper parameterization (see companion paper in this conference).
- They are able to identify minimum phase common zeros among the different channels [12].

All these advantages leads us to think that LP methods are good candidates for blind channel identification, at least as a startup method.

LP methods consist in two main part, the first one identifies the noiseless equivalent AR model, the second part deduces the channel from the LP filter coefficients. The niceties of LP are that both steps are robust, the first to AR model order overestimation and the second to channel order overestimation. We will first present the different approaches to LP for channel estimation, identify the robustness features and then propose a global approach where the three first cited advantages are combined. The first part of the algorithm consists of a multichannel singular Levinson algorithm with AR order estimation and the second part of the Weighted LP approach introduced by [4].

2 Data Model

Consider linear digital modulation over a linear channel with additive Gaussian noise. Assume that we have p transmitters at a certain carrier frequency and m antennas receiving mixtures of the signals. We shall assume that $m > p$. The received signals can be written in the baseband as

$$y_i(t) \;=\; \sum_{j=1}^{p}\sum_{k} a^j(k)h_i^j(t-kT) \;+\; v_i(t) \qquad (1)$$

*The work of Luc Deneire is supported by the EC by a Marie-Curie Fellowship (TMR program) under contract No ERBFMBICT950155

where the $a^j(k)$ are the transmitted symbols from source j, T is the common symbol period, $h_i^j(t)$ is the (overall) channel impulse response from transmitter j to receiver antenna i. Assuming the $\{a^j(k)\}$ and $\{v_i(t)\}$ to be jointly (wide-sense) stationary, the processes $\{y_i(t)\}$ are (wide-sense) cyclostationary with period T. If $\{y_i(t)\}$ is sampled with period T, the sampled process is (wide-sense) stationary. Sampling in this way leads to an equivalent discrete-time representation. We could also obtain multiple channels in the discrete-time domain by oversampling the continuous-time received signals, see [7],[11].

We assume the channels to be FIR. In particular, after sampling we assume the (vector) impulse response from source j to be of length N^j. Without loss of generality, we assume the first non-zero vector impulse response sample to occur at discrete-time zero. Let $N = \sum_{j=1}^{p} N^j$ and $N^1 = \max_j(N^j)$. The discrete-time received signal can be represented in vector form as

$$\begin{aligned} \boldsymbol{y}(k) &= \sum_{j=1}^{p}\sum_{i=0}^{N^j-1} \boldsymbol{h}^j(i)a^j(k{-}i) + \boldsymbol{v}(k) \\ &= \sum_{i=0}^{N^1-1} \boldsymbol{h}(i)\boldsymbol{a}(k{-}i) + \boldsymbol{v}(k) \\ &= \sum_{j=1}^{p} \boldsymbol{H}^j A^j_{N^j}(k) + \boldsymbol{v}(k) \;=\; \boldsymbol{H}\boldsymbol{A}_N(k) + \boldsymbol{v}(k) \end{aligned} \tag{2}$$

$$\begin{aligned} &\boldsymbol{y}(k) = \left[y_1^H(k)\cdots y_m^H(k)\right]^H, \\ &\boldsymbol{v}(k) = \left[v_1^H(k)\cdots v_m^H(k)\right]^H, \\ &\boldsymbol{h}^j(k) = \left[h_1^{jH}(k)\cdots h_m^{jH}(k)\right]^H \\ &\boldsymbol{H}^j = \left[\boldsymbol{h}^j(N^j{-}1)\cdots \boldsymbol{h}^j(0)\right], \boldsymbol{H} = \left[\boldsymbol{H}^1\cdots\boldsymbol{H}^p\right], \\ &\boldsymbol{h}(k) = \left[\boldsymbol{h}^1(k)\cdots\boldsymbol{h}^p(k)\right], \boldsymbol{H}_i^j = \text{line } i \text{ of } \boldsymbol{H}^j \\ &\boldsymbol{a}(k) = \left[a^{1H}(k)\cdots a^{pH}(k)\right]^H, \\ &A_n^j(k) = \left[a^{jH}(k{-}n{+}1)\cdots a^{jH}(k)\right]^H, \\ &\boldsymbol{A}_N(k) = \left[A_{N^1}^{1H}(k)\cdots A_{N^p}^{pH}(k)\right]^H \end{aligned} \tag{3}$$

where superscript ^{H} denotes Hermitian transpose.

