DIGRAPH FOURIER TRANSFORM VIA SPECTRAL DISPERSION MINIMIZATION

Rasoul Shafipour[†]*, Ali Khodabakhsh*[‡]*, Gonzalo Mateos*[†]*, and Evdokia Nikolova*[‡]

[†]Dept. of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA
[‡]Dept. of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX, USA

ABSTRACT

We address the problem of constructing a graph Fourier transform (GFT) for both undirected and directed graphs (digraphs), which decomposes graph signals into different modes of variation with respect to the underlying network. Accordingly, we seek orthonormal bases that yield maximally-spread frequency components in the graph spectral domain to better capture low, medium and high frequencies. To that end, we advocate a two-step design whereby we: (i) find the maximum directed variation (i.e., frequency on a digraph) a candidate basis vector can attain; and (ii) minimize a smooth spectral dispersion function over the achievable frequency range to obtain the desired spread GFT basis. Both steps involve non-convex, orthonormality-constrained optimization problems, which are efficiently tackled via a provably convergent, feasible optimization method on the Stiefel manifold. We illustrate the effectiveness of the novel GFT construction algorithm through numerical tests on synthetic and real-world graphs.

***Index Terms*—** Graph signal processing, graph Fourier transform, directed graphs, Stiefel manifold optimization, total variation.

1. INTRODUCTION

Network data indexed by the nodes of a graph are becoming increasingly ubiquitous, with examples ranging from measurements of neural activities at different regions of the brain [1, 2], to economic activity observed over a network of production flows between industrial sectors [3]. It is only natural that complex signals with irregular structure become of interest, and the goal of graph signal processing (GSP) is to develop algorithms that leverage this relational structure; see [4, 5] for tutorial treatments. From this vantage point, signal processing tasks such as filtering [2, 5–8], sampling and reconstruction [3, 9, 10], spectrum estimation [11], (blind) filter identification [12, 13], as well as signal representations [14, 15], have been reexamined under the purview of GSP.

An instrumental GSP tool is the graph Fourier transform (GFT), which decomposes a graph signal into orthonormal components describing different modes of variation with respect to the graph topology. Here we aim to generalize the GFT to directed graphs (digraphs); see also [16–18]. We build on a novel notion of signal variation (frequency) over digraphs and find the maximum possible frequency ($f_{\max}$) that a unit-norm graph signal can achieve. We design a digraph (D)GFT such that the resulting frequencies (i.e., the directed variation of the sought orthonormal bases) distribute as evenly as possible across $[0, f_{\max}]$. Beyond offering parsimonious representations of slowly-varying signals on digraphs, a DGFT with spread frequency components can facilitate more interpretable frequency analyses and aid filter design in the graph spectral domain.

To position our contributions in the context of related work, we first introduce some basic GSP notions and terminology. We consider a weighted digraph $\mathcal{G} = (\mathcal{V}, \mathbf{A})$, where $\mathcal{V}$ is the set of nodes (i.e., vertices) with cardinality $|\mathcal{V}| = N$, and $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the graph adjacency matrix with entry A_{ij} denoting the edge weight from node i to node j. We assume that the graph is connected and has no self loops; i.e. $A_{ii} = 0$, and the edge weights are non-negative ($A_{ij} \geq 0$). For an undirected graph $\mathbf{A}$ is symmetric, and the positive semi-definite combinatorial Laplacian matrix is $\mathbf{L} := \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is the diagonal degree matrix with $D_{ii} = \sum_j A_{ji}$. A graph signal $\mathbf{x} : \mathcal{V} \mapsto \mathbb{R}^N$ can be represented as a vector of size N, where component x_i denotes the signal value at node $i \in \mathcal{V}$.

Related work. For undirected graphs, the GFT of signal $\mathbf{x}$ is often defined as $\tilde{\mathbf{x}} = \mathbf{V}^T\mathbf{x}$, where $\mathbf{V} := [\mathbf{v}_1, \ldots, \mathbf{v}_N]$ comprises the eigenvectors of the Laplacian [4, 8]. Defining the total variation of the signal $\mathbf{x}$ with respect to the Laplacian $\mathbf{L}$ as

$$\text{TV}(\mathbf{x}) = \mathbf{x}^T\mathbf{L}\mathbf{x} = \sum_{i,j=1,j>i}^{N} A_{ij}(x_i - x_j)^2 \tag{1}$$

