Spread and Sparse: Learning Interpretable Transforms for Bandlimited Signals on Directed Graphs

Rasoul Shafipour and Gonzalo Mateos

***Abstract*—We address the problem of learning a sparsifying graph Fourier transform (GFT) for compressible signals on directed graphs (digraphs). Blending the merits of Fourier and dictionary learning representations, the goal is to obtain an orthonormal basis that captures spread modes of signal variation with respect to the underlying network topology, and yields parsimonious representations of bandlimited signals. Accordingly, we learn a data-adapted dictionary by minimizing a spectral dispersion criterion over the achievable frequency range, along with a sparsity-promoting regularization term on the GFT coefficients of training signals. An iterative algorithm is developed which alternates between minimizing a smooth objective over the Stiefel manifold, and soft-thresholding the graph-spectral domain representations of the signals in the training set. A frequency analysis of temperature measurements recorded across the contiguous United States illustrates the merits of the novel GFT design.**

***Index Terms*—** Graph signal processing, graph Fourier transform, directed graphs, dictionary learning.

I. Introduction

Network data supported on the vertices of a graph $\mathcal{G}$ are becoming ubiquitous across disciplines spanning the bio-behavioral sciences and engineering. Examples range from measurements of neural activities at different regions of the brain [1], to vehicle traces over transportation networks [2]. Such data, in a snapshot, can be thought of as graph signals represented by vectors indexed by the N nodes of $\mathcal{G}$. In this context, the goal of graph signal processing (GSP) is to broaden the scope of traditional signal and information processing by developing algorithms that fruitfully exploit the complex relational structure of said signals. Accordingly, generalizations of signal processing tasks have been explored in the literature; see [3] for a recent tutorial treatment.

Focusing on signal representations, noteworthy contributions include the graph Fourier transform (GFT) and the design of data-adapted dictionaries incorporating the graph stucture. Bringing together these two advances while building on our previous work [4], here we address the problem of learning a sparsifying digraph (D)GFT basis (dictionary) for compressible signals on directed networks. In particular, we first use our recently proposed approach to capture the notion of signal variation (frequency) over digraphs and to approximately find the maximum attainable frequency ($f_{\max}$) [4]. Then we design an orthonormal digraph (D)GFT basis such that: (i) the resulting frequencies (i.e., the directed variation of the sought orthonormal basis vectors) distribute as evenly as possible over $[0, f_{\max}]$; and (ii) a given set of training bandlimited graph signals can be sparsely represented using a few basis vectors. This design would facilitate more interpretable frequency analyses and aid universal filter design in the spectral domain, while offering parsimonious representations of signals on digraphs.

Work in this paper was supported by the NSF awards CCF-1750428 and ECCS-1809356. The authors are with the Dept. of ECE, University of Rochester. Emails: {rshafipo,gmateosb}@ece.rochester.edu

Related work. Learning (overcomplete) dictionaries to sparsely represent signals has been an active area of research. For example, the K-SVD method in [5] can be applied to graph signals as vectors in $\mathbb{R}^N$; however, such methods do not explicitly account for $\mathcal{G}$'s topology; see also [6]. In the GSP literature for undirected graphs, the workhorse GFT approach based on eigenvectors of the Laplacian matrix has been shown to sparsely represent smooth graph signals [7]. A GFT framework combining aspects of signal variation and energy to design general orthonormal transforms for graph signals is proposed in [8]. An efficient algorithm to learn parametric dictionaries for signals over undirected networks is put forth in [9]; see also the online dictionary-learning approach in [10] developed to predict partially-observed dynamic graph processes such as link loads on communication networks.

A more general GFT definition is based on the Jordan decomposition of adjacency matrix $\mathbf{A} = \mathbf{V}\mathbf{J}\mathbf{V}^{-1}$, where the spectral representation of graph signal $\mathbf{x}$ is $\tilde{\mathbf{x}} = \mathbf{V}^{-1}\mathbf{x}$ [11]. While this definition is valid for digraphs, the associated notion of signal variation in [11] does not ensure that constant signals have zero variation. Moreover, $\mathbf{V}$ is not necessarily orthonormal and Parseval's identity does not hold; see also [12]. A noteworthy directed GFT construction was put forth in [13], which is based on minimization of the (convex) Lovász extension of the graph cut size, subject to orthonormality constraints on the desired basis. However, the GFT basis vectors in [13] may fail to yield signal representations capturing different modes of signal variation with respect to $\mathcal{G}$; see [4, Section III-A]. Recently, we have proposed an orthonormal DGFT in [4] whereby we learn a set of basis vectors corresponding to spread frequency modes. However, none of these methods for digraphs are data adaptive, since the dictionaries only account for the graph structure.

