Blind joint MIMO channel and data estimation based on regularized ML

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

Blind Joint MIMO Channel and Data Estimation based on Regularized ML

Nacerredine Lassami[a,b], Abdeldjalil Aïssa-El-Bey[b], Karim Abed-Meraim[c]

[a] *Ecole Militaire Polytechnique, LTS, BP 17 Bordj El Bahri, Algiers, Algeria.*
[b] *IMT Atlantique, Lab-STICC UMR CNRS 6285, F-29238 Brest, France.*
[c] *Université d'Orléans, PRISME, 12 Rue de Blois, 45067 Orleans, France.*

Abstract

The problem of blind joint FIR-MIMO channel and data estimation is addressed in this paper. Based on a regularized DML (Deterministic Maximum Likelihood) formulation of the problem, a bilinear approach is used in order to estimate jointly the channel impulse responses and the input data. This regularization is introduced as a penalty function added to the classical DML criterion representing the *a priori* information about the problem in order to enhance the accuracy of the estimation. Two types of priors information are considered for the transmitted data: the finite alphabet simplicity or the sparsity. The sparsity prior was also considered for channel impulse responses. The key advantage of the proposed criteria is their convexity when optimized alternatively over the channel and the input data. The proposed approach allows to improve further the estimation accuracy of such a blind estimation problem but suffers from a relatively high computational cost. Hence, a reduced complexity implementation of the latter has been proposed at the end of the paper, in an adaptive scheme for high dimensional or streaming data situations.

Keywords: Blind MIMO identification, deterministic ML, regularized estimation, joint channel and data estimation, sparsity, simplicity.

Email addresses: lassami.nacerredine@gmail.com (Nacerredine Lassami), abdeldjalil.aissaelbey@imt-atlantique.fr (Abdeldjalil Aïssa-El-Bey), karim.abed-meraim@univ-orleans.fr (Karim Abed-Meraim)

1. Introduction

Blind system identification (BSI) problems are receiving considerable interest from both signal processing and communications communities since accurate channel estimates are likely to become more challenging in future generation wireless systems. Furthermore, blind techniques are necessary in some communication systems where the synchronization between the receiver and the transmitter is not possible and training sequences are not exploitable. This can be due to security problems such is the case in military communication systems or to the impossibility to influence the input such as the problem of extracting the fetal electrocardiogram (fECG) or in speech separation. Blind techniques present a reduced need for overhead information which increases the bandwidth efficiency. Although, even if some training sequence exists, combining them with blind techniques, in a semi-blind scheme, often leads to improved performance [1–8].

Recently, the problem of joint channel estimation and data detection has been the subject of much research and different solutions have been proposed depending on the considered application. For instance, in [5] the authors analyzed the performance of the data-aided channel estimation scheme in a multi-cell large antenna system and they can effectively suppress the contamination effect and achieve improved performance in large antenna systems. A solution to enhance the reliability of a single-input single-output (SISO) underwater acoustic communication system was proposed in [9] based on a maximum likelihood (ML) method for both channel estimation and data detection followed by a data-projection step. In the case of SISO orthogonal frequency-division multiplexing (SISO-OFDM) model and under the assumption of sparse channel, a recursive algorithm was proposed in [6] based on sparse Bayesian learning and Kalman filtering for the joint channel estimation and data detection. In [7], an iterative frequency-domain joint channel estimation and data detection solution was proposed for the problem of faster-than-Nyquist signaling in a SISO system. A Bayes-optimal inference approach based on the bilinear generalized approximate

message passing (BiG-AMP) algorithm was used in [4] to address the problem of joint channel and data estimation in a multiple-input multiple-output (MIMO) system. Recently, a low-complexity semi-blind scheme is proposed in [8] for joint channel estimation and data detection on sphere manifold MIMO systems with high-order quadrature amplitude modulation (QAM) signaling. It uses training symbols with zero-forcing detector to get an initial channel and data estimation, then, a Riemannian conjugate gradient method is used followed by another zero-forcing detector. The MIMO-OFDM system is considered in [10], where the authors proposed an adaptive blind scheme for joint channel and data estimation based on parallel factor decomposition (PARAFAC).

The joint channel estimation and data detection (data-aided) problem is generally formulated as a deterministic maximum likelihood (DML) estimation where the input signals are considered as part of the unknown parameters with the channel coefficients. The ML methods have the advantage of being efficient for high signal-to-noise ratio (SNR) [11]. The ML methods usually cannot be obtained in closed form and require an optimization in the presence of local minima. In addition, the dimension of the problem increases with the sample size, which makes this approach not practical for large data size applications. However, ML approaches can be made very effective by including initialization procedure such as least squares channel estimate based on training symbols, the blind channel subspace method or other sub-optimal approaches. As major contributions to DML methods, we can cite the two-step maximum likelihood (TSML) [12], the iterative quadratic maximum likelihood (IQML) [13] and its dual algorithm proposed in [14] which considered single-input multiple-output (SIMO) finite impulse response (FIR) channels. Another DML method, the maximum likelihood block algorithm (MLBA), has been proposed in [15] where for both the channels and the symbols, least squares estimation is performed in an alternating manner. Following the same formulation, the Maximum Likelihood Adaptive Algorithm (MLAA) is derived in [16] which presents reduced-complexity in term of computation cost.

Another important feature of the considered DML formulation is the facility

to introduce any further information about the channel impulse response or the input signal to the cost function. This priors depends on the considered system and adding them to the DML formulation as regularization will likely enhance the estimation accuracy. For instance, most of the digital communications are based on transmitted signals that belong to a finite-alphabet set. The finite alphabet property was first considered with DML formulation in [17–19]. Unfortunately, the convergence of such methods is often not guaranteed due to the increased number of the local minima caused by the incorporation of the finite alphabet property. The solution to such a problem is computationally expensive and can be achieved using the Viterbi-like algorithms. In order to reduce the computational complexity, Bayesian maximum a posteriori approaches were proposed such as in [4, 20]. Recently, the simplicity property was considered for the recovery of finite-alphabet signals as is the case in [21] for large-scale MIMO systems. The sparsity prior of the transmitted signals was also considered in different applications such as in [22, 23] for blind source separation when the sources are known to be sparse or can be sparsely represented. The sparsity of the channel impulse response can also be used with the DML formulation. In fact, it seems likely that in case of very long impulse response and sparse channel, most of the state-of-the-art methods perform poorly. Such sparse channels can be encountered in many communication applications including High-Definition television (HDTV) channels and underwater acoustic channels [24].

Paper contributions

In this paper, we aim to estimate blindly both the channel state information and the transmitted data in the case of a MIMO FIR channel (convolutive system). We propose a blind scheme based on a regularized DML formulation by considering priors on either or both the channel impulse response and the transmitted data. First, a blind second-order subspace method [25] is used to estimate the channel followed by a resolving ambiguity step which is necessary in second-order method. We will use for the ambiguity removal, either the sparsity assumption in case of sparse channel prior or the independence assumption of

the transmitted signals in case of non-sparse channel. The key advantage of the proposed regularized DML formulation is its convexity when optimized alternatively over the channel and the input data. Two regularized quadratic convex problems will be iteratively optimized to estimate the data in the first one and the channel in the second. Different priors are considered in our work: the finite alphabet simplicity, the sparsity of the transmitted data, or the sparsity of the channels finite responses. The proposed constrained ML-like approach allows to improve further the estimation accuracy of such a blind estimation problem but suffers from a relatively high computational cost. Hence, an adaptive version is also proposed to reduce the numerical complexity and to handle the case of time varying channel. The specific contributions of this paper include:

- Providing a unified convex regularized DML formulation for different types of priors: sparsity or simplicity of input signals and sparsity of channel impulse responses while considering the MIMO convolutive model. To the best of our knowledge, this is the first work that solves the joint channel and input data estimation in a fully blind context using simultaneously channel sparsity and input signal simplicity priors.
- Resolving the ambiguity of blind second-order subspace method in case of sparse channel impulse responses.
- A reduced complexity implementation of the proposed constrained ML-like approach, proposed in an adaptive scheme for high dimensional or streaming data situations.

2. Data model

Our focus is on the MIMO channels case in a blind identification context. We aim to estimate jointly the channel state information and the transmitted data using only the observation data and some additional prior information. 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 N_t input

sequences $\mathbf{s}_1(t), ..., \mathbf{s}_{N_t}(t)$, and yields N_r output sequences $\mathbf{x}_1(t), ..., \mathbf{x}_{N_r}(t)$ at time t. We assume that the discrete channel between the N_t transmit antennas and the N_r receive antennas are modeled as a $N_r \times N_t$ FIR filter with L as the upper bound on the orders of these channels i.e. $\mathbf{H} = [\mathbf{H}^T(0), ..., \mathbf{H}^T(L)]^T$ with

$$\mathbf{H}(l) = \begin{bmatrix} h_{11}(l) & \dots & h_{1N_t}(l) \\ \vdots & & \vdots \\ h_{N_r1}(l) & \dots & h_{N_rN_t}(l) \end{bmatrix}$$

The system can be described by:

$$\mathbf{x}(t) = \sum_{k=0}^{L} \mathbf{H}(k)\mathbf{s}(t-k) + \mathbf{n}(t) = [\mathbf{H}(z)]\mathbf{s}(t) + \mathbf{n}(t) \tag{1}$$

where

$$[\mathbf{H}(z)] = \sum_{k=0}^{L} \mathbf{H}(k) z^{-k} \tag{2}$$

and $\mathbf{n}(t)$ is an additive N_r-dimensional white noise, independent from the symbol sequences with $\mathbb{E}[\mathbf{n}(t)\mathbf{n}^T(t)] = \sigma^2 \mathbf{I}_{N_r}$, where σ^2 is the unknown noise power.

We denote stacked observations over a window of length N by the vector $\mathbf{x}_N(t) = [\mathbf{x}^T(t), \mathbf{x}^T(t-1), ..., \mathbf{x}^T(t-N+1)]^T$. The following linear model holds:

$$\mathbf{x}_N(t) = \mathcal{T}(\mathbf{H})\mathbf{s}_N(t) + \mathbf{n}_N(t) \tag{3}$$

with $\mathbf{s}_N(t) = [\mathbf{s}^T(t), \mathbf{s}^T(t-1), ..., \mathbf{s}^T(t-N-L)]^T$ and $\mathbf{n}_N(t) = [\mathbf{n}^T(t), ..., \mathbf{n}^T(t-N+1)]^T$. The $(NN_r) \times ((N+L+1)N_t)$ block-Toeplitz matrix $\mathcal{T}(\mathbf{H})$ associated with the filter $\mathbf{H}$ is given by:

$$\mathcal{T}(\mathbf{H}) = \begin{bmatrix} \mathbf{H}(0) & \dots & \mathbf{H}(L) & & 0 \\ & \ddots & & \ddots & \\ 0 & & \mathbf{H}(0) & \dots & \mathbf{H}(L) \end{bmatrix} \tag{4}$$

From now on, we make the following important assumptions:

- $\mathbf{H}(z)$ is irreducible ($\text{Rank}(\mathbf{H}(z)) = N_t$; $\forall z$) and $\mathbf{H}(L)$ is of full column rank
- The channels have a known maximum order L.

