A Directed Graph Fourier Transform with Spread Frequency Components

Rasoul Shafipour, Student Member, IEEE, Ali Khodabakhsh, Student Member, IEEE, Gonzalo Mateos, 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: (i) find the maximum directed variation (i.e., a novel notion of frequency on a digraph) a candidate basis vector can attain; and (ii) minimize a smooth spectral dispersion function over the achievable frequency range to obtain the desired spread DGFT basis. Both steps involve non-convex, orthonormality-constrained optimization problems, which are efficiently tackled via a provably convergent, feasible optimization method on the Stiefel manifold. We 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 [11], vehicle trajectories over road networks [6], or, infectious states of individuals in a population affected by an epidemic [15], 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 [23], [26]. From this vantage point, generalizations of traditional signal processing tasks such as filtering [13], [26], [30], [38], [40], sampling and reconstruction [4], [18], statistical GSP and spectrum estimation [9], [19], [24], (blind) filter identification [31], [34] as well as signal representations [35], [39], [42] have been recently explored under the GSP purview [23].

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 basis 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 [27], [28]. We first propose a novel notion of signal variation (frequency) over digraphs and find an approximation of the maximum possible frequency \( f_{\text{max}} \) that a unit-norm graph signal can achieve.

We design a digraph (DGFT) such that the resulting frequencies (i.e., the directed variation of the sought orthonormal bases) distribute as evenly as possible across \([0, f_{\text{max}}]\). 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 [17] 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 \( \mathcal{G} = (\mathcal{V}, \mathcal{A}) \), where \( \mathcal{V} \) is the set of nodes (i.e., vertices) with cardinality \( |\mathcal{V}| = N \), and \( \mathcal{A} \subseteq \mathbb{R}^{N \times N} \) 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 \( \mathcal{A} \) is symmetric, and the positive semi-definite Laplacian matrix takes the form \( \mathcal{L} = \mathcal{D} - \mathcal{A} \), where \( \mathcal{D} \) is the diagonal degree matrix with \( D_{ii} = \sum_j A_{ij} \). A graph signal \( x : \mathcal{V} \mapsto \mathbb{R}^N \) 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 \( x \) is often defined as \( \hat{x} = \mathcal{V}^T x \), where \( \mathcal{V} = [v_1, \ldots, v_N] \) comprises the eigenvectors of the Laplacian \( \mathcal{L} \). 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 $x$ with respect to the Laplacian $L$ as
\begin{equation}
TV(x) = x^T L x = \sum_{i,j=1}^{N} A_{i,j} (x_i - x_j)^2. \tag{1}
\end{equation}
The total variation $TV(x)$ is a smoothness measure, quanti-
ifying how much the signal $x$ changes with respect to the
expectation on variability that is encoded by the weights $A$.
Consider the total variation of the eigenvectors $v_k$, which is
given by $TV(v_k) = \lambda_k$, the $k$th Laplacian eigenvalue. Hence,
eigenvectors $0 = \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_N$ can be viewed as graph
frequencies, indicating how the GFT bases (i.e., the frequency
components) vary over the graph $G$. Note that there may be
more than one eigenvector corresponding to a graph frequency
in case of having repeated eigenvalues. Moreover, frequency
components associated with close eigenvalues can often be
quite dissimilar and focus on different parts of the graph $[27]$.