then it follows that the total variation of eigenvector $\mathbf{v}_k$ is $\text{TV}(\mathbf{v}_k) = \lambda_k$, the k^{th} Laplacian eigenvalue. Hence, eigenvalues $0 = \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_N$ can be viewed as graph frequencies, indicating how the GFT bases vary over the graph. Note that there may be more than one eigenvector corresponding to a graph frequency in case of having repeated eigenvalues. Extensions of the combinatorial Laplacian to digraphs have also been proposed [19]. However, eigenvectors of the directed Laplacian generally fail to yield spread frequency components as we illustrate in Section 4. 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 frequency representation of graph signal $\mathbf{x}$ is $\tilde{\mathbf{x}} = \mathbf{V}^{-1}\mathbf{x}$ [17]. While valid for digraphs, the associated notion of signal variation in [17] does not ensure that constant signals have zero variation. Moreover, $\mathbf{V}$ is not necessarily orthonormal and Parseval's identity does not hold. From a computational standpoint, obtaining the Jordan decomposition is expensive and often numerically unstable; see also [20]. Recently, a fresh look to the GFT for digraphs was put forth in [18] based on minimization of the (convex) Lovász extension of the graph cut size, subject to orthonormality constraints on the desired bases. The optimization procedure in [18] is computationally expensive due to repeated singular value decompositions, and can be infeasible for large graphs. Also, the definition of cut size (and its Lovász extension which can be interpreted as a graph directed variation measure) is based on a bipartition of the graph, while the network may have multiple (more than two) clusters. While the GFT bases in [18] tend to be constant across clusters of the graph, in general they may fail to yield signal representations capturing different modes of signal variation with respect to $\mathcal{G}$; see [16, Remark 1] for an example of this phenomenon.

Contributions. Here we design a DGFT with the following desirable properties: P1) The bases offer notions of frequency and signal variation over digraphs which are also consistent with those used for subsumed undirected graphs. P2) Frequencies are designed to be (approximately) equidistributed in $[0, f_{\max}]$, to better capture low, middle, and high frequencies. P3) Bases are orthonormal so Parse-

Work in this paper was supported by the NSF grants CCF-1750428, CCF-1331863, CCF-1350823, and CCF-1733832.

val's identity holds and signal energy is preserved in the vertex and graph frequency domains. Moreover, the inverse DGFT can be easily computed via inner products. In [16] we made a first attempt to construct a DGFT basis with the aforementioned properties P1)-P3), through a greedy basis selection approach with performance guarantees. Here instead we leverage a feasible method for optimization with orthonormality constraints (outlined in Section 3), to directly minimize a smooth measure of spectral dispersion over the Stiefel manifold formulated in Section 2. The novel algorithm is scalable and offers convergence guarantees to stationary points.

2. PRELIMINARIES AND PROBLEM STATEMENT

In this section we first introduce an extended notion of signal variation and graph frequencies for digraphs. Then, we state the problem as one of finding orthonormal bases with evenly distributed frequencies in the graph spectral domain.

Signal variation on digraphs. Our goal is to find N orthonormal bases capturing different modes of variation over the graph $\mathcal{G}$. We collect these desired bases in a matrix $\mathbf{U} := [\mathbf{u}_1, \ldots, \mathbf{u}_N] \in \mathbb{R}^{N\times N}$, where $\mathbf{u}_k \in \mathbb{R}^N$ represents the kth frequency component. For undirected graphs, the quantity $\mathrm{TV}(\mathbf{x})$ in (1) measures how signal $\mathbf{x}$ varies over the network with Laplacian $\mathbf{L}$. This motivates defining a more general notion of signal variation for digraphs, called directed variation (DV), as (cf. [16, eq. (2)]; see also [18])

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

where $[x]_+ := \max(0, x)$. To gain insights on (2), consider a graph signal $\mathbf{x} \in \mathbb{R}^N$ on digraph $\mathcal{G}$ and suppose a directed edge represents the direction of signal flow from a larger value to a smaller one. Thus, an edge from node i to node j (i.e., $A_{ij} > 0$) contributes to $\mathrm{DV}(\mathbf{x})$ only if $x_i > x_j$. Notice that if $\mathcal{G}$ is undirected, then $\mathrm{DV}(\mathbf{x}) \equiv \mathrm{TV}(\mathbf{x})$. Analogously to the undirected case, we define the frequency $f_k := \mathrm{DV}(\mathbf{u}_k)$ as the directed variation of the basis $\mathbf{u}_k$.

