Blind identification of sparse SIMO channels using maximum a posteriori approach

Abdeldjalil Aissa El Bey, Karim Abed-Meraim

BLIND IDENTIFICATION OF SPARSE SIMO CHANNELS USING MAXIMUM A POSTERIORI APPROACH

Abdeldjalil Aïssa-El-Bey[1] *and Karim Abed-Meraim*[2,3]

[1]TELECOM Bretagne, SC Department, Technopôle Brest-Iroise, 29285, Brest, France
[2]University of Sharjah, ECUOS/ECE Department, 27272 Sharjah, UAE
[3]TELECOM ParisTech, TSI Department, 37-39 rue Dareau 75014, Paris, France

ABSTRACT

In this paper, we are interested in blind identification of sparse single-input multiple-output (SIMO) systems. A maximum a posteriori approach is considered using generalized Laplacian distribution for the channel coefficients. This leads to a cost function given by the deterministic maximum likelihood (ML) criterion penalized by 'a sparsity measure' term expressed by the ℓ_p norm of the channel coefficient vector. A simple but efficient optimization algorithm using gradient technique with optimal step-size is proposed. The simulations show that the proposed method outperforms the ML technique in terms of estimation error and is robust against channel order overestimation errors.

1. INTRODUCTION

Blind system identification (BSI) is a fundamental signal processing technology aimed at retrieving a system's unknown information from its outputs only. This problem has received a lot of attention in the signal processing literature and a plethora of methods and techniques have been proposed to solve the BSI over the last two decades [1–3]. Techniques for BSI can generally be classified into two main classes (i) higher order statistical (HOS) and (ii) second order statistical (SOS) methods. Although HOS methods [1] were proposed for BSI due to the rich information, large number of observation samples are required. As a result, SOS methods such as [4] have become more popular. Comparison between SOS and HOS methods have been presented in [3]. Unfortunately, these methods have demonstrated their limitation when channel impulse response is very long and sparse (e.g. HF communication, echo cancelation, etc).

Estimation of sparse long channels (i.e. channels with small number of nonzero coefficients but a large span of delays) is considered in this paper. Such sparse channels are encountered in many communication applications : High-Definition television (HDTV) channels are hundreds of data symbols long but there are only a few nonzero taps [5]. Hilly terrain delay profile has a small number of multipath in the broadband wireless communication [6] and underwater acoustic channels are also known to be sparse [7]. We propose exploiting the sparse nature of the channel via a maximum a posteriori (MAP) approach using a generalized Laplacian distribution to model the sparse channel taps. As will be shown in the sequel, the MAP criterion combines the ML cost function with the ℓ_p norm constraint ($0 < p \leq 1$) of the channel impulse response, which is considered by many authors as a good sparsity measure, e.g. [8, 9].

In the following section, we discuss the data model that formulates our problem. Next, we review the deterministic ML method for blind SIMO channel identification before using it to introduce the MAP solution. In Section 5, some simulations are undertaken to validate our algorithm and illustrate its robustness against channel overestimation errors and to compare its performance to other existing BSI techniques.

2. PROBLEM FORMULATION

The problem addressed in this paper is to determine the sparse impulse response of a SIMO system in a blind way, i.e. only the observed system outputs are available and used without assuming knowledge of the specific input signal.

Consider a mathematical model where the input and the output are discrete, the system is driven by a single-input sequence $s(n)$ and yields M output sequences $x_1(n), \ldots, x_M(n)$, and the system has finite impulse responses (FIR's) $h_i(n) \neq 0$, for $n = 0, \ldots, L$ and $i = 1, \ldots, M$. Such a system model can be described as follows :

$$\left\{ \begin{array}{rcccl} x_1(n) & = & s(n) * h_1(n) & + & w_1(n) \\ x_2(n) & = & s(n) * h_2(n) & + & w_2(n) \\ & \vdots & & & \\ x_M(n) & = & s(n) * h_M(n) & + & w_M(n) \end{array} \right. \quad (1)$$

where $*$ denotes linear convolution and $\boldsymbol{w}(n) = [w_1(n), \ldots, w_M(n)]^T$ is an additive spatial white noise, i.e. $\mathbb{E}[\boldsymbol{w}(n)\boldsymbol{w}(n)^H] = \sigma^2 \boldsymbol{I}_M$ where $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and the conjugate transpose, respectively and $\boldsymbol{I}_M$ is the $M \times M$ identity matrix. In vector form, equation (1) can be expressed as :

