2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)

Semi-blind Sparse Channel Estimation and Data Detection by Successive Convex Approximation

Ouahbi Rekik
L2TI, UR 3043, F-93430,
Université Sorbonne Paris Nord
Villetaneuse, France
ouahbi.rekik@univ-paris13.fr

Karim Abed-Meraim
PRISME Laboratory
Université d'Orléans
Orléans, France
karim.abed-meraim@univ-orleans.fr

Marius Pesavento
Communication Systems Group
Darmstadt University of Technology
Darmstadt, Germany
mpesa@nt.tu-darmstadt.de

Anissa Mokraoui
L2TI, UR 3043, F-93430, Université Sorbonne Paris Nord
Villetaneuse, France. E-mail: anissa.mokraoui@univ-paris13.fr

***Abstract*—The aim of this paper is to propose a semi-blind solution, for joint sparse channel estimation and data detection, based on the successive convex approximation approach. The optimization is performed on an approximate convex problem, rather than the original nonconvex one. By exploiting available data and system structure, an iterative procedure is proposed where the channel coefficients and data symbols are updated simultaneously at each iteration. Also an optimized step size, introduced according to line search procedure, is used for convergence improvement with guaranteed convergence to a stationary point. Simulation results show that the proposed solution exhibits fast convergence with very attractive channel and data estimation performance.**

***Index Terms*—semi-blind channel and data estimation, optimal step size, SCA, ADMM, BCD.**

I. Introduction

Channel estimation and data detection is a key task in wireless communications systems and different approaches have been proposed for joint channel and data estimation (e.g. [1], [2]). Moreover, In many wireless communications, growing experimental studies have showed that many practical channels exhibit sparsity as the delay spread could be very large but the number of distinguishable multi-path delays is usually small. In such a case, sparse channel estimation should be considered (see [3] and references therein).

A solution to perform channel identification and data recovery, by taking into account prior information about the channel and/or the data, would be solving a regression-based optimization problem (OP). Many techniques have been proposed for the problem of constrained linear regression, such as the fast iterative soft thresholding algorithm (FISTA) [4], the block coordinate descent (BCD) algorithm [5], the alternating direction method of multiplier (ADMM) [6] or the parallel best-response with exact line search algorithm [7]. In our context, the sequential update algorithm seems to be a good alternative to overcome difficulties arising from the nonconvexity of the problem, however, a major drawback of the sequential update is the induced large delay, because the update of a block variable cannot be performed until its predecessor block variable is updated (BCD algorithm [5]). In such a case, the delay may be very large with big number of data blocks.

Conseqently, an appropriate approach would be the successive convex approximation framework ([8] and references therein), where a sequence of successively refined approximate problems are solved, while preserving the algorithm's convergence to a stationary point of the original OP.

On the other hand, it bears mentioning that, in practice some training sequences, known by the transmitter and the receiver, are usually sent periodically within the wireless network frames besides the unknown data. Thus, the focus of this paper is to propose a solution for semi-blind sparse channel estimation and data recovery by considering a single-carrier single-input multiple-output (SIMO) system[1]. The motivation for adopting a semi-blind approach is to make use of available sequences and to avoid the different difficulties and issues that emerge from the blind process such as the inherent ambiguity of blind processing [9]. The proposed solution is based on the successive convex approximation (SCA) framework along with the majorization-maximization (MM) approach [10]. An iterative procedure is performed where channel coefficients and data symbols are estimated simultaneously, at each iteration, with an optimized step size introduced for improving the convergence speed.

II. System model

In this paper, a single-input multiple-output (SIMO) convolutive communications system is considered, where the r-th system output, $r = 1, \cdots, N_r$, is given by:

$$y_r(k) = \sum_{n=0}^{M} h_r(n)s(k-n) + v_r(k), \tag{1}$$

where h_r refers to the r-th channel finite impulse response of size $M+1$, $s(k)$ represents the transmitted symbols and $v_r(k)$

[1]This solution can easily be extended to the multi-user case (i.e. convolutive MIMO systems).

978-1-7281-5478-7/20/$31.00 ©2020 IEEE

is a white Gaussian noise with variance σ_v^2. By considering N_s received samples, one can write (1) as follows:

$$\mathbf{Y} = \mathbf{H}\mathbf{S} + \mathbf{V}, \tag{2}$$

where $\mathbf{Y}, \mathbf{V} \in \mathbb{C}^{N_r \times N_s}$, $\mathbf{H} \in \mathbb{C}^{N_r \times (M+1)}$, and $\mathbf{S} \in \mathbb{C}^{(M+1)\times N_s}$ is a Toeplitz matrix. Moreover, and without loss of generality, N_p training symbols (pilots) are sent at the beginning of the data frame, so that the transmitted and received symbols are given by $\bar{\mathbf{S}} = [\mathbf{S}_p, \mathbf{S}]$ and $\bar{\mathbf{Y}} = [\mathbf{Y}_p, \mathbf{Y}]$ respectively.