Extensions of the combinatorial Laplacian to digraphs have
also been proposed $[3], [30]$. 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 $A = VJV^{-1}$, where the frequency repre-
sentation of signal $x$ is $\tilde{x} = V^{-1}x$ $[27]$. While valid for
digraphs, the associated notion of signal variation in $[27]$ does
not ensure that constant signals have zero variation. Moreover,$V$ 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
$[7]$. Recently, a fresh look at the GFT for digraphs was put
forth in $[28]$ based on minimization of the (convex) Lovász
extension of the graph cut size, subject to orthonormality
constraints on the desired bases. While the GFT bases in $[28]$
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 $G$; see Section $[III-A]
for an example of this phenomenon. An encompassing GFT
framework was proposed in $[10]$, 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
bases 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 $[0, f_{\text{max}}]$,
and thus better capture low, middle, and high frequencies.
P3) Bases 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 bases. Motivation and challenges associated
with the proposed spectral-dispersion minimization approach
are discussed in Section $[III]$ We then propose two algorithm-
ic 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 $G$. In this
setting, we show that the spectral-dispersion minimization
problem can be cast as supermodular (frequency) set optimiza-
tion. 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. Numerical
tests with both synthetic and real-world digraphs corroborate
the effectiveness of our DGFT algorithms in yielding (near)
maximally-spread frequency components (Section $[VI]$). Conclu-
sembling remarks are given 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
$X$ and a (column) vector $x$ are denoted by $X_{ij}$ and $x_i$, respectively. Sets are represented by calligraphic capital letters.

The notation $T$ stands for transposition and $1_N$ represents the
$N \times N$ identity matrix, while $1_N x$ denotes the $N \times 1$ vector
of all ones. For a vector $x$, $\text{diag}(x)$ is a diagonal matrix whose
ith diagonal entry is $x_i$. Lastly, $\|x\| = (\sum_i x_i^2)^{1/2}$ denotes the Euclidean norm of $x$, while $\text{trace}(X) = \sum_i X_{ii}$ stands
for the trace of a square matrix $X$ whose Frobenius norm is
$\|X\|_F = \text{trace}(XX^T)^{1/2}$.

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

Our goal is to find $N$ orthonormal bases capturing different
modes of variation over the graph $G$. We collect these desired
bases in a matrix $U := [u_1, \ldots, u_N] \in \mathbb{R}^{N \times N}$, where $u_k \in
\mathbb{R}^N$ represents the kth frequency component. This means that
the DGFT of a graph signal $x \in \mathbb{R}^N$ with respect to $G$ is the
signal $\tilde{x} = [\tilde{x}_1, \ldots, \tilde{x}_N]^T$ defined as $\tilde{x} = U^Tx$. The inverse
(i)DGFT of $x$ is given by $x = U \tilde{x} = \sum_{k=1}^N \tilde{x}_k u_k$, which
allows one to synthesize $x$ as a sum of orthogonal frequency
components $u_k$. The contribution of $u_k$ to the signal $x$ is the
DGFT coefficient $\tilde{x}_k$.

For undirected graphs, the quantity $TV(x)$ in $[11]$ measures
how signal $x$ varies over the network with Laplacian $L$. This
motivates defining a more general notion of signal variation
for digraphs, called directed variation (DV), as
\begin{equation}
DV(x) := \sum_{i,j=1}^{N} A_{i,j} [x_i - x_j]^2, \tag{2}
\end{equation}
where \([x]_+ := \max(0, x)\) denotes projection onto the non-negative reals. To gain insight on \([6]\), consider a graph signal \(x \in \mathbb{R}^N\) 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 \(\text{DV}(x)\) only if \(x_i > x_j\). Accordingly, one in general has that \(\text{DV}(x) \neq \text{DV}(-x)\) and we will exploit this property later. Notice that if \(\mathcal{G}\) is undirected, then \(\text{DV}(x) \equiv \text{TV}(x)\).

Analogously to the undirected case, we define the frequency \(f_k := \text{DV}(\mathbf{u}_k)\) as the directed variation of the basis \(\mathbf{u}_k\).

**B. Challenges facing spread frequency components**

Similar to the discrete spectrum of periodic time-varying signals, by designing the bases we would ideally like to have \(N\) equidistributed graph frequencies forming an arithmetic sequence

\[
 f_k = \text{DV}(\mathbf{u}_k) = \frac{k - 1}{N - 1} f_{\text{max}}, \quad k = 1, \ldots, N
\]

where \(f_{\text{max}}\) 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

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

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

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

where \(\lambda_{\text{max}}\) 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

\[
 \mathbf{u}_{\text{max}} = \arg\max_{\|\mathbf{u}\| = 1} \text{DV}(\mathbf{u}) \quad \text{and} \quad f_{\text{max}} := \text{DV}(\mathbf{u}_{\text{max}}).
\]