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

where $\boldsymbol{h}(z) = \sum_{k=0}^{L} \boldsymbol{h}(k) z^{-k}$ is an unknown causal FIR $M \times 1$ transfer function satisfying $\boldsymbol{h}(z) \neq 0, \forall z$. Given a finite set of observation vectors $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T)$ and based on the channel entries co-primness (i.e. $\boldsymbol{h}(z) \neq 0 \ \forall z$), the objective here is to estimate the channel coefficients vector $\boldsymbol{h} = [\boldsymbol{h}(0)^T, \cdots, \boldsymbol{h}(L)^T]^T$ up to a scalar constant (this is an inherent indeterminacy of the blind system identification problem as shown in [4]).

3. MAXIMUM-LIKELIHOOD APPROACH

The deterministic maximum-likelihood (ML) is a classical approach applicable to parameter estimation problems where the probability density function (PDF) of the available data is known. Assuming that the system output vector is corrupted by additive white Gaussian noise vector, the system output vector (i.e. vector given by the stack of all observation samples) can be written as

$$\boldsymbol{x} = \boldsymbol{\mathcal{H}}_M \boldsymbol{s} + \boldsymbol{w} \tag{3}$$

and its PDF given by

$$f(\boldsymbol{x}|\boldsymbol{h}) = \frac{1}{\pi^T \sigma^{2T}} \exp\left(-\frac{1}{\sigma^2}\|\boldsymbol{x} - \boldsymbol{\mathcal{H}}_M \boldsymbol{s}\|_2^2\right) \tag{4}$$

where σ^2 is the variance of each element of $\boldsymbol{w}$ and $\boldsymbol{\mathcal{H}}_M$ is a block Sylvester matrix [2] ($\boldsymbol{\mathcal{H}}_M = [\boldsymbol{H}_1^T, \cdots, \boldsymbol{H}_M^T]^T$ where $\boldsymbol{H}_i$ is the Sylvester matrix of the i-th channel). The ML estimates of $\boldsymbol{\mathcal{H}}_M$ and $\boldsymbol{s}$ are given by those arguments that maximize the PDF $f(\boldsymbol{x})$

$$\begin{aligned} (\boldsymbol{\mathcal{H}}_M, \boldsymbol{s}) &= \arg\max_{\boldsymbol{\mathcal{H}}_M, \boldsymbol{s}} f(\boldsymbol{x}|\boldsymbol{h}) && (5)\\ &= \arg\min_{\boldsymbol{\mathcal{H}}_M, \boldsymbol{s}} \left\{\|\boldsymbol{x} - \boldsymbol{\mathcal{H}}_M \boldsymbol{s}\|_2^2\right\} && (6) \end{aligned}$$

where proper constraints on $\boldsymbol{\mathcal{H}}_M$ and $\boldsymbol{s}$ are imposed. Note that such ML criterion is equivalent to the least-squares (LS) criterion, for which the knowledge of the PDF of $\boldsymbol{x}$ is not necessary. For any given $\boldsymbol{\mathcal{H}}_M$, the ML estimate that minimizes the quadratic function $\|\boldsymbol{x} - \boldsymbol{\mathcal{H}}_M \boldsymbol{s}\|_2^2$ is known to be

$$\widehat{\boldsymbol{s}} = \left(\boldsymbol{\mathcal{H}}_M^H \boldsymbol{\mathcal{H}}_M\right)^{-1} \boldsymbol{\mathcal{H}}_M^H \boldsymbol{x}. \tag{7}$$

Under the necessary identifiability condition, the matrix $\boldsymbol{\mathcal{H}}_M$ is known to have full column rank [10]. Using this estimate in equation (6) yields

$$\widehat{\boldsymbol{\mathcal{H}}}_M = \arg\min_{\boldsymbol{\mathcal{H}}_M} \left\{\|(\boldsymbol{I}_M - \boldsymbol{\mathcal{P}}_H)\boldsymbol{x}\|_2^2\right\} \tag{8}$$

where $\boldsymbol{\mathcal{P}}_H$ is the orthogonal projection matrix onto the range of $\boldsymbol{\mathcal{H}}_M$, i.e.

$$\boldsymbol{\mathcal{P}}_H = \boldsymbol{\mathcal{H}}_M \left(\boldsymbol{\mathcal{H}}_M^H \boldsymbol{\mathcal{H}}_M\right)^{-1} \boldsymbol{\mathcal{H}}_M^H.$$

