Compressive 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*—Compressive sensing was proposed to process electroencephalography (EEG) signals. However, most of the current methods follow the traditional sparse signal recovery which decomposes the signal into a dictionary multiplying a sparse or block-sparse vector, and each channel of EEG signals is processed individually. In this paper, we propose to exploit the EEG signals' cosparsity to recover the single-channel EEG signal. The cosparsity results from an analysis matrix multiplying the measurement vector. The 2nd-order derivative matrix is chosen to be the analysis operator for EEG signals' cosparsity. Since multi-channel EEG signals have a low rank structure, we further enhance the signal recovery performance by exploiting the cosparsity and low rank property of multi-channel EEG signals simultaneously. An optimization model with L0 norm and Schatten-0 norm is proposed to enforce cosparsity and low rank structures in the reconstructed signals. Both convex relaxation and global consensus optimization with alternating direction method of multipliers are used to compute the optimization model. Numerical experiments based on real-life EEG signals show the performance improvement in term of both accuracy and computational complexity. It proves that the proposed method should be a better candidate for compressive sensing of EEG signals.**

***Index Terms*—wireless body-area network (WBAN), multi-channel electroencephalogram (EEG), compressive sensing, cosparse signal recovery, low rank matrix recovery, alternating direction method of multipliers (ADMM).**

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. Continuous EEG monitoring usually requires large amount of data to be sampled and transmitted, which leads to large physical size for sampling and storage and high power consumption [2] [3]. Physically large scale device would not be portable; and excessive power consumption would make the long time recording very hard.

Compressive sensing (CS) was proposed to deal with this challenge. Rather than first samples the analog signal at Nyquist rate and discard most in the compression, it directly acquires the digital compressive measurements at a lower sampling rate [4], and recovers the digital signals by nonlinear algorithms from the compressive measurements. With the incorporation of lower power CS analog-to-digital converters (ADC) and of less wireless links for data transmission, the power consumption in wireless EEG sensors network would be reduced.

In practical CS ADC, the random sub-Nyquist measurements vector $\mathbf{x}$ is acquired directly from the analog signal $x(t)$ [5]. For demonstration convenience, we can formulate the random sub-sampling in discrete form as [6]:

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

where $\mathbf{x} \in \mathbb{R}^{N\times 1}$ is the Nyquist measurements vector which can be regarded as the original signal sampled at Nyquist rate, and $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ is the sampling matrix. Here $M \ll N$, which is the reason that $\mathbf{y}$ is the compressive measurements vector of $\mathbf{x}$. When the sampling matrix has some properties, such as null space property, restricted isometry property, incoherence conditions, classical CS can successfully recover the signal $\mathbf{x}$ from its compressive measurements $\mathbf{y}$ [4]. A lot of random sub-samplers has been developed, such as the Random Demodulator (RD), the Random-Modulation Pre-Integrator (RMPI), the Non-Uniform Sampler (NUS), Random Convolution, the Modulated Wideband Converter (MWC). A comprehensive overview can be found in [5]. Most of them can be categorized by different kinds of sampling random matrices, such as Bernoulli matrix, Circulant Matrix, sparse binary matrix, partial identity matrix, etc.

Recovering the EEG signal from compressive measurements is an underdetermined linear inverse problem. Classical methods are based on signal's sparse representation which can be formulated as

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

where $\mathbf{\Psi} \in \mathbb{R}^{N\times P}$ is the representation matrix (dictionary) with

This work was supported by Research Council KUL: GOA/10/09 MaNet, CoE PFV/10/002 (OPTEC); PhD/postdoc grants: Flemish Government: F-WO: projects: G.0427.10N (Integrated EEG-fMRI), G.0108.11 (Compressed Sensing) G.0869.12N (Tumor imaging) G.0A5513N (Deep brain stimulation); PhD/Postdoc grants; IWT: projects: TBM 080658-MRI (EEG-fMRI), TBM 110697-NeoGuard; iMinds Medical Information Technologies SBO 2014, ICON: NXT Sleep; Flanders Care: Demonstratieproject Tele-Rehab III (2012-2014); Belgian Federal Science Policy Office: IUAP P7/19/ (DYSCO, 'Dynamical systems, control and optimization', 2012-2017); Belgiuan Foreign Affairs-Development Cooperation: VLIR UOS programs; EU: RECAP 209G within INTERREG IVB NWE programme, EU MC ITN TRANSACT 2012 (316679), ERC Advanced Grant: BIOTENSORS (339804), ERASMUS EQR: Community service engineer (539642-LLP-1-2013).

Yipeng Liu and Sabine Van Huffel are with ESAT-STADIUS Division / iMinds Medical IT Department, Dept. of Electrical Engineering, University of Leuven, Kasteelpark Arenberg 10, box 2446, 3001 Leuven, Belgium. (emails: yipeng.liu@esat.kuleuven.be; sabine.vanhuffel@esat.kuleuven.be)

Maarten De Vos is with Department of Engineering, Oxford University, United Kingdom. (email: maarten.devos@eng.ox.ac.uk)

$N \leqslant P$, and $\theta \in \mathbb{R}^{P\times 1}$ is a sparse vector with most of its entries are zero or almost zero. To sparsely represent EEG signals, the possible choice of dictionaries can be wavelets, B-spline, Gabor dictionary, etc [2] [3].

