A SCHUR METHOD FOR MULTIUSER MULTICHANNEL BLIND IDENTIFICATION

Luc Deneire and *Dirk T.M. Slock*

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

ABSTRACT

We address the problem of blind multiuser multichannel identification in a Spatial Division Multiple Access (S.D.M.A.) context. Using a stochastic model for the input symbols and only second order statistics, we develop a simple algorithm, based on the Generalized Schur algorithm to apply LDU decomposition of the covariance matrix of the received data. We show that this method leads to identification of the channel, up to a unitary mixture matrix. Furthermore, the identification algorithm is shown to be robust to channel length overestimation and approaches the performance of the Weighted Linear Prediction (WLP) method [1], at low computational cost.

1. INTRODUCTION

Blind multichannel identification has received considerable interest over the last decade. In particular, second-order methods have raised a lot of attention, due to their ability to perform channel identification with relatively short data bursts. Among these methods, we can distinguish the deterministic methods, where the input symbols are considered deterministic and the stochastic methods, where the input symbols are considered stochastic. Using the deterministic model leads to a dynamical indeterminacy [3, 5] as opposed to the stochastic model which leads to the identification of the channel up to a unitary static mixture matrix [3]. Subsequent source separation can then be performed by other classical methods or by resorting to known symbols (i.e. performing semi-blind identification).

We show that LDU decomposition of the covariance matrix leads to the identification of the channel (up to a unitary mixing matrix) and that performing this decomposition with a Schur algorithm yields good performance. Moreover, this identification procedure is inherently robust to channel length overestimation, which makes our method a serious candidate for bootstrap blind multiuser multichannel identification.

2. DATA MODEL AND NOTATIONS

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

$$y_i(t) \;=\; \sum_{j=1}^{p}\sum_{k} a_j(k)h_{ij}(t-kT)+v_i(t),\; i=1,\cdots,m \tag{1}$$

where the $a_j(k)$ are the transmitted symbols from source j, T is the common symbol period, $h_{ij}(t)$ 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.

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. The discrete-time received signal can be represented in vector form as

$$\boldsymbol{y}(k)=\boldsymbol{H}\boldsymbol{A}_N(k)+\boldsymbol{v}(k);\boldsymbol{H}=[\boldsymbol{H}_1\cdots\boldsymbol{H}_p] \tag{2}$$

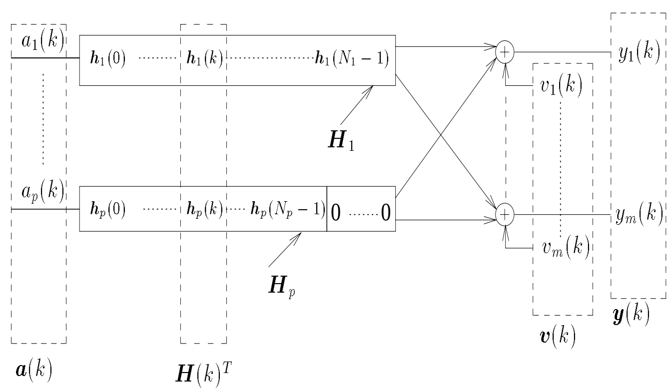


Figure 1: Notations.

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 :

$$\boldsymbol{Y}_M(k)=\mathcal{T}_M\left(\boldsymbol{H}\right)\,\boldsymbol{A}_{N+p(M-1)}(k)+\boldsymbol{V}_M(k) \tag{3}$$

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 $\mathcal{T}_M(\boldsymbol{H})$ is the multichannel multiuser

The work of Luc Deneire is supported by the EC by a Marie-Curie Fellowship (TMR program) under contract No ERBFMBICT950155 and the work of Dirk Slock by Laboratoires d'Electronique PHILIPS under contract LEP 95FAR008 and EURECOM Institute.

convolution matrix of $\boldsymbol{H}$, with M block lines ($\mathcal{T}_M(\boldsymbol{H}) = [\mathcal{T}_M(\boldsymbol{H}_1)\ldots\mathcal{T}_M(\boldsymbol{H}_p)]$, where $\mathcal{T}_M(\boldsymbol{H}_j)$ is block Toeplitz). The input symbols are also i.i.d., zero mean and independent from one user to another, so we can write $R_{AA} = \sigma_a^2 I$.

