

# A Directed Graph Fourier Transform With Spread Frequency Components

Rasoul Shafipour<sup>✉</sup>, *Student Member, IEEE*, Ali Khodabakhsh<sup>✉</sup>, *Student Member, IEEE*,  
Gonzalo Mateos<sup>✉</sup>, *Senior Member, IEEE*, and Evdokia Nikolova

**Abstract**—We study the problem of constructing a graph Fourier transform (GFT) for directed graphs (digraphs), which decomposes graph signals into different modes of variation with respect to the underlying network. Accordingly, to capture low, medium, and high frequencies we seek a digraph (D)GFT such that the orthonormal frequency components are as spread as possible in the graph spectral domain. To that end, we advocate a two-step design whereby we 1) find the maximum directed variation (i.e., a novel notion of frequency on a digraph) a candidate basis vector can attain and 2) minimize a smooth spectral dispersion function over the achievable frequency range to obtain the desired spread DGFT basis. Both steps involve non-convex, orthonormality-constrained optimization problems, which are efficiently tackled via a feasible optimization method on the Stiefel manifold that provably converges to a stationary solution. We also propose a heuristic to construct the DGFT basis from Laplacian eigenvectors of an undirected version of the digraph. We show that the spectral-dispersion minimization problem can be cast as supermodular optimization over the set of candidate frequency components, whose orthonormality can be enforced via a matroid basis constraint. This motivates adopting a scalable greedy algorithm to obtain an approximate solution with quantifiable worst-case spectral dispersion. We illustrate the effectiveness of our DGFT algorithms through numerical tests on synthetic and real-world networks. We also carry out a graph-signal denoising task, whereby the DGFT basis is used to decompose and then low pass filter temperatures recorded across the United States.

**Index Terms**—Graph signal processing, graph Fourier transform, directed variation, greedy algorithm, Stiefel manifold.

## I. INTRODUCTION

NETWORK processes such as neural activities at different regions of the brain [13], vehicle trajectories over road networks [6], or, infectious states of individuals in a population

affected by an epidemic [17], can be represented as graph signals supported on the vertices of the adopted graph abstraction to the network. Under the natural assumption that the signal properties relate to the underlying graph (e.g., when observing a network diffusion or percolation process), the goal of graph signal processing (GSP) is to develop algorithms that fruitfully exploit this relational structure [26], [29]. From this vantage point, generalizations of traditional signal processing tasks such as filtering [15], [29], [33], [40], [42], sampling and reconstruction [4], [21], statistical GSP and spectrum estimation [11], [22], [27], (blind) filter identification [34], [37] as well as signal representations [38], [41], [44] have been recently explored under the GSP purview [26].

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. The GFT provides a method to equivalently represent a graph signal in two different domains – the graph domain, consisting of the nodes in the graph, and the frequency domain, represented by the frequency components of the graph. Therefore, signals can be manipulated in the frequency domain to induce different levels of interactions between neighbors in the network. Here we aim to generalize the GFT to directed graphs (digraphs); see also [30], [31]. We first propose a novel notion of signal variation (frequency) over digraphs and find an approximation of 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 basis vectors) 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 spectral domain. In a way, to achieve these goals we advocate a form of regularity in the DGFT-induced frequency domain. A different perspective is to consider an irregular (dual) graph support, which as argued in [20] can offer complementary merits and insights.

**Related work:** To position our contributions in context, we first introduce some basic GSP notions and terminology. 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 edge weights in  $\mathcal{G}$  are non-negative ( $A_{ij} \geq 0$ ). For an undirected graph,  $\mathbf{A}$  is symmetric, and the positive semi-definite Laplacian matrix

Manuscript received April 6, 2018; revised September 4, 2018 and November 16, 2018; accepted November 27, 2018. Date of publication December 10, 2018; date of current version January 4, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Pierre Borgnat. This work was supported by the NSF awards under Grant CCF-1750428, Grant ECCS-1809356, Grant CCF-1216103, Grant CCF-1331863, Grant CCF-1350823, and Grant CCF-1733832. This paper was presented in part at the Fifth IEEE Global Conference on Signal and Information Processing, Montreal, QC, Canada, November 2017 and in part at the Forty-Third IEEE International Conference on Acoustics, Speech, and Signal Processing, Calgary, AB, Canada, April 2018. (*Corresponding author: Rasoul Shafipour*.)

R. Shafipour and G. Mateos are with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627 USA (e-mail: rshafipo@ur.rochester.edu; gmateosb@ece.rochester.edu).

A. Khodabakhsh and E. Nikolova are with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712 USA (e-mail: ali.kh@utexas.edu; nikolova@austin.utexas.edu).

Digital Object Identifier 10.1109/TSP.2018.2886151

takes the form  $\mathbf{L} := \mathbf{D} - \mathbf{A}$ , 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 length  $N$ , where entry  $x_i$  denotes the signal value at node  $i \in \mathcal{V}$ .

For undirected graphs, the GFT of signal  $\mathbf{x}$  is often defined as  $\tilde{\mathbf{x}} = \mathbf{V}^T \mathbf{x}$ , where  $\mathbf{V} := [\mathbf{v}_1, \dots, \mathbf{v}_N]$  comprises the eigenvectors of the Laplacian [12], [26], [42], [44]. Interestingly, in this setting the GFT encodes a notion of signal variability over the graph akin to the notion of frequency in Fourier analysis of temporal signals. To understand this analogy, define the total variation of the signal  $\mathbf{x}$  with respect to the Laplacian  $\mathbf{L}$  as

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

The total variation  $\text{TV}(\mathbf{x})$  is a smoothness measure, quantifying how much the signal  $\mathbf{x}$  changes with respect to the expectation on variability that is encoded by the weights  $\mathbf{A}$ . Consider the total variation of the eigenvectors  $\mathbf{v}_k$ , which is given by  $\text{TV}(\mathbf{v}_k) = \lambda_k$ , the  $k$ th Laplacian eigenvalue. Hence, eigenvalues  $0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_N$  can be viewed as graph frequencies, indicating how the GFT basis vectors (i.e., the frequency components) vary over the graph  $\mathcal{G}$ . 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 [5], [39]. However, eigenvectors of the directed Laplacian generally fail to yield spread frequency components as we illustrate in Section VI. 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  $\tilde{\mathbf{x}} = \mathbf{V}^{-1} \mathbf{x}$  [30]. While valid for digraphs, the associated notion of signal variation in [30] 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 often numerically unstable; see also [7] and [10] for recent stabilizing alternatives. On a related note, a class of energy-preserving graph-shift operators with desirable properties are constructed in [9]. Starting from the adjacency matrix of an arbitrary graph, the idea is to preserve eigenvectors (hence resulting in the same GFT as in [30]) and replace the eigenvalues with pure phase shifts. Recently, a fresh look at the GFT for digraphs was put forth in [31] based on minimization of the (convex) Lovász extension of the graph cut size, subject to orthonormality constraints on the desired basis. While the GFT basis vectors in [31] 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 Section III-A for an example of this phenomenon. An encompassing GFT framework was proposed in [12], which combines aspects of signal variation and energy to design general orthonormal transforms for graph signals.

*Contributions and paper outline:* Here we design a novel DGFT that has the following desirable properties: P1) The basis graph signals provide notions of frequency and signal variation over digraphs which are also consistent with those typically used for subsumed undirected graphs. P2) Frequency components are designed to be (approximately) equidistributed

in  $[0, f_{\max}]$ , and thus better capture low, middle, and high frequencies. P3) Basis vectors are orthonormal so Parseval's identity holds and inner products are preserved in the vertex and graph frequency domains. Moreover, the inverse DGFT can be easily computed. To formalize our design goal via a well-defined criterion, in Section II we introduce a smooth spectral dispersion function, which measures the spread of the frequencies associated with candidate DGFT basis vectors. Motivation and challenges associated with the proposed spectral-dispersion minimization approach are discussed in Section III. We then propose two algorithmic approaches with complementary strengths, to construct a DGFT basis with the aforementioned properties P1)–P3). We first leverage a provably-convergent feasible method for orthonormality-constrained optimization, to directly minimize the smooth spectral-dispersion cost over the Stiefel manifold (Section IV). In Section V, we propose a DGFT heuristic whereby we restrict the set of candidate frequency components to the (possibly sign reversed) eigenvectors of the Laplacian matrix associated with an undirected version of  $\mathcal{G}$ . In this setting, we show that the spectral-dispersion minimization problem can be cast as supermodular (frequency) set optimization. Moreover, we show that orthonormality can be enforced via a matroid basis constraint, which motivates the adoption of a scalable greedy algorithm to obtain an approximate solution with provable worst-case performance guarantees.

Relative to the conference precursors to this paper [35], [36], here we offer a more comprehensive and unified exposition of the aforementioned algorithmic approaches, comparing their computational complexity and their performance against state-of-the-art GFT methods for digraphs (Section VI). This way, results become clearer and new insights as well as conclusions can be drawn. We include all proofs and technical details missing from [35], [36], as well as new examples and results that shed light on the properties of directed variation, the maximum attainable frequency on a digraph, and the proposed dispersion minimization problem. Numerical tests with both synthetic and real-world digraphs corroborate the effectiveness of our DGFT algorithms in yielding (near) maximally-spread frequency components (Section VI). Concluding remarks and future research directions are outlined in Section VII, while some technical details are deferred to the Appendix.

*Notation:* Bold capital letters refer to matrices and bold lowercase letters represent vectors. The entries of a matrix  $\mathbf{X}$  and a (column) vector  $\mathbf{x}$  are denoted by  $X_{ij}$  and  $x_i$ , respectively. Sets are represented by calligraphic capital letters. The notation  $^T$  stands for transposition and  $\mathbf{I}_N$  represents the  $N \times N$  identity matrix, while  $\mathbf{1}_N$  denotes the  $N \times 1$  vector of all ones. For a vector  $\mathbf{x}$ ,  $\text{diag}(\mathbf{x})$  is a diagonal matrix whose  $i$ th diagonal entry is  $x_i$ . The binary-valued indicator function of a statement  $S$  is denoted by  $\mathbb{I}\{S\}$ , where  $\mathbb{I}\{S\} = 1$  if  $S$  holds true and  $\mathbb{I}\{S\} = 0$  otherwise. Lastly,  $\|\mathbf{x}\| = (\sum_i x_i^2)^{1/2}$  denotes the Euclidean norm of  $\mathbf{x}$ , while  $\text{trace}(\mathbf{X}) = \sum_i X_{ii}$  stands for the trace of a square matrix  $\mathbf{X}$  whose Frobenius norm is  $\|\mathbf{X}\|_F = [\text{trace}(\mathbf{X}\mathbf{X}^T)]^{1/2}$ .

## II. PRELIMINARIES AND PROBLEM STATEMENT

In this section we extend the notion of signal variation to digraphs and accordingly define graph frequencies. We then

state the problem of finding an orthonormal DGFT basis with evenly distributed frequencies in the graph spectral domain.

### A. Signal Variation on Digraphs

As per P1–P3, our goal is to find  $N$  orthonormal vectors capturing different modes of variation over the graph  $\mathcal{G}$ . We collect these desired vectors in a matrix  $\mathbf{U} := [\mathbf{u}_1, \dots, \mathbf{u}_N] \in \mathbb{R}^{N \times N}$ , where  $\mathbf{u}_k \in \mathbb{R}^N$  represents the  $k$ th frequency component. This means that the DGFT of a graph signal  $\mathbf{x} \in \mathbb{R}^N$  with respect to  $\mathcal{G}$  is the signal  $\tilde{\mathbf{x}} = [\tilde{x}_1, \dots, \tilde{x}_N]^T$  defined as  $\tilde{\mathbf{x}} = \mathbf{U}^T \mathbf{x}$ . The inverse (i)DGFT of  $\tilde{\mathbf{x}}$  is given by  $\mathbf{x} = \mathbf{U} \tilde{\mathbf{x}} = \sum_{k=1}^N \tilde{x}_k \mathbf{u}_k$ , which allows one to synthesize  $\mathbf{x}$  as a sum of orthogonal frequency components  $\mathbf{u}_k$ . The contribution of  $\mathbf{u}_k$  to the signal  $\mathbf{x}$  is the DGFT coefficient  $\tilde{x}_k$ .

For undirected graphs, the quantity  $\text{TV}(\mathbf{x})$  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

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

where  $[x]_+ := \max(0, x)$  denotes projection onto the non-negative reals. Notice that  $[x]_+^2 = ([x]_+)^2$ , and for brevity we omit the parenthesis. Similar to the total variation (1), the proposed directed variation measure is non-negative, non-linear, but convex and differentiable. 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  $\text{DV}(\mathbf{x})$  only if  $x_i > x_j$ . Accordingly, one in general has that  $\text{DV}(\mathbf{x}) \neq \text{DV}(-\mathbf{x})$  and we will exploit this property later. Notice that if  $\mathcal{G}$  is undirected, then  $\text{DV}(\mathbf{x}) \equiv \text{TV}(\mathbf{x})$  because from  $A_{ij}[x_i - x_j]_+^2$  and  $A_{ji}[x_j - x_i]_+^2$ , one will be zero and the other one will be equal to  $A_{ij}(x_i - x_j)^2$ .

Analogously to the undirected case, we define the frequency  $f_k := \text{DV}(\mathbf{u}_k)$  as the directed variation of the vector  $\mathbf{u}_k$  (cf. the desired property P1 stated in Section I).

### B. Challenges Facing Spread Frequency Components

Similar to the discrete spectrum of periodic time-varying signals, by designing the basis signals we would ideally like to have  $N$  equidistributed graph frequencies forming an arithmetic sequence (cf. property P2 stated in Section I)

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

where  $f_{\max}$  is the maximum variation attainable by a unit-norm signal on  $\mathcal{G}$ . Such a spread frequency distribution could facilitate more interpretable spectral analyses of graph signals (where it is apparent what low, medium and high frequencies mean), and also aid filter design in the graph spectral domain.