III. PROPOSED CHANNEL ESTIMATION AND DATA RECOVERY FRAMEWORK

This section is dedicated to the formulation of the appropriate OP describing the aforementioned scenario, then the derivation of the proposed solution.

A. Problem formulation

In what follows, this paper considers the problem of joint channel estimation and data recovery, by taking into account the channel sparsity as prior information. Also, a number of data symbols (pilots) are assumed to be known by the transmitter and the receiver. To do so, one can formulate an appropriate optimization problem (OP) that incorporates all available model information and requirements. Basically, our OP is composed of the estimation error of a data matching function and regularization (penalty) terms for promoting, in the solution, a certain structure known *a priori*. Hence, the following OP is considered:

$$\underset{\mathbf{H},\mathbf{S}}{\text{minimize}} \ \frac{1}{2}\|\bar{\mathbf{Y}} - \mathbf{H}\bar{\mathbf{S}}\|_F^2 + \mu\|\mathbf{H}\|_1, \tag{3}$$

where the l_1 norm enforces the sparsity of the channel response $\mathbf{H}$, with a regularization constant μ, whereas the term $\|\bar{\mathbf{Y}} - \mathbf{H}\bar{\mathbf{S}}\|_F^2$ can be expressed as the sum of a data (blind) and a pilot-based terms as follows:

$$\begin{aligned} \|\bar{\mathbf{Y}} - \mathbf{H}\bar{\mathbf{S}}\|_F^2 &= \|[\mathbf{Y}_p, \mathbf{Y}] - \mathbf{H}[\mathbf{S}_p, \mathbf{S}]\|_F^2 \\ &= \|\mathbf{Y}_p - \mathbf{H}\mathbf{S}_p\|_F^2 + \|\mathbf{Y} - \mathbf{H}\mathbf{S}\|_F^2. \end{aligned} \tag{4}$$

where $\mathbf{Y}_p$ and $\mathbf{S}_p$ refer respectively to the received and transmitted pilot symbols.

Moreover, given equation (1), the data matrix $\mathbf{S}$ has a Teoplitz structure, which can be added as a constraint to our OP (given by (3)). To do so, we define the matrix $\mathbf{S}_L$ which is formed by removing the last row and column of $\mathbf{S}$, and the matrix $\mathbf{S}_F$ which is formed by removing the first row and column of $\mathbf{S}$. In such a case, we have: $\mathbf{S}_L = \mathbf{S}_F$ where:

$$\mathbf{S}_L = \mathbf{J}_{L_1}\mathbf{S}\mathbf{J}_{L_2}, \tag{5}$$
$$\mathbf{S}_F = \mathbf{J}_{F_1}\mathbf{S}\mathbf{J}_{F_2}, \tag{6}$$

where $\mathbf{J}_{L_1} = [\mathbf{I}_M, \mathbf{0}_{M\times 1}]$, $\mathbf{J}_{L_2} = [\mathbf{I}_{N_s-1}, \mathbf{0}^T_{(N_s-1)\times 1}]^T$, $\mathbf{J}_{F_1} = [\mathbf{0}_{M\times 1}, \mathbf{I}_M]$, $\mathbf{J}_{F_2} = [\mathbf{0}^T_{(N_s-1)\times 1}, \mathbf{I}_{N_s-1}]^T$, $\mathbf{I}$ and $\mathbf{0}$ refer to the identity matrix and all-zero matrix, respectively.

Thus, we can write:

$$\mathbf{S}_L - \mathbf{S}_F = \tilde{\mathbf{M}}\text{vec}(\mathbf{S}), \tag{7}$$

where $\tilde{\mathbf{M}} = \mathbf{J}^T_{L_2} \otimes \mathbf{J}_{L_1} - \mathbf{J}^T_{F_2} \otimes \mathbf{J}_{F_1}$ ($\otimes$ being the Kronecker product) and vec(.) denotes the matrix vectorization operator. Consequently, the optimization problem, given by equation (3), can be re-expressed as follows:

$$\underset{\mathbf{H},\mathbf{S}}{\text{minimize}} \ \frac{1}{2}\|\bar{\mathbf{Y}} - \mathbf{H}\bar{\mathbf{S}}\|_F^2 + \frac{\lambda}{2}\|\tilde{\mathbf{M}}\text{vec}(\mathbf{S})\|_F^2 + \mu\|\mathbf{H}\|_1. \tag{8}$$

N.B.: This OP can be easily extended to the multi-user case (i.e. MIMO system) by modifying properly the matrix $\tilde{\mathbf{M}}$.