3. IDENTIFIABILITY

Effective Number of Channels

In the subsequent developments, we will often consider irreducible and column reduced channels, (i.e. such that $\mathbf{h}(z) = [\mathbf{h}_1(z)\cdots\mathbf{h}_p(z)]$ is full-rank $\forall z$ and $[\boldsymbol{h}_1(N_1-1)\ldots\boldsymbol{h}_p(N_p-1)]$ is full-rank), which is equivalent to forcing the tall (i.e. more lines than columns) matrix $\mathcal{T}(\boldsymbol{H})$ to be full column rank. Obviously, if $\boldsymbol{H}_N$ is not full rank, $\mathcal{T}(\boldsymbol{H})$ can not be full column rank under the same matrix size conditions and one must consider a reduced number of channels (equal to the rank of $\boldsymbol{H}_N$), which we will call effective number of channels. From here on, irreducible channels will mean irreducible channels with m being the effective number of channels.

One must identify $\theta = [\boldsymbol{H}\ \sigma_v^2]$ from

$$R_{YY}(\theta) = \mathcal{T}(\boldsymbol{H})\mathcal{T}^H(\boldsymbol{H})\sigma_a^2 + \sigma_v^2 I \qquad (4)$$

Sufficient condition *[3] In the stochastic model, the m-channel $\boldsymbol{H}$ is identifiable blindly up to a unitary static mixture factor by second-order statistics if*

(i) The channel is irreducible and column reduced.

(ii) $M \geq \underline{L}+1$

where $\underline{L} = \left\lceil \frac{N-p}{m-p} \right\rceil$ ($= 0$ *for* $\frac{0}{0}$)

4. LDU FACTORIZATION OF A COVARIANCE MATRIX

From (4), under the identifiability conditions, we can identify σ_v^2 as the singular vector corresponding to the minimum singular value of R_{YY}. Let $\widetilde{\boldsymbol{Y}}$ be the prediction error, then, as $\boldsymbol{Y}$ can be perfectly predicted in the absence of noise, the covariance of the error can be written as

$$R_{YY} - R_{Y\tilde{Y}}R^{\#}_{\tilde{Y}\tilde{Y}}R_{\tilde{Y}Y} = 0 \Rightarrow R_{Y\tilde{Y}}R^{\#}_{\tilde{Y}\tilde{Y}}R_{\tilde{Y}Y} = U^H D U \qquad (5)$$

where $^{\#}$ denotes a generalized inverse.

Consider we perform a block triangularization, examination of the rank of $R_{YY} - \sigma_v^2 I = \mathcal{T}(\boldsymbol{H})\mathcal{T}^H(\boldsymbol{H})\sigma_a^2$ leads to, for block i

$$\text{rank}\,(D_i) = \begin{cases} = p & , i \geq \underline{L} \\ = m - \underline{m} \in \{p+1,\ldots,m\} & , i = \underline{L}-1 \\ = m & , i < \underline{L}-1 \end{cases} \qquad (6)$$

where $\underline{m} = (m-p)\underline{L} - N - p \in \{0,1,\ldots,m-1-p\}$.

As the prediction of $\boldsymbol{Y} = \boldsymbol{Y}_M(k)$ is perfect from instant $\underline{L}+1$ on, in $t > \underline{L}$, $\tilde{\boldsymbol{y}}(k-t)$ contains the emitted symbols, apart from a unitary matrix, which is consistent with the rank profile of D_i.

