ROBUSTNESS OF BLIND SUBSPACE BASED TECHNIQUES USING ℓ_p QUASI-NORMS

Abla Kammoun[1] *Abdeljalil Aissa El Bey*[2], *Karim Abed-Meraim*[1] *and Sofiène Affès*[3]

[1] Télecom ParisTech France, [2] Télécom-Bretagne, [3] INRS-EMT Canada

ABSTRACT

It has been very recently noted that it is possible to recover the blind channel estimate in case of channel order overmodeling by using ℓ_p quasi norms. But, to the best of our knowledge, there is, until now, no theoretical results that investigate this issue. In this paper, we propose to study the robustness of subspace blind methods using ℓ_p quasi-norms in the noiseless case and for nonsparse channels. More particularly, we provide conditions that ensures channel identifiability and study their frequency of occurence with respect to the system parameters.

1. INTRODUCTION

In current communication systems, channel estimation is essential, since it enables data detection without the 3dB loss incurred in the case of noncoherent estimation. Over the last decades, a special interest has been devoted to blind channel estimation techniques for their high spectral efficiency as compared to their training-based counterparts. As long as the channel order is correctly estimated, the channel can be uniquely identified using blind methods, but once an error on the estimation of the channel order occurs, identifiability is no longer possible for many existing blind methods. This is for instance the case of conventional subspace-based methods, which are known to exhibit a significant sensitivity to channel order overmodeling [1]. Actually, in the noiseless case, the channel can be identified as the vector that spans the 1-dimensional kernel of a matrix denoted by $\mathbf{Q}$ which can be estimated by using solely second-order statistics. But when the channel order is overestimated, the kernel of the matrix $\mathbf{Q}$ is no longer a line but rather a vector space whose dimension depends on the overestimated order, thereby raising a new issue: how to estimate the right direction among all the vectors that span the kernel of $\mathbf{Q}$?

To deal with this problem, a large effort was devoted to either add to conventional subspace techniques a feature that estimates efficiently the channel order [2], or to propose new methods that are robust to channel-order overmodeling. In this context, a new technique for blind channel estimation of sparse channels has been recently proposed. This technique takes into account the sparsity criterion so as to select among the possible vectors the vector that exhibits the lowest ℓ_p quasi-norm $0 < p \leq 1$. It was noted that using this technique, Cross relation as well as blind deterministic maximum likelihood based methods become robust to channel-order overmodeling as far as sparse channels are concerned, [3, 4]. However, for nonsparse channels, no results are available so far, to the best of our knowledge. Yet, we strongly believe that introducing likewise a sparsity criterion shall enhance the channel identifiability probability. Actually in this case, one can note that overmodeling the channel is equivalent to zero-padding the channel vector, thus making it artificially sparse. Moreover, as far as subspace methods are concerned, it can be shown that the zero-padded channel vector is the one that exhibits the most sparsity. In light of this consideration, we claim that selecting the vector that minimizes the ℓ_p quasi-norm should often yield the desired channel vector response (up to a scalar ambiguity).

In this paper, we propose to study the robustness of subspace methods using ℓ_p quasi-norms for nonsparse channels. We derive the necessary and sufficient condition for channel identifiability when considering the ℓ_1 norm as well as a sufficient condition when considering the ℓ_p quasi-norm $0 < p < 1$. Then, we derive lower bounds on the probability that these conditions are satisfied. Using these lower bounds, we study the effect of the system parameters on the channel identifiability probability. For instance, we note that in the ℓ_1 norm problem, increasing the number of antennas improves significantly the channel identifiability probability, in contrast to increasing the number of channel coefficients, which tends to reduce it.

2. A BRIEF REVIEW ON SUBSPACE-BASED METHODS FOR SIMO SYSTEMS

For the reader's convenience, we review hereafter the subspace based method for Single Input Multiple Output (SIMO) systems [1].

In a SIMO system, if s_k denotes the unit-power transmitted signal, the M receiving antennas observe the following signal:

$$\mathbf{y}_k = \sum_{l=0}^{L} \mathbf{h}_l s_{k-l} + \mathbf{v}_k,$$

where $\mathbf{h}_l$ is the channel impulse response vector at the l-th tap and $\mathbf{v}_k$ denotes the additive Gaussian noise. Let $\mathbf{h} = [\mathbf{h}_0^{\mathrm{T}}, \cdots, \mathbf{h}_L^{\mathrm{T}}]^{\mathrm{T}}$ be the channel vector parameter. Stacking n observations of vector $\mathbf{y}_k$ in a $(n+1)M$ vector $\overline{\mathbf{y}}_k = [\mathbf{y}_k^{\mathrm{T}}, \cdots, \mathbf{y}_{k-n}^{\mathrm{T}}]^{\mathrm{T}}$, we will get:

$$\overline{\mathbf{y}}_k = \boldsymbol{\mathcal{I}}_n(\mathbf{h})\mathbf{s}_k + \mathbf{v}_k,$$

where $\boldsymbol{\mathcal{I}}_n(\mathbf{h})$ is the $M(n+1) \times (L+n+1)$ block-toeplitz matrix :

$$\boldsymbol{\mathcal{I}}_n(\mathbf{h}) = \begin{pmatrix} \boxed{\mathbf{h}_0 \cdots \mathbf{h}_L} & & 0 \\ & \boxed{\mathbf{h}_0 \cdots \mathbf{h}_L} & \\ & \ddots & \ddots \\ 0 & & \boxed{\mathbf{h}_0 \cdots \mathbf{h}_L} \end{pmatrix}.$$

The covariance matrix of the received signal $\overline{\mathbf{y}}_k$ can be expressed as:

$$\mathbf{R} = \mathbb{E}\overline{\mathbf{y}}_k\overline{\mathbf{y}}_k^{\mathrm{H}} = \boldsymbol{\mathcal{I}}_n(\mathbf{h})\boldsymbol{\mathcal{I}}_n^{\mathrm{H}}(\mathbf{h}) + \sigma^2\mathbf{I}_{(n+1)M}.$$

Assuming that the subchannels of vector $\mathbf{h}$ have no zeros in common and $n \geq L$, the rank of $\boldsymbol{\mathcal{I}}_n(\mathbf{h})$ is equal to $L+n+1$. Hence, there are $L+n+1$ eigenvalues of $\mathbf{R}$ that correspond to the signal subspace (non null eigenvalues of $\mathcal{I}_n(\mathbf{h})$), whereas the remaining eigenvalues correspond to the noise subspace. Denote by $\mathbf{\Pi}$ the orthogonal projector on the noise subspace, and by $\mathcal{D}$ the operator given by:

$$\mathcal{D}: \mathcal{M}_{M(n+1)\times M(n+1)}(\mathbb{C}) \rightarrow \mathcal{M}_{M(n+1)(L+1)\times M(L+1)}(\mathbb{C})$$

$$\mathbf{M} = [\mathbf{M}_0, \cdots, \mathbf{M}_n] \mapsto \begin{pmatrix} \mathbf{M}_0 & & 0 \\ \vdots & \ddots & \\ \mathbf{M}_n & & \mathbf{M}_0 \\ & \ddots & \vdots \\ 0 & & \mathbf{M}_n \end{pmatrix},$$

The blind subspace estimator can be defined as:

$$\widehat{\mathbf{h}} = \min_{\|\mathbf{f}\|_2=1} \mathbf{f}^{\mathrm{H}}\mathbf{Q}\mathbf{f},$$

with:

$$\mathbf{Q} = \mathcal{D}(\mathbf{\Pi})^{\mathrm{H}}\mathcal{D}(\mathbf{\Pi}).$$

In case of channel overmodeling, the kernel of matrix $\mathbf{Q}$ is a vector space with dimension equal to $\delta = L' - L$ (L' being the overestimated order), which is spanned by the channel vector as well as all its $\delta - 1$ delayed copies [1]. In other words, the kernel is spanned by the following $(\delta + L + 1)M \times (\delta + 1)$ block-Toeplitz matrix $\mathbf{H}$:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_0 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{h}_L & \ddots & \mathbf{h}_0 \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{h}_L \end{bmatrix}$$

3. CONDITIONS FOR CHANNEL IDENTIFIABILITY

One can note that the columns of the matrix $\mathbf{H}$ represent the sparsest vectors of the kernel of matrix $\mathbf{Q}$. In fact, any linear combination of vectors of $\mathbf{H}$ will yield almost surely vectors that are less sparse as they contain less zeros. Hence, the channel vector can be selected as the one that solves the following combinatorial optimization problem:

$$(P_0) \qquad \min_{\mathbf{x},\mathbf{x}_1=1,\mathbf{Q}\mathbf{x}=\mathbf{0}} \|\mathbf{x}\|_0 \qquad (1)$$

where $\mathbf{x}_1$ denotes the first entry of $\mathbf{x}$, $\|.\|_0$ is the ℓ_0 quasi-norm that returns the number of coefficients where the vector is not equal to zero. However, solving (P_0) requires generally an intractable combinatorial search, thus reducing its interest for real-time applications.

