



**HAL**  
open science

# **Adaptive Blind Identification of Sparse SIMO Channels using Maximum a Posteriori Approach**

Nacerredine Lassami, Abdeldjalil Aissa El Bey, Karim Abed-Meraim

► **To cite this version:**

Nacerredine Lassami, Abdeldjalil Aissa El Bey, Karim Abed-Meraim. Adaptive Blind Identification of Sparse SIMO Channels using Maximum a Posteriori Approach. Asilomar 2018 : 52nd Conference on Signals, Systems, and Computers, Oct 2018, Pacific Grove, Ca, United States. <10.1109/ACSSC.2018.8645077>. <hal-01847560>

**HAL Id: hal-01847560**

**<https://hal.science/hal-01847560v1>**

Submitted on 23 Jul 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Adaptive Blind Identification of Sparse SIMO Channels using Maximum a Posteriori Approach

Nacerredine Lassami

*IMT Atlantique*

*Lab-STICC, UBL*

29238 Brest, France

nacerredine.lassami@imt-atlantique.fr

Abdeljalil Aïssa-El-Bey

*IMT Atlantique*

*Lab-STICC, UBL*

29238 Brest, France

abdeldjalil.aissaelbey@imt-atlantique.fr

Karim Abed-Meraim

*University of Orléans*

*PRISME Laboratory*

45067 Orleans, France

karim.abed-meraim@univ-orleans.fr

**Abstract**—In this paper, we are interested in adaptive blind channel identification of sparse single input multiple output (SIMO) systems. A generalized Laplacian distribution is considered to enhance the sparsity of the channel coefficients with a maximum a posteriori (MAP) approach. The resulting cost function is composed of the classical deterministic maximum likelihood (ML) term and an additive  $\ell_p$  norm of the channel coefficient vector which represents the sparsity penalization. The proposed adaptive optimization algorithm is based on a simple gradient step. Simulations show that our method outperforms the existing adaptive versions of cross-relation (CR) method.

**Index Terms**— Adaptive algorithms, sparse channel, SIMO channel, blind identification.

## I. INTRODUCTION

The blind system identification (BSI) techniques aim to estimate a systems unknown information from its output only. This technology is particularly suitable for applications where all the available data are generated from an unknown system driven by an unknown input (SIMO system). The need for BSI arises from a number of applications such as in data communications [1], speech recognition [2], image restoration [3], seismic signal processing [4], etc. The BSI problem has received a lot of attention over the last two decades and many efficient solutions exist in the literature. We can distinguish two main classes of BSI methods: higher order statistics (HOS) and second order statistics (SOS) techniques. In general, HOS-based methods require large sample sizes to achieve better estimation performances than the SOS-based methods [5]. Among the famous SOS-based techniques, one can cite the cross-relation (CR) method [6], the subspace method [7], and the two-step maximum likelihood (TSML) method [8]. Unfortunately, it seems likely that in case of very long impulse response and sparse channel, these methods perform poorly. Such sparse channels can be encountered in many communication applications including High-Definition television (HDTV) channels and underwater acoustic channels. Recently, solutions have been proposed to handle this case by adapting the standard blind identification methods with the sparse case. To this end, they either constrain the desired solution by adding an  $\ell_p$  norm constraint (with  $0 < p \leq 1$ ) to the cost function [9] [10], or they use a specular channel parametric model [11]. Furthermore, the main objective of the derived algorithms is to be used in real-time applications,

therefore, they should be computationally simple and can be adaptively implemented. A solution was proposed in [12] to deal adaptively with such system identification based on the CR method in the case of SIMO channel.

In this paper, we extend the work in [9] by considering the SIMO case of a time varying sparse channel. A generalized Laplacian distribution is considered to enhance the sparsity of the channel coefficients with a Maximum a Posteriori (MAP) approach. Then, An adaptive technique based on gradient descent method is proposed to estimate efficiently the sparse channel coefficients. The robustness against channel overestimation errors will also be discussed.

The rest of this paper is organized as follows. In the next section, we present the considered data model, then Section III reviews the ML approach. In section IV, the MAP approach is discussed in addition to the used gradient descent algorithm. In section V, the simulation comparisons with the adaptive sparse CR method [12] are provided followed by some concluding remarks in Section VI.