With the random sampling model (1) and sparse representation model (2), we can exploit the sparsity to recover the signal $\mathbf{x}$ from the compressive measurements $\mathbf{y}$ by finding the sparsest solution of the underdetermined linear system. The sparsest solution means the solution has fewest nonzero entries. The optimization model for sparse signal recovery can be formulated as:

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

where $\|\theta\|_0$ is the pseudo-ℓ_0 norm which counts the number of nonzero entries, i.e. $\|\theta\|_0 = \#\{\theta_n \neq 0, n = 1, 2, \cdots, N\}$. Most of the current methods for biomedical signal recovery from compressive 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 [7] [8] [9] [10]. Besides, [3] found that EEG is not sparse enough in the transformed domains, and proposed to exploit block-sparsity by block sparse Bayesian learning (BSBL) to recover EEG signals [3] [11].

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. [12] proposed to jointly process multi-channel EEG signals by allowing slightly different phases of the dictionaries in different channels. A classical way it to assume that multiple channels share a similar support of the sparse vector. This generalizes the single measurement vector (SMV) problem straightforwardly to a multiple measurement vector (MMV) problem [11] [13]. [14] 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 compressive EEG sensor [2]; 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 measurements, [15] used a wavelet-based volumetric coding method, while [16] [17] and [18] exploited the low rank structure in matrix/tensor form and achieved better performance. Since most of the multi-channel 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.

Differently from the classical sparse signal model that decomposes the signal into a linear combination of a few atoms chosen from a predefined dictionary, a new signal model for CS, called cosparse signal model, was proposed [19] [20]. In this new representation, an analysis operator multiplying the measurements leads to a sparse vector. In [21], instead of sparsity, its counterpart, called cosparsity, was exploited to recover ECG signals. It was further generalized to multi-channel ECG signals by assuming that multi-channel signals have similar supports [22].

In this paper, we analyze the cosparsity of EEG signals. The 2nd-order derivative matrix is chosen to be the cosparsity 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 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 relaxation; and the other ones 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. 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 major contributions of this paper can be summarized as follows:

First, we propose to exploit cosparsity for single-channel EEG signal recovery from compressive measurements. It is the first time that cosparse signal recovery is introduced to EEG signal processing. Numerical results demonstrate that cosparse signal recovery methods are more accurate.

Second, we propose to simultaneously exploit the low rank structure and cosparsity structure to reconstruct multi-channel EEG signals. In the newly formed optimization model, the ℓ_0 norm constraint and the Schatten-0 norm constraint encourage the cosparsity and low rank structures respectively. Both convex relaxation and alternating direction method of multipliers (ADMM) are taken to solve the optimization problem. The theoretical convergence and computational complexity of ADMM are briefly analyzed.

The rest of the paper is organized as follows. In section II, EEG signals' cosparse and low-rank structures are analyzed. Section III presents a optimization model to exploit both cosparsity and low rank data structure to recovery the EEG signals. In Section IV, two methods are given to solve the optimization problem. One is convex relaxation with higher accuracy, and the other one is ADMM with lower computational complexity. In Section V, numerical experiments are used to demonstrate the proposed methods' performance improvement. Section VI draws the conclusion.

II. Cosparse and Low-rank EEG Signal Models

A. Cosparse single-channel EEG signals

The cosparse analysis signal model exploits the sparsity differently compared to the classical sparse synthesis signal model. It can be formulated as:

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

where $\mathbf{\Omega} \in \mathbb{R}^{Q\times N}$ is the cosparse representation matrix with $N \leqslant Q$, and $\rho \in \mathbb{R}^{Q\times 1}$ is the cosparse vector if most of its entries are nearly zero.

The sparse synthesis signal model, which has been intensively studied for many years, expects the signal $\mathbf{x}$ to be a linear combination of a few of columns from a pre-defined matrix $\mathbf{\Psi}$, while the recently proposed cosparse analysis signal model encourages the signal $\mathbf{x}$ to be orthogonal to most rows of the analysis dictionary $\mathbf{\Omega}$, and therefore leads to cosparsity in the vector $\mathbf{\Omega x}$.

Current research shows that the EEG signal does not have a satisfactory sparse representation for CS [3]. Therefore, we may turn to a cosparse signal model (4) and its corresponding optimization model for compressive EEG sensing [21]. Similarly to sparse synthesis signal model, the performance of cosparsity based CS depends on the sampling matrix and cosparse analysis matrix [23].

One sufficient condition on the measurement matrix is called D-RIP. A theoretical result implies that any matrix which satisfies the RIP will satisfy the D-RIP when multiplied by a random sign matrix [24]. Therefore we can take advantage of almost all the current CS-ADC architectures as mentioned before.

To achieve a better reconstruction performance for EEG signals, a suitable analysis dictionary should be selected too. Current candidates include the shift-invariant wavelet transform, and the finite difference operator, etc. The finite difference operator, which is closely connected to the total variation method, is a very popular choice. However, the corresponding structure encouraged by the total variation method is approximate piece-wise constant whereas the 1st-order difference matrix is chosen to be the cosparse analysis dictionary. As can be seen in Fig. 1, most EEG signals are approximately piecewise linear. Piecewise linearity means the 1st-order derivatives are piecewise constant, which implies that it is better to choose the 2nd-order difference matrix as the cosparse analysis dictionary.