An alternative is to consider the optimization problem:

$$(P_p) \qquad \min_{\mathbf{x},\mathbf{x}_1=1,\mathbf{Q}\mathbf{x}=\mathbf{0}} \|\mathbf{x}\|_p \qquad (2)$$

where $\|\mathbf{x}\|_p$ denotes the ℓ_p quasi-norm: $\|\mathbf{x}\|_p = \left(\sum_i |\mathbf{x}_i|^p\right)^{\frac{1}{p}}$. It should be mentioned that this approach has been extensively studied by the compressed sensing theory [5] and applied to many fields like image processing [6] and communication systems [7, 8]. For all these applications, the problem is usually put under the form:

$$\min_{\mathbf{x},\mathbf{\Phi}\mathbf{x}=\mathbf{y}} \|\mathbf{x}\|_p$$

where $\mathbf{\Phi}$ is a matrix independently distributed from vector $\mathbf{x}$. This is different from our case, since matrix $\mathbf{Q}$ is a function of vector $\mathbf{h}$.

Therefore, all the theoretical results that have been derived in compressed sensing theory should be adapted to our context, and cannot be applied directly. Thereby, in this paper, we propose to make use of the structure of our problem so as to derive new results about the channel identifiability conditions and evaluate their frequency of occurence. Taking into account the structure of our problem, we can deduce that (P_p) is equivalent to:

$$\min_{\mathbf{s},\mathbf{s}_1=1} \|\mathbf{H}\mathbf{s}\|_p^p \Leftrightarrow \min_{\mathbf{s}} \| \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_L \\ \vdots \\ \mathbf{0} \end{bmatrix} + \widetilde{\mathbf{H}}\mathbf{s}\|_p^p$$

where $\widetilde{\mathbf{H}}$ is the $(\delta+L)M \times \delta$ block-Toeplitz matrix that has the same shape as $\mathbf{H}$. Before proceeding, we shall partition $\widetilde{\mathbf{H}} = \left[\frac{\mathbf{A}}{\mathbf{B}}\right]$, where $\mathbf{A}$ (resp. $\mathbf{B}$) represents the first ML (resp. the last $M(\delta+1)$) rows of $\widetilde{\mathbf{H}}$.

3.1. ℓ_1 norm

Unlike the ℓ_p quasi-norm,($p < 1$), the ℓ_1 norm is convex. So in this case, it is possible to derive a necessary and sufficient condition for channel identifiability, which can be stated by the following theorem. For simplicity, we consider here the real case, i.e $\mathbf{h} \in \mathbb{R}^{M(L+1)}$.

Theorem 1. *Necessary and sufficient condition*

Let $\mathbf{v} = \left[\mathrm{sign}(\mathbf{h}_1)^{\mathrm{T}}, \cdots, \mathrm{sign}(\mathbf{h}_L)^{\mathrm{T}}\right]^{\mathrm{T}}$ *and assume that* $L > \delta \geq 1$, *then the necessary and sufficient condition for channel identifiability can be expressed as:*

$$\frac{|\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{s}|}{\|\mathbf{B}\mathbf{s}\|_1} \leq 1 \;\; \forall \mathbf{s} \in \mathbb{R}^{\delta}.$$

3.2. ℓ_p quasi-norm

Since the ℓ_p quasi-norm is a nonconvex function, the problem might have many local minima. Nevertheless, We still can find a sufficient condition that ensures that the channel can be identified as a local minimum of (2). This result is stated in the following theorem:

Theorem 2. *Sufficient condition*

Let $\mathbf{v} = p \begin{bmatrix} \mathrm{sign}(\mathbf{h}_1) \\ \vdots \\ \mathrm{sign}(\mathbf{h}_L) \end{bmatrix} \bullet \begin{bmatrix} |\mathbf{h}_1|^{p-1} \\ \vdots \\ |\mathbf{h}_L|^{p-1} \end{bmatrix}$ *where* $\bullet$ *denotes the Hadamard (element by element) product. If the following condition is satisfied:*

$$\frac{|\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{s}|}{\|\mathbf{B}\mathbf{s}\|_1} \leq 1 \;\; \forall \mathbf{s} \in \mathbb{R}^{\delta} \qquad (3)$$

then the channel can be identified as a local minimum of (2).

