Compressed Sensing for Implantable Neural Recordings Using Co-sparse Analysis Model and Weighted ℓ_1-Optimization

Biao Sun, *Member, IEEE*, Wenfeng Zhao, *Member, IEEE*, and Xinshan Zhu*, *Member, IEEE*

Abstract

Reliable and energy-efficient wireless data transmission remains a major challenge in resource-constrained wireless neural recording tasks, where data compression is generally adopted to relax the burdens on the wireless data link. Recently, Compressed Sensing (CS) theory has successfully demonstrated its potential in neural recording application. The main limitation of CS, however, is that the neural signals have no good sparse representation with commonly used dictionaries and learning a reliable dictionary is often data dependent and computationally demanding. In this paper, a novel CS approach for implantable neural recording is proposed. The main contributions are: 1) The co-sparse analysis model is adopted to enforce co-sparsity of the neural signals, therefore overcoming the drawbacks of conventional synthesis model and enhancing the reconstruction performance. 2) A multi-fractional-order difference matrix is constructed as the analysis dictionary, thus avoiding the dictionary learning procedure and reducing the need for previously acquired data and computational resources. 3) By exploiting the statistical priors of the analysis coefficients, a weighted analysis ℓ_1-minimization (WALM) algorithm is proposed to reconstruct the neural signals. Experimental results on Leicester neural signal database reveal that the proposed approach outperforms the state-of-the-art CS-based methods. On the challenging high compression ratio task, the proposed approach still achieves high reconstruction performance and spike classification accuracy.

B. Sun and X. Zhu are with the School of Electrical Engineering and Automation, Tianjin University, Tianjin, 300072, China. Email: {sunbiao, xszhu}@tju.edu.cn.

W. Zhao is with the Department of Electrical and Computer Engineering, National University of Singapore, 117583, Singapore. Email: elezhwf@nus.edu.sg.

arXiv:1602.00430v1 [cs.IT] 1 Feb 2016

Index Terms

Compressed sensing, implantable neural recording, co-sparse analysis model, fractional order difference sequence, weighted ℓ_1-minimization

I. INTRODUCTION

Large-scale, multi-channel extracellular neural recording simultaneously from various brain regions are desired to investigate the neural activities from different neuron ensembles, local circuits and brain networks [1], [2]. Such technological capabilities would advance the understanding of brain functions, and moreover, brain-machine interfaces and translational neurotechnologies would become feasible for sophisticated prosthetic devices and disease treatment. In conventional static recording scenario, large amounts of neural data are generated (on the order of tens of Megabytes per second) and tethered cables or wires are commonly adopted for data streaming purposes. However, its applications would be limited owing to tissue infection for subcutaneous, chronic recording tasks as well as in the neuroscience experiments to study awake and free behaving animals models [3], [4]. Wireless neural recording devices overcome the above-mentioned limitations and would greatly expand the research and application scenarios.

Nevertheless, wireless neural recording devices would compromise among various system-level considerations, including system complexity, power budget and volume miniaturization, and component-level design aspects, such as neural recording amplifiers and analog-to-digital converters (ADCs), neural signal processors, data transceivers and antenna designs. Arguably, the most challenging component in a wireless neural recording device is a reliable, high-throughput and energy-efficient wireless data link, as the wireless link dominates the system channel count, resolution, and energy-efficiency. Although continuous progress is made on data rate and energy efficiency of RF transceivers, it is still prohibitive to adopt such wireless links due to practical limitations of experimental and clinical procedures. Another straightforward approach is to perform on-chip compression before transmission to relax the bandwidth constraints, such as spike detection based approaches [5]–[8] and lossy data compression via DWT [9], etc. These approaches significantly reduce the neural data that needed to be transmitted, yet the compression hardware overhead of the on-chip resources and excessive power consumption cannot be neglected.

Recently, the field of Compressed Sensing (CS) [10], [11] has shown potential in achieving compression and reconstruction performance comparable to the previous approaches but with simpler hardware resources [12]–[14]. The CS approach requires a set of the random measurements of the original signals, and avoids the need for dedicated DSPs and leaves most of the computational burden to off-chip processing. Its main challenge, however, is that the spike segments are not sparse on common dictionaries such as

Discrete Cosine Transform (DCT) basis and Discrete Fourier Transform (DFT) basis. Reconstructing the spikes using these dictionaries will severely degrade the performance. Therefore, a careful design of the sparsifying dictionary is needed to guarantee the compression performance [15]. To alleviate this issue, various dictionary-learning based algorithms are proposed for neural data compression [12], [13]. Zhang *et al.* [12] proposed learning dictionaries using K-SVD and developed a signal-dependent CS approach to compress the data. Suo *et al.* [13] proposed to use the recorded neural data directly as the sparsity dictionary. However, these algorithms are computational demanding and highly signal-dependent. This indicates that iterative training processes are required during practical neural recording applications, which is unfavorable in most experimental settings.

This paper proposes a novel CS framework for implantable neural recordings that is capable of recovering neural spikes with high compression ratio but avoiding the sparsity dictionary learning procedure. The main contributions of the work are as follows.

i) Instead of using conventional synthesis model in CS, the analysis model is adopted to enforce co-sparsity of neural signals, overcoming drawbacks of conventional model and enhancing the reconstruction performance. To our best knowledge, this is the first time that neural signal reconstruction problem is solved by using the analysis model of CS.

ii) Based on the piecewise smooth structures in neural signals, a multiple-fractional-order-difference matrix is constructed as the analysis dictionary. It not only has high co-sparsity with neural signals but also avoids the dictionary learning procedure, saving both computational resources and data storage space.

iii) The statistical priors of the analysis coefficients among difference orders are deduced. The associated reconstruction algorithm, dubbed weighted analysis ℓ_1-minimization (WALM), is proposed to improve the reconstruction performance by embedding the multiple orders knowledge within penalty weights.

The remainder of this paper is organized as follows. Section II describes the CS-based implantable neural recording system architecture, the relevant background of synthesis model and co-sparse analysis model. Section III introduces the proposed construction method of the multiple-fractional-order-difference dictionary. Section IV covers the weighted analysis ℓ_1-minimization algorithm for neural spike reconstruction. In Section V, the experimental results are presented and compared to state-of-the-art CS-based reconstruction methods. Section VI concludes the paper.