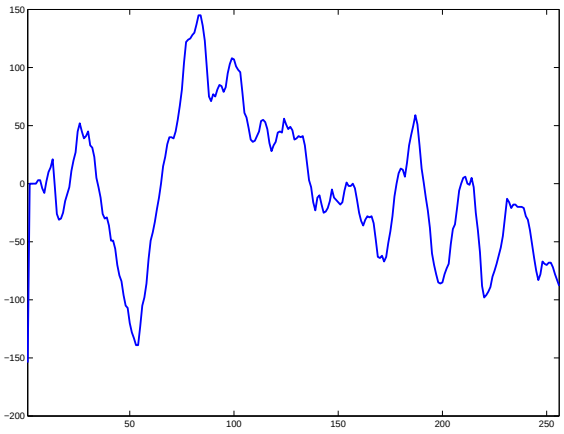


Fig. 1: A segment of EEG signal [25] [26].

B. Low-rank multi-channel EEG signals

The rank of a matrix is the maximum number of linearly independent columns (or rows) of the matrix. Low rank property means that the rank of the matrix is much smaller than both the number of rows and columns. In matrix computation, low rank also means the number of significant singular values is much smaller than number of rows or columns. The singular value decomposition (SVD) of a low rank matrix $\mathbf{X} \in \mathbb{R}^{N\times P}$ can be represented as

$$\mathbf{X} = \mathbf{U\Sigma V}^T = \sum_{n=1}^{\min(N,P)} \sigma_n \mathbf{u}_n \mathbf{v}_n^T \tag{5}$$

where $\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_N \end{bmatrix}$, $\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_P \end{bmatrix}$ are unitary matrices, and $\mathbf{\Sigma} \in \mathbb{R}^{N\times P}$ is a diagonal matrix with nonnegative real numbers $\sigma_n, \mathrm{n} = 1, 2, \cdots, \min(N, P)$ which are called singular values, and most of the singular values are zero or almost zero.

Motivated by previous work on low rank based multi-channel EEG compression [16] [17] and [18], a low rank structure may be used to regularize the optimization model for multi-channel EEG signal recovery from compressive measurements. In multi-channel EEG recording, spatially adjacent channels are significantly correlated. To exploit spatial correlation, the multi-channel EEG signals are put columnwise into a matrix. The newly formed EEG data matrix has only a few nonzero singular values. Fig. 2a shows the 64-channel EEG signals, and Fig. 2b shows the singular values of the data matrix [26] [27]. We can see that only a few singular values are significantly larger than zero, which meets the requirement of low rank approximation based matrix recovery [28]. The smaller the number of the nonzero singular values of data matrix, the better performance the low rank matrix recovery method can achieve. If the channels are more correlated, probably the nonzero singular values of the data matrices would be less, and the estimation would be more accurate.

III. Simultaneous Cosparsity and Low Rank Optimization

The optimization model for cosparse signal recovery can be formulated as

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} \ \|\mathbf{\Omega x}\|_0 \\ &\text{subject to} \ \ \mathbf{y} = \mathbf{\Phi x} \end{aligned} \tag{6}$$

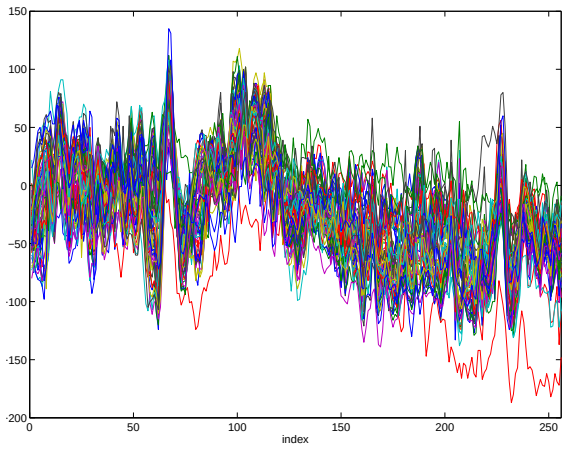
Here we call (6) the analysis L0 optimization. When the EEG system records P channels simultaneously, the extension of analysis L0 optimization to multi-channel data is:

$$\begin{aligned} &\underset{\mathbf{X}}{\text{minimize}} \ \|\mathrm{vec}\,(\mathbf{\Omega X})\|_0 \\ &\text{subject to} \ \ \mathbf{Y} = \mathbf{\Phi X} \end{aligned} \tag{7}$$

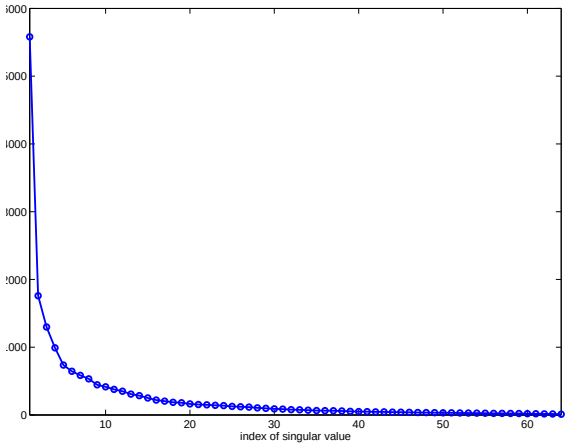
where vec($\mathbf{X}$) puts all the columns of $\mathbf{X}$ into one column vector sequentially. Solving the analysis L0 optimization is an NP-hard problem. A series of solvers are summarized in [20], such as analysis L1 optimization, greedy analysis pursuit (GAP), etc.