Challenges facing spread frequency components. Similar to the discrete spectrum of periodic time-varying signals, by designing the bases we would ideally like to have N equidistributed graph frequencies forming an arithmetic sequence

$$f_k = \mathrm{DV}(\mathbf{u}_k) = \frac{k-1}{N-1} f_{\max}, \quad k = 1, \ldots, N \tag{3}$$

where $f_{\max}$ is the maximum variation of a unit-norm signal on $\mathcal{G}$.

However, attaining the exact frequencies in (3) may be impossible for irregular graph domains. This can be clearly seen for undirected graphs, where one has the additional constraint that the summation of frequencies is constant, since $\sum_{k=1}^N f_k = \sum_{k=1}^N \mathrm{TV}(\mathbf{u}_k) = \mathrm{trace}(\mathbf{L})$.

Moreover, one needs to determine the maximum frequency $f_{\max}$ that a unit-norm basis can attain. For undirected graphs, one has

$$f_{\max}^u = \max_{\|\mathbf{u}\|=1} \mathrm{TV}(\mathbf{u}) = \max_{\|\mathbf{u}\|=1} \mathbf{u}^T\mathbf{L}\mathbf{u} = \lambda_{\max}, \tag{4}$$

where $\lambda_{\max}$ is the largest eigenvalue of the Laplacian matrix $\mathbf{L}$. However, finding the maximum *directed variation* is in general challenging, since one needs to solve the (non-convex) spherically-constrained problem

$$\mathbf{u}_{\max} = \operatorname*{argmax}_{\|\mathbf{u}\|=1} \mathrm{DV}(\mathbf{u}) \quad \text{and} \quad f_{\max} := \mathrm{DV}(\mathbf{u}_{\max}). \tag{5}$$

Notice that $f_{\max}$ is upper-bounded by $\lambda_{\max}$. This is because dropping the direction of any edge can not decrease the directed variation.

Going back to the design of $\mathbf{U}$, to cover the whole spectrum of variations one would like to set $\mathbf{u}_1 = \mathbf{u}_{\min} := \frac{1}{\sqrt{N}}\mathbf{1}_N$ (normalized all ones vector of length N) and $\mathbf{u}_N = \mathbf{u}_{\max}$. Next we state a criterion for the design of the remaining bases.

Problem statement. Consider the spectral dispersion function

$$\delta(\mathbf{U}) := \sum_{i=1}^{N-1} [\mathrm{DV}(\mathbf{u}_{i+1}) - \mathrm{DV}(\mathbf{u}_i)]^2 \tag{6}$$

that measures how well spread the corresponding frequencies are over $[0, f_{\max}]$. Having fixed the first and last columns of $\mathbf{U}$, it follows that the dispersion function $\delta(\mathbf{U})$ is minimized when the free DV values are selected to form an arithmetic sequence between $\mathrm{DV}(\mathbf{u}_1) = 0$ and $\mathrm{DV}(\mathbf{u}_N) = f_{\max}$, consistent with our design goal.

All in all, rather than going after frequencies exactly equidistributed as in (3), our idea is to minimize the spectral dispersion

$$\begin{aligned} \min_{\mathbf{U}} \ & \sum_{i=1}^{N-1} [\mathrm{DV}(\mathbf{u}_{i+1}) - \mathrm{DV}(\mathbf{u}_i)]^2 \\ \text{subject to} \ & \mathbf{U}^T\mathbf{U} = \mathbf{I} \\ & \mathbf{u}_1 = \mathbf{u}_{\min} \\ & \mathbf{u}_N = \mathbf{u}_{\max}. \end{aligned} \tag{7}$$

Problem (7) is feasible since one can show that $\mathbf{u}_{\max}$ [cf. (5)] is orthogonal to the constant vector $\mathbf{u}_{\min} = \frac{1}{\sqrt{N}}\mathbf{1}_N$[1]. But finding the global optimum of (7) is challenging due to the non-convexity arising from the orthonormality (a.k.a. Stiefel manifold) constraints. The objective function $\delta(\mathbf{U})$ is smooth though, and so there is hope of finding good stationary solutions by leveraging recent advances in manifold optimization. In the next section we build on an effective feasible method for optimization with orthogonality constraints [22], to solve judiciously modified forms of problems (5) and (7) to directly find the maximum frequency along with the disperse bases.