Contributions. Here we propose a novel approach to sparsely represent bandlimited signals in the graph spectral domain, while the learnt atoms explicitly capture different modes of variation with respect to the underlying digraph. We design a novel DGFT basis with the following desirable properties: P1) The basis vectors (dictionary atoms) provide notions of frequency and signal variation over digraphs. P2) Frequency modes are designed to be well spread from zero to $f_{\max}$, thus better capturing low, middle, and high frequencies to enhance interpretability of frequency analyses. P3) The learnt

dictionary is orthonormal so the transform preserves inner products in the vertex and graph frequency domain; i.e., Parseval's identity holds. P4) Bandlimited signals can be sparsely represented in the graph spectral domain using a few atoms of the basis (dictionary); i.e., the DGFT provides parsimonious representations of historical signals as well as of unseen samples from the subjacent generative network process. In [4] we constructed a DGFT basis with properties P1)-P3). On top of those, here we further make our DGFT data-adaptive to sparsely represent a set of bandlimited graph signals at hand. To that end, in Section II we formulate the problem of learning a dictionary adhering to P1)-P4). Next, we develop an alternating algorithm by leveraging a feasible method for optimization with orthonormality constraints and embedded soft-thresholding operations (outlined in Section III). With the aid of computer simulations, the convergence behavior and effectiveness of the proposed algorithm is examined in Section IV. Concluding remarks are given in Section V.

II. Preliminaries and Problem Statement

In this section, we first introduce some basic GSP concepts and terminology. Then, we briefly review a notion of signal variation and graph frequencies on digraphs which we utilize to make frequency modes as spread as possible. Lastly, we state the problem of learning a sparsifying (orthonormal) DGFT basis with well-dispersed modes in the graph spectral domain.

Notation and terminology. We consider a weighted digraph $\mathcal{G} = (\mathcal{V}, \mathbf{A})$, where $\mathcal{V}$ is the set of nodes with cardinality $|\mathcal{V}| = N$, and $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the graph adjacency matrix with A_{ij} denoting the edge weight from node i to node j. A graph signal $\mathbf{x} : \mathcal{V} \mapsto \mathbb{R}^N$ can be represented as a vector of length N, where the ith entry denotes the signal value at node i. For a given network topology $\mathbf{A}$, suppose we observe P compressible (e.g., bandlimited) graph signals collected in a matrix $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_P] \in \mathbb{R}^{N\times P}$. Then our goal is to find N orthonormal basis vectors capturing different modes of variation with respect to $\mathcal{G}$, while sparsely representing the signals $\mathbf{X}$. We collect these desired basis vectors in a matrix $\mathbf{U} := [\mathbf{u}_1, \cdots, \mathbf{u}_N] \in \mathbb{R}^{N\times N}$, where $\mathbf{u}_k \in \mathbb{R}^N$ represents the kth frequency component. This means that the DGFT of a graph signal $\mathbf{x}$ is $\tilde{\mathbf{x}} = \mathbf{U}^T\mathbf{x}$. The inverse DGFT is $\mathbf{x} = \mathbf{U}\tilde{\mathbf{x}} = \sum_{k=1}^{N} \tilde{x}_k\mathbf{u}_k$, which allows one to synthesize $\mathbf{x}$ as a linear combination of orthogonal frequency modes $\mathbf{u}_k$.

A signal directed variation measure. To measure how a basis vector $\mathbf{u}$ varies over the network and define graph frequencies, we adopt the following notion of signal directed variation (DV) over digraphs [14]

$$\mathrm{DV}(\mathbf{u}) := \sum_{i,j=1}^{N} A_{ij}[u_i - u_j]_+^2, \tag{1}$$

where $[x]_+ = \max(0, x)$. One can then generalize the notion of frequency to digraphs as the directed variation of the basis signal $\mathbf{u}_k$; i.e., $f_k := \mathrm{DV}(\mathbf{u}_k)$.

