Compressed Sensing of Multi-Channel EEG Signals: The Simultaneous Cosparsity and Low Rank Optimization

Yipeng Liu*, *Member, IEEE*, Maarten De Vos, *Member, IEEE*,
and Sabine Van Huffel, *Fellow, IEEE*

Abstract

Goal: This paper deals with the problems that some EEG signals have no good sparse representation and single channel processing is not computationally efficient in compressed sensing of multi-channel EEG signals. *Methods*: An optimization model with L0 norm and Schatten-0 norm is proposed to enforce cosparsity and low rank structures in the reconstructed multi-channel EEG signals. Both convex relaxation and global consensus optimization with alternating direction method of multipliers are used to compute the optimization model. *Results*: The performance of multi-channel EEG signal reconstruction is improved in term of both accuracy and computational complexity. *Conclusion*: The proposed method is a better candidate than previous sparse signal recovery methods for compressed sensing of EEG signals. *Significance*: The proposed method enables successful compressed sensing of EEG signals even when the signals have no good sparse representation. Using compressed sensing would much reduce the power consumption of wireless EEG system.

Index Terms

alternating direction method of multipliers (ADMM), compressed sensing, cosparse signal recovery, low rank matrix recovery, multi-channel electroencephalogram (EEG).

I. Introduction

Wireless body sensor networks take spatially distributed sensors to acquire physiological signals, and transmit them over wireless links to a central unit for signal processing [1]. The electroencephalogram (EEG) signal is one of the most frequently used biomedical signals. It has important applications in medical healthcare, brain computer interfacing (BCI), and so on [2]. Continuous EEG monitoring usually requires large amount of data to be sampled and transmitted, which leads to large size of batteries. The recording unit of the wireless portable EEG systems is powered with batteries, and the physical size of the batteries sets the overall device size and operational lifetime. A physically too large device would not be portable; and excessive battery power consumption would make the long time wireless recording very hard [3] [4] [5].

Compressed sensing (CS) was proposed to deal with this challenge. Rather than first sample the analog signal at Nyquist rate and discard most in the compression, it directly acquires the digital compressed measurements at a lower sampling rate, and recovers the digital signals by nonlinear algorithms from the compressed measurements [6]. CS relies on the assumption that the signal vector $\mathbf{x}$ is compressed by a random matrix $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ (measurement or sampling matrix) in discrete form as [6] [7]:

$$\mathbf{y} = \mathbf{\Phi x}, \tag{1}$$

This work was supported by FWO of Flemish Government: G.0108.11 (Compressed Sensing); Belgian Federal Science Policy Office: IUAP P7/19/ (DYSCO, 'Dynamical systems, control and optimization', 2012-2017); ERC Advanced Grant: BIOTENSORS (339804). *Asterisk indicates corresponding author*.

*Yipeng Liu is with School of Electronic Engineering / Center for Information in BioMedicine, University of Electronic Science and Technology of China (UESTC), Xiyuan Avenue 2006, Western High-Tech Zone, Chengdu, 611731, China. Most of this work has been done when he was with ESAT-STADIUS Division / iMinds Medical IT Department, Dept. of Electrical Engineering, University of Leuven, Kasteelpark Arenberg 10, box 2446, 3001 Leuven, Belgium. (e-mail: yipengliu@uestc.edu.cn)

Maarten De Vos is with Institute of Biomedical Engineering, Department of Engineering, University of Oxford, Oxford, United Kingdom.

Sabine Van Huffel is with ESAT-STADIUS Division / iMinds Medical IT Department, Dept. of Electrical Engineering, University of Leuven, Kasteelpark Arenberg 10, box 2446, 3001 Leuven, Belgium.

Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

where $\mathbf{y}$ is the random sub-Nyquist compressed measurement. Here $M \ll N$, which means that it is sampled at a greatly reduced rate. If $\mathbf{x}$ is sparse, its recovery only requires the compressed signal $\mathbf{y}$ and the sampling matrix Φ. If it is not sparse, the signal $\mathbf{x}$ should be represented (transformed) using a representation matrix (dictionary) $\mathbf{\Psi} \in \mathbb{R}^{N\times P}$ with $N \leqslant P$ and a sparse vector $\theta \in \mathbb{R}^{P\times 1}$ with most of its entries zero or almost zero as:

$$\mathbf{x} = \mathbf{\Psi}\theta. \tag{2}$$

With the compressed measurement $\mathbf{y}$, sampling matrix $\mathbf{\Phi}$ and dictionary $\mathbf{\Psi}$, we can recover $\mathbf{x}$ by (2) after computing θ by:

$$\begin{aligned} &\underset{\theta}{\text{minimize}} \, \|\theta\|_0 \\ &\text{subject to} \quad \mathbf{y} = \mathbf{\Phi}\mathbf{\Psi}\theta \end{aligned}, \tag{3}$$

where $\|\theta\|_0$ is the pseudo-ℓ_0 norm which counts the number of nonzero entries, i.e. $\|\theta\|_0 = \#\{\theta_n \neq 0, \ n = 1, 2, \cdots, N\}$. The signal $\mathbf{x}$ is called K-sparse when the number of nonzero entries is K. Most of the current methods for biomedical signal recovery from compressed samples are based on the solution of the ℓ_0 programming problem (3), such as, basis pursuit (BP), orthogonal matching pursuit (OMP), iterative hard thresholding (IHT), etc [4] [8] [9]. Besides, [5] found that some EEG signals are not sparse in any sparse transformed domains, and proposed to exploit block-sparsity by block sparse Bayesian learning (BSBL) to recover EEG signals [5].

