Efficient Low-Rank Spectrotemporal Decomposition using ADMM

Gabriel Schamberg, Demba Ba, Mark Wagner, and Todd Coleman

***Abstract*— Classical time-evolving spectral analysis techniques utilize a sliding window approach that fails to exploit overarching spectrotemporal structures that are known to occur in many real-world signals. In particular, many biological signals have the distinct quality of having few defining spectral characteristics. We propose an algorithm for efficiently estimating a low-rank spectrotemporal decomposition. While existing approaches yield such representations by penalizing a group norm of successive differences, our insight is that a penalty that promotes low-rank yields more flexible representations and an efficient distributed implementation. We demonstrate on simulated time series data and human electroencephalogram (EEG) recordings that this low-rank spectrotemporal decomposition can provide a spectral representation of time series that highlights salient features and reduces the effects of noise.**

I. Introduction

With the onset of advanced biotechnologies such as brain computer interfaces (BCIs), there is an increasing need for accurate, concise, and efficient spectrotemporal representations of biological time series. Non-invasive EEG recordings are a popular choice for such applications, though they present issues with regard to noise and spectral resolution [1]. To accommodate these issues, we draw from the growing field of sparse recovery and propose a method for estimating a low-rank spectrotemporal decomposition of time series in order to extract the dominant spectral features of EEG recordings, and more generally any suitable biological time series. One of the most popular examples of this approach is illustrated by the Netflix problem [2], where recommender systems that use low-rank matrices perform well given the tendency of users to have few factors influencing their movie preferences. Similarly, many biological signals may have few active frequency bands or a select number of time windows in which we would expect to see spectral activity. In the development of the low-rank spectrotemporal decomposition we consider the method of spectrotemporal pursuit [3], which obtains a time-evolving spectral representation that is sparse in frequency and smooth across time. As a byproduct of the assumptions made by its penalty function, the spectrotemporal pursuit estimate yields a low-rank decomposition. By directly imposing a low-rank prior on the matrix of first differences of the coefficients, the proposed algorithm makes fewer inherent assumptions on the specific nature of the sparse structure of a signal and in turn enables a solution with lower complexity and greater flexibility.

II. Problem Formulation

We seek to find the set of time-evolving frequency coefficients $\mathbf{x} \in \mathbb{R}^{K\times N}$ that best represent an observed time series $\mathbf{y} \in \mathbb{R}^T$, where K is the desired number of frequency bins, T is the number of samples in the time series, and N is the number of length M time windows in $\mathbf{y}$. Without loss of generality, we consider time series of length $T = N \cdot M$. Using this notation, we assume that the n^{th} window of the observed data is given by:

$$\mathbf{y}_n = \mathbf{F}\mathbf{x}_n + \mathbf{v}_n$$

where $\mathbf{y}_n \in \mathbb{R}^M$, $\mathbf{x}_n \in \mathbb{R}^K$ is the collection of frequency coefficients for the n^{th} time window, and $\mathbf{v}_n \sim \mathcal{N}(0, \sigma^2) \in \mathbb{R}^M$ is the measurement noise. Lastly, $\mathbf{F} \in \mathbb{R}^{M\times K}$ is the real Fourier basis with entries $\mathbf{F}_{m,k} = \cos(2\pi m \frac{k}{K})$ and $\mathbf{F}_{m,k+\frac{K}{2}} = \sin(2\pi m \frac{k}{K})$, for $k = 0, 1, \ldots, \frac{K}{2} - 1$ and $m = 0, 1, \ldots, M-1$. In circumstances when we allow $K > M$, we are oversampling the Fourier domain and $\mathbf{F}$ as defined above is not a proper basis. Using these definitions, we formulate the problem as the following ℓ_2-norm minimization:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \sum_{n=1}^{N} \|\mathbf{y}_n - \mathbf{F}\mathbf{x}_n\|_2^2$$

This problem can be solved in closed form, yielding a discrete time Fourier transform of $\mathbf{y}_n$. While this formulation does not provide any direct advantage over traditional spectral analysis techniques from a performance standpoint, it enables us to impose a penalty function on estimates of $\mathbf{x}$ that do not conform to our prior knowledge of the signal.