Furthermore, denoting $U(i,j)$ as the (i,j) block of U, $U^H = R_{Y\tilde{Y}}$ implies :

$$\begin{aligned} U^H(i,j) &= \mathrm{E}\left\{\boldsymbol{y}(k-i)\tilde{\boldsymbol{y}}^H(k-j)\right\} \\ &= \sum_{l=0}^{N_1-1} \boldsymbol{H}(l)\mathrm{E}\left\{\boldsymbol{a}(k-i-l)T\ \boldsymbol{a}^H(k-j)\right\} \\ &\quad \text{for } i,j > \underline{L} \\ &= T\ \boldsymbol{H}(j-i)\sigma_a^2 \end{aligned} \qquad (7)$$

where T is a unitary matrix. Hence, we can identify the channel, up to a unitary matrix, by triangularization of the covariance matrix of the received signal.

5. USE OF THE GENERALIZED SCHUR ALGORITHM

5.1. Some basics

The displacement of a $n \times n$ Hermitian matrix is defined as $\nabla R \triangleq R - Z_\nu R Z_\nu^H$, where Z_ν is a $n \times n$ lower shift matrix with ones on the ν^{th} subdiagonal [1]. The rank r of ∇R is called the displacement rank and can be shown to be equal to $2m$ for the covariance matrix $R_{YY}(-\sigma_v^2 I)$. Moreover, we can factor ∇R_{YY} as $\nabla R_{YY} = G\Sigma G^H$ where $\Sigma = (I_m \oplus -I_m)$ is called the signature matrix and G the generator of R_{YY}. One can easily check that, denoting the blocs $r_i \triangleq R_{YY}^{-1/2}(0)R_{YY}(i)$, R_{YY} of size $Km \times Km$ and $\mathcal{R}_{YY} = [r_{i-j}]_{i,j}$.

$$\nabla\mathcal{R}_{YY} = \begin{bmatrix} r_0 & 0 \\ r_1 & r_1 \\ \vdots & \vdots \\ r_{K-1} & r_{K-1} \end{bmatrix} \begin{bmatrix} I_m & 0 \\ 0 & -I_m \end{bmatrix} \begin{bmatrix} r_0 & 0 \\ r_1 & r_1 \\ \vdots & \vdots \\ r_{K-1} & r_{K-1} \end{bmatrix}^H$$

Proceeding by block, the generalized Schur algorithms starts with the generator $G^{(0)} = G$, forms

$$\begin{aligned} G^{(1)\,H} &= S^{(1)} \begin{bmatrix} 0 & r_0^{(0)} & r_1^{(0)} & \cdots & r_{K-1}^{(0)} \\ 0 & r_1^{(0)} & r_2^{(0)} & \cdots & r_K^{(0)} \end{bmatrix} \\ &= \begin{bmatrix} 0 & r_0^{(1)} & r_1^{(1)} & \cdots & r_{K-1}^{(1)} \\ 0 & 0 & \tilde{r}_2^{(1)} & \cdots & \tilde{r}_K^{(1)} \end{bmatrix} \end{aligned} \qquad (8)$$

where $S^{(1)}$ is a block hyperbolic Householder transformation (such that it is Σ unitary : i.e. $S^{(1)}\Sigma S^{(1)\,H} = \Sigma$). Then $G^{(1)}$ is the generator of the Schur complement of $\mathcal{R}_{YY}$ with respect to r_0. Continuing this process further, we get $\mathcal{R}_{YY} = U^H D U$ where

$$U = \begin{bmatrix} r_0^{(0)} & r_1^{(0)} & \cdots & r_{K-1}^{(0)} \\ 0 & r_0^{(1)} & \ddots & r_{K-2}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_0^{(K-1)} \end{bmatrix}$$

[1] broader definitions can be found in [6, 2].

5.2. Applying it to LDU factorization of $\widehat{R}_{YY}$

After the $\underline{L}^{\text{th}}$ iteration, the p first columns of the generator contains the channel, the subsequent $m-p$ columns being zero.