- Number of sensors is strictly greater than the number of sources $N_r > N_t$
- Large enough observation Window $N > (L+1)N_t$

Under this assumptions, $\mathcal{T}(\mathbf{H})$ is a full rank column matrix and the channel matrix $\mathbf{H}$ is identifiable using second order statistics up to a $N_t \times N_t$ constant full rank matrix [25].

3. Blind deterministic maximum likelihood DML estimation approach

The DML approach assumes no statistical model for the input sequences. In other words, both the channel matrix $\mathbf{H}$ and the input source vectors $\mathbf{s}_i(t)$ $i = 1, ..., N_t$ are parameters to be estimated. Considering the MIMO FIR model expressed in above, the DML problem can be stated as: given the observation $\mathbf{x}_N(t)$, we want to estimate:

$$\left\{\hat{\mathbf{H}}, \hat{\mathbf{s}}_N(t)\right\} = \arg\max \quad f\left(\mathbf{x}_N; \mathbf{H}, \mathbf{s}_N\right) \tag{5}$$

where $f\left(\mathbf{x}_N; \mathbf{H}, \mathbf{s}_N\right)$ is the density function of the observation vectors parameterized by both the channel matrix $\mathbf{H}$ and the input sources vector $\mathbf{s}_N$. In the case of zero-mean Gaussian noise with covariance $\sigma^2\mathbf{I}$, the above DML estimator lead to the nonlinear least squares optimization:

$$\left\{\hat{\mathbf{H}}, \hat{\mathbf{s}}_N(t)\right\} = \underset{\mathcal{T}(\mathbf{H}), \mathbf{s}_N(t)}{\arg\min} \left\{\left\|\mathbf{x}_N(t) - \mathcal{T}(\mathbf{H})\mathbf{s}_N(t)\right\|_2^2\right\} \tag{6}$$

The above function can also be written as :

$$\left\{\hat{\mathbf{H}}, \hat{\mathbf{s}}_N(t)\right\} = \underset{\mathcal{T}(\mathbf{H}), \mathbf{s}_N(t)}{\arg\min} \left\{\left\|\mathbf{x}_N(t) - \mathcal{F}\left(\mathbf{s}_N(t)\right)\mathbf{h}\right\|_2^2\right\} \tag{7}$$

where $\mathbf{h} = \text{vec}(\mathbf{H})$ is the vectorized version of the matrix $\mathbf{H}$. The operator $\mathcal{F}\left(.\right)$ transforms a vector $\mathbf{s}_N(t)$ into an $NN_r \times (L+1)N_tN_r$ matrix, in such a way that:

$$\mathcal{F}\left(\mathbf{s}_N(t)\right)\mathbf{h} = \mathcal{T}(\mathbf{H})\mathbf{s}_N(t). \tag{8}$$

Under the considered assumptions, $\mathcal{T}(\mathbf{H})$ is a full rank column matrix and the variables $\left(\mathbf{H}, \mathbf{s}_N(t)\right)$ are identifiable up to a $N_t \times N_t$ constant full rank

matrix [25] which we will denote $\mathbf{Q}$. The identifiability condition for the DML approach is the same as that for the deterministic second-order subspace methods. The reason is that, when the noise is Gaussian (assumed true in DML), all information about the channel in the likelihood function resides in the second-order moments of the observations. Readers are referred to [25] for a more detailed discussion about the identifiability of the MIMO FIR channel with the noise subspace method.

Resolving the ambiguity caused by the matrix $\mathbf{Q}$ in second-order based blind channel estimation of MIMO-FIR channels is equivalent to instantaneous Blind Source Separation BSS problem [26]. Hence, the indeterminacy can be reduced to a complex-valued diagonal matrix and a real-valued permutation matrix.

The blind DML criterion as stated in Eq. (6) and Eq. (7) is a complicated and a non-convex optimization problem when estimating jointly both parameters $\Big(\mathbf{H}, \mathbf{s}_N(t)\Big)$. However, it is convex for each unknown parameter. Hence, Eq. (6) and Eq. (7) can be resolved by alternatively optimizing over $\mathbf{H}$ and $\mathbf{s}_N(t)$. This idea was proposed in many previous papers [13, 15, 16, 20] but only in the SIMO case.

4. Regularized blind deterministic maximum likelihood DML estimation approach

In the case where the system is not time-varying and if the data sequence is long enough so that a reliable statistical model can be built, then, a statistical maximum likelihood SML method should be used, since in that case, the statistical method outperforms the deterministic one in terms of estimation accuracy. However, in a fast fading environment, data related to a given channel are not numerous which makes a reliable statistical estimation less effective. In such a situation, the symbols are assumed arbitrary and a DML method is used instead. The regularized DML approach is a trade-off between DML and SML approaches, where we take advantage of some prior information about input data or/and transmission channels to enhance the DML results. Next, we

present different types of priors and how to integer them into the DML criterion.

4.1. Simplicity and the finite alphabet property

The finite alphabet property in digital communication signals was considered in [17–19] with the DML approach. In this case, an alternating optimization is used to minimize the criterion given by:

$$\hat{\mathbf{s}}_N(t) = \arg\min_{\mathbf{s}_N(t)\in\mathcal{S}} \Big\| \mathbf{x}_N(t) - \mathcal{T}(\mathbf{H})\mathbf{s}_N(t) \Big\|_2^2 \tag{9}$$

$$\hat{\mathbf{H}} = \arg\min_{\mathbf{h}} \Big\| \mathbf{x}_N(t) - \mathcal{F}\Big(\mathbf{s}_N(t)\Big)\mathbf{h} \Big\|_2^2 \tag{10}$$

where $\mathbf{s}_N(t)$ belongs to a complex-valued finite alphabet $\mathcal{S}$. It can be decomposed from its real and imaginary parts as $\mathbf{s}_N(t) = \mathcal{R}(\mathbf{s}_N(t)) + j\mathcal{I}(\mathbf{s}_N(t))$ where $(\mathcal{R}(\mathbf{s}_N(t)), \mathcal{I}(\mathbf{s}_N(t))) \in \mathcal{A}^{N_t(N+L)} \times \mathcal{A}^{N_t(N+L)}$ and $\mathcal{A} = \{\alpha_1, ..., \alpha_m\}$ is a real-valued finite alphabet of cardinal m. The optimization in Eq. (10) is a linear-least squares problem whereas the optimization in Eq. (9) is computationally expensive and can be achieved by using the Viterbi algorithm. The convergence of such approaches is not guaranteed due to the numerous local minima induced by the discrete set constraint. We propose to replace this constraint by the signal simplicity. Simplicity property was first introduced in [27] where we say that a signal is *simple* if most of its elements are equal to the extremes of the finite alphabet. Recently, the authors in [21] proposed a simplicity-based detector for finite alphabet source separation in both determined and underdetermined large-scale MIMO systems. They relaxed the problem of the finite alphabet constraint as stated in Eq. (9) to convex box constraint depending only on the constellation extremes $\{\alpha_1, \alpha_m\}$ (e.g. $\{-1, +1\}$ for BPSK and QPSK modulations, or $\{-3, +3\}$ for 16-QAM modulation). The new optimization problem is given by:

$$\hat{\mathbf{r}} = \arg\min_{\mathbf{r}} \Big\| \underline{\mathbf{x}}_N(t) - \underline{\mathcal{T}}(\mathbf{H})\mathbf{B}_\alpha\mathbf{r} \Big\|_2^2 \quad \text{subject to} \quad \mathbf{B}_1\mathbf{r} = \mathbf{1}_{2N}, \quad \mathbf{r} \geq 0 \tag{11}$$

where $\underline{\mathbf{s}}_N(t) = \mathbf{B}_\alpha\mathbf{r}$. The matrices $\mathbf{B}_\alpha$ and $\mathbf{B}_1$ are defined by: $\mathbf{B}_\alpha = \mathbf{I}_{2N_t(N+L)} \otimes [\alpha_1, \alpha_m]$ and $\mathbf{B}_1 = \mathbf{I}_{2N_t(N+L)} \otimes [1, 1]$ where $(. \otimes .)$ stands for the Kronecker

product. The underlined notation is reserved to the complex-real transformation where $\underline{\mathcal{T}}(\mathbf{H})$, $\underline{\mathbf{x}}_N(t)$ and $\underline{\mathbf{s}}_N(t)$ are given by

$$\underline{\mathcal{T}}(\mathbf{H}) = \begin{bmatrix} \mathcal{R}(\mathcal{T}(\mathbf{H})) & -\mathcal{I}(\mathcal{T}(\mathbf{H})) \\ \mathcal{I}(\mathcal{T}(\mathbf{H})) & \mathcal{R}(\mathcal{T}(\mathbf{H})) \end{bmatrix}$$

$$\underline{\mathbf{x}}_N(t) = \begin{bmatrix} \mathcal{R}(\mathbf{x}_N(t)) \\ \mathcal{I}(\mathbf{x}_N(t)) \end{bmatrix} \quad \underline{\mathbf{s}}_N(t) = \begin{bmatrix} \mathcal{R}(\mathbf{s}_N(t)) \\ \mathcal{I}(\mathbf{s}_N(t)) \end{bmatrix}$$

In addition to the convexity of this formulation, the computation cost of the resulting detector does not depend on the constellation size [21]. The simplicity is used generally in case of modulations where the constellation is shaped as a square box such in the case of BPSK, QPSK and M-QAM modulations where M is a square number (e.g., 16-QAM, 64-QAM). Note that the performance of the simplicity approach depends on the number of the points in the square box defined by the boundary of the constellation.