## II. DATA MODEL AND PROBLEM FORMULATION

Our focus is on the SIMO sparse channel case in a blind identification context. We aim to estimate the channel state information using only the observation data and the sparsity information of the channel coefficients. Considering a mathematical model where the input and the output are both discrete but we have access only to the output. The system is driven by the input sequence  $s(n)$  and yields  $M$  output sequences  $x_1(n), \dots, x_M(n)$  at time  $n$  where the  $x_i(n)$  correspond to the  $i$ -th output observation. The system is defined by a finite impulse responses (FIRs)  $h_i(n) \neq 0$ , for  $n = 0, \dots, L$  and  $i = 1, \dots, M$  and can be described as follows

$$\begin{cases} x_1(n) = s(n) * h_1(n) + w_1(n) \\ x_2(n) = s(n) * h_2(n) + w_2(n) \\ \vdots \\ x_M(n) = s(n) * h_M(n) + w_M(n) \end{cases} \quad (1)$$

where  $*$  denotes linear convolution and  $\mathbf{w}(n) = [w_1(n), \dots, w_M(n)]^T$  is an additive spatial white noise. It is assumed that the maximum order of the  $M$  channels is  $L$  and the number of available output samples of each

channel is  $N$ . Note that the channel order  $L$  will be assumed to be known throughout this paper unless stated otherwise in the case of an over-estimation.

In vector form, equation (1) can be expressed as:

$$\mathbf{x}(n) = \sum_{k=0}^L \mathbf{h}(k)s(n-k) + \mathbf{w}(n) \quad (2)$$

where  $\mathbf{x}(n) = [x_1(n), x_2(n), \dots, x_M(n)]^T$ . The transfer function of the system  $\mathbf{h}(z) = \sum_{k=0}^L \mathbf{h}(k)z^{-k}$  corresponds to an unknown causal FIR satisfying  $\mathbf{h}(z) \neq 0, \forall z$ . The aim of BSI is to estimate the channel coefficients vector  $\mathbf{h} = [\mathbf{h}(0)^T, \dots, \mathbf{h}(L)^T]^T$  up to a scalar constant (for more details about the indeterminacy of the problem, see [13]).

### III. MAXIMUM LIKELIHOOD APPROACH

The deterministic maximum likelihood (ML) is a classical approach applicable to parameter estimation problems where the source signal is considered as being deterministic and the probability density function (PDF) of the available data is known. Stacking all observation samples into a single vector  $\mathbf{x} = [\mathbf{x}_1^T, \dots, \mathbf{x}_M^T]^T$  with  $\mathbf{x}_i = [x_i(0), \dots, x_i(N-1)]^T$  for  $i = 1, \dots, M$  and assuming that the system output vector is corrupted by additive white Gaussian noise vector allows us to write:

$$\mathbf{x} = \mathcal{H}_M \mathbf{s} + \mathbf{w} \quad (3)$$

with its PDF given by

$$f(\mathbf{x}|\mathbf{h}, \mathbf{s}) = \frac{1}{(\pi\sigma^2)^N} \exp\left(-\frac{1}{\sigma^2} \|\mathbf{x} - \mathcal{H}_M \mathbf{s}\|_2^2\right) \quad (4)$$

where  $\mathbf{s} = [s(-L), s(-L+1), \dots, s(N-1)]^T$ ,  $\sigma^2$  is the variance of the white noise  $\mathbf{w}$  and  $\mathcal{H}_M = [\mathbf{H}_1^T, \dots, \mathbf{H}_M^T]^T$  is a block Sylvester matrix,  $\mathbf{H}_i$  being the Sylvester matrix of the  $i$ -th channel. It then follows that the channel output vector  $\mathbf{x}$  is Gaussian distributed with the mean vector  $\hat{\mathbf{x}} = \mathcal{H}_M \mathbf{s}$  and the covariance matrix  $\mathbf{R}_x = \sigma^2 \mathbf{I}$ . The ML criterion is expressed as:

$$(\mathcal{H}_M, \mathbf{s}) = \arg \min_{\mathcal{H}_M, \mathbf{s}} f(\mathbf{x}|\mathbf{h}) \quad (5)$$

$$= \arg \max_{\mathcal{H}_M, \mathbf{s}} \left\{ \|\mathbf{x} - \mathcal{H}_M \mathbf{s}\|_2^2 \right\} \quad (6)$$

Let's define  $\mathcal{G}_M$  by

$$\mathcal{G}_2^H = [-\bar{\mathbf{H}}_1, \bar{\mathbf{H}}_2] \quad (7)$$

and

$$\mathcal{G}_q^H = \left[ \begin{array}{c|c} \mathcal{G}_{q-1}^H & 0 \\ \hline -\bar{\mathbf{H}}_q & \bar{\mathbf{H}}_1 \\ \vdots & \vdots \\ 0 & -\bar{\mathbf{H}}_q \quad \bar{\mathbf{H}}_{q-1} \end{array} \right] \quad (8)$$

for  $q = 3, \dots, M$ . Where  $\bar{\mathbf{H}}_q$  is the top-left  $(N-L) \times N$  sub-matrix of  $\mathbf{H}_q$ .