Throughout the paper, boldface capital letters (e.g., A) denote matrices, boldface lowercase letters (e.g., x) denote vectors, not bold letters (e.g., c) denote scalars, and boldface calligraphic letters (e.g., $\mathcal{I}$) specify number sets. For a vector x, we use x_i to denote its ith entry, and we use $\|x\|_2$, $\|x\|_1$, and $\|x\|_0$ to indicate its ℓ_2, ℓ_1, and ℓ_0 norms, respectively. For a matrix A, we use A_i to denote its ith
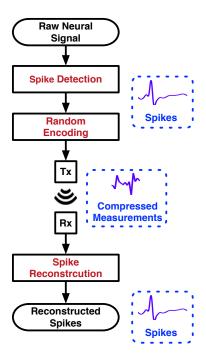
Fig. 1. Diagram of the compressed sensing system for implant neural recording.

row or column depending on the situation it is used. For a set $\mathcal{I}$, we use $|\mathcal{I}|$ to indicate its cardinality. For a random variable a, its probability distribution function (*pdf*) is denoted by $p(a)$, and its standard deviation is indicated by σ_a.

II. SYSTEM OVERVIEW AND SPARSE MODELS

A. System Overview

CS-based wireless neural recording system architecture is briefly depicted in Fig. 1. The recorded raw neural data are first conditioned into appropriate signal amplitude and bandwidth through amplifiers, filters and digitized via Nyquist-rate ADCs. Second, the neural spike events are detected through threshold crossing techniques and aligned temporally ("Spike Detection" in the figure). The aligned segments containing the spikes are then compressed via randomized encoding circuit ("Random Encoding" in the figure) based on the compressive sensing theory, and the compressed data are transmitted via wireless transmitters (e.g., Bluetooth, Zig-Bee, Wi-Fi). On the receiver side, the random measurements of aligned spikes are reconstructed through some specific algorithms ("Spike Reconstruction" in the figure) at workstations or fusion centers.

B. Compressed Sensing and Synthesis Model

Compressed sensing is an emerging low-rate sampling scheme for the signals that are known to be sparse or compressible in certain basis. Assume a signal $\boldsymbol{x} \in \mathbb{R}^n$ is measured by a simple matrix-vector multiplication,

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n}$ is called the sensing matrix, $\boldsymbol{y} \in \mathbb{R}^m$ is the compressed measurement vector, $\boldsymbol{e} \in \mathbb{R}^m$ denotes the measurement noise. Usually Eq. (1) is undetermined, i.e., $m < n$, and the ratio m/n is called the compression ratio of CS. In this undetermined system, the signal $\boldsymbol{x}$ cannot be uniquely retrieved from sensing matrix $\boldsymbol{\Phi}$ and measurements $\boldsymbol{y}$. However, if the $\boldsymbol{x}$ can be described using a synthesis model [16], i.e,

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\theta}, \tag{2}$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{n \times n}$ is a pre-defined dictionary, and the signal's representation $\boldsymbol{\theta} \in \mathbb{R}^n$ is assumed to be k-sparse, i.e.,

$$\|\boldsymbol{\theta}\|_0 \triangleq |\mathrm{supp}(\boldsymbol{\theta})| = k \ll n, \tag{3}$$

or is well-approximated by a k-sparse vector. The name "synthesis" comes from the relation (2), with the obvious interpretation that the model describes a way to synthesize a signal. Therefore, based on Eq. (1) and (2), the compressed measurements $\boldsymbol{y}$ can be represented as

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\theta} + \boldsymbol{n} = \boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{e}, \tag{4}$$

where $\boldsymbol{A} = \boldsymbol{\Phi}\boldsymbol{\Psi}$. Due to the sparse prior knowledge of $\boldsymbol{\theta}$, it is possible to estimate $\boldsymbol{\theta}$ via the ℓ_0 minimization formulation as

$$\hat{\boldsymbol{\theta}} = \min\|\boldsymbol{\theta}\|_0, \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\theta}\|_2^2 < \epsilon, \tag{5}$$

where ϵ is the tolerance of noise or modeling errors. Calculating the solution is very hard because Eq. (5) is an NP-hard problem [11]. Generally, one seeks the solution of a relaxed convex optimization problem [17], in which $\|\boldsymbol{\theta}\|_0$ is replaced with $\|\boldsymbol{\theta}\|_1$ as

$$\hat{\boldsymbol{\theta}} = \min\|\boldsymbol{\theta}\|_1, \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\theta}\|_2^2 < \epsilon. \tag{6}$$

Under the condition of Restricted Isometry Property (RIP) [18], minimizing ℓ_1 has been theoretically proven to be equivalent to minimizing ℓ_0. Moreover, ℓ_1 minimization is convex and can be solved within polynomial time. After estimating the sparse coefficient $\boldsymbol{\theta}$, the original signal $\boldsymbol{x}$ can be recovered directly by Eq. (2).

C. Co-sparse Analysis Model

While the synthesis model has been extensively studied, its "twin model" that takes an analysis point of view has been left aside almost untouched [19]. The alternative assumes that for a signal of interest, the *analysis coefficients* vector

$$\boldsymbol{z} = \boldsymbol{\Omega}\boldsymbol{x} \tag{7}$$

is expected to be sparse, where $\boldsymbol{\Omega} \in \mathbb{R}^{l \times n}$ is a possibly redundant analysis dictionary ($l \geq n$), and the ratio $\rho = l/n$ is called the redundant ratio. The co-sparsity of a signal $\boldsymbol{x}$ with respect to $\boldsymbol{\Omega}$ is defined as the number of zeros in the vector $\boldsymbol{z}$, i.e.,

$$k_{\text{co}} = l - \|\boldsymbol{z}\|_0, \tag{8}$$

and the index set of the zero entries of $\boldsymbol{z}$ is called the co-support of $\boldsymbol{x}$. It is worth noting that for a square and invertible dictionary, the synthesis and the analysis models are the same with $\boldsymbol{\Psi} = \boldsymbol{\Omega}^{-1}$ [19]. While the analysis model may seem similar to the synthesis counterpart one, it is in-fact very different when dealing with a redundant dictionary $p > n$ [20]. The traditional synthesis model puts an emphasis on the non-zeros of the sparse vector $\boldsymbol{\theta}$, but the co-sparse analysis model draws its strength from the zeros of the analysis vector $\boldsymbol{z}$. The optimization problem for co-sparse signal recovery can be formulated as

$$\hat{\boldsymbol{x}} = \min\|\boldsymbol{\Omega}\boldsymbol{x}\|_0, \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 < \epsilon. \tag{9}$$

Here we call (9) the analysis ℓ_0-minimization. There is also a classical way to relax the nonconvex ℓ_0 norm into convex ℓ_1 norm, i.e.,

$$\hat{\boldsymbol{x}} = \min\|\boldsymbol{\Omega}\boldsymbol{x}\|_1, \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 < \epsilon. \tag{10}$$

We call (10) the analysis ℓ_1-minimization (AL1). Several sufficient conditions theoretically guarantee the successful recovery of the original signal from the compressed measurement using (10), such as the restricted isometry property adapted to the dictionary (D-RIP), restricted orthogonal projection property (ROPP), etc [20]–[22].