Contrary to the traditional sparse or block-sparse signal model, the cosparse signal model uses an analysis operator multiplying the measurement to produce a sparse vector [10]:

$$\rho = \mathbf{\Omega}\mathbf{x}, \tag{4}$$

where $\mathbf{\Omega} \in \mathbb{R}^{Q\times N}$ is the cosparse representation matrix (analysis dictionary) with $N \leqslant Q$, and $\rho \in \mathbb{R}^{Q\times 1}$ is the cosparse vector if most of its entries are nearly zero. Several sufficient conditions theoretically guarantee the successful recovery of the cosparse signal from the compressed measurement, such as the restricted isometry property adapted to the dictionary (D-RIP), restricted orthogonal projection property (ROPP), etc [10] [11] [12]. When $N = P$, an equivalent cosparse signal model to the sparse signal model can be found by letting $\mathbf{\Omega}$=$\mathbf{\Psi}^{-1}$; but there is no such an equivalent when $N < P$. The traditional sparse synthesis model puts an emphasis on the non-zeros of the sparse vector θ, but the cosparse analysis model draws its strength from the zeros of the analysis vector ρ.

The cosparse signal recovery has some unique advantages in CS based EEG systems. First, the sparse signal recovery (3) gets the best estimate of the sparse vector θ; but the cosparse signal recovery (5) gets the EEG signal's best estimate directly. Second, theoretically the sparse signal recovery (3) requires the columns of the representation matrix $\mathbf{\Psi}$ to be incoherent, but the cosparse way (5) allows the coherence of the cosparse representation matrix $\mathbf{\Omega}$, which can result in super resolution of the EEG signal estimate [11]. Third, the EEG signal can hardly be sparsely represented [5]. However, data analysis shows that the EEG signals are approximately piecewise linear [13], as shown in Fig. 1, which implies the signal fits the cosparse signal model (4) well with the 2nd order difference matrix as the cosparse analysis dictionary. Therefore, the cosparse signal recovery should be more appropriate for CS of EEG signals.

Since nearly all types of EEG systems have multiple channels, it can be taken for granted that it is better to jointly process the multi-channel EEG signals. [14] proposed to jointly process multi-channel EEG signals by allowing slightly different phases of the dictionaries in different channels. Another classical way assumes that multiple channels share a similar support of sparse vector. This generalizes the single measurement vector (SMV) problem straightforwardly to a multiple measurement vector (MMV) problem [15] [16]. [17] proposed to incorporate preprocessing and entropy coding in the sampling to reduce the redundance in correlated multi-channel signals, but the added preprocessing and encoder would increase the power consumption in EEG sampling [4]; and the procedure can hardly be realized for analog signals, which implies the analog EEG signals should be sampled at Nyquist sampling rate in the beginning. To

compress the multi-channel EEG signals from the complete digital measurement, [18] used a wavelet-based volumetric coding method, while [19] exploited the low rank structure in matrix/tensor form and achieved better performance.

Since most of the multi-channel EEG signals are more or less correlated with each other, the low rank structure based compression method motivates the use of low rank data structure in CS of multi-channel EEG signals too. The multi-channel EEG signals are put columnwise into a matrix. Our EEG data analysis finds that the newly formed EEG data matrix has only a few nonzero singular values.

In this paper, the 2nd-order difference matrix is chosen to be the cosparse analysis dictionary, which tries to enforce the approximate piecewise linear structure. Exploiting additionally the low rank structure, we can further enhance the signal recovery performance by exploiting the cosparsity of single channel EEG signals and the low rank property of multi-channel EEG signals simultaneously in the framework of multi-structure CS. The ℓ_0 norm and Schatten-0 norm based optimization model is used to encourage cosparsity and low rank structure in the reconstructed signals. Two methods are proposed to solve the multi-criterion optimization problem. One relaxes it to a convex optimization; and the other one transforms it into a global consensus optimization problem. The alternating direction method of multipliers (ADMM) is used to solve it efficiently. The convergence and computational complexity are briefly analyzed. In numerical experiments, a group of real-life EEG data is used to test the algorithms' performance of both single-channel and multi-channel EEG signal recovery methods. Numerical results show that the cosparse signal recovery method and simultaneous cosparsity and low-rank (SCLR) optimization achieve the best performance in term of mean squared error (MSE) and mean cross-correlation (MCC) in single channel and multi-channel EEG signal recovery respectively.

The rest of the paper is organized as follows. Section II presents an optimization model to exploit both cosparsity and low rank data structures to recover the EEG signals. In Section III, two methods are given to solve the optimization problem, i.e. convex relaxation and alternating direction method of multipliers (ADMM). In Section IV, numerical experiments are used to demonstrate the proposed methods' performance improvement. Section V draws the conclusion.

II. Simultaneous Cosparsity and Low Rank Optimization Model

The optimization model for cosparse signal recovery can be formulated as [10]:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \; & \|\mathbf{\Omega x}\|_0 \\ \text{subject to} \; & \mathbf{y} = \mathbf{\Phi x} \end{aligned}. \tag{5}$$

Here we call (5) the analysis L0 optimization. When the EEG system records R channels simultaneously, the extension of analysis L0 optimization to multi-channel data is:

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \; & \|\text{vec}\,(\mathbf{\Omega X})\|_0 \\ \text{subject to} \; & \mathbf{Y} = \mathbf{\Phi X} \end{aligned}, \tag{6}$$

where $\mathbf{X} \in \mathbb{R}^{N\times R}$, and vec($\mathbf{X}$) puts all the columns of $\mathbf{X}$ into one column vector sequentially. A series of solvers are summarized in [10].