To gain more insight on (1), consider an arbitrary graph signal $\mathbf{x} \in \mathbb{R}^N$ on digraph $\mathcal{G}$, where a directed edge denotes the direction of signal flow from a larger value to a smaller one. So, a directed edge from node i to node j would contribute to the $\mathrm{DV}(\mathbf{x})$ only if the inequality $x_i > x_j$ holds true between the corresponding vertices (cf. the desirable property P1 alluded to in Section I).

Spread frequencies. Similar to the discrete spectrum of periodic time-varying signals, we seek N (approximately) equidistributed graph frequencies covering the whole viable spectrum to capture low, medium, and high frequencies. To cover the whole spectrum of variations, we set $\mathbf{u}_1 = \mathbf{u}_{\min} = \frac{1}{\sqrt{N}}$ for capturing the minimum frequency (i.e., DC component) and $\mathbf{u}_N = \mathbf{u}_{\max} := \arg\max_{\|\mathbf{u}\|_2=1} \mathrm{DV}(\mathbf{u})$, where $f_{\max} := \mathrm{DV}(\mathbf{u}_{\max})$ is the maximum attainable directed variation by a unit-norm vector. In this work, we rely on the recently proposed method in [4, Algorithm 1] to find $\mathbf{u}_{\max}$. Having fixed the first and last columns of $\mathbf{U}$ and thus their corresponding frequencies, ideally one would like the free frequencies (i.e., $f_2, \cdots, f_{N-1}$) to form an arithmetic sequence between $f_1 = 0$ and $f_N = f_{\max}$, yielding maximally-spread frequency modes as in the DFT. However, this might not always be feasible as discussed in [4, Section II-B].

Instead, by taking into account the graph structure and fixed $\mathbf{u}_1$ and $\mathbf{u}_N$, one can consider the spectral dispersion function

$$\delta(\mathbf{U}) := \sum_{i=1}^{N-1} [\mathrm{DV}(\mathbf{u}_{i+1}) - \mathrm{DV}(\mathbf{u}_i)]^2 \tag{2}$$

which is minimized when the free frequencies form an arithmetic sequence between $f_1 = 0$ and $f_N = f_{\max}$ (cf. the desirable property P2 stated in Section I). Such a spread frequency distribution which captures different modes of variation with respect to the graph would facilitate more interpretable spectral analyses of graph signals and also aid filter design in the graph spectral domain; see also [4] for illustrations.

Problem statement. In addition to the aforementioned design considerations, we would like the DGFT basis to sparsely represent a given set of bandlimited training signals $\mathbf{X} \in \mathbb{R}^{N\times P}$ in the spectral (dual) domain. Sparsity can be promoted by minimizing $\|\mathbf{U}^T\mathbf{X}\|_0 = \|\tilde{\mathbf{X}}\|_0$, where the ℓ_0-(pseudo) norm $\|.\|_0$ counts the number of non-zero entries in $\tilde{\mathbf{X}}$. Since the ℓ_0-norm criterion is in general NP-hard to optimize, we minimize its closest convex approximant $\|\tilde{\mathbf{X}}\|_1 = \sum_{i,j}|\tilde{X}_{ij}|$ instead. Sparse recovery via ℓ_1-norm minimization has shown remarkable success over the last decade or so, since the convex relaxation often entails no loss of optimality.

Consolidating all the mentioned criteria, the DGFT design as a dictionary learning task can be stated as

$$\begin{aligned} \min_{\mathbf{U}} \quad & \Psi(\mathbf{U}) := \delta(\mathbf{U}) + \mu\|\mathbf{U}^T\mathbf{X}\|_1 \\ \text{subject to} \quad & \mathbf{U}^T\mathbf{U} = \mathbf{I}_N, \\ & \mathbf{u}_1 = \mathbf{u}_{\min}, \\ & \mathbf{u}_N = \mathbf{u}_{\max}, \end{aligned} \tag{3}$$

where regularization parameter $\mu > 0$ trades-off the dispersion of the frequency modes versus the sparsity of the training signals' DGFT coefficients. For the extreme value of $\mu = 0$, (3) finds the most dispersed frequency components in $[0, f_{\max}]$. By increasing μ we compromise the dispersion for sparser representations of the observed signals' DGFT coefficients.