III. ANALYSIS DICTIONARY CONSTRUCTION

In this section, we focus on the construction of the analysis dictionary $\boldsymbol{\Omega}$ so that the analysis coefficients $\boldsymbol{\Omega}\boldsymbol{x}$ are sparse. It is worth noting that when dealing with a square (and invertible) matrix $\boldsymbol{\Omega}$, the analysis model is completely equivalent to the synthesis one, and in such a case, the synthesis-dictionary construction methods can be used to build $\boldsymbol{\Omega}$. In this work, we concentrate on the redundant case $l > n$, where the two models depart, and where the analysis model becomes more powerful.

A. *Multiple-integer-order-difference Matrix*

Prior works show that many types of signals [23], [24], e.g., EEG and ECG signals, often reveals approximately piecewise smooth structure [25]–[28]. This structure exhibits gradient sparsity, i.e., signals will become sparse when differenced with some specific orders. Moreover, investigations of the statistical properties using the available implantable neural signals show that neural spikes are also approximately piecewise smooth, implying that neural spikes fit the co-sparse signal model (7) well with the integer-order-difference (IOD) sequence [29] as analysis dictionary. For an n-length signal $\boldsymbol{x}$, the IOD sequence of $\boldsymbol{x}$ is defined as

$$\Delta^r(\boldsymbol{x}) = \sum_{k=0}^{r}(-1)^k \binom{r}{k} x_{i+k}, \ \ i = 1, \ldots, n, \tag{11}$$

where $\Delta^r(\cdot)$ denotes IOD operator and $r \in \mathbb{N}$ is difference order. The IOD sequence can be reformulated into matrix form $\boldsymbol{D}^r$ as proposed in [29]. For simplicity, we do not distinguish between the two forms in the rest of the paper, and we use matrix form to build the analysis dictionary.

Fig. 2 presents a histogram of the co-sparsities of 1000 spikes[1] using 2nd order IOD matrix as the analysis dictionary. As can be seen, the co-sparsities are all strictly high. Furthermore, to construct a redundant analysis dictionary, we seek to promote co-sparsity over multiple orders difference sequence rather than a single one, and propose using a multiple-integer-order-difference (MIOD) matrix composed of a concatenation of q IOD matrices, i.e.,

$$\boldsymbol{\Omega}_{\text{MIOD}} \triangleq \frac{1}{\sqrt{q}}[\boldsymbol{D}^{r_0}, \boldsymbol{D}^{r_1}, \ldots, \boldsymbol{D}^{r_{q-1}}]^T, \tag{12}$$

where $\boldsymbol{D}^{r_i}$ denotes the IOD matrix of order r_i. The corresponding order set is

$$\mathcal{R} = \{r_0, r_1, \ldots, r_{q-1}\} \in \mathbb{N}^{q \times 1}. \tag{13}$$

Given the MIOD matrix, we propose the following prior, proportional to the multiple-order sparsity as

$$\|\boldsymbol{\Omega}_{\text{MIOD}}\boldsymbol{x}\|_0 = \sum_{r_i \in \mathcal{R}} \|\boldsymbol{D}^{r_i}\boldsymbol{x}\|_0. \tag{14}$$

It is worth noting that in this setting the analysis coefficients of each order contain all the signal information, which cannot be formulated in a synthesis model.

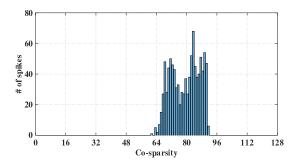[1]The spikes are randomly chosen from Leicester Easy2 dataset [30].

Fig. 2. A histogram of the effective co-sparsities of the 1000 spikes using 2nd order IOD matrix as the analysis dictionary. The spikes are randomly chosen from Leicester Easy2 dataset. The length of each spike is $n = 128$.
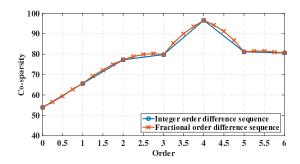

Fig. 3. The average co-sparsities of the 1000 spikes using IOD matrix (blue-circle curve) and FOD matrix (red-times curve) as the analysis dictionary, respectively. The spikes are randomly chosen from Leicester Easy2 dataset. The length of each spike is $n = 128$.

B. Multiple-fractional-order-difference Matrix

Although the MIOD matrix promotes higher co-sparsity than the single-order one, how to choose the order set appropriately remains a problem. The blue-circle curve in Fig. 3 shows the co-sparsity of 1000 spikes using IOD matrix as the analysis dictionary, from which we notice that the 4-th order analysis coefficients have maximum co-sparsity. If adding difference matrix of other orders to $\mathbf{\Omega}$, however, the co-sparsity of the analysis coefficients will not be optimal, and the signal reconstruction performance will severely degrade.

To solve this problem, we build the multiple-fractional-orders-difference (MFOD) matrix using the fractional-order-difference (FOD) sequence, which is a generalization of IOD sequence to fractional order. The FOD sequence is defined as

$$\Delta^f(\boldsymbol{x}) = \sum_{k=0}^{\infty} (-1)^k \frac{\Gamma(f+1)}{k!\Gamma(f-k+1)} x_{i+k}, \ i = 1, \ldots, n, \tag{15}$$

where $\Gamma(\cdot)$ denotes the gamma function. It is easy to verify that the summation in (15) is convergent. If f is a nonnegative integer, then the infinite sum defined in (15) reduces to a finite sum, i.e.,

$$\Delta^f(\boldsymbol{x}) = \sum_{k=0}^{f}(-1)^k\frac{\Gamma(f+1)}{k!\Gamma(f-k+1)}x_{i+k}, \ i = 1,\ldots,n, \tag{16}$$

and this operator generalizes the one defined in (11). The red-times curve in Fig. 3 shows the co-sparsity of the 1000 spikes using FOD matrix as the analysis dictionary. Based on FOD matrix, the MFOD matrix can be constructed as

$$\boldsymbol{\Omega}_{\text{MFOD}} \triangleq \frac{1}{\sqrt{q}}[\boldsymbol{D}^{f_0}, \boldsymbol{D}^{f_1}, \ldots, \boldsymbol{D}^{f_{q-1}}]^T, \tag{17}$$

with its order set

$$\boldsymbol{\mathcal{F}} = \{f_0, f_1, \ldots, f_{q-1}\} \in \mathbb{R}^{q \times 1}. \tag{18}$$