The generalized Schur algorithm can only work with strongly nonsingular and definite positive matrices (i.e. whose principal minors are all definite positive). Solutions have been proposed for singular matrices [4, 7] but introduce some additional complexity, which we want to avoid, hence, the choice of the estimator of the correlation matrix is critical. Here, we choose the "biased estimator" of the correlation sequence

$$\widehat{R}_{YY}(i) = \frac{1}{M}\sum_{t=1}^{M} \boldsymbol{y}(t)\boldsymbol{y}^H(t+i),$$

because the matrix formed with these estimators is, by construction, block Toeplitz and definite positive. The "unbiased estimator" (where the scaling factor is $1/(M-i)$) leads to an indefinite matrix and the "sample covariance matrix" ($\widehat{R}_{YY} = \frac{1}{M}\sum_{t=0}^{M-1} \boldsymbol{Y}(t)\boldsymbol{Y}^H(t)$, where $\boldsymbol{y}(k)=0$ for $k => M$) has a displacement rank of $2(m+1)$, having thus a slightly different structure as the true covariance matrix.

Furthermore, due to the fact that it is a biased estimator, the singular value spectrum does not present an abrupt breakdown at the theoretic rank of R_{YY}, but presents a relatively "smooth" decay of the singular values. This will lead to late occurence of bad conditioning in the Schur algorithm and prevent us from using the methods necessary in the case of singular matrices. Besides this numerical advantage, we will also be able, provided the size of $\widehat{R}_{YY}$ is big enough (2 to 3 ($\underline{L}m \times \underline{L}m$)), to overestimate the channel length without numerical problems, as we will be able to perform the Schur iteration further than the $\underline{L}^{\text{th}}$ iteration.

The drawback of this behaviour is that, after the $\underline{L}^{\text{th}}$ iteration, though singular in the exact case, the $(\underline{L}+1)^{\text{th}}$ block of the block diagonal is far from singular and, hence, the p first columns of the generator can not pretend to contain the channel exactly. A simple, though possibly expensive ($\mathcal{O}(pm^2mK)$) way, is then to perform an SVD on the m first columns of the generator and take the p first columns, which correspond to the non-singular part of $D_{\underline{L}+1}$.

SCHUR ESTIMATION PROCEDURE

1. Calculate the estimate of R_{YY}

$$\widehat{R}_{YY}(i) = \frac{1}{M}\sum_{t=1}^{M} \boldsymbol{y}(t)\boldsymbol{y}^H(t+i).$$

   then $\widehat{R}_{YY} - \lambda_{min}(\widehat{R}_{YY})I$

2. Calculate $\widehat{R}_{YY}^{1/2}$.
3. Proceed with the Schur iterations as in (8) until the $\underline{L}^{\text{th}}$ iteration.
4. Either
   - collect the p first columns of the generator.
   - calculate the p first left singular vectors of the generator.

6. COMPLEXITY

In this section, we evaluate the order of magnitude of complexity of the Schur algorithm, apart from the calculation of $\widehat{R}_{YY}$, which is common to all algorithms.