Although the minimization in (8) is computationally much more efficient than that in (6), it is still highly nonlinear. Therefore, the computation of (8) has to be iterative in nature. Many iterative optimization approaches such as [11, 12] can be applied to compute (8). Below, however, we consider a more elegant technique. Define

$$\boldsymbol{\mathcal{G}}_2^H = \left[-\overline{\boldsymbol{H}}_1,\ \overline{\boldsymbol{H}}_2\right] \tag{9}$$

and

$$\boldsymbol{\mathcal{G}}_q^H = \left[\begin{array}{ccc|c} & \boldsymbol{\mathcal{G}}_{q-1}^H & & \boldsymbol{0} \\ \hline -\overline{\boldsymbol{H}}_q & & \boldsymbol{0} & \overline{\boldsymbol{H}}_1 \\ & \ddots & & \vdots \\ \boldsymbol{0} & & -\overline{\boldsymbol{H}}_q & \overline{\boldsymbol{H}}_{q-1} \end{array}\right] \tag{10}$$

where $q = 3, \ldots, M$ and $\overline{\boldsymbol{H}}_q$ is the top-left $(T-L) \times T$ submatrix of $\boldsymbol{H}_q$. Then, provided that all channels do not share a common zero and $T \geq 2L$ (for $M = 2$, only $T > L$ is required), an orthogonal complement matrix of the generalized Sylvester matrix $\boldsymbol{\mathcal{H}}_M$ is $\boldsymbol{\mathcal{G}}_M$, i.e.

$$\boldsymbol{\mathcal{P}}_H + \boldsymbol{\mathcal{P}}_G = \boldsymbol{I}$$

where $\boldsymbol{\mathcal{P}}_H$ and $\boldsymbol{\mathcal{P}}_G$ denote the orthogonal projection matrices onto range $\boldsymbol{\mathcal{G}}_M$ and $\boldsymbol{\mathcal{H}}_M$, respectively. Under this condition, the minimization of (8) becomes

$$\begin{aligned} \widehat{\boldsymbol{h}} &= \arg\min_{\boldsymbol{h}} \left\{\|\boldsymbol{\mathcal{P}}_G \boldsymbol{x}\|_2^2\right\} && (11)\\ &= \arg\min_{\boldsymbol{h}} \left\{\boldsymbol{x}^H \boldsymbol{\mathcal{P}}_G \boldsymbol{x}\right\} && (12)\\ &= \arg\min_{\boldsymbol{h}} \left\{\boldsymbol{x}^H \boldsymbol{\mathcal{G}}_M \left(\boldsymbol{\mathcal{G}}_M^H \boldsymbol{\mathcal{G}}_M\right)^{\#} \boldsymbol{\mathcal{G}}_M^H \boldsymbol{x}\right\} && (13) \end{aligned}$$

where $\boldsymbol{h}$ is the vector of all channels impulse responses and the superscript $(\cdot)^{\#}$ denotes a Moore-Penrose pseudoinverse operator. The following is a matrix form of the commutativity property of linear convolution :

$$\boldsymbol{\mathcal{G}}_M \boldsymbol{x} = \boldsymbol{\mathcal{X}}_M \boldsymbol{h} \tag{14}$$

where $\boldsymbol{\mathcal{X}}_M$ is defined by :

$$\boldsymbol{\mathcal{X}}_2 = [\boldsymbol{X}_2,\ -\boldsymbol{X}_1] \tag{15}$$

and

$$\boldsymbol{\mathcal{X}}_q = \left[\begin{array}{ccc|c} & \boldsymbol{\mathcal{X}}_{q-1} & & \boldsymbol{0} \\ \hline \boldsymbol{X}_q & & \boldsymbol{0} & -\boldsymbol{X}_1 \\ & \ddots & & \vdots \\ \boldsymbol{0} & & \boldsymbol{X}_q & -\boldsymbol{X}_{q-1} \end{array}\right]$$

with $q = 3, \ldots, M$ and :

$$\boldsymbol{X}_q = \begin{bmatrix} x_q(L) & \cdots & x_q(0) \\ \vdots & & \vdots \\ x_q(T-1) & \cdots & x_q(T-L-1) \end{bmatrix}.$$

Combining (13) with (14) yields

$$\widehat{\boldsymbol{h}} = \arg\min_{\boldsymbol{h}} \left\{\boldsymbol{h}^H \boldsymbol{\mathcal{X}}_M^H \left(\boldsymbol{\mathcal{G}}_M^H \boldsymbol{\mathcal{G}}_M\right)^{\#} \boldsymbol{\mathcal{X}}_M \boldsymbol{h}\right\} \tag{16}$$

This expression suggests the following two-step ML method.