3. MINIMIZING DISPERSION IN A STIEFEL MANIFOLD

Here we show how to find a disperse Fourier basis for signals on digraphs, by bringing to bear a feasible method for optimization of differentiable functions over the Stiefel manifold [22]. Specifically, we take a two step approach whereby: i) we find $f_{\max}$ and its corresponding basis $\mathbf{u}_{\max}$ by solving (5); and ii) we solve (7) to find well-spread frequency components $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_N]$ satisfying $\mathbf{u}_1 = \mathbf{u}_{\min} = \frac{1}{\sqrt{N}}\mathbf{1}_N$ and $\mathbf{u}_N = \mathbf{u}_{\max}$. Similar feasible methods have been also successfully applied to a wide variety of applications, such as low-rank matrix approximations, Independent Component Analysis, and subspace tracking, to name a few [23].

The general iterative method of [22] deals with an orthogonality constrained problem of the form

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times p}} \mathcal{F}(\mathbf{U}), \quad \text{subject to} \quad \mathbf{U}^T\mathbf{U} = \mathbf{I}, \tag{8}$$

where $\mathcal{F}(\mathbf{U}) : \mathbb{R}^{n\times p} \to \mathbb{R}$ is assumed to be differentiable, just like $\delta(\mathbf{U})$ in (6). Given a feasible point $\mathbf{U}_k$ at iteration k and gradient $\mathbf{G}_k = \nabla\mathcal{F}(\mathbf{U}_k)$, one follows the update rule

$$\mathbf{U}_{k+1}(\tau) = \left(\mathbf{I} + \frac{\tau}{2}\mathbf{B}_k\right)^{-1}\left(\mathbf{I} - \frac{\tau}{2}\mathbf{B}_k\right)\mathbf{U}_k, \tag{9}$$

where $\mathbf{B}_k := \mathbf{G}_k\mathbf{U}_k^T - \mathbf{U}_k\mathbf{G}_k^T$ is a skew-symmetric ($\mathbf{B}_k^T = -\mathbf{B}_k$) projection of the gradient onto the constraint's tangent space. Update rule (9) in known as the Cayley transform which preserves orthogonality ($\mathbf{U}_{k+1}^T\mathbf{U}_{k+1} = \mathbf{I}$), since $(\mathbf{I}+\frac{\tau}{2}\mathbf{B}_k)^{-1}$ and $\mathbf{I}-\frac{\tau}{2}\mathbf{B}_k$ commute.

[1] A proof is omitted here due to lack of space, details can be found in [21].

Algorithm 1 Feasible Method for Spectral Dispersion Minimization

Input: Adjacency matrix $\mathbf{A}$, regularization parameter $\lambda > 0$ and tolerance $\epsilon > 0$.
Find $\mathbf{u}_{\max}$ (details in [21, Alg. 1]) and set $\mathbf{u}_{\min} = \frac{1}{\sqrt{N}}\mathbf{1}_N$
Initialize $k = 0$ and orthonormal $\mathbf{U}_0$ at random.
repeat
  Compute gradient $\mathbf{G}_k \in \mathbb{R}^{N\times N}$ from (11).
  Set $\mathbf{B}_k = \mathbf{G}_k {\mathbf{U}_k}^T - \mathbf{U}_k {\mathbf{G}_k}^T$.
  Find τ_k chosen via the Armijo-Wolfe conditions.
  Update $\mathbf{U}_{k+1}(\tau_k) = (\mathbf{I} + \frac{\tau_k}{2}\mathbf{B}_k)^{-1}(\mathbf{I} - \frac{\tau_k}{2}\mathbf{B}_k)\mathbf{U}_k$.
  $k \leftarrow k + 1$.
until $\|\mathbf{U}_k - \mathbf{U}_{k-1}\|_F \leq \epsilon$
Return $\hat{\mathbf{U}} = \mathbf{U}_k$.

Noteworthy properties are: i) $\mathbf{U}_{k+1}(0) = \mathbf{U}_k$; ii) $\mathbf{U}_{k+1}(\tau)$ in (9) is smooth in τ; and iii) $\frac{d}{d\tau}\mathbf{U}_{k+1}(\tau)$ is the projection of $-\mathbf{G}_k$ into $\mathbf{B}_k$ at $\mathbf{U}_k$. Most importantly, iii) ensures that the update (9) is a descent path for a proper step size τ; see [22] for more details. In particular, one such step size τ can be obtained through a curvilinear search satisfying the Armijo-Wolfe conditions [24]. Theorem 2 in [22] asserts that the overall procedure converges to a stationary point of $\mathcal{F}(\mathbf{U})$, while generating feasible points at every iteration.