Under the necessary identifiability condition [5] and by using projection techniques and commutativity property of linear convolution, equation (6) yields the equivalent problem:

$$\hat{\mathbf{h}} = \arg \min_{\|\mathbf{h}\|_2=1} \{\mathbf{h}^H \mathcal{X}_M^H (\mathcal{G}_M^H \mathcal{G}_M)^{\#} \mathcal{X}_M \mathbf{h}\} \quad (9)$$

where  $(\cdot)^{\#}$  refers to the pseudo inverse operator and  $\mathcal{X}_M$  is defined by

$$\mathcal{X}_2 = [\mathbf{X}_2, -\mathbf{X}_1] \quad (10)$$

and

$$\mathcal{X}_q = \left[ \begin{array}{c|c} \mathcal{X}_{q-1} & 0 \\ \hline \mathbf{X}_q & -\mathbf{X}_1 \\ \vdots & \vdots \\ 0 & \mathbf{X}_q \quad -\mathbf{X}_{q-1} \end{array} \right] \quad (11)$$

for  $q = 3, \dots, M$ , where  $\mathbf{X}_q$  is given by:

$$\mathbf{X}_q = \begin{bmatrix} x_q(L) & \dots & x_q(0) \\ \vdots & & \vdots \\ x_q(N-1) & \dots & x_q(N-L-1) \end{bmatrix}. \quad (12)$$

The cost function is optimized under the constraint  $\|\mathbf{h}\|_2 = 1$  to avoid the scalar indeterminacy. The TSML method [8] uses an iterative two-step estimation procedure to solve equation (9) efficiently as shown below:

- Minimize  $\mathbf{h}_c = \arg \min_{\|\mathbf{h}\|_2=1} \{\mathbf{h}^H \mathcal{X}_M^H \mathcal{X}_M \mathbf{h}\}$
- Minimize  $\mathbf{h} = \arg \min_{\mathbf{h}} \{\mathbf{h}^H \mathcal{X}_M^H (\mathcal{G}_c^H \mathcal{G}_c)^{\#} \mathcal{X}_M \mathbf{h}\}$

where  $\mathcal{G}_c$  is constructed from  $\mathbf{h}_c$  according to (8) and (7). Although The CR method was developed based on an algebraic insight into the  $M$  channel system, the first step of the TSML method is equivalent to the CR method. This links between the two methods makes our later comparison with the adaptive CR method [12] more coherent.

### IV. MAXIMUM A POSTERIORI APPROACH

The main idea of the Maximum a Posteriori (MAP) approach is to estimate  $\mathbf{h}$  using its conditional probability distribution as follows:

$$\hat{\mathbf{h}}_{MAP} = \arg \max_{\mathbf{h}} \left\{ \frac{f(\mathbf{x}|\mathbf{h})f(\mathbf{h})}{\int f(\mathbf{x}|\mathbf{h}')f(\mathbf{h}')d\mathbf{h}'} \right\} \quad (13)$$

$$= \arg \max_{\mathbf{h}} \{f(\mathbf{x}|\mathbf{h})f(\mathbf{h})\} \quad (14)$$

Generally speaking, the MAP allows us to exploit prior information about the desired parameter. Hence, one needs to a priori know the probability distribution function  $f(\mathbf{h})$  of the channel vector. This a priori depends on the application context and its physical environment. In our case, it is the channel vector sparsity that we model by representing the channel PDF with the generalized Laplacian distribution given by:

$$f(\mathbf{h}) = \left[ 2\beta\Gamma\left(1 + \frac{1}{p}\right) \right]^{-M(L+1)} \exp\left(-\frac{\|\mathbf{h}\|_p^p}{\beta^p}\right) \quad (15)$$

where  $\beta > 0$  is a scale parameter,  $0 < p \leq 1$  and  $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ ,  $z > 0$  is the Gamma function. With this PDF, one increases the chances to get channel coefficients close