To relate the maximum frequencies in \([3]\) and \([6]\), for a given digraph \(\mathcal{G} = (\mathcal{V}, \mathcal{A})\) consider its underlying undirected graph \(\mathcal{G}^u = (\mathcal{V}, \mathcal{A}^u)\), 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 \(f_{\text{max}}\) is upper-bounded by \(f_{\text{max}}^u = \lambda_{\text{max}}\), the spectral radius of the Laplacian of \(\mathcal{G}^u\). This is because dropping the direction of any edge can not decrease the directed variation of a signal; see also Section **III-B** for additional insights on the maximum signal variation over a given digraph.

**C. Problem statement**

Going back to the design of \(\mathbf{U}\), to cover the whole spectrum of variations one would like to set \(\mathbf{u}_1 = \mathbf{u}_{\text{min}} := \frac{1}{\sqrt{N}} \mathbf{1}_N\) (normalized all ones vector of length \(N\); i.e., a constant signal) and \(\mathbf{u}_N = \mathbf{u}_{\text{max}}\) in \([6]\). As a criterion for the design of the remaining bases, consider the spectral dispersion function

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

that measures how well spread the corresponding frequencies are over \([0, f_{\text{max}}]\). 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 \(\text{DV}(\mathbf{u}_N) = f_{\text{max}}\), consistent with our design goal.

Rather than going after frequencies exactly equidistributed as in \([6]\), our idea is to minimize the spectral dispersion

\[
 \min_{\mathbf{U}} \sum_{i=1}^{N-1} [\text{DV}(\mathbf{u}_{i+1}) - \text{DV}(\mathbf{u}_i)]^2
\]

subject to \(\mathbf{U}^T \mathbf{U} = \mathbf{I}_N\),

\[
 \mathbf{u}_1 = \mathbf{u}_{\text{min}}, \quad \mathbf{u}_N = \mathbf{u}_{\text{max}}.
\]

Problem \([8]\) is feasible since we show in Appendix \([A]\) that \(\mathbf{u}_{\text{max}}\) defined in \([6]\) is orthogonal to the constant vector \(\mathbf{u}_{\text{min}} = \frac{1}{\sqrt{N}} \mathbf{1}_N\). However, finding the global optimum of \([8]\) is challenging due to the non-convexity arising from the orthonormality (Stiefel manifold) constraints. The objective function \(\delta(\mathbf{U})\) 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 \([41]\), to solve judiciously modified forms of problems \([6]\) and \([8]\) to directly find the maximum frequency along with the disperse bases. 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 \([7]\). 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.

1. Similar to \([28]\), the focus here is on designing a real-valued DGFT basis. Still, the optimization methods in Section **IV** can accommodate complex-valued Stiefel manifold constraints, and generalizations of the directed variation measure in \([3]\) can be devised to e.g., recover the DFT basis when \(\mathcal{G}\) is the directed cycle representing the support of periodic discrete-time signals.
analytically, and then provide a general 1/2-approximation to \( f_{\text{max}} \) 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

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

was introduced in [28] as the convex Lovász extension of the graph cut size, whose minimization can facilitate identifying graph clusters [15, Ch. 4]. Different from [28], the directed variation measure (9) is not smooth and for undirected graphs it boils down to a so-called graph absolute variation

\[
TV_1(x) := \sum_{i,j=1}^{N} A_{ij} |x_i - x_j| \quad \text{[cf. (1)]}; \tag{10}
\]

To obtain a GFT for digraphs, the approach in [28] is to solve the orthogonality-constrained problem

\[
\min_{U} \sum_{i=1}^{N} DV'(u_i), \quad \text{subject to } U^T U = I_N. \tag{10}
\]

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 \( G \).