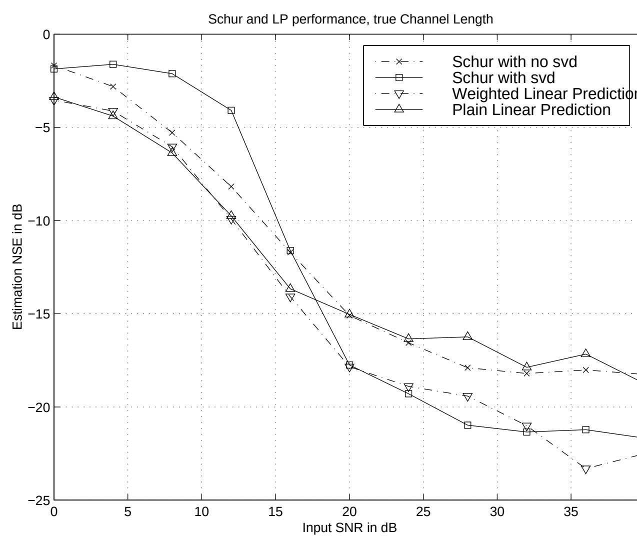
Evaluating the complexity for each step as : $\lambda_{min}(\widehat{R}_{YY})$: $\mathcal{O}(Km)$; $\widehat{R}_{YY}^{1/2}$: $\mathcal{O}(m^2)$; multiply the first block column : $\mathcal{O}(Km^3)$; $\underline{L}$ Schur iterations [2] : $\mathcal{O}(\underline{L}m(2m)Km)$; optional SVD : $\mathcal{O}(Km^3)$. In total, this gives a complexity which is linear in the size of the covariance matrix ($\mathcal{O}(2(\underline{L}+1)Kmm^2)$).

If we compare to the PLP-WLP, where the complexity is dominated by a Pseudo-inverse ($\mathcal{O}((Km)^3)$) and the solution of a least-square problem involving structured matrices ($\mathcal{O}(4p(Km)^2)$), we get a far less complex algorithm with comparable performances.

7. SIMULATIONS

In order to evaluate the performance of the algorithms, we have computed the Normalized MSE (NMSE) on the estimated channels, averaged over 200 Monte Carlo runs. We have used a randomly generated channel with $p=2$ users, $N_1=3$ and $N_2=4$, and $m=4$ subchannels. The symbols are i.i.d. BPSK and the data length is $M=250$. The mixing matrix has been estimated afterwards as $T = \frac{\boldsymbol{H}_t\boldsymbol{H}_t^H}{\widehat{\boldsymbol{H}}_t\boldsymbol{H}_t^H}$ for the Schur method (where $\boldsymbol{H}_t \triangleq [\boldsymbol{H}_{1t}^H \cdots \boldsymbol{H}_{pt}^H]^H$ and $\boldsymbol{H}_{it} \triangleq [\boldsymbol{h}_i^T(j)]_{j=1,N_p}$) and similarly on $\boldsymbol{h}(0)$ for the WLP-PLP methods.

We evaluate the performance of the Schur algorithm, with and without SVD performed on the generator and compare them to the Weighted and Plain Linear Prediction algorithm [1].

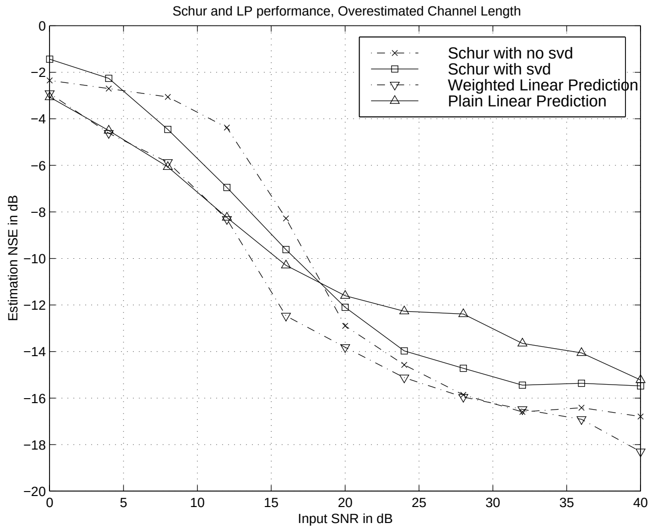


Comparison between Schur and WLP algorithms true channel length.

Curves show that the Schur algorithm, although very simple, gives comparable performance as the PLP and WLP method, this last one was shown to be only slightly sub-optimal.