- Step 1 : $\widehat{\boldsymbol{h}}_c = \arg\min_{\|\boldsymbol{h}\|_2 = 1} \left\{\boldsymbol{h}^H \boldsymbol{\mathcal{X}}_M^H \boldsymbol{\mathcal{X}}_M \boldsymbol{h}\right\}$
- Step 2 : $\widehat{\boldsymbol{h}}_e = \arg\min_{\|\boldsymbol{h}\|_2 = 1} \left\{\boldsymbol{h}^H \boldsymbol{\mathcal{X}}_M^H \left(\boldsymbol{\mathcal{G}}_c^H \boldsymbol{\mathcal{G}}_c\right)^{\#} \boldsymbol{\mathcal{X}}_M \boldsymbol{h}\right\}$, where $\boldsymbol{\mathcal{G}}_c$ is $\boldsymbol{\mathcal{G}}_M$ constructed from $\widehat{\boldsymbol{h}}_c$ according to equations (9) and (10).

The first step comes from (16) by setting the weighting matrix $\left(\boldsymbol{\mathcal{G}}_M^H \boldsymbol{\mathcal{G}}_M\right)^{\#}$ to an identity matrix. It can be shown that Step 1 of the algorithm yields the exact estimate of $\boldsymbol{h}$ in the absence of noise (or when the noise is white and the data length is infinite) and that Step 2 of the algorithm yields the optimum (ML) estimate of $\boldsymbol{h}$ at a relatively high signal-to-noise ratio (SNR).

4. MAXIMUM A POSTERIORI APPROACH

In this section, we introduce a Maximum a Posteriori (MAP) probability approach which estimates the probability distribution of $\boldsymbol{h}$ as follows

$$\widehat{\boldsymbol{h}}_{MAP} = \arg\max_{\boldsymbol{h}} \left\{ \frac{f(\boldsymbol{x}|\boldsymbol{h}) f(\boldsymbol{h})}{\int f(\boldsymbol{x}|\boldsymbol{h}') f(\boldsymbol{h}') d\boldsymbol{h}'} \right\} \quad (17)$$

$$= \arg\max_{\boldsymbol{h}} \{ f(\boldsymbol{x}|\boldsymbol{h}) f(\boldsymbol{h}) \} \quad (18)$$

To approximate the channel distribution in (18), we adopt the generalized Laplacian distribution model, which can be mathematically represented as follows :

$$f(\boldsymbol{h}) = \left[2\beta\Gamma\left(1+\frac{1}{p}\right) \right]^{-M(L+1)} \exp\left(-\frac{\|\boldsymbol{h}\|_p^p}{\beta^p} \right) \quad (19)$$

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.
The new prior distribution gives more weight to values that are close to zero, thereby encouraging the model to set many latent variables to (or close to) zero. This makes it ideal for learning sparse representations.
The combination of equations (18) and (19) leads to the following objective function :

$$\mathcal{J}(\boldsymbol{h}) = \boldsymbol{h}^H \boldsymbol{\mathcal{X}}_M^H \left(\boldsymbol{\mathcal{G}}_c^H \boldsymbol{\mathcal{G}}_c \right)^{\#} \boldsymbol{\mathcal{X}}_M \boldsymbol{h} + \lambda \|\boldsymbol{h}\|_p^p \quad (20)$$

where $\lambda = \sigma^2/\beta^p$ is a weighting parameter which controls the trade-off between approximation error and sparsity. The first term is the ML criterion and the second term is the penalty term, which minimizes the ℓ_p norm of the channel impulse response $\boldsymbol{h}$. It is well known that the concavity of this ℓ_p norm function yields the sparse solution [13].
Therefore, the desired solution of $\boldsymbol{h}$ is determined by minimizing the cost function $\mathcal{J}(\boldsymbol{h})$ under the unit norm constraint $\|\boldsymbol{h}\|_2 = 1$:

$$\widehat{\boldsymbol{h}} = \arg \min_{\|\boldsymbol{h}\|_2 = 1} \left\{ \boldsymbol{h}^H \boldsymbol{\mathcal{X}}_M^H \left(\boldsymbol{\mathcal{G}}_c^H \boldsymbol{\mathcal{G}}_c \right)^{\#} \boldsymbol{\mathcal{X}}_M \boldsymbol{h} + \lambda \|\boldsymbol{h}\|_p^p \right\}. \quad (21)$$