We consider additive temporally and spatially white Gaussian circular noise $\boldsymbol{v}(k)$ with $R_{vv}(k-i) = \mathrm{E}\left\{\boldsymbol{v}(k)\boldsymbol{v}^H(i)\right\} = \sigma_v^2 I_m \delta_{ki}$. Assume we receive M samples :

$$\boldsymbol{Y}_M(k) \;=\; \mathcal{T}_M^p(\boldsymbol{H})\; \boldsymbol{A}_{N+p(M-1)}(k{+}M{-}1) + \boldsymbol{V}_M(k) \tag{4}$$

where $\boldsymbol{Y}_M(k) = \left[\boldsymbol{Y}^H(k)\cdots\boldsymbol{Y}^H(k{+}M{-}1)\right]^H$ and $\boldsymbol{V}_M(k)$ is defined similarly whereas $\mathcal{T}_M^p(\boldsymbol{H})$ is the multichannel multiuser convolution matrix of $\boldsymbol{H}$, with M block lines. Therefore, the structure of the covariance matrix of the received signal $\boldsymbol{Y}(k)$ is

$$R_{YY} = \mathcal{T}_M^p(\boldsymbol{H})\, R_{AA}\mathcal{T}_M^{pH}(\boldsymbol{H}) + \sigma_v^2 I_{mM} \tag{5}$$

where $R_{AA} = \mathrm{E}\left\{\boldsymbol{A}_{N+p(M-1)}(k)\boldsymbol{A}^H_{N+p(M-1)}(k)\right\}$. From here on, we will assume white sources with power $\sigma_a^2 (R_{AA} = \sigma_a^2 I)$.

3 FIR Zero-Forcing Equalization

We consider an equalizer $F(z)$ such that $F(z)H(z) = \text{diag}\{z^{-n_1}\ldots z^{-n_p}\}$, which can be written in the time-domain as

$$\boldsymbol{F}\mathcal{T}_L^p(\boldsymbol{H}) = \begin{bmatrix} 0\cdots1\cdots0 & \cdots & 0\cdots0 \\ \vdots & \ddots & \vdots \\ 0\cdots0 & \cdots & 0\cdots1\cdots0 \end{bmatrix} \tag{6}$$

$F(z)$ is a $p\times m$ filter of order L.

(6) is a system of $p(N+p(L-1))$ equations in Lmp unknowns. The minimum length L of the FIR equalizer is such that the system (6) is exactly or under-determined. Hence

$$L \geq \underline{L} = \left\lceil \frac{N-p}{m-p} \right\rceil \tag{7}$$

We assume than $\boldsymbol{H}$ has full rank if $N \geq m$, otherwise, there is lack of channel diversity (space or time diversity according to the manner the channels were obtained) and only a subset of the channels is relevant.

4 LP and Equalization

4.1 Noise-free Linear Prediction

Consider the problem of predicting $\boldsymbol{y}(k)$ from $\boldsymbol{Y}_L(k-1)$, where $\boldsymbol{Y}_L(k-1)$ is considered noiseless (in "real life", we will use $R_{YY} = R_{YY} - \sigma_v^2 I$). The prediction error can be written as

$$\widetilde{\boldsymbol{y}}(k)|_{\boldsymbol{Y}_L(k-1)} = \boldsymbol{y}(k) - \widehat{\boldsymbol{y}}(k)|_{\boldsymbol{Y}_L(k-1)} = \boldsymbol{P}_L\boldsymbol{Y}_{L+1}(k) \tag{8}$$

with $\boldsymbol{P}_L = [\boldsymbol{P}_{L,L}\cdots\boldsymbol{P}_{L,1}\;\boldsymbol{P}_{L,0}]$, $\boldsymbol{P}_{L,0} = I_m$. Minimizing the prediction error variance leads to the following optimization problem

$$\min_{\boldsymbol{P}_L:\boldsymbol{P}_{L,0}=I_m} \boldsymbol{P}_L R_{YY}\boldsymbol{P}_L^H \;=\; \sigma^2_{\tilde{y},L} \tag{9}$$

hence

$$\boldsymbol{P}_L R_{YY} \;=\; \left[0\cdots0\;\; \sigma^2_{\tilde{y},L}\right]. \tag{10}$$

All this holds for $L \geq \underline{L}$. As a function of L, the rank profile of $\sigma^2_{\tilde{y},L}$ behaves like

$$\text{rank}\left(\sigma^2_{\tilde{y},L}\right) \begin{cases} = p & , L \geq \underline{L} \\ = m{-}\underline{m} \in \{p+1,\ldots,m\} & , L = \underline{L}{-}1 \\ = m & , L < \underline{L}{-}1 \end{cases} \tag{11}$$

where $\underline{m} = m\underline{L}{-}(\underline{L}{+}N{-}1) \in \{0,1,\ldots,m-1-p\}$ represents the degree of singularity of $R_{YY,\underline{L}}$.