to zero. Combining equations (4), (14) and (15) leads to the following objective function:

$$\mathcal{J}(\mathbf{h}) = \mathbf{h}^H \mathcal{X}_M^H (\mathcal{G}_M^H \mathcal{G}_M)^\# \mathcal{X}_M \mathbf{h} + \lambda \|\mathbf{h}\|_p^p \quad (16)$$

where  $\lambda = \frac{\sigma^2}{\beta^p}$  is a good approximation to the 'optimal' weighting parameter which controls the trade-off between the ML term and the penalty term. The cost function is optimized under the same constraint  $\|\mathbf{h}\|_2 = 1$  to avoid the scalar indeterminacy. The minimization of such a problem is computationally expensive and may be even intractable when the channel impulse responses are long and the number of channels is large.

For a slowly varying channel, one can reduce the computational cost and track the channel variations by using a stochastic adaptive gradient technique to solve the previous minimization problem efficiently.

Let  $\mathbf{h}(t)$  be the solution after the  $t$ -th iteration, then the solution at the  $(t+1)$ -th iteration is given by:

$$\mathbf{h}(t+1) = \mathbf{h}(t) - \mu \nabla \mathcal{J}(\mathbf{h}(t)) \quad (17)$$

where  $\mu$  is a small positive optimization step and the gradient  $\nabla \mathcal{J}(\mathbf{h}(t))$  is given by:

$$\nabla \mathcal{J}(\mathbf{h}(t)) = 2\mathcal{X}_M^H(t+1)(\mathcal{G}_M^H(t)\mathcal{G}_M(t))^\# \mathcal{X}_M(t+1)\mathbf{h}(t) + \lambda p \tilde{\mathbf{h}}(t) \quad (18)$$

where  $\tilde{h}_i = \text{sign}(h_i) |h_i|^{p-1}$  for  $i = 1, \dots, M(L+1)$ .

We define the matrix:

$$\mathcal{Q}_w(t+1) = \mathcal{X}_M^H(t+1)\mathbf{W}(t)\mathcal{X}_M(t+1) \quad (19)$$

with  $\mathbf{W}(t) = (\mathcal{G}_M^H(t)\mathcal{G}_M(t))^\#$ .

A more elegant way to update  $\mathcal{Q}_w(t+1)$  is to use the adaptive exponential window

$$\mathcal{X}_M(t+1) = [\sqrt{\gamma}\mathcal{X}_M(t)^T, \bar{\mathcal{X}}_M(t+1)^T]^T \quad (20)$$

with  $0 < \gamma < 1$  being a forgetting factor and  $\bar{\mathcal{X}}_M(t+1)$  is given by:

$$\bar{\mathcal{X}}_2(t+1) = [\bar{\mathbf{x}}_2(t+1), -\bar{\mathbf{x}}_1(t+1)] \quad (21)$$

and

$$\bar{\mathcal{X}}_q(t+1) = \left[ \begin{array}{c|c} \bar{\mathbf{x}}_{q-1}(t+1) & 0 \\ \hline \bar{\mathbf{x}}_q(t+1) & -\bar{\mathbf{x}}_1(t+1) \\ & \vdots \\ 0 & \bar{\mathbf{x}}_q(t+1) \\ & -\bar{\mathbf{x}}_{q-1}(t+1) \end{array} \right] \quad (22)$$

for  $q = 3, \dots, M$  with  $\bar{\mathbf{x}}_q(t) = [x_q(t), \dots, x_q(t-L)]$ .

Hence, combining equation (19) and (20) yields:

$$\begin{aligned} \mathcal{Q}_w(t+1) &= \gamma \mathcal{X}_M^H(t)\mathbf{W}(t)\mathcal{X}_M(t) \\ &+ \bar{\mathcal{X}}_M^H(t+1)\mathbf{W}(t)\bar{\mathcal{X}}_M(t+1) \end{aligned} \quad (23)$$

Using the approximation  $\mathcal{Q}_w(t) \approx \mathcal{X}_M^H(t)\mathbf{W}(t)\mathcal{X}_M(t)$ , we can rewrite:

$$\mathcal{Q}_w(t+1) = \gamma \mathcal{Q}_w(t) + \bar{\mathcal{X}}_M^H(t+1)\mathbf{W}(t)\bar{\mathcal{X}}_M(t+1) \quad (24)$$

The algorithm can be summarized as follows:

1) Update  $\mathbf{W}(t) = (\mathcal{G}_M^H(t)\mathcal{G}_M(t))^\#$

2) Update  $\mathcal{Q}_w(t+1)$  using (24)

3) Update  $\mathbf{h}(t+1)$  using (17) and (18)

4) Normalize  $\mathbf{h}(t+1)$  such that  $\|\mathbf{h}(t+1)\|_2 = 1$

Note that the proposed algorithm can be modified by using another optimization descent method. We can also change sparsity prior which will lead us to change the final objective function. Next, we will provide two ideas that are used to extend this work: derive an optimal gradient step and approximate the update of  $\mathbf{W}(t)$  in order to further reduce the computational complexity.

#### A. Computational complexity reduction

The complexity of the proposed algorithm is dominated by the computation of the pseudo inverse  $\mathbf{W}(t) = (\mathcal{G}_M^H(t)\mathcal{G}_M(t))^\#$  where  $\mathcal{G}_M(t)$  is defined in (7) and (8). One way to reduce the cost of this step is to use the previous gradient step and the fact that we are using a small optimization step  $\mu$  to approximate the pseudo inverse. Actually, one can use the first order approximation of the pseudo inverse of the sum of two matrices  $\mathcal{A} + \mu\mathcal{B}$  of size  $(n \times m)$  which is given by :

$$\begin{aligned} (\mathcal{A} + \mu\mathcal{B})^\# &\approx \mathcal{A}^\# - \mu\mathcal{A}^\#\mathcal{B}\mathcal{A}^\# \\ &+ \mu\mathcal{A}^\#\mathcal{A}^{\#T}\mathcal{B}^T(I_n - \mathcal{A}\mathcal{A}^\#) \\ &+ \mu(I_m - \mathcal{A}^\#\mathcal{A})\mathcal{B}^T\mathcal{A}^{\#T}\mathcal{A}^\# + \mathcal{O}(\mu^2) \end{aligned} \quad (25)$$

In our case, we have

$$\mathbf{h}(t) = \frac{\mathbf{h}(t-1) - \mu \nabla \mathcal{J}(\mathbf{h}(t-1))}{\|\mathbf{h}(t-1) - \mu \nabla \mathcal{J}(\mathbf{h}(t-1))\|_2} \quad (26)$$

The linear construction of  $\mathcal{G}_M^H(t)$  from  $\mathbf{h}(t)$  as entry allows us to write :

$$\mathcal{G}_M^H(t) = \frac{\mathcal{G}_M^H(t-1) - \mu \nabla \mathcal{G}_M^H(t-1)}{\|\mathbf{h}(t-1) - \mu \nabla \mathcal{J}(\mathbf{h}(t-1))\|_2} \quad (27)$$

with  $\nabla \mathcal{G}_M^H(t-1)$  is the same construction as is (7) and (8) with the gradient  $\nabla \mathcal{J}(\mathbf{h}(t-1))$  as entry. which leads to:

$$\begin{aligned} \mathcal{G}_M^H(t)\mathcal{G}_M(t) &\approx \frac{1}{\|\mathbf{h}(t-1) - \mu \nabla \mathcal{J}(\mathbf{h}(t-1))\|_2^2} \times \{ \\ &\mathcal{G}_M^H(t-1)\mathcal{G}_M(t-1) \\ &- \mu \mathcal{G}_M^H(t-1)\nabla \mathcal{G}_M(t-1) \\ &- \mu \nabla \mathcal{G}_M^H(t-1)\mathcal{G}_M(t-1) \} + \mathcal{O}(\mu^2) \end{aligned} \quad (28)$$

Separating terms with and without  $\mu$  in (28) and using them as  $\mathcal{B}$  and  $\mathcal{A}$  respectively in (25) enables us to approximate  $(\mathcal{G}_M^H(t)\mathcal{G}_M(t))^\#$  without an expensive complexity.

**Remark:** Another way to reduce the numerical cost would be to choose a noise subspace generating matrix  $\hat{\mathcal{G}}_M$  that is 'non redundant' so that matrix  $\hat{\mathcal{G}}_M^H \hat{\mathcal{G}}_M$  is invertible which allows us to replace the pseudo-inversion by the relatively simpler matrix inversion. To build matrix  $\hat{\mathcal{G}}_M$ , one can follow similar steps as for the minimum CR method in [6].