Define

$$d \triangleq \max |f_i - f_j|, \quad f_i, f_j \in \boldsymbol{\mathcal{F}}, \quad i \neq j \tag{19}$$

as the maximum order distance. It presents a trade-off between the co-sparsity of the analysis coefficients and the mutual coherence of $\boldsymbol{\Omega}_{\text{MFOD}}$. A small d increases the co-sparsity of the analysis coefficients but leads to the condition that each $\boldsymbol{D}^i$ is highly coherent with the other order difference matrix, and vice versa. Our observations on different d suggest that $d \in [\frac{1}{4}, \frac{1}{2}]$ is a good compromise.

Remark 1: The redundant ratio ρ presents a trade-off between signal reconstruction accuracy and computational cost. A large ρ increases the reconstruction performance but consumes more computational resources. The work in [20] showed that $\rho \in [2, 4]$ is a good compromise.

IV. WEIGHTED ANALYSIS ℓ_1-MINIMIZATION

Having constructed the multiple fractional orders difference matrix $\boldsymbol{\Omega}$ as the analysis dictionary, we herein propose a weighted analysis ℓ_1-minimization (WALM) method to reconstruct the original neural spike $\boldsymbol{x}$. Considering the CS measurement model $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{e}$, we assume that the components of $\boldsymbol{e}$ are independent and identically distributed (i.i.d.) Gaussian variables with unknown variance σ_e^2, and the entry z_i of $\boldsymbol{z} = \boldsymbol{\Omega}\boldsymbol{x}$ is independent and has a Laplacian distribution with standard deviation σ_i, i.e.,

$$p(z_i) = \frac{1}{\sqrt{2}\sigma_i}\exp\left(-\frac{\sqrt{2}\|z_i\|_1}{\sigma_i}\right), \quad i = 1,\ldots,l. \tag{20}$$

Note that reconstructing $\boldsymbol{x}$ and $\boldsymbol{z}$ are identical. Therefore, we first infer $\boldsymbol{z}$ from $\boldsymbol{y}$ by maximizing the conditional probability distribution $p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\Phi})$, which can be expressed by Bayes's rule as

$$p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\Phi}) \propto p(\boldsymbol{y}|\boldsymbol{\Phi}, \boldsymbol{z})p(\boldsymbol{z}). \tag{21}$$
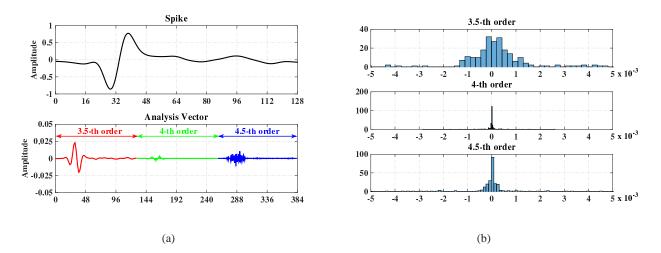
Fig. 4. (a) From top to bottom, the original neural spike and the corresponding analysis coefficients of 3.5-th, 4-th, and 4.5-th order difference. (b) From top to bottom, the histograms of analysis coefficients of the 3.5-th, 4-th, and 4.5-th order, respectively.

Because the noise e is assumed to be Gaussian, the likelihood function is given by

$$p(\boldsymbol{y}|\boldsymbol{\Phi},\boldsymbol{z}) \propto \exp\left(-\frac{\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\|_2^2}{2\sigma_e^2}\right). \tag{22}$$

Hence, maximizing the posterior distribution $p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{\Phi})$ leads to

$$\begin{aligned}\hat{\boldsymbol{z}}_{\text{MAP}} &= \arg\max_{\boldsymbol{z}} p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{\Phi}) \\ &= \arg\max_{\boldsymbol{z}}\left(\log p(\boldsymbol{y}|\boldsymbol{\Phi},\boldsymbol{z}) + \sum_i \log p(z_i)\right).\end{aligned} \tag{23}$$

By substituting $\boldsymbol{z} = \boldsymbol{\Omega}\boldsymbol{x}$ into (23), we obtain

$$\hat{\boldsymbol{x}}_{\text{MAP}} = \arg\min_{\boldsymbol{x}}\left(\frac{\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\|_2^2}{2\sigma_e^2} + \sum_i \frac{\sqrt{2}\|\boldsymbol{\Omega}_i\boldsymbol{x}\|_1}{\sigma_i}\right), \tag{24}$$

where $\boldsymbol{\Omega}_i$ denotes the ith row of $\boldsymbol{\Omega}$, $i \in \{1,\ldots,p\}$. The problem in (24) is equivalent to

$$\hat{\boldsymbol{x}}_{\text{MAP}} = \arg\min_{\boldsymbol{x}}\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \lambda\|\text{diag}(\boldsymbol{w})\boldsymbol{\Omega}\boldsymbol{x}\|_1, \tag{25}$$

where $\boldsymbol{w} = [\frac{1}{\sigma_1},\frac{1}{\sigma_2},\ldots,\frac{1}{\sigma_l}]$ and λ denotes a tuning parameter. Hence, the ℓ_1-minimization in (10) can be interpreted as the MAP estimation under the hypothesis that all σ_i are equal.

However, for our multiple fractional orders analysis matrix $\boldsymbol{\Omega}$, the hypothesis that the entries of $\boldsymbol{z}$ have equal standard deviations does not reflect this fact. To illustrate this argument, Fig. 4(a) plots the analysis coefficients of the 3.5-th, 4-th, and 4.5-th order difference for a typical neural spike[2]. At the same time,

[2]The spike was randomly chosen from the Leicester Easy2 dataset [30].

the corresponding histograms of the analysis coefficients in the three orders are simultaneously reported in Fig. 4(b).