4.2 LP and inverse of R_{YY}

Note that multichannel linear prediction corresponds to block triangular factorization of (some generalized) inverse of R_{YY}. Indeed,

$$
\begin{array}{c} L_L\, R_{YY}\, L_L^H = D_L\ ,\ (L_L)_{i,j} = \boldsymbol{P}_{i-1,i-j}\ , \\ (D_L)_{i,i} = \sigma^2_{\tilde{y},i-1} \end{array} \quad (12)
$$

where L_L is block lower triangular and D_L is block diagonal. (A slight generalization to the singular case of) the multichannel Levinson algorithm can be used to compute the prediction quantities and hence the triangular factorization above in a fast way. In the case that $R_{YY,\underline{L}}$ is singular, some precaution is necessary in the determination of the last block coefficient $P_{\underline{L},\underline{L}}$, which is not unique (see [8]). Similar singularities will then arise at higher orders.

4.3 Other LP algorithms

Rewriting equation (10) at the correct order as

$$
[-\boldsymbol{Q}_{\underline{L}}|I_m] = \left[\begin{array}{c|c} R_{YY,\underline{L}} & r^H \\ \hline r & r_0 \end{array}\right] = [0\cdots 0|\sigma_y^2] \quad (13)
$$

gives :

$$
\left\{\begin{array}{l} \sigma_y^2 = r_0 - r(R_{YY,\underline{L}})^{-1} r^H \\ \boldsymbol{Q}_{\underline{L}} = r(R_{YY,\underline{L}})^{-1} \end{array}\right. \quad (14)
$$

In this method, the inverse is replaced by a pseudo-inverse in the overestimated case, which gives slightly different results than the singular multichannel Levinson algorithm (this is another choice for the non-unique $\mathbf{P}(z)$ in the case of singular correlation matrix). The drawback of this method is that the pseudo-inverse resorts to computationally intensive SVD.

The main advantage of this method, besides its robustness to order overestimation, is that it allows the use of a correlation of a smoothing window K (i.e. a correlation matrix size $MK \times MK$) bigger than than $\underline{L}$, opposed to the Levinson method. The correct use of the pseudo-inverse, as will be shown hereunder, corresponds to the use of the signal subspace part of it and can be seen as resorting to the correlation matrix cleaned from its noise subspace (the whole algorithm has strong connections with [13]).

To get this averaging effect, but without the cost of the SVD, we propose the following "simplified" method, which, when the order is correctly estimated, relies on :

$$
[0\cdots 0| - \boldsymbol{Q}_{\underline{L}}|I_m] \left[\begin{array}{c|c} *\cdots * & * \\ \ddots & \vdots \\ *\cdots * & * \\ \hline & * \\ R_{\text{rect}} & \vdots \\ \hline r & * \end{array}\right] = [0\cdots 0|\sigma_y^2] \quad (15)
$$

which, solved in a least-squares manner, gives :

$$
\boldsymbol{Q}_{\underline{L}} = (R_{\text{rect}} R_{\text{rect}}^H)^{-1} R_{\text{rect}} r^H \quad (16)
$$

Further investigation should lead to a weighted least-squares solution.

4.4 LP filter as ZF equalizer

Consider the noise-free received signal, which is a singular multivariate MA process, then for $L = \underline{L}$ we have

$$
\boldsymbol{y}(k) + \sum_{i=1}^{\underline{L}} P_{\underline{L},i}\boldsymbol{y}(k-i) = \widetilde{\boldsymbol{y}}_{\underline{L}}(k) = \boldsymbol{h}(0)\,\boldsymbol{a}(k) \quad (17)
$$

so that the prediction error is a singular white noise. This means that the noise-free received signal $\boldsymbol{y}(k)$ is also a singular multivariate AR process. Hence

$$
\boldsymbol{P}_L = \left[\cdots 0\ \ \boldsymbol{P}_{\underline{L}}\right]\ ,\ \ \sigma^2_{\tilde{y},L} = \sigma^2_{\tilde{y},\underline{L}}\ ,\ L > \underline{L}\ . \quad (18)
$$

Hence the factors L_L and D_L in the factorization (12) become block Toeplitz after $\underline{L}$ lines.

For $L = \underline{L}$, $\sigma_y^2 = \sigma_a^2 \boldsymbol{h}^H(0)\boldsymbol{h}(0)$ allows us to find $\boldsymbol{h}(0)$ up to a unitary matrix. We see from (8) and from $\widetilde{\boldsymbol{y}}(k)|_{\boldsymbol{Y}_{L\geq\underline{L}}(k-1)} = \boldsymbol{a}(k)$ that $\frac{\boldsymbol{h}^H(0)}{\boldsymbol{h}^H(0)\boldsymbol{h}(0)}\boldsymbol{P}_{\underline{L}}$ is a zero-delay ZF equalizer. Along with the preceding section, this gives us a ZF equalizer of minimum length.