4.2. Sparsity of the input signals

The sparsity property of the input signals was introduced earlier as a contrast function for blind source separation [22]. In this paper, the sparsity is introduced as a regularization constraint to Eq. (6) in order to enhance the estimation and improve the robustness to outliers. If the sources are sparse, the regularization can be carried out directly in the time domain by considering the least absolute shrinkage and selection (LASSO) problem [28] (or basis pursuit equivalent problem [28]). In this case, an alternating optimization is used to minimize the criterion given by:

$$\hat{\mathbf{s}}_N(t) = \arg\min_{\mathbf{s}_N(t)} \left\| \mathbf{x}_N(t) - \mathcal{T}(\mathbf{H})\mathbf{s}_N(t) \right\|_2^2 + \lambda_s \left\| \mathbf{s}_N(t) \right\|_1 \tag{12}$$

$$\hat{\mathbf{H}} = \arg\min_{\mathbf{h}} \left\| \mathbf{x}_N(t) - \mathcal{F}\Big(\mathbf{s}_N(t)\Big)\mathbf{h} \right\|_2^2 \tag{13}$$

where λ_s is a weighting parameter that controls the trade-off between approximation error and sparsity level of the input signals.

This formulation can also be interpreted as a Maximum a Posteriori (MAP) approach where the input signals $\mathbf{s}_N$ are estimated using its conditional probability distribution as follows:

$$\hat{\mathbf{s}}_N = \arg\max_{\mathbf{s}_N} \left\{ \frac{f\Big(\mathbf{x}_{N_r}|\mathbf{s}_N\Big) f\Big(\mathbf{s}_N\Big)}{\int f\Big(\mathbf{x}_{N_r}|\mathbf{s'}_N\Big) f\Big(\mathbf{s'}_N\Big) d\mathbf{s'}_N} \right\} \tag{14}$$

$$= \arg\max_{\mathbf{s}_N} \left\{ f\Big(\mathbf{x}_{N_r}|\mathbf{s}_N\Big) f\Big(\mathbf{s}_N\Big) \right\} \tag{15}$$

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 $f(\mathbf{s}_N)$ of the input signals vector. This *a priori* depends on the application context and its physical environment. In our case, it is the transmitted vector sparsity that we model by considering the input signals components i.i.d. as the Generalized Gaussian Distribution (GGD) given by:

$$f(\mathbf{s}_N) = \Big(\frac{p}{2\beta\Gamma(\frac{1}{p})}\Big)^{-N_t(N+L+1)} \exp\Big(-\frac{\|\mathbf{s}_N\|_p^p}{\beta^p}\Big) \tag{16}$$

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. Using this pdf, one increases the chances to get signal coefficients close to zero. Next, we will consider the special case of $p = 1$ which is equivalent to the Laplace prior distribution because of the convexity of the ℓ_1-norm.

The problem in Eq. (12) is convex but has no closed-form solution. The Least Angle Regression (LARS) [29] is a less greedy version of traditional forward selection methods for model selection problems and solves the LASSO problem efficiently.

In the case where the input signals are not sparse in the time domain, we can use sparse representations methods in order to transform the signals into a dictionary where they are more sparse. For instance, speech signals have a more sparse representations in the time-frequency domain than in the time domain, therefore, the Short-time Fourier transform (STFT) transform is used in this case. If we denote by $\mathbf{\Phi}$ the considered transformation (dictionary) matrix, then

the problem in Eq. (12) can be rewritten as:

$$\hat{\mathbf{r}} = \arg\min_{\mathbf{r}} \left\| \mathbf{x}_N(t) - \mathcal{T}(\mathbf{H})\mathbf{\Phi}\mathbf{r} \right\|_2^2 + \lambda_r \left\| \mathbf{r} \right\|_1 \tag{17}$$

with $\underline{\mathbf{s}}_N(t) = \mathbf{\Phi}\mathbf{r}$.

4.3. Sparsity of the channel impulse response

The sparsity of the channel impulse response was first studied in the case of SIMO systems, then extended to the MIMO, e.g. [24]. In order to exploit the sparsity *a priori* information of the channel impulse response, we introduce an additional cost function based on the GGD model of channel coefficients in the same manner as we did previously in case of sparse input signals. Under the assumption that all the component of $\mathbf{H}$ are `i.i.d`, the GGD model is expressed in the same way as in Eq. (16). This model enhances the sparsity of the channel $\mathbf{H}$. Taking the logarithm of the *a posteriori* estimator leads to the objective function:

$$\hat{\mathbf{H}} = \arg\min_{\mathbf{h}} \left\| \mathbf{x}_N(t) - \mathcal{F}\Big(\mathbf{s}_N(t)\Big)\mathbf{h} \right\|_2^2 + \lambda_h \left\| \mathbf{h} \right\|_q^q \tag{18}$$

where λ_h is a weighting parameter that controls the trade-off between approximation error and sparsity represented by the ℓ_q-norm of the channel impulse response. This function has a LASSO-like formulation for the case $q = 1$ and can be solved using the LARS method.

Actually, representing the sparsity of the source signals or the channel coefficients with the ℓ_1-norm minimization is suitable due to the convexity property. However, another heuristic penalization function (generally non-convex) was used in the literature to enhance the sparsity such as the Reweighted-ℓ_1 [30] criterion, which generally out performs the ℓ_1-based criterion. The cost function, in this case, is given by:

$$\hat{\mathbf{H}} = \arg\min \left\| \mathbf{x}_N(t) - \mathcal{F}\Big(\mathbf{s}_N(t)\Big)\mathbf{h} \right\|_2^2 + \lambda_h \sum_{i=1}^{N_r N_t (L+1)} \log\Big(|h(i)| + \epsilon\Big) \tag{19}$$

where $\epsilon > 0$ is relatively small positive constant. Both cost functions in Eq. (18) and Eq. (19) are optimized under the constraint $\mathbf{H}(1, j) = 1$ for $j = 1, ..., N_t$ to avoid the trivial null solutions.

4.4. Initialization and resolving ambiguities in subspace-based blind identification of MIMO channels

As we stated before, ML methods solutions usually cannot be obtained in closed-form and require optimization in the presence of local minima. The dimension of the problem to optimize depends on the sample size which is problematic in case of large data size applications. As a solution, generally ML approaches can be very effective if they are initialized by some suboptimal procedure. In our case, we have chosen the MIMO second-order based subspace method [25] because of its nice convergence properties.

The choice of the subspace method as an initialization procedure will lead to another problem in our formulation which concerns the ambiguity of the blind identification MIMO problem itself. In the MIMO case ($N_t > 1$), the channel matrix $\mathbf{H}$ is estimated up to a $N_t \times N_t$ full rank matrix $\mathbf{Q}$ i.e.

$$\mathbf{H} = \hat{\mathbf{H}}_{\mathrm{SS}}\mathbf{Q} \tag{20}$$

where $\hat{\mathbf{H}}_{\mathrm{SS}}$ is the estimated channel with the second order subspace method [25]. Hence, the input signal is also estimated up to full rank matrix $\tilde{\mathbf{Q}}^{-1}$ expressed as:

$$\mathbf{s}_N(t) = \tilde{\mathbf{Q}}^{-1}\hat{\mathbf{s}}_N(t)_{\mathrm{SS}} \tag{21}$$

where

$$\mathcal{T}(\mathbf{H}) = \mathcal{T}(\hat{\mathbf{H}}_{\mathrm{SS}})\tilde{\mathbf{Q}} \tag{22}$$

and

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & 0 & \dots & 0 \\ 0 & \mathbf{Q} & \dots & 0 \\ \vdots & & \ddots & 0 \\ 0 & 0 & \dots & \mathbf{Q} \end{bmatrix} \tag{23}$$

Hence, we can not start the iterative optimization of Eq. (6) or any other regularized without resolving this ambiguity. This ambiguity can not be resolved without using-higher order methods or using additional information about the system. In a total blind scheme, the best we can aim for, is to estimate the

channel matrix $\mathbf{H}$ up to a scaling and permutation matrices i.e. $\hat{\mathbf{H}}$ is a solution if:

$$\hat{\mathbf{H}} = \mathbf{HP\Lambda} \tag{24}$$

where $\mathbf{P}$ is a permutation matrix and $\mathbf{\Lambda}$ is a non-singular complex diagonal matrix.

The authors in [26] show that resolving this ambiguity can be treated as an instantaneous Blind Source Separation BSS problem and they solve it using independent component analysis (ICA) under the assumption that transmitted sequences are statistically independent and non-Gaussian (which is generally true for communication sources). Hence, the indeterminacy can be reduced to a complex-valued diagonal scaling matrix and a permutation matrix.

In the case of sparse MIMO channel $\mathbf{H}$, the problem is similar to the sparse principal subspace estimation problem discussed in our previous work [31]. In this case, the subspace is represented by the estimation $\hat{\mathbf{H}}_{\mathrm{SS}}$ and we search for the rotation matrix $\mathbf{Q}$ that leads to the sparse channel $\mathbf{H}$ (i.e. $\mathbf{H} = \hat{\mathbf{H}}_{\mathrm{SS}}\mathbf{Q}$). In a more recent result [32], we have discussed the conditions to have a unique solution that corresponds to the sparse MIMO channel $\mathbf{H}$. Without going deep into details, our results stated that if we have high level of sparsity of the MIMO channel $\mathbf{H}$ and a large number of receivers ($(L+1)N_r > CN_t \log(N_t)$ with C is a positive constant related to the sparsity of the channel), then we can estimate the sparse channel $\mathbf{H}$ up to complex diagonal scaling matrix and permutation matrix. If these conditions are satisfied, we can use iteratively the second step of the algorithm SS-FAPI [32] (System matrix Sparsity based on Fast Approximated Power Iterations) in order to estimate the sparse channel matrix $\mathbf{H}$ from the subspace channel method estimation $\hat{\mathbf{H}}_{\mathrm{SS}}$. The SS-FAPI algorithm is an adaptive algorithm based on a two step approach, where the first one uses the FAPI algorithm [33] for the adaptive extraction of an orthonormal basis of the principal subspace equivalent to $\hat{\mathbf{H}}^t_{\mathrm{SS}}$ (changes over the time t). Then, an estimation of the desired sparse weight matrix $\mathbf{H}^t$ is done in the second step using the natural gradient scheme. Taking iteratively the second

Algorithm 1 Regularized blind DML joint estimation

Requires: $\mathbf{x}_N(t)$, regularization parameters and number of iterations N_{iter}.

Ensures: $\hat{\mathbf{H}}$ and $\hat{\mathbf{s}}_N(t)$.