A. Nuclear Norm Penalty

Our motivation to consider a time-frequency decomposition via a joint minimization over a classical Fourier measurement model and a penalty function comes from a method termed spectrotemporal pursuit [3], wherein the penalty function is imposed over the differences in frequency coefficients in adjacent time windows. Specifically, by defining the matrix $\mathbf{w}$ with columns $\mathbf{w}_n \triangleq \mathbf{x}_n - \mathbf{x}_{n-1}$, [3] demonstrates that the following penalty function ensures sparsity across frequencies and continuity across time:

$$\sum_{k=1}^{K} \left(\sum_{n=1}^{N} \mathbf{w}_{k,n}^2 \right)^{\frac{1}{2}}$$

Notice that the inner sum can be viewed as a ℓ_2-norm on the change in frequency coefficients over time in a given frequency band (enforcing continuity) and the outer sum can be viewed as a ℓ_1-norm across frequencies. As a result, many rows of $\mathbf{w}$ will be exceedingly close to zero. Thus, many of the rows of $\mathbf{w}$ are linearly dependent and $\mathbf{w}$ is necessarily low rank.

Here we consider an alternative penalty function that does not make any explicit assumptions regarding the spectral or temporal structure of the signal, but instead isolates the aspect of spectrotemporal pursuit which ensures $\mathbf{w}$ is low rank. It is well known that direct rank-minimization is an NP-hard problem. We instead consider penalizing by the *nuclear norm* of $\mathbf{w}$, which is given by the sum of the singular values of $\mathbf{w}$. It has been shown that minimizing the nuclear norm yields the smallest convex envelope of the rank-minimization problem and can recover the exact minimum rank solution under certain assumptions [4].

From a Bayesian inference perspective, this equates to assuming a prior probability density function (PDF) on the random matrix $\mathbf{w}$ given by:

$$f_\beta(\mathbf{w}) = \frac{1}{Z_\beta} e^{-\beta\|\mathbf{w}\|_*} \tag{1}$$

where $\|\mathbf{w}\|_*$ is the nuclear norm of $\mathbf{w}$, $\beta > 0$ is a parameter to control the extent to which the rank of $\mathbf{w}$ is minimized, and Z_β is an appropriate normalizing constant. It is clear that $p_\beta(\mathbf{w}) \geq 0$

$\forall \mathbf{w} \in \mathbb{R}^{K \times N}$. To show the existence of $Z_\beta < \infty$, we consider the following bound:

$$Z_\beta = \int_{\mathbb{R}^{K \times N}} e^{-\beta \|\mathbf{X}\|_*} d\mathbf{X} \leq \int_{\mathbb{R}^{K \times N}} e^{-\beta \|\mathbf{X}\|_F} d\mathbf{X} \tag{2}$$

where $\|\mathbf{X}\|_F$ is the Frobenius norm of $\mathbf{X}$. The proof of this inequality uses the interpretation of the Frobenius norm as the ℓ_2-norm of the singular values and proceeds to follow the proof of the well known fact that for $\mathbf{v} \in \mathbb{R}^n$, $\|\mathbf{v}\|_2 \leq \|\mathbf{v}\|_1$. Next, we note that the latter integral in (2) can be equivalently rewritten as an integral over vectors in $\mathbb{R}^{K \cdot N}$ using the ℓ_2-norm in place of the Frobenius norm. This integral evaluates to the normalizing constant of the prior distribution used in [3] for the special case of $K = 1$. Thus, we can conclude that function given by (1) is a valid PDF.

We consider finding the maximum a posteriori (MAP) estimate of the frequency coefficients $\mathbf{x}$ given the time series observation $\mathbf{y}$:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} - \log p(\mathbf{y} \mid \mathbf{x}) - \log p(\mathbf{x}) \tag{3}$$

Defining the prior distribution on $\mathbf{x}$ to be $p(\mathbf{x}) = f_\beta(\mathbf{x}\mathbf{A})$, where $\mathbf{A}$ is the identity matrix with the superdiagonal set to -1, we get that $p(\mathbf{x}) = f_\beta(\mathbf{w})$. Substituting our Gaussian measurement model for $-\log p(\mathbf{x} \mid \mathbf{y})$ yields our target optimization problem:

$$\begin{aligned} \hat{\mathbf{x}} = \underset{\mathbf{x},\mathbf{w}}{\operatorname{argmin}} \quad & \left(\sum_{n=1}^{N} \|\mathbf{y}_n - \mathbf{F}\mathbf{x}_n\|_2^2 \right) + \beta \|\mathbf{w}\|_* \\ s.t. \quad & \mathbf{w}_n = \mathbf{x}_n - \mathbf{x}_{n-1} \end{aligned} \tag{4}$$

Below we detail an algorithm for efficiently computing $\hat{\mathbf{x}}$ given an observed time series $\mathbf{y}$ using the nuclear norm penalty.