5 LP and Identification

5.1 Identifiability

The channel can be found from

$$
\begin{array}{rl} & \boldsymbol{P}_{\underline{L}}\,\mathrm{E}\left\{\boldsymbol{Y}_{\underline{L}+1}(k)\boldsymbol{Y}_N^H(k+N-1)\right\} \\ = & \sigma_a^2\,\boldsymbol{h}(0)[\boldsymbol{h}^H(0)\cdots\boldsymbol{h}^H(N-1)] \end{array} \quad (19)
$$

or from $\mathbf{P}_{\underline{L}}(z)\mathbf{H}(z) = \boldsymbol{h}(0) \ \Rightarrow\ \mathbf{H}(z) = \mathbf{P}_{\underline{L}}^{-1}(z)\,\boldsymbol{h}(0)$ using the lattice parameterization for $\mathbf{P}_{\underline{L}}(z)$ obtained with the Levinson algorithm.

Consider for a moment that we do not have channel overestimation problems (and that the singularities are properly handled), then $\mathbf{P}(z)$ is consistently estimated and the fundamental equation is

$$\mathbf{P}(z)\mathbf{H}(z) = \boldsymbol{h}(0) \quad (20)$$

where $\boldsymbol{h}(0)$ is computed from $\sigma_{\tilde{y}}^2 = \sigma_a^2 \boldsymbol{h}^H(0)\boldsymbol{h}(0)$ up to a unitary matrix (say U). Obviously, $\mathbf{H}'(z) = \mathbf{H}(z)U$ fulfills (20), which is a fundamental limitation of the second order methods. Identification of the unitary matrix must be done by resorting to higher order statistics, by finding the innovations of the AR process and applying a source separation to these. Taking into account the whiteness of the sources allows then proper identification of $\boldsymbol{h}(0)$. Some refinements appear when the orders of the channels of the different users are different [4].

5.2 Weighted Linear Prediction

Alternatively, given $\boldsymbol{h}(0)$ and $\boldsymbol{P}_{\underline{L}}$, we can solve for the channel impulse response $\boldsymbol{H}$ from $\mathbf{P}(z)\mathbf{H}(z) = \boldsymbol{h}(0)$, using a weighted least squares procedure [4] :

$$\widehat{\boldsymbol{H}} = \arg\min_{\boldsymbol{H}} ||W^{.5}(\mathcal{T}^{tH}(\boldsymbol{P})\boldsymbol{H} - [\hat{\boldsymbol{h}}(0)^H, 0\cdots 0]^H)||^2 \quad (21)$$

We retain here the "practical" algorithm proposed by the this author, where the weighting matrix is :

$$W = I \otimes (\hat{\sigma}_{\tilde{y}}^2 + \hat{\sigma}_v^2 I_m)^{-1} \quad (22)$$

which is some weighting between the signal innovations subspace, which would be sufficient if the order were known and the noise innovations subspace, which yields some robustness through the regularization of the LS equations system.

6 LP Order overestimation

6.1 "Pseudo-inverse" method

In this method, equation (14) becomes :

$$\begin{cases} \sigma_{\tilde{y}}^2 = r_0 - r(R_{YY,\underline{L}})^{-\#} r^H \\ \boldsymbol{Q}_{\underline{L}} = r(R_{YY,\underline{L}})^{\#} \end{cases} \quad (23)$$

where $^{\#}$ denotes the Moore-Penrose pseudo-inverse.

This pseudo-inverse relies on the correct separation of the signal and noise subspaces of the correlation matrix. When working with the true value of the noise power (or the ML estimate, which can be computed as the mean value of the noise singular values), simulations and some theoretical considerations (e.g. appendix F of [4]) show that incorrect separation of this signals do not worsen the performance of the channel estimation. More precisely, this is due to the following conjecture :

$$\widehat{P}_* = \widehat{P} + \widehat{P}^- + O(\frac{1}{\sqrt{L}}) \quad \widehat{P}^- R_{YY} = 0 \quad (24)$$

Unfortunately, in "practical" situations, simulations do not agree with this. Indeed, when overestimating the channel length ($N' > N$), the noise power is underestimated :

$$\hat{\sigma}_v^2 = \sum_{i=N'+L-1}^{mL} \lambda_i < \sum_{i=N+L-1}^{mL} \lambda_i = \hat{\sigma}_{v,ML}^2 \quad (25)$$

where λ_i are the eigenvalues of R_{YY}, which leads to :

$$\widehat{P}_* = \widehat{P} + \widehat{P}^- + \Delta\hat{\sigma}_v^2 I + O(\frac{1}{\sqrt{L}}) \quad P^- R_{YY} = 0 \quad (26)$$