## B. Gradient step optimization

The choice of the optimization step  $\mu$  is really important to achieve good convergence performance. Choosing  $\mu$  too large can cause the divergence of the algorithm and in the same time, a too small  $\mu$  will induce a poor convergence rate. To avoid these problems, one can use the fact that we are using a gradient descent method which allows us to derive an optimal step that minimizes:

$$\hat{\mu} = \arg \min_{\mu} \mathcal{J}\{\mathbf{h}(t+1)\} \quad (29)$$

Let's start by replacing  $\mathbf{h}(t+1)$  by its formula from (17) in (16) which yields:

$$\begin{aligned} \mathcal{J}(\mathbf{h}(t+1)) = & (\mathbf{h}(t) - \mu \nabla \mathcal{J}(\mathbf{h}(t)))^H \mathbf{Q}_w(t+1) (\mathbf{h}(t) - \mu \nabla \mathcal{J}(\mathbf{h}(t))) \\ & + \lambda \|\mathbf{h}(t) - \mu \nabla \mathcal{J}(\mathbf{h}(t))\|_p^p \end{aligned} \quad (30)$$

hence, the derivative of the previous formula w.r.t.  $\mu$  is:

$$\begin{aligned} \frac{\partial \mathcal{J}(\mathbf{h}(t+1))}{\partial \mu} = \mathcal{K}(\mu) = & \{2(\mu \nabla \mathcal{J}(\mathbf{h}(t)) - \mathbf{h}(t))^H \mathbf{Q}_w(t+1) - \mu \lambda \tilde{\mathbf{r}}^H\} \\ & \times \nabla \mathcal{J}(\mathbf{h}(t)) \end{aligned} \quad (31)$$

where

$$\tilde{r}_i = \text{sign}(h_i(t) - \mu \nabla \mathcal{J}(\mathbf{h}(t))_i) |h_i(t) - \mu \nabla \mathcal{J}(\mathbf{h}(t))_i|^{p-1}$$

for  $i = 1, \dots, M(L+1)$ .

Finally, we can use a Newton method to approximate the optimal solution of (29) at every iteration with

$$\mu_t = \mu_{t-1} - \mathcal{K}(\mu_{t-1}) \frac{\mu_{t-1} - \mu_{t-2}}{\mathcal{K}(\mu_{t-1}) - \mathcal{K}(\mu_{t-2})} \quad (32)$$

The resulting algorithm is summarized in the table below.

---

### Algorithm 1 MAP-adapt & AMAP-adapt

---

**Require:**  $\mathbf{x}(t)$  data vector,  $\gamma$  the forgetting factor,  $\lambda$  the sparsity parameter,  $\mathbf{h}(t-1)$ ,  $\mathbf{Q}_w(t-1)$  and  $\mathbf{W}(t-1)$  from the previous iteration.

**Ensure:**  $\mathbf{Q}_w(t)$ ,  $\mathbf{W}(t)$  and the sparse channel  $\mathbf{h}(t)$

- 1: Update  $\mathbf{W}(t)$  by  $\begin{cases} (\mathcal{G}_M^H(t) \mathcal{G}_M(t))^\# & \text{if MAP-adapt} \\ \text{use (25) (28)} & \text{if AMAP-adapt} \end{cases}$
  - 2: Build  $\bar{\mathcal{X}}_M(t)$  as explained in (21) (22)
  - 3: Update  $\mathbf{Q}_w(t)$  using (24)
  - 4: Calculate the gradient  $\nabla \mathcal{J}(\mathbf{h}(t-1))$  given by (18)
  - 5: Update  $\mathbf{h}(t)$  with (17)  $\begin{cases} \text{fixed } \mu \\ \text{optimal } \mu \text{ using (32)} \end{cases}$
  - 6: Normalize  $\mathbf{h}(t+1)$  w.r.t.  $\|\mathbf{h}(t+1)\|_2 = 1$
- 

## V. SIMULATIONS

To assess the performance of the proposed algorithm, we consider a SIMO system with  $M$  outputs represented by a polynomial transfer function of degree  $L$ . The channel impulse response is a sparse sequence of random variables with Bernoulli-Gaussian distribution generated by the MATLAB function SPRANDN. In our simulation, we have used a

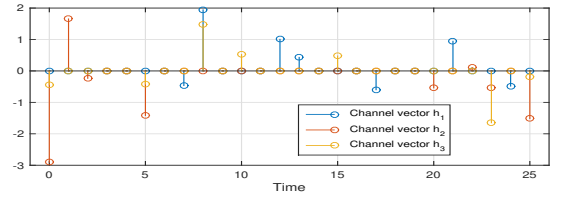


Fig. 1: Example of channel used in simulation with  $M = 3$  and  $L = 25$ .