Note that the orthonormality (Stiefel manifold) constraints in (3) ensures the desirable property P3.

Problem (3) is feasible (i.e., $\mathbf{u}_{\max} \perp \mathbf{u}_{\min}$) as shown in [4, Proposition 3]. However, finding the global optimum of (3) is challenging due to the non-convexity arising from the orthonormality constraints as well as the objective function [due to the cross-terms $\mathrm{DV}(\mathbf{u}_i)\mathrm{DV}(\mathbf{u}_{i+1})$]. While the spectral dispersion $\delta(\mathbf{U})$ is smooth, the non-differentiability of ℓ_1-norm regularization term hinders a direct adoption of methods for effectively minimizing differentiable cost functions over the Stiefel manifold like the one developed in [15].

In the next section we adopt a variable splitting technique that facilitates an alternating-minimization algorithm to solve (3). The iterates are obtained by combing a (feasible) method over the Stiefel manifold [15] along with sparsity-inducing soft-thresholding operations.

III. Alternating-Minimization Scheme

Here we develop an iterative algorithm to find the orthonormal basis $\mathbf{U}$ that solves (3). Our first idea is to replace $\mathbf{U}^T\mathbf{X}$ with an auxiliary variable $\mathbf{Y} \in \mathbb{R}^{N\times P}$ and enforce the equality constraint $\mathbf{Y} = \mathbf{U}^T\mathbf{X}$ via a quadratic penalty term in the objective. Accordingly, we reformulate (3) as

$$\begin{aligned} \min_{\mathbf{U},\mathbf{Y}} \quad & \delta(\mathbf{U}) + \mu\|\mathbf{Y}\|_1 + \frac{\gamma}{2}\|\mathbf{Y} - \mathbf{U}^T\mathbf{X}\|_F \\ \text{subject to} \quad & \mathbf{U}^T\mathbf{U} = \mathbf{I}_N \\ & \mathbf{u}_1 = \mathbf{u}_{\min}, \\ & \mathbf{u}_N = \mathbf{u}_{\max}, \end{aligned} \tag{4}$$

and we adopt a block-coordinate descent approach, which solves (4) cyclically over each variable $\mathbf{U}$ and $\mathbf{Y}$ while fixing the other variable to its most up-to-date value.

Optimization in the Stiefel manifold. At each iteration $k = 0, 1, 2\ldots$ we fix $\mathbf{Y} = \mathbf{Y}_k$ and update $\mathbf{U}_{k+1}$ by bringing to bear a feasible method for minimizing differentiable functions over the Stiefel manifold [15]. Specifically, we would like to solve the orthogonality-constrained, smooth problem

$$\begin{aligned} \min_{\mathbf{U}} \quad & \delta(\mathbf{U}) + \frac{\gamma}{2}\|\mathbf{Y} - \mathbf{U}^T\mathbf{X}\|_F \\ \text{subject to} \quad & \mathbf{U}^T\mathbf{U} = \mathbf{I}_N \\ & \mathbf{u}_1 = \mathbf{u}_{\min}, \\ & \mathbf{u}_N = \mathbf{u}_{\max}. \end{aligned} \tag{5}$$

However, the general feasible method of [15] is tailored for orthogonality constrained problems of the form

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times p}} \phi(\mathbf{U}), \quad \text{subject to} \quad \mathbf{U}^T\mathbf{U} = \mathbf{I}_p, \tag{6}$$

where $\phi(\mathbf{U}) : \mathbb{R}^{n\times p} \to \mathbb{R}$ is assumed to be differentiable. To cast the optimization (5) in the favorable form of (6), we penalize the objective in (5) with a measure of the constraint violations to obtain

$$\begin{aligned} \mathbf{U}_{k+1} := \arg\min_{\mathbf{U}} \phi(\mathbf{U}) := \delta(\mathbf{U}) + \frac{\gamma}{2}\|\mathbf{Y}_k - \mathbf{U}^T\mathbf{X}\|_F^2 \\ + \frac{\lambda}{2}\big(\|\mathbf{u}_1 - \mathbf{u}_{\min}\|^2 + \|\mathbf{u}_N - \mathbf{u}_{\max}\|^2\big) \\ \text{subject to} \quad \mathbf{U}^T\mathbf{U} = \mathbf{I}_N, \end{aligned} \tag{7}$$