Hence

$$\hat{\sigma}_{\tilde{y}}^2 = \sigma_a^2 \boldsymbol{h}^H(0)\boldsymbol{h}(0) + \Delta\hat{\sigma}_v^2 r r^H \quad (27)$$

which will introduce an estimation error on $\boldsymbol{h}(0)$. This error is a function of the channel itself via r and can be hardly specified statistically. Moreover, the error on the prediction filter will also affect the channel estimate via the WLS estimator.

6.2 "Levinson" method

In the Levinson method, in a noiseless context, the prediction coefficients of overestimated order become zero. Without going into the details of the algorithm, the calculations rely on the backward and forward prediction error powers (this latter corresponding to $\sigma_{\tilde{y}}^2$), and on some pseudo-inverse and the rank of these powers. To get the correct prediction filter, detection of this rank is necessary, even if the type of pseudo-inverse is of no influence on the prediction error variances' values. When this rank is not correctly estimated, there is no result as in the "pseudo-inverse" method (one can consider that the use of the "Levinson" method leads to a minimum length generalized inverse, while the "pseudo-inverse" method leads to a minimum norm generalized inverse, which has more robustness virtues). Nevertheless, the values of the prediction error variances will give us a good method for determination of the AR order.

7 Order estimation

7.1 Channel order estimation based on λ_i

In a subspace fitting algorithm, it is natural to try to estimate the order by examining the eigenvalues of the correlation matrix. In the Direction of Arrival context, Wax developed a method described in [14] and based on Information Theoretic criterions. Unfortunately, this method does not apply here. He further worked out a general method in [15], but this method resorts to non linear minimization and its complexity does not suit our purposes.

7.2 AR order estimation

Order estimation for vector AR($L < \infty$) processes are usually based either on test statistics of the ideally zero prediction coefficients or the prediction error variances. Another class of estimation procedures rely on Information Theoretic considerations and were initiated by Akaike, Rissanen, Hannan and Quinn (see [9] and references therein). These methods use the prediction error variances. We will use these latter results and adapt them to the singular case.

The classical Infomation Theoretic Criteria try to minimize :

$$\begin{array}{rl} \text{AIC} & \log|\sigma^2_{\tilde{y},k}| + \dfrac{2km^2}{L} \\ \text{HQ} & \log|\sigma^2_{\tilde{y},k}| + \dfrac{2km^2 \log\log L}{L} \\ \text{MDL} & \log|\sigma^2_{\tilde{y},k}| + \dfrac{km^2 \log L}{2L} \end{array} \quad (28)$$

where k are the candidate orders and $|.|$ denotes the determinant.

These expression are based on the maximized log-likelihood of the prediction error, with different bias correction terms based on the number of free parameters (km^2) and the length of the data burst used (L). Rissanen's MDL (Minimum Description Length) and Hannan and Quinn's HQ criteria give strongly consistent estimates for true AR(L) processes and Akaike's AIC criterion has a tendency to overestimate the order. As we will use a subsequent procedure which is robust to order overestimation, we will use the latter in the simulations.

In the singular case, using classic results concerning singular normal multivariate distributions, these criteria extend to our case, but gave poor results, so we propose the following modification.

Remind that $\tilde{y}(n) = \boldsymbol{h}(0)\boldsymbol{a}(n)$ where $R_{aa} = \sigma^2_a I_m$. This mean that we can reason on the equivalent lower dimension $\boldsymbol{h}^H(0)\tilde{\boldsymbol{y}}(n) = \boldsymbol{h}^H(0)\boldsymbol{h}(0)\boldsymbol{a}(n)$ and its prediction error. Since the sources have been considered uncorrelated, we can further consider the uncorrelated prediction error powers, the total prediction pozer being the sum of the single prediction error powers, i.e. the trace of this matrix. Hence, the $|\sigma^2_{\tilde{y},k}|$ by trace($\sigma^2_{\tilde{y},k}$).

More involved and heavier related methods are presented (for the non-singular case) in [5].

8 Overall algorithm

The overall algorithm(s) proposed simply use the AR order selection + inspection of the eigenvalues of the next to last $\sigma^2_{\tilde{y}}$ to get a channel length estimate. Once we have these order estimates, we proceed with the various classical WLP algorithms where the prediction coefficients are estimated according to the various methods described. The ultimate choice will depend on the price we are willing to pay for estimation accuracy.