III. MAP Estimation with ADMM

As a means of efficiently evaluating the MAP estimate of the time-evolving Fourier coefficients $\mathbf{x} \in \mathbb{R}^{K \times N}$, we propose an approach for the following general class of problems involving the estimation of a latent variable $\mathbf{x}$, given observations $\mathbf{y}$:

$$\begin{aligned} \hat{\mathbf{x}} = \underset{\mathbf{x},\mathbf{w}}{\operatorname{argmin}} \quad & \left(\sum_{n=1}^{N} l_n(\mathbf{x}_n) \right) + \beta \phi(\mathbf{w}) \\ s.t. \quad & \mathbf{w}_n = \mathbf{x}_n - \mathbf{x}_{n-1} \end{aligned} \tag{5}$$

where $l_n(\mathbf{x}_n) \triangleq -\log p(\mathbf{y}_n | \mathbf{x}_n)$ is the log-likelihood given by our measurement model, and $\phi(\mathbf{w})$ is a general penalty function that utilizes prior knowledge on the temporal dynamics of the latent variable.

We can see that the objective function in (5) has two terms acting on different variables. The constraints prevent these terms from being solved directly in a separable manner, but use of the Alternating Direction Method of Multipliers (ADMM) allows the problem to be decomposed into subproblems that can be considered independently and solved iteratively [5]. For reasons that will soon be clear, we introduce an auxiliary variable that will allow the latent coefficients $\mathbf{x}$ and the difference matrix $\mathbf{w}$ to be separated in the constraints:

$$\begin{aligned} \hat{\mathbf{x}} = \underset{\mathbf{x},\mathbf{w},\mathbf{z}}{\operatorname{argmin}} \quad & \left(\sum_{n=1}^{N} l_n(\mathbf{x}_n) \right) + \beta \phi(\mathbf{w}) \\ s.t. \quad & \mathbf{x}_n = \mathbf{z}_n \\ & \mathbf{w}_n = \mathbf{z}_n - \mathbf{z}_{n-1} \end{aligned}$$

Defining $\mathbf{c}_n \triangleq \mathbf{z}_n - \mathbf{z}_{n-1} - \mathbf{w}_n$ and $\mathbf{d}_n \triangleq \mathbf{z}_n - \mathbf{x}_n$, it is clear that the problem constraints become $\mathbf{c}_n = \mathbf{d}_n = 0$. Using this notation we introduce the augmented Lagrangian:

$$\begin{aligned} L_\rho(\mathbf{x}, \mathbf{z}, \mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\alpha}) = \beta \phi(\mathbf{w}) + \Bigg(& \sum_{n=1}^{N} l_n(\mathbf{x}_n) \\ & + \boldsymbol{\lambda}_n^T \mathbf{c}_n + \boldsymbol{\alpha}_n^T \mathbf{d}_n + \frac{\rho}{2} \|\mathbf{c}_n\|_2^2 + \frac{\rho}{2} \|\mathbf{d}_n\|_2^2 \Bigg) \end{aligned} \tag{6}$$

where $\boldsymbol{\lambda}_n$ and $\boldsymbol{\alpha}_n$ are the n^{th} columns of the Lagrange multipliers $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha} \in \mathbb{R}^{N \times K}$, respectively. Using standard ADMM techniques [5], an estimate of $\mathbf{x}$ can be obtained by iteratively updating each optimization variable independently:

$$\begin{aligned} \mathbf{x}^{(m+1)} &= \underset{\mathbf{x}}{\operatorname{argmin}}\ L_\rho \left(\mathbf{x}, \mathbf{z}^{(m)}, \mathbf{w}^{(m)}, \boldsymbol{\lambda}^{(m)}, \boldsymbol{\alpha}^{(m)} \right) \\ \mathbf{z}^{(m+1)} &= \underset{\mathbf{z}}{\operatorname{argmin}}\ L_\rho \left(\mathbf{x}^{(m+1)}, \mathbf{z}, \mathbf{w}^{(m)}, \boldsymbol{\lambda}^{(m)}, \boldsymbol{\alpha}^{(m)} \right) \\ \mathbf{w}^{(m+1)} &= \underset{\mathbf{w}}{\operatorname{argmin}}\ L_\rho \left(\mathbf{x}^{(m+1)}, \mathbf{z}^{(m+1)}, \mathbf{w}, \boldsymbol{\lambda}^{(m)}, \boldsymbol{\alpha}^{(m)} \right) \\ \boldsymbol{\lambda}_n^{(m+1)} &= \boldsymbol{\lambda}_n^{(m)} + \rho \left(\mathbf{z}_n^{(m+1)} - \mathbf{z}_{n-1}^{(m+1)} - \mathbf{w}_n^{(m+1)} \right) \\ \boldsymbol{\alpha}_n^{(m+1)} &= \boldsymbol{\alpha}_n^{(m)} + \rho \left(\mathbf{z}_n^{(m+1)} - \mathbf{x}_n^{(m+1)} \right) \end{aligned} \tag{7}$$