Not surprisingly, finding a DGFT basis 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

$$\sum_{k=1}^N f_k = \sum_{k=1}^N \text{TV}(\mathbf{u}_k) = \text{trace}(\mathbf{U}^T \mathbf{L} \mathbf{U}) = \text{trace}(\mathbf{L}). \quad (4)$$

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

$$f_{\max}^u := \max_{\|\mathbf{u}\|=1} \text{TV}(\mathbf{u}) = \max_{\|\mathbf{u}\|=1} \mathbf{u}^T \mathbf{L} \mathbf{u} = \lambda_{\max}, \quad (5)$$

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

$$\mathbf{u}_{\max} = \underset{\|\mathbf{u}\|=1}{\text{argmax}} \text{DV}(\mathbf{u}) \quad \text{and} \quad f_{\max} := \text{DV}(\mathbf{u}_{\max}). \quad (6)$$

To relate the maximum frequencies in (5) and (6), for a given digraph  $\mathcal{G} = (\mathcal{V}, \mathbf{A})$  consider its *underlying undirected graph*  $\mathcal{G}^u = (\mathcal{V}, \mathbf{A}^u)$ , obtained by replacing all directed edges in  $\mathcal{G}$  with undirected ones via  $A_{ij}^u = A_{ji}^u := \max(A_{ij}, A_{ji})$ . Notice then that  $f_{\max}$  is upper-bounded by  $f_{\max}^u = \lambda_{\max}$ , the spectral radius of the Laplacian of  $\mathcal{G}^u$ . This is formally proved in Proposition 2, but it is essentially because dropping the direction of any edge can not decrease the directed variation of a signal.

### C. Problem Statement

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  $N$ , i.e., a constant signal) and  $\mathbf{u}_N = \mathbf{u}_{\max}$  in (6). As a criterion for the design of the remaining basis vectors, consider the spectral dispersion function

$$\delta(\mathbf{U}) := \sum_{i=1}^{N-1} [\text{DV}(\mathbf{u}_{i+1}) - \text{DV}(\mathbf{u}_i)]^2 \quad (7)$$

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 directed variation values are selected to form an arithmetic sequence between  $\text{DV}(\mathbf{u}_1) = 0$  and  $\text{DV}(\mathbf{u}_N) = f_{\max}$ , consistent with our design goal; recall P2 in Section I.

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

$$\begin{aligned} \min_{\mathbf{U}} \quad & \sum_{i=1}^{N-1} [\text{DV}(\mathbf{u}_{i+1}) - \text{DV}(\mathbf{u}_i)]^2 \\ \text{subject to} \quad & \mathbf{U}^T \mathbf{U} = \mathbf{I}_N, \\ & \mathbf{u}_1 = \mathbf{u}_{\min}, \\ & \mathbf{u}_N = \mathbf{u}_{\max}. \end{aligned} \quad (8)$$

Problem (8) is feasible since we show in Appendix A that  $\mathbf{u}_{\max}$  defined in (6) is orthogonal to the constant vector  $\mathbf{u}_{\min} = \frac{1}{\sqrt{N}} \mathbf{1}_N$ . There is no need to explicitly add a constraint enforcing that basis vectors  $\mathbf{u}_i$  are ordered, meaning that their respective  $\text{DV}(\mathbf{u}_i)$  values increase with  $i$ . The solution of (8) will be ordered, otherwise one could simply sort the candidate directed variation values and decrease the objective. As an alternative definition of graph signal energy, one could adopt



generalized orthonormality constraints  $\mathbf{U}^T \mathbf{M} \mathbf{U} = \mathbf{I}_N$  for some positive definite  $\mathbf{M}$  [12]. In any case, finding the global optimum of (8) is challenging due to the non-convexity arising from the orthonormality (Stiefel manifold) constraints as well as the objective function [due to the cross-terms  $DV(\mathbf{u}_i)DV(\mathbf{u}_{i+1})$ ]. The spectral dispersion  $\delta(\mathbf{U})$  is smooth though, and so there is hope of finding good stationary solutions by bringing to bear recent advances in manifold optimization.

In Section IV we build on a feasible method for optimization with orthogonality constraints [43], to solve judiciously modified forms of problems (6) and (8) to directly find the maximum frequency along with the disperse basis vectors. But before delving into algorithmic solutions, in the next section we expand on the motivation behind spectral dispersion minimization. We also offer additional graph-theoretic insights on the maximum directed variation a unit norm signal can achieve.

### III. ON SPREAD AND MAXIMUM DIGRAPH FREQUENCIES

Here, we further motivate the advocated DGFT design by first showing how state-of-the-art methods may fail to offer signal representations capturing different modes of signal variation with respect to  $\mathcal{G}$ . For undirected graphs where the Laplacian eigenbasis has well documented merits, we then show that the said GFT in general does not minimize the spectral dispersion measure (7). Even for such graphs where  $DV(\mathbf{x}) \equiv TV(\mathbf{x})$ , notice that an identity akin to  $TV(\mathbf{x}) = \sum_{k=1}^N |\tilde{x}_k|^2 TV(\mathbf{u}_k)$  does not hold here. This is because the basis  $\mathbf{U}$  obtained by solving (8) does not involve eigenvectors related to a signal variation operator [12]. In Section III-B we revisit problem (6), namely that of finding the maximum frequency over a given digraph. We first identify some graph families for which the maximum directed variation can be obtained analytically, and then provide a general 1/2-approximation to  $f_{\max}$  that will serve as a first step for a greedy DGFT construction algorithm in Section V.

#### A. Motivation for Spread Frequencies

*Directed graphs:* The directed variation measure

$$DV'(\mathbf{x}) := \sum_{i,j=1}^N A_{ij} [x_i - x_j]_+ \quad (9)$$

was introduced in [31] as the convex Lovász extension of the graph cut size, whose minimization can facilitate identifying graph clusters [17, Ch. 4]. Different from (2), the directed variation measure (9) is not smooth and for undirected graphs it boils down to a so-termed graph absolute variation  $TV_1(\mathbf{x}) := \sum_{i,j=1}^N A_{ij} |x_i - x_j|$  [cf. (1)]; see also [3], [30]. To obtain a GFT for digraphs, the approach in [31] is to solve the orthogonality-constrained problem

$$\min_{\mathbf{U}} \sum_{i=1}^N DV'(\mathbf{u}_i), \quad \text{subject to } \mathbf{U}^T \mathbf{U} = \mathbf{I}_N. \quad (10)$$

An attractive feature of this construction is that it can offer parsimonious representations of graph signals exhibiting smooth structure within clusters (i.e., densely connected subgraphs) of the underlying graph  $\mathcal{G}$ .

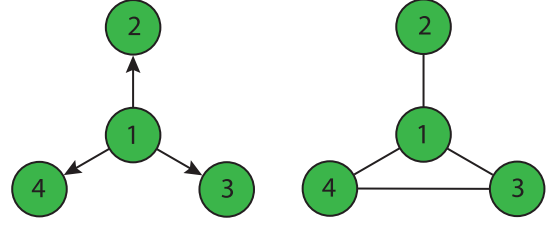


Fig. 1. Toy directed (left) and undirected (right) graphs used to motivate the advocated DGFT design based on spectral dispersion minimization. For the shown digraph, the GFT construction of [31] fails to yield signal expansions with respect to different modes of variation. The Laplacian eigenbasis of the shown undirected graph does not minimize the dispersion criterion in (8).

Consider the digraph with  $N = 4$  nodes shown in Fig. 1 (left). An optimal GFT basis  $\mathbf{U}$  solving (10) takes the form

$$\mathbf{U} = \begin{bmatrix} 0.5 & c & c & c \\ 0.5 & a & 0 & b \\ 0.5 & b & a & 0 \\ 0.5 & 0 & b & a \end{bmatrix},$$

where  $a = (1 + \sqrt{5})/4 \approx 0.8090$ ,  $b = (1 - \sqrt{5})/4 \approx -0.3090$ , and  $c = -0.5$ . These values satisfy  $a + b + c = 0$ ,  $a^2 + b^2 + c^2 = 1$ , and  $c^2 + ab = 0$ , which implies the orthonormality of  $\mathbf{U}$ . As a result, for all columns  $\mathbf{u}_k$  of  $\mathbf{U}$  one has  $DV'(\mathbf{u}_k) = 0$ ,  $k = 1, \dots, 4$ , and hence the inverse GFT synthesis formula  $\mathbf{x} = \mathbf{U}\tilde{\mathbf{x}}$  fails to offer an expansion of  $\mathbf{x}$  with respect to *different* modes of variation (e.g., low and high graph frequencies).

*Undirected graphs:* When the edges of the graph are undirected, the workhorse GFT approach is to project the signals onto the eigenvectors of the graph Laplacian  $\mathbf{L}$ ; see e.g., [26], [42], [44]. A pertinent question is whether the Laplacian-based GFT minimizes the spectral dispersion function in (7), where  $DV(\mathbf{x}) \equiv TV(\mathbf{x})$  because  $\mathcal{G}$  is undirected. To provide an answer, we consider the graph shown in Fig. 1 (right) and form its Laplacian matrix  $\mathbf{L}$ . Denote the eigenvectors of  $\mathbf{L}$  as  $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4]$ , with corresponding frequencies  $[\lambda_1, \lambda_2, \lambda_3, \lambda_4] = [0, 1, 3, 4]$ . To determine the DGFT basis  $\mathbf{U}$  satisfying our design criteria in (8), we first set  $\mathbf{u}_1 := \mathbf{v}_1$  and  $\mathbf{u}_4 := \mathbf{v}_4$  which represent the frequency components associated with minimum and maximum frequencies, respectively. Since the Laplacian eigenvalues are distinct and we are searching for an orthonormal basis, then it follows that  $\mathbf{v}_2$  and  $\mathbf{v}_3$  span the hyperplane containing  $\mathbf{u}_2$  and  $\mathbf{u}_3$ . There is a single degree of freedom to specify  $\mathbf{u}_2$  and  $\mathbf{u}_3$  on that plane, namely a simultaneous rotation of  $\mathbf{v}_2$  and  $\mathbf{v}_3$ . Accordingly, all the feasible basis vectors  $\{\mathbf{u}_2, \mathbf{u}_3\}$  will be of the form

$$[\mathbf{u}_2 \quad \mathbf{u}_3] = [\mathbf{v}_2 \quad \mathbf{v}_3] \mathbf{R}_\theta = [\mathbf{v}_2 \quad \mathbf{v}_3] \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix},$$

where  $\mathbf{R}_\theta$  rotates vectors counterclockwise by an angular amount of  $\theta$ . Collecting the sought basis vectors in  $\mathbf{U}_\theta = [\mathbf{v}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{v}_4]$ , in Fig. 2(a) we plot the dispersion function defined in (7) as a function of  $\theta$ . It is apparent from Fig. 2(a) that  $\delta(\mathbf{V})$  obtained from eigenvectors of the Laplacian is not a global minimizer of the spectral dispersion function. As shown in Fig. 2(a), the optimum basis is  $\mathbf{U} := \mathbf{U}_\theta|_{\theta \approx 0.4214}$ .

To further compare the frequency components obtained, the horizontal lines in Fig 2(b) depict both sets of frequencies

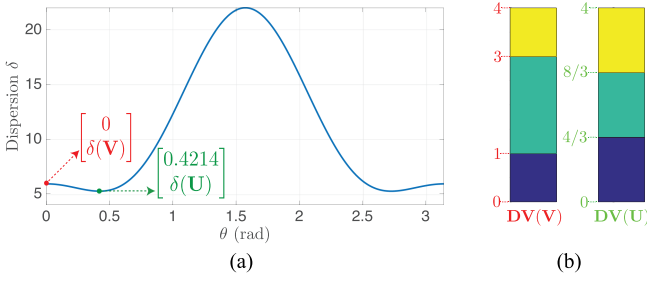


Fig. 2. (a) Spectral dispersion function (7) versus  $\theta$  in radians, for the undirected graph in Fig. 1 (right). Apparently, the solution  $\mathbf{U}$  of (8) offers a more disperse basis than the eigenvectors  $\mathbf{V}$  of the Laplacian matrix. (b) Colored boxes show the consecutive frequency differences obtained by  $\mathbf{V}$  (left) and  $\mathbf{U}$  (right), while the specific directed variation values correspond to the horizontal boundary lines. In this particular case, the optimum basis  $\mathbf{U}$  yields exactly equidistributed graph frequencies in  $[0, 4]$ , as defined in (3).

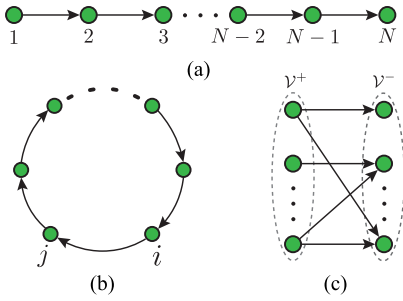


Fig. 3. Some special families of digraphs for which  $f_{\max}$  as defined in (6) can be obtained analytically. (a) Directed path. (b) Directed cycle. (c) Unidirectional bipartite graph, where all the edges go from vertices in  $\mathcal{V}^+$  to those in  $\mathcal{V}^-$ .

$\{\text{DV}(\mathbf{v}_k)\}_{k=1}^4$  and  $\{\text{DV}(\mathbf{u}_k)\}_{k=1}^4$ . As expected, the optimized GFT basis  $\mathbf{U}$  gives rise to frequencies that are more uniformly spread in the graph spectral domain. Moreover, for this particular example the graph frequencies  $\{\text{DV}(\mathbf{u}_k)\}_{k=1}^4$  form an arithmetic sequence in  $[0, 4]$ , which for general graphs may be infeasible as discussed in Section II-B. In Sections IV and V we propose two approaches with complementary strengths to find spread frequencies for arbitrary digraphs.

