Symmetric Minimum Noise Subspace for Multi-Input Multi-Output System Identification

Anahid SAFAVI, Karim ABED-MERAIM

Dept. TSI, ENST (Telecom Paris)
46 Rue Barrault, Paris 75013, France
safavi,abed@tsi.enst.fr

Résumé – Dans cet article nous considérons la classe des méthodes d'identification aveugle multicapteurs du sous espace minimum (MNS). L'algorithme MNS est une version rapide de l'algorithme sous-espace. Nous développons ici une version dite symétrique de la méthode MNS (que l'on notera SMNS) qui a l'avantage d'être plus robuste et plus performante que la méthode MNS originale au prix d'une légère augmentation du coût de calcul. Nous proposons aussi, une implémentation en bloc efficace et rapide de l'algorithme SMNS. Finalement, nous montrons que, sous certaines hypothèses supplémentaires sur la fonction de transfert du système, il est possible d'aller au-delà du minimum dans le sens où moins de $q-p$ vecteurs bruit (q étant le nombre de sorties et p le nombre d'entrées) seraient suffisants pour l'identification du canal.

Abstract – This contribution deals with a particular family of blind system identification methods, referred to as Minimum Noise Subspace (MNS). MNS method is a computationally fast version of subspace method. Here, we develop a symmetric version of MNS method (referred to as SMNS) that has the advantage of better robustness and estimation accuracy at the cost of a slight increase of the computational cost in comparison with original MNS. In the same time, we present and compare different algorithms for the block implementation of the SMNS. Finally, we show that under certain additional assumptions on the channel transfer functions, we can go beyond the minimum in the sense that less than $q-p$ (q being the number of outputs and p the number of inputs) noise vectors are sufficient for unique identification of channel parameters.

1 Introduction

Blind system identification is important in many fields of applications and particularly in mobile communication systems. Recently, the world has seen a great research interest in blind identification of single-input multi-output (SIMO) and multi-input multi-output (MIMO) systems using second order statistics (SOS) based methods [1]. The SOS based approach is attractive because it requires much less samples than the traditional higher order statistic (HOS) based approach.

Among the existing SOS based methods, the subspace (SS) method shown in [2, 6] appears to be one of the most robust to noise. One of the important advantages of SS method is its deterministic property. That is, the channel parameters can be recovered perfectly in absence of noise, using only a finite set of data samples, without any statistical assumption over input data. Therefore, SS method is promising for applications where only a few number of output data are available, or the input data is arbitrary.

A major drawback of the SS method is its high computational cost due to the eigendecomposition of a large dimensional matrix. Recently, a computationally attractive subspace method called Minimum Noise Subspace (MNS) has been proposed for MIMO system identification [5]. This method computes the noise subspace via a set of noise vectors which are computed in parallel from a set of combinations of system outputs that form a basis of the rational noise subspace.

In this paper, we introduce the concept of Symmetric MNS (SMNS) method, we study the effective implementation of the MNS and SMNS and we compare their respective estimation performances.

2 The MNS method: a review

Let $\mathbf{y}(n)$ be a q-variate discrete-time stationary time-series given by

$$\mathbf{y}(n)=\sum_{k=0}^{M}\mathbf{H}(k)\mathbf{s}(n-k)+\mathbf{w}(n) \tag{1}$$

where $\mathbf{H}(z)\stackrel{\text{def}}{=}\sum_{k=0}^{M}\mathbf{H}(k)z^{-k}$ is an unknown causal FIR $q\times p$ transfer function satisfying $\text{Rank}(\mathbf{H}(z))=p$ for each z and $\text{Rank}(\mathbf{H}(M))=p$ ($\mathbf{H}(z)$ is column reduced [3]), $\mathbf{s}(n)$ is a p-dimensional unknown process and $\mathbf{w}(n)$ is an additive q-dimensional (spatially and temporally) white noise, i.e. $E(\mathbf{w}(n)\mathbf{w}^*(l))=\delta(n-l)\sigma^2\mathbf{I}_q$.