Algorithm 1 Sparsifying DGFT with spread frequency modes

1: **Input:** Adjacency matrix $\mathbf{A}$, signals $\mathbf{X} \in \mathbb{R}^{N\times P}$, and parameters λ, μ, γ, ϵ_1, $\epsilon_2 > 0$.
2: Find $\mathbf{u}_{\max}$ using [4, Algorithm 1] and set $\mathbf{u}_{\min} = \frac{1}{\sqrt{N}}\mathbf{1}_N$.
3: **Initialize** $k = 0$, $\mathbf{Y}_0 \in \mathbb{R}^{N\times P}$ at random.
4: **repeat**
5: **U-update:**
6: **Initialize** $t = 0$ and orthonormal $\hat{\mathbf{U}}_0 \in \mathbb{R}^{N\times N}$.
7: **repeat**
8: Compute gradient $\mathbf{H}_t := \nabla\phi(\hat{\mathbf{U}}_t) \in \mathbb{R}^{N\times N}$.
9: Form $\mathbf{B}_t = \mathbf{H}_t\hat{\mathbf{U}}_t^T - \hat{\mathbf{U}}_t\mathbf{H}_t^T$.
10: Select τ_t satisfying Armijo-Wolfe conditions [16].
11: Update $\hat{\mathbf{U}}_{t+1}(\tau_t) = (\mathbf{I}_N + \frac{\tau_t}{2}\mathbf{B}_t)^{-1}(\mathbf{I}_N - \frac{\tau_t}{2}\mathbf{B}_t)\hat{\mathbf{U}}_t$.
12: $t \leftarrow t + 1$.
13: **until** $\|\hat{\mathbf{U}}_t - \hat{\mathbf{U}}_{t-1}\|_F / \|\hat{\mathbf{U}}_{t-1}\|_F \le \epsilon_1$
14: **Return** $\mathbf{U}_k = \hat{\mathbf{U}}_t$
15: **Y-update:**
16: $\mathbf{Y}_k = \mathrm{sgn}(\mathbf{U}_k^T\mathbf{X}) \circ (|\mathbf{U}_k^T\mathbf{X}| - \mu/\gamma)_+$.
17: $k \leftarrow k + 1$.
18: **until** $\|\mathbf{U}_k^T\mathbf{X} - \mathbf{U}_{k-1}^T\mathbf{X}\|_1 / \|\mathbf{U}_{k-1}^T\mathbf{X}\|_1 \le \epsilon_2$
19: **Return** $\hat{\mathbf{U}} = \mathbf{U}_k$.

where $\lambda > 0$ is chosen large enough to ensure $\mathbf{u}_1 = \mathbf{u}_{\min}$ and $\mathbf{u}_N = \mathbf{u}_{\max}$. Leveraging a solver similar to the one developed in [4, Section IV], one can guarantee convergence to a stationary point of $\phi(\mathbf{U})$ by virtue of [15, Theorem 2], while generating feasible points in the Stiefel manifold at every inner iteration $t = 0, 1, 2, \ldots$.