### B. Maximum Directed Variation

As mentioned in Section II-B, one challenge in finding an approximately equidistributed set of frequencies on a digraph  $\mathcal{G}(\mathcal{V}, \mathbf{A})$  is to calculate the maximum frequency  $f_{\max}$ . The spherically-constrained problem (6) is non-convex and in general challenging to solve [cf. (5) for subsumed undirected graphs, whose solution is the spectral radius of the Laplacian]. The following proposition asserts that for some particular classes of digraphs (depicted in Fig. 3), the value of  $f_{\max}$  can be obtained analytically.

**Proposition 1:** Let  $f_{\max}$  be the maximum directed variation that a unit norm vector can attain as defined in (6).

- 1) Let  $\mathcal{G}$  be the *directed path* (dipath) depicted in Fig. 3(a), i.e., a digraph whose adjacency matrix has nonzero entries  $A_{ij} > 0$  only for  $j = i + 1$ ,  $i < N$ . Then,  $f_{\max} = 2 \max_{i,j} A_{ij}$ .
- 2) Let  $\mathcal{G}$  be the *directed cycle* depicted in Fig. 3(b), i.e., a digraph whose adjacency matrix has nonzero entries

$A_{ij} > 0$  only for  $j = \text{mod}_N(i) + 1$ , where  $\text{mod}_N(x)$  denotes the modulus (remainder) obtained after dividing  $x$  by  $N$ . Then,  $f_{\max} = 2 \max_{i,j} A_{ij}$ .

- 3) Let  $\mathcal{G}$  be a *unidirectional bipartite graph* as depicted in Fig. 3(c), i.e., a bipartite digraph where  $\mathcal{V} = \mathcal{V}^+ \cup \mathcal{V}^-$ ,  $\mathcal{V}^+ \cap \mathcal{V}^- = \emptyset$ , and whose adjacency matrix may only have nonzero entries  $A_{ij} > 0$  for  $i \in \mathcal{V}^+$  and  $j \in \mathcal{V}^-$ . Let  $\mathbf{L}$  be the Laplacian matrix of the underlying undirected graph  $\mathcal{G}^u$  [recall the discussion following (6)], with spectral radius  $\lambda_{\max}$ . Then,  $f_{\max} = \lambda_{\max}$ .

*Proof:* See the Appendices B, C, and D.  $\blacksquare$

Beyond the special digraph families in Proposition 1, it is not clear how to solve (6). Next we show that  $f_{\max}$  is upper-bounded by  $\lambda_{\max}$ , the spectral radius of the Laplacian of the underlying undirected graph  $\mathcal{G}^u$ ; see Section II-B.

**Proposition 2:** For a digraph  $\mathcal{G}$ , recall its underlying undirected graph  $\mathcal{G}^u$  and the spectral radius  $\lambda_{\max}$  of its Laplacian  $\mathbf{L}$ . Then, the maximum directed variation  $f_{\max}$  defined in (6) is upper-bounded by  $f_{\max}^u = \lambda_{\max}$ .

*Proof:* From the definition of directed variation (6), we have

$$\begin{aligned} f_{\max} &= \sum_{i,j=1}^N A_{ij} [u_{\max,i} - u_{\max,j}]_+^2 \\ &= \sum_{\substack{i,j=1 \\ i < j}}^N A_{ij} [u_{\max,i} - u_{\max,j}]_+^2 + A_{ji} [u_{\max,j} - u_{\max,i}]_+^2. \end{aligned}$$

Since  $A_{ij}, A_{ji} \leq \max\{A_{ij}, A_{ji}\} = A_{ij}^u$ , then it follows that

$$\begin{aligned} f_{\max} &\leq \sum_{\substack{i,j=1 \\ i < j}}^N A_{ij}^u \left( [u_{\max,i} - u_{\max,j}]_+^2 + [u_{\max,j} - u_{\max,i}]_+^2 \right) \\ &= \sum_{i,j=1, i < j}^N A_{ij}^u (u_{\max,i} - u_{\max,j})^2 = \text{TV}(\mathbf{u}_{\max}). \end{aligned}$$

Finally, because  $\text{TV}(\mathbf{u}_{\max})$  is upper-bounded by  $\lambda_{\max}$  in (5), we conclude that  $f_{\max} \leq \lambda_{\max}$ .  $\blacksquare$

For general digraphs, Proposition 3 specifies how to find a basis vector  $\tilde{\mathbf{u}}$  with an approximate  $\tilde{f}_{\max} := \text{DV}(\tilde{\mathbf{u}})$  which is at least half of  $f_{\max}$ . Once more, the underlying undirected graph  $\mathcal{G}^u$  and the leading eigenvector of its Laplacian matrix will prove instrumental to obtain the desired approximation.

**Proposition 3:** For a digraph  $\mathcal{G}$ , recall its underlying undirected graph  $\mathcal{G}^u$  and the spectral radius  $\lambda_{\max}$  of its Laplacian  $\mathbf{L}$ . Let  $\mathbf{u}$  be the dominant eigenvector of  $\mathbf{L}$ , i.e., the unit-norm vector  $\mathbf{u}$  such that  $\mathbf{L}\mathbf{u} = \lambda_{\max}\mathbf{u}$ . Then, a worst-case 1/2-approximation to  $f_{\max}$  is given by

$$\tilde{f}_{\max} := \max \{ \text{DV}(\mathbf{u}), \text{DV}(-\mathbf{u}) \} \geq \frac{f_{\max}}{2}. \quad (11)$$

*Proof:* First recall that

$$\lambda_{\max} = \mathbf{u}^T \mathbf{L} \mathbf{u} = \sum_{\substack{i,j=1 \\ i < j}}^N A_{ij}^u (u_i - u_j)^2 = \frac{1}{2} \sum_{i,j=1}^N A_{ij}^u (u_i - u_j)^2.$$

Since  $A_{ij}^u \leq A_{ij} + A_{ji}$  and  $(u_i - u_j)^2 = [u_i - u_j]_+^2 + [u_j - u_i]_+^2$ ,

$$\lambda_{\max} \leq \frac{1}{2} \sum_{i,j=1}^N (A_{ij} + A_{ji}) \left( [u_i - u_j]_+^2 + [u_j - u_i]_+^2 \right) \\ = \text{DV}(\mathbf{u}) + \text{DV}(-\mathbf{u}).$$

In conclusion, at least one of  $\text{DV}(\mathbf{u})$  or  $\text{DV}(-\mathbf{u})$  is larger than  $\lambda_{\max}/2$ , and this completes the proof since  $\lambda_{\max} \geq f_{\max}$ . ■

In practice, we can compute  $\max \{\text{DV}(\mathbf{u}), \text{DV}(-\mathbf{u})\}$  for *any* eigenvector  $\mathbf{u}$  of the Laplacian matrix. This will possibly give a higher frequency in  $\mathcal{G}$ , while preserving the 1/2-approximation.

In Section V we will revisit the result in Proposition 3, to motivate a greedy heuristic to construct a disperse DGFT basis from Laplacian eigenvectors of  $\mathcal{G}^u$ . But before that, in the next section we develop a DGFT algorithm that in practice returns near-optimal solutions to the spectral dispersion minimization problem (7). While computationally more demanding than the recipe in Proposition 3, we show in Section IV-A that the adopted framework for orthogonality-constrained optimization can be as well used to accurately approximate  $f_{\max}$ .

#### IV. 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 [43]. Specifically, following the specification in Section II-C we take a two step approach whereby: i) we find  $f_{\max}$  and its corresponding basis vector  $\mathbf{u}_{\max}$  by solving (6); and ii) we solve (8) to find well-spread frequency components  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_N]$  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 [1].

The general iterative method of [43] deals with an orthogonality constrained problem of the form

$$\min_{\mathbf{U} \in \mathbb{R}^{n \times p}} \phi(\mathbf{U}), \quad \text{subject to} \quad \mathbf{U}^T \mathbf{U} = \mathbf{I}_p, \quad (12)$$

where  $\phi(\mathbf{U}) : \mathbb{R}^{n \times p} \rightarrow \mathbb{R}$  is assumed to be differentiable, just like  $\delta(\mathbf{U})$  in (7). Given a feasible point  $\mathbf{U}_k$  at iteration  $k = 0, 1, 2, \dots$  and the gradient  $\mathbf{G}_k = \nabla \phi(\mathbf{U}_k)$ , one follows the update rule

$$\mathbf{U}_{k+1}(\tau) = \left( \mathbf{I}_n + \frac{\tau}{2} \mathbf{B}_k \right)^{-1} \left( \mathbf{I}_n - \frac{\tau}{2} \mathbf{B}_k \right) \mathbf{U}_k, \quad (13)$$

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 (13) is known as the Cayley transform which preserves orthogonality (i.e.,  $\mathbf{U}_{k+1}^T \mathbf{U}_{k+1} = \mathbf{I}_p$ ), since  $(\mathbf{I}_n + \frac{\tau}{2} \mathbf{B}_k)^{-1}$  and  $\mathbf{I}_n - \frac{\tau}{2} \mathbf{B}_k$  commute. Other noteworthy properties of the update are: i)  $\mathbf{U}_{k+1}(0) = \mathbf{U}_k$ ; ii)  $\mathbf{U}_{k+1}(\tau)$  in (13) is smooth as a function of the step size  $\tau$ ; and iii)  $\frac{d}{d\tau} \mathbf{U}_{k+1}(0)$  is the projection of  $-\mathbf{G}_k$  into the tangent space of the Stiefel manifold at  $\mathbf{U}_k$ .

Most importantly, iii) ensures that the update (13) is a descent path for a proper step size  $\tau$ , which can be obtained through a

---

#### Algorithm 1: Directed Variation Maximization.

---

- 1: **Input:** Adjacency matrix  $\mathbf{A}$  and parameter  $\epsilon > 0$ .
  - 2: **Initialize**  $k = 0$  and unit-norm  $\mathbf{u}_0 \in \mathbb{R}^N$  at random.
  - 3: **repeat**
  - 4:   Evaluate objective  $\phi(\mathbf{u}_k) := -\text{DV}(\mathbf{u}_k)$  in (2).
  - 5:   Compute gradient  $\bar{\mathbf{g}}_k \in \mathbb{R}^N$  via (15).
  - 6:   Form  $\mathbf{B}_k = \bar{\mathbf{g}}_k \mathbf{u}_k^T - \mathbf{u}_k \bar{\mathbf{g}}_k^T$ .
  - 7:   Select  $\tau_k$  satisfying conditions (14a) and (14b).
  - 8:   Update  $\mathbf{u}_{k+1}(\tau_k) = (\mathbf{I}_N + \frac{\tau_k}{2} \mathbf{B}_k)^{-1} (\mathbf{I}_N - \frac{\tau_k}{2} \mathbf{B}_k) \mathbf{u}_k$ .
  - 9:    $k \leftarrow k + 1$ .
  - 10: **until**  $\|\mathbf{u}_k - \mathbf{u}_{k-1}\| \leq \epsilon$
  - 11: **Return**  $\mathbf{u}_{\max} := \mathbf{u}_k$  and  $f_{\max} := \text{DV}(\mathbf{u}_{\max})$ .
- 

curvilinear search satisfying the Armijo-Wolfe conditions

$$\phi(\mathbf{U}_{k+1}(\tau_k)) \leq \phi(\mathbf{U}_{k+1}(0)) + \rho_1 \tau_k \phi'_\tau(\mathbf{U}_{k+1}(0)) \quad (14a)$$

$$\phi'_\tau(\mathbf{U}_{k+1}(\tau_k)) \geq \rho_2 \phi'_\tau(\mathbf{U}_{k+1}(0)), \quad (14b)$$

where  $0 < \rho_1 < \rho_2 < 1$  are two parameters [25]. One can show that if  $\phi(\mathbf{U}_{k+1}(\tau))$  is continuously differentiable and bounded below as is the case for problems (6) and (8), then there exists a  $\tau_k$  satisfying (14a) and (14b). Moreover, the derivative of  $\phi(\mathbf{U}_{k+1}(\tau))$  at  $\tau = 0$  is given by  $\phi'_\tau(\mathbf{U}_{k+1}(0)) = -1/2 \|\mathbf{B}_k\|_F$ ; see [43] for additional details. All in all, the iterations (13) are well defined and by implementing the aforementioned curvilinear search, [43, Theorem 2] asserts that the overall procedure converges to a stationary point of  $\phi(\mathbf{U})$ , while generating feasible points in the Stiefel manifold at every iteration.

##### A. Directed Variation Maximization

As the first step to find the DGFT basis, we obtain  $f_{\max}$  by using the feasible approach to minimize  $-\text{DV}(\mathbf{u})$  over the sphere  $\{\mathbf{u} \in \mathbb{R}^N \mid \mathbf{u}^T \mathbf{u} = 1\}$  [cf. (6)]. The gradient  $\bar{\mathbf{g}} := \nabla \text{DV}(\mathbf{u}) \in \mathbb{R}^N$  has entries  $\bar{g}_i$ ,  $1 \leq i \leq N$ , given by

$$\bar{g}_i = 2 \left( \mathbf{A}_{\cdot i}^T [\mathbf{u} - u_i \mathbf{1}_N]_+ - \mathbf{A}_{\cdot i} [u_i \mathbf{1}_N - \mathbf{u}]_+ \right), \quad (15)$$

where  $\mathbf{A}_{\cdot i}$  denotes the  $i$ th column of the adjacency matrix  $\mathbf{A}$ , and  $\mathbf{A}_{\cdot i}$  the  $i$ th row.

The algorithm starts from a random unit-norm vector and then via (13) it takes a descent path towards a stationary point. The overall procedure is tabulated under Algorithm 1. It is often prudent to run the iterations multiple times using random initializations, and retain the solution that yields the least cost. Although Algorithm 1 only guarantees convergence to a stationary point of the directed variation cost, in practice we have observed that it tends to find  $f_{\max} = \text{DV}(\mathbf{u}_{\max})$  exactly if the number of initializations is chosen large enough; see Section VI. While finding  $f_{\max}$  is of interest in its own right, our focus next is on using the obtained  $\mathbf{u}_{\max}$  to formulate and solve the spectral dispersion minimization problem (8).