Finding $\mathbf{u}_{\max}$ and $f_{\max}$. As the first step to find the DGFT bases, we obtain $f_{\max}$ by using the feasible approach to minimize $-\mathrm{DV}(\mathbf{u})$ over $\{\mathbf{u} \mid \mathbf{u}^T\mathbf{u} = 1\}$ [cf. (5)]. The gradient $\bar{\mathbf{g}} := \frac{\partial}{\partial \mathbf{u}}\mathrm{DV}(\mathbf{u}) \in \mathbb{R}^N$ has entries given by ($\mathbf{A}_{\cdot i}$ is the ith column of $\mathbf{A}$, $\mathbf{A}_{i\cdot}$ the ith row)

$$\bar{g}_i = 2\left[\mathbf{A}_{\cdot i}^T(\mathbf{u} - u_i\mathbf{1}_N)_+ - \mathbf{A}_{i\cdot}(u_i\mathbf{1}_N - \mathbf{u})_+\right], \quad 1 \leq i \leq N.$$

The algorithm initiates at a random unit-norm vector and then via (9) it takes a descent path towards a stationary point (detailed iterations can be found in [21, Alg. 1]). It is often prudent to run the algorithm multiple times using random initializations, and retain the solution that yields the least cost. Although the algorithm only guarantees converges to stationary points, in practice it tends to find $f_{\max} = \mathrm{DV}(\mathbf{u}_{\max})$ exactly if the number of initializations is chosen properly; see Section 4. The limiting basis $\mathbf{u}_{\max}$ is then used for the spectral dispersion minimization step.

Spectral dispersion minimization. As the second and final step, we aim at finding the orthonormal basis $\mathbf{U}$ that minimizes the spectral dispersion (6). To write the optimization problem (7) in the form of (8) and apply the previously outlined feasible method, we penalize the objective with a measure of the constraint violations to obtain

$$\min_{\mathbf{U}} \delta(\mathbf{U}) + \frac{\lambda}{2}\left(\|\mathbf{u}_1 - \mathbf{u}_{\min}\|_2^2 + \|\mathbf{u}_N - \mathbf{u}_{\max}\|_2^2\right)$$
$$\text{subject to} \quad \mathbf{U}^T\mathbf{U} = \mathbf{I}, \tag{10}$$

where λ is made large enough to ensure $\mathbf{u}_1 = \mathbf{u}_{\min}$ and $\mathbf{u}_N = \mathbf{u}_{\max}$. The overall procedure is tabulated under Algorithm 1, where the gradient matrix $\mathbf{G} := \frac{\partial}{\partial \mathbf{U}}\delta(\mathbf{U}) \in \mathbb{R}^{N\times N}$ has columns given by

$$\begin{aligned}
\mathbf{g}_1 &= [\mathrm{DV}(\mathbf{u}_1) - \mathrm{DV}(\mathbf{u}_2)]\,\bar{\mathbf{g}}(\mathbf{u}_1) + \lambda(\mathbf{u}_1 - \mathbf{u}_{\min}),\\
\mathbf{g}_i &= [-\mathrm{DV}(\mathbf{u}_{i+1}) + 2\mathrm{DV}(\mathbf{u}_i) - \mathrm{DV}(\mathbf{u}_{i-1})]\,\bar{\mathbf{g}}(\mathbf{u}_i),\ i \in [2, N-1]\\
\mathbf{g}_N &= [\mathrm{DV}(\mathbf{u}_{N-1}) - \mathrm{DV}(\mathbf{u}_N)]\,\bar{\mathbf{g}}(\mathbf{u}_N) + \lambda(\mathbf{u}_N - \mathbf{u}_{\max}).
\end{aligned} \tag{11}$$

Once more, it is convenient to run the algorithm multiple times and retain the least disperse DGFT basis.

4. NUMERICAL RESULTS

To assess the performance of the proposed algorithm, we construct the GFT bases for two different digraphs and compare their resulting modes of variation with existing approaches.

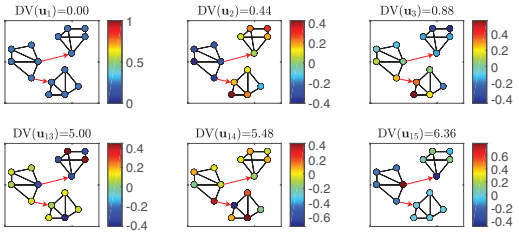

Fig. 1: First and last three DGFT bases obtained using Algorithm 1. Respective DV values (frequencies) are shown as well.