4. PROBABILISTIC ANALYSIS

In this section we will study the effect of the system parameters on the channel identifiability probability. We assume that the channel coefficients are drawn from the Gaussian distribution with mean 0 and variance $\frac{1}{L+1}$. To determine a lower bound on the channel identifiability probability, we will rely on the techniques derived in [9, 10]. Actually, in the same way as [9], we recast the probability conditions in an other form as stated by the following theorem:

Theorem 3. *Let* $\mathbf{d}^*$ *be the value that minimizes :*

$$\min \ \|\mathbf{d}\|_\infty$$
$$\text{subject to } \mathbf{B}^{\mathrm{T}}\mathbf{d} = \mathbf{A}^{\mathrm{T}}\mathbf{v}$$

Then, the channel can be identified if and only if $\|\mathbf{d}\|_\infty \leq 1$.

The new formulation given by theorem 3 is interesting in the sense that it allows geometric interpretation of the channel identifiability condition. Actually, it follows from theorem 3 that the channel identifiability holds for a given channel realization if and only if there is a vector $\mathbf{d}$ on the cube $Q = [-1, 1]^{\delta M}$ such that $\mathbf{B}^{\mathrm{T}}\mathbf{d} = \mathbf{A}^{\mathrm{T}}\mathbf{v}$, i.e, $\mathbf{A}^{\mathrm{T}}\mathbf{v}$ belongs to the image of the cube generated by $\mathbf{B}^{\mathrm{T}}$. Since $\mathrm{rank}(\mathbf{B}) = \delta$ almost surely, the channel identifiability will hold if the following conditions are satisfied:

- The image of the cube by $\mathbf{B}^{\mathrm{T}}$ contains a ball of radius α
- The vector $\mathbf{A}^{\mathrm{T}}\mathbf{v}$ satisfies $\|\mathbf{A}^{\mathrm{T}}\mathbf{v}\|_2 \leq \alpha$.

Let $\mathcal{P}$ denote the probability that the channel identifiability holds and E_α^1 and E_α^2 be the events given by:

$$E_\alpha^1 = \left\{\text{The image of the cube by } \mathbf{B}^{\mathrm{T}} \text{contains a ball of radius } \alpha\right\},$$
$$E_\alpha^2 = \left\{\|\mathbf{A}^{\mathrm{T}}\mathbf{v}\|_2 \leq \alpha\right\}.$$

Then, $\mathcal{P}$ can be lower bounded as:

$$\mathcal{P} \geq \mathbb{P}\left\{\bigcup_\alpha E_\alpha^1 \cap E_\alpha^2\right\} \geq \max_\alpha \mathbb{P}\left(E_\alpha^1 \cap E_\alpha^2\right).$$

4.1. ℓ_1 norm

In the following, we propose to determine a lower bound on the probability of the events E_α^1 and E_α^2, while considering the ℓ_1 norm minimization. We will consider first the relatively easy case $\delta = 1$ and after that the more general case $\delta \leq \min(M, L-1)$. For $\delta \geq M$, we have not been able to derive a lower bound on the probability of channel identifiability, but we conjecture that the effect of the system parameters remains the same.

Before going any further, let us, first, write the event E_α^1 in an other equivalent way [9]:

$$E_\alpha^1 = \left\{\min_{\mathbf{x}} \frac{\|\mathbf{B}\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \geq \alpha\right\}.$$

4.1.1. Case when $\delta = 1$

When $\delta = 1$, it is easy to see that $\mathbf{v}^{\mathrm{T}}\mathbf{A}$ is a real standard Gaussian random variable with mean 0 and variance $\frac{LM}{L+1}$. Hence we have;

$$\mathbb{P}\left(E_\alpha^2\right) = \mathbb{P}\left(|\mathbf{v}^{\mathrm{T}}\mathbf{A}| \leq \alpha\right) \tag{4}$$
$$= \mathbb{P}\left(|\mathbf{v}^{\mathrm{T}}\mathbf{A}|^2 \leq \alpha^2\right) = \frac{1}{\sqrt{\pi}}\gamma\left(\frac{1}{2}, \frac{\alpha^2(L+1)}{2LM}\right) \tag{5}$$

where $\gamma(a, x)$ is the lower incomplete gamma function given by:

$$\gamma(a, x) = \int_0^x \exp(-t)t^{a-1}dt.$$

On the other hand, using standard concentration inequalities for normal variables [11], we show that, for every $\epsilon \in [0, 1]$, we have :

$$\mathbb{P}\left(\|\mathbf{h}_L\|_1 \geq M\sqrt{\frac{2}{\pi(L+1)}}(1-\epsilon)\right) \geq 1 - \exp\left(-\frac{M\epsilon^2}{\pi}\right). \tag{6}$$

Since for $\delta = 1$, the events E_α^1 and E_α^2 are independent, we get after combining (6) and (5), and setting $\alpha = M\sqrt{\frac{2}{\pi(L+1)}}(1-\epsilon)$, the following theorem:

Theorem 4. *For* $\delta = 1$, *the probability* $\mathcal{P}$ *that channel identifiability occurs is greater than:*

$$\mathcal{P} \geq \max_{\epsilon \in [0,1]} \left(1 - \exp\left(-\frac{M\epsilon^2}{\pi}\right)\right) \frac{1}{\sqrt{\pi}}\gamma\left(\frac{1}{2}, \frac{2M}{L}(1-\epsilon)^2\right) \tag{7}$$

Remark 1. Under the assumption that the random variables $\|\mathbf{v}^{\mathrm{T}}\mathbf{A}\|_2$ and $\|\mathbf{h}_L\|_1 = \min_{\mathbf{x}} \frac{\|\mathbf{B}\mathbf{x}\|_1}{\|\mathbf{x}\|_2}$ are concentrated around their expected values with high probability (this assumption is valid in general for standard random distributions), one can understand intuitively the effect of the system parameters M and L on the probability for channel identifiability. Actually, given that $\mathbb{E}\|\mathbf{h}_L\|_1 = M\sqrt{\frac{2}{\pi(L+1)}}$, we deduce that we can find, a ball of radius r_1 of order $\mathcal{O}\left(\frac{M}{\sqrt{L}}\right)$ that is contained in the image of the cube Q by $\mathbf{B}^{\mathrm{T}}$ with a high probability. In the same way, given that the expected value of $|\mathbf{v}^{\mathrm{T}}\mathbf{A}|$ is of the order $\mathcal{O}\left(\sqrt{M}\right)$, we can find a ball of radius $r_2 = \mathcal{O}\left(\sqrt{M}\right)$ that contains the vector $\mathbf{v}^{\mathrm{T}}\mathbf{A}$, with a high probability. Since channel identifiability occurs when $r_1 \geq r_2$, we deduce that as M increases, and L decreases, channel identifiability should be more likely to occur.

4.1.2. Case when $\delta > 1$ and $\delta \leq \min(L-1, M)$

When $\delta > 1$, the problem becomes more difficult, since $\mathbf{v}^{\mathrm{T}}\mathbf{A}$ is no longer Gaussian and $\min_{\mathbf{x}} \frac{\|\mathbf{B}\mathbf{x}\|_1}{\|\mathbf{x}\|_2}$ has no closed-form expression. Besides E_α^1 and E_α^2 are no longer independent, thus making our computations less tighter. But, as we can see later, even if the lower bound probability is too loose, one can still draw conclusions about the effect of the system parameters on the channel identifiability probability.

Let us now deal with the probability of the event E_α^1.

$$\mathbb{P}\left(E_\alpha^1\right) = \left\{\min_{\mathbf{x}} \frac{\|\mathbf{B}\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \geq \alpha\right\}.$$

Since $\delta \leq \min(L-1, M)$, it can be shown that:

$$\min_{\mathbf{x}} \frac{\|\mathbf{B}\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \geq \min_{\mathbf{x}} \frac{\|\widetilde{\mathbf{B}}\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \tag{8}$$

where $\widetilde{\mathbf{B}} = [\mathbf{h}_L, \cdots, \mathbf{h}_{L-\delta+1}]$. Consequently,

$$\mathbb{P}\left(E_\alpha^1\right) \geq \mathbb{P}\left\{\min_{\mathbf{x}} \frac{\|\widetilde{\mathbf{B}}\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \geq \alpha\right\}. \tag{9}$$

To determine a lower bound on the probability $\mathbb{P}\left(E_\alpha^1\right)$, we will use the following result:

Theorem 5. *[12] Let* Φ *be a* $M \times \delta$ *Gaussian matrix, with iid entries, i.e,* $\phi_{i,j} \sim \mathcal{N}(0, \sigma^2)$. *Let* $1 > \kappa > 0$ *and choose* $\eta, \epsilon > 0$ *such that* $\kappa = \frac{\eta+\epsilon}{1-\epsilon}$. *Then*

$$\|\Phi x\|_1 \geq M\sigma\sqrt{\frac{2}{\pi}}(1-\kappa)\|\mathbf{x}\|_2$$

holds uniformly for $\mathbf{x} \in \mathbb{R}^\delta$ *with probability exceeding*

$$1-(1+2/\epsilon)^\delta \exp\left(-\frac{\eta^2 M}{2c^2}\right)$$

where $c = (31/40)^{\frac{1}{4}}\,(1.13+\sqrt{\pi})$.