##### B. Spectral Dispersion Minimization

As the second and final step, here we develop an algorithm to find the orthonormal basis  $\mathbf{U}$  that minimizes the spectral dispersion (7). To cast the optimization problem (8) in the form of (12) and apply the previously outlined feasible method, we penalize the objective  $\delta(\mathbf{U})$  with a measure of the constraint

**Algorithm 2:** Spectral Dispersion Minimization.

- 1: **Input:** Adjacency matrix  $\mathbf{A}$ , parameters  $\lambda > 0$  and  $\epsilon > 0$ .
- 2: Find  $\mathbf{u}_{\max}$  using Algorithm 1 and set  $\mathbf{u}_{\min} = \frac{1}{\sqrt{N}}\mathbf{1}_N$ .
- 3: **Initialize**  $k = 0$  and orthonormal  $\mathbf{U}_0 \in \mathbb{R}^{N \times N}$  at random.
- 4: **repeat**
- 5:   Evaluate objective  $\phi(\mathbf{U}_k)$  in (16).
- 6:   Compute gradient  $\mathbf{G}_k \in \mathbb{R}^{N \times N}$  via (17).
- 7:   Form  $\mathbf{B}_k = \mathbf{G}_k \mathbf{U}_k^T - \mathbf{U}_k \mathbf{G}_k^T$ .
- 8:   Select  $\tau_k$  satisfying conditions (14a) and (14b).
- 9:   Update  $\mathbf{U}_{k+1}(\tau_k) = (\mathbf{I}_N + \frac{\tau_k}{2}\mathbf{B}_k)^{-1}(\mathbf{I}_N - \frac{\tau_k}{2}\mathbf{B}_k)\mathbf{U}_k$ .
- 10:    $k \leftarrow k + 1$ .
- 11: **until**  $\|\mathbf{U}_k - \mathbf{U}_{k-1}\|_F \leq \epsilon$
- 12: **Return**  $\hat{\mathbf{U}} = \mathbf{U}_k$ .

violations to obtain

$$\begin{aligned} \min_{\mathbf{U}} \quad & \phi(\mathbf{U}) := \delta(\mathbf{U}) + \frac{\lambda}{2} \left( \|\mathbf{u}_1 - \mathbf{u}_{\min}\|^2 \right. \\ & \left. + \|\mathbf{u}_N - \mathbf{u}_{\max}\|^2 \right) \\ \text{subject to} \quad & \mathbf{U}^T \mathbf{U} = \mathbf{I}_N, \end{aligned} \quad (16)$$

where  $\lambda > 0$  is chosen large enough to ensure  $\mathbf{u}_1 = \mathbf{u}_{\min}$  and  $\mathbf{u}_N = \mathbf{u}_{\max}$ . Since we can explicitly monitor whether the constraints are enforced, the choice of  $\lambda$  is often made via trial and error; see also Section VI for an empirical demonstration. The resulting iterations are tabulated under Algorithm 2, where the gradient matrix  $\mathbf{G} := \nabla \phi(\mathbf{U}) \in \mathbb{R}^{N \times N}$  has columns given by

$$\begin{aligned} \mathbf{g}_1 &= [\mathbf{DV}(\mathbf{u}_1) - \mathbf{DV}(\mathbf{u}_2)] \bar{\mathbf{g}}(\mathbf{u}_1) + \lambda(\mathbf{u}_1 - \mathbf{u}_{\min}) \\ \mathbf{g}_i &= [2\mathbf{DV}(\mathbf{u}_i) - \mathbf{DV}(\mathbf{u}_{i+1}) - \mathbf{DV}(\mathbf{u}_{i-1})] \bar{\mathbf{g}}(\mathbf{u}_i), \quad 1 < i < N \\ \mathbf{g}_N &= [\mathbf{DV}(\mathbf{u}_{N-1}) - \mathbf{DV}(\mathbf{u}_N)] \bar{\mathbf{g}}(\mathbf{u}_N) + \lambda(\mathbf{u}_N - \mathbf{u}_{\max}), \end{aligned} \quad (17)$$

where the entries of  $\bar{\mathbf{g}}$  are specified in (15). Once more, it is convenient to run the algorithm multiple times and retain the least disperse DGFT basis  $\hat{\mathbf{U}}$ .

While provably convergent to a stationary point of (16), Algorithm 2 does not offer guarantees on the global optimality of the solution  $\hat{\mathbf{U}}$ . Still, numerical tests in Section VI corroborate the effectiveness of the proposed optimization strategy and its robustness with respect to the initialization. The computational complexity of Algorithm 2 is  $\mathcal{O}(N^3)$  per iteration due to the matrix inversion involved in the calculation of the Cayley transform. In the next section we propose a lightweight heuristic to construct spread DGFT basis using the eigenvectors of  $\mathcal{G}^u$ 's Laplacian matrix. But before moving on, a remark on the relationship between the DGFT and the DFT of discrete-time signals is in order.

*Remark 1 (Relationship with the DFT):* The scope of the proposed DGFT framework [in particular the notion of directed variation in (2)] is limited to real-valued signals and basis vectors. Accordingly, for the directed cycle graph in Fig. 3(b) which represents the support of periodic discrete-time signals, one would fail to obtain the (complex-valued) DFT as the solution of (8). Still, one can modify the definition of directed variation

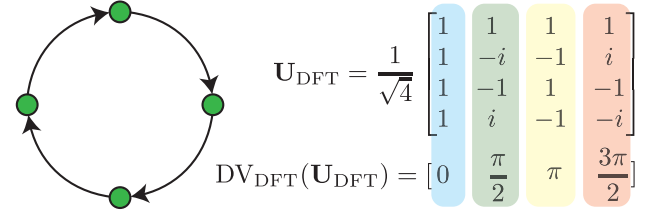


Fig. 4. (left) A directed cycle graph with  $N = 4$  nodes. (right) The 4-point DFT basis is recovered as the optimal solution of (19).

in (2) to recover the classical DFT when  $\mathcal{G}$  is a directed cycle. To this end, for a unit-norm complex-valued  $\mathbf{x}$  consider

$$\mathbf{DV}_{\text{DFT}}(\mathbf{x}) := \frac{1}{\sum_{j,k=1}^N A_{jk}} \sum_{j,k=1}^N A_{jk} [\text{mod}_{2\pi}(\omega_j - \omega_k)], \quad (18)$$

where  $i = \sqrt{-1}$  is the imaginary number and  $\omega_j \in [0, 2\pi)$  denotes the phase angle of  $x_j$ , the signal value at node  $j$ . Using this definition, one can formulate the following complex-valued spectral dispersion minimization problem [cf. (8)]

$$\begin{aligned} \min_{\mathbf{U}} \quad & \sum_{r=1}^N [\mathbf{DV}_{\text{DFT}}(\mathbf{u}_{r+1}) - \mathbf{DV}_{\text{DFT}}(\mathbf{u}_r)]^2 \\ \text{subject to} \quad & \mathbf{U}^H \mathbf{U} = \mathbf{I}_N, \\ & \mathbf{u}_1 = \frac{1}{\sqrt{N}}\mathbf{1}_N, \\ & \mathbf{DV}_{\text{DFT}}(\mathbf{u}_{N+1}) = 2\pi, \end{aligned} \quad (19)$$

where  $^H$  stands for conjugate transposition. Notice that  $\mathbf{u}_{N+1}$  is not an optimization variable, and the third constraint fixes  $[2\pi - \mathbf{DV}_{\text{DFT}}(\mathbf{u}_N)]^2$  as the last (i.e.,  $r = N$ ) summand of the objective function.

When  $\mathcal{G}$  is the unweighted, directed cycle with  $N$  nodes, the global optimum  $\mathbf{U}_{\text{DFT}}^*$  of (19) is the  $N$ -point DFT matrix, with entries  $U_{\text{DFT},jk}^* = \frac{1}{\sqrt{N}} e^{-\frac{2\pi i j k}{N}}$ . The frequencies form an arithmetic sequence, while covering the whole frequency range  $[0, 2\pi)$ ; see Fig. 4. It is fair to say that the objective function of (19) does not subsume its counterpart (7) as a special case for real signals. While an iterative solver is not needed in this particular case, the feasible method in [43] can accommodate complex-valued Stiefel manifold constraints.

## V. A DIGRAPH FOURIER TRANSFORM HEURISTIC

As an alternative to the feasible method discussed in the previous section, here we consider the underlying undirected graph  $\mathcal{G}^u$  and use the eigenvectors of its Laplacian matrix to construct a disperse set of frequencies. The reason for using the eigenvectors of  $\mathbf{L}$  is that i) they are widely used for undirected graphs having good localization properties in the vertex domain; ii) we can modify the frequencies (in the digraph  $\mathcal{G}$ ) by flipping the sign of each Laplacian eigenvector; and iii) they can be used to approximate  $f_{\max}$  within a factor of 1/2 as asserted in Proposition 3. Unlike the undirected case, the directed variation of any eigenvector  $\mathbf{u}$  will in general be different from the variation of  $-\mathbf{u}$ ; so we can pick the one that we desire without compromising the orthogonality constraint.



Fixing  $f_1 = 0$  and  $f_N = \tilde{f}_{\max}$  from (11), in lieu of (3) we will henceforth construct a disperse set of frequencies by using the eigenvectors of  $\mathbf{L}$ . Let  $\mathbf{f}_i := \text{DV}(\mathbf{u}_i)$  and  $\bar{\mathbf{f}}_i := \text{DV}(-\mathbf{u}_i)$ , where  $\mathbf{u}_i$  is the  $i$ th eigenvector of  $\mathbf{L}$ . Define the set of all candidate frequencies as  $\mathcal{F} := \{f_i, \bar{f}_i : 1 < i < N\}$ . The goal is to select  $N - 2$  frequencies from  $\mathcal{F}$  such that together with  $\{0, \tilde{f}_{\max}\}$  they form our well-spread Fourier frequencies. To preserve orthonormality, we would select exactly one from each pair  $\{f_i, \bar{f}_i\}$ . We will argue later that this induces a matroid basis constraint on a supermodular (frequency) set minimization problem described next.

#### A. Frequency Selection via Supermodular Minimization

To find the DGFT basis in accordance to the design criterion in Section II-C, we define a spectral dispersion (set) function that measures how well spread the corresponding frequencies are over  $[0, \tilde{f}_{\max}]$ . For a candidate frequency set  $\mathcal{S} \subseteq \mathcal{F}$ , let  $s_1 \leq s_2 \leq \dots \leq s_m$  be the elements of  $\mathcal{S}$  in non-decreasing order, where  $m = |\mathcal{S}|$ . Then we define the dispersion of  $\mathcal{S}$  as

$$\delta(\mathcal{S}) = \sum_{i=0}^m (s_{i+1} - s_i)^2, \quad (20)$$

where  $s_0 = 0$  and  $s_{m+1} = \tilde{f}_{\max}$  [cf. (7)]. It can be verified that  $\delta(\mathcal{S})$  is a monotone non-increasing function, which means that for any sets  $\mathcal{S}_1 \subseteq \mathcal{S}_2$ , we have  $\delta(\mathcal{S}_1) \geq \delta(\mathcal{S}_2)$ . For a fixed value of  $m$ , one can show that  $\delta(\mathcal{S})$  is minimized when the  $s_i$ 's form an arithmetic sequence, consistent with our design goal in (3). Hence, we seek to minimize  $\delta(\mathcal{S})$  through a set function optimization procedure. In Lemma 1 we show that the dispersion function (20) has the supermodular property. First, for completeness we define submodularity/supermodularity.

**Definition 1 (Submodularity):** Let  $\mathcal{S}$  be a finite ground set. A set function  $f : 2^{\mathcal{S}} \mapsto \mathbb{R}$  is submodular if:

$$f(\mathcal{T}_1 \cup \{e\}) - f(\mathcal{T}_1) \geq f(\mathcal{T}_2 \cup \{e\}) - f(\mathcal{T}_2), \quad (21)$$

for all subsets  $\mathcal{T}_1 \subseteq \mathcal{T}_2 \subseteq \mathcal{S}$  and any element  $e \in \mathcal{S} \setminus \mathcal{T}_2$ .

Equation (21) is also known as the *diminishing returns* property. It means that adding a single element  $e$  results in less gain when added to a bigger set  $\mathcal{T}_2$ , compared to adding the same element to a subset of  $\mathcal{T}_2$  like  $\mathcal{T}_1$ . The diminishing returns property arises in many science and engineering applications including facility location, sensor placement, and feature selection [24], where adding a new sensor/feature or opening a new location becomes increasingly less beneficial as one has more and more of them already available. A set function  $f$  is said to be supermodular if  $-f$  is submodular, i.e., (21) holds in the other direction. Roughly speaking, for supermodular functions, items have more value when bundled together.

**Lemma 1:** The spectral dispersion function  $\delta : 2^{\mathcal{F}} \mapsto \mathbb{R}$  defined in (20) is a supermodular function.