Reconstructing the EEG matrix from the compressed measurements by exploiting the low rank structure can be formulated as:

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \; & \|\mathbf{X}\|_{\text{Schatten}-0} \\ \text{subject to} \; & \mathbf{Y}=\mathbf{\Phi X} \end{aligned}, \tag{7}$$

where $\|\mathbf{X}\|_{\text{Schatten}-0}$ is the Schatten-0 norm which counts the number of the nonzero singular values of $\mathbf{X}$ [21]. A variety of methods to solve it can be found in [22].

Motivated by the fact that many EEG signals have both cosparsity and low rank structure, we propose to simultaneously exploit these two data structures in multi-channel EEG signal reconstruction from the compressed measurement. Both ℓ_0 norm and Schatten-0 norm based constraints are used in the optimization

model. Combining with the linear data fitting constraint, we can formulate the simultaneous cosparsity and low rank (SCLR) optimization model as follows:

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \; & \|\text{vec}\,(\mathbf{\Omega X})\|_0 + \|\mathbf{X}\|_{\text{Schatten}-0} \\ \text{subject to} \; & \mathbf{Y}=\mathbf{\Phi X} \end{aligned} . \tag{8}$$

III. SOLUTIONS

A. Convex relaxation

To solve the SCLR optimization (8), one classical way relaxes the nonconvex ℓ_0 norm and Schatten-0 norm into convex ℓ_1 norm and Schatten-1 norm respectively, where the ℓ_1 norm sums all the absolute values of the entries, i.e. $\|\mathbf{x}\|_1 = \sum_{n=1}^{N} |x_n|$. The Schatten-1 norm is called nuclear norm too, and sums all the singular values of the data matrix, i.e. $\|\mathbf{X}\|_{\text{Schatten}-1} = \|\mathbf{X}\|_* = \sum_{n=1}^{\min(N,P)} \sigma_n$. The newly formed convex simultaneous cosparsity and low rank (CSCLR) optimization model can be formulated as:

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \; & \|\text{vec}\,(\mathbf{\Omega X})\|_1 + \|\mathbf{X}\|_* \\ \text{subject to} \; & \mathbf{Y}=\mathbf{\Phi X} \end{aligned} . \tag{9}$$

Similarly to the reformulation from $\text{minimize}_{\mathbf{x}} \|\mathbf{x}\|_1$ to $\text{minimize}_{\mathbf{x},\mathbf{e}>0} \; \mathbf{1}^T\mathbf{e}$, subject to $-\mathbf{e} \prec \mathbf{x} \prec \mathbf{e}$ due to the definition of the ℓ_1 norm, we can re-formulate the ℓ_1 norm minimization into its equivalent linear programming in (9) [23]. By introduction of new nonnegative variables $\mathbf{e}$ and f, (9) can be expressed as:

$$\begin{aligned} \underset{\mathbf{X},\mathbf{e}>0,\; f\geq 0}{\text{minimize}} \; & \mathbf{1}^T\mathbf{e} + f \\ \text{subject to} \; & \mathbf{Y}=\mathbf{\Phi X} \\ & \|\mathbf{X}\|_* \leq f \\ & -\mathbf{e} \prec \text{vec}\,(\mathbf{\Omega X}) \prec \mathbf{e} \end{aligned} , \tag{10}$$

where $\mathbf{1} \in \mathbb{R}^{QR\times 1}$ is a column vector with all the entries being 1.

The nuclear norm constraint can be replaced by its linear matrix inequality (LMI) equivalent; and the approximation constraints can also be expressed via LMIs using Schur complements [24]. The obtained optimization model is:

$$\begin{aligned} \underset{\mathbf{X},\mathbf{e}>0, f\geq 0}{\text{minimize}} \; & \mathbf{1}^T\mathbf{e} + 2f \\ \text{subject to} \; & \mathbf{Y}=\mathbf{\Phi X} \\ & -\mathbf{e} \prec \text{vec}\,(\mathbf{\Omega X}) \prec \mathbf{e} \\ & \begin{bmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{B} \end{bmatrix} \geq 0 \\ & \text{Tr}\,(\mathbf{A}) + \text{Tr}\,(\mathbf{B}) < f \end{aligned} , \tag{11}$$

where $\mathbf{A} = \mathbf{A}^T$ and $\mathbf{B} = \mathbf{B}^T$ are new variables. (11) is a semi-definite programming (SDP) which can be solved by interior-point method [23] [24]. The software CVX can compute the solution in this way [25].