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

$$\begin{aligned} &\underset{\mathbf{X}}{\text{minimize}} \ \|\mathbf{X}\|_{\text{Schatten}-0} \\ &\text{subject to} \ \ \mathbf{Y}{=}\mathbf{\Phi X} \end{aligned} \tag{8}$$



(a) Multi-channel EEG signals with 64 channels



(b) Singular values of the multi-channel EEG data matrix

Fig. 2: A segment of multi-channel EEG signals and the corresponding singular values of the EEG data matrix [26] [27].

where $\|\mathbf{X}\|_{\text{Schatten}-0}$ is the Schatten-0 norm which counts the number of the nonzero singular values of $\mathbf{X}$ [29]. A variety of methods to solve it can be found in [30], such as interior point method for semi-definite programming (SDP), projected subgradient Methods, singular value thresholding, etc.

To simultaneously exploit the cosparsity and low rank structure in multi-channel EEG signal reconstruction from the compressive measurements, both ℓ_0 norm and Schatten-0 norm based constraints should be 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:

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \ & \|\text{vec}\,(\mathbf{\Omega X})\|_0 + \|\mathbf{X}\|_{\text{Schatten}-0} \\ \text{subject to} \ & \mathbf{Y}=\mathbf{\Phi X} \end{aligned} \tag{9}$$

IV. SOLUTIONS

A. Convex relaxation

To solve the SCLR optimization (9), 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_0 norm sums all the absolute values of the entries, i.e. $\|\mathbf{x}\|_1 = \sum_{n=1}^{N} |x_n|$; and the Schatten-0 norm is called nuclear norm too, which sums all the singular values of the data matrix, i.e. $\|\mathbf{X}\|_{\text{Schatten}-0} = \|\mathbf{X}\|_* = \sum_{n=1}^{\min(N,P)} \sigma_n$. The newly formed convex simultaneous cosparsity and low rank (CSCLR) optimization model can be formulated as:

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \ & \|\text{vec}\,(\mathbf{X})\|_1 + \|\mathbf{X}\|_* \\ \text{subject to} \ & \mathbf{Y}=\mathbf{\Phi X} \end{aligned} \tag{10}$$

By introduction of new nonnegative variables $\mathbf{e}$ and f, (10) can be expressed as:

$$\begin{aligned} \underset{\mathbf{X},\mathbf{e}>0,f\geq 0}{\text{minimize}} \ & \mathbf{1}^T\mathbf{e} + f \\ \text{subject to} \ & \mathbf{Y}=\mathbf{\Phi X} \\ & \|\mathbf{X}\|_* \leq f \\ & -\mathbf{e} \prec \text{vec}\,(\mathbf{\Omega X}) \prec \mathbf{e} \end{aligned} \tag{11}$$

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

The nuclear norm constraint can be replaced by its linear matrix inequality (LMI) equivalent; the approximation constraints can also be expressed via LMIs using Schur complements [31]. The obtained optimization model is:

$$\begin{aligned} \underset{\mathbf{X},\mathbf{e}>0,f\geq 0}{\text{minimize}} \ & \mathbf{1}^T\mathbf{e} + 2f \\ \text{subject to} \ & \mathbf{Y}=\mathbf{\Phi X} \\ & -\mathbf{e} \prec \text{vec}\,(\mathbf{\Omega X}) \prec \mathbf{e} \\ & \begin{bmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{B} \end{bmatrix} \geq 0 \\ & \text{Tr}\,(\mathbf{A}) + \text{Tr}\,(\mathbf{B}) < f \end{aligned} \tag{12}$$

where $\mathrm{A} = \mathrm{A}^T$ and $\mathbf{B} = \mathbf{B}^T$ are new variables. (12) is an SDP which can be solved by interior-point method.

B. ADMM

Besides the classical SDP way, another method, called ADMM, can be used to solve the SCLR optimization [32]. With individual constraints on the same variables in each constraint, (10) can be rewritten into global consensus optimization with local variables $\mathbf{X}_i$, $i = 1,2$ and a common global variable $\mathbf{X}$ as:

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

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

$$\begin{aligned} \underset{\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}}{\text{minimize}} \ & \|\text{vec}\,(\mathbf{\Omega X}_1)\|_1 + \|\mathbf{X}_2\|_* \\ \text{subject to} \ & \bar{\mathbf{Y}}=\bar{\mathbf{\Phi}}\mathbf{X}_1; \bar{\mathbf{Y}}=\bar{\mathbf{\Phi}}\mathbf{X}_2 \end{aligned} \tag{14}$$

where

$$\bar{\mathbf{Y}} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{X} \end{bmatrix} \tag{15}$$

$$\bar{\mathbf{\Phi}} = \begin{bmatrix} \mathbf{\Phi} \\ \mathbf{I} \end{bmatrix} \tag{16}$$