1: Find $\hat{\mathbf{H}}_{\text{SS}}$ using MIMO subspace method [24]

2: Resolve the ambiguity of the subspace method by using either:
$\begin{cases} \text{Sparsity of } \mathbf{H} \text{ with the second step of SS-FAPI.} \\ \text{Independence of the transmitted signals with ICA methods.} \end{cases}$

3: **for** $k = 1 \rightarrow N_{\text{iter}}$ **do**

4: **if** Signal simplicity prior **then**

5: Resolve Eq. (11)

6: **else if** Signal sparsity prior **then**

7: Resolve Eq. (12) or Eq. (17) depending on the sparse domain

8: **end if**

9: **if** Channel sparsity prior **then**

10: Resolve Eq. (18)

11: **else**

12: Resolve Eq. (10)

13: **end if**

14: **end for**

step of SS-FAPI is equivalent to fixing the $\hat{\mathbf{H}}^t_{\text{SS}}$ over t and looking for the best $\mathbf{Q}$ to get the sparset channel matrix $\mathbf{H}$.

Finally, the resulting scaling and permutation indeterminacy is inherent to the blind identification problem and requires additional information (semi-blind) in order to be resolved. In our case, we generalize the same assumptions used in the SIMO case such as having the first row of $\mathbf{H}$ equal to ones and sort its columns depending on their ℓ_2-norm. The full regularized blind DML approach is summarized in Algorithm 1.

Complexity analysis

Note that we consider here the dominant costs and hence, the computational complexity of the subspace method and the ambiguity resolving step (steps 1-2

of Algorithm 1) are disregarded. Steps 3-14 of algorithm 1 are based on convex optimization problems. In order to solve these convex problems, we have used the CVX MATLAB Toolbox which is based on the interior point method whose complexity is a function of the length of the unknown vector and the number of constraints [34]. It is shown in [35] that interior-point methods can solve a convex optimization problem defined over $\mathbb{R}^{n_1}$ and subject to n_2 constraints in a number of steps or iterations that is almost always in the range between 10 and 100. Each step requires $max(n_1^3, n_1^2 n_2, F)$ operations, where F is the cost of evaluating the first and second derivatives of the objective and constraint functions. In our case, this is reduced to $\mathcal{O}(N^3)$ for all the cases depending on the considered prior in Algorithm 1 (recall that N is bigger than L, N_r and N_t). This complexity can be reduced further, according to [35], if the problem has some structure which is the case of sparse matrices $\mathcal{T}(\mathbf{H})$ and $\mathcal{F}\Big(\mathbf{s}_N(t)\Big)$.

4.5. Adaptive regularized blind DML estimation approach

In real applications, channels are usually time-varying, which means that we need to develop an adaptive solution where processing is faster compared to acquisition. We provide here an adaptive solution based on the proposed regularized DML approach. First, the subspace MIMO method is used to estimate the channel response of the first observed data block. Then, an iterative estimation of the transmitted data and the channel response is done using the appropriate regularized DML approach depending on the considered case (simplicity, signal sparsity, or channel sparsity). The processing is done block by block. These blocks of size $\tilde{N}$ need to respect identifiability conditions and can be chosen with some overlapping. To enhance the processing speed, we propose to use the alternating direction method of multipliers (ADMM) algorithm [36] for the optimization of the convex regularized objective functions of the DML approach. The ADMM algorithm can solve convex optimization problems by breaking them into smaller problems, each of which is then easier to handle, especially when we use proximal operators. We have chosen to use ADMM algorithm because it is less time-consuming compared to interior point solutions.

Algorithm 2 Adaptive regularized blind DML joint estimation

Requires: $\mathbf{x}_{\tilde{N}}(t)$ the t^{th} block of size $\tilde{N}$, regularization parameters and number of iterations N_{iter}.

Ensures: $\hat{\mathbf{H}}(t)$ and $\hat{\mathbf{s}}_{\tilde{N}}(t)$.

1: **if** $t = 0$ **then**

Find $\hat{\mathbf{H}}_{\text{SS}}(0)$ using MIMO subspace method [24]

Resolve the ambiguity of the subspace method by using either:

$\begin{cases} \text{Sparsity of } \mathbf{H} \text{ with the second step of SS-FAPI.} \\ \text{Independence of the transmitted signals with ICA methods.} \end{cases}$

2: **end if**

3: Apply steps $3 \rightarrow 14$ of Algorithm 1 on $\mathbf{x}_{\tilde{N}}(t)$ while taking as initialization $\hat{\mathbf{H}}(t-1)$ to get $\hat{\mathbf{H}}(t)$ and $\hat{\mathbf{s}}_{\tilde{N}}(t)$

More details about the rate of convergence and how to use ADMM for solving convex problems such as LASSO can be found in [36]. The proposed adaptive algorithm is summarized in Algorithm 2 Table.

5. Simulation results

To assess the performance of the proposed solutions, we consider the MIMO problem with N_t transmitters and N_r receivers and the channels are represented by a polynomial transfer function of maximum degree L. The channels' impulse responses are generated randomly according to a Gaussian distribution for the non-sparse case. Sparse channels impulse responses are generated according to the Bernoulli-Gaussian distribution. In our simulation, we have used a sparsity level of 30%, which means that 30% of vector $\mathbf{h}$ entries are non-zero.

The input signals are a random binary sequences and the additive white Gaussian noise has a variance σ^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 (NMSE) criterion (generalized from the SIMO case to the MIMO) which was used in many previous work such as in [24]. In our case, due to the inherent scaling indeterminacy of the full blind scheme, we have

normalized the entries of the first row of $\mathbf{H}$ to be equal to one. Hence, we can use directly the standard NMSE (i.e. without scale ambiguity removal) defined as:

$$NMSE = \frac{1}{K}\sum_{k=1}^{K}\frac{\|\hat{\mathbf{h}}_k - \mathbf{h}\|^2}{\|\mathbf{h}\|^2} \tag{25}$$

where $\hat{\mathbf{h}}_k = \text{vec}(\hat{\mathbf{H}}_k)$ is the estimated channel vector corresponding to the k^{th} Monte-Carlo simulation and $K = 100$ is the number of Monte-Carlo runs. All the proposed criteria are convex with respect to the channel or the input data separately. Hence, for alternative optimization, we have used the Matlab CVX toolbox [34].

In the first simulation, we consider a real-valued case with a Gaussian noise, a real non-sparse channel with $N_t = 3$, $N_r = 8$ and $L = 10$. We consider a BPSK constellation which will be used for the simplicity prior with $T = 1000$ received symbols. We have used as performance factors the NMSE of the estimated channels and the Bit Error Rate (BER) of the estimated input signals. The proposed regularized DML method which in this case considers the simplicity prior (BPSK constellation) is compared to the initialization subspace method, the Least-Squares LS solution (no regularization) and the oracle solution. In the latter, the exact input signal is used to estimate the channel matrix and the exact channel matrix is used to estimate the transmitted signal. We have considered one iteration of Algorithm 1 for the regularized DML ($N_{iter} = 1$) and two iterations for the LS solutions.

It is clear in Fig. 1 that using the simplicity prior in the DML formulation enhances the performance in terms of NMSE and BER. Furthermore, we can reach the oracle performance for a certain SNR level (above 13dB). Same observations can be made in Fig. 2 where we have considered QPSK modulation, circularly symmetric complex Gaussian noise and a complex non-sparse channel with $(N_t = 3, N_r = 8, L = 10)$ and $T = 1000$ symbols. However, compared to the BPSK case, we require a much higher SNR level to see the impact of the proposed approach.

In both previous simulations, we have considered only one iteration in our

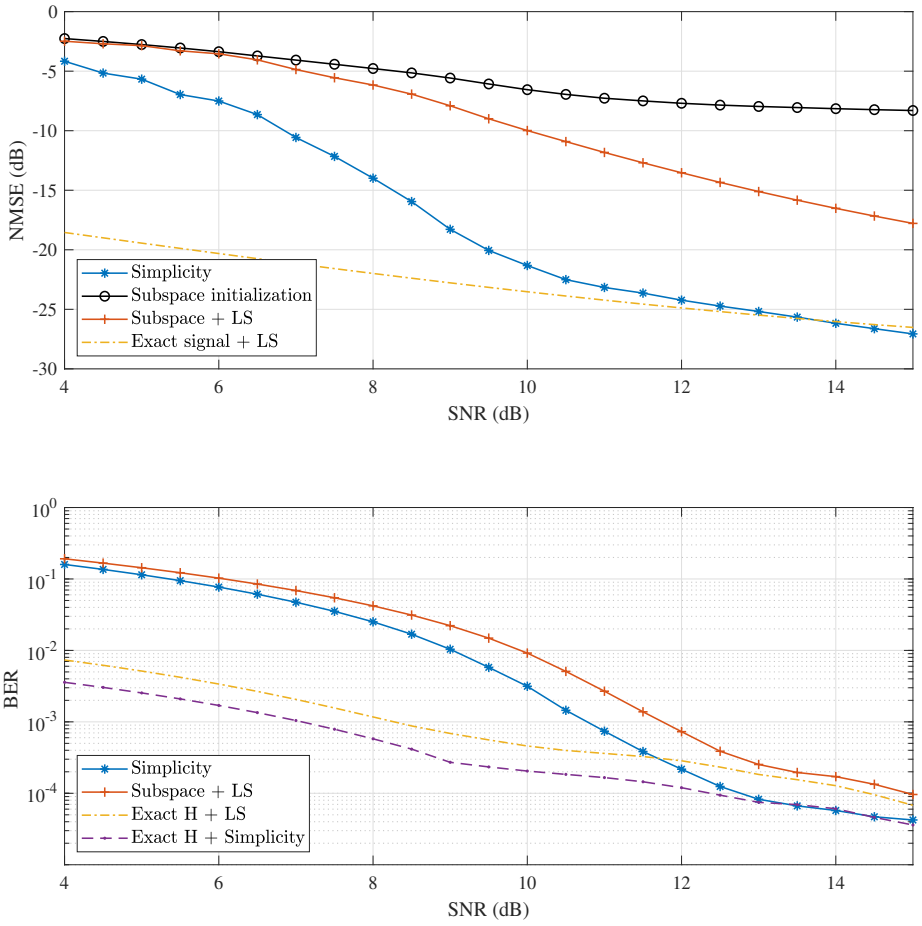

Figure 1: NMSE in dB and BER performance of the regularized-DML versus SNR for BPSK modulation and $(N_t = 3, N_r = 8, L = 10)$ and $T = 1000$.

regularized DML approach. However, we can use multiple iterations to have better performance. Fig. 3 shows the results of using 1, 2, 5 and 10 iterations of the DML approach based on the simplicity prior for the case of BPSK constellation, $(N_t = 3, N_r = 8, L = 10)$ and $T = 1000$ symbols. It is clear that increasing the number of iterations improves the performance, especially for the first ones. In terms of NMSE, we can observe that the amount of SNR needed to reach the oracle performance is going lower with the increase of the iterations number. In terms of BER, the same can be noted, in addition to outperforming the LS