Clearly, the standard deviations of analysis coefficients of distinct orders are not identical. To cope with this issue, we divide the standard deviation vector $\boldsymbol{w}$ into multiple groups to incorporate the aforementioned multiple orders prior. Suppose $\boldsymbol{\Omega}$ is constructed of q difference matrices with fractional orders, then $\boldsymbol{w}$ can be partitioned into q groups, i.e.,

$$\boldsymbol{w} = \left[\boldsymbol{w}_{G_0}^T, \boldsymbol{w}_{G_1}^T, \ldots, \boldsymbol{w}_{G_{q-1}}^T \right]^T, \tag{26}$$

where $\boldsymbol{w}_{G_0}^T, \boldsymbol{w}_{G_1}^T, \ldots, \boldsymbol{w}_{G_{q-1}}^T$ represent the standard deviations corresponding to q orders, respectively. Note that all σ_i of the same group are equal. As the variance of analysis coefficients tends to decrease first and increase across orders, we propose to model the variance across orders with quadratic functions as

$$\sigma_{f_i}^2 = c_i 2^{-2a_i f_i^2 - 2b_i f_i}, \quad i = 0, \ldots, q-1, \tag{27}$$

where a_i, b_i and c_i are the model parameters, and f_i is the difference order. In this model, the σ_i are made equal for all coefficients within an order, and $\sigma_{f_i}^2$ refers to the variance of the analysis coefficients at order f_i. Therefore, we have

$$\boldsymbol{w}_{G_i} = \sigma_{f_i} = \frac{2^{a_i f_i^2 + b_i f_i}}{\sqrt{c_i}}, \quad i = 0, \ldots, q-1. \tag{28}$$

As the entries of $\boldsymbol{w}$ only depend on the value of $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$, problem (25) can be solved after these parameters are calculated. This leads us to propose a training stage to estimate these values. The first part predicts the standard deviations σ_{f_i} using maximum likelihood estimation. Once the variances are estimated, simple quadratic regression can be employed to solve for a_i, b_i, and c_i in the following equation, derived from (27),

$$\log_2 \sigma_{f_i}^2 = \log_2 c_i - 2a_i f_i^2 - 2b_i f_i. \tag{29}$$

An example of the fitted regression curve is shown in Fig. 5.

After building $\boldsymbol{w}$, the problem (25) can be easily solved using ℓ_1-minimization algorithms. The complete WALM algorithm can be outlined as Algorithm 1.

Remark 2: The proposed WALM is a nearly signal independent approach. Although it requires a training step to estimate the regression parameters, the amount of data needed is much less than that of signal dependent approaches such as [12] and [13]. Therefore, WALM can significantly reduce the data storage and computational resource.
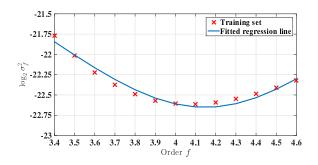
Fig. 5. Scatter plot of fractional order f versus variance $\log_2 \sigma_f^2$ and fitted regression curve. For each order, the variance was averaged over the training dataset that contains 100 spikes randomly chosen from Leicester Easy2 dataset.

Remark 3: In contrast to the canonical AL1 method, the main advantage of WALM is the incorporation of the multiple orders prior in analysis coefficients, including the positions of nonzero coefficients and its standard deviations between neighboring difference orders, which will allow the number of measurements to be significantly reduced without leading to ambiguity.

Algorithm 1 Weighted Analysis ℓ_1-Minimization

Input: $\boldsymbol{y}, \boldsymbol{\Phi}, q$

1: **for** $i = 1, \ldots, q-1$ **do**

2:      Construct $\boldsymbol{D}^{f_i}$ by using (15)

3: **end for**

4: Construct $\boldsymbol{\Omega}$ by using (17)

5: **for** $i = 1, \ldots, q-1$ **do**

6:      Estimate a_i, b_i, and c_i by using (29)

7:      Construct $\boldsymbol{w}_{G_i}^T$ by using (28)

8: **end for**

9: Construct $\boldsymbol{w}$ by using (26)

10: Solve $\hat{\boldsymbol{x}} = \arg \min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \lambda \|\text{diag}(\boldsymbol{w})\boldsymbol{\Omega}\boldsymbol{x}\|_1$

Output: recovered signal $\hat{\boldsymbol{x}}$

V. EXPERIMENT VALIDATION

In this section, we examine the performance of the proposed WALM method against state-of-the-art compressed sensing schemes for implant neural recording.

A. Experimental Setup

We use the Leicester neural signal database [30], which contains 20 synthesized datasets. Each dataset contains spikes from three different neurons with different noise levels. The datasets are categorized by the spike sorting difficulty levels, such as Leicester Difficult1, Difficult2, Easy1, and Easy2. We take 128 samples around each spike to form the signal frame. To simplify the comparison, we retain the signal containing only one spike. A Random i.i.d Bernoulli matrix is chosen as the sensing matrix because it guarantees excellent reconstruction quality and implementation efficiency [31]. All signals are compressed and reconstructed for 20 times, using a different sensing matrix in each trial. The results are then averaged across all trials. To measure the reconstruction quality, we employ the percentage root-mean-square difference (PRD) to quantify the error percentage between the original x and the reconstructed $\hat{x}$:

$$\text{PRD} = \frac{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2}{\|\boldsymbol{x}\|_2} \times 100\%. \tag{30}$$

For physiological signal reconstruction, Zigel *et al.* [32] classified the different values of PRD based on the signal quality perceived by specialists. In this work, PRD value below 5% is regarded as "good" reconstruction quality.

The following state-of-the-art CS algorithms have been chosen for performance comparison.

1) Basis Pursuit De-noising (BPDN) described in (6). The orthonormal basis of Daubechies-4 wavelet was used as the sparsity dictionary. For the BPDN implementation, we used the solvers *SolveBP* from the SparseLab toolbox [33].

2) Block Sparse Bayesian Learning (BSBL) proposed by Zhang *et al.* [14]. We used the solver *BSBL-BO* for BSBL implementation.

3) Analysis ℓ_1-minimization (AL1) algorithm described in (10). For the AL1 implementation, we used the *CVX* toolbox [34] from Stanford University.

4) Signal Dependent Neural Compressed Sensing (SDNCS) method proposed in [12]. It used a sparse representation dictionary learned from data. For this method, each dataset was divided into training section and test section, composed of 20% and 80% of the dataset. The training section was used to construct the sparse representation dictionary, whereas the test section was used to evaluate its performance[3].