Consider the spatio-temporal variables:

$$\begin{aligned}\mathbf{y}_N^{(i)} &= \mathcal{T}_N(\mathbf{H}_{i,:})\mathbf{s}_N+\mathbf{w}_N^{(i)},\\ \mathbf{y}_N &= \mathcal{T}_N(\mathbf{H})\mathbf{s}_N+\mathbf{w}_N\end{aligned} \tag{2}$$

where

$$\begin{aligned}\mathbf{y}_N^{(i)} &= [y_i(n),\cdots,y_i(n+N-1)]^T,\\ \mathbf{y}_N &= [\mathbf{y}_N^{(1)T},\cdots,\mathbf{y}_N^{(q)T}]^T,\\ \mathbf{s}_N &= [\mathbf{s}_N^{(1)}(n),\cdots,\mathbf{s}_N^{(p)}(n)]^T,\\ \mathbf{s}_N^{(i)}(n) &= [s_i(n),\cdots,s_i(n+N+M_i-1)]^T,\\ \mathcal{T}_N(\mathbf{H}_{i,:}) &= [\mathcal{T}_N(\mathbf{H}_{i,1}),\cdots,\mathcal{T}_N(\mathbf{H}_{i,p})],\\ \mathcal{T}_N(\mathbf{H}) &= [\mathcal{T}_N(\mathbf{H}_{1,:})^T,\cdots,\mathcal{T}_N(\mathbf{H}_{q,:})^T]^T,\end{aligned}$$

$\mathcal{T}_N(\mathbf{H}_{i,j})$ being the $N\times(N+M)$ Sylvester matrix associated with the scalar polynomial $\mathbf{H}_{i,j}(z)$ (see [2]). Let $\mathcal{R}_N$ be the

covariance matrix of $\mathbf{y}_N(n)$:

$$\begin{aligned}\mathcal{R}_N &\stackrel{\text{def}}{=} E(\mathbf{y}_N(n)\mathbf{y}_N^*(n)) \\ &= \mathcal{T}_N(\mathbf{H})\mathcal{S}\mathcal{T}_N^*(\mathbf{H}) + \sigma^2\mathbf{I}_{qN} \end{aligned} \quad (3)$$

where $\mathcal{S} \stackrel{\text{def}}{=} E(\mathbf{s}_N(n)\mathbf{s}_N^*(n))$ is assumed to be positive definite. For $N > pM$, the noise-free covariance matrix is rank deficient and its eigen-decomposition is given by:

$$\mathcal{R}_N = \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^* + \sigma^2\mathbf{U}_n\mathbf{U}_n^* \quad (4)$$

where $\text{Range}(\mathbf{U}_s) = \text{Range}(\mathcal{T}_N(\mathbf{H}))$ is the signal subspace (as $\mathbf{\Lambda}_s$ is the diagonal matrix of the $p(N+M)$ largest eigenvalues of $\mathcal{R}_N$) and $\text{Range}(\mathbf{U}_n) = \text{Range}(\mathcal{T}_N(\mathbf{H}))^\perp$ is the noise subspace. The following orthogonality relation

$$\mathbf{U}_n^*\mathcal{T}_N(\mathbf{H}) = 0 \quad (5)$$

is the keystone of the SS method to identify the transfer function $\mathbf{H}(z)$ up to a $p \times p$ constant matrix [6].

The SS method is expensive in computational cost due to the above subspace decomposition. A major contribution in [5] is to show that first, only $q - p$ properly chosen noise vectors are as sufficient as the whole noise subspace $\text{Range}(\mathbf{U}_n)$ for (5) to yield a consistent estimate of $\mathbf{H}(z)$ and, second, under a mild additional assumption, each of the $q-p$ noise vectors can be found by computing the least eigenvector of a covariance matrix corresponding to a distinct $(p+1)$-tuples of channel outputs given by a *properly connected sequence* (PCS) defined as [5]:

Definition 1 *Denote the q system outputs by a set of members $m_1, \cdots, m_q$. A combination of $p+1$ ($\leq$ q) members $t = (m_{i_1}, \cdots, m_{i_{p+1}})$ is called a $(p+1)$-tuple. A sequence of $q-p$ tuples is said to be properly connected if each tuple in the sequence consists of p members shared by its preceding tuples and another member not shared by its preceding tuples. A properly connected sequence (PCS) is denoted by $S(p,q) = \{t_1, t_2, \cdots, t_{q-p}\}$ where*

$$\begin{aligned} &t_i = (m_{i_1}, \cdots, m_{i_p}, m_{i_{p+1}}),\ \ 1 \leq i \leq q-p \\ &\{m_{i_1}, \cdots, m_{i_p}\} \subset \{m_{j_k} |\ j < i,\ 1 \leq k \leq p+1\} \\ &m_{i_{p+1}} \notin \{m_{j_k} |\ j < i,\ 1 \leq k \leq p+1\}. \end{aligned}$$

3 The SMNS method

The idea of SMNS is to use q noise vectors instead of $q-p$ as in the original MNS is such a way that all system outputs are used identically. Indeed, in the original MNS method, certain system outputs are used more than others depending on the chosen PCS. This might lead to poor estimation performances if the system outputs that are used most correspond to the 'worst system channels'. This raises the problem of the 'best' choice of PCS. In the SMNS we avoid that problem by choosing the following sub-systems for the computation of the noise vectors:

$$\left\{ \begin{array}{l} (1, 2, \cdots, p+1) \\ (2, 3, \cdots, p+2) \\ \quad \vdots \\ (q-p, \cdots, q) \\ (q-p+1, \cdots, q, 1) \\ \quad \vdots \\ (q, 1, \cdots, p) \end{array} \right. \quad (6)$$

Note that the first $q-p$ tuples correspond to a PCS and the last p tuples correspond to the additional redundancy we introduce to guarantee that all system outputs are used $p+1$ times (i.e. we guarantee here a certain symmetry between the system outputs).

More precisely, the SMNS estimation method proceeds as follows:

- For each tuple of channel outputs $(m_{i_1}, \cdots, m_{i_{p+1}})$, we compute the covariance matrix:

$$\mathbf{R}^{(i)} = \frac{1}{T-N+1}\sum_{n=1}^{T-N+1} \underline{\mathbf{y}}_i(n)\underline{\mathbf{y}}_i^*(n) \quad (7)$$

  T being the sample size and

$$\underline{\mathbf{y}}_i(n) \stackrel{\text{def}}{=} [\mathbf{y}_N^{(m_{i_1})}(n)^T, \cdots \mathbf{y}_N^{(m_{i_{p+1}})}(n)^T]^T.$$

  Then, we compute its least dominant eigenvector $\underline{\mathbf{v}}_i$.

- Let $\underline{\mathbf{v}}_i = [\underline{\mathbf{v}}_{i_1}^T, \cdots, \underline{\mathbf{v}}_{i_{p+1}}^T]^T$ where each subvector has the dimension $N \times 1$ (i.e., $\underline{\mathbf{v}}_{i_k} = [\underline{\mathbf{v}}_{i_k}(0), \cdots, \underline{\mathbf{v}}_{i_k}(N-1)]^T$, for $k = 1, \cdots, p+1$). Then define "zero-padded" $qN \times 1$ vectors

$$\mathbf{v}_i = \left[\begin{array}{c} \mathbf{v}_{i,1} \\ \vdots \\ \mathbf{v}_{i,q} \end{array} \right] \quad \text{where} \quad \mathbf{v}_i(k) = \left\{ \begin{array}{ll} \underline{\mathbf{v}}_{i_j} & \text{if } k = i_j \\ \mathbf{0} & \text{otherwise} \end{array} \right. \quad (8)$$

  and form a $qN \times q$ matrix $\mathbf{V}_n$ of the q vectors $\{\mathbf{v}_i\}$, i.e., $\mathbf{V}_n = [\mathbf{v}_1, \cdots, \mathbf{v}_q]$.