B. Proposed solution

To solve the aforementionned OP, the successive convex approximation approach proposed in [8] is adopted. Basically, this approach deals with OP given by:

$$\underset{\mathbf{Z}\in\mathcal{Z}}{\text{minimize}} \ h(\mathbf{Z}) = f(\mathbf{Z}) + g(\mathbf{Z}), \tag{9}$$

where $f(.)$ is a smooth nonconvex function, $g(.)$ is a regularization nonsmooth function and $\mathcal{Z}$ is a convex set.

First, an upper bound of the original function h is constructed, by the standard MM method, then, a convex approximation of this upper bound is defined, based on the standard SCA framework. Finally, the obtained function is minimized so that it has the same optimal points as the original one (see [8] for proofs). Also, a line search based procedure is introduced for calculating an optimal step size.

In the sequel, the following notation is adopted:

$$\mathbf{Z} = (\mathbf{H}, \mathbf{S}), \tag{10}$$
$$f(\mathbf{Z}) = \frac{1}{2}\|\bar{\mathbf{Y}} - \mathbf{H}\bar{\mathbf{S}}\|_F^2 + \frac{\lambda}{2}\|\tilde{\mathbf{M}}\text{vec}(\mathbf{S})\|_F^2, \tag{11}$$
$$g(\mathbf{Z}) = \mu\|\mathbf{H}\|_1. \tag{12}$$

One can notice that $f(\mathbf{H}, \mathbf{S})$ is not jointly convex w.r.t. $(\mathbf{H}, \mathbf{S})$ but it is individual convex in $\mathbf{H}$ and $\mathbf{S}$. This leads to the best OP approximation: given fixed values $\mathbf{Z}^t = (\mathbf{H}^t, \mathbf{S}^t)$ at iteration t, the original nonconvex function $f(\mathbf{Z})$ is upper bounded by a proximal convex function $\tilde{f}(\mathbf{Z}, \mathbf{Z}^t)$ given by:

$$\tilde{f}(\mathbf{Z}, \mathbf{Z}^t) = \tilde{f}_H(\mathbf{H}, \mathbf{Z}^t) + \tilde{f}_S(\mathbf{S}, \mathbf{Z}^t), \tag{13}$$

where

$$\tilde{f}_H(\mathbf{H}, \mathbf{Z}^t) = f(\mathbf{H}, \mathbf{S}^t) = \frac{1}{2}\|\bar{\mathbf{Y}} - \mathbf{H}\bar{\mathbf{S}}^t\|_F^2, \tag{14}$$

$$\tilde{f}_S(\mathbf{S}, \mathbf{Z}^t) = f(\mathbf{H}^t, \mathbf{S}) = \frac{1}{2}\|\mathbf{Y} - \mathbf{H}^t\mathbf{S}\|_F^2 + \frac{\lambda}{2}\|\tilde{\mathbf{M}}\text{vec}(\mathbf{S})\|_F^2, \tag{15}$$

and $\mathbf{S}^t$ (resp. $\mathbf{H}^t$) refers to a fixed value of the parameter $\mathbf{S}$ (resp. $\mathbf{H}$). In such a case, at iteration t, the approximate problem consists of minimizing:

$$\underset{\mathbf{H},\mathbf{S}}{\text{minimize}} \ \tilde{f}(\mathbf{Z}, \mathbf{Z}^t) + g(\mathbf{Z}). \tag{16}$$

Since $\tilde{f}(\mathbf{Z}, \mathbf{Z}^t)$ is a convex function w.r.t. $\mathbf{Z}$ and $g(\mathbf{Z})$ is convex in $\mathbf{H}$, the approximate problem, given by equation (16), is strongly convex and has a unique globally optimal solution, which is denoted by $\mathbb{B}\mathbf{Z}^t = (\mathbb{B}_H\mathbf{Z}^t, \mathbb{B}_S\mathbf{Z}^t)$. Moreover, the approximate problem (16) is separable w.r.t. variables $\mathbf{H}$ and $\mathbf{S}$, so that it can be decomposed into smaller problems that can be solved in parallel:

$$\mathbb{B}_H\mathbf{Z}^t = \arg\min_{\mathbf{H}} \ \tilde{f}_H(\mathbf{H}, \mathbf{Z}^t) + g(\mathbf{H}), \tag{17}$$
$$\mathbb{B}_S\mathbf{Z}^t = \arg\min_{\mathbf{S}} \ \tilde{f}_S(\mathbf{S}, \mathbf{Z}^t). \tag{18}$$