Direct minimization is computationally intensive and may be even intractable when the channel impulse responses are long and the number of channels is large. Here, a two step stochastique gradient technique is proposed to solve this minimization problem efficiently :
Step 1 : Similarly to the previous ML solution, we set the weighting matrix $\left(\boldsymbol{\mathcal{G}}_M^H \boldsymbol{\mathcal{G}}_M \right)^{\#}$ to the identity. Consequently, the cost function becomes :

$$\mathcal{J}_1(\boldsymbol{h}) = \boldsymbol{h}^H \boldsymbol{\mathcal{X}}_M^H \boldsymbol{\mathcal{X}}_M \boldsymbol{h} + \lambda \|\boldsymbol{h}\|_p^p. \quad (22)$$

By using a stochastic gradient minimization, we can write the solution as :

$$\boldsymbol{h}_{k+1} = \boldsymbol{h}_k - \mu \nabla \mathcal{J}_1(\boldsymbol{h}_k) \quad (23)$$

where μ is a small positive step size and ∇ is a gradient operator. The gradient of $\mathcal{J}_1(\boldsymbol{h})$ is given by :

$$\nabla \mathcal{J}_1(\boldsymbol{h}) = \frac{\partial \mathcal{J}_1(\boldsymbol{h})}{\partial \boldsymbol{h}} = 2\, \boldsymbol{\mathcal{X}}_M^H \boldsymbol{\mathcal{X}}_M \boldsymbol{h} + \lambda\, p\, \widetilde{\boldsymbol{h}} \quad (24)$$

where

$$\widetilde{h}(i) = \text{sign}(h(i))\ |h(i)|^{p-1}. \quad (25)$$

The unit norm constraint is to ensure that the iterative algorithm does not converge to a trivial solution with all zero elements. It is applied after each updating step. Therefore the update equation is given by :

$$\boldsymbol{h}_{k+1} = \frac{\boldsymbol{h}_k - \mu \left(2\, \boldsymbol{\mathcal{X}}_M^H \boldsymbol{\mathcal{X}}_M \boldsymbol{h}_k + \lambda\, p\, \widetilde{\boldsymbol{h}}_k \right)}{\left\| \boldsymbol{h}_k - \mu \left(2\, \boldsymbol{\mathcal{X}}_M^H \boldsymbol{\mathcal{X}}_M \boldsymbol{h}_k + \lambda\, p\, \widetilde{\boldsymbol{h}}_k \right) \right\|_2}. \quad (26)$$

Step 2 : According to the ML approach, we use the first step channel estimate to compute the weighting matrix $\left(\boldsymbol{\mathcal{G}}_M^H \boldsymbol{\mathcal{G}}_M \right)^{\#}$ leading to the cost function :

$$\mathcal{J}_2(\boldsymbol{h}) = \boldsymbol{h}^H \boldsymbol{\mathcal{X}}_M^H \left(\boldsymbol{\mathcal{G}}_M^H \boldsymbol{\mathcal{G}}_M \right)^{\#} \boldsymbol{\mathcal{X}}_M \boldsymbol{h} + \lambda \|\boldsymbol{h}\|_p^p. \quad (27)$$

Therefore, by using a stochastic gradient minimization, the iterative solution can be written as :

$$\boldsymbol{h}_{k+1} = \boldsymbol{h}_k - \mu \nabla \mathcal{J}_2(\boldsymbol{h}_k). \quad (28)$$

In order to determine the gradient in equation (28), we take a derivative of $\mathcal{J}_2(\boldsymbol{h})$ with respect to $\boldsymbol{h}$:

$$\nabla \mathcal{J}_2(\boldsymbol{h}) = 2\, \boldsymbol{\mathcal{X}}_M^H \left(\boldsymbol{\mathcal{G}}_M^H \boldsymbol{\mathcal{G}}_M \right)^{\#} \boldsymbol{\mathcal{X}}_M \boldsymbol{h} + \lambda\, p\, \widetilde{\boldsymbol{h}}, \quad (29)$$

and thus, the update equation is given by :

$$\boldsymbol{h}_{k+1} = \frac{\boldsymbol{h}_k - \mu \left(2\, \boldsymbol{\mathcal{X}}_M^H \left(\boldsymbol{\mathcal{G}}_M^H \boldsymbol{\mathcal{G}}_M \right)^{\#} \boldsymbol{\mathcal{X}}_M \boldsymbol{h}_k + \lambda\, p\, \widetilde{\boldsymbol{h}}_k \right)}{\left\| \boldsymbol{h}_k - \mu \left(2\, \boldsymbol{\mathcal{X}}_M^H \left(\boldsymbol{\mathcal{G}}_M^H \boldsymbol{\mathcal{G}}_M \right)^{\#} \boldsymbol{\mathcal{X}}_M \boldsymbol{h}_k + \lambda\, p\, \widetilde{\boldsymbol{h}}_k \right) \right\|_2}.$$