B. ADMM

Besides the classical SDP, another method, called ADMM, can be used to solve the SCLR optimization [26]. With individual constraints on the same variables in each constraint, (9) can be rewritten into a global consensus optimization with local variables $\mathbf{X}_i$, i = 1,2 and a common global variable $\mathbf{X}$ as:

$$\begin{aligned} \underset{\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}}{\text{minimize}} \; & \|\text{vec}\,(\mathbf{\Omega X}_1)\|_1 + \|\mathbf{X}_2\|_* \\ \text{subject to} \; & \mathbf{X}=\mathbf{X}_1; \mathbf{X}=\mathbf{X}_2; \mathbf{Y} = \mathbf{\Phi X} \end{aligned} . \tag{12}$$

Here the new constraints are that all the local variables should be equal. It is equivalent to:

$$\begin{aligned} \underset{\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}}{\text{minimize}} \; & \|\text{vec}\,(\mathbf{\Omega X}_1)\|_1 + \|\mathbf{X}_2\|_* \\ \text{subject to} \; & \bar{\mathbf{Y}}=\bar{\mathbf{\Phi}}\mathbf{X}_1; \bar{\mathbf{Y}}=\bar{\mathbf{\Phi}}\mathbf{X}_2 \end{aligned} , \tag{13}$$

where

$$\bar{\mathbf{Y}} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{X} \end{bmatrix}, \tag{14}$$

$$\bar{\mathbf{\Phi}} = \begin{bmatrix} \mathbf{\Phi} \\ \mathbf{I} \end{bmatrix}. \tag{15}$$

The corresponding augmented Lagrangian of (13) is:

$$\begin{aligned} &L_\rho\left(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Z}_1, \mathbf{Z}_2\right) = \|\mathrm{vec}\left(\mathbf{\Omega}\mathbf{X_1}\right)\|_1 + \|\mathbf{X}_2\|_* \\ &+\mathrm{vec}(\mathbf{Z}_1)^T \mathrm{vec}\left(\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_1\right) + \mathrm{vec}(\mathbf{Z}_2)^T \mathrm{vec}\left(\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_2\right), \\ &+\tfrac{\rho}{2}\left\|\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_1\right\|_F^2 + \tfrac{\rho}{2}\left\|\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_2\right\|_F^2 \end{aligned} \tag{16}$$

where $\rho > 0$, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are dual variables. The resulting ADMM algorithm in the scaled dual form is the following

$$\mathbf{X}_1^{t+1} := \arg\min_{\mathbf{X}_1}\left(\|\mathrm{vec}\left(\mathbf{\Omega}\mathbf{X}_1\right)\|_1 + \frac{\rho}{2}\left\|\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_1 + \bar{\mathbf{\Phi}}\mathbf{U}_1^t\right\|_F^2\right), \tag{17}$$

$$\mathbf{X}_2^{t+1} := \arg\min_{\mathbf{X}_2}\left(\|\mathbf{X}_2\|_* + \frac{\rho}{2}\left\|\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_2 + \bar{\mathbf{\Phi}}\mathbf{U}_2^t\right\|_F^2\right), \tag{18}$$

$$\mathbf{X}^{t+1} = \frac{1}{2}\left(\mathbf{X}_1^{t+1} + \mathbf{X}_2^{t+1}\right), \tag{19}$$

$$\begin{aligned} \mathbf{U}_1^{t+1} &= \mathbf{U}_1^t + \left(\mathbf{X}_1^{t+1} - \mathbf{X}_1^t\right) \\ \mathbf{U}_2^{t+1} &= \mathbf{U}_2^t + \left(\mathbf{X}_2^{t+1} - \mathbf{X}_2^t\right) \end{aligned}, \tag{20}$$

where $\mathbf{U}_1 = 1/\rho\mathbf{Z}_1$ and $\mathbf{U}_2 = 1/\rho\mathbf{Z}_2$ are scaled dual variables. In the proposed ADMM algorithm for SCLR optimization, two steps separately optimize over variables generally, i.e. updating the prime variables $\mathbf{X}_1$ and $\mathbf{X}_2$, updating the scaled dual variables $\mathbf{U}_1$ and $\mathbf{U}_2$. In this iterative algorithm, the variables are updated in an alternating fashion.

For both (17) and (18), there are many computationally efficient algorithms [10] [22]. For example, analysis L1 optimization, greedy analysis pursuit (GAP) can be used to solve (17); to solve (18), SDP method or singular value thresholding (SVT) can be used. The solutions of (19) and (20) are straightforwardly easy. The ADMM for SCLR optimization is summarized in Algorithm 1.

Algorithm 1: ADMM for the SCLR optimization

- Set $t := 0$, a small scalar $\eta > 0$, $\mathbf{U}_1^0$, $\mathbf{U}_2^0$, $\mathbf{X}^0$, $T_{\max}$, ρ;

repeat

- step 1: update of the analysis L1 optimization: $\mathbf{X}_1^{t+1} := \arg\min_{\mathbf{X}_1}\left(\|\mathrm{vec}\left(\mathbf{\Omega}\mathbf{X}_1\right)\|_1 + \frac{\rho}{2}\left\|\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_1 + \bar{\mathbf{\Phi}}\mathbf{U}_1^t\right\|_F^2\right)$;
- step 2: update of the low rank optimization: $\mathbf{X}_2^{t+1} := \arg\min_{\mathbf{X}_2}\left(\|\mathbf{X}_2\|_* + \frac{\rho}{2}\left\|\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_2 + \bar{\mathbf{\Phi}}\mathbf{U}_2^t\right\|_F^2\right)$;
- step 3: update of the global variable: $\mathbf{X}^{t+1} = \frac{1}{2}\left(\mathbf{X}_1^{t+1} + \mathbf{X}_2^{t+1}\right)$;
- step 4: update of the dual variables: $\begin{aligned} \mathbf{U}_1^{t+1} &= \mathbf{U}_1^t + \left(\mathbf{X}_1^{t+1} - \mathbf{X}_1^t\right) \\ \mathbf{U}_2^{t+1} &= \mathbf{U}_2^t + \left(\mathbf{X}_2^{t+1} - \mathbf{X}_2^t\right) \end{aligned}$;
- step 5: set $t := t + 1$;

until $\frac{\|\mathbf{X}^{t+1} - \mathbf{X}^t\|_F}{\|\mathbf{X}^{t+1}\|_F \|\mathbf{X}^t\|_F} \le \eta$ *or* $t = T_{\max}$;

- Algorithm ends and return $\mathbf{X}^{t+1}$.