The resulting iterations to solve (7) [and thus (5)] are tabulated under Algorithm 1; see steps $6 - 14$ which are performed until a stopping criterion is met. The gradient matrix $\mathbf{H}_t := \nabla\phi(\hat{\mathbf{U}}_t)$ at step 9 can be computed as $\mathbf{H}_t = \mathbf{G}_t - \gamma\mathbf{X}(\mathbf{Y}_k - \hat{\mathbf{U}}_t^T\mathbf{X})^T$, where $\mathbf{G}_t$ is the dispersion function gradient calculated in [4, Equation (17)]. Matrix $\mathbf{B}_t := \mathbf{H}_t\hat{\mathbf{U}}_t^T - \hat{\mathbf{U}}_t\mathbf{H}_t^T$ at step 10 is a skew-symmetric projection of the gradient $\mathbf{H}_t$ onto the constraint's tangent space. The update rule at step 12 is known as the Cayley transform which preserves the orthonormality for $\mathbf{U}_{t+1}$, since $(\mathbf{I} + \frac{\tau}{2}\mathbf{B}_t)^{-1}$ and $\mathbf{I} - \frac{\tau}{2}\mathbf{B}_t$ commute. This update rule is a descent path for a proper step size τ; see [15] for more details. In particular, one such step size τ can be obtained through a curvilinear search satisfying the Armijo-Wolfe conditions [16]. Steps $6 - 14$ ensure that the update rule at 12th step is a descent path and [15, Theorem 2] asserts that the overall procedure converges to a stationary point of $\phi(\mathbf{U})$.

Spread and sparse DGFT. After finding a stationary solution $\mathbf{U}_k$ at iteration k, we keep the the latest DGFT basis fixed and solve the following problem with respect to $\mathbf{Y}$:

$$\min_{\mathbf{Y}} \quad \mu\|\mathbf{Y}\|_1 + \frac{\gamma}{2}\|\mathbf{Y} - \mathbf{U}_k^T\mathbf{X}\|_F^2. \tag{8}$$

For fixed $\mathbf{U}_k$, the unknown variable $\mathbf{Y}$ in (8) is component-wise separable. In particular, $\mathbf{Y}_k$ can be obtained in a simple closed form in terms of soft-thresholding operations, namely

$$\mathbf{Y}_k = \mathrm{sign}(\mathbf{U}_k^T\mathbf{X}) \circ \big[|\mathbf{U}_k^T\mathbf{X}| - \mu/\gamma\big]_+, \tag{9}$$

Fig. 1. Graph signal of average annual temperature in Fahrenheit for the contiguous US states digraph.

where $\circ$ and $\mathrm{sgn}(.)$ are the element-wise product and sign function, respectively; see also step 16 in Algorithm 1.

In summary, the overall procedure in Algorithm 1 entails alternating updates for $\mathbf{U}$ [i.e., by solving (7)] and $\mathbf{Y}$ [i.e., solving (8)] until a certain termination condition is met. This results in an approximate solution to the original problem (3), and thus a data-driven DGFT basis with well-dispersed frequencies that can sparsely represent signals from the underlying process. We generally observe convergence within a reasonable number of iterations; see the next section for an empirical demonstration. Deriving formal convergence guarantees for the alternating scheme would be a valuable extension, which is a topic of ongoing investigation.

IV. Preliminary Numerical Test

To gain insights on the behavior of the proposed data-driven DGFT, here we carry out simulations on real-world graph signals. Specifically, we evaluate the performance of Algorithm 1 in yielding (near) maximally-spread frequency components with sparse representations for given (bandlimited) training signals as well as a test signal from a similar process.

We consider a digraph of 48 contiguous United States (excluding Alaska and Hawaii which are not connected by land). A directed edge joins two states if they share a border, and the direction of the arc is set so that the state whose barycenter has a lower latitude points to the one with higher latitude – consistent with the temperature flow. We also consider the states annual average temperature used as the test signal $\mathbf{x} \in \mathbb{R}^{48}$ which is shown in Fig. 1. To train the proposed Algorithm 1, we use average monthly temperature over the past $\sim$ 60 years of each state as the training signals[1] $\mathbf{X} \in \mathbb{R}^{48\times 12}$. Our goal is to sparsely represent the signals while capturing meaningful, spread, and broad modes of variation with respect to the graph. To that end, we run Algorithm 1 on the training signals $\mathbf{X}$ for the established graph model.

First we use the Monte-Carlo method to study the convergence properties of the algorithm. As the iterations of Algorithm 1 evolve in Fig. 2-(a), we show the objective function $\Psi(\mathbf{U})$ in (3). We do so for 10 different initializations and report the median as well as the first and third quartiles versus number of iterations. While still there is no theoretical guarantees on the overall convergence of Algorithm 1, we observe that the realizations converge to limiting values with small variability in this practical setting.