9 Simulations

9.1 Channel estimation

In order to characterize the robustness of the WLP to channel length overestimation, we made various simulations illustrating the different effects of channel lenght estimation errors and of the overall algorithm.

The performance measure is the Normalized Root MSE (NRMSE) which is computed over 100 Monte Carlo runs as

$$\text{NRMSE} = \sqrt{\frac{1}{100}\sum_{i=1}^{100} \boldsymbol{h}^H P^{-}_{\widehat{\boldsymbol{h}}^{(i)}} \boldsymbol{h}/||\boldsymbol{h}||^2}$$

where $\boldsymbol{h}^H P^{-}_{\widehat{\boldsymbol{h}}} \boldsymbol{h} = \min_\alpha ||\alpha\widehat{\boldsymbol{h}} - \boldsymbol{h}||^2$. We use the real channel $\boldsymbol{h}$ with $N = 6$, $m = 4$ and $p = 1$ which was used in [2].

The symbols are i.i.d. QPSK, and the data length is $L = 250$. The SNR is defined as $(||\boldsymbol{h}||^2\sigma^2_a)/(mM\sigma^2_v)$.

The eigenvalues profile of the correlation matrix is reproduced hereunder, which gives an idea of how easy it should be to determine order at the different SNR's.

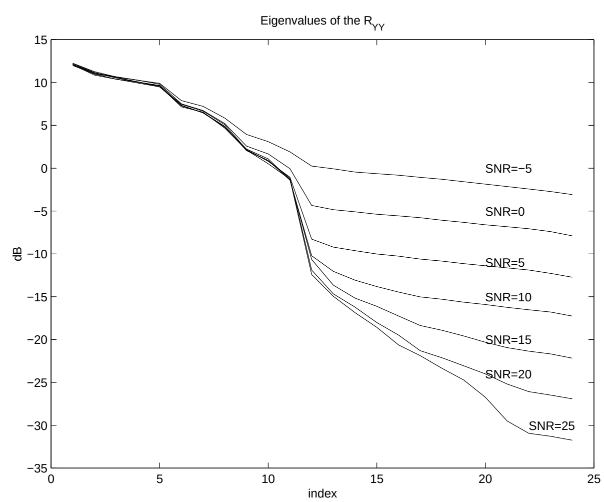

Figure 1: Eigenvalues of R_{YY}.

Use of $\hat{\sigma}^2_{v,ML}$ The simulations agree with the conjecture that, using the ML estimate of the noise power, there is no loss of performance due to overestimation of the channel length. The smoothing window is $K = 6$ (i.e. R_{YY} of size $mK \times mK$).

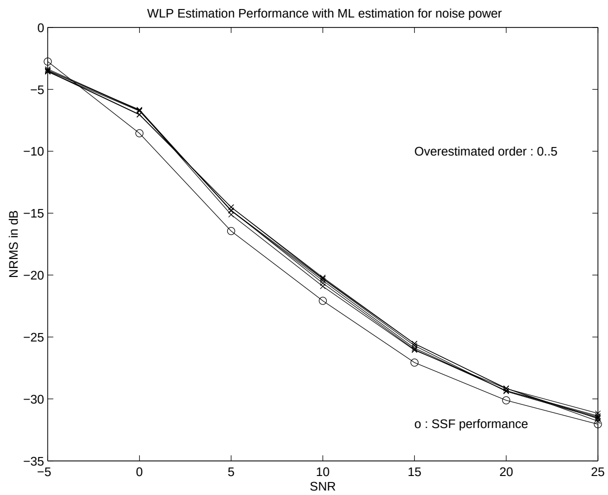

Figure 2: Performance WLP using $\hat{\sigma}^2_{v,ML}$

Use of $\hat{\sigma}^2_v$ Here, the simulations show clearly the influence of $\Delta\hat{\sigma}^2_v$, mostly at high SNR, where order estimation is the easiest to perform. We first show the results with the minimum smoothing window and then for a smoothing window $K = \hat{N}$. Comparison of these clearly favor non-minimum smoothing window.