- Estimate the channel parameter vector:

$$\text{h} \stackrel{\text{def}}{=} vec\left([\mathbf{H}_1^T, \cdots, \mathbf{H}_q^T]^T\right)$$

  where $\mathbf{H}_i = [\mathbf{H}_i^T(0), \cdots, \mathbf{H}_i^T(M)]^T$ by minimizing the least squares criterion

$$\hat{\text{h}} = \arg\min \|\mathcal{T}_N^*(\mathbf{F})\mathbf{V}_n\|^2 \quad (9)$$

  under a suitable constraint (see [4] for more details).

4 Efficient implementation

The main advantage of MNS-like methods is that large matrix eigendecomposition is avoided and the noise vectors are computed in a parallel scheme as the least eigenvectors of covariance matrices corresponding to the chosen $(p+1)$-tuples of system outputs.

Eventhough, computing one single eigenvector costs $O(L^2)$ (L

being the size of considered matrices), we found out that existing algorithms (e.g., [7, 8]) for extracting minor subspaces or minor eigenvectors are unefficient and slowly convergent in comparison with those dedicated to principal subspaces or principal eigenvectors. For this reason, we propose here to compute from the beginning the inverse matrices using RLS-type technique followed by a power method to extract the principal eigenvectors of each of the considered matrices. Therefore, using Schur inversion lemma, the first step of SMNS algorithm is replaced by the following: For $t = 1, \cdots, T - N + 1$

$$\begin{aligned} \mathbf{z}_i(t) &= \mathbf{P}_i(t-1)\underline{\mathbf{y}}_i(t) \\ \mathbf{P}_i(t) &= \mathbf{P}_i(t-1) - \frac{\mathbf{z}_i(t)\mathbf{z}_i^*(t)}{1 + \underline{\mathbf{y}}_i^*(t)\mathbf{z}_i(t)} \end{aligned} \quad (10)$$

where $\mathbf{P}_i$ represents the inverse matrix of $\mathbf{R}^{(i)}$. The initialization is done by choosing $\mathbf{P}_i(0) = \alpha\mathbf{I}$, α being a small positive scalar[1]. The least eigenvector of $\mathbf{R}^{(i)}$ becomes the principal eigenvector of $\mathbf{P}_i$ that can be computed using power iterations according to: For $k \geq 1$

$$\begin{aligned} \mathbf{w}_i(k) &= \mathbf{P}_i\mathbf{w}_i(k-1) \\ \mathbf{w}_i(k) &:= \mathbf{w}_i(k)/\|\mathbf{w}_i(k)\| \end{aligned}$$

where $\mathbf{w}_i(k)$ represents the desired eigenvector estimate at the kth iteration. The initialization vector $\mathbf{w}_i(0)$ is chosen randomly.

To solve equation (9) in the single input case, one can use a similar RLS+power iteration algorithm applied to the quadratic form matrix of criterion (9). The latter can be rewritten as:

$$\|\mathcal{T}_N^*(\mathbf{f})\mathbf{V}_n\|^2 = \mathbf{f}^* \left(\sum_{i=1}^{q} \mathcal{D}_N(\mathbf{v}_i)^* \mathcal{D}_N(\mathbf{v}_i) \right) \mathbf{f} \quad (11)$$

where

$$\begin{aligned} \mathcal{D}_N(\mathbf{v}_i) &= [\mathcal{D}_N(\mathbf{v}_{i,1}), \cdots, \mathcal{D}_N(\mathbf{v}_{i,q})] \\ \mathcal{D}_N(\mathbf{v}_{i,j}) &= \begin{bmatrix} v_{i,j}(0) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ v_{i,j}(N-1) & & & v_{i,j}(0) \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & v_{i,j}(N-1) \end{bmatrix}. \end{aligned}$$

However, in the multi-input case, the solution of (9) depends on the considered constraint and thus it cannot be computed necessarily as a least eigenvector of a given matrix [4].

5 Discussion

Performance comparison: Comparatively, the SMNS and MNS have almost the same order of computational cost except if $q \gg (q-p)$ in which case the SMNS becomes computationally more expensive.

[1] Note that the eigenvectors of $\mathbf{R}^{(i)}$ coincide with those of $\mathbf{R}^{(i)} + \alpha\mathbf{I}$.