---

[3]The implementation of SDNCS can be downloaded from http://etienne.ece.jhu.edu/projects/neural_cs/CS_code.rar
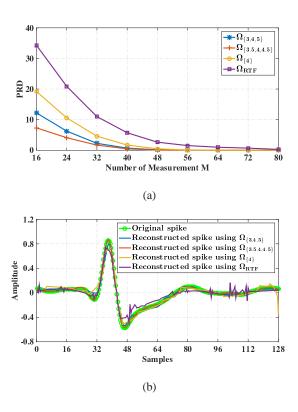
(a)

(b)

Fig. 6. (a) Averaged PRDs over all spikes from Leicester Easy1 dataset versus the different number of measurements M for $\mathbf{\Omega_{\{3,4,5\}}}$, $\mathbf{\Omega_{\{3.5,4,4.5\}}}$, $\mathbf{\Omega_{\{4\}}}$, and $\mathbf{\Omega_{RTF}}$, respectively. (b) Original spike and reconstructed spikes using $\mathbf{\Omega_{\{3,4,5\}}}$, $\mathbf{\Omega_{\{3.5,4,4.5\}}}$, $\mathbf{\Omega_{\{4\}}}$, and $\mathbf{\Omega_{RTF}}$, respectively.

B. The Advantage of MFOD Matrix

To evaluate the effectiveness of the proposed MFOD matrix, we compared it with IOD matrix, MIOD matrix, and the random tight frame (RTF) proposed in [20]. In this experiment, we constructed IOD matrix $\mathbf{\Omega_{\{4\}}}$ as (11), MIOD matrix $\mathbf{\Omega_{\{3,4,5\}}}$ as (12), and MFOD matrix $\mathbf{\Omega_{\{3.5,4,4.5\}}}$ as (17). The Leicester Easy1 dataset was chosen for evaluation. For all the four dictionaries, the AL1 algorithm was used to reconstruct the spikes. The average PRDs over all spikes for the four dictionaries and a spike reconstruction example are shown in Fig. 6(a) and 6(b), respectively. Among the four different dictionaries, the $\mathbf{\Omega_{RTF}}$ has the worst performance. It is mainly because the $\mathbf{\Omega_{RTF}}$ is a general analysis dictionary and does not exploit any statistical information of neural spikes. Furthermore, we can note that $\mathbf{\Omega_{\{3.5,4,4.5\}}}$ and $\mathbf{\Omega_{\{3,4,5\}}}$ outperform $\mathbf{\Omega_{\{4\}}}$ due to their redundancy. In addition, $\mathbf{\Omega_{\{3.5,4,4.5\}}}$ has better reconstruction accuracy than $\mathbf{\Omega_{\{3,4,5\}}}$, especially when the number of measurements is very small. It is mainly because the analysis coefficients with $\mathbf{\Omega_{\{3.5,4,4.5\}}}$ will be sparser than that of $\mathbf{\Omega_{\{3,4,5\}}}$, and the high sparsity reduces the number of measurements for signal reconstruction.

TABLE I

PROBABILITIES OF RECONSTRUCTION FOR "GOOD" QUALITY UNDER THE DIFFERENT NUMBER OF MEASUREMENTS M

(%).

M	Easy1 dataset					Difficult1 dataset				
	BPDN	BSBL	AL1	SDNCS	WALM	BPDN	BSBL	AL1	SDNCS	WALM
16	0	0	30.8	93.3	92.5	0	0	28.8	93.1	92.2
24	0	2.1	54.2	95.6	95.1	0	1.3	52.1	94.9	94.3
32	0	52.9	83.8	97.9	97.7	0	48.5	82.4	96.2	96.3
40	9.3	60.8	92.1	100	100	6.4	58.3	90.2	98.8	98.9
48	42.9	68.4	94.3	100	100	33.4	66.7	93.0	100	100
56	52.9	75.8	98.8	100	100	48.2	74.3	98.7	100	100
64	68.2	80.8	100	100	100	63.3	78.4	100	100	100
72	84.5	93.2	100	100	100	82.8	92.0	100	100	100
80	92.7	98.3	100	100	100	91.2	97.9	100	100	100

C. Average PRD and the Probability of "Good" Reconstruction

Then we evaluate the performance of the proposed WALM algorithm versus the number of measurements. The $\Omega_{\{3.5,4,4.5\}}$ was used as the analysis dictionary for both AL1 and WALM. For the Leicester neural signal database, the Easy1 and Difficult1 datasets are chosen for evaluation. The experimental results are shown in Fig. 7, where each point indicates the average PRD of all spikes at a specified number of measurements. At the same time, Table I reports the probability of "good" reconstruction quality in different situations. First of all, we can observe that analysis model based algorithms outperform synthesis model based ones in terms of both averaged PRD and the probability of "good" reconstruction quality. Moreover, due to the incorporation of the multiple orders prior in analysis coefficients, the WALM algorithm performs better than the canonical AL1 algorithm, especially when the number of measurements is very small. The WALM algorithm has the averaged PRD less than 5% for all numbers of measurements, and it achieves more than 92% of "good" reconstruction quality with $M = 16$. As a comparison, BPDN, BSBL, and AL1 cannot recover so many spikes both in Easy1 and Difficult1 datasets with "good" reconstruction quality under this condition. Both SDNCS and WALM algorithms show "good" reconstruction quality (SDNCS even slightly outperforms WALM when $M = 16$), while the proposed WALM algorithm simplifies sparse dictionary learning with much fewer computational resources, which is preferred for practical neural recording experiments.

To observe the PRD variance across individual datasets, Fig. 8 shows the box plots for these algorithms
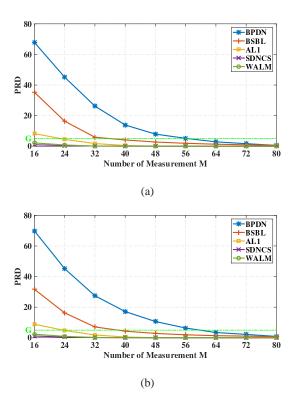
(a)

(b)

Fig. 7. PRD averaged over all spikes from (a) Easy1 dataset, (b) Difficult1 dataset, versus the different number of measurements M for BPDN, BSBL, AL1, SDNCS, and WALM, respectively. The green dash-dotted line denotes the "good" PRD bound at 5%.

when the number of measurements is $M = 32$. On each box, the central mark indicates the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points. Obviously, both for Easy1 and Difficult1 datasets, the WALM algorithm adopted the multiple fractional orders analysis dictionary outperforms the other algorithms. Once more, SDNCS has similar performance as WALM. Moreover, although the PRD variances of the two datasets are similar, the number of outliers of Difficult1 dataset is more than that of Easy1 dataset.