In order to compute $\mathbb{B}_H\mathbf{Z}^t$ while $g(\mathbf{H})$ is not differentiable, the elements of $\mathbf{H}$ are updated element-wise according to:

$$\mathbb{B}_H\mathbf{Z}^t = \text{diag}(\bar{\mathbf{S}}^t\bar{\mathbf{S}}^{t^H})\mathbb{S}_\mu\left(\text{diag}(\bar{\mathbf{S}}^t\bar{\mathbf{S}}^{t^H})\mathbf{H}^{t^H} - \bar{\mathbf{S}}^t(\bar{\mathbf{S}}^{t^H}\mathbf{H}^H - \bar{\mathbf{Y}}^H)\right), \tag{19}$$

where $\mathbb{S}_\mu(\mathbf{X})$ is an element-wise soft-thresholding function so that its complex (i,j)-th element is given by $[\text{real}(X_{i,j})-\mu]^+ - [-\text{real}(X_{i,j})-\mu]^+ + j[[\text{imag}(X_{i,j})-\mu]^+ - [-\text{imag}(X_{i,j})-\mu]^+]$ with $[x]^+ = \max(x,0)$.

On the other hand, we have:

$$\begin{aligned}\mathbb{B}_s\mathbf{Z}^t &= \arg\min_{\mathbf{s}} \frac{1}{2}\|\mathbf{y}-\tilde{\mathbf{H}}^t\mathbf{s}\|_2^2 + \frac{\lambda}{2}\|\tilde{\mathbf{M}}\mathbf{s}\|_F^2 \\ &= (\tilde{\mathbf{H}}^H\tilde{\mathbf{H}} + \lambda\tilde{\mathbf{M}}^H\tilde{\mathbf{M}})^{-1}\tilde{\mathbf{H}}^H\mathbf{y}, \end{aligned} \tag{20}$$

where $\tilde{\mathbf{H}} = \mathbf{I}\otimes\mathbf{H}$, $\mathbf{y} = \text{vec}(\mathbf{Y})$ and $\mathbf{s} = \text{vec}(\mathbf{S})$ (or equivalently $\mathbf{S} = \text{unvec}(\mathbf{s})$).

By using these optimal solutions, the variables update is given by:

$$\mathbf{H}^{t+1} = \mathbf{H}^t + \gamma(\mathbb{B}_H\mathbf{Z}^t - \mathbf{H}^t), \tag{21}$$
$$\mathbf{S}^{t+1} = \mathbf{S}^t + \gamma(\mathbb{B}_S\mathbf{Z}^t - \mathbf{S}^t), \tag{22}$$

where $\gamma \in [0,1]$ is the algorithm's step size (see [8] for more details).

As can be noticed, $\mathbf{H}$ and $\mathbf{S}$ are updated simultaneously at each iteration based only on the old solutions of the approximate problems (19) and (20), respectively. Also, the approximate problem (16) can be solved efficiently because the optimal solutions are provided in analytic expressions.

C. Optimal step size computation

One can notice that in equations (21) and (22), the choice of the step size is crucial for the convergence speed and accuracy. Therefore an optimal step size would notably improve such characteristics. For this, a line search can be adopted to obtain an optimal step size value as follows:

$$\gamma_{opt} = \arg\min_{\gamma\in[0,1]} [f(\mathbf{Z}^t + \gamma(\mathbb{B}\mathbf{Z}^t - \mathbf{Z}^t)) + g(\mathbf{Z}^t + \gamma(\mathbb{B}\mathbf{Z}^t - \mathbf{Z}^t))]. \tag{23}$$

Although it is a scalar problem, it has no closed form solution due to the non differentiable function g. To overcome this limitation, we use the Jensen's inequality:

$$g(\mathbf{Z}^t + \gamma(\mathbb{B}\mathbf{Z}^t - \mathbf{Z}^t)) \le g(\mathbf{Z}^t) + \gamma g(\mathbb{B}\mathbf{Z}^t - \mathbf{Z}^t)). \tag{24}$$

One can notice that, the function on the right hand side of (24) is differentiable and linear with respect to γ. Hence, a closed form expression, of an approximate optimal step size, is obtained by mimimizing the following polynomial function:

$$\begin{aligned}\gamma_{opt} &= \arg\min_{\gamma\in[0,1]} [f(\mathbf{Z}^t + \gamma(\mathbb{B}\mathbf{Z}^t - \mathbf{Z}^t)) + \gamma(g(\mathbb{B}_H\mathbf{Z}) - g(\mathbf{H}^t))] \\ &= \arg\min_{\gamma\in[0,1]} \{\frac{1}{4}a\gamma^4 + \frac{1}{3}b\gamma^3 + \frac{1}{2}c\gamma^2 + d\gamma\}, \end{aligned} \tag{25}$$