**Proof:** Consider two subsets  $\mathcal{S}_1, \mathcal{S}_2$  such that  $\mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq \mathcal{F}$ , and a single element  $e \in \mathcal{F} \setminus \mathcal{S}_2$ . Let  $s_1^L$  and  $s_1^R$  be the largest value smaller than  $e$  and the smallest value greater than  $e$  in  $\mathcal{S}_1 \cup \{0, \tilde{f}_{\max}\}$ , respectively (i.e.,  $e$  breaks the gap between  $s_1^L$  and  $s_1^R$ ). Similarly, let  $s_2^L$  and  $s_2^R$  be defined for  $\mathcal{S}_2$ . Since  $\mathcal{S}_1 \subseteq \mathcal{S}_2$ , then  $s_1^L \leq s_2^L \leq e \leq s_2^R \leq s_1^R$ . The result follows by

comparing the marginal values

$$\begin{aligned} \delta(\mathcal{S}_1 \cup \{e\}) - \delta(\mathcal{S}_1) &= (s_1^R - e)^2 + (e - s_1^L)^2 - (s_1^R - s_1^L)^2 \\ &= -2(s_1^R - e)(e - s_1^L) \\ &\leq -2(s_2^R - e)(e - s_2^L) \\ &= \delta(\mathcal{S}_2 \cup \{e\}) - \delta(\mathcal{S}_2). \end{aligned} \quad \blacksquare$$

Recalling the orthonormality constraint, we define  $\mathcal{B}$  to be the set of all subsets  $\mathcal{S} \subseteq \mathcal{F}$  that satisfy  $|\mathcal{S} \cap \{f_i, \bar{f}_i\}| = 1$ ,  $i = 2, \dots, N - 1$ . Then, frequency selection from  $\mathcal{F}$  boils down to solving

$$\min_{\mathcal{S}} \delta(\mathcal{S}), \quad \text{subject to } \mathcal{S} \in \mathcal{B}. \quad (22)$$

Next, in Lemma 2 we show that the constraint in (22) is a matroid basis constraint. To state that result, we first define the notions of matroid and partition matroid.

**Definition 2 (Matroid):** Let  $\mathcal{S}$  be a finite ground set and let  $\mathcal{I}$  be a collection of subsets of  $\mathcal{S}$ . The pair  $\mathcal{M} = (\mathcal{S}, \mathcal{I})$  is a matroid if the following properties hold:

- Hereditary Property: If  $\mathcal{T} \in \mathcal{I}$ , then  $\mathcal{T}' \in \mathcal{I}$  for all  $\mathcal{T}' \subseteq \mathcal{T}$ .
- Augmentation Property: If  $\mathcal{T}_1, \mathcal{T}_2 \in \mathcal{I}$  and  $|\mathcal{T}_1| < |\mathcal{T}_2|$ , then there exists  $e \in \mathcal{T}_2 \setminus \mathcal{T}_1$  such that  $\mathcal{T}_1 \cup \{e\} \in \mathcal{I}$ .

The collection  $\mathcal{I}$  is called the set of independent sets of the matroid  $\mathcal{M}$ . A maximal independent set is a basis. One can show that all the basis vectors of a matroid have the same cardinality.

A matroid is a powerful structure used in combinatorial optimization, generalizing the notion of linear independence in vector spaces. Indeed, it is not hard to observe that if  $\mathcal{S}$  is a set of (not necessarily independent) vectors, then the linearly independent subsets of  $\mathcal{S}$  form a valid independent family that satisfies the above two properties. The uniform matroid, graphic matroid, and partition matroid are other examples of matroids. The latter one will be useful in the sequel.

**Definition 3 (Partition matroid [32]):** Let  $\mathcal{S}$  denote a finite set and let  $\mathcal{S}_1, \dots, \mathcal{S}_m$  be a partition of  $\mathcal{S}$ , i.e., a collection of disjoint sets such that  $\mathcal{S}_1 \cup \dots \cup \mathcal{S}_m = \mathcal{S}$ . Let  $d_1, \dots, d_m$  be a collection of non-negative integers. Define a set  $\mathcal{I}$  by  $\mathcal{A} \in \mathcal{I}$  iff  $|\mathcal{A} \cap \mathcal{S}_i| \leq d_i$  for all  $i = 1, \dots, m$ . Then,  $\mathcal{M} = (\mathcal{S}, \mathcal{I})$  is called the partition matroid.

All elements are now in place to establish that the orthonormality constraint in (22) is a partition matroid basis constraint.

**Lemma 2:** There exists a (partition) matroid  $\mathcal{M}$  such that the set  $\mathcal{B}$  in (22) is the set of all basis vectors of  $\mathcal{M}$ .

**Proof:** Recall Definition 3 and set  $\mathcal{S} := \mathcal{F}$ ,  $\mathcal{S}_i := \{f_i, \bar{f}_i\}$  and  $d_i := 1$  for all  $i = 2, \dots, N - 1$ , to get a partition matroid  $\mathcal{M} = (\mathcal{F}, \mathcal{I})$ . The basis vectors of  $\mathcal{M}$ , which are defined as the maximal elements of  $\mathcal{I}$ , are those subsets  $\mathcal{A} \subseteq \mathcal{F}$  that satisfy  $|\mathcal{A} \cap \{f_i, \bar{f}_i\}| = 1$  for all  $i = 2, \dots, N - 1$ , which are the elements of  $\mathcal{B}$ .  $\blacksquare$

#### B. Greedy Algorithm for DGFT Basis Selection

Lemmas 1 and 2 assert that (22) is a supermodular minimization problem subject to a matroid basis constraint. Since supermodular minimization is NP-hard and hard to approximate to any factor [16], [23], we create a submodular function  $\tilde{\delta}(\mathcal{S})$  and use the algorithms for submodular maximization to find a set of disperse basis  $\mathbf{U}$ . In particular, we define



**Algorithm 3:** Greedy Spectral Dispersion Minimization.

---

```

1: Input: Set of possible frequencies  $\mathcal{F}$ .
2: Initialize  $\mathcal{S} = \emptyset$ .
3: repeat
4:    $e \leftarrow \operatorname{argmax}_{f \in \mathcal{F}} \{ \tilde{\delta}(\mathcal{S} \cup \{f\}) - \tilde{\delta}(\mathcal{S}) \}$ .
5:    $\mathcal{S} \leftarrow \mathcal{S} \cup \{e\}$ .
6:   Delete from  $\mathcal{F}$  the pair  $\{f_i, \bar{f}_i\}$  that  $e$  belongs to.
7: until  $F = \emptyset$ 

```

---

$$\tilde{\delta}(\mathcal{S}) := \tilde{f}_{\max}^2 - \delta(\mathcal{S}), \quad (23)$$

which is a non-negative (increasing) submodular function, because  $\delta(\emptyset) = \tilde{f}_{\max}^2$  is an upper bound for  $\delta(\mathcal{S})$ . There are several results for maximizing submodular functions under matroid constraints for both the non-monotone [19] and monotone cases [2], [8]. We adopt the greedy algorithm of [8] due to its simplicity (tabulated under Algorithm 3), which provides a  $1/2$ -approximation guarantee (Theorem 1).

The algorithm starts with an empty set  $\mathcal{S}$ . In each iteration, it finds the element  $e$  that produces the biggest gain in terms of increasing  $\tilde{\delta}(\mathcal{S})$ . Then it deletes the pair that  $e$  belongs to, because the other element in that pair cannot be chosen by virtue of the matroid constraint. The running time of the algorithm is  $\mathcal{O}(N^2)$ , in addition to the  $\mathcal{O}(N^3)$  cost of computing the Laplacian eigenvectors.

*Theorem 1 ([8]):* Let  $\mathcal{S}^*$  be the solution of problem (22) and  $\mathcal{S}^g$  be the output of the greedy Algorithm 3. Then,

$$\tilde{\delta}(\mathcal{S}^g) \geq \frac{1}{2} \times \tilde{\delta}(\mathcal{S}^*).$$

Notice that Theorem 1 offers a worst-case guarantee, and Algorithm 3 is usually able to find near-optimal solutions in practice.

In summary, the greedy DGFT basis construction algorithm entails the following steps. First, we form  $\mathcal{G}^u$  and find the eigenvectors of the graph Laplacian  $\mathbf{L}$ . Second, the set  $\mathcal{F}$  is formed by calculating the directed variation for each eigenvector  $\mathbf{u}_i$  and its negative  $-\mathbf{u}_i$ ,  $i = 2, \dots, N-1$ . Finally, the greedy Algorithm 3 is run on the set  $\mathcal{F}$ , and the output determines the set of frequencies as well as the orthonormal set of DGFT basis vectors comprising  $\mathbf{U}$ .

*Remark 2 (Computational complexity):* While the heuristic method is computationally more efficient than the feasible method in Algorithm 2, it still requires a full diagonalization of the graph Laplacian matrix which costs  $\mathcal{O}(N^3)$ . The complexity of  $\mathcal{O}(N^3)$  appears to be a bottleneck for large graphs in all the state-of-the-art existing methods [12], [30], [31], [39]. For particular cases speedups may be obtained by exploiting eigenvector routines for very sparse matrices, by relying on truncated decompositions (which could suffice to approximately decompose smooth signals), or e.g., through greedy approximate diagonalization to compute the Laplacian eigenbasis at lower cost [18]. While certainly a very interesting and fundamental problem, there has been little progress to date when it comes to realizing the vision of a “fast” GFT, even for undirected graphs.

## VI. NUMERICAL RESULTS

Here we carry out computer simulations on three graphs to assess the performance of the algorithms developed to construct

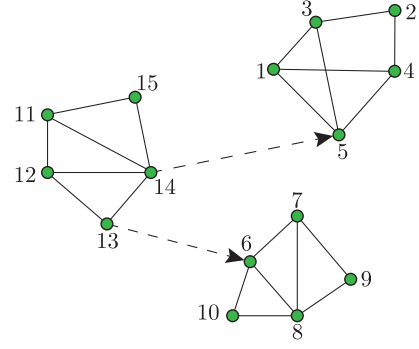


Fig. 5. Synthetic digraph with  $N = 15$  nodes and 2 directed edges depicted as dashed arrows [31].

a DGFT with spread frequency components. We also compare these basis signals with other state-of-the-art GFT methods.

*Synthetic digraph:* Using Algorithms 2 and 3 we construct respective DGFTs for an unweighted digraph  $\mathcal{G}$  with  $N = 15$  nodes shown in Fig. 5, and compare them with the GFT put forth in [31] that relies on an augmented Lagrangian optimization method termed PAMAL, as well as with the eigenbasis of the combinatorial Laplacian for directed graphs in [5]. To define said combinatorial Laplacian for digraphs  $\mathbf{L}_d$ , 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  $\mathbf{\Pi} = \text{diag}(\boldsymbol{\pi})$  be the diagonal matrix with the stationary distribution  $\boldsymbol{\pi}$  of the random walk on the diagonal. Using these definitions, the combinatorial Laplacian for directed graphs in [5] is given by  $\mathbf{L}_d := \mathbf{\Pi} - (\mathbf{\Pi P} + \mathbf{P}^T \mathbf{\Pi})/2$ .

One would expect that the proposed DGFT approaches – which directly optimize the spectral dispersion metric – yield: i) a more spread set of graph frequencies; also ii) spanning a wider range of directed variations. This is indeed apparent from Fig. 6, which depicts the distribution of frequencies (shown as vertical lines) for all GFT methods being compared. In particular, notice how the DGFT basis obtained via direct minimization of the dispersion cost (Algorithm 2) yields an almost equidistributed set of graph frequencies. To further quantify this assertion, we first rescale the directed variation values to the  $[0, 1]$  interval and calculate their dispersion using (7). The results are reported in Table I, which confirms that Algorithms 2 and 3 yield a better frequency spread (i.e., a smaller dispersion). While computationally more demanding, Algorithm 2 yields a more spread set of graph frequencies when compared to the greedy Algorithm 3, since it minimizes dispersion over a larger set [cf. (16) and (22)]. Finally, Fig. 7 shows the frequency components obtained via Algorithm 2. Each subplot depicts one basis vector (column) of the resulting DGFT matrix  $\mathbf{U}$ , along with its corresponding directed variation values defined in (2). It is apparent that the first vectors exhibit less variability than the higher frequency components. Moreover one can see that lower frequency components have the additional desired property of being roughly constant over network clusters; see also the design in [31].

We also use Monte-Carlo simulations to study the effect of  $\lambda$  [cf. (16)] on the convergence properties of our algorithms. After fixing  $\mathbf{u}_{\max}$  and  $\mathbf{u}_{\min}$ , we run Algorithm 2 for different values of  $\lambda$  to find the DGFT basis of the graph in Fig. 5. The spectral dispersion in (7) and the Euclidean distance between  $\mathbf{u}_N$  ( $\mathbf{u}_1$ ) and

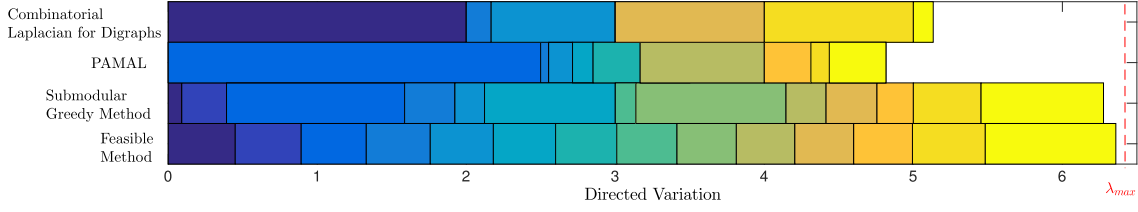


Fig. 6. Comparison of directed variations (i.e., graph frequencies) for the synthetic digraph in Fig. 5 and different GFT methods: eigenvectors of the combinatorial Laplacian matrix  $\mathbf{L}_d$  introduced in [5]; augmented Lagrangian method (PAMAL) in [31]; proposed greedy heuristic (Algorithm 3); and feasible method (Algorithm 2). Colored boxes show the difference between two consecutive frequencies for each method, while the specific directed variation values correspond to the vertical boundary lines. Ideally one would like to have  $N - 1$  equal-sized boxes, but we argued that this is not always achievable. Notice how Algorithm 2 comes remarkably close to such a specification.