A lot of convergence results exist for ADMM in the literature [26]. Generally, the convergence to optimum can be guaranteed when the epigraph of g_i:

$$\text{epi}g_i = \{(\mathbf{X}, \varepsilon) \mid g_i(\mathbf{X}) \leq \varepsilon,\ i = 1, 2.\} \tag{21}$$

is a closed nonempty convex set, where $g_1(\mathbf{X}) = \|\text{vec}(\mathbf{\Omega X})\|_1$, $g_2(\mathbf{X}) = \|\mathbf{X}\|_*$, and the unaugmented Lagrangian

$$\begin{aligned} &L_{\rho=0}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Z}_1, \mathbf{Z}_2) = \|\text{vec}(\mathbf{\Omega X_1})\|_1 + \|\mathbf{X}_2\|_* \\ &+\text{vec}(\mathbf{Z}_1)^T \text{vec}\left(\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_1\right) + \text{vec}(\mathbf{Z}_2)^T \text{vec}\left(\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_2\right) \end{aligned} \tag{22}$$

has a saddle point. The proof can be found in [27].

The ADMM decomposes the optimization model with multiple constraints into several ones with fewer constraints. There could be some fast algorithms for these new optimization models. Besides, it allows multiple steps in one iteration to be processed in parallel. With a multi-core processor, the computational time can be decreased. Previous experience shows that a few iterations will often produce acceptable results of practical use.

IV. Numerical Experiments

To demonstrate the performance of the possible methods for EEG signal recovery from the compressed measurement, we perform two groups of numerical experiments. The details about the data materials and subjects are given in section IV-A. In section IV-B, we test the performance of two cosparse signal recovery methods for single-channel EEG signals in different kinds of situations, i.e. analysis L1 optimization and GAP. Some other algorithms are tested to make comparison, such as BSBL which is reported to be the best of all the current candidates for EEG signal recovery from compressed measurement [5], and OMP which is a proper representative of the classical sparse signal recovery algorithms [4]. In section IV-C, a group of multi-channel EEG signals are recovered by the proposed algorithms for SCLR optimization, as well as simultaneous orthogonal matching pursuit (SOMP) [28], BSBL [5] [16], and simultaneous greedy analysis pursuit (SGAP) [29].

In all experiments, as argued by our analysis in section I, the 2nd-order difference matrix is chosen to be the analysis dictionary for cosparse EEG signal recovery. The Gaussian matrix is chosen to be the sampling matrix for CS of EEG signals. The sparse dictionaries of OMP and SOMP are Daubechies wavelets [4].

To measure the compression degree, the subsampling ratio (SSR) is defined as:

$$\text{SSR} = \frac{M}{N} \times 100\%. \tag{23}$$

To quantify the difference between high-dimensional values implied by the estimator and the true values of the quantity being estimated, two different evaluation functions are often used in EEG signal processing. One is the mean squared error (MSE) which measures the average of the squares of the errors. The error is the amount by which the value implied by the estimator differs from the quantity to be estimated. Here we can formulate it as:

$$\text{MSE} = \sum_{l=1}^{L} \frac{\left\|\hat{\mathbf{X}}_l - \mathbf{X}\right\|_F^2}{LNR}, \tag{24}$$

where $\mathbf{X}$ is the true EEG data with R channels and each channel has length N, $\hat{\mathbf{X}}_l$ is its estimate in the l-th experiment, and L is the number of experiments. Both $\mathbf{X}$ and $\hat{\mathbf{X}}_l$ are normalized by their Frobenius norms respectively. When $R = 1$, the matrix $\mathbf{X}$ is degenerated into a vector $\mathbf{x}$. In that case, MSE can be used to evaluate single-channel EEG signal reconstruction evaluation. The MSE has variants of other equivalent forms, such as mean L2 error [30], percent of root-mean-square difference (PRD) [4].

Another evaluation function is the mean cross-correlation (MCC). It is equivalent to the Structural SIMilarity index (SSIM), which measures the similarity of two waveforms [31] [4] [5]. It can be formulated as:

$$\text{MCC} = \sum_{l=1}^{L} \frac{\text{vec}(\mathbf{X})^T \text{vec}\left(\hat{\mathbf{X}}_l\right)}{L\|\mathbf{X}\|_F \left\|\hat{\mathbf{X}}_l\right\|_F}. \tag{25}$$

A. Data material and subjects

The used EEG data is the CHB-MIT scalp EEG database which is online available in the *Physiobank* database: http://www.physionet.org/cgi-bin/atm/ATM [32] [33]. Collected at the Children's Hospital Boston, these EEG recordings are from pediatric subjects with intractable seizures. Subjects were monitored without anti-seizure medication in order to characterize their seizures and assess their candidacy for surgical intervention. All the recordings were collected from 22 subjects (5 males, ages 3-22; and 17 females, ages 1.5-19). All used datasets consist of 23-channel EEG recordings which were sampled at 256 samples per second with 16-bit resolution. The international 10-20 system of EEG electrode positions and nomenclature was used for these recordings. More details about the EEG database can be found [32]. In our experiments, the EEG recording $chb01_31.edf$ has been selected to demonstrate the recovery algorithms' performance.