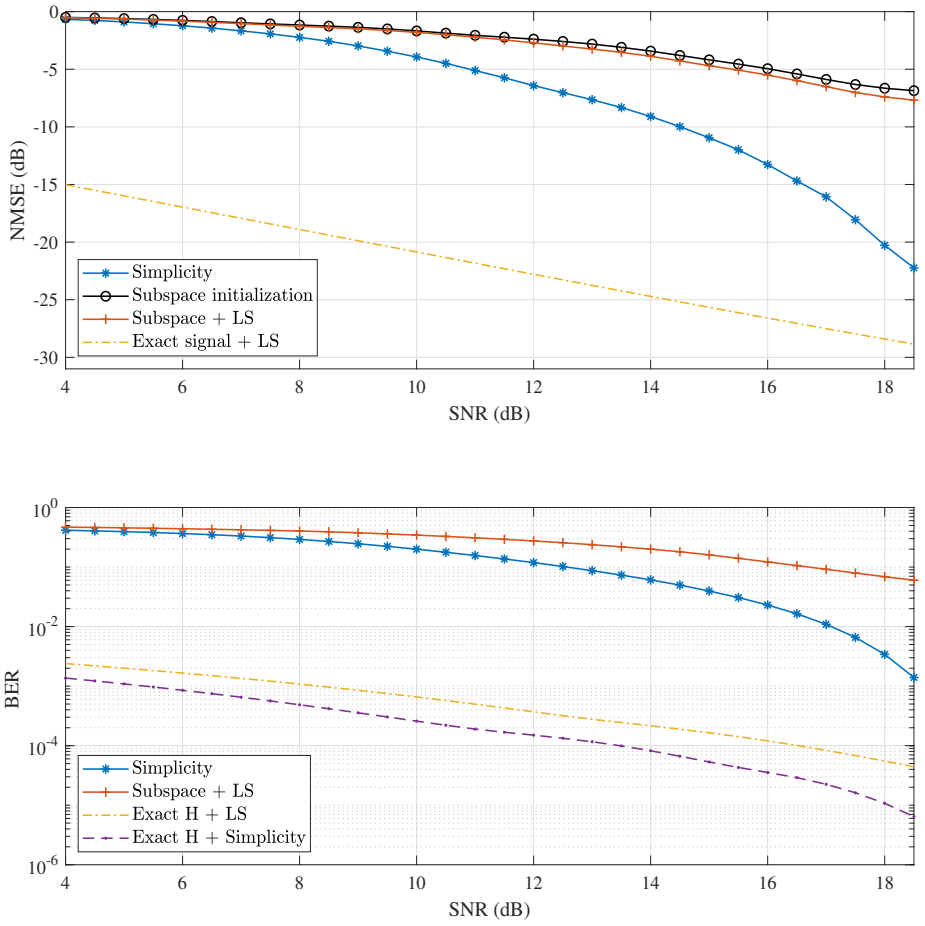

Figure 2: NMSE in dB and BER performance of the regularized-DML versus SNR for QPSK modulation and $(N_t = 3, N_r = 8, L = 10)$ and $T = 1000$.

oracle performance. In the next simulation, we consider changing the number of receivers N_r form 3 to 7 while keeping $N_t = 2$, $L = 2$ and $T = 400$ constant. Figure 4 shows the evolution of the NMSE and BER for different values of SNR (5, 10 and 15 dB) versus the number of receivers N_r. Increasing the number of receivers N_r will result in a significant performance gain especially at low SNR levels. However, if the number of receivers is very large, one has to use a wider data block (greater T) to maintain the estimation accuracy of such large dimensional systems.

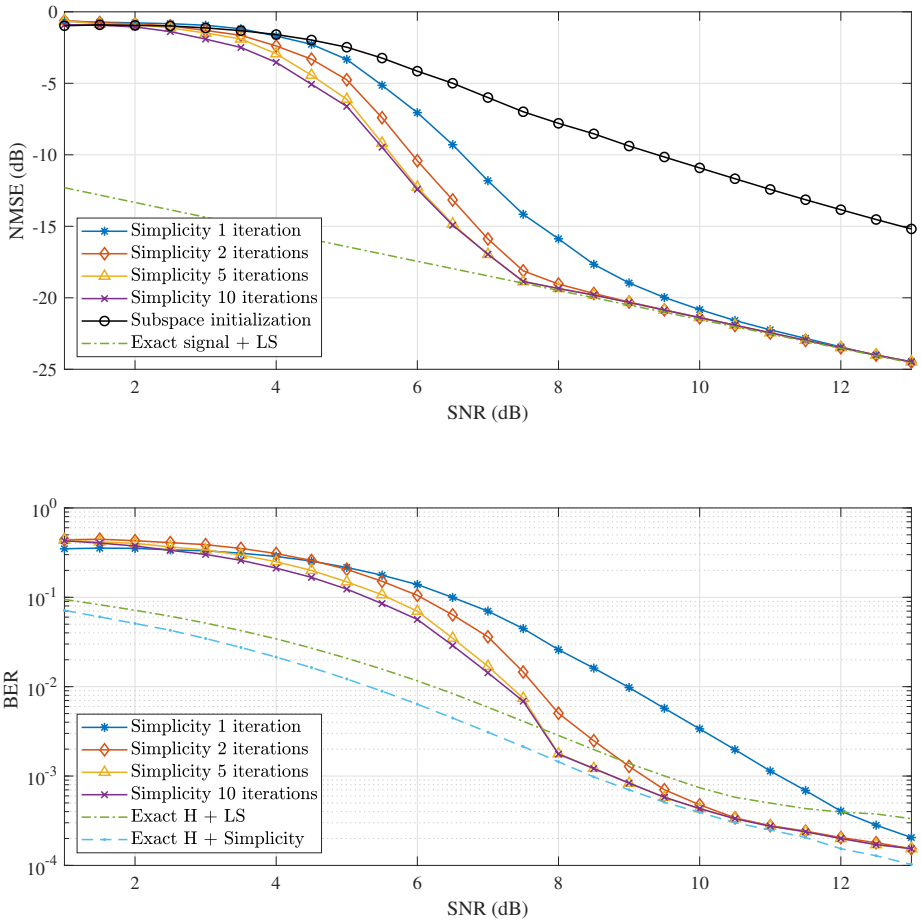

Figure 3: NMSE in dB and BER performance of the regularized-DML with different number of iterations versus SNR for BPSK modulation and $(N_t = 3, N_r = 8, L = 10)$ and $T = 1000$.

Now, we consider in addition to the simplicity prior, the sparsity of the channels impulse responses. We reuse the same parameters of the first simulation with longer channels $L = 20$ which have %30 of their elements non-zeros. Fig. 5 shows the influence of combining both the simplicity of the input signals and the sparsity of the channels ($\lambda_h = 0.5$) on the NMSE and the BER. Adding the sparsity prior allows us to reach even better performance in terms of NMSE and BER in both the oracle case and the full blind case. In order to investigate the effect of the sparsity level (the ratio between the number of non-zero elements of the

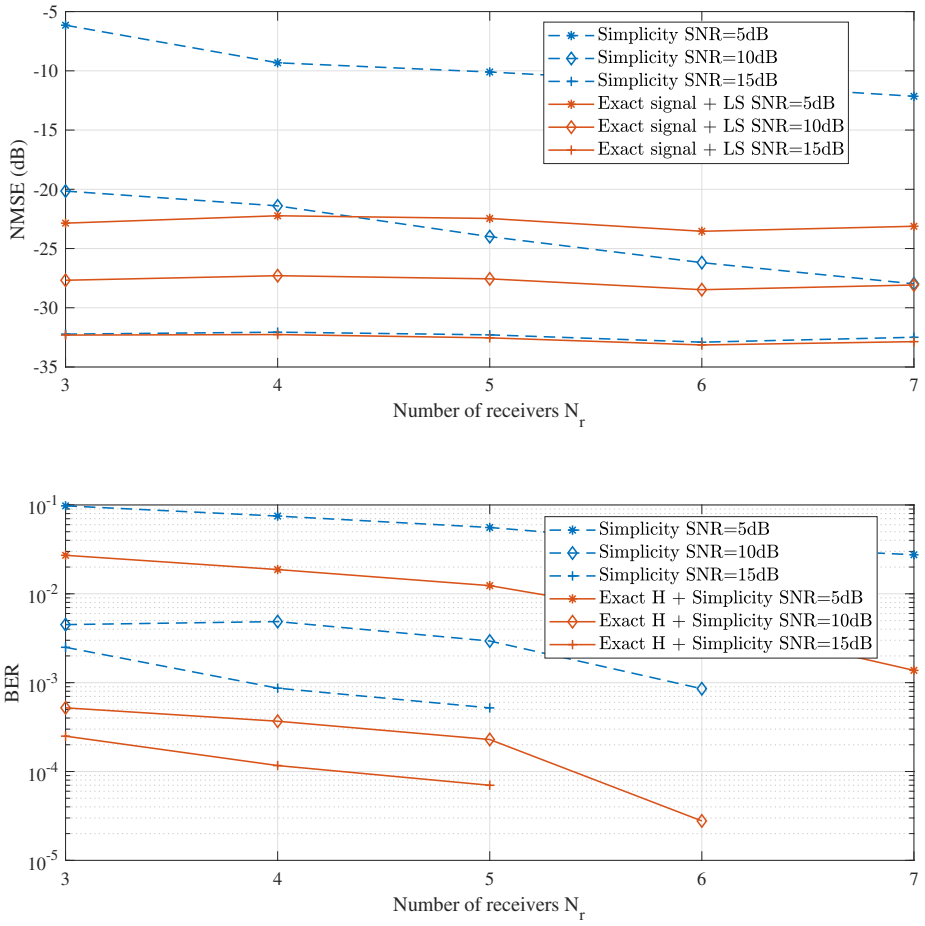

Figure 4: NMSE in dB and BER performance of the regularized-DML for different SNR values versus the number of receivers for QPSK modulation and $(N_t = 2, L = 2)$ and $T = 400$.

channel matrix $\mathbf{H}$ over the total number of elements) on the performance of the proposed solution, we consider the parameters $N_r = 5; N_t = 2, L = 2, T = 400$, QPSK modulation and SNR = 12 dB. Figure 6 shows the NMSE and BER versus the sparsity level ratio changing from 0.1 to 0.9. We have kept the LASSO parameter fixed at $\lambda_h = 0.5$ for all the sparsity levels which explains that the performance is better for certain sparsity levels (mainly between 0.3 and 0.6) and severely deteriorates for other levels. Note that the important choice of the LASSO parameter λ_h depends on the sparsity level and the SNR and its

optimal tuning remains an open problem.