TABLE I  
SPECTRAL DISPERSION  $\delta(\mathbf{U})$  OF OBTAINED BASIS  $\mathbf{U}$  USING DIFFERENT ALGORITHMS FOR THE SYNTHETIC DIGRAPH IN FIG. 5

Method	Dispersion
Combinatorial Laplacian for directed graphs [5]	0.256
PAMAL [31]	0.301
Submodular Greedy (Alg. 3)	<b>0.118</b>
Feasible Method (Alg. 2)	<b>0.077</b>

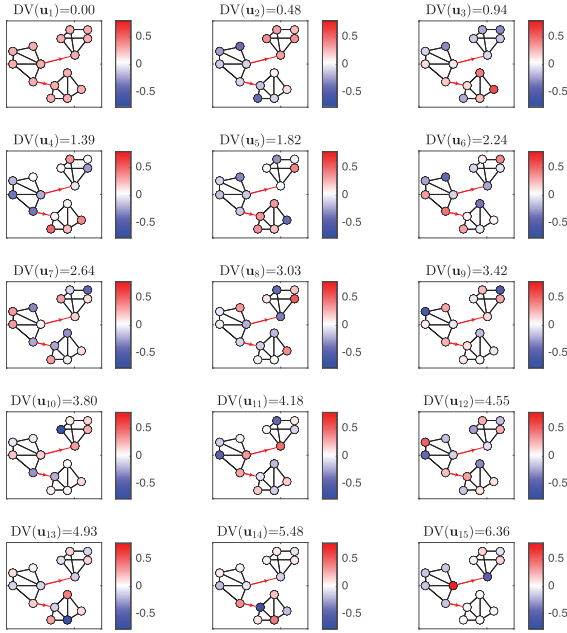


Fig. 7. DGFT basis vectors obtained using Algorithm 2 for the synthetic digraph in Fig. 5, along with their respective directed variation values (frequencies). Notice how the frequency components associated with lower frequencies are roughly constant over node clusters.

$\mathbf{u}_{\max}(\mathbf{u}_{\min})$  averaged over 100 Monte-Carlo simulations are shown in Fig. 8. Apparently, for the extreme value of  $\lambda = 0$  frequencies tend to be as close as possible, resulting in the smallest value of  $\delta(\mathbf{U})$ . However, this solution is not feasible for problem (8). By increasing  $\lambda$  we trade-off dispersion for feasibility, making  $\mathbf{u}_1$  and  $\mathbf{u}_N$  closer to  $\mathbf{u}_{\min}$  and  $\mathbf{u}_{\max}$ , respectively. This way, all constraints are satisfied resulting in a broader spectrum approximately spanning  $[0, f_{\max}]$ . For  $\lambda > 10^2$  in Fig 8,  $\mathbf{u}_1$  and  $\mathbf{u}_N$  become fixed and the remaining basis vectors spread as evenly as possible in the viable spectral band. Further exploring these trade-offs (including the localization properties of the resulting basis vectors) is certainly an interesting direction, which is beyond the scope of this paper and its DGFT design goals.

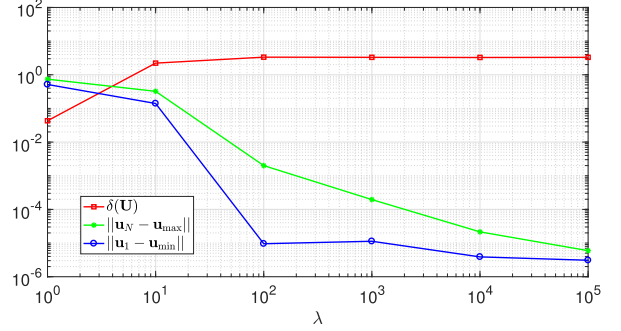


Fig. 8. Effect of  $\lambda$  on the constraint violations  $\|\mathbf{u}_1 - \mathbf{u}_{\min}\|$ ,  $\|\mathbf{u}_N - \mathbf{u}_{\max}\|$  and spectral dispersion  $\delta(\mathbf{U})$  in (16). We run Algorithm 2 for the graph in Fig. 5, and average the results over 100 Monte-Carlo simulations. For larger enough values of  $\lambda$  (here  $10^2$ ), we observe that the dispersion does not change and the last two constraints in (8) are (almost) satisfied.

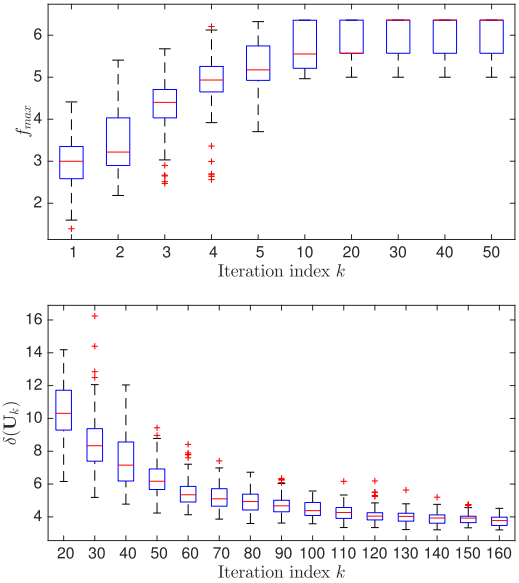


Fig. 9. (top) Convergence behavior of Algorithm 1 for finding the maximum directed variation  $f_{\max}$  of unit-norm signals on the synthetic digraph in Fig. 5. The boxes show the median and the 25th and 75th percentiles of  $f_{\max}$  vs. the number of iterations  $k$ , obtained by running 100 Monte-Carlo simulations based on independent initializations. (bottom) Likewise, but when using Algorithm 2 to minimize the spectral dispersion  $\delta(\mathbf{U})$  in (7).

In Fig. 9 (top) we show the evolution of iterates for the feasible method in [43], when used to find the maximum directed variation (i.e.,  $f_{\max}$ ) for the same 15-node graph in Fig. 5. We do so for 100 different (random) initializations and report the median as well as the first and third quartiles versus the

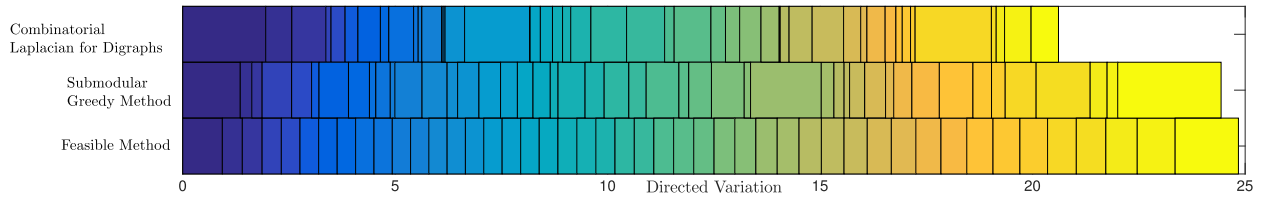


Fig. 10. Comparison of directed variations (i.e., graph frequencies) for the structural brain network and different GFT methods: eigenvectors of the combinatorial Laplacian matrix  $\mathbf{L}_d$  introduced in [5]; proposed greedy heuristic (Algorithm 3); and feasible method (Algorithm 2). Colored boxes show the difference between two consecutive frequencies for each method, while the specific directed variation values correspond to the vertical boundary lines.

number of iterations. We observe that all the realizations converge [43, 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 (8). 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. 9 (bottom) shows the median, first, and third quartiles of the dispersion function iterates  $\delta(\mathbf{U}_k)$ , when minimized using Algorithm 2. Again, 100 different Monte-Carlo simulations are considered and we observe that all of them robustly converge to limiting values with small variability. This suggests that in practice we can run Algorithm 2 with different random initializations and retain the most spread frequency components among the obtained candidate solutions.

*Structural brain graph:* Next we consider a real brain graph representing the anatomical connections of the macaque cortex, which was studied in [13], [28] for example. The network consists of  $N = 47$  nodes and 505 edges (among which 121 of them are directed). The vertices represent different hubs in the brain, and the edges capture directed information flow among them. To corroborate that our resulting basis vectors are well distributed in the graph spectral domain, Fig. 10 depicts the distribution of all the frequencies for the examined algorithms except for the PAMAL algorithm which did not converge within a reasonable time. In Fig. 10, each vertical line indicates the directed variation (frequency) associated with a basis vector. Once more, the proposed algorithms are effective in terms of finding well dispersed and non-repetitive frequencies, which in this context could offer innovative alternatives for filtering of brain signals leading to potentially more interpretable graph frequency analyses [14]. While certainly interesting, such a study is beyond the scope of this paper.

*Contiguous United States:* Finally, we consider a digraph of the  $N = 48$  so-termed contiguous United States (excluding Alaska and Hawaii, which are not connected by land with the other states). 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, i.e., from South to North (S–N). We also consider the average annual temperature of each state as the signal  $\mathbf{x} \in \mathbb{R}^{48}$  shown in Fig. 11.<sup>1</sup> It is apparent from the temperature map that the states closer to the Equator (i.e., with lower latitude) have higher average temperatures. This justifies the adopted latitude-based graph construction scheme, to better capture a notion of flow through the temperature field.

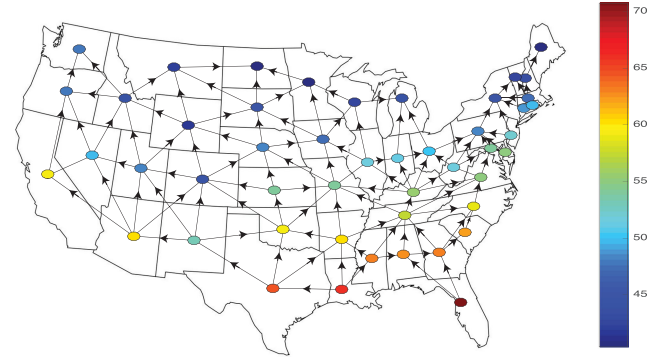


Fig. 11. Graph signal of average annual temperature in Fahrenheit for the contiguous US states. In the depicted digraph, a directed edge joins two states if they share a border, and the direction of the arc is set so that the state whose barycenter has lower latitude points to the one with higher latitude.

We determine a DGFT basis for this digraph via spectral dispersion minimization using Algorithm 2. The resulting first and last 4 frequency modes are depicted in Fig. 13. The first four vectors are smooth as expected. The last basis vectors are smooth as well in the majority of the graph, but there exist a few nodes in them such that a highly connected vertex is significantly warmer than its northern neighbors, or, colder than its southern neighbors. For instance, in  $\mathbf{u}_{48}$  Kentucky has markedly high temperature. This high-frequency modes can indeed help towards filtering out noisy measurements, as these spikes can be due to anomalous events or defective sensors. It is worth mentioning that spikes in the low-frequency modes (i.e., at nodes with no outgoing edges) may not be removed with an ideal low-pass filter, due to the intricacies of the adopted digraph model. For example, Vermont has a large signal value in  $\mathbf{u}_2$  and  $\mathbf{u}_3$  to render  $\mathbf{U}$  orthonormal, but spikes in Vermont (a node with no outgoing edges) have no effect on the directed variation. To alleviate this limitation one could add lightly-weighted outgoing edges from dead ends, an interesting exploration that is well-beyond the scope of this paper.

To corroborate the effectiveness of the obtained basis in a denoising task, we aim to recover the temperature signal from noisy measurements  $\mathbf{y} = \mathbf{x} + \mathbf{n}$ , where the additive noise  $\mathbf{n}$  is a zero-mean, Gaussian random vector with covariance matrix  $10\mathbf{I}_N$ . To that end, we use a low-pass graph filter with frequency response  $\tilde{\mathbf{h}} = [\tilde{h}_1, \dots, \tilde{h}_N]^T$ , where  $\tilde{h}_i = \mathbb{I}\{i \leq w\}$  and  $w$  is the prescribed spectral window size. The filter retains the first  $w$  components of the signal's DGFT, and we approximate the noisy temperature signal by

$$\hat{\mathbf{x}} = \mathbf{U} \text{diag}(\tilde{\mathbf{h}}) \tilde{\mathbf{y}} = \mathbf{U} \text{diag}(\tilde{\mathbf{h}}) \mathbf{U}^T \mathbf{y}. \quad (24)$$

<sup>1</sup>Temperature data obtained from <https://www.ncdc.noaa.gov>



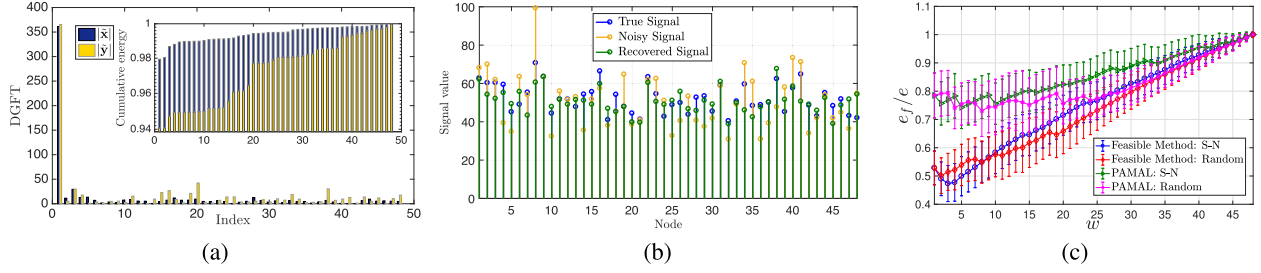


Fig. 12. Denoising a temperature signal supported in the contiguous US digraph in Fig. 11. (a) DGFT of the original signal ( $\tilde{x}$ ) and the noisy signal ( $\tilde{y}$ ), along with their cumulative energy distribution across frequencies. (b) A sample realization of the true, noisy, and recovered temperature signal for  $w = 3$ . (c) Ratio of average recovery error using a low-pass filter to recovery error without filtering versus the window size for the two digraphs with South to North (S-N) and random directed edges using the feasible method and PAMAL algorithm in [31]. As expected, on average the proposed method outperforms the PAMAL algorithm. Moreover, the lowest error is obtained when filtering in the S-N digraph, which better captures the temperature flow from states with lower to higher latitudes.