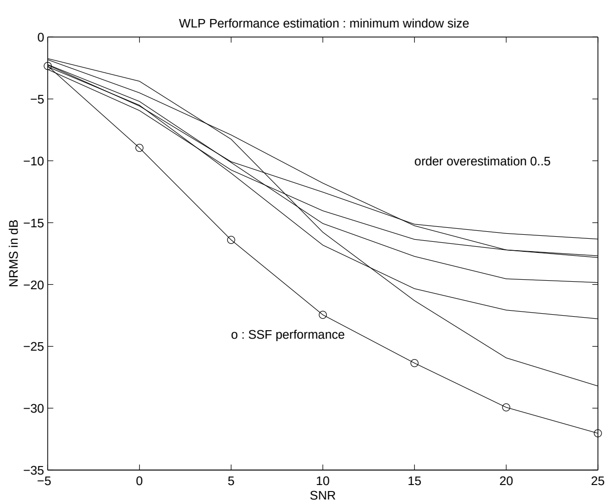

Figure 3: WLP without order estimation or knowledge, minimum K

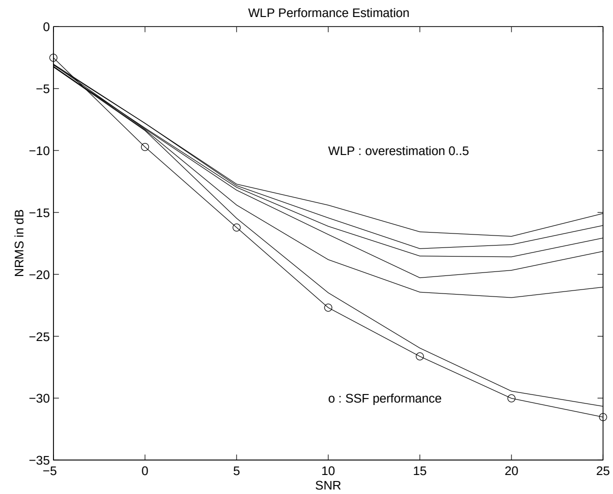

Figure 4: WLP without order estimation or knowledge, $K = \hat{N}$

Use of the prior order estimation This final figure compares the different practical WLP algorithms, namely the WLP without order estimation, WLP with AR order estimation, WLP with channel order estimation followed by a 'pseudo-inverse' LP modeling method or a 'Levinson' LP modeling method. In this latter case, we use a two-step procedure where the first step consists of the channel length estimation and the second step is again LP modeling by Levinson, based on the covariance matrix of minimum size ($R_{YY,\underline{L}}$) and $\hat{\sigma}^2_v = \lambda_{min}(R_{YY,\underline{L}})$ The results agree with what we expected, namely the channel order estimation greatly improve the performance and the 'pseudo-inverse'LP method is far better then the Levinson method on the raw correlation matrix.

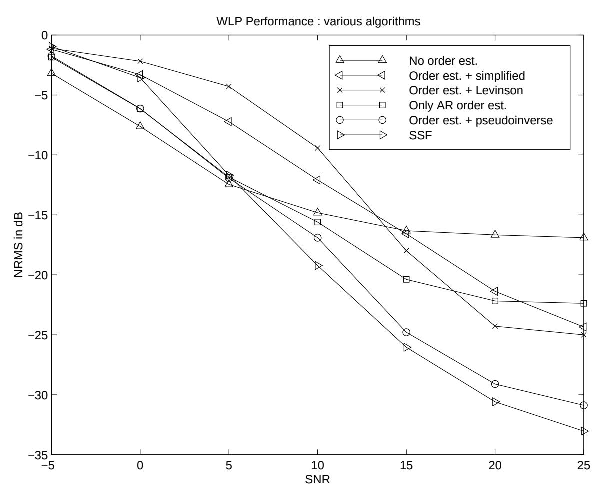

Figure 5: WLP with order estimation

9.2 LP order estimation

Hereunder, we reproduce the tests for a channel with $m = 4$ sub-channels, one user and length $N = 6$, leading to an AR of true order 2. The results are rather convincing for relatively low SNR. The noise power is computed under the assumption that we have an AR(6) process.

AIC	SNR						
ord.	25	20	15	10	5	0	-5
1	0	0	0	0	0	1	69
2	99	99	82	11	2	23	15
3	0	0	18	87	86	61	3
4	0	0	0	0	1	2	0
5	0	0	0	0	1	1	2
6	1	1	0	2	10	12	11

MDL	SNR						
ord.	25	20	15	10	5	0	-5
1	0	0	0	0	0	2	99
2	100	100	100	56	33	79	1
3	0	0	0	44	67	19	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0

HQ	SNR						
ord.	25	20	15	10	5	0	-5
1	0	0	0	0	0	17	100
2	100	100	100	86	78	81	0
3	0	0	0	14	22	2	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0

These are the results of the tests for a channel with $m = 4$ sub-channels, 2 users and length $N = 12$, leading to an AR of true order 5. The results are rather convincing for high to moderate SNR. The noise power is computed under the assumption that we have an AR(10) process.