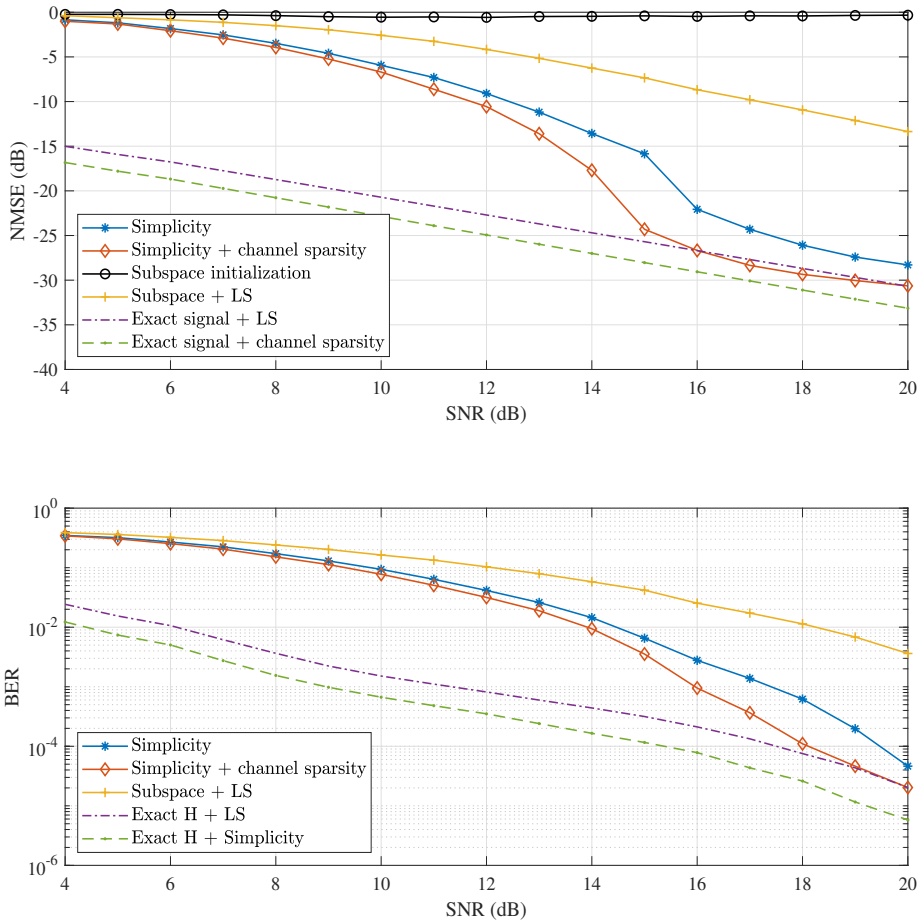

Figure 5: NMSE in dB and BER performance of the regularized-DML versus SNR for BPSK modulation and sparse channel with $(N_t = 3, N_r = 8, L = 20)$ and $T = 1000$.

Most of the works in the literature consider the semi-blind case and the instantaneous channel model due to its relation with the OFDM formulation. Even though the contexts are different (i.e. semi-blind instead of blind and instantaneous channel instead of convolutive channel), we compare our solution with the recent work proposed in [8] in the particular case of instantaneous (no memory) channel. We consider a non-sparse channel with $N_t = 4$, $N_r = 10$ and $L = 0$. We consider a QPSK constellation which will be used for the simplicity

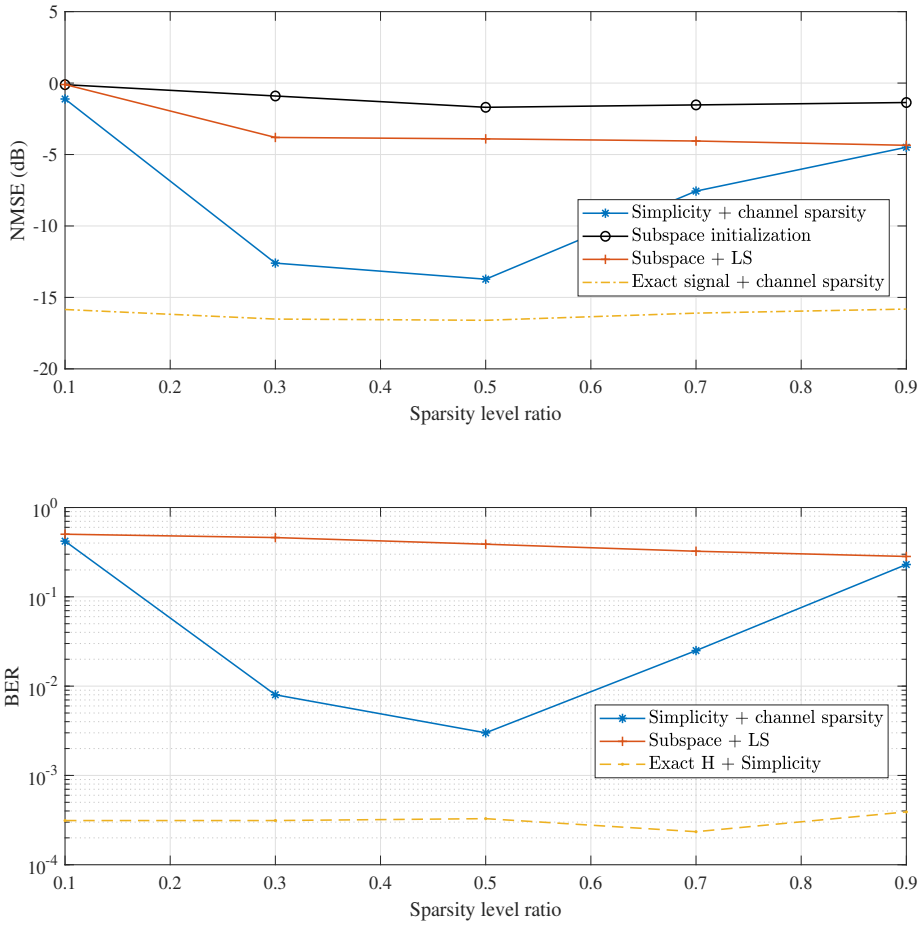

Figure 6: NMSE in dB and BER performance of the regularized-DML versus sparsity level ratio for QPSK modulation and sparse channel with $(N_t = 2, N_r = 5, L = 2)$ and $T = 400$.

prior with $T = 250$ received symbols. The solution in [8] is semi-blind and needs some initial training data. Hence, we have chosen, for each user, the first $N_t + 2 = 6$ symbols as training sequence. Figure 7 shows that the semi-blind method labeled as ref. [8] is slightly better than our solution. However, this small performance difference is admissible since our approach is fully blind compared to ref. [8]. Next, we consider the sparse channel case with parameters $N_t = 4$, $N_r = 15$, $L = 0$, $T = 250$ and QPSK constellation. Figure 8 shows how the performance of the semi-blind method in ref. [8] deteriorates in the

case of sparse channel as compared to the proposed one, especially when we are combining the data simplicity and the channel sparsity priors.

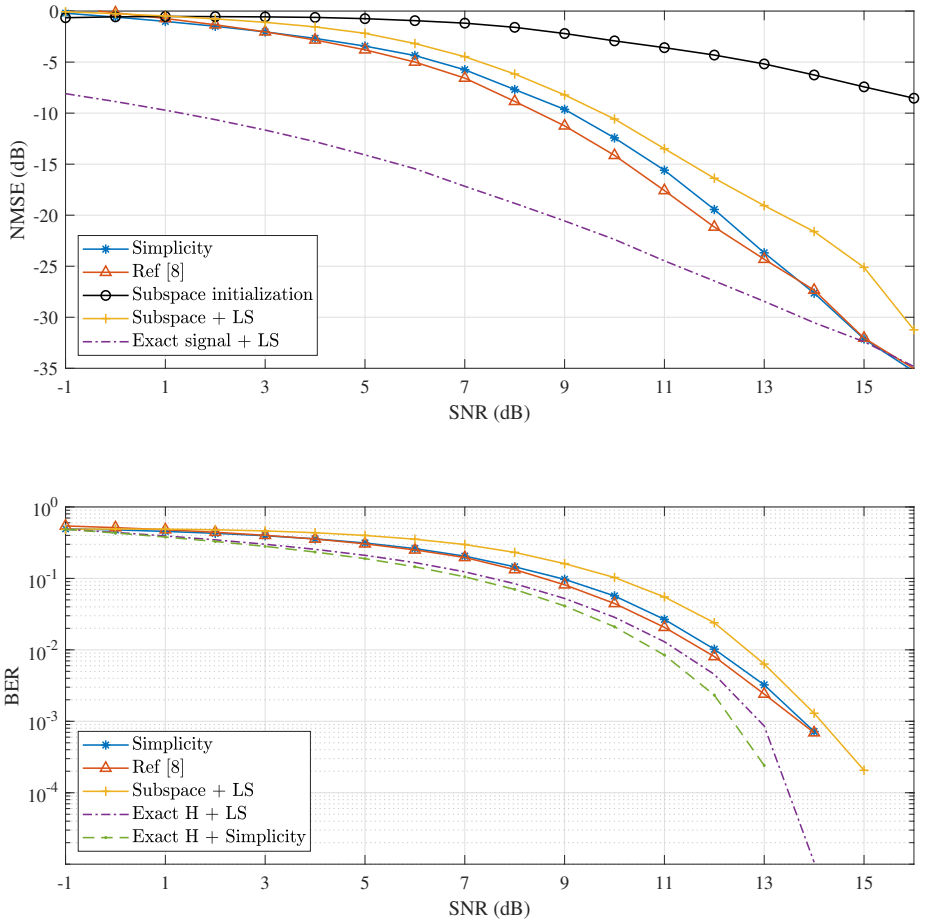

Figure 7: NMSE in dB BER performance of the regularized-DML versus SNR for QPSK modulation with $(N_t = 4, N_r = 10, L = 0)$ and $T = 250$.

In the case of sparse signals, we will use the mean squared error (MSE) $MSE = \frac{1}{T}\mathbb{E}\Big\{\|\mathbf{s}_N - \hat{\mathbf{s}}_N\|_2^2\Big\}$ between the original signal and the estimated one as a performance factor. Fig. 9 shows the improved performance of the proposed approach in case of sparse *a priori* on the input signals compared to the least-squares method with $(N_t = 3, N_r = 8, L = 10)$ and $T = 1000$.