The augmented Lagrangian of (14) is:

$$\begin{aligned} &L_\rho(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Z}_1, \mathbf{Z}_2) = \|\mathrm{vec}(\mathbf{\Omega X_1})\|_1 + \|\mathbf{X}_2\|_* \\ &+\mathrm{vec}(\mathbf{Z}_1)^T \mathrm{vec}\left(\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_1\right) + \mathrm{vec}(\mathbf{Z}_2)^T \mathrm{vec}\left(\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_2\right) \\ &+\frac{\rho}{2}\left\|\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_1\right\|_F^2 + \frac{\rho}{2}\left\|\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_2\right\|_F^2 \end{aligned} \tag{17}$$

The resulting ADMM algorithm of the scaled dual form is the following

$$\mathbf{X}_1^{t+1} := \arg\min_{\mathbf{X}_1}\left(\|\mathrm{vec}(\mathbf{\Omega X}_1)\|_1 + \frac{\rho}{2}\left\|\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_1 + \bar{\mathbf{\Phi}}\mathbf{U}_1^t\right\|_F^2\right) \tag{18}$$

$$\mathbf{X}_2^{t+1} := \arg\min_{\mathbf{X}_2}\left(\|\mathbf{X}_2\|_* + \frac{\rho}{2}\left\|\bar{\mathbf{Y}} - \bar{\mathbf{\Phi}}\mathbf{X}_2 + \bar{\mathbf{\Phi}}\mathbf{U}_2^t\right\|_F^2\right) \tag{19}$$

$$\mathbf{X}^{t+1} = \frac{1}{2}\left(\mathbf{X}_1^{t+1} + \mathbf{X}_2^{t+1}\right) \tag{20}$$

$$\begin{aligned} \mathbf{U}_1^{t+1} &= \mathbf{U}_1^t + \left(\mathbf{X}_1^{t+1} - \mathbf{X}_1^t\right) \\ \mathbf{U}_2^{t+1} &= \mathbf{U}_2^t + \left(\mathbf{X}_2^{t+1} - \mathbf{X}_2^t\right) \end{aligned} \tag{21}$$

For both (18) and (19), there are many computationally efficient algorithms [20] [30]. For example, to solve (18) analysis L1 optimization, GAP can be used; and to solve (19), SDP method, singular value thresholding can be used. 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:
  $\mathbf{X}_1^{t+1} := \arg\min_{\mathbf{X}_1}\left(\|\mathrm{vec}(\mathbf{\Omega X}_1)\|_1 + \frac{\rho}{2}\left\|\mathbf{Y} - \bar{\mathbf{\Phi}}\mathbf{X}_1 + \bar{\mathbf{\Phi}}\mathbf{U}_1^t\right\|_F^2\right)$;
- step 2: update of the low rank optimization:
  $\mathbf{X}_2^{t+1} := \arg\min_{\mathbf{X}_2}\left(\|\mathbf{X}_2\|_* + \frac{\rho}{2}\left\|\mathbf{Y} - \bar{\mathbf{\Phi}}\mathbf{X}_2 + \bar{\mathbf{\Phi}}\mathbf{U}_2^t\right\|_F^2\right)$;
- 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:
  $\mathbf{U}_1^{t+1} = \mathbf{U}_1^t + \left(\mathbf{X}_1^{t+1} - \mathbf{X}_1^t\right)$;
  $\mathbf{U}_2^{t+1} = \mathbf{U}_2^t + \left(\mathbf{X}_2^{t+1} - \mathbf{X}_2^t\right)$;
- 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 discussed in the literature. Generally, the convergence to optimum can be guaranteed when the epigraph of g_i:

$$\mathrm{epi} g_i = \{(\mathbf{X}, \varepsilon) \ \ | g_i(\mathbf{X}) \le \varepsilon\} \tag{22}$$

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

$$\begin{aligned} &L_{\rho=0}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Z}_1, \mathbf{Z}_2) = \|\mathrm{vec}(\mathbf{\Omega X_1})\|_1 + \|\mathbf{X}_2\|_* \\ &+\mathrm{vec}(\mathbf{Z}_1)^T \mathrm{vec}\left(\mathbf{Y}_1 - \bar{\mathbf{\Phi}}\mathbf{X}_1\right) + \mathrm{vec}(\mathbf{Z}_2)^T \mathrm{vec}\left(\mathbf{Y}_2 - \bar{\mathbf{\Phi}}\mathbf{X}_2\right) \end{aligned} \tag{23}$$

has a saddle point. The proof can be found in [33].

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

V. Numerical Experiments

To demonstrate the performance of the possible methods for EEG signal recovery from the compressive measurements, we perform two groups of numerical experiments. In section V-A, 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 of the current algorithms are tested in order to make comparison, such as BSBL which is reported to be the best of all the current candidates for EEG signal recovery from compressive measurements [3], orthogonal matching pursuit (OMP) which is a proper representative of the classical sparse signal recovery algorithms [2]. In section V-B, a group of multi-channel EEG signals are recovered by the proposed SCLR optimization algorithms, as well as SOMP, SGAP, BSBL.