Note that the filter  $\mathbf{H} := \mathbf{U} \text{diag}(\tilde{\mathbf{h}}) \mathbf{U}^T$  will in general not be expressible as a polynomial of the graph's adjacency matrix  $\mathbf{A}$  (or some other graph-shift operator [29]), since DGFT modes need not be eigenvectors of the graph. Such a structure can be desirable to implement the filtering operation in a distributed fashion, and polynomial graph filter approximations of arbitrary lineal operators like  $\mathbf{H}$  have been studied in [33].

Fig. 12(a) compares the original signal and the noisy measurements in the graph spectral domain induced by the DGFT. The original signal is low-pass bandlimited, compared to the noisy signal which spans a broader range of frequencies due to the white noise. To better observe the low-pass property of the original signal, we also plot the cumulative energy of both the original and the noisy signals, defined by the percentage of the total energy present in the first  $i$  frequency components for  $i = 1, \dots, N$ . It is apparent from Fig. 12(a) that the first few components of  $\mathbf{x}$  capture most of its energy.

Fig. 12(b) shows a realization of the noisy graph signal  $\mathbf{y}$  superimposed with the denoised temperature profile  $\hat{\mathbf{x}}$  obtained using (24) with  $w = 3$ , and the original signal  $\mathbf{x}$ . Filter design and the choice of  $w$  is beyond the scope of this paper, but the average recovery error  $e_f = \|\hat{\mathbf{x}} - \mathbf{x}\|/\|\mathbf{x}\|$ , over 1000 Monte-Carlo simulations of independent noise, attains a minimum of approximately 12% and Fig. 12(b) shows  $\hat{\mathbf{x}}$  closely approximates  $\mathbf{x}$ .

For this denoising task, we compare Algorithm 2 with a state-of-the-art approach in [31]. The PAMAL algorithm [31, Algorithm 2] fails to converge to an orthonormal basis within a reasonable time; however, we terminate the routine after 100 iterations and work with the obtained basis. To assess the importance of the network model, we repeat the whole denoising experiment using a baseline contiguous US digraph (cf. Fig. 11), where we choose the direction of edges uniformly at random. To assess performance, we compute the relative recovery error with and without low-pass filtering as  $e_f$  and  $e = \|\mathbf{n}\|/\|\mathbf{x}\|$ , respectively. Fig. 12(c) depicts  $e_f/e$  versus  $w$  averaged over 1000 Monte-Carlo simulations, which demonstrates the effectiveness of graph filtering in the dual domain enabled by the DGFT. Notice that our proposed approach outperforms the PAMAL method for both network models. Also, the minimum recovery error justifies the choice of  $w = 3$ . The results from the feasible method (S-N graph) are consistent with the energy plot, as the first 3–4 components of  $\mathbf{x}$  capture almost 99% of the energy, and increasing the window size will only enhance

the noise. As expected, Fig. 12(c) shows that, on average, the performance degrades for the randomized US network using the feasible method, because the S-N digraph in Fig. 11 better captures the temperature flow. While not shown here to avoid repetition, similar results with slightly higher recovery errors can be obtained using the greedy Algorithm 3.

For additional comparisons we also consider the following state-of-the-art GFT approaches for digraphs: i) filter design using adjacency matrix [30]; and ii) the directed Laplacian in [39]. We apply the filter design in [30, Section V-B] to both the adjacency matrix and directed Laplacian defined as  $\mathbf{D}_{\text{in}} - \mathbf{A}$ , where  $\mathbf{D}_{\text{in}}$  is the diagonal in-degree matrix. In the constructed S-N digraph, the adjacency matrix and directed Laplacian have only three and eight distinct eigenvalues, respectively. This limits the degrees of freedom for filter design, in which the low-pass graph filter is constructed through inverse polynomial interpolation using the distinct eigenvalues of the graph-shift operator. Such an ill-posed design gives rise to filters that amplify various high-frequency modes where only noise is present, resulting in large reconstruction errors. Furthermore, the Jordan decomposition required in [39] is numerically unstable, an undesirable effect that can be reduced using the Schur decomposition as proposed in [10]. Although both the Jordan decomposition and the Schur-based block diagonalization preserve the subspaces for each eigenvalue, the degrees-of-freedom challenge remains as we only have a few distinct eigenvalues. Since the recovery errors obtained using [30, Section V] or [39] are very variable and orders of magnitude higher than those reported in Fig. 12(c) (e.g., the minimum average error over 1000 realizations is  $4.58 \times 10^4$  for the directed Laplacian), we have decided against plotting them to avoid hindering the clarity of the figure. Overall, these comparisons corroborate the merits of adopting spread (hence distinct) and orthogonal frequency modes, that also capture the notion of signal variation over a digraph.

Lastly, we compare our method against a Laplacian-based denoiser on the *undirected* version of the contiguous US graph, namely  $\mathcal{G}^u$ . The said method utilizes the eigenvectors of the Laplacian to implement an ideal low-pass filter as in (24). For zero-mean Gaussian noise with variance 10 at each node, the Laplacian eigenvectors result in a recovery error comparable to the one obtained via DGFT-based filtering on the S-N digraph (approximately 12%). However, for other noise distributions (e.g., exponential, Poisson, Gamma, and uniform) the proposed approach outperforms the Laplacian-based low-pass

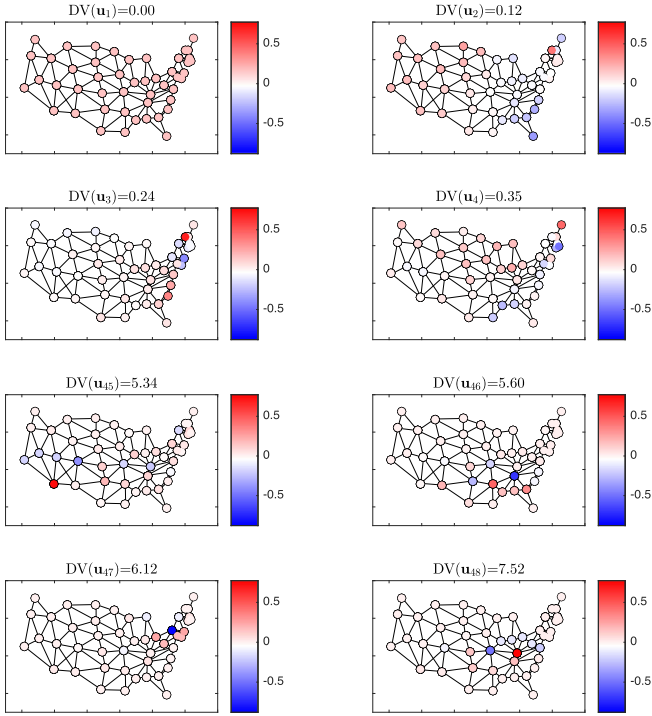


Fig. 13. DGFT basis vectors obtained using Algorithm 2 for the contiguous United States digraph in Fig. 11, along with their respective directed variation values (frequencies). Eight frequency components are shown, corresponding to the lowest and highest four frequencies in the graph.

filter on  $\mathcal{G}^u$ . For example, the most notable performance difference is observed for additive noise drawn from an exponential distribution with mean 5 (i.e., rate  $1/5$ ). In this case, our method results in a minimum average recovery error of approximately 9% (the optimal bandwidth is  $w = 7$ ), while the Laplacian-based denoiser on  $\mathcal{G}^u$  still yields an error of around 12%. In conclusion, the performance degrades using the Laplacian of  $\mathcal{G}^u$ , especially for non-Gaussian noise where a digraph model (and induced DV-based denoiser) is more effective. Similar general trends were obtained for other graph-shift operators on  $\mathcal{G}^u$ .

## VII. CONCLUSION

We considered the problem of finding an orthonormal set of graph Fourier basis signals for digraphs. The starting point was to introduce a novel measure of directed variation to capture the notion of frequency on digraphs. Our DGFT design is to construct orthonormal frequency modes that take into account the underlying digraph structure, span the entire frequency range, and that are as evenly distributed as possible in the graph spectral domain to better capture notions of low, medium and high frequencies. 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 of orthonormal basis vectors. To tackle the resulting non-convex problems, we developed two algorithms with complementary strengths to compute near-optimal solutions. First, we used a feasible method for optimization with orthogonality constraints, which offers provable convergence guarantees to stationary points of the spectral dispersion function. Second, we proposed a greedy heuristic to approximately minimize this dispersion using the eigenvectors of the Laplacian matrix of the

underlying undirected graph. The greedy algorithm offers theoretical approximation guarantees by virtue of matroid theory and results for submodular function optimization. The overall DGFT construction pipeline is validated on a synthetic digraph with three communities as well as on a structural brain network. Finally, we show how the proposed DGFT facilitates the design of a low-pass filter used to denoise a real-world temperature signal supported on a network of the US contiguous states.

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 here). Furthermore, it would be valuable to quantify or bound the optimality gap for the stationary solution of the feasible method in the Stiefel manifold.

## APPENDIX A

### FEASIBILITY OF PROBLEM (8)

The following proposition ensures that the spectral dispersion minimization problem (8) is feasible.

**Proposition 4:** The unit-norm basis vector  $\mathbf{u}_{\max}$  defined in (6) is orthogonal to the constant vector  $\mathbf{u}_{\min} := \frac{1}{\sqrt{N}}\mathbf{1}_N$ .

*Proof:* Since  $\mathbf{u}_{\min} := \frac{1}{\sqrt{N}}\mathbf{1}_N$ , we will show that  $\mathbf{u}_{\max}^T \mathbf{1}_N = 0$ . Arguing by contradiction, suppose that the sum of the entries in  $\mathbf{u}_{\max}$  is not zero. We show that  $DV(\mathbf{u}_{\max})$  can be improved in that case, which contradicts the optimality of  $\mathbf{u}_{\max}$ .

Without loss of generality assume that  $\mathbf{u}_{\max}^T \mathbf{1}_N = \epsilon > 0$ , and define  $\bar{\mathbf{u}} := \mathbf{u}_{\max} - \frac{\epsilon}{N}\mathbf{1}_N$ . First, note that  $DV(\mathbf{u}_{\max}) = DV(\bar{\mathbf{u}})$ , since adding (subtracting) a constant to (from) all coordinates will not change the directed variation. Second,

$$\|\bar{\mathbf{u}}\|^2 = \mathbf{u}_{\max}^T \mathbf{u}_{\max} - \frac{2\epsilon}{N} \mathbf{u}_{\max}^T \mathbf{1}_N + \left(\frac{\epsilon}{N}\right)^2 \mathbf{1}_N^T \mathbf{1}_N = 1 - \frac{\epsilon^2}{N}.$$

Therefore, we have a new vector  $\bar{\mathbf{u}}$  with the same directed variation but smaller norm. Now we can scale this vector as  $\alpha\bar{\mathbf{u}}$  (with  $\alpha > 1$ ) to obtain a normalized vector with  $DV(\alpha\bar{\mathbf{u}}) = \alpha^2 DV(\bar{\mathbf{u}})$ , which improves upon  $\mathbf{u}_{\max}$ . ■

## APPENDIX B

### PROOF OF PROPOSITION 1-1

We prove by induction (on the length of path) that the maximum frequency on a dipath is twice the maximum edge weight. Let  $x_1, x_2, \dots, x_N$  be the signal values on the dipath of length  $N - 1$ , with directed edges going from  $i$  to  $i + 1$ ,  $i = 1, \dots, N - 1$ . For the base case of  $N = 2$ , we have to maximize  $A_{12}(x_1 - x_2)^2$  subject to  $x_1 \geq x_2$  and  $x_1^2 + x_2^2 = 1$ . The solution to this optimization problem is  $x_1 = \sqrt{2}/2$  and  $x_2 = -\sqrt{2}/2$  which evaluates to  $A_{12}(x_1 - x_2)^2 = 2A_{12}$ . For the inductive step, assume that the claim is true for a dipath of length  $N - 1$ . We show that it should be the case for  $N$  edges as well. If  $x_N \leq x_{N+1}$  in the optimal solution for  $N$  edges, then for the last edge  $[x_N - x_{N+1}]_+ = 0$  and the optimal directed variation is obtained from the first  $N - 1$  edges, which is twice their largest edge weight by assumption. Indeed, note that  $A_{N(N+1)}$  cannot be the maximum edge weight in this case, otherwise setting  $x_N = \sqrt{2}/2$  and  $x_{N+1} = -\sqrt{2}/2$  would improve the optimal solution and violates the

assumption of  $x_N \leq x_{N+1}$ . Next, we assume  $x_N > x_{N+1}$ . In this case, we claim that  $x_N$  should also be greater than or equal to  $x_{N-1}$ . If not, we have  $x_{N-1} > x_N > x_{N+1}$ , which cannot be an optimal solution. To see this, we can swap the value of  $x_N$  with either  $x_{N-1}$  or  $x_{N+1}$  and improve the directed variation because either  $A_{(N-1)N}(x_{N-1} - x_{N+1})^2$  or  $A_{N(N+1)}(x_{N-1} - x_{N+1})^2$  is greater than  $A_{(N-1)N}(x_{N-1} - x_N)^2 + A_{N(N+1)}(x_N - x_{N+1})^2$ . Finally, if  $x_N > x_{N+1}$  and  $x_N \geq x_{N-1}$ , then the edge  $(N-1, N)$  does not contribute to the directed variation and the path is divided into two sections. Since both  $\|\mathbf{u}\|^2$  and  $DV(\mathbf{u})$  scale quadratically with  $\mathbf{u}$ , one can show (in general) that one of the optimal solutions should be achieved by only the variations of a set of connected edges; otherwise it is better to void the section with the lower ratio of objective to norm, and scale up the other section. This means that in our dipath example, once the edge  $(N-1, N)$  has zero objective, we can also make one of the two sections zero. The claim is then true by the inductive assumption. The achievability of the maximum directed variation follows by setting  $\pm\sqrt{2}/2$  on the edge with largest weight. ■

#### APPENDIX C

##### PROOF OF PROPOSITION 1-2

There should be at least one edge  $(i, j)$  in the cycle for which  $[x_i - x_j]_+ = 0$ , otherwise we obtain a closed loop of strict inequalities among consecutive  $x_i$  values which is impossible. Given that edge  $(i, j)$  has zero directed variation, the rest of the cycle can be viewed as a dipath which has directed variation of at most 2 times the largest edge weight by Proposition 1-1). The same argument ensures the achievability of the solution. ■

#### APPENDIX D

##### PROOF OF PROPOSITION 1-3

Let  $\mathcal{V}^+$  and  $\mathcal{V}^-$  be the two node partitions of the unidirectional bipartite graph, where the edges are constrained to go from  $\mathcal{V}^+$  to  $\mathcal{V}^-$ . First, we show that in the optimal solution maximizing the directed variation we must have  $x_i \geq 0$  for all  $i \in \mathcal{V}^+$ , and  $x_i \leq 0$  for all  $i \in \mathcal{V}^-$ . Otherwise, assume that there exists some node  $j \in \mathcal{V}^-$  with  $x_j > 0$ . Then we can improve the directed variation by setting  $x_j = 0$ , because  $j$  has only incoming edges and decreasing  $x_j$  will not decrease the variation on such edges (and we gain some slack in the norm constraint by this change). Similarly, we arrive at a contradiction if some node in  $\mathcal{V}^+$  has negative value.