sparsity level of 30%, which means that 30% of the vector  $\mathbf{h}$  entries are nonzeros as illustrated in Figure 1. The input signal is a random binary sequence and the additive white Gaussian noise has a variance  $\sigma^2$  chosen according to the target signal to noise ratio  $SNR = 10 \log(\frac{\|\mathbf{h}\|_2^2}{\sigma^2})$ . The used performance factor is the normalized mean-square error criterion given by

$$NMSE = \frac{1}{Nr} \sum_{k=1}^{Nr} 1 - \left( \frac{\hat{\mathbf{h}}_k^H \mathbf{h}}{\|\hat{\mathbf{h}}_k\|_2 \|\mathbf{h}\|_2} \right)^2 \quad (33)$$

with  $Nr = 100$  is the number of Monte-Carlo runs. For the comparison, we use the CR-based algorithm [12] which we will refer to as SCR-adapt. MAP-adapt will refer to the first proposed algorithm and AMAP-adapt will refer to the version that uses the pseudo inverse approximation.

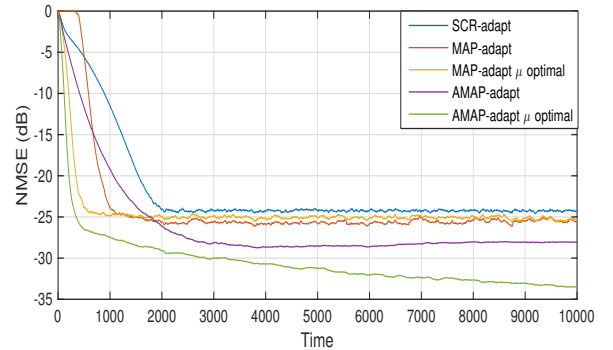


Fig. 2: NMSE in dB versus time for  $(M = 3, L = 15)$  and  $SNR = 20dB$ .

Figure 2 shows the NMSE evolution as function of time for the parameters  $M = 3, L = 15, \gamma = 0.98$  and  $SNR = 20$ . The optimization step is chosen fixed  $\mu = 0.0001$ , then we use the optimal one calculated according to the previous section. It is clearly shown that the proposed algorithms outperform the SCR-adapt algorithm. In addition, the following observations can be made out of this experiment: (i) the AMAP method is better than the MAP one due to the fact that the considered approximation allows us to control the slow variation of the up-date vector estimate contrary to the adaptive MAP using exact (batch) pseudo-inversion. In other words, the AMAP-adapt is better aligned with the spirit of the gradient method than the MAP-adapt. (ii) As expected, the optimal step size allows increasing the convergence rate of the considered

algorithms. (iii) the best results are obtained by the AMAP-adapt with optimal step size which has the joint advantages of reduced complexity, faster convergence rate and lower steady state error as compared to the other methods.

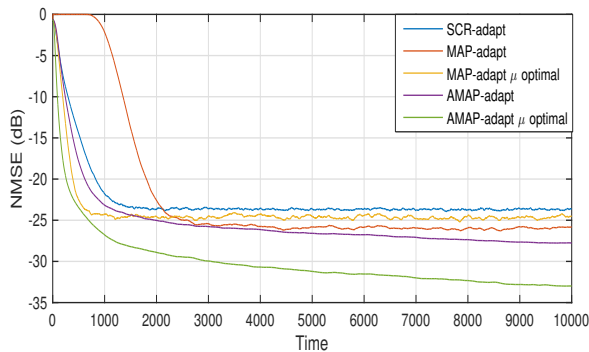


Fig. 3: NMSE in  $dB$  versus time for  $(M = 3, L = 30)$  and  $SNR = 20dB$ .

Although, we have changed the parameters to  $M = 3, L = 30, \gamma = 0.95$  and  $SNR = 20$ , we can see in Figure 3 relatively the same NMSE evolution over time. The AMAP-adapt algorithm with optimal  $\mu$  remains the best among all considered algorithms.