The used EEG data is from the *Physiobank* database [25] [26]. It consists of 23-channel EEG recordings from pediatric subjects with intractable seizures. All signals 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.

In all experiments, as argued by our analysis in section II-A, 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 are Daubechies wavelets.

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

$$\mathrm{SSR} = \frac{M}{N} \times 100\% \tag{24}$$

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:

$$\mathrm{MSE} = \sum_{l=1}^{L} \frac{\left\|\hat{\mathbf{X}}_l - \mathbf{X}\right\|_F^2}{LNP} \tag{25}$$

where $\mathbf{X}$ is the true EEG data with P channels each of length N, $\hat{\mathbf{X}}_l$ is its estimate in the l-th Monte Carlo simulation, and L is the number of Monte Carlo simulations. Both $\mathbf{X}$ and $\hat{\mathbf{X}}_l$ are normalized by their Frobenius norms respectively.

When $P = 1$, 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 [21], percent of root-mean-square difference(PRD) [2]. Another evaluation function is mean cross-correlation (MCC). It is equivalent to the Structural SIMilarity index (SSIM), which measures the similarity of two waveforms [34] [2] [3]. It can be formulated as:

$$\text{MCC} = \sum_{l=1}^{L} \frac{\text{vec}(\mathbf{X})^T \text{vec}\left(\hat{\mathbf{X}}_l\right)}{L\|\mathbf{X}\|_F \left\|\hat{\mathbf{X}}_l\right\|_F} \tag{26}$$

A. Single channel EEG signal recovery

To find the best methods for single channel EEG signal reconstruction, $L = 500$ segments of EEG data are used. This data are from 1 of the 23 channels. The length of the data is $N = 256$. Each segment of EEG signal is normalized by its ℓ_2 norm. Four methods are employed to do the EEG signal reconstruction, namely OMP, BSBL, GAP, and analysis L1 optimization.

To show how the proposed methods work, we take a segment of EEG signal and reconstruct it from the compressive measurement with SSR = 0.35. The reconstructed and real signals are shown in Fig. 3. 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.

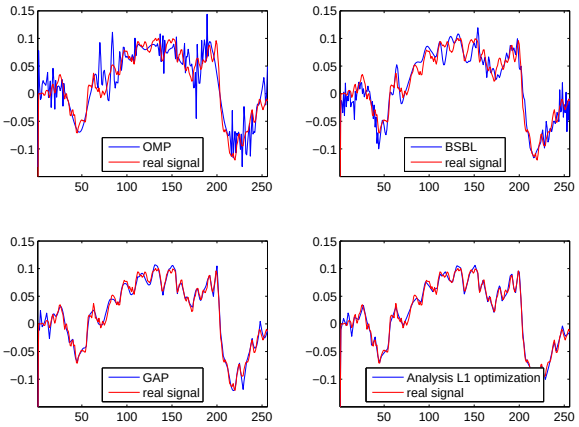
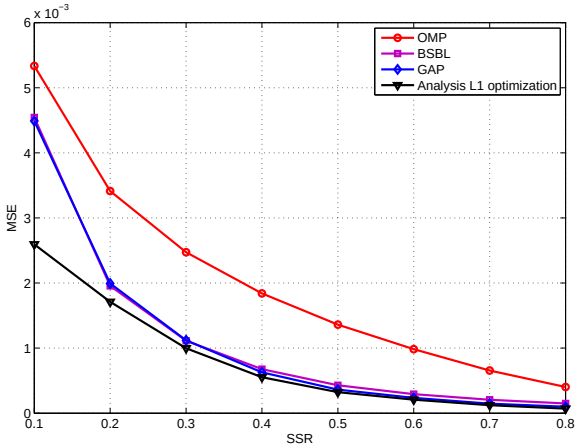
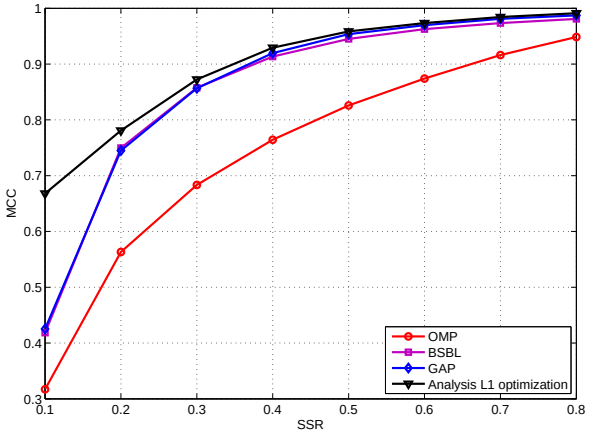


Fig. 3: EEG signals reconstructed by OMP, BSBL, GAP, and analysis L1 optimization with SSR = 0.35.