Fig. 2-(b) depicts the heat maps of the non-DC components absolute values for the trained $\hat{\mathbf{X}}$ obtained using Algorithm 1 when $\mu \neq 0$ (data-adaptive) (left) and when $\mu = 0$ (i.e., method in [4] which does not take the signals into consideration) (right). Each column in Fig. 2-(b) depicts a sample's spectral representation excluding the DC component. As we can see, we achieve sparser representations in the Fourier domain for properly chosen μ.

In addition, Fig. 2-(c) compares the test signal ($\mathbf{x}$) representations in the graph spectral domain ($\tilde{\mathbf{x}}$) using two mentioned approaches. Notice that $\tilde{\mathbf{x}}$ is naturally lowpass bandlimited and can be captured by a few atoms of the obtained basis even when $\mu \neq 0$ due to the interpretability originated from the design considerations (cf. $\delta(\mathbf{U})$ in (2) and (3)). However, for the data-adapted algorithm (i.e. $\mu \neq 0$), the test signal $\mathbf{x}$ is more sparsely representable, which showcases the effectiveness of the obtained data-driven approach. To better appreciate this, we also plot the cumulative energy distributions of both methods, defined by the percentage of the total energy present in the first i frequency components for $i = 1, \cdots, N$. It is apparent from Fig. 2-(c) that the first few components of $\tilde{\mathbf{x}}$ capture more of its energy when $\mu \neq 0$.

Finally, Fig. 3 depicts the distribution of all the frequencies for the basis vectors obtained by Algorithm 1 and its counterpart in [4] which is the special case of Algorithm 1 with $\mu = 0$. In Fig. 3, each vertical line indicates the directed variation (frequency) associated with a basis vector. Although we compromised the spectral dispersion for having sparse representations, but Fig. 3 shows that we can still achieve relatively well dispersed frequencies.

V. Conclusion

We considered the problem of finding a sparsifying orthonormal Fourier basis for bandlimited graph signals supported on a digraph. The notion of frequency was captured via a recently proposed directed variation measure of basis vectors over directed networks. Then frequency modes were designed to evenly span the entire viable frequency range (with respect to the graph), while at the same time sparsely representing the given training signals. To that end, we developed an iterative algorithm that alternates between a feasible method over the Stiefel manifold to minimize a spectral dispersion criterion, and a proximal operator stemming from a sparsity-promoting regularization term on the DGFT coefficients of the training signals. The overall procedure to find a desirable basis was validated on real-world temperature signals supported on a directed network of the US contiguous states.

With regards to future directions, a formal convergence guarantee for the developed algorithm is of interest. Furthermore, it would be a significant improvement to exploit the

[1]Temperature data obtained from http://www.weatherbase.com and https://www.ncdc.noaa.gov, respectively

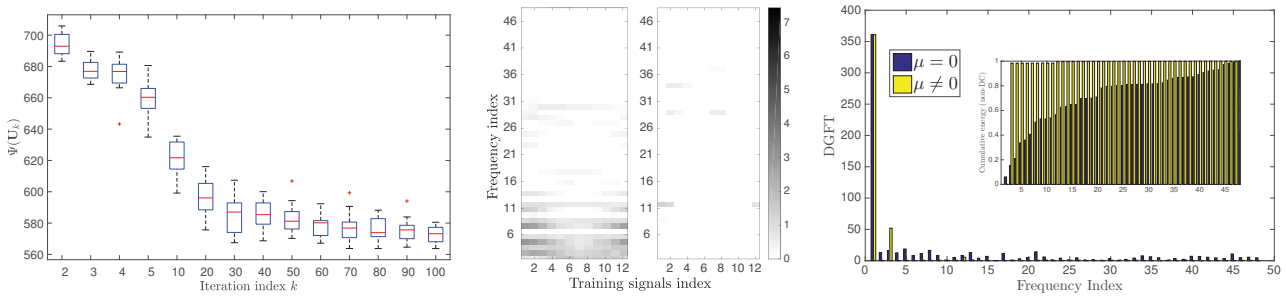

Fig. 2. (a) Median and the 25th and 75th percentiles of the objective function in (3) vs. the number of iterations k in Algorithm 1, obtained by running 10 Monte-Carlo simulations. (b) Non-DC components of the trained signals DGFT ($\tilde{\mathbf{X}}$) using Algorithm 1 when $\mu = 0$ (left) and when $\mu \neq 0$ (right). (c) DGFT of the test signal ($\tilde{\mathbf{x}}$) one with and one without sparsity regularization, along with their cumulative energy distribution across frequencies.