Applying theorem 5, we get that for every $1 \geq \kappa > 0$, and $\alpha^* = M\sqrt{\frac{2}{\pi(L+1)}}(1-\kappa)$ we have :

$$\mathbb{P}\left(E^1_{\alpha^*}\right) \geq 1-(1+2/\epsilon)^\delta \exp\left(-\frac{\eta^2 M}{2c^2}\right) \tag{10}$$

where ϵ and η are positive reals satisfying $\kappa = \frac{\eta+\epsilon}{1-\epsilon}$.

Remark 2. Note that in contrast to ϵ, increasing η improves the lower bound probability. Consequently, the values of η and ϵ can be set so as to maximize the lower bound probability.

According to Markov inequality, $\mathbb{P}\left(E^2_\alpha\right)$ can be written as:

$$\begin{aligned}\mathbb{P}\left(E^2_{\alpha^*}\right) &= \mathbb{P}\left\{\|\mathbf{A}^T\mathbf{v}\|_2 \leq \alpha^*\right\}\\ &\geq 1-\frac{\mathbb{E}\left(\|\mathbf{A}^T\mathbf{v}\|_2^2\right)}{(\alpha^*)^2}\\ &\geq 1-\frac{M\delta(2L-\delta+1)}{2\left(\alpha^*\right)^2(L+1)}\\ &\geq 1-\frac{\pi\delta\left(2L-\delta+1\right)}{4M\left(1-\kappa\right)^2}. \end{aligned} \tag{11}$$

Using (6) and (11), the lower bound on the channel identifiability can be lower bounded by:

$$\mathcal{P} \geq \max_{\substack{\kappa,\epsilon,\eta\\ \kappa=\frac{\eta-\epsilon}{1+\epsilon}}} 1-(1+2/\epsilon)^\delta \exp\left(-\frac{\eta^2M}{2c^2}\right)-\frac{\pi\delta\left(2L-\delta+1\right)}{4M\left(1-\kappa\right)^2}. \tag{12}$$

Although the provided bound is not tight, it provides information about the impact of the system parameters on the channel indetifiability probability. One can obviously see that increasing the number of antennas M improves the channel identifiability probability, in contrast to the system parameters L and δ which tend to decrease it.

4.2. ℓ_p quasi-norm

In this section, we will consider, only the case when $\delta = 1$. Let $\mathbf{h}_i = [h_{i,1},\cdots,h_{i,M}]$ and $s_{i,j} = \text{sign}(h_{i,j})\text{sign}(h_{i-1,j})$. Then, the probability that E^2_α is satified can be expressed as:

$$\mathbb{P}\left(E^2_\alpha\right) = \mathbb{P}\left\{\left|p\sum_{k=1}^{L}\sum_{j=1}^{M}|h_{i,j}|^{p-1}\,|h_{i-1,j}|\,s_{i,j}\right| \leq \alpha\right\}.$$

One can note that as p tends to zero, $p\sum_{k=1}^{L}\sum_{j=1}^{M}|h_{i,j}|^{p-1}\,|h_{i-1,j}|\,s_{i,j}$ converges almost surely to zero, thus implying that $\mathbb{P}\left(E^2_\alpha\right)$ tends to one as p tends to zero.

Let us deal with the complementary event of E^2_α. According to Markov inequality, $\mathbb{P}\left({}^cE^2_\alpha\right)$ can be upper bounded by:

$$\mathbb{P}\left({}^cE^2_\alpha\right) \leq \frac{\mathbb{E}\left|p\sum_{i=1}^{L}\sum_{j=1}^{M}|h_{i,j}|^{p-1}|h_{i-1,j}|s_{i,j}\right|^{\frac{1}{2}}}{\alpha^{\frac{1}{2}}}.$$

To remove the expectation on the Rademacher sequence $s_{i,j}$, we will use the Kintchine's inequality that can be stated as follows:

Lemma 1. *[13] Let* $\mathbf{x}$ *be a vector in* $\mathbb{R}^n$*, and* s_j *a Rademacher sequence. Then, we have:*

$$\mathbb{E}\sqrt{\left|\sum_{j=1}^{n}s_j\mathbf{x}_j\right|} \leq C\left[\sum_{j=1}^{n}x_j^2\right]^{\frac{1}{4}}$$

where $C = \frac{\Gamma(\frac{3}{4})}{2^{\frac{3}{4}}\Gamma(\frac{3}{2})}$.