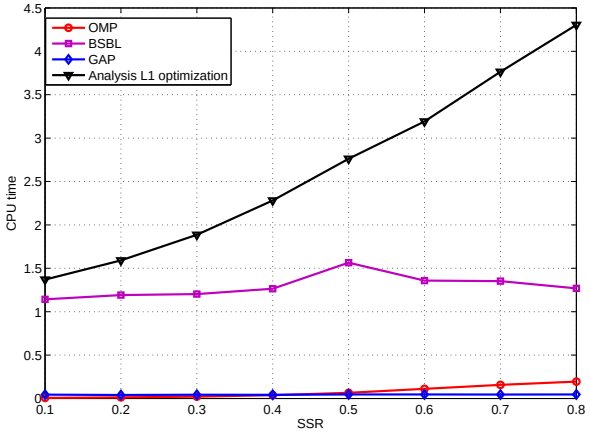
Fig. 4a, and Fig. 4b give the values of MSE and MCC of GAP, OMP, BSBL with different sub-sampling ratio. We can see that analysis L1 optimization, GAP and BSBL have similar accuracy performance, and they outperform OMP. Analysis L1 optimization is slightly more accurate than GAP, and GAP is lightly more accurate than BSBL. In Fig. 4c, we can see the greedy algorithms GAP and OMP are much faster than BSBL and analysis L1 optimization, and GAP is even a little faster than OMP. Therefore, if we only care about the accuracy, the analysis L1 optimization is the best choice. Otherwise, if both accuracy and computational complexity are important, GAP should be a better choice.



(a) MSE



(b) MCC



(c) CPU time

Fig. 4: Different performance evaluations of single channel EEG signal recovery from compressive measurements with different SSRs: (a) MSE vs SSR; (b) MCC vs SSR; (c) CPU Time vs SSR.

B. Multi-channel EEG signal recovery

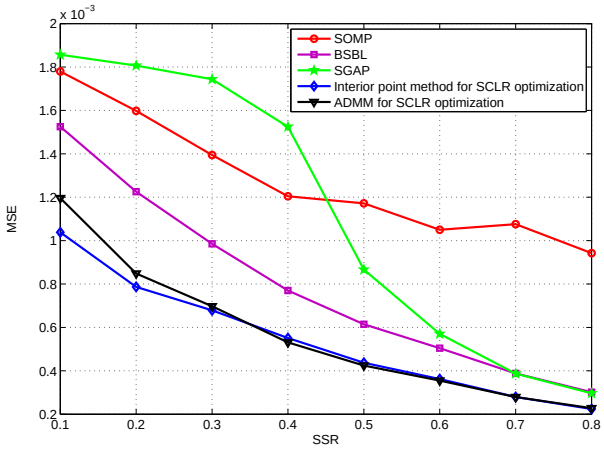
In these experiments, most of the parameters are selected as in section V-A. For multi-channel EEG data, the number of channels are $P = 23$. $L = 50$ segments of 23-channel EEG data are used. Two SCLR optimization algorithms are used. Interior point method for SCLR optimization and ADMM for SCLR optimization 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, SGAP.

Fig. 5a, 5b and 5c display the values of MSE, MCC and CPU times of an interior point method for SCLR optimization, ADMM for SCLR optimization, BSBL, SOMP, and SGAP with different values of sub-sampling ratio. 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. Comparing the speed of these two solutions for SCLR optimization, the ADMM for SCLR optimization is faster. In Fig. 5c we can see that the greedy algorithms SOMP and SGAP are much faster than the rest. But considering that their accuracy is much worst 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 current methods.

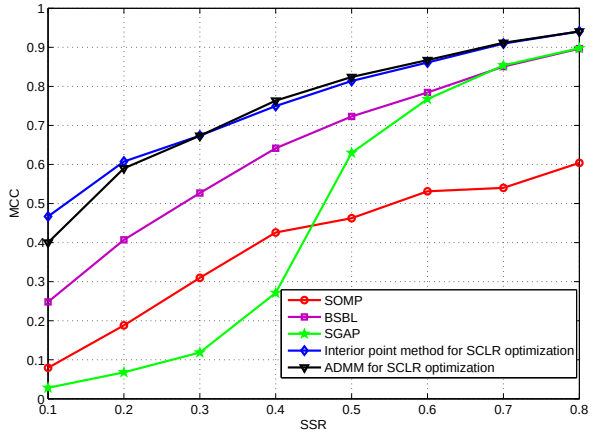
The two assumed data structures in the SCLR optimization method are cosparsity and low rank. Since the 2nd-order different matrix is chosen to be the dictionary, the cosparsity is approximately piecewise linearity. The cosparsity requires the each signal has only a few elements whose differences are significant. The low rank structure requires that the data matrix is of low rank. The low rank relates to the linear correlation between channels. If the signal with EEG signal's characteristic patterns depends on either one of this data structure, it should be effective. In additional numerical experiments, we find the SCLR optimization method performs similarly on EEG data with only seizure segments.