With this information, we know that all the summands  $A_{ij}[x_i - x_j]_+^2$  in the objective are indeed equal to  $A_{ij}(x_i - x_j)^2$ , because  $i \in \mathcal{V}^+$ ,  $j \in \mathcal{V}^-$ , and  $x_i \geq 0 \geq x_j$ . Therefore, we can replace the cost function with the total variation and solve the following optimization problem instead

$$\begin{aligned} \max_{\mathbf{x}} \quad & TV(\mathbf{x}) = \mathbf{x}^T \mathbf{L} \mathbf{x} \\ \text{subject to} \quad & \mathbf{x}^T \mathbf{x} = 1 \\ & x_i \geq 0, \quad i \in \mathcal{V}^+ \\ & x_j \leq 0, \quad j \in \mathcal{V}^- \end{aligned} \quad (25)$$

Assume that we relax problem (25) by dropping the inequality constraints. Once we do that, the solution will be  $\lambda_{\max}$ . The next lemma shows this relaxation entails no loss of optimality.

**Lemma 3:** For an undirected bipartite graph  $\mathcal{G} = (\mathcal{V}_1, \mathcal{V}_2, \mathbf{A})$  with Laplacian matrix  $\mathbf{L}$ , let  $\mathbf{u} \in \mathbb{R}^N$  be the dominant eigenvector of  $\mathbf{L}$  (corresponding to  $\lambda_{\max}$ ). Then  $\mathbf{u}$  has the same sign over the coordinates of each partition (i.e., non-negative for  $\mathcal{V}_1$  and non-positive for  $\mathcal{V}_2$  or vice versa).

**Proof:** If by contradiction that were not the case, we could change the signs (maintaining the absolute values and vector norm) to be positive in one partition (say  $\mathcal{V}_1$ ) and negative in the other ( $\mathcal{V}_2$ ). This change does not decrease the absolute difference between signal values on nodes incident to each edge, and contradicts the fact that  $\mathbf{u}$  maximizes  $TV(\mathbf{u}) = \mathbf{u}^T \mathbf{L} \mathbf{u} = \sum_{i,j=1,j>1}^N A_{ij}(u_i - u_j)^2$ . ■

By virtue of Lemma 3, either  $\mathbf{u}$  or  $-\mathbf{u}$  is feasible in (25) and attains the optimum objective value  $\lambda_{\max}$ . ■

#### REFERENCES

- [1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2009.
- [2] G. Calinescu, C. Chekuri, M. Pál, and J. Vondrák, "Maximizing a monotone submodular function subject to a matroid constraint," *SIAM J. Comput.*, vol. 40, no. 6, pp. 1740–1766, 2011.
- [3] S. Chen, A. Sandryhaila, J. M. F. Moura, and J. Kovačević, "Signal recovery on graphs: Variation minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4609–4624, Sep. 2015.
- [4] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6510–6523, Dec. 2015.
- [5] F. Chung, "Laplacians and the Cheeger inequality for directed graphs," *Ann. Combinatorics*, vol. 9, no. 1, pp. 1–19, 2005.
- [6] J. A. Deri and J. M. F. Moura, "New York City taxi analysis with graph signal processing," in *Proc. IEEE Global Conf. Inf. Process.*, Washington, DC, USA, Dec. 2016, pp. 1275–1279.
- [7] 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, Sep. 2017.
- [8] M. L. Fisher, G. L. Nemhauser, and L. A. Wolsey, "An analysis of approximations for maximizing submodular set functions—I," *Math. Program.*, vol. 14, no. 1, pp. 265–294, 1978.
- [9] A. Gavili and X. P. Zhang, "On the shift operator, graph frequency, and optimal filtering in graph signal processing," *IEEE Trans. Signal Process.*, vol. 65, no. 23, pp. 6303–6318, Dec. 2017.
- [10] B. Girault, "Signal processing on graphs—contributions to an emerging field," Ph.D. dissertation, Ecole Normale Supérieure de Lyon-ENS LYON, Lyon, France, 2015.
- [11] B. Girault, "Stationary graph signals using an isometric graph translation," in *Proc. Eur. Signal Process. Conf.*, Nice, France, Aug. 2015, pp. 1516–1520.
- [12] B. Girault, A. Ortega, and S. Narayanan, "Irregularity-aware graph Fourier transforms," *IEEE Trans. Signal Process.*, vol. 66, no. 4, pp. 5746–5761, 2018.
- [13] C. J. Honey, R. Kötter, M. Breakspear, and O. Sporns, "Network structure of cerebral cortex shapes functional connectivity on multiple time scales," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 24, pp. 10240–10245, 2007.
- [14] 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, Oct. 2016.
- [15] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Autoregressive moving average graph filtering," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 274–288, Jan. 2017.
- [16] J. A. Kelner and E. Nikolova, "On the hardness and smoothed complexity of quasi-concave minimization," in *Proc. IEEE Symp. Found. Comput. Sci.*, 2007, pp. 472–482.
- [17] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models* (Series in Statistics). Berlin, Germany: Springer, 2009.
- [18] L. Le Magoarou, R. Gribonval, and N. Tremblay, "Approximate fast graph Fourier transforms via multilayer sparse approximations," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 2, pp. 407–420, Jun. 2018.



- [19] J. Lee, V. S. Mirrokni, V. Nagarajan, and M. Sviridenko, "Non-monotone submodular maximization under matroid and knapsack constraints," in *Proc. ACM Symp. Theory Comput.*, 2009, pp. 323–332.
- [20] G. Leus, S. Segarra, A. Ribeiro, and A. G. Marques, "The dual graph shift operator: Identifying the support of the frequency domain," 2017, arXiv:1705.08987 [cs.IT].
- [21] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1832–1843, Apr. 2016.
- [22] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary graph processes and spectral estimation," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5911–5926, Nov. 2017.
- [23] S. Mittal and A. S. Schulz, "An FPTAS for optimizing a class of low-rank functions over a polytope," *Math. Program.*, vol. 141, pp. 103–120, 2013.
- [24] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions—I," *Math. Program.*, vol. 14, no. 1, pp. 265–294, 1978.
- [25] J. Nocedal and S. J. Wright, *Numerical Optimization* (Series in Operations Research and Financial Engineering). Berlin, Germany: Springer, 2006.
- [26] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vanderghenst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, 2018.
- [27] N. Perraudin and P. Vanderghenst, "Stationary signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3462–3477, Jul. 2017.
- [28] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.
- [29] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.
- [30] 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, Jun. 2014.
- [31] 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, Sep. 2017.
- [32] A. Schrijver, *Combinatorial Optimization: Polyhedra and Efficiency*, vol. 24. Berlin, Germany: Springer, 2002.
- [33] S. Segarra, A. G. Marques, and A. Ribeiro, "Optimal graph-filter design and applications to distributed linear network operators," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4117–4131, Aug. 2017.
- [34] S. Segarra, G. Mateos, A. G. Marques, and A. Ribeiro, "Blind identification of graph filters," *IEEE Trans. Signal Process.*, vol. 65, no. 5, pp. 1146–1159, Mar. 2017.
- [35] R. Shafipour, A. Khodabakhsh, G. Mateos, and E. Nikolova, "A digraph Fourier transform with spread frequency components," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Montreal, QC, Canada, Nov. 2017, pp. 583–587.
- [36] R. Shafipour, A. Khodabakhsh, G. Mateos, and E. Nikolova, "Digraph Fourier transform via spectral dispersion minimization," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Calgary, AB, Canada, Apr. 2018, pp. 6284–6288.
- [37] R. Shafipour, S. Segarra, A. G. Marques, and G. Mateos, "Network topology inference from non-stationary graph signals," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, New Orleans, LA, USA, Mar. 2017, pp. 5870–5874.
- [38] D. I. Shuman, C. Wiesmeyr, N. Holighaus, and P. Vanderghenst, "Spectrum-adapted tight graph wavelet and vertex-frequency frames," *IEEE Trans. Signal Process.*, vol. 63, no. 16, pp. 4223–4235, Aug. 2015.
- [39] R. Singh, A. Chakraborty, and B. Manoj, "Graph Fourier transform based on directed Laplacian," in *Proc. IEEE Int. Conf. Signal Process. Commun.*, 2016, pp. 1–5.
- [40] O. Teke and P. P. Vaidyanathan, "Extending classical multirate signal processing theory to graphs-Part I: Fundamentals," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 409–422, Jan. 2017.
- [41] 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, Aug. 2014.
- [42] N. Tremblay, P. Gonçalves, and P. Borgnat, "Design of graph filters and filterbanks," *Cooperative and Graph Signal Processing*. Amsterdam, The Netherlands: Elsevier, 2018, pp. 299–324.
- [43] Z. Wen and W. Yin, "A feasible method for optimization with orthogonal constraints," *Math. Program.*, vol. 142, no. 1/2, pp. 397–434, 2013.
- [44] X. Zhu and M. Rabbat, "Approximating signals supported on graphs," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Kyoto, Japan, Mar. 2012, pp. 3921–3924.



**Rasoul Shafipour** (S'15) received the B.Sc. degree in electrical engineering (communication systems) from the Sharif University of Technology, Tehran, Iran, in 2015 and the M.Sc. degree from the University of Rochester, Rochester, NY, in 2017. Since August 2015, he has been working toward the Ph.D. degree in electrical and computer engineering at the University of Rochester. His research interests include signal processing on graphs, network science, machine learning, optimization, and algorithms. He obtained the sixth rank in the Iranian nation-wide university entrance exam among more than 270,000 participants in 2010. He is a recipient of the University of Rochester's Donald M. and Janet C. Barnard fellowship and a best student paper award at the 2018 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).



**Ali Khodabakhsh** (S'16) received the B.Sc. degree in electrical engineering (communication systems) from the Sharif University of Technology, Tehran, Iran, in 2014, and the M.Sc. degree in electrical and computer engineering from the University of Texas (UT) at Austin, Austin, TX, USA, in 2017. He is currently working toward the Ph.D. degree with UT Austin, where he is affiliated with the Wireless Networking and Communications Group. During Spring 2018, he was a visiting scholar with the Simons Institute, the University of California Berkeley, Berkeley, CA, USA. His research interest is in combinatorial optimization and approximation algorithms with focus on applications in signal processing, power systems, and electricity market. He was the recipient of a best student paper award at ICASSP 2018.



**Gonzalo Mateos** (SM'17) received the B.Sc. degree in electrical engineering from Universidad de la República (Udelar), Montevideo, Uruguay, in 2005, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Minnesota (UofM), Minneapolis, MN, USA, in 2009 and 2011, respectively. Currently, he is an Assistant Professor with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA, where he is also affiliated with the Geogron Institute for Data Science. From 2003 to 2006, he was an Assistant with the Department of Electrical Engineering, Udelar. From 2004 to 2006, he was a System Engineer with Asea Brown Boveri (ABB), Montevideo, Uruguay. During 2013, he was a visiting scholar with the CS Department, Carnegie Mellon University. His research interests lie in the areas of statistical learning from Big Data, network science, wireless communications, and graph signal processing. He is currently an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS, and is a member of the IEEE SigPort Editorial Board. He was the recipient of the NSF CAREER Award in 2018, the 2017 IEEE Signal Processing Society Young Author Best Paper Award (as senior co-author), and Best Paper Awards at SPAWC 2012, SSP 2016, and ICASSP 2018. His doctoral work has been recognized with the 2013 UofM's Best Dissertation Award (Honorable Mention) across all Physical Sciences and Engineering areas.



**Evdokia Nikolova** received the B.A. degree in applied mathematics with economics from Harvard University, Cambridge, MA, USA, the M.S. degree in mathematics from Cambridge University, Cambridge, U.K., and the Ph.D. degree in computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. She is an Assistant Professor with the Department of Electrical and Computer Engineering, the University of Texas at Austin, Austin, TX, USA, where she is a member of the Wireless Networking and Communications Group. Her research aims to improve the design and efficiency of complex systems (such as networks and electronic markets), by integrating stochastic, dynamic, and economic analysis. Her recent work examines how human risk aversion transforms traditional computational models and solutions. One of her algorithms has been adapted in the MIT CarTel project for traffic-aware routing. She currently focuses on developing algorithms for risk mitigation in networks, with applications to transportation and energy. She is a recipient of an NSF CAREER award.