D. Performance of Classification using Reconstructed Spikes

To further illustrate the performance of WALM, we carried out a spike classification experiment using reconstructed spikes. The Leicester Easy1 and Difficult1 datasets were chosen for evaluation. Firstly, all spikes were compressed with $M = 16$ and reconstructed using the five algorithms. Then Principal component analysis (PCA) and [35] wavelet decomposition [30] methods were used to extract features from reconstructed spikes in Easy1 and Difficult1 datasets, respectively. Finally, the first 10 features of
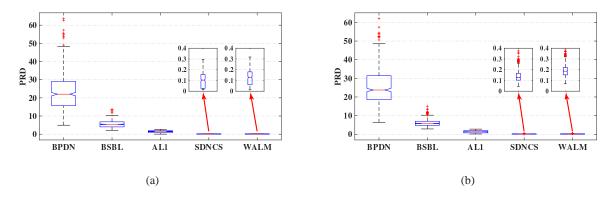
Fig. 8. Box plots for all spikes from (a) Easy1 dataset, (b) Difficult1 dataset, for BPDN, BSBL, AL1, SDNCS, and WALM, respectively, when the number of measurements is $M = 32$.

each spike were used for classification by superparamagnetic clustering (SPC) [30] algorithm.

Fig. 9 and Fig. 10 show the three-dimensional (3D) projections of the first three features of reconstructed spikes using the five algorithms in the Easy1 and Difficult1 datasets, respectively. In all cases, the classification was done automatically with SPC and is represented in different colors. For the Easy1 dataset, we observe that it is possible to identify the three clusters clearly using the spikes reconstructed by all five algorithms. Furthermore, the features of spikes reconstructed by WALM and SDNCS are more accurate than that of the other algorithms, implying that WALM and SDNCS have better reconstruction performance. For the Difficult1 dataset, we observe that only the spikes reconstructed by WALM and SDNCS can be clearly classified, whereas the classification got failed using the spikes reconstructed by AL1, BPDN, and BSBL.

Spike classification accuracy was also used as a performance metric, calculated as a percentage of the total number of spikes correctly classified. The classification results were compared with the ground truth labels contained in the datasets. The spike classification results for Easy1 and Difficult1 are shown in Fig. 11. We observe that WALM and SDNCS outperform the other three algorithms for spike classification. Even the number of measurements is only 16, WALM can achieve above 99% and above 92% classification accuracy for Easy1 and Difficult1 datasets, respectively. Moreover, WALM provides a reliable solution which yields better reconstruction and classification performance with much fewer computational resource and pre-acquired data than SDNCS.
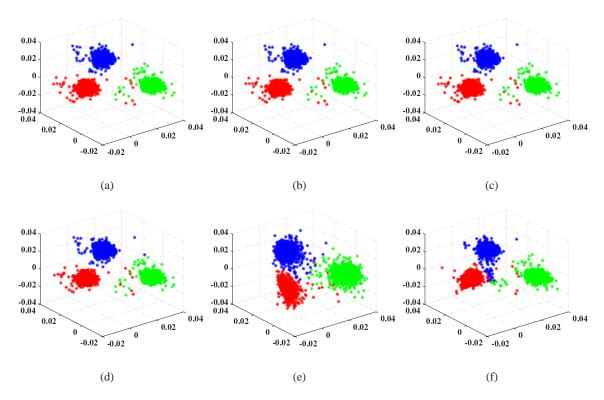
Fig. 9. Scatter plot of the first three features of (a) original spikes in Easy1 dataset, (b) spikes reconstructed by WALM, (c) spikes reconstructed by SDNCS, (d) spikes reconstructed by AL1, (e) spikes reconstructed by BPDN, and (f) spikes reconstructed by BSBL, respectively. The features were extracted using PCA. Points are colored according to the cluster to which they are assigned. The number of measurements is $M = 16$.

VI. CONCLUSION

This paper proposed a novel compressed sensing method for implantable neural recording. The proposed method enforces sparsity of neural spikes not by the traditional synthesis model, but by the analysis model with a multiple fractional orders difference matrix as its analysis dictionary. Therefore, the pre-acquired data and computational resource for dictionary learning will be significantly reduced. Besides, by exploiting statistical priors of the analysis coefficients among difference orders, a weighted analysis ℓ_1-minimization algorithm was proposed to reconstruct neural spikes. Experimental results proved the efficacy of the proposed method for neural signal reconstruction.

REFERENCES

[1] I. H. Stevenson and K. P. Kording, "How advances in neural recording affect data analysis," *Nature neuroscience*, vol. 14, no. 2, pp. 139–142, 2011.
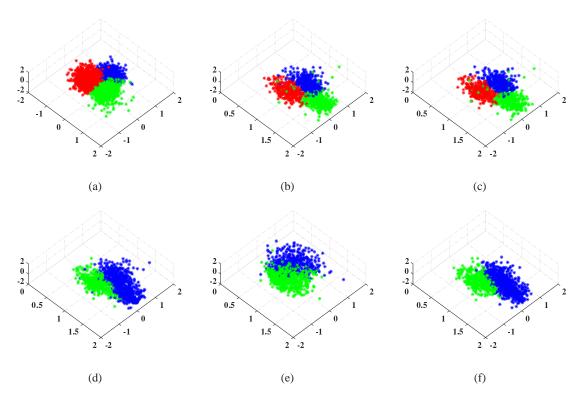
Fig. 10. Scatter plot of the first three features of (a) original spikes in Difficult1 dataset, (b) spikes reconstructed by WALM, (c) spikes reconstructed by SDNCS, (d) spikes reconstructed by AL1, (e) spikes reconstructed by BPDN, and (f) spikes reconstructed by BSBL, respectively. The features were extracted using wavelet decomposition. Points are colored according to the cluster to which they are assigned. The number of measurements is $M = 16$.

[2] A. Berényi, Z. Somogyvari, A. J. Nagy, L. Roux, J. D. Long, S. Fujisawa, E. Stark, A. Leonardo, T. D. Harris, and G. Buzsáki, "Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals," *Journal of neurophysiology*, vol. 111, no. 5, pp. 1132–1149, 2014.