In section IV-B, $L = 500$ segments of EEG data are used, i.e. $\mathbf{x}_l \in \mathbb{R}^{N\times 1},\ l = 1, 2, \cdots, L$. They are taken from all the R=23 channels sequentially. The length of each segment of the EEG data $\mathbf{x}$ is $N = 256$. Each segment of EEG data is normalized by its ℓ_2 norm.

In section IV-C, $L = 50$ segments of 23-channel EEG data are used, i.e. $\mathbf{X}_l \in \mathbb{R}^{N\times R},\ l = 1, 2, \cdots, L$. In each segment of the EEG data matrix $\mathbf{X}$, the number of sampling points is $N \times R = 256 \times 23$. Each segment of EEG data is normalized by its Frobenius norm.

B. Single channel EEG signal recovery

To show how the proposed cosparse signal recovery methods work, we take a segment of single channel EEG signal and reconstruct it from the compressed measurement with SSR = 0.35. The reconstructed and real signals are shown in Fig. 1. We can see that the reconstructed signals from GAP and the analysis L1 optimization methods fit the real signal better than those from the classical OMP and BSBL methods.

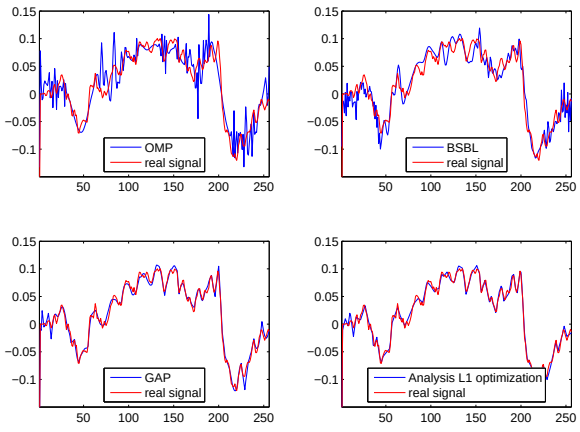


Fig. 1: EEG signals reconstructed by OMP, BSBL, GAP, and analysis L1 optimization with SSR = 0.35.

Fig. 2a, and Fig. 2b give the values of MSE and MCC of GAP, OMP, BSBL with different SSRs. We can see that analysis L1 optimization, GAP and BSBL have similar accuracy performance, and they outperform OMP. Analysis L1 optimization is slightly more accurate than GAP, and GAP is slightly more

accurate than BSBL. Fig. 2c shows that the greedy algorithms GAP and OMP are much faster than BSBL and analysis L1 optimization, and GAP is even slightly faster than OMP. Therefore, if we only care about the accuracy, the analysis L1 optimization is the best choice; and if both accuracy and computational complexity are important, GAP should be a better choice.

C. Multi-channel EEG signal recovery

In these experiments, most of the parameters are selected as in section IV-B. Two algorithms for SCLR optimization are used, i.e. interior point method for SCLR optimization and ADMM for SCLR optimization with experienced choices of the parameters $T_{max} = 5$, $\rho = 1$, $\eta = 0.05$. In comparison with the proposed methods, 3 other popular multi-channel sparse / cosparse signal recovery methods are taken too. i.e. BSBL, SOMP and SGAP.

Fig. 3a, 3b and 3c display the values of MSE, MCC and CPU times of the interior point method for the SCLR optimization, ADMM for SCLR optimization, BSBL, SOMP, and SGAP with different values of SSR. We can see that the interior point method for SCLR optimization, ADMM for SCLR optimization have similar accuracy performance, and they outperform the other ones in accuracy. Comparing the speed of these two solutions for SCLR optimization, the ADMM for SCLR optimization is faster. In Fig. 3c we can see that the greedy algorithms SOMP and SGAP are much faster than the rest. But their accuracy is much worse and not acceptable. Therefore, we recommend that the ADMM for SCLR optimization should be a better candidate for multi-channel EEG signal recovery than the other methods.

V. Conclusion

With the 2nd-order difference matrix as the cosparse analysis dictionary, the EEG signals' cosparsity is exploited for the single-channel EEG signal recovery from compressed measurements. To further enhance the performance, cosparsity and low rank structure are jointly used in the multi-channel EEG signal recovery. In the proposed new optimization model, the ℓ_0 norm constraint is used to encourage cosparsity while Schatten-0 norm constraint is used for low rank structure. To solve the optimization model, two methods are used. One approximates it by relaxing the ℓ_0 and Schatten-0 norms into ℓ_1 norm and nuclear norm respectively, which leads to a convex optimization. The other way is ADMM which divides the multiple criterion optimization into several connected single criterion optimizations in the form of global consensus optimization. Each single criterion optimization can be solved by a series of existing efficient methods. In numerical experiments, EEG signals' cosparsity for CS is proved by the single-channel EEG data based results; and the multi-channel EEG data results show that the SCLR optimization outperforms all the previous methods.