It follows that the low-rank spectrotemporal decomposition can be computed by iteratively solving five subproblems, two of which are trivial vector sums. We will see that the $\mathbf{x}$ and $\mathbf{w}$ updates involve minimizing over the measurement model and penalty function with an added proximal operator [6]. Noting that the linear Lagrange multiplier term and the ℓ_2-norm resulting from augmenting the Lagrangian can be combined into a single quadratic term, this quadratic term serves to ensure that the iterative updates are converging to a solution that preserves the relationship between $\mathbf{x}$ and $\mathbf{w}$ given by the constraint in the original problem.

While ADMM is known to require many iterations to converge to a precise solution, it achieves modest accuracy within a relatively small number of iterations [5]. In the following sections we derive solutions for each of the updates used in the low-rank spectrotemporal decomposition problem.

A. Measurement Model Update

We begin with the update of the estimate of $\mathbf{x}$ based on the measurement model:

$$\begin{aligned} \mathbf{x}^{(m+1)} = \underset{\mathbf{x}}{\operatorname{argmin}} & \sum_{n=1}^{N} l_n(\mathbf{x}_n) \\ & + \boldsymbol{\alpha}_n^{(m)^T} \left(\mathbf{z}_n^{(m)} - \mathbf{x}_n \right) + \frac{\rho}{2} \left\| \mathbf{z}_n^{(m)} - \mathbf{x}_n \right\|_2^2 \end{aligned} \tag{8}$$

noting that in the context of this update, $\boldsymbol{\alpha}_n^{(m)}$ and $\mathbf{z}_n^{(m)}$ are the values from the previous iteration and can therefore be treated as constants. In (8), the value of the auxiliary variable $\mathbf{z}$ is apparent. By separating $\mathbf{x}$ and $\mathbf{w}$ in the constraints, we have delegated consideration of the temporal structure of $\mathbf{x}$ to the $\mathbf{z}$ update. As a result, the $\mathbf{x}$ update can be separated into N completely separable problems:

$$\begin{aligned} \mathbf{x}_n^{(m+1)} = \underset{\mathbf{x}_n \in \mathbb{R}^K}{\operatorname{argmin}} & \ l_n(\mathbf{x}_n) \\ & + \boldsymbol{\alpha}_n^{(m)^T} \left(\mathbf{z}_n^{(m)} - \mathbf{x}_n \right) + \frac{\rho}{2} \left\| \mathbf{z}_n^{(m)} - \mathbf{x}_n \right\|_2^2 \end{aligned}$$

Plugging in the quadratic measurement model for our spectrotemporal decomposition, we can collapse the update to a single term by first expanding and then combining terms and completing the square:

$$\begin{aligned}\mathbf{x}_n^{(m+1)} &= \underset{\mathbf{x}_n}{\operatorname{argmin}} \|\mathbf{y}_n - \mathbf{F}\mathbf{x}_n\|_2^2 \\ &\quad + \boldsymbol{\alpha}_n^{(m)T}\left(\mathbf{z}_n^{(m)} - \mathbf{x}_n\right) + \frac{\rho}{2}\left\|\mathbf{z}_n^{(m)} - \mathbf{x}_n\right\|_2^2 \\ &= \underset{\mathbf{x}_n}{\operatorname{argmin}} \left\|\mathbf{x}_n + \mathbf{C}^{-1}\mathbf{b}\right\|_{\mathbf{C}^{-1}}^2\end{aligned}$$