[3] D. A. Schwarz, M. A. Lebedev, T. L. Hanson, D. F. Dimitrov, G. Lehew, J. Meloy, S. Rajangam, V. Subramanian, P. J. Ifft, Z. Li *et al.*, "Chronic, wireless recordings of large-scale brain activity in freely moving rhesus monkeys," *Nature methods*, vol. 11, no. 6, pp. 670–676, 2014.

[4] M. Yin, D. A. Borton, J. Komar, N. Agha, Y. Lu, H. Li, J. Laurens, Y. Lang, Q. Li, C. Bull *et al.*, "Wireless neurosensor for full-spectrum electrophysiology recordings during free behavior," *Neuron*, vol. 84, no. 6, pp. 1170–1182, 2014.

[5] A. Rodriguez-Perez, J. Ruiz-Amaya, M. Delgado-Restituto, and A. Rodriguez-Vazquez, "A low-power programmable neural spike detection channel with embedded calibration and data compression," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 6, no. 2, pp. 87–100, 2012.

[6] M. S. Chae, Z. Yang, M. R. Yuce, L. Hoang, and W. Liu, "A 128-channel 6 mw wireless neural recording ic with spike feature extraction and uwb transmitter," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 17, no. 4, pp. 312–321, 2009.

[7] B. Gosselin and M. Sawan, "An ultra low-power cmos automatic action potential detector," *Neural Systems and*
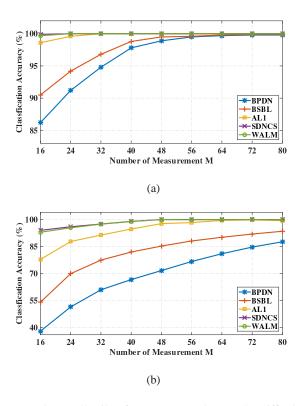
(a)

(b)

Fig. 11. Classification accuracy averaged over all spikes from (a) Easy1 dataset, (b) Difficult1 dataset, versus different numbers of measurements M for BPDN, BSBL, AL1, SDNCS, and WALM, respectively.

Rehabilitation Engineering, IEEE Transactions on, vol. 17, no. 4, pp. 346–353, 2009.

[8] B. Gosselin, A. E. Ayoub, J.-F. Roy, M. Sawan, F. Lepore, A. Chaudhuri, and D. Guitton, "A mixed-signal multichip neural recording interface with bandwidth reduction," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 3, no. 3, pp. 129–141, 2009.

[9] K. G. Oweiss, A. Mason, Y. Suhail, A. M. Kamboh, and K. E. Thomson, "A scalable wavelet transform vlsi architecture for real-time signal processing in high-density intra-cortical implants," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 54, no. 6, pp. 1266–1278, 2007.

[10] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[11] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, 2006.

[12] J. Zhang, Y. Suo, S. Mitra, S. P. Chin, S. Hsiao, R. F. Yazicioglu, T. D. Tran, and R. Etienne-Cummings, "An efficient and compact compressed sensing microsystem for implantable neural recordings," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 8, no. 4, pp. 485–496, 2014.

[13] Y. Suo, J. Zhang, T. Xiong, P. S. Chin, R. Etienne-Cummings, and T. D. Tran, "Energy-efficient multi-mode compressed sensing system for implantable neural recordings," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 8, no. 5, pp. 648–659, 2014.

[14] Z. Zhang and B. Rao, "Extension of sbl algorithms for the recovery of block sparse signals with intra-block correlation,"

Signal Processing, IEEE Transactions on, vol. 61, no. 8, pp. 2009–2015, 2013.

[15] C. Bulach, U. Bihr, and M. Ortmanns, "Evaluation study of compressed sensing for neural spike recordings," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 3507–3510.

[16] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM review*, vol. 51, no. 1, pp. 34–81, 2009.

[17] S. Becker, J. Bobin, and E. J. Candès, "Nesta: a fast and accurate first-order method for sparse recovery," *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 1–39, 2011.

[18] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589–592, 2008.

[19] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse problems*, vol. 23, no. 3, p. 947, 2007.

[20] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[21] 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.

[22] 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.

[23] Y. Liu, M. De Vos, and S. Van Huffel, "Compressed sensing of multichannel eeg signals: The simultaneous cosparsity and low-rank optimization," *Biomedical Engineering, IEEE Transactions on*, vol. 62, no. 8, pp. 2055–2061, 2015.

[24] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical imaging and vision*, vol. 20, no. 1-2, pp. 89–97, 2004.

[25] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Knowledge and information Systems*, vol. 3, no. 3, pp. 263–286, 2001.

[26] Y. Jiang, T. Lan, and D. Zhang, "A new representation and similarity measure of time series on data mining," in *Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on*. IEEE, 2009, pp. 1–5.

[27] J. Zhou, G. Ye, and D. Yu, "A new method for piecewise linear representation of time series data," *Physics Procedia*, vol. 25, pp. 1097–1103, 2012.

[28] C. Yan, J. Fang, L. Wu, and S. Ma, "An approach of time series piecewise linear representation based on local maximum, minimum and extremum," *Journal of Information & Computational Science*, vol. 10, no. 9, pp. 2747–2756, 2013.

[29] M. Et and R. Çolak, "On some generalized difference sequence spaces," *Soochow Journal of Mathematics*, vol. 21, no. 4, pp. 377–386, 1995.

[30] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering," *Neural computation*, vol. 16, no. 8, pp. 1661–1687, 2004.

[31] F. Chen, A. P. Chandrakasan, and V. M. Stojanović, "Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors," *Solid-State Circuits, IEEE Journal of*, vol. 47, no. 3, pp. 744–756, 2012.

[32] Y. Zigel, A. Cohen, and A. Katz, "The weighted diagnostic distortion (wdd) measure for ecg signal compression," *Biomedical Engineering, IEEE Transactions on*, vol. 47, no. 11, pp. 1422–1430, 2000.

[33] D. Donoho, V. Stodden, and Y. Tsaig, "Sparselab. 2007," *See http://sparselab. stanford. edu..(Accessed January 23, 2014.)*, 2008.

[34] M. Grant, S. Boyd, and Y. Ye, "Cvx: Matlab software for disciplined convex programming," 2008.

[35] I. Jolliffe, *Principal component analysis*.   Wiley Online Library, 2002.