Concerning the channel estimation error, we have a significant gain in favor of the SMNS that can be explained by the fact that all system outputs are used equally (which is not the case in the MNS) and the fact that for a fixed number of sensors q, the larger the number of sources is, the smaller is the number of noise vectors used by the MNS for channel identification. Eventhough, this is theoretically justified, the estimation performances are affected seriously as the parameter estimation problem becomes harder and thus requires 'more efforts' when the number of sources increases.

Convergence rate: The average convergence rate of the RLS + power iteration method is high (typically 5 to 10 iterations are sufficient) and overcomes the one of the optimal step size gradient method in [7]. This has been observed in different scenarios (see figures 3-5) at low and high SNRs and for single and multiple ($p = 2$) input cases. Also, we have observed that the average convergence rate remains quite high (see figure 6) even if the noise subspace dimension is larger than one (this is the case if we choose $N > pM + 1$)[2].

Beyond the minimum: Note that, under additional assumptions on the channel transfer functions, one can use less than $q - p$ noise vectors (recall that, without additional assumptions, $q-p$ is the minimum number of noise vectors needed to achieve unique channel identification). As an example, let consider the case $p = 1$ and assume that the channels are such that $\mathrm{h}_i(z)$ and $\mathrm{h}_j(z)$ are co-prime[3] for all $i \neq j$. In that case, only $\lceil \frac{q}{2} \rceil + 1$ noise vectors are sufficient for unique channel identification where $\lceil . \rceil$ represents the integer rounding towards plus infinity. In fact, one can observe that each of the following system output pairs: $(1, 2), (3, 4), \cdots, (q-1, q)$ allows a unique identification of the corresponding channel polynomials, i.e., $(\mathrm{h}_1(z), \mathrm{h}_2(z))$, $(\mathrm{h}_3(z), \mathrm{h}_4(z)), \cdots, (\mathrm{h}_{q-1}(z), \mathrm{h}_q(z))$ up to unknown scalar constants $\alpha_1, \alpha_2, \cdots, \alpha_{\lceil \frac{q}{2} \rceil}$, respectively. To get rid of these scalar constant indeterminacy, we need to use one extra relation that links all system outputs together.

6 Performance evaluation

We present here some numerical simulations to assess the performance of our SMNS algorithm. Figure 1 (resp. figure 2) compares performances of SMNS and MNS methods for $p = 1$ (resp. $p = 2$) sources and $q = 6$ channels of degree $M = 2$. Statistics are evaluated over 100 Monte-Carlo runs with $T = 250$ samples and the channels are generated randomly (following a complex gaussian distribution for each channel coefficient) at each run.

Figures 3-6 show the convergence rate of the RLS+power iteration method and the optimal step size gradient method in [7] in different scenarios corresponding to the single input case with an SNR of 30dB, the single input case with an SNR of 10dB, the 2-input case with an SNR of 30dB, and the single input case when the noise subspace dimension is strictly larger than one, respectively. In all these scenarios, one can observe a relatively high convergence rate of the proposed algorithm.

[2] Note that the convergence rate of power methods is exponential in terms of the ratio of the two largest eigenvalues (λ_2/λ_1) of $\mathbf{P}_i$.

[3] This is a stronger assumption than that used previously where only 'global' co-primeness is required, i.e. $\mathrm{h}_1(z), \mathrm{h}_2(z), \cdots, \mathrm{h}_q(z)$ do not share common

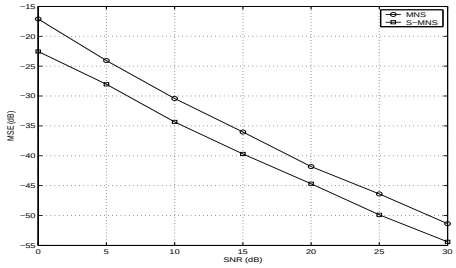

Figure 1: Comparison between MNS and SMNS for $p = 1$ and $q = 6$.