In the next simulation, in addition to the sparsity of the input signals, we

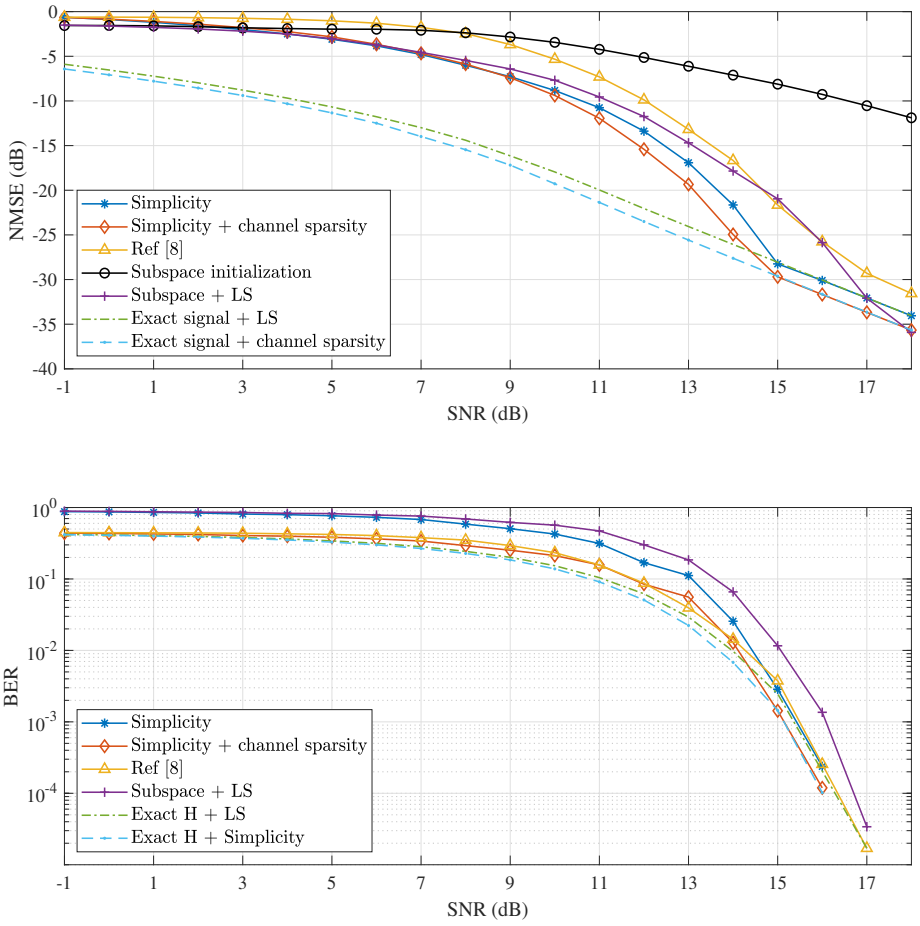

Figure 8: NMSE in dB BER performance of the regularized-DML versus SNR for QPSK modulation and sparse channel with $(N_t = 4, N_r = 20, L = 0)$ and $T = 250$.

consider longer $L = 20$ and sparse channels with 30% of non-zeros elements. Hence, we will use the sparse regularization DML approach on both the channels and the signals. Fig. 10 illustrates how introducing the double sparsity prior enhances the performance. The last simulation considers the case of the sparse signal and the time-varying sparse channel where we introduce a sudden change to the channel matrix $\mathbf{H}$ after 2000 samples. We consider $N_t = 3, N_r = 5, L = 12$ and $SNR = 8dB$ or $SNR = 12dB$. We propose to use the adaptive regularized DML approach with blocks of size 200 and 50% overlapping. We compare the

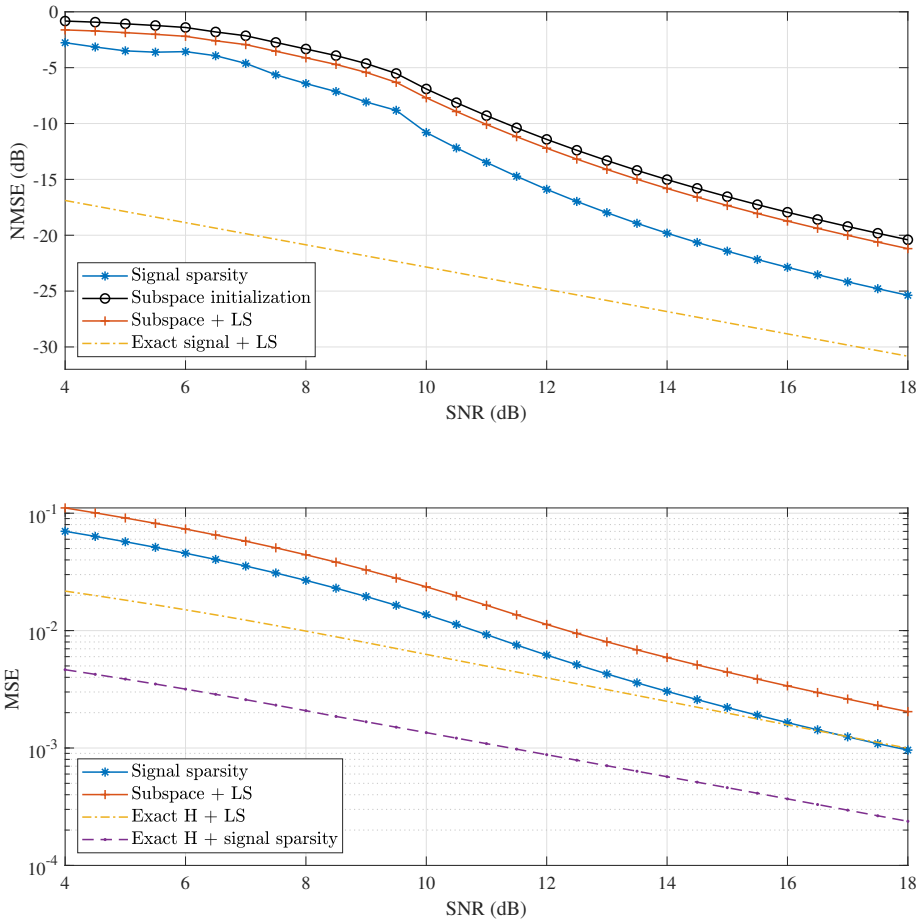

Figure 9: NMSE in dB and MSE performance of the regularized-DML versus SNR for sparse signals with $(N_t = 3, N_r = 8, L = 10)$ and $T = 1000$.

proposed procedure to the oracle cases, where the channel matrix $\mathbf{H}$ or the transmitted data $\mathbf{S}$ are known with or without regularization. We compare also to the case where we use the subspace method as initialization to each block followed by iterative least-squares without regularization. Fig. 11 confirms the superiority of the proposed approach and how its performance approaches the oracle performance for both $SNR = 8dB$ and $SNR = 12dB$. It shows also how the sudden change in the channel impacts the performance. One can reduce the time reaction of the proposed adaptive approach by combining

the previous channel estimation with a subspace-based new estimation at the start of each block. This will enhance the tracking capability but increase the complexity (apply subspace method at each block). The use of ADMM optimization technique reduce considerably the calculation time compared to using CVX toolbox. For instance, CVX toolbox-based solution needs on average 10s to solve the problem compared to 0.2s for the ADMM-based solution. One should keep in mind that the solution proposed by CVX toolbox is based on interior point methods which should give more precise results (the error is not important compared to the calculation time difference).

In the previous simulations, we have chosen ad-hoc parameters λ_s and λ_h. This choice has an important influence on the performance of the proposed formulation. Choosing the "best" LASSO regularization parameter which depends on the sparsity level and the SNR, is an open problem. Some solutions can be adopted from the literature such as those presented in [28].

6. Conclusion

In this paper, we have considered the problem of blind identification of MIMO FIR systems. We have presented a bilinear approach based on the regularized DML formation of the problem. This formulation has the advantage of alternatively estimate the channel impulse responses and the transmitted data while adding the *a priori* information about the problem as a regularization penalty. Different *a priori* are considered: the finite alphabet simplicity or the sparsity of the transmitted data, the sparsity of the channels finite responses. As an initialization of our blind framework, we have proposed to use the subspace blind identification method followed by a step to resolve its full rank matrix ambiguity for both sparse and non-sparse channels. Then, an iterative convex optimization is applied over the channel and the transmitted data. An adaptive version is also proposed to handle the case of time varying channel. Simulations results showed that the proposed method has improved the estimation accuracy which reaches the same level as the oracle solutions above a certain SNR.

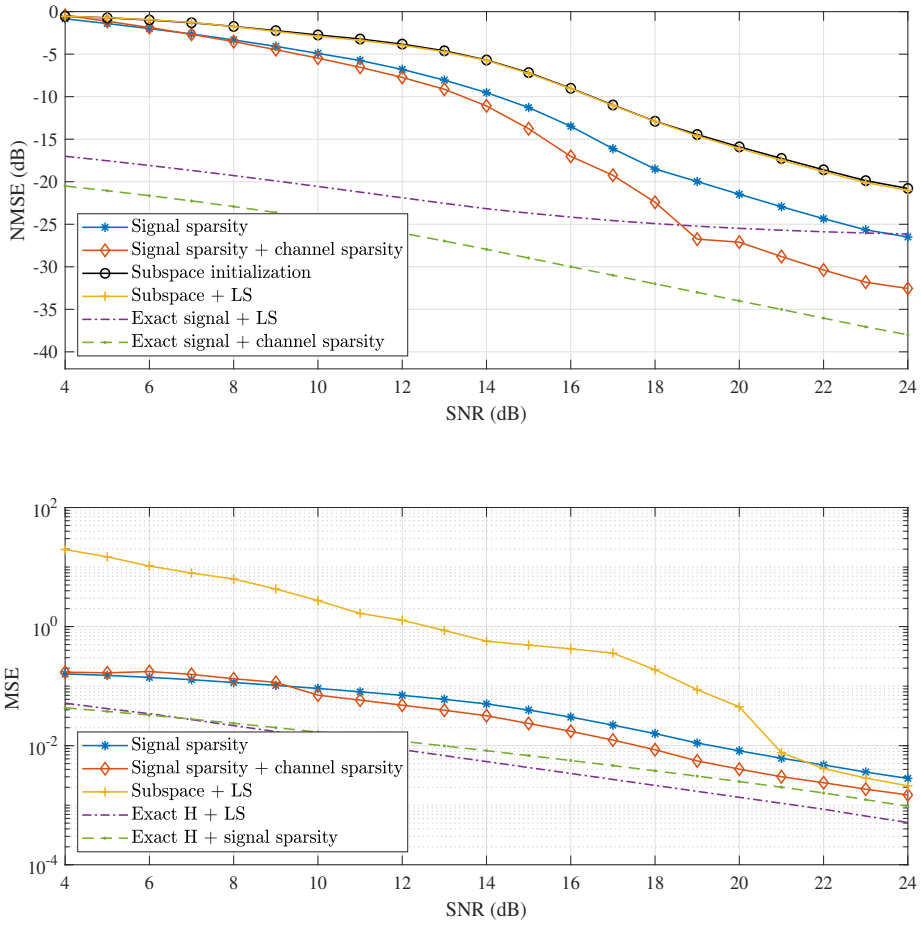

Figure 10: NMSE in dB and MSE performance of the regularized-DML versus SNR for sparse signals and sparse channels with $(N_t = 3, N_r = 8, L = 20)$ and $T = 1000$.