Method	Dispersion
Directed Laplacian [19]	0.256
PAMAL [18]	0.301
Submodular Greedy Algorithm [16]	0.118
Feasible Method (Alg. 1)	**0.076**

Table 1: Spectral dispersion $\delta(\mathbf{U})$ of obtained bases $\mathbf{U}$ using different algorithms for the synthetic graph shown in Fig. 1.

Synthetic graph. First, using Algorithm 1 we construct the DGFT for an unweighted digraph $\mathcal{G}$ with $N = 15$ nodes shown in Fig. 1, and compare it with the GFT given in [18], the eigenvectors of the directed Laplacian defined in [19], and the DGFT obtained by submodular greedy optimization in [16]. To define the directed Laplacian in [19], consider a random walk on the graph with transition probability matrix $\mathbf{P} = \mathbf{D}_{\text{out}}^{-1}\mathbf{A}$, where $\mathbf{D}_{\text{out}}$ is the diagonal matrix of node out-degrees. Let $\mathbf{\Pi} = \mathrm{diag}(\boldsymbol{\pi})$ be the diagonal matrix with the stationary distribution $\boldsymbol{\pi}$ of the random walk on the diagonal.[2] Using these definitions, the directed Laplacian is given by

$$\mathbf{Q} = \mathbf{\Pi} - \frac{\mathbf{\Pi P} + \mathbf{P}^T\mathbf{\Pi}}{2}. \tag{12}$$

Fig. 1 shows the results from Algorithm 1, after finding the last basis $\mathbf{u}_{\max}$ through a similar procedure [i.e. feasible method for (5)]. Each subplot shows one basis vector (column) of the resulting $\mathbf{U}$, and the corresponding DV is calculated using (2). It is apparent that the first bases exhibit significantly less variability than the three higher frequency components.

To corroborate that the resulting DGFT bases are well distributed in the graph spectral domain, Fig. 2a depicts the distribution of all the frequencies for the four examined algorithms. In Fig. 2a, each vertical line indicates the directed variation (frequency) associated with a basis. As expected, the proposed approach – which directly optimizes the spectral dispersion metric – yields a set of almost equidistributed graph frequencies. As additional figure of merit, we first rescale the DV values to the $[0, 1]$ interval and calculate their dispersion using (6). The resulting values are reported in Table 1, which confirms that Algorithm 1 yields a better frequency spread (i.e., smaller dispersion). While other approaches may find repeated frequencies (as can be inferred from the total number of vertical lines in Fig. 2a), the proposed DGFT construction returns N well dispersed frequencies.

We also use the Monte-Carlo method to study the convergence properties of our algorithms. In Fig. 3 (top) we show the evolution of iterates for the feasible method in [22], when used to find the maximum DV (i.e., $f_{\max}$) for the same 15-node graph. We do so for 100 different (random) initializations and report the median as well as the first and third quartiles versus the number of iterations. We

[2] Vector π is a stationary distribution for the random walk if $\pi = \mathbf{P}^T\pi$.