4.1 Optimal step size

In order to avoid divergence, a conservatively small μ is usually used, which inevitably sacrifies the convergence speed of the iterative algorithm. In this section, we will derive an optimal step size for the MAP algorithm and hence propose a variable step size MAP algorithm. To find an optimal step size μ for each iteration we propose to use a line search method. More precisely, we choose a line search, in which μ is chosen to minimize $\mathcal{J}_i,\ i = 1, 2$

$$\mu = \arg\min_{\mu} \{ \mathcal{J}_i (\boldsymbol{h} - \mu \nabla \mathcal{J}_i(\boldsymbol{h})) \} \quad (30)$$

The criterion in the $(k+1)^{th}$ iteration is written as follow :

$$\mathcal{J}_i(\boldsymbol{h}_{k+1}) = \boldsymbol{h}_{k+1}^H \boldsymbol{\mathcal{Q}}_i \boldsymbol{h}_{k+1} + \lambda \|\boldsymbol{h}_{k+1}\|_p^p \quad (31)$$

with $\boldsymbol{\mathcal{Q}}_1 = \boldsymbol{\mathcal{X}}_M^H \boldsymbol{\mathcal{X}}_M$ and $\boldsymbol{\mathcal{Q}}_2 = \boldsymbol{\mathcal{X}}_M^H \left(\boldsymbol{\mathcal{G}}_M^H \boldsymbol{\mathcal{G}}_M \right)^{\#} \boldsymbol{\mathcal{X}}_M$. By replacing $\boldsymbol{h}_{k+1}$ by (23) or (28) we rewrite equation (31) as :

$$\mathcal{J}_i(\boldsymbol{h}_{k+1}) = (\boldsymbol{h}_k - \mu \nabla \mathcal{J}_i(\boldsymbol{h}_k))^H \boldsymbol{\mathcal{Q}}_i (\boldsymbol{h}_k - \mu \nabla \mathcal{J}_i(\boldsymbol{h}_k)) + \lambda \|\boldsymbol{h}_k - \mu \nabla \mathcal{J}_i(\boldsymbol{h}_k)\|_p^p \quad (32)$$

we take a derivative of $\mathcal{J}_i(\boldsymbol{h}_{k+1})$ with respect to μ :

$$\frac{\partial \mathcal{J}_i(\boldsymbol{h}_{k+1})}{\partial \mu} = \mathcal{F}_i(\mu) =$$

$$\left[\mu(2\nabla \mathcal{J}_i(\boldsymbol{h}_k)^H \boldsymbol{\mathcal{Q}}_i - \lambda\ p\ \widetilde{\boldsymbol{r}}_k^H) - 2\boldsymbol{h}_k^H \boldsymbol{\mathcal{Q}}_i\right] \nabla \mathcal{J}_i(\boldsymbol{h}_k)$$

where

$$\widetilde{r}(i) = sign\left(h(i) - \mu \nabla \mathcal{J}(\boldsymbol{h})(i)\right) |h(i) - \mu \nabla \mathcal{J}(\boldsymbol{h})(i)|^{p-1}.$$

Therefore, the optimal step size in each iteration is obtained in the form :

$$\mu_k = \mu_{k-1} - \mathcal{F}_i(\mu_{k-1}) \frac{\mu_{k-1} - \mu_{k-2}}{\mathcal{F}_i(\mu_{k-1}) - \mathcal{F}_i(\mu_{k-2})}\ , \quad (33)$$

where we use an approximate Newton approach for solving (30).