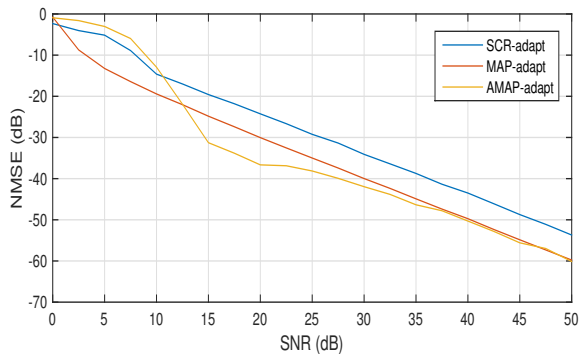


Fig. 4: NMSE in  $dB$  versus SNR en  $dB$  for  $(M = 3, L = 15)$ .

In Figure 4, the NMSE is plotted versus the SNR after  $T = 10000$  snapshots for a SIMO system with  $M = 3$  and  $L = 15$ . Generally, our algorithms MAP-adapt and AMAP-adapt keep having better performance than SCR-adapt algorithm for moderate and high SNRs.

In order to illustrate the behavior of the AMAP algorithm in case of overestimation of the channel order, we consider a system with  $M = 3, L = 60$  and  $SNR = 20$ . Figure 5 shows the robustness of the AMAP-adapt algorithm for a channel order overestimation by 15 and 30, respectively.

## VI. CONCLUSION

In this paper, adaptive blind identification of sparse SIMO channels has been considered. A sparsity prior on the channel coefficients was considered within a MAP approach to solve the BSI problem. Based on an adaptive gradient descent

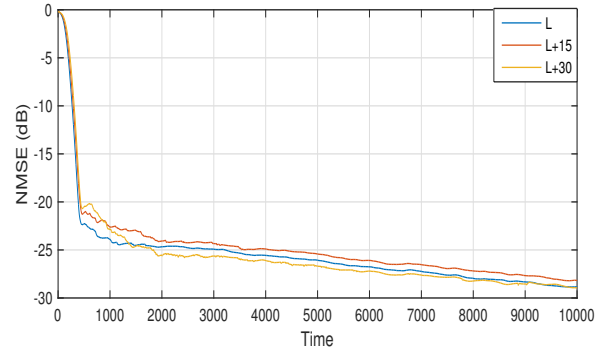


Fig. 5: NMSE in  $dB$  of AMAP-adapt algorithm versus time for  $(M = 3, L = 60)$  and  $SNR = 20dB$ .

method, the proposed algorithm has improved the estimation accuracy as compared to the adaptive CR method. Improvements has been proposed to enhance the computational complexity and the convergence speed of our solution.

## REFERENCES

- [1] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for ofdm systems," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1699–1712, 2002.
- [2] Yiteng Huang, J. Benesty, and Jingdong Chen, "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 882–895, 2005.
- [3] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 43–64, 1996.
- [4] A. K. Takahata, E. Z. Nadalin, R. Ferrari, L. T. Duarte, R. Suyama, R. R. Lopes, J. M. T. Romano, and M. Tygel, "Unsupervised processing of geophysical signals: A review of some key aspects of blind deconvolution and blind source separation," *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 27–35, 2012.
- [5] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1310–1322, 1997.
- [6] A. Aïssa-El-Bey, M. Grebici, K. Abed-Meraim, and A. Belouchrani, "Blind system identification using cross-relation methods: further results and developments," in *7th International Symposium on Signal Processing and Its Applications (ISSPA)*, July 2003, vol. 1, pp. 649–652.
- [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*, vol. 43, no. 2, pp. 516–525, 1995.
- [8] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 661–672, 1996.
- [9] A. Aïssa-El-Bey and K. Abed-Meraim, "Blind identification of sparse SIMO channels using maximum a posteriori approach," in *16th European Signal Processing Conference (EUSIPCO)*, August 2008, pp. 1–5.
- [10] A. Aïssa-El-Bey and K. Abed-Meraim, "Blind SIMO channel identification using a sparsity criterion," in *9th IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Recife, Brazil, July 2008, pp. 271–275.
- [11] L. Perros-Meilhac, E. Moulines, K. Abed-Meraim, P. Chevalier, and P. Duhamel, "Blind identification of multipath channels: a parametric subspace approach," *IEEE Transactions on Signal Processing*, vol. 49, no. 7, pp. 1468–1480, 2001.
- [12] A. Aïssa-El-Bey, K. Abed-Meraim, and C. Laot, "Adaptive blind estimation of sparse SIMO channels," in *International Workshop on Systems, Signal Processing and their Applications (WOSSPA)*, May 2011, pp. 348–351.
- [13] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2982–2993, 1995.