Applying the Kintchine's inequality, we get:

$$\mathbb{E}\left|\sum_{i=1}^{L}\sum_{j=1}^{M}|h_{i,j}|^{p-1}|h_{i-1,j}|s_{i,j}\right|^{\frac{1}{2}} \leq C\sum_{i=1}^{L}\mathbb{E}\left[\sum_{j=1}^{M}|h_{i,j}|^{2(p-1)}|h_{i-1,j}|^2\right]^{\frac{1}{4}} \tag{13}$$

Since $p < 1$, $x \mapsto |x|^p$ is a concave function when $x > 0$, we can therefore prove, using Jensen inequality, that:

$$\sum_{j=1}^{M}|h_{i,j}|^{2(p-1)}|h_{i-1,j}|^2 \leq \left(\sum_{j=1}^{M}|h_{i-1,j}|^2\right)^p\left(\sum_{j=1}^{M}\left|\frac{h_{i-1,j}}{h_{i,j}}\right|^2\right)^{1-p} \tag{14}$$

$$\overset{(a)}{\leq} p\sum_{j=1}^{M}|h_{i-1,j}|^2+(1-p)\sum_{j=1}^{M}\frac{|h_{i-1,j}|^2}{|h_{i,j}|^2} \tag{15}$$

where the second inequality referred to as (a) follows from the fact that $a^pb^q \leq pa+qb$ whenever a and b are positive and $p+q=1$. Combining (13) and (15), we get :

$$\begin{aligned}&\mathbb{E}\left[p\sum_{i=1}^{L}\left|\sum_{j=1}^{M}|h_{i,j}|^{p-1}|h_{i-1,j}|s_{i,j}\right|\right]^{\frac{1}{2}}\\ &\leq C\sqrt{p}\sum_{i=1}^{L}\mathbb{E}\left(p\sum_{j=1}^{M}|h_{i-1,j}|^2\right)^{\frac{1}{4}}+\sqrt{p}(1-p)^{\frac{1}{4}}\sum_{j=1}^{M}\mathbb{E}\frac{\sqrt{|h_{i-1,j}|}}{\sqrt{|h_{i,j}|}}\\ &\leq CL\sqrt{p}\left(\frac{2p}{L+1}\right)^{\frac{1}{4}}\frac{\Gamma(M/2+\frac{1}{4})}{\Gamma(M/2)}+\frac{CLM\sqrt{p}(1-p)^{\frac{1}{4}}}{\pi}\Gamma(3/4)\Gamma(\frac{1}{4}).\end{aligned}$$

Consequently, for $\alpha = M\sqrt{\frac{2}{\pi(L+1)}}(1-\epsilon)$

$$\mathbb{P}\left(E^2_\alpha\right) \geq 1-\frac{A_1p^{\frac{3}{4}}}{\sqrt{1-\epsilon}}-\frac{A_2\sqrt{p}\,(1-p)^{\frac{1}{4}}}{\sqrt{1-\epsilon}} \tag{16}$$

where

$$\begin{aligned}A_1 &= \frac{2^{\frac{1}{4}}L\Gamma(M/2+1/4)\Gamma(3/4)}{\Gamma(M/2)\sqrt{M}\pi^{\frac{1}{4}}}\\ A_2 &= \frac{L(L+1)^{\frac{1}{4}}\sqrt{M}\Gamma(1/4)(\Gamma(3/4))^2}{\pi^{\frac{5}{4}}}.\end{aligned}$$

We note that unlike A_1, A_2 tends to increase with M. Hence, the lower bound probability does not always decrease with M. Combining (10) and (16), it can be proved that the channel identifiability is lower bounded by:

$$\mathcal{P} \geq \max_{\epsilon}\left(1-\exp(-\frac{M\epsilon^2}{\pi})\right)\left(1-\frac{A_1p^{\frac{3}{4}}}{\sqrt{1-\epsilon}}-\frac{A_2\sqrt{p}\,(1-p)^{1/4}}{\sqrt{1-\epsilon}}\right). \tag{17}$$

5. SIMULATION RESULTS

5.1. ℓ_1 norm

We present here simulation results for the ℓ_1 norm. Fig. 1 displays the effect of the system parameters L and M on the lower bound probability that we have computed by maximizing (7) numerically. We note that, as expected, increasing the number of antennas tends to enhance the channel identifiability probability.