where terms independent of γ are omitted, and:

$$a = 2\|\Delta\mathbf{H}\Delta\mathbf{S}\|_F^2, \tag{26}$$
$$b = 3\text{tr}(\text{real}(\Delta\mathbf{H}\Delta\mathbf{S}(\mathbf{H}\Delta\mathbf{S} + \Delta\mathbf{H}\mathbf{S})^H)), \tag{27}$$
$$\begin{aligned} c &= 2\text{tr}(\text{real}(\Delta\mathbf{H}\Delta\mathbf{S}(\mathbf{H}\mathbf{S}-\mathbf{Y})^H)) + \|\mathbf{H}\Delta\mathbf{S} \\ &+ \Delta\mathbf{H}\mathbf{S}\|_2^2 + \lambda\|\tilde{\mathbf{M}}\Delta\mathbf{s}\|_2^2, \end{aligned} \tag{28}$$
$$\begin{aligned} d &= \text{tr}(\text{real}((\mathbf{H}\Delta\mathbf{S} + \Delta\mathbf{H}\mathbf{S})(\mathbf{H}\mathbf{S}-\mathbf{Y})^H)) \\ &+ \lambda\mathbf{s}^H\tilde{\mathbf{M}}^H\tilde{\mathbf{M}}\Delta\mathbf{s} + \mu(\|\mathbb{B}_H\mathbf{Z}^t\|_1 - \|\mathbf{H}^t\|_1), \end{aligned} \tag{29}$$

where tr(.) is the trace of a matrix and $\Delta\mathbf{H} = \mathbb{B}_H\mathbf{Z}^t - \mathbf{H}^t$, $\Delta\mathbf{S} = \mathbb{B}_S\mathbf{Z}^t - \mathbf{S}^t$, and $\Delta\mathbf{s} = \mathbb{B}_s\mathbf{Z}^t - \mathbf{s}^t$.

The proposed solution will be named ST-SCA for Soft Thresholding Successive Convex Approximation algorithm. This method is resumed in **Algorithm** 1.

Algorithm 1 *the proposed ST-SCA algorithm*

Initialization:

1: pilot-based initialization for $\mathbf{H}$, zero-forcing equalization for the initialization of $\mathbf{S}$ and stop criterion ϵ;

Processing:

2: Compute $\mathbb{B}_H\mathbf{Z}^t$ and $\mathbb{B}_S\mathbf{Z}^t$ according to (19) and (20);
3: Compute the optimal step size according to (25);
4: Update $\mathbf{H}$ and $\mathbf{S}$ according to (21) and (22);
5: While $\mid tr((\mathbb{B}\mathbf{Z}^t - \mathbf{Z}^t)^H\nabla f(\mathbf{Z}^t)) + g(\mathbb{B}\mathbf{Z}^t) - g(\mathbf{Z}^t) \mid \ge \epsilon$ repeat from step 2.

D. Alternative solutions

In what follows, and for benchmarking, the OP given by equation (8) can be optimized by using the widely used ADMM approach. To do so, (8) can be reformulated as:

$$\begin{aligned} &\underset{\mathbf{A},\mathbf{B},\mathbf{S}}{minimize} \quad \tfrac{1}{2}\|\bar{\mathbf{Y}} - \mathbf{A}\bar{\mathbf{S}}\|_F^2 + \tfrac{\lambda}{2}\|\tilde{\mathbf{M}}\text{vec}(\mathbf{S})\|_F^2 + \mu\|\mathbf{B}\|_1, \\ &\text{subject to } \mathbf{A} = \mathbf{B}. \end{aligned} \tag{30}$$

Consequently, the augmented Lagrangian of (30) is given by:

$$\begin{aligned}\mathcal{L}(\mathbf{A},\mathbf{B},\mathbf{S},\mathbf{\Pi}) &= \frac{1}{2}\|\bar{\mathbf{Y}} - \mathbf{A}\bar{\mathbf{S}}\|_F^2 + \frac{\lambda}{2}\|\tilde{\mathbf{M}}\text{vec}(\mathbf{S})\|_F^2 + \mu\|\mathbf{B}\|_1 \\ &+ tr(\mathbf{\Pi}^H(\mathbf{A}-\mathbf{B})) + \frac{c}{2}\|\mathbf{A}-\mathbf{B}\|_F^2, \end{aligned} \tag{31}$$

where the matrix $\mathbf{\Pi}$ is the dual variable (the Lagrange multiplier) which adds the constraint to the cost function and c is a positive constant. Within ADMM framework, the variables are updated as follows:

$$\mathbf{A}^{t+1} = \arg\min_{\mathbf{A}} \mathcal{L}(\mathbf{A},\mathbf{B}^t,\mathbf{S}^t,\mathbf{\Pi}^t), \tag{32}$$
$$\mathbf{B}^{t+1} = \arg\min_{\mathbf{B}} \mathcal{L}(\mathbf{A}^{t+1},\mathbf{B},\mathbf{S}^t,\mathbf{\Pi}^t), \tag{33}$$
$$\mathbf{S}^{t+1} = \arg\min_{\mathbf{S}} \mathcal{L}(\mathbf{A}^{t+1},\mathbf{B}^{t+1},\mathbf{S},\mathbf{\Pi}^t), \tag{34}$$
$$\mathbf{\Pi}^{t+1} = \mathbf{\Pi}^t + c(\mathbf{A}^{t+1} - \mathbf{B}^{t+1}). \tag{35}$$

The solutions to the above OP are as given by:

$$\mathbf{A}^{t+1} = (\bar{\mathbf{Y}}(\bar{\mathbf{S}}^t)^H + c\mathbf{B}^t - \mathbf{\Pi}^t)(\bar{\mathbf{S}}^t(\bar{\mathbf{S}}^t)^H + c\mathbf{I})^{-1}, \tag{36}$$
$$\mathbf{B}^{t+1} = \mathbb{S}_{\frac{\mu}{c}}(\mathbf{A}^{t+1} + \frac{(\mathbf{\Pi}^t)^H}{c}), \tag{37}$$
$$\mathbf{s}^{t+1} = ((\tilde{\mathbf{A}}^{t+1})^H\tilde{\mathbf{A}}^{t+1} + \lambda\tilde{\mathbf{M}}^H\tilde{\mathbf{M}})^{-1}(\tilde{\mathbf{A}}^{t+1})^H\mathbf{y}, \tag{38}$$

where $\mathbf{s}^{t+1} = \text{vec}(\mathbf{S}^{t+1})$ and $\tilde{\mathbf{A}} = \mathbf{I}\otimes\mathbf{A}$.

Also a BCD-based solution is considered in the sequel (for simulation comparison) for solving OP (8) where the channel matrix is updated row-wise, whereas the data symbols are updated by considering the channel's matrix as fixed.

It is important to notice that the numerical cost of the proposed algorithm increases significantly mainly due to the number of the observation samples. To reduce this cost, due mainly to equation (20), one should exploit the block-tridiagonal structure of $(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}} + \lambda\tilde{\mathbf{M}}^H\tilde{\mathbf{M}})$ and its quasi block-Toeplitz property for its fast inversion [11]. Another alternative would be to replace the term associated to the data matrix structure by a term associated to the channel matrix. For example, one can use the Cross-Relation (CR) quadratic criterion [12] for blind SIMO channel estimation according to:

$$\underset{\mathbf{H},\mathbf{S}}{\text{minimize}}\ \frac{1}{2}\|\bar{\mathbf{Y}}-\mathbf{H}\bar{\mathbf{S}}\|_F^2+\frac{\lambda}{2}\mathbf{h}^H\mathbf{Q}_{CR}\mathbf{h}+\mu\|\mathbf{H}\|_1, \tag{39}$$

where $\mathbf{h}=\text{vec}(\mathbf{H})$ and $\mathbf{Q}_{CR}$ is the quadratic form (obtained from the observations) associated to the CR method. In the case of large sample sizes, the minimization of the latter cost function is much cheaper than the minimization of (8).

IV. Performance analysis and discussion

This section highlights the performance of the proposed solution for channel estimation and data detection. A pilot-based initialization is used for channel coefficients, which are, in turn, used for zero-forcing equalization to get the initial data values. Also, for comparison, we have used a fully pilot-based channel estimator (i.e. we assume the data symbols known for benchmarking), and a subspace-based (SS) channel and data estimators [13], [14]. The pilots and data symbols are drawn from a 4-QAM modulation, whereas the channel coefficients are generated randomly using i.i.d. unit-power, zero-mean, Gaussian distribution, with some randomly chosen null coefficients (60% of columns in our case) to model the channel sparsity. The results are averaged over 100 Monte Carlo runs, and the performance is assessed through the normalized mean squared error (NMSE) and an average symbol error rate (SER). Simulation parameters are summarized in TABLE I, unless otherwise mentioned.

Parameters	Specifications
Number of transmitters	N_t = 1
Number of receive antennas	N_r = 20
Number of data symbols	N_d = 100
Number of pilot symbols	N_p = 10
Channel's taps	M = 15
Cost function constants	λ=0.25 $\|\mathbf{Y}\|$ ($\|\mathbf{Y}\|$: spectral norm of $\mathbf{Y}$) μ=1.8 $c=10^4$ (for ADMM-based algorithm)

TABLE I: Simulation parameters.