where $\mathbf{C} \triangleq \mathbf{F}^T\mathbf{F} + \frac{\rho}{2}\mathbf{I}$ and $\mathbf{b} \triangleq -\frac{1}{2}\left(\mathbf{F}^T\mathbf{y}_n + \boldsymbol{\alpha}_n^{(m)} + \rho\mathbf{z}_n^{(m)}\right)$. We can see that the problem reduces to minimizing the Mahalanobis distance between $\mathbf{x}_n$ and $-\mathbf{C}^{-1}\mathbf{b}$ with the weight matrix $\mathbf{C}$. Given that this distance is non-negative, it is minimized when the argument is zero, yielding the straightforward solution $\mathbf{x}_n^{(m+1)} = -\mathbf{C}^{-1}\mathbf{b}$. It should be noted that in the case when $K > M$, the matrix $\mathbf{F}^T\mathbf{F}$ is rank deficient, but careful selection of the parameter ρ can ensure the existence of $\mathbf{C}^{-1}$. We note that $\mathbf{C}^{-1}$ can be computed once, as $\mathbf{F}$ and β are not altered between ADMM updates. As a result, the measurement model update is reduced to N matrix multiplications.

B. Proximal Update

The update equation for $\mathbf{z}$ does not depend on the measurement model or the penalty function and is therefore always of the form:

$$\begin{aligned}\mathbf{z}^{(m+1)} &= \underset{\mathbf{z}}{\operatorname{argmin}} \sum_{n=1}^{N} \boldsymbol{\lambda}_n^{(m)T}\left(\mathbf{z}_n - \mathbf{z}_{n-1} - \mathbf{w}_n^{(m)}\right) \\ &\quad + \boldsymbol{\alpha}_n^{(m)T}\left(\mathbf{z}_n - \mathbf{x}_n^{(m+1)}\right) \\ &\quad + \frac{\rho}{2}\left\|\mathbf{z}_n - \mathbf{z}_{n-1} - \mathbf{w}_n^{(m)}\right\|_2^2 \\ &\quad + \frac{\rho}{2}\left\|\mathbf{z}_n - \mathbf{x}_n^{(m+1)}\right\|_2^2\end{aligned} \tag{9}$$

Again noting that a quadratic term and linear term can be combined into a single 2-norm by completing the square, we simplify to an equivalent problem with only two terms by combining the terms containing $\mathbf{z}_n - \mathbf{z}_{n-1}$ and separately combining the terms containing $\mathbf{z}_n$. Beginning with the $\mathbf{z}_n$ terms, we have:

$$\begin{aligned}&\sum_{n=1}^{N} \boldsymbol{\alpha}_n^{(m)T}\left(\mathbf{z}_n - \mathbf{x}_n^{(m+1)}\right) + \frac{\rho}{2}\left\|\mathbf{z}_n - \mathbf{x}_n^{(m+1)}\right\|_2^2 \\ &\propto \sum_{n=1}^{N} \mathbf{z}_n^T\mathbf{z}_n + 2\left(\frac{\boldsymbol{\alpha}_n^{(m)}}{\rho} - \mathbf{x}_n^{(m+1)}\right)^T \mathbf{z}_n \\ &= \sum_{n=1}^{N} \left\|\mathbf{z}_n + \left(\frac{\boldsymbol{\alpha}_n^{(m)}}{\rho} - \mathbf{x}_n^{(m+1)}\right)\right\|_2^2\end{aligned}$$

For the terms containing $\mathbf{z}_n - \mathbf{z}_{n-1}$, we see that we can use the above result with the following substitutions: $\mathbf{z}_n \to \mathbf{z}_n - \mathbf{z}_{n-1}$, $\boldsymbol{\alpha} \to \boldsymbol{\lambda}$, and $\mathbf{x} \to \mathbf{w}$. Thus, we get:

$$\begin{aligned}&\sum_{n=1}^{N} \boldsymbol{\lambda}_n^{(m)T}\left(\mathbf{z}_n - \mathbf{z}_{n-1} - \mathbf{w}_n^{(m)}\right) + \frac{\rho}{2}\left\|\mathbf{z}_n - \mathbf{z}_{n-1} - \mathbf{w}_n^{(m)}\right\|_2^2 \\ &\propto \sum_{n=1}^{N} \left\|\mathbf{z}_n - \mathbf{z}_{n-1} + \left(\frac{\boldsymbol{\lambda}_n^{(m)}}{\rho} - \mathbf{w}_n^{(m)}\right)\right\|_2^2\end{aligned}$$