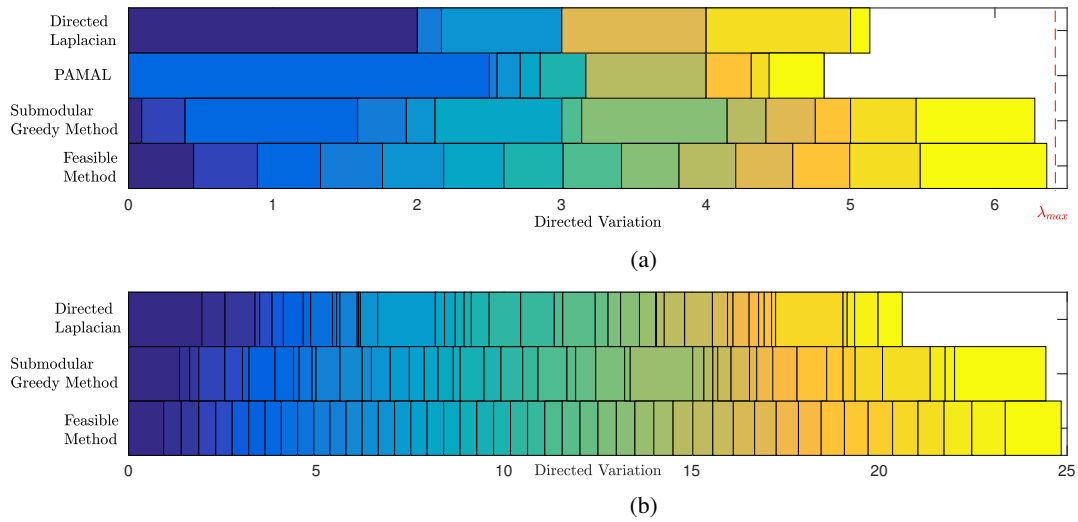

Fig. 2: (a) Comparison of directed variations for different methods in the synthetic graph in Fig. 1: eigenvectors of the directed Laplacian matrix introduced in [19], augmented Lagrangian method of [18] (PAMAL), submodular greedy algorithm in [16], and feasible method (Algorithm 1). Colored boxes show the difference between two consecutive frequencies for each method. Apparently, Algorithm 1 yields the most equidistributed graph frequencies. (b) Comparison of directed variations in the brain graph: similar to (a) except for the PAMAL algorithm which did not converge in a feasible time.

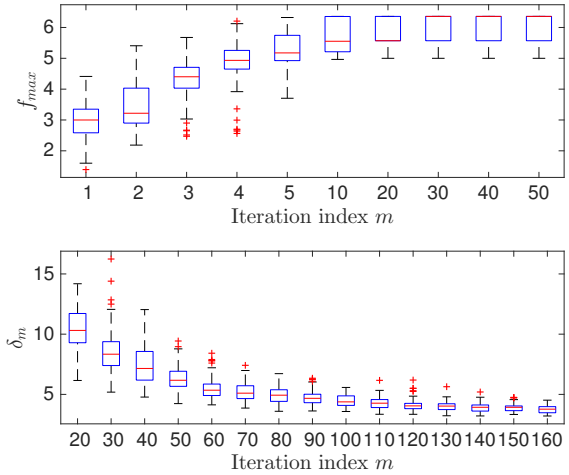

Fig. 3: (top) Convergence behaviour of the proposed feasible method for finding the maximum directed variation $f_{\max}$. The boxes show the median and the 25th and 75th percentiles of $f_{\max}$ vs. the number of iterations, obtained by running 100 Monte-Carlo simulations based on independent initializations. (bottom) Likewise, but when using Algorithm 1 to minimize the spectral dispersion $\delta(\mathbf{U})$ in (6).

observe that all the realizations converge [22, Theorem 2], but there is a small variation among the limiting values. This is expected because the feasible method is not guaranteed to converge to the global optimum of the non-convex problem (7). It is worth mentioning that after about 10 iterations, the exact value of $f_{\max}$ is achieved by a quarter of the realizations (and this improves to half of the realizations with about 30 iterations). Similarly, Fig. 3 (bottom) shows the median, first, and third quartiles of the dispersion function iterates $\delta(\mathbf{U}_k)$, when minimized using Algorithm 1. Again, 100 different Monte-Carlo simulations are considered and we observe that all of them converge to limiting values with small variability. This means that in practice, we can run Algorithm 1 with different random initializations and retain the most spread frequency components.

Brain graph. We also consider a real brain graph representing the anatomical connections of the macaque cortex, which was studied e.g. in [1,25]. The network consists of $N = 47$ nodes and 505 edges (among which 121 links are directed). The vertices represent different hubs in the brain, and the edges capture directed information flow among them. Fig. 2b compares the distribution of frequencies for the different methods, except for the PAMAL algorithm which did not converge within a reasonable time. Again, our proposed algorithm outperforms other approaches in terms of finding well dispersed and non-repetitive frequencies, which demonstrates its potential effectiveness for filtering of brain signals and enhanced interpretability of graph frequency analyses [2].

5. CONCLUSION

We considered the problem of finding an orthonormal set of graph Fourier bases for digraphs, which is also naturally applicable to undirected networks. First, we introduced a measure of directed variation to capture the notion of frequency on digraphs. Our DGFT design is to construct orthonormal bases that span the entire frequency range and for which frequency components are as evenly distributed as possible while taking into account the network structure. To that end, we defined a spectral dispersion function to quantify the quality of any feasible solution compared to our ideal design, and minimized this criterion over the Stiefel manifold. To tackle the resulting non-convex problems, we used a feasible method for optimization with orthogonality constraints, which offers provable convergence guarantees to the stationary points. The overall pipeline for finding desirable bases is validated on synthetic and structural brain networks.