Fig. 1 investigates the performance of channel estimation, in terms of NMSE w.r.t. SNR, of the different techniques described previously. One can notice that the proposed solution ($\mathbf{H}_{ST-SCA}$) outperforms the ADMM-based ($\mathbf{H}_{ADMM}$), the BCD-based ($\mathbf{H}_{BCD}$) and the subspace-based ($\mathbf{H}_{SS}$) solutions, while becoming close to the fully pilot-based one ($\mathbf{H}_{Pilots}$) for high SNRs.

Fig. 2 assesses the performance of data estimation, in terms of NMSE w.r.t. SNR, of the different techniques described previously. Note that for $\mathbf{S}_{Pilots}$, a zero forcing is applied by using the estimated channel matrix $\mathbf{H}_{Pilots}$. It can be seen that the proposed solution ($\mathbf{S}_{ST-SCA}$) performs better than all other techniques $\mathbf{S}_{ADMM}$, $\mathbf{S}_{BCD}$, $\mathbf{S}_{SS}$ and $\mathbf{S}_{Pilots}$.

By using the estimated data symbols, a hard decision is performed to obtain a 4-QAM symbols. Hence, Fig. 3 illustrates the obtained SER w.r.t. SNR of the different solutions described previously. One can notice that the results obtained in Fig. 2 are confirmed here.

Fig. 4 illustrates the behavior of the cost function, given in (8), w.r.t. the number of iterations needed for convergence. One can notice the gain obtained by using an optimal step size illustrated by ST-SCA (optimal step size), compared to the use of a fixed step size ($\gamma=0.06$) illustrated by ST-SCA (fixed step size). On the other hand, it is shown that a similar behavior is observed for the proposed solution (ST-SCA (optimal step size)) and the BCD-based algorithm, whereas slightly lower number of iterations is observed for the ADMM-based solution. Nevertheless, small number of iterations are needed for convergence for the three techniques, in this context. However, it is worth noting that the choice of the appropriate constant c for the ADMM-based solution remains a very hard task that influences the algorithm's convergence. Whereas for the BCD-based solution, the sequential update may incur large delays, especially for high dimensions (see TABLE II), which becomes not suitable for real-time processing.

TABLE II illustrates the CPU time (in seconds) needed

channel's memory	ST-SCA	BCD	ADMM
M = 15	2.0296	1.0991	1.9881
M = 80	26.1254	30.7365	27.0012

TABLE II: CPU time (in seconds) for one iteration.

for one iteration of the proposed algorithm (ST-SCA), the ADMM-based and the BCD-based solutions. In this comparison we considered a brute-force implementation of the algorithms (i.e. without exploiting the close to Toeplitz, block tridiagonal structure of the involved matrices). Compared to the ADMM-based solution, ST-SCA has similar time consumption, but still performs better in terms of channel and data estimation as illustrated in Figures 1, 2 and 3. However, one can notice that the BCD-based solution needs more time for longer channels (e.g. $M=80$). Note also that the CPU time given in TABLE II has been calculated for a sequential implementation, and can be further reduced, by around the half, for ST-SCA when using parallel processing or multithreading, since the variable update at iteration $t+1$ depends only on the variables of the t-th iteration.

V. Conclusion

This paper proposed a semi-blind solution for joint sparse channel estimation and data recovery, by solving an appropriate optimization problem. The proposed solution is based on the successive convex approximation approach, where the optimization is performed on an approximate convex problem, rather than dealing with the original nonconvex one. An OP is formulated, based on available data (pilots) and convolutive system structure, then an iterative procedure is proposed where the channel coefficients and data symbols are updated (estimated) simultaneously, at each iteration. Along the iterations, an optimized step size procedure is introduced for convergence improvement with guaranteed convergence to a stationary point. Simulation results show that ST-SCA outperforms state-of-the-art techniques by exhibiting moderate-complexity, fast-

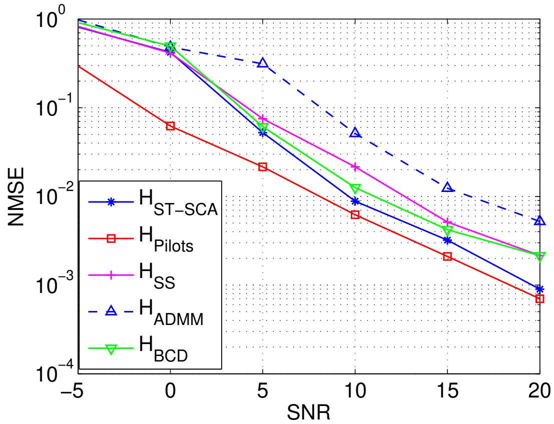

Fig. 1: NMSE of channel matrix $\mathbf{H}$ estimate vs. SNR.