Combining the two yields the update equation:

$$\mathbf{z}^{(m+1)} = \underset{\mathbf{z}}{\operatorname{argmin}} \sum_{n=1}^{N} \|\mathbf{z}_n - \mathbf{z}_{n-1} - \mathbf{a}_n\|_2^2 + \|\mathbf{z}_n - \mathbf{b}_n\|_2^2$$

where we define $\mathbf{a}_n \triangleq \mathbf{w}_n^{(m)} - \boldsymbol{\lambda}_n^{(m)}/\rho$ and $\mathbf{b}_n \triangleq \mathbf{x}_n^{(m+1)} - \boldsymbol{\alpha}_n^{(m)}/\rho$. To further simplify, we define the vectorized versions of matrices $\mathbf{z}$, $\mathbf{a}$, and $\mathbf{b}$ as $\tilde{\mathbf{z}}$, $\tilde{\mathbf{a}}$, and $\tilde{\mathbf{b}}$ respectively, where each is a length $N \cdot K$ column vector. Furthermore, we define a matrix $\mathbf{G} \in \mathbb{R}^{KN \times KN}$ which consists of N $K \times K$ identity matrices along the diagonal and $N-1$ negative $K \times K$ identity matrices on the subdiagonal. Noting that $\operatorname{vec}(\mathbf{z}_n - \mathbf{z}_{n-1}) = \mathbf{G}\tilde{\mathbf{z}}$, we write the equivalent optimization problem for the vectorized version of $\mathbf{z}$ as:

$$\tilde{\mathbf{z}}^{(m+1)} = \underset{\tilde{\mathbf{z}}}{\operatorname{argmin}} \|G\tilde{\mathbf{z}} - \tilde{\mathbf{a}}\|_2^2 + \left\|\tilde{\mathbf{z}} - \tilde{\mathbf{b}}\right\|_2^2$$

Using standard quadratic minimization techniques we find the closed form solution:

$$\tilde{\mathbf{z}}^{(m+1)} = \left(\mathbf{G}^T\mathbf{G} + \mathbf{I}\right)^{-1}\left(\mathbf{G}^T\tilde{\mathbf{a}} + \tilde{\mathbf{b}}\right)$$

Inversion of this matrix can be costly for large N and K, so we instead note that $\left(\mathbf{G}^T\mathbf{G} + \mathbf{I}\right)$ has a block-tridiagonal structure and use an efficient algorithm for solving block-tridiagonal matrix problems [7]. Furthermore, this matrix has a simple structure that is determined solely by N and K, so we are able to take many shortcuts in the algorithm to further reduce the complexity. As shown in [7], this algorithm can be equivalently interpreted as a Kalman smoother. This follows intuitively if we view $\mathbf{z}$ as a latent process with noisy observations $\mathbf{b}_n$ and with transitions distributed according to independent normal distributions centered at $\mathbf{a}_n$.

C. Penalty Update

In general, the update equation for $\mathbf{w}$ is given by:

$$\begin{aligned}\mathbf{w}^{(m+1)} &= \underset{\mathbf{w}}{\operatorname{argmin}}\ \beta\phi(\mathbf{w}) \\ &\quad + \left(\sum_{n=1}^{N} \boldsymbol{\lambda}_n^{(m)T}\left(\mathbf{z}_n^{(m+1)} - \mathbf{z}_{n-1}^{(m+1)} - \mathbf{w}_n\right)\right. \\ &\quad \left. + \frac{\rho}{2}\left\|\mathbf{z}_n^{(m+1)} - \mathbf{z}_{n-1}^{(m+1)} - \mathbf{w}_n\right\|_2^2\right)\end{aligned} \tag{10}$$

As in the previous updates, we can combine the second and third terms by completing the square. Additionally, the sum of squared ℓ_2-norms can be more compactly represented with a single squared Frobenius norm. Thus, substituting the nuclear norm for the general penalty function ϕ, the $\mathbf{w}$ update equation becomes:

$$\mathbf{w}^{(m+1)} = \underset{\mathbf{w} \in \mathbb{R}^{K \times N}}{\operatorname{argmin}} \frac{2\beta}{\rho}\|\mathbf{w}\|_* + \left\|\mathbf{w} - \left(\mathbf{q} + \frac{\boldsymbol{\lambda}^{(m)}}{\rho}\right)\right\|_F^2 \tag{11}$$

where $\mathbf{q}$ is a matrix whose columns are defined by $\mathbf{q}_n \triangleq \mathbf{z}_n^{(m+1)} - \mathbf{z}_{n-1}^{(m+1)}$.