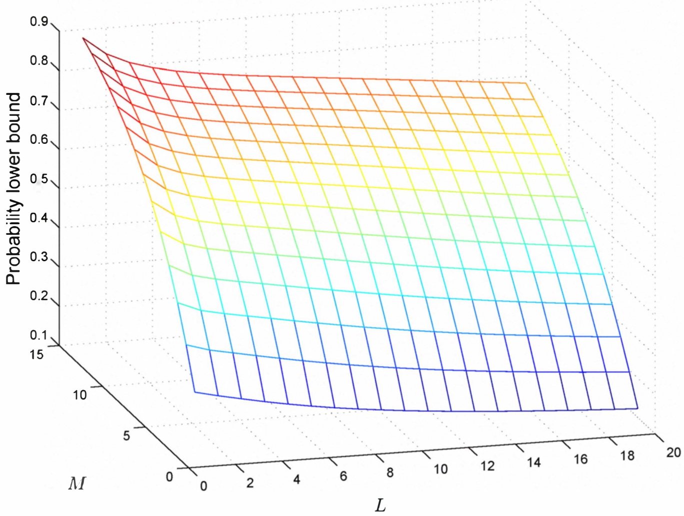

Fig. 1. Impact of the system parameters L and M on the lower bound probability.

5.2. ℓ_p quasi-norm

For the ℓ_p quasi-norm, we study the effect of the parameter p on the lower bound probability. We set the system parameters M and L to 6 and 3, and we vary p from 10^{-3} to 10^{-6}. Fig. 2 displays the lower bound with respect to p. We note that as p tends to zero, the lower bound probability increases.

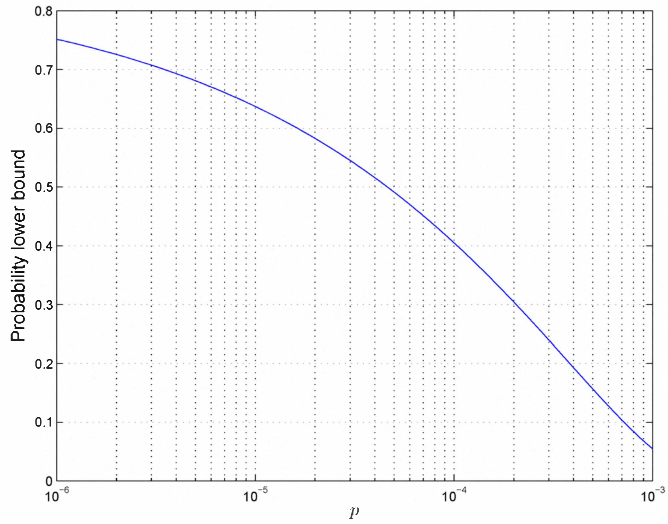

Fig. 2. Lower bound probability with respect to p.

6. CONCLUSION

This paper analyses the robustness of certain blind channel identification methods when using ℓ_p quasi-norms. Necessary and sufficient conditions of channel identifiability are provided. Lower bounds on the channel identifiability probability are derived. Even though they are not tight, they provide some useful insights on the impact of the number of sensors M, the channel length L and the quasi-norm parameter p on the identifiability conditions.

7. REFERENCES

[1] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A Subspace Algorithm for Certain Blind Identification Problems," *IEEE Transactions on Information Theory*, vol. 43, no. 2, pp. 499–511, March 1997.

[2] A. Gorokhov, M. Kristensson, and B. Ottersten, "Robust Blind Second-Order Deconvolution," *IEEE Signal Processing Letters*, vol. 6, no. 1, pp. 13–16, January 1999.

[3] A. Aissa El Bey and K. Abed-Meraim, "Blind Identification of Sparse SIMO Channels Using Maximum a Posteriori Approach," *EUSIPCO*, August 2008.

[4] A. Aissa El Bey and K. Abed-Meraim, "Blind SIMO Channel Identification Using a Sparsity Criterion," *SPAWC*, July 2008.

[5] E. J. Candès and M. B. Wakin, "An Introduction to Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[6] P. Blomgren and T. F. Chan, "Color TV: Total Variation Methods for Restoration of Vector-Valued Images," *IEEE Transactions on Image Processing*, vol. 7, pp. 304–309, March 1998.

[7] E. J. Candès and T. Tao, "Decoding by Linear Programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, December 2005.

[8] A. C. Gilbert and J. A. Tropp, "Applications of Sparse Approximation in Communications," *IEEE Int. Symp. Information Theory*, pp. 1000–1004, September 2005.

[9] R. Gribonval and K. Schnass, "Dictionary Identification-Sparse Matrix-Factorisation via ℓ_1-Minimisation," *arxiv preprint 0904.4774*, April 2009.

[10] R. Gribonval and K. Schnass, "Dictionary Identifiability from Few Training Samples," *EUSIPCO*, August 2008.

[11] M. Ledoux and M. Talagrand, *Probability in Banach Spaces. Isoperimetry and processes*, Springer-Verlag, 1991.

[12] Rick Chartrand and Valentina Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 035020, pp. 1–14, 2008.

[13] S. Foucart and M. J. Lai, "Sparse Recovery with Pre-Gaussian Random Matrices," 2009.