[2]for one iteration the complexity is $\mathcal{O}(rmKm)$ where r is the displacement rank

We further explore the performance when the channel length has been overestimated ($\widehat{N}_1 = \widehat{N}_2 = 6$),

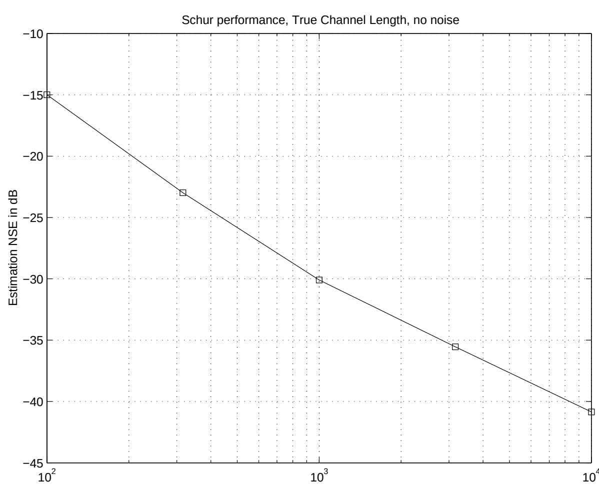


Comparison between Schur and WLP algorithms overestimated channel length.

Again, the performances are rather close and the robustness of the Schur method is as good as the robustness of the Linear Prediction Method.

In both figures, it can be seen that performing a SVD on the generator gives (as expected) better results at high SNR, in which case we attain the WLP performance. At low SNR, the performance is degraded.

Figures show also that the performances of the Schur method are "smoother" than those of the WLP method, indeed, simulations show that this latter algorithm leads to NMSE's that have larger confidence intervals than the first one.



Performance of Schur algorithm with respect to the number of input symbols

It is worth mentioning that, as we have a stochastic point of view, we have a flooring effect at high SNR, due to the assumption that $R_{AA} = \sigma_a^2 I$. The figure here above shows the performance increasing with the number of input symbols.

8. CONCLUSIONS

We have introduced a Schur method to identify multiuser multichannels blindly in a computationaly efficient way. This algorithm is recursive in order, and can be coupled to a channel length estimator and source detector by examining the diagonal of the LDU of $R_{YY} - \sigma_v^2 I$. Even if this channel length is overestimated, the algorithm provides a good estimate of the channel. Furthermore, an adaptive version could be developed by using standard signal processing techniques.

The performances of this simple algorithm are shown to be close to the Weighted Linear Prediction method, which is near optimal [1] and presents the same robustness characteristics.

Moreover, the computational complexity of the Schur algorithm is linear in the size of the covariance matrix, opposed to the cubic complexity of the Linear Prediction methods, which renders this method very attractive for bootstrap blind multiuser multichannel identification.

9. REFERENCES

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

[2] P. COMON. Structured matrices and inverses. In A. Bojanczyk and G. Cybenko, editors, *Linear Algebra for Signal Processing*, volume 69 of *IMA Volumes in Mathematics and its Applications*, pages 1–16. Springer Verlag, 1995.

[3] Luc Deneire and Dirk Slock. "Identifiability conditions for blind and semi-blind multiuser multichannel identification". In *9th IEEE Signal Processing Workshop On Statistical Signal And Array Processing*, Portland, Oregon, USA, September 1998.

[4] K.A Gallivan, S. Thirumalai, P. Van Dooren, and V. Vermaut. "High Performance Algorithms for Toeplitz and block Toeplitz matrices". *Linear Algebra and its Applications*, 1994.

[5] A. Gorokhov and Ph. Loubaton. "Subspace based techniques for second order blind separation of convolutive mixtures with temporally correlated sources". *IEEE Trans. on Circuits and Systems*, July 1997.

[6] J. Chun. *"Fast Array Algorithms for Structured Matrices"*. PhD thesis, Information Systems Laboratory, Stanford University, June 1989.

[7] Ali H. Sayed and Thomas Kailath. "A Look-ahead Block Schur Algorithm for Toeplitz-Like Matrices". *SIAM J. Matrix Anal. Appl.*, 16(2), April 1995.