References

[1] C. Bachmann et al, "Low-power wireless sensor nodes for ubiquitous long-term biomedical signal monitoring," *Communications Magazine, IEEE*, vol. 50, no. 1, pp. 20-27, 2012.

[2] M. De Vos et al, "Towards a truly mobile auditory brainCcomputer interface: Exploring the P300 to take away," *International journal of psychophysiology*, vol. 91, no. 1, pp.46-53, 2014.

[3] S. Debener et al, "How about taking a low-cost, small, and wireless EEG for a walk?," *Psychophysiology*, vol. 49, no. 11, pp.1617-1621, 2012.

[4] A. M. Abdulghani et al, "Compressive sensing scalp EEG signals: implementations and practical performance," *Medical and biological engineering and computing*, vol. 50, no. 11, pp. 1137-1145, 2012.

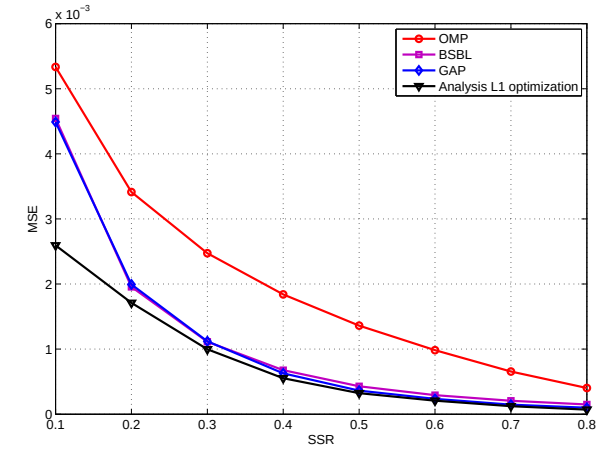
[5] Z. Zhang et al, "Compressed sensing of EEG for wireless telemonitoring with low energy consumption and inexpensive hardware," *Biomedical Engineering, IEEE Transactions on*, vol. 60, no. 1, pp. 221-224, 2013.

[6] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge University Press, 2012.

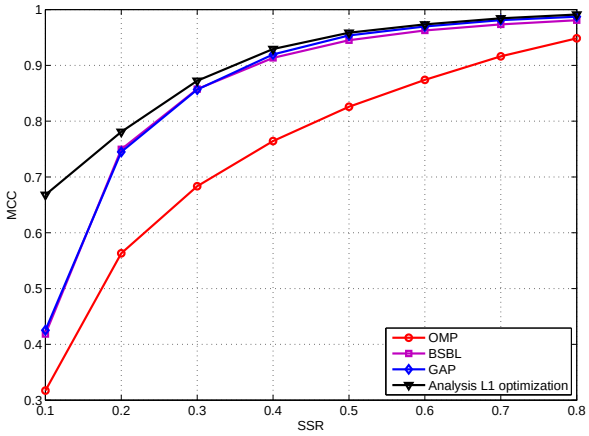
[7] S. R. Becker, "Practical compressed sensing: modern data acquisition and signal processing," Ph.D. dissertation, California Institute of Technology, CA, 2011.

[8] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948-958, 2010.

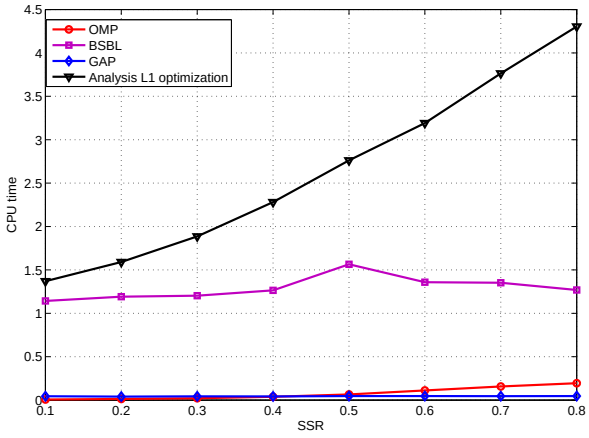
[9] H. Mamaghanian et al, "Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 9, pp. 2456-2466, 2011.