With regards to future directions, the complexity of finding the maximum frequency ($f_{\max}$) on a digraph is an interesting open question. If NP-hard, it will be interesting to find the best achievable approximation factor (a $1/2$-approximation was given in [16]). Furthermore, it would be a significant improvement if we can bound the optimality gap for the stationary solution of the proposed algorithm.

6. REFERENCES

[1] Christopher J. Honey, Rolf Kötter, Michael Breakspear, and Olaf Sporns, "Network structure of cerebral cortex shapes functional connectivity on multiple time scales," *Proceedings of the National Academy of Sciences*, vol. 104, no. 24, pp. 10240–10245, 2007.

[2] Weiyu Huang, Leah Goldsberry, Nicholas F. Wymbs, Scott T. Grafton, Danielle S. Bassett, and Alejandro Ribeiro, "Graph frequency analysis of brain signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 7, pp. 1189–1203, 2016.

[3] Antonio G. Marques, Santiago Segarra, Geert Leus, and Alejandro Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1832–1843, 2016.

[4] David I. Shuman, Sunil K. Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[5] Aliaksei Sandryhaila and José M. F. Moura, "Discrete signal processing on graphs," *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1644–1656, 2013.

[6] Oguzhan Teke and Palghat P. Vaidyanathan, "Extending classical multirate signal processing theory to graphs-Part I: Fundamentals," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 409–422, 2017.

[7] Oguzhan Teke and Palghat P. Vaidyanathan, "Extending classical multirate signal processing theory to graphs-Part II: M-channel filter banks," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 423–437, 2017.

[8] Nicolas Tremblay, Paulo Gonçalves, and Pierre Borgnat, "Design of graph filters and filterbanks," 2017.

[9] Siheng Chen, Rohan Varma, Aliaksei Sandryhaila, and Jelena Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6510–6523, 2015.

[10] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3462–3477, July 2017.

[11] Antonio G. Marques, Santiago Segarra, Geert Leus, and Alejandro Ribeiro, "Stationary graph processes and spectral estimation," *IEEE Transactions on Signal Processing*, vol. 65, no. 22, pp. 5911–5926, Nov. 2017.

[12] Rasoul Shafipour, Santiago Segarra, Antonio G. Marques, and Gonzalo Mateos, "Network topology inference from non-stationary graph signals," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[13] Santiago Segarra, Gonzalo Mateos, Antonio G. Marques, and Alejandro Ribeiro, "Blind identification of graph filters," *IEEE Transactions on Signal Processing*, vol. 65, no. 5, pp. 1146–1159, 2017.

[14] Dorina Thanou, David I. Shuman, and Pascal Frossard, "Learning parametric dictionaries for signals on graphs," *IEEE Transactions on Signal Processing*, vol. 62, no. 15, pp. 3849–3862, 2014.

[15] Xiaofan Zhu and Michael Rabbat, "Approximating signals supported on graphs," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 3921–3924.

[16] Rasoul Shafipour, Ali Khodabakhsh, Gonzalo Mateos, and Evdokia Nikolova, "A digraph Fourier transform with spread frequency components," in *Proc. of IEEE Global Conf. on Signal and Information Processing*, Nov 2017.

[17] Aliaksei Sandryhaila and José M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3042–3054, June 2014.

[18] Stefania Sardellitti, Sergio Barbarossa, and Paolo Di Lorenzo, "On the graph fourier transform for directed graphs," *J. Sel. Topics Signal Processing*, vol. 11, no. 6, pp. 796–811, 2017.

[19] Fan Chung, "Laplacians and the cheeger inequality for directed graphs," *Annals of Combinatorics*, vol. 9, no. 1, pp. 1–19, 2005.

[20] Joya A. Deri and José M. F. Moura, "Spectral projector-based graph fourier transforms," *J. Sel. Topics Signal Processing*, vol. 11, no. 6, pp. 785–795, 2017.

[21] Rasoul Shafipour, Ali Khodabakhsh, Gonzalo Mateos, and Evdokia Nikolova, "A graph Fourier transform with spread spectral components," *IEEE Transactions on Signal Processing (submitted)*, 2017.

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

[23] P-A Absil, Robert Mahony, and Rodolphe Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, 2009.

[24] Nocedal Jorge and J Wright Stephen, "Numerical Optimization," 2006.

[25] Mikail Rubinov and Olaf Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.