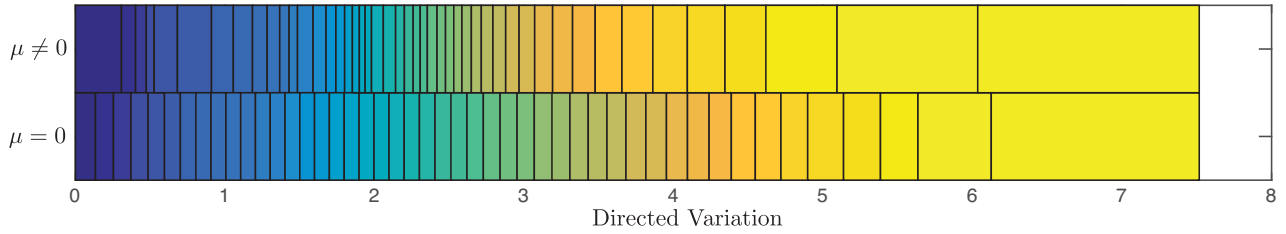

Fig. 3. Comparison of directed variations using Algorithm 1 for a tuned $\mu \neq 0$ and $\mu = 0$, where the latter boils down to the method in [4]. Colored boxes show the difference between two consecutive frequencies for each method, while the frequency values correspond to the vertical boundary lines.

knowledge on the passband of the bandlimited signals to yield better spectral dispersion versus the sparsity trade-off.

REFERENCES

[1] W. Huang, L. Goldsberry, N. F. Wymbs, S. T. Grafton, D. S. Bassett, and A. Ribeiro, "Graph frequency analysis of brain signals," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 7, pp. 1189–1203, 2016.

[2] J. A. Deri and J. M. F. Moura, "New York City taxi analysis with graph signal processing," in *Proc. IEEE Global Conf. on Signal and Information Process.*, Washington, DC, Dec. 7-9, 2016, pp. 1275–1279.

[3] A. Ortega, P. Frossard, J. Kovacevic, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges and applications," *Proc. of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[4] R. Shafipour, A. Khodabakhsh, G. Mateos, and E. Nikolova, "A Directed Graph Fourier Transform with Spread Frequency Components," *IEEE Trans. Signal Process. (to appear; see also arXiv:1804.03000 [eess.SP])*, 2018.

[5] M. Aharon, M. Elad, A. Bruckstein *et al.*, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, p. 4311, 2006.

[6] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.

[7] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, 2013.

[8] B. Girault, A. Ortega, and S. Narayanan, "Irregularity-aware graph Fourier transforms," *IEEE Trans. Signal Process. (submitted; see also arXiv:1802.10220 [eess.SP])*, 2018.

[9] D. Thanou, D. I. Shuman, and P. Frossard, "Learning parametric dictionaries for signals on graphs," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3849–3862, 2014.

[10] P. A. Forero, K. Rajawat, and G. B. Giannakis, "Prediction of partially observed dynamical processes over networks via dictionary learning," *IEEE Trans. Signal Process.*, vol. 62, no. 13, pp. 3305–3320, July 2014.

[11] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, June 2014.

[12] J. A. Deri and J. M. F. Moura, "Spectral projector-based graph Fourier transforms," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 6, pp. 785–795, 2017.

[13] S. Sardellitti, S. Barbarossa, and P. Di Lorenzo, "On the graph Fourier transform for directed graphs," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 6, pp. 796–811, 2017.

[14] R. Shafipour, A. Khodabakhsh, G. Mateos, and E. Nikolova, "A digraph Fourier transform with spread frequency components," in *Proc. IEEE Global Conf. on Signal and Information Process.*, Montreal, Canada, Nov. 14-16, 2017, pp. 583–587.

[15] Z. Wen and W. Yin, "A feasible method for optimization with orthogonality constraints," *Mathematical Programming*, vol. 142, no. 1-2, pp. 397–434, 2013.

[16] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering, 2006.