5. SIMULATION

We present here some numerical simulations to assess the performance of the proposed algorithm. We consider a SIMO system with $M = 3$ outputs represented by polynomial transfer function of degree $L = 64$. The channel impulse response is a sparse sequence of random variables with Bernoulli-Gaussian distribution [8] :

$$f(h_i) = p_i \delta(h_i) + (1 - p_i) \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-h_i^2 / 2\sigma_i^2\right)$$

generated by the MATLAB function SPRANDN. We used the parameters $p_i = 0.5$ and $\sigma_i = 1$. The input signal is a 4-QAM i.i.d. sequence of length $T = 512$. The observation is corrupted by addition white Gaussian noise with a variance σ^2 chosen such that the $SNR = \frac{\|\boldsymbol{h}\|^2}{\sigma^2}$ varies in the range $[5, 50]\ dB$. The weighting parameter for the MAP algorithm is chosen as $\lambda = 1$. Statistics are evaluated over $N_r = 200$ Monte-Carlo runs and estimation performance are given by the normalized mean-square error criterion :

$$\begin{aligned} NMSE &= \frac{1}{N_r} \sum_{r=1}^{N_r} \min_{\alpha} \left(\frac{\|\alpha \widehat{\boldsymbol{h}}_r - \boldsymbol{h}\|^2}{\|\boldsymbol{h}\|^2} \right) \\ &= \frac{1}{N_r} \sum_{r=1}^{N_r} 1 - \left(\frac{\widehat{\boldsymbol{h}}_r^H \boldsymbol{h}}{\|\widehat{\boldsymbol{h}}_r\| \|\boldsymbol{h}\|} \right)^2 , \end{aligned}$$

where $\widehat{\boldsymbol{h}}_r$ denotes the estimated channel coefficient vector at the r^{th} Monte-Carlo run and α is a scalar factor that compensates for the scale indeterminacy of the BSI problem.

In figure 1, the normalized mean-square error is plotted versus the SNR for the proposed algorithm and the ML algorithm. It is clearly shown that our algorithm (MAP) performs better in terms of the normalized mean-square error especially for moderate and high SNR.

In figure 2, we represent the evolution of the cost function in dB as a function of the iteration number for the gradient with fixed and optimal step size techniques. It is shown that the optimal step size technique converges much faster than the fixed step size one.

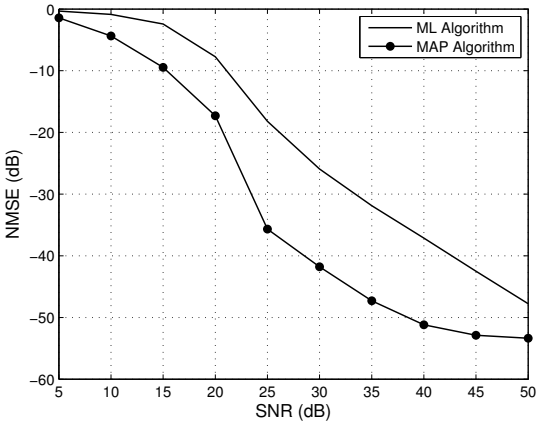

FIG. 1 – Normalized mean-square error (NMSE) versus the SNR for SIMO system with 3 sensors : comparison between ML and the proposed MAP algorithm.

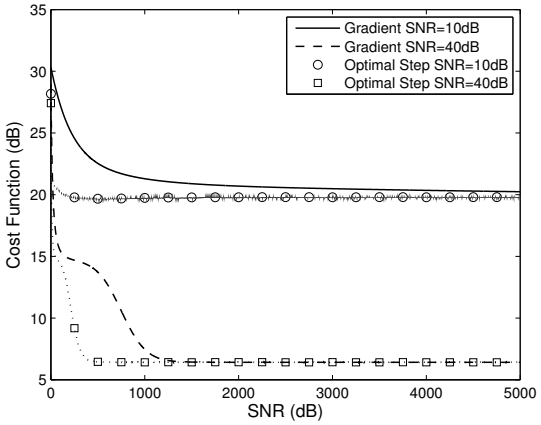

FIG. 2 – Evolution of the cost function in dB as a function of the iteration number.

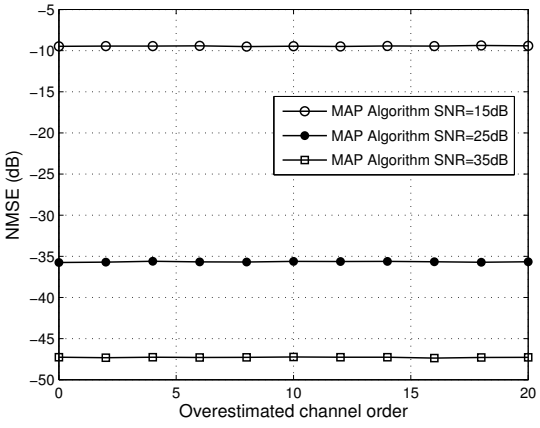

FIG. 3 – Normalized mean-square error (NMSE) versus the overestimated channel order for SIMO system with 3 sensors and for different value of the SNR.