VI. Conclusion

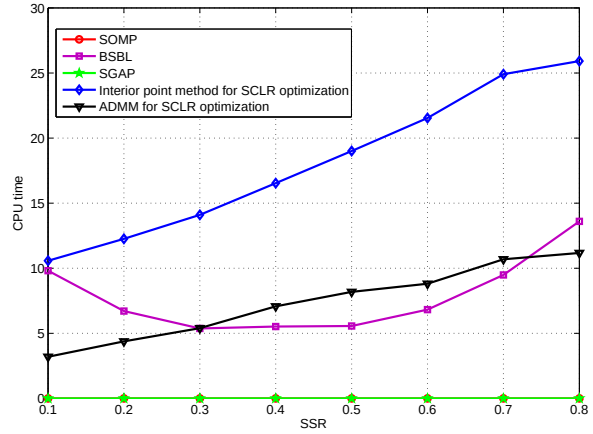
With the 2nd-order difference matrix as the cosparse analysis dictionary, the EEG signals' cosparsity is exploited for the single-channel EEG signal recovery from compressive measurements. To further enhance the performance, cosparsity and low rank structure are jointly used in the multi-channel EEG signal recovery. In the proposed new optimization model, the ℓ_0 norm constraint is used to enforce cosparse structure 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 existed 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 shows that the SCLR optimization outperforms all the previous methods.



(a) MSE



(b) MCC



(c) CPU time

Fig. 5: Different performance evaluations of multi-channel EEG signal recovery from compressive measurements with different SSRs: (a) MSE vs SSR; (b) MCC vs SSR; (c) CPU Time vs SSR.

References

[1] C. Bachmann, M. Ashouei, V. Pop, M. Vidojkovic, H. Groot, and B. Gyselinckx, "Low-power wireless sensor nodes for ubiquitous long-

term biomedical signal monitoring," *Communications Magazine, IEEE*, vol. 50, no. 1, pp. 20–27, 2012.

[2] A. M. Abdulghani, A. J. Casson, and E. Rodriguez-Villegas, "Compressive sensing scalp eeg signals: implementations and practical performance," *Medical & biological engineering & computing*, vol. 50, no. 11, pp. 1137–1145, 2012.

[3] Z. Zhang, T.-P. Jung, S. Makeig, and B. D. Rao, "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.

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

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

[6] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Springer, 2013.

[7] 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.

[8] S. Aviyente, "Compressed sensing framework for eeg compression," in *Statistical Signal Processing, 2007. SSP'07. IEEE/SP 14th Workshop on*. IEEE, 2007, pp. 181–184.

[9] A. M. Abdulghani, A. J. Casson, and E. Rodriguez-Villegas, "Quantifying the performance of compressive sensing on scalp eeg signals," in *Applied Sciences in Biomedical and Communication Technologies (ISABEL), 2010 3rd International Symposium on*. IEEE, 2010, pp. 1–5.

[10] S. Senay, L. F. Chaparro, M. Sun, and R. J. Sclabassi, "Compressive sensing and random filtering of eeg signals using slepian basis," *Proceedings of the EURASIP EUSIPCO*, vol. 8, 2008.

[11] B. Liu, Z. Zhang, H. Fan, and Q. Fu, "Compression via compressive sensing: A low-power framework for the telemonitoring of multi-channel physiological signals," *arXiv preprint arXiv:1309.4136*, 2013.

[12] P. J. Durka, A. Matysiak, E. M. Montes, P. V. Sosa, and K. J. Blinowska, "Multichannel matching pursuit and eeg inverse solutions," *Journal of neuroscience methods*, vol. 148, no. 1, pp. 49–59, 2005.

[13] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *Signal Processing, IEEE Transactions on*, vol. 53, no. 7, pp. 2477–2488, 2005.

[14] 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.

[15] K. Srinivasan, J. Dauwels, and M. R. Reddy, "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.

[16] J. Dauwels, K. Srinivasan, M. Ramasubba Reddy, and A. Cichocki, "Multi-channel eeg compression based on 3d decompositions," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 637–640.

[17] ——, "Multi-channel eeg compression based on matrix and tensor decompositions," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 629–632.

[18] J. Dauwels, K. Srinivasan, M. Reddy, and A. Cichocki, "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.

[19] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "Cosparse analysis modeling-uniqueness and algorithms," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 5804–5807.

[20] ——, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[21] Y. Liu, M. De Vos, I. Gligorijevic, V. Matic, Y. Li, and S. Van Huffel, "Multi-structural signal recovery for biomedical compressive sensing," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2794–2805, 2013.

[22] Y. Avonds, Y. Liu, and S. Van Huffel, "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, submitted.

[23] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.

[24] F. Krahmer and R. Ward, "New and improved johnson-lindenstrauss embeddings via the restricted isometry property," *SIAM Journal on Mathematical Analysis*, vol. 43, no. 3, pp. 1269–1281, 2011.

[25] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[26] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[27] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "Bci2000: a general-purpose brain-computer interface (bci) system," *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 6, pp. 1034–1043, 2004.

[28] E. Candès and B. Recht, "Simple bounds for recovering low-complexity models," *Mathematical Programming*, vol. 141, no. 1-2, pp. 577–589, 2013.

[29] A. Rohde, A. B. Tsybakov *et al.*, "Estimation of high-dimensional low-rank matrices," *The Annals of Statistics*, vol. 39, no. 2, pp. 887–930, 2011.

[30] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.

[31] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *American Control Conference, 2001. Proceedings of the 2001*, vol. 6. IEEE, 2001, pp. 4734–4739.

[32] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "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.

[33] J. Eckstein and D. P. Bertsekas, "On the douglasłrachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1-3, pp. 293–318, 1992.

[34] 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.