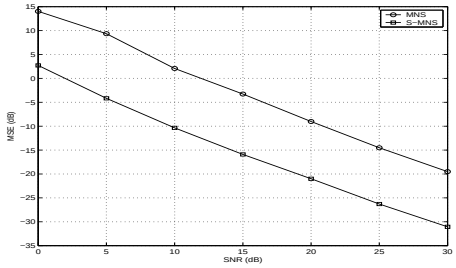

Figure 2: Comparison between MNS and SMNS for $p = 2$ and $q = 6$.

References

[1] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proceedings of the IEEE* ,Volume: 85 Issue: 8, pp. 1310 -1322, Aug. 1997.

[2] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE-T-SP*, pp. 516–525, Feb. 1995.

[3] T. Kailath, *Linear Systems*. Prentice-Hall, Inc., 1980.

[4] A. Gorokhov and P. Loubaton, "Subspace based techniques for second order blind separation of convolutive mixtures with temporally correlated sources." *IEEE-T-CS: Fund. Theory. & Appl.*, Sept. 1997.

[5] K. Abed-Meraim and Y. Hua, "Blind identification of MIMO system using minimum noise subspace", *IEEE-T-SP*, vol. 45, pp. 254–258, no. 1, Jan. 1997.

[6] K. Abed-Meraim, Ph. Loubaton and E. Moulines "A Subspace algorithm for certain blind identification problems" *IEEE-T-IT*, Mar. 1997.

[7] J. Manton, "A new algorithm for computing the extreme eigenvectors of a complex hermitian matrix", *Proc. of IEEE Work. on Stat. Sig. Proc. (SSP), Singapore*, pp. 225–228, Aug. 2001.

[8] K. Abed-Meraim, S. Attallah, A. Chkeif and Y. Hua, "Orthogonal Oja Algorithm," *IEEE-SP Letters*, vol. 7, no. 5, May 2000.

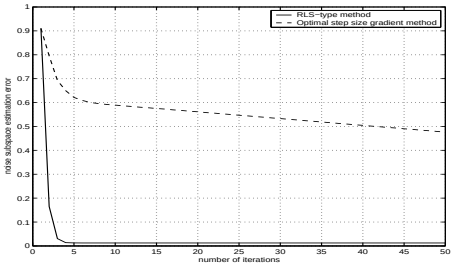

Figure 3: Convergence rate comparison: for $p = 1$, $q = 2$ and SNR$=$ 30dB

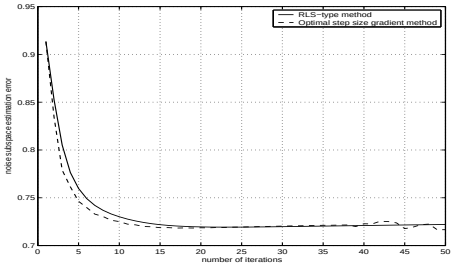

Figure 4: Convergence rate comparison: for $p = 1$, $q = 2$ and SNR$=$ 10dB

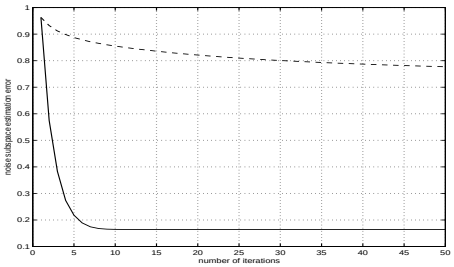

Figure 5: Convergence rate comparison: for $p = 2$, $q = 3$ and SNR$=$ 30dB

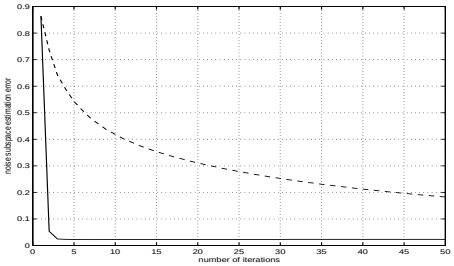

Figure 6: Convergence rate comparison: for $p = 1$, $q = 2$, SNR$=$ 30dB and $N = M + 2$

zeros.