(a) MSE



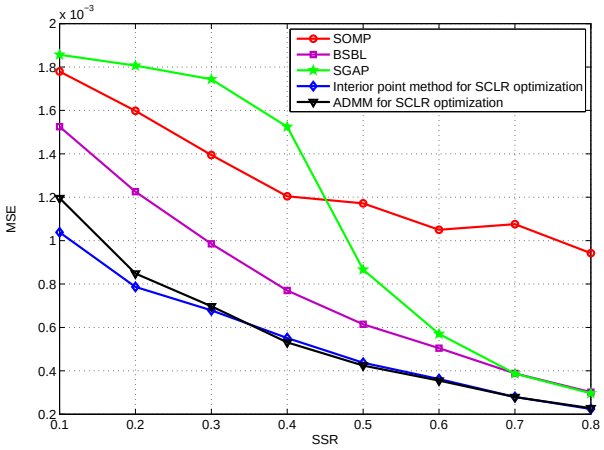
(b) MCC



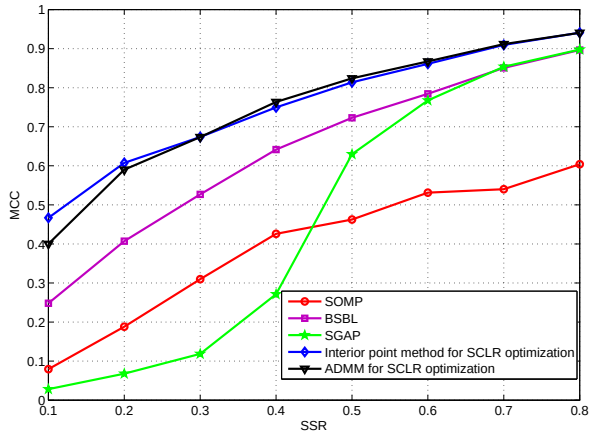
(c) CPU time

Fig. 2: Differences in average performance evaluation of single channel EEG signal recovery from compressed measurements with different SSRs using 500 different single channel EEG segments: (a) MSE vs SSR; (b) MCC vs SSR; (c) CPU Time vs SSR.

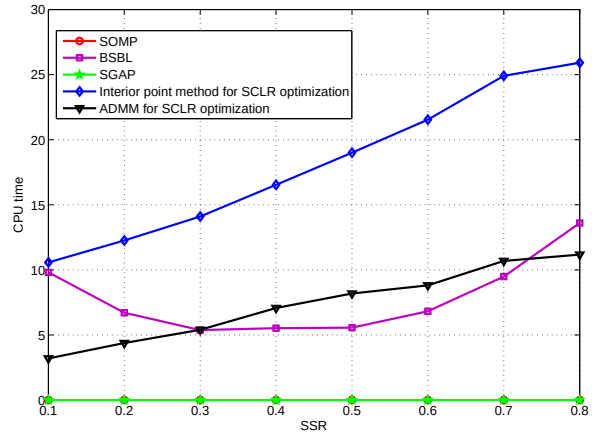
[10] S. Nam et al, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30-56, 2013.



(a) MSE



(b) MCC



(c) CPU time

Fig. 3: Differences in average performance evaluation of multi-channel EEG signal recovery from compressed measurement with different SSRs using 50 different multi-channel EEG segments: (a) MSE vs SSR; (b) MCC vs SSR; (c) CPU Time vs SSR.

[11] E. J. Candes et al, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59-73, 2011.

[12] T. Peleg and M. Elad, "Performance guarantees of the thresholding algorithm for the cosparse analysis model," *Information Theory, IEEE Transactions on*, vol. 59, no. 3, pp. 1832-1845, 2013.
[13] C. Yan et al, " An approach of time series piecewise linear representation based on local maximum, minimum and extremum," *Journal of Information and Computational Science*, vol. 10, no. 9, pp. 2747-2756, 2013.
[14] P. J. Durka et al, "Multichannel matching pursuit and EEG inverse solutions," *Journal of Neuroscience Methods*, vol. 148, no. 1, pp. 49-59, 2005.
[15] S. F. Cotter et al, "Sparse solutions to linear inverse problems with multiple measurement vectors," *Signal Processing, IEEE Transactions on*, vol. 53, no. 7, pp. 2477-2488, 2005.
[16] Z. Zhang et al, "Spatiotemporal sparse Bayesian learning with applications to compressed sensing of multichannel physiological signals," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 22, no. 6, pp. 1186-1197, 2014.
[17] S. Fauvel and R. K. Ward, "An energy efficient compressed sensing framework for the compression of electroencephalogram signals," *Sensors*, vol. 14, no. 1, pp. 1474-1496, 2014.
[18] K. Srinivasan et al, "Multichannel EEG compression: Wavelet-based image and volumetric coding approach," *Biomedical and Health Informatics, IEEE Journal of*, vol. 17, no. 1, pp. 113-120, 2013.
[19] J. Dauwels et al, "Near-lossless multichannel EEG compression based on matrix and tensor decompositions," *Biomedical and Health Informatics, IEEE Journal of*, vol. 17, no. 3, pp. 708-714, 2013.
[20] E. Candès and B. Recht, "Simple bounds for recovering low-complexity models," *Mathematical Programming*, vol. 141, no. 1-2, pp. 577-589, 2013.
[21] A. Rohde and A. Tsybakov, "Estimation of high-dimensional low-rank matrices," *The Annals of Statistics*, vol. 39, no. 2, pp. 887-930, 2011.
[22] B. Recht et al, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471-501, 2010.
[23] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2009.
[24] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM review*, vol. 38, no. 1, pp. 49-95, 1996.
[25] M. Grant et al, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," *Recent Advances in Learning and Control*, pp. 95-110, 2012.
[26] S. Boyd 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.
[27] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1-3, pp. 293-318, 1992.
[28] J. A. Tropp et al, "Algorithms for simultaneous sparse approximation. part I: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572-588, 2006.
[29] Y. Avonds et al, "Simultaneous greedy analysis pursuit for compressive sensing of multi-channel ecg signals," in *Engineering in Medicine and Biology Society (EMBC), 2014 Annual International Conference of the IEEE*. IEEE, 2014, pp. 6385-6388.
[30] Y. Liu et al, "Multi-structural signal recovery for biomedical compressive sensing," *Biomedical Engineering, IEEE Transactions on*, vol. 60, no. 10, pp. 2794-2805, 2013.
[31] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? a new look at signal fidelity measures," *Signal Processing Magazine, IEEE*, vol. 26, no. 1, pp. 98-117, 2009.
[32] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.
[33] A. L. Goldberger et al, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215-e220, 2000.