This update can be solved with very low complexity using a singular value soft-thresholding algorithm [8]. Given that the nuclear norm only penalizes $\mathbf{w}$ based on its singular values, it is clear that the left and right singular vectors of $\mathbf{w}^{(m+1)}$ can be taken directly from the matrix $\mathbf{q} + \boldsymbol{\lambda}^{(m)}/\rho$.

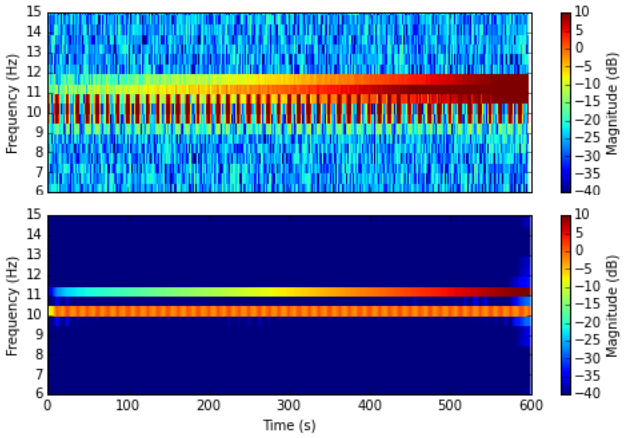

Fig. 1: Comparison of methods with $\sigma_v^2 = 1$ - Spectrogram with window length of 1000 and 25% window overlap (top) and low-rank spectrotemporal decomposition with $\beta = 4100$ and $\rho = 100$ after 50 iterations (bottom).

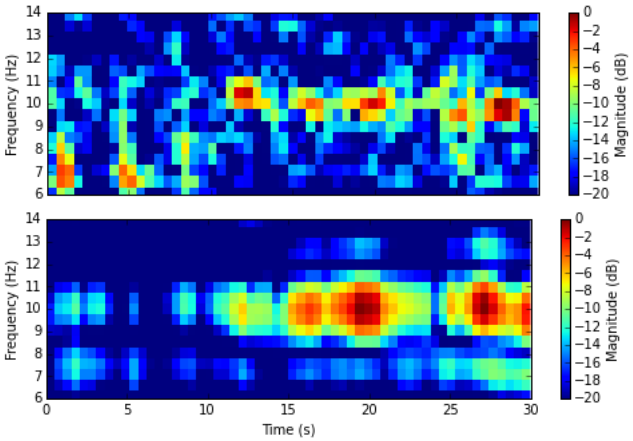

Fig. 2: Comparison of methods on human EEG recordings - Spectrogram with window length of 256 and 75% window overlap (top) and low-rank spectrotemporal decomposition with $\beta = 37{,}000$ and $\rho = 100$ after 50 iterations (bottom).

IV. Results

A. High Resolution Simulation

We begin by recreating a toy example from [3] which uses a time series with extremely distinct spectral structure. The discrete time series is given by:

$$y(t) = 10\cos^8(2\pi f_0 t)\sin(2\pi f_1 t) + 10\exp\left(4\frac{t-T}{T}\right)\cos(2\pi f_2 t) + v(t)$$

where $f_0 = 0.4$ Hz, $f_1 = 10$ Hz, $f_2 = 11$ Hz, $T = 600$ s, and $v(t) \sim \mathcal{N}(0, \sigma_v^2)$ is independent and identically distributed Gaussian noise. Note that T is consistent with the earlier definition (number of samples) and $y(t)$ is a scalar, not to be confused with the vector $\mathbf{y}_n = [y((n-1)M+1), \ldots, y(nM)]^T$. The result is two amplitude modulated signals at neighboring frequencies, one with an oscillating envelope and the other with an exponential envelope.

In Fig. 1 we see a stark difference in the spectral resolution of a traditional spectrogram and the low-rank spectrotemporal decomposition. The traditional spectrogram suffers from significant spectral leakage issues, making it very difficult to distinguish between the activity occurring at 10 and 11 Hz, while the low-rank spectrotemporal decomposition has a clear separation between the two frequency bands.