References

[1] M. Abuthinien, S. Chen, A. Wolfgang, L. Hanzo, Joint maximum likelihood channel estimation and data detection for MIMO systems, in: 2007 IEEE International Conference on Communications, 2007, pp. 5354–5358.

[2] M. Abuthinien, S. Chen, L. Hanzo, Semi-blind joint maximum likelihood channel estimation and data detection for MIMO systems, IEEE Signal Processing Letters 15 (2008) 202–205.

[3] H. R. Palally, S. Chen, W. Yao, L. Hanzo, Particle swarm optimisation aided

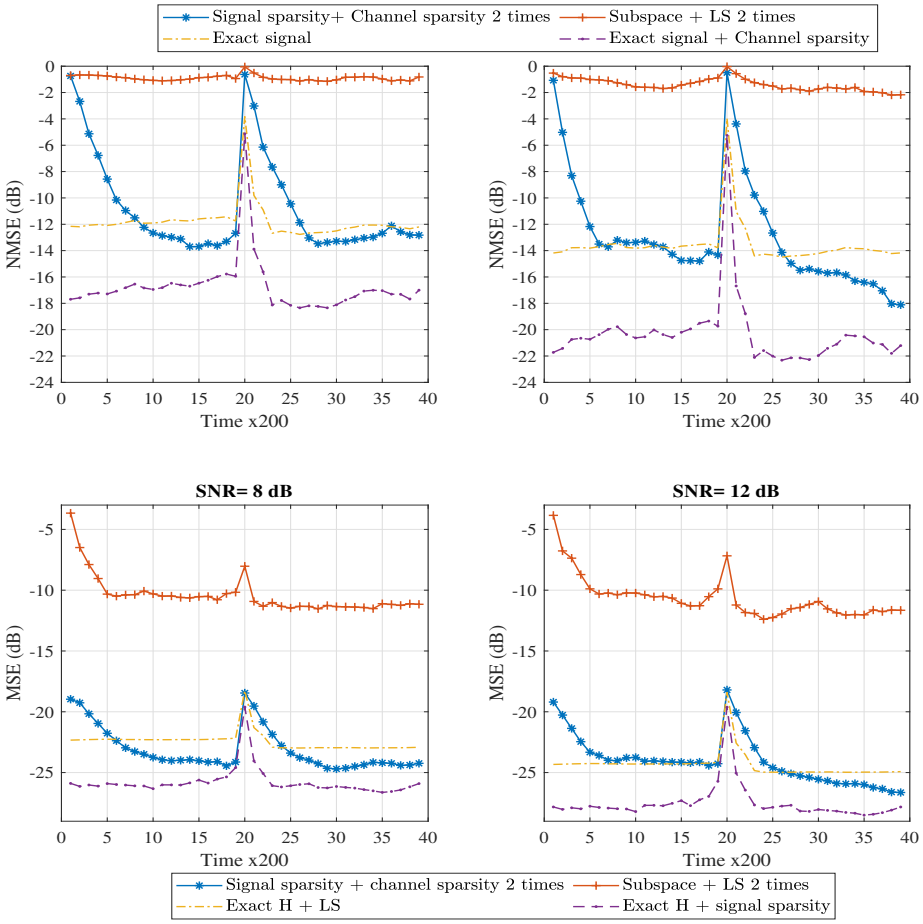

Figure 11: NMSE in dB and MSE in dB performance of the adaptive regularized-DML versus time in steps (step=100 samples) for sparse signals and sparse channels with $(N_t = 3, N_r = 5, L = 12)$, $SNR = 8dB$ (left) and $SNR = 12dB$ (right).

semi-blind joint maximum likelihood channel estimation and data detection for MIMO systems, in: IEEE Workshop on Statistical Signal Processing, 2009, pp. 309–312.

[4] C. Wen, C. Wang, S. Jin, K. Wong, P. Ting, Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs, IEEE Transactions on Signal Processing 64 (10) (2016) 2541–2556.

[5] J. Ma, L. Ping, Data-aided channel estimation in large antenna systems, in: 2014

IEEE International Conference on Communications (ICC), 2014, pp. 4626–4631.

[6] R. Prasad, C. R. Murthy, B. D. Rao, Joint approximately sparse channel estimation and data detection in OFDM systems using sparse bayesian learning, IEEE Transactions on Signal Processing 62 (14) (2014) 3591–3603.

[7] T. Ishihara, S. Sugiura, Iterative frequency-domain joint channel estimation and data detection of faster-than-nyquist signaling, IEEE Transactions on Wireless Communications 16 (9) (2017) 6221–6231.

[8] X. Hong, J. Gao, S. Chen, Semi-blind joint channel estimation and data detection on sphere manifold for MIMO with high-order QAM signaling, Journal of the Franklin Institute 357 (9) (2020) 5680 – 5697.

[9] M. I. Torrico, M. Li, D. A. Pados, Joint channel estimation and data detection on commercially available underwater acoustic modems, in: OCEANS 2014 - TAIPEI, 2014, pp. 1–5.

[10] R.-N. Yang, W.-T. Zhang, S.-T. Lou, Joint adaptive blind channel estimation and data detection for MIMO-OFDM systems, Wireless Communications and Mobile Computing 2020 (2020) 1–9.

[11] L. Tong, S. Perreau, Multichannel blind identification: from subspace to maximum likelihood methods, Proceedings of the IEEE 86 (10) (1998) 1951–1968.

[12] Y. Hua, Fast maximum likelihood for blind identification of multiple FIR channels, IEEE Transactions on Signal Processing 44 (3) (1996) 661–672.

[13] D. T. M. Slock, C. B. Padias, Further results on blind identification and equalization of multiple FIR channels, in: International Conference on Acoustics, Speech, and Signal Processing, Vol. 3, 1995, pp. 1964–1967.

[14] M. Feder, J. A. Catipovic, Algorithms for joint channel estimation and data recovery-application to equalization in underwater communications, IEEE Journal of Oceanic Engineering 16 (1) (1991) 42–55.

[15] D. Gesbert, P. Duhamel, S. Mayrargue, Blind least-squares approaches for joint data/channel estimation, in: IEEE Digital Signal Processing Workshop Proceedings, 1996, pp. 450–453.

[16] F. Alberge, P. Duhamel, M. Nikolova, Adaptive solution for blind identification/equalization using deterministic maximum likelihood, IEEE Transactions on Signal Processing 50 (4) (2002) 923–936.

[17] N. Seshadri, Joint data and channel estimation using fast blind trellis search techniques, in: [Proceedings] GLOBECOM '90: IEEE Global Telecommunications Conference and Exhibition, Vol. 3, 1990, pp. 1659–1663.

[18] M. Ghosh, C. L. Weber, Maximum-likelihood blind equalization, in: Proc. SPIE, Adaptive Signal Processing, Vol. 1565, 1991, pp. 188–195.

[19] S. Talwar, M. Viberg, A. Paulraj, Blind estimation of multiple co-channel digital signals using an antenna array, IEEE Signal Processing Letters 1 (2) (1994) 29–31.

[20] F. Alberge, M. Nikolova, P. Duhamel, Blind identification/equalization using deterministic maximum likelihood and a partial prior on the input, IEEE Transactions on Signal Processing 54 (2) (2006) 724–737.

[21] Z. Hajji, A. Aïssa-El-Bey, K. Amis, Simplicity-based recovery of finite-alphabet signals for large-scale MIMO systems, Digital Signal Processing 80 (2018) 70 – 82.

[22] M. Zibulevsky, B. A. Pearlmutter, Blind source separation by sparse decomposition in a signal dictionary, Neural Computation 13 (4) (2001) 863–882.

[23] R. Gribonval, M. Zibulevsky, Sparse component analysis, in: Handbook of Blind Source Separation, Academic Press, 2010, pp. 367 – 420.

[24] A. Aïssa-El-Bey, D. Kimura, H. Seki, T. Taniguchi, Blind and semi-blind sparse channel identification in MIMO OFDM systems, in: 2011 IEEE International Conference on Communications (ICC), 2011, pp. 1–5.

[25] K. Abed-Meraim, P. Loubaton, E. Moulines, A subspace algorithm for certain blind identification problems, IEEE Transactions on Information Theory 43 (2) (1997) 499–511.

[26] S. Visuri, V. Koivunen, Resolving ambiguities in subspace-based blind receiver for MIMO channels, in: Asilomar Conference on Signals, Systems and Computers, Vol. 1, 2002, pp. 589–593.

[27] D. L. Donoho, J. Tanner, Counting faces of randomly projected polytopes when the projection radically lowers dimension, Journal of the American Mathematical Society 22 (1) (2009) 1–53.

[28] T. Hastie, R. Tibshirani, M. Wainwright, Statistical Learning with Sparsity: The LASSO and Generalizations, Chapman & Hall/CRC, 2015.

[29] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, The Annals of Statistics 32 (2) (2004) 407–451.

[30] E. J. Candès, M. B. Wakin, S. P. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, Journal of Fourier Analysis and Applications 14 (5) (2008) 877–905.

[31] N. Lassami, K. Abed-Meraim, A. Aïssa-El-Bey, Low cost subspace tracking algorithms for sparse systems, in: 25th European Signal Processing Conference (EUSIPCO), 2017, pp. 1400–1404.

[32] N. Lassami, A. Aïssa-El-Bey, K. Abed-Meraim, Low cost sparse subspace tracking algorithms, Signal Processing 173 (2020) 107522.

[33] R. Badeau, B. David, G. Richard, Fast approximated power iteration subspace tracking, IEEE Transactions on Signal Processing 53 (8) (2005) 2931–2941.

[34] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1, `http://cvxr.com/cvx` (Mar. 2014).

[35] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004. doi:10.1017/CBO9780511804441.

[36] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Foundations and Trends in Machine Learning 3 (1) (2011) 1122.