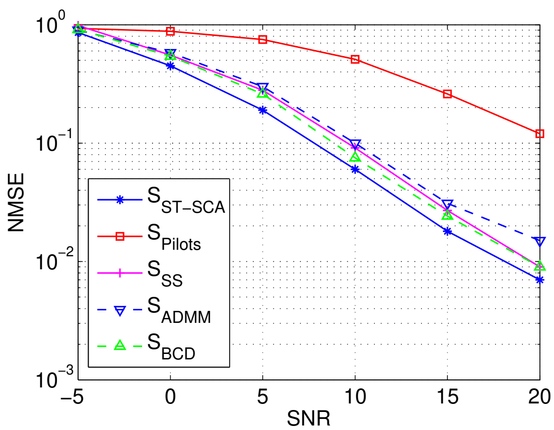

Fig. 2: NMSE of data matrix $\mathbf{S}$ estimate vs. SNR.

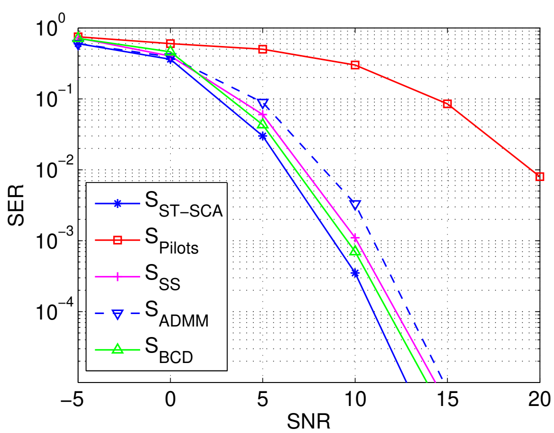

Fig. 3: SER vs. SNR.

convergence and a promizing channel and data estimation accuracy. Moreover, the adopted approach is suitable for parallel processing or multithreading since all variables are updated simultaneously.

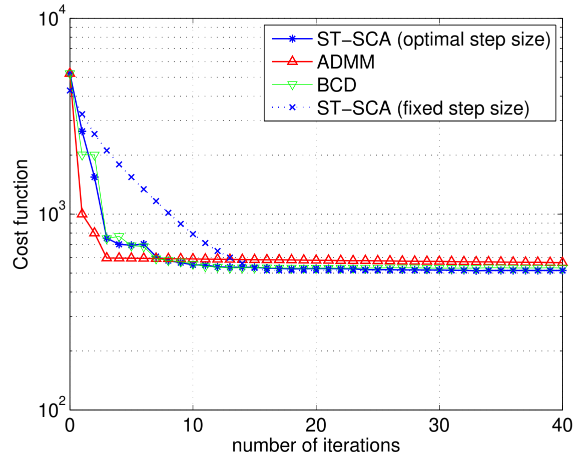

Fig. 4: Cost function vs. number of iterations at SNR $= 10dB$.

REFERENCES

[1] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint channel estimation and data detection in MIMO-OFDM systems: A sparse Bayesian learning approach," *IEEE Transactions on signal processing*, vol. 63, no. 20, pp. 5369–5382, 2015.

[2] A. Mishra, A. K. Jagannatham, and L. Hanzo, "Sparse Bayesian learning-aided joint sparse channel estimation and ML sequence detection in space-time trellis coded MIMO-OFDM systems," *IEEE Transactions on Communications*, 2019.

[3] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint approximately sparse channel estimation and data detection in OFDM systems using sparse bayesian learning," *IEEE Transactions on Signal Processing*, vol. 62, no. 14, pp. 3591–3603, 2014.

[4] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[5] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of optimization theory and applications*, vol. 109, no. 3, pp. 475–494, 2001.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[7] Y. Yang and M. Pesavento, "A unified successive pseudoconvex approximation framework," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3313–3328, 2017.

[8] Y. Yang, M. Pesavento, S. Chatzinotas, and B. Ottersten, "Successive convex approximation algorithms for sparse signal estimation with nonconvex regularizations," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1286–1302, 2018.

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

[10] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, 2016.

[11] M. G. Reuter and J. C. Hill, "An efficient, block-by-block algorithm for inverting a block tridiagonal, nearly block Toeplitz matrix," *Computational Science & Discovery*, vol. 5, no. 1, p. 014009, 2012.

[12] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Transactions on signal processing*, vol. 43, no. 12, pp. 2982–2993, 1995.

[13] 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, 1997.

[14] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications," *IEEE Transactions on signal processing*, vol. 45, no. 11, p. 2714–2723, 1995.