AIC	SNR						
ord.	25	20	15	10	5	0	-5
1	0	0	0	0	0	70	100
2	0	0	0	0	0	1	0
3	0	0	0	0	1	9	0
4	0	0	0	0	57	20	0
5	87	97	100	100	39	0	0
6	13	3	0	0	0	0	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0
10	0	0	0	0	3	0	0

MDL	SNR						
ord.	25	20	15	10	5	0	-5
1	0	0	0	0	30	100	100
2	0	0	0	0	0	0	0
3	0	0	0	0	18	0	0
4	0	0	0	3	49	0	0
5	100	100	100	97	3	0	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0

HQ	SNR						
ord.	25	20	15	10	5	0	-5
1	0	0	0	3	92	100	100
2	0	0	0	0	0	0	0
3	0	0	0	0	6	0	0
4	0	0	0	17	2	0	0
5	100	100	100	72	0	0	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0

10 Conclusions

We have investigated Linear Prediction robustness characteristics, comparing various methods yielding the prediction coefficients of a singular AR($L < \infty$) process. The "pseudo-inverse" method is robust, but its performance critically relies on the noise power estimation and on the use of a big enough smoothing window. The "Levinson" method needs to perform order estimation to prove robust. This led to the development of performant order estimation algorithms at almost no cost, and yielding the prediction coefficients as a by-product. Although computationally far less demanding than the first method, it does not use a non-minimum smoothing window, yielding poorer channel estimates. In order to conjugate low computational complexity and better performance, we propose a solution where a bigger smoothing window is used but without the need of an SVD, this method should be further refined with the use of a weighting matrix and thus attain comparable performance to the best method.

References

[1] *"Special Issue on Signal Processing for Advanced Communications"*, volume 45. IEEE Transactions on Signal Processing, January 1997.

[2] K. Abed-Meraim, J.-F. Cardoso, A. Y. Gorokhov, P. Loubaton, and E. Moulines. "On Subspace Methods for Blind Identification of Single-Input Multiple-Output FIR Systems". *IEEE Transactions on Signal Processing*, 45(1):42–55, January 1997.

[3] Karim Abed-Meraim, Wanzhi Qiu, and Yingbo Hua. "Blind System Identification". *Proceedings of the IEEE*, 85(8):1310–1322, August 1997.

[4] Alexei Gorokhov. *"Séparation autodidacte des mélanges convolutifs: méthodes du second ordre"*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, 1997.

[5] Changhua Chen, Richard A. Davis, and Peter J. Brockwell. "Order Determination for Multivariate Autoregressive Processes Using Resampling Methods". *Journal of Multivariate Analysis*, 57(2):175–190, May 1996.

[6] Steven D. Halford and Georgios B. Giannakis. "Channel order Determination based on Sample Cyclic Correlations". In *Twenty-Eight Asilomar Conference on Signal, Systems & Computers*, October 1994.

[7] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue. "Subspace Methods for the Blind Identification of Multichannel FIR filters". *IEEE Transactions on Signal Processing*, 43(2):516–525, February 1995.

[8] Boaz Porat. *"Contributions to the Theory and Applications of Lattice Filters"*. PhD thesis, Stanford University, August 1982.

[9] Gregory C. Reinsel. *"Elements of Multivariate Time Series Analysis"*. Springer Series in Statistics. Springer-Verlag, 1993.

[10] Dirk T.M. Slock. "Blind Joint Equalization of Multiple Synchronous Mobile Users Using Oversampling and/or Multiple Antennas". In *Twenty-Eight Asilomar Conference on Signal, Systems & Computers*, October 1994.

[11] D.T.M. Slock. "Blind Fractionally-Spaced Equalization, Perfect-Reconstruction Filter Banks and Multichannel Linear Prediction". In *Proc ICASSP 94 Conf.*, Adelaide, Australia, April 1994.

[12] A. Touzni and I. Fijalkow. "Robustness of Blind Fractionally-Spaced Identification/Equalizqtion to Loss of Channel Disparity". In *Proc. ICASSP*, volume 5, pages 3937–3940, Munich, Germany, April 1997.

[13] D.W. Tufts and R. Kumaresan. "Estimation of frequencies of multiple sinusoids : making linear prediction perform like maximum likelihood". *Proceedings of the IEEE*, 70:975–989, 1982.

[14] M. Wax and T. Kailath. "Detection of signals by information theoretic criteria". *IEEE Trans. Acoust., Speech, Signal Processing*, 33:387–392, 1985.

[15] M. Wax and I. Ziskind. "Detection of the number of coherent signals by the MDL principle". *IEEE Trans. Acoust., Speech, Signal Processing*, 37:1190–1196, 1989.