In figure 3, we represent the evolution of the NMSE in dB as a function of the overestimated channel order for the MAP algorithm. This figure illustrates the robustness of our algorithm against channel order overestimation errors.

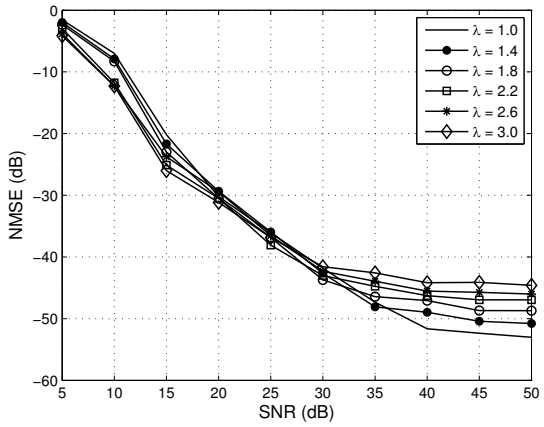

FIG. 4 – Normalized mean-square error (NMSE) versus the SNR for SIMO system with 3 sensors : performance of our MAP algorithm for different value of the regularization parameter λ.

In figure 4, we represent the NMSE as a function of the SNR for different values of the weighting parameter λ. We observe that, for large SNRs, small λ values are preferred, while for low SNRs, the large λ values are those leading to the best channel estimation accuracy. From this observation we plan for our futur works to study the optimization of the parameter λ in the MAP algorithm.

6. CONCLUSION

This paper introduces a generalized version of the ML method for the blind estimation of sparse and long SIMO channel impulse responses. In the proposed method, we use a channel sparsity measure together with the ML criterion to improve the estimation quality and to take into account the sparsity of the channel. A gradient type technique with optimized step size has been considered for the optimization of the proposed cost function. Besides its improved performance, the new BSI method is robust against channel order overestimation errors.

REFERENCES

[1] J. A. Cadzow, "Blind deconvolution via cumulant extrema," *IEEE Signal Processing Magazine*, vol. 6, no. 13, pp. 24–42, May 1996.

[2] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1310–1322, August 1997.

[3] L. Tong and S. Perreau, "Multichannel blind identification : from subspace to maximum likelihood methods," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1951–1968, October 1998.

[4] 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, December 1995.

[5] W. Schreiber, "Advanced television systems for terrestrial broadcasting : Some problems and some proposed solutions," *Proceedings of the IEEE*, vol. 83, no. 6, pp. 958–981, June 1995.

[6] S. Ariyavisitakul, N. Sollenberger, and L. Greenstein, "Tap selectable decision-feedback equalization," *IEEE Transactions on Communications*, vol. 45, no. 12, pp. 1497–1500, December 1997.

[7] M. Kocic, D. Brady, and M. Stojanovic, "Sparse equalization for real-time digital underwater acoustic communications," in *Proc. OCEANS*, San Diego, USA, October 1995, vol. 3, pp. 1417–1422.

[8] M. Zibulevsky and Y. Zeevi, "Extraction of a source from multichannel data using sparse decomposition," *Neurocomputing*, vol. 49, no. 1, pp. 163–173, December 2002.

[9] A. Ossadtchi and S. Kadambe, "Over-complete blind source separation by applying sparse decomposition and information theoretic based probabilistic approach," in *Proc. ICASSP*, 2001, vol. 5, pp. 2801–2804.

[10] A. Hua and M. Wax, "Strict identifiability of multiple FIR channels driven by an unknown arbitrary sequence," *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 756–759, March 1996.

[11] K. Abed-Meraim and E. Moulines, "A maximum likelihood solution to blind identification of multichannel FIR filters," in *Proc. EUSIPCO*, Edinburgh, Scotland, 1994, vol. 2, pp. 1011–1014.

[12] D. Starer and A. Nehorai, "Passive localization of near-field sources by path following," *IEEE Transactions on Signal Processing*, vol. 42, no. 3, pp. 677–680, March 1994.

[13] K. Kreutz-Delgado and B. D. Rao, "Sparse basis selection, ICA, and majorization : towards a unified perspective," in *Proc. ICASSP*, March 1999, vol. 2, pp. 1801–1804.