B. Experimental Data

Next we observe the results of using the low-rank spectrotemporal decomposition on human EEG recordings collected using adhesive flexible electronic sensors developed in [9]. Specifically we consider a 30 second time series consisting of 10 seconds where the subject's eyes are open followed by 20 seconds where the subject's eyes are closed, at which point we expect to see an alpha rhythm in the 10-12 Hz frequency band. During the first 10 seconds of the recording, we witness three spikes in spectral activity resulting from the subject's eye blinks. The data was sampled at 500 Hz and passed through a 6-14 Hz digital bandpass filter before computing the spectral estimates.

The results from both the spectrogram and the low-rank spectrotemporal decomposition are shown in Fig. 2. Both methods were normalized to have a maximum at 0 dB and are shown with a 20 dB dynamic range. We can see that the low-rank spectrotemporal decomposition exaggerates the activity in the alpha band while dampening the noise in the remainder of the spectrum. However, unlike the toy example, there is not an increased sharpness in the spectral resolution. This follows from the fact that the nuclear norm penalty is not making any inherent assumptions with regard to specific spectral or temporal structures in the data. As a result, in signals where the spectral sparsity is as precise as the toy example, the low-rank decomposition captures that quality effectively, while the estimate of more complex signals may manifest differently depending on that signal's dominant characteristics.

Fig. 3 demonstrates the ability of the low-rank spectrotemporal decomposition to differentiate between the eyes-open and eyes-closed states. We can see that while the traditional spectrogram does not yield a significant shift in the average magnitude in the alpha frequency band form the eye-open to eyes-closed state, there is a nearly 10 dB difference on average for the low-rank decomposition. To illustrate this point we consider a simple threshold for each method and measure the percentage of samples that the threshold incorrectly classifies as eyes-open or eyes-closed samples. For this example, the best performing thresholds yield 20% error for the traditional spectrogram and just under 7% percent error for the low-rank spectrotemporal decomposition. We see that our algorithm accentuates the defining spectral characteristics of this particular EEG recording.

It should be noted that the number of iterations and the parameters ρ and β used in generating the estimates above were chosen experimentally to obtain what we deemed to be the most desirable results. In future work we intend to determine convergence based on criteria detailed in [5].

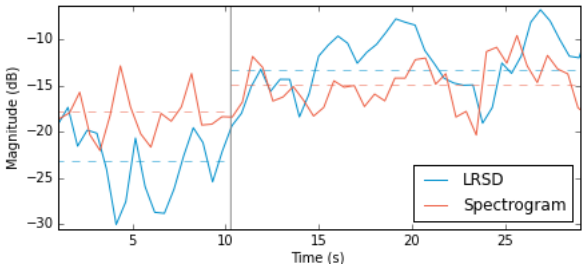

Fig. 3: Average power in the alpha frequency band (8-12 Hz) for the spectrogram (red) and low-rank spectrotemporal decomposition (blue) and their corresponding time averages before and during the closed eyes period (dashed).

REFERENCES

[1] J. d. R. Millán and J. Carmena, "Invasive or noninvasive: Understanding brain-machine interface technology," *IEEE Engineering in Medicine and Biology Magazine*, vol. 29, no. EPFL-ARTICLE-150426, pp. 16–22, 2010.

[2] K.-C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of Optimization*, vol. 6, no. 615-640, p. 15, 2010.

[3] D. Ba, B. Babadi, P. L. Purdon, and E. N. Brown, "Robust spectrotemporal decomposition by iteratively reweighted least squares," *Proceedings of the National Academy of Sciences*, vol. 111, no. 50, E5336–E5345, 2014.

[4] M. Fornasier, H. Rauhut, and R. Ward, "Low-rank matrix recovery via iteratively reweighted least squares minimization," *SIAM Journal on Optimization*, vol. 21, no. 4, pp. 1614–1640, 2011.

[5] 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.

[6] N. Parikh and S. P. Boyd, "Proximal algorithms.," *Foundations and Trends in optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[7] A. Y. Aravkin, J. V. Burke, and G. Pillonetto, "Optimization viewpoint on kalman smoothing with applications to robust and sparse estimation," in *Compressed Sensing & Sparse Filtering*, Springer, 2014, pp. 237–280.

[8] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and bregman iterative methods for matrix rank minimization," *Mathematical Programming*, vol. 128, no. 1-2, pp. 321–353, 2011.

[9] D. Y. Kang, Y.-S. Kim, G. Ornelas, M. Sinha, K. Naidu, and T. P. Coleman, "Scalable microfabrication procedures for adhesive-integrated flexible and stretchable electronic sensors," *Sensors*, vol. 15, no. 9, pp. 23 459–23 476, 2015.