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

\[
U = \begin{bmatrix} 0.5 & c & c & c \\ 0.5 & a & 0 & b \\ 0.5 & b & a & 0 \\ 0.5 & b & a & 0 \end{bmatrix},
\]

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 \( U \). As a result, for all columns \( u_k \) of \( U \) one has \( DV'(u_k) = 0, \ k = 1, \ldots, 4 \), and hence the inverse GFT synthesis formula \( x = U^*x \) fails to offer an expansion of \( x \) 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 \( L \); see e.g., [23], [40], [42]. A pertinent question is whether the Laplacian-based GFT minimizes the spectral dispersion function in (7), where

\[
DV(x) \equiv TV(x) \quad \text{because } G \text{ is undirected}. \tag{11}
\]

To provide an answer, we consider the graph shown in Fig. 1 (right). Denote the eigenvectors of the Laplacian matrix as \( V = [v_1, v_2, v_3, v_4] \) with frequencies \( \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\} = [0, 1, 3, 4] \). To determine the DGFT basis \( U \) satisfying our design criteria in (8), we first set \( u_1 := v_1 \) and \( u_4 := v_4 \) 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 \( v_2 \) and \( v_3 \) span the hyperplane containing \( u_2 \) and \( u_3 \). There is a single degree of freedom to specify \( u_2 \) and \( u_3 \) on that plane, namely a simultaneous rotation of \( v_2 \) and \( v_3 \). Accordingly, all the feasible basis vectors \{\( u_2, u_3 \)\} will be of the form

\[
\begin{bmatrix} u_2 \\ u_3 \\ v_2 \\ v_3 \end{bmatrix} = R_\theta \begin{bmatrix} v_2 \\ v_3 \\ \cos \theta \sin \theta \\ \sin \theta \cos \theta \end{bmatrix}, \tag{11}
\]

where \( R_\theta \) rotates vectors counterclockwise by an angular amount of \( \theta \). Collecting the sought bases in \( U_\theta = [v_1, u_2, u_3, v_4] \), in Fig. 2 (a) we plot the dispersion function

\[
\delta(U_\theta) = \left( TV(v_4) - \max(TV(u_2), TV(u_3)) \right)^2 + \left( \max(TV(u_2), TV(u_3)) - \min(TV(u_2), TV(u_3)) \right)^2 + \left( \min(TV(u_2), TV(u_3)) - TV(v_1) \right)^2, \tag{12}
\]

as a function of \( \theta \). Note that the maximum and minimum operators in (12) are needed to sort the frequencies associated with \( u_2 \) and \( u_3 \). It is apparent from Fig. 2 (a) that \( \delta(V) \) 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 \( U := U_\theta|_{\theta = 0.4214} \).

To further compare the frequency components obtained, the horizontal lines in Fig. 2 (b) depict both sets of frequencies \{\( DV(v_k) \)\}_{k=1}^4 and \{\( DV(u_k) \)\}_{k=1}^4. As expected, the optimized GFT basis \( U \) gives rise to frequencies that are more uniformly spread in the graph spectral domain. Moreover, for this particular example the graph frequencies \{\( DV(u_k) \)\}_{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 III and IV
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 $G(V, A)$ is to calculate the maximum frequency $f_{\text{max}}$. The spherically-constrained problem (6) is 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 $f_{\text{max}}$ can be obtained analytically.

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

1) Let $G$ 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, $f_{\text{max}} = 2 \max_{i,j} A_{ij}$.

2) Let $G$ 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, $f_{\text{max}} = 2 \max_{i,j} A_{ij}$.

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

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

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

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

In Section V, we will revisit the result in Proposition 1 to motivate a greedy heuristic to construct a disperse DGFT basis from Laplacian eigenvectors of $G^u$. 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 1, we show the adopted framework for orthogonality-constrained optimization can be as well used to accurately approximate $f_{\text{max}}$.

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 $\mathbb{H}$. Specifically, following the specification in Section [41] we take a two step approach whereby: i) we find $f_{\text{max}}$ and its corresponding basis $u_{\text{max}}$ by solving (6); and ii) we solve (8) to find well-spread frequency components $U = [u_1, \ldots, u_N]$ satisfying $u_1 = u_{\text{min}} = \frac{1}{\sqrt{N}} \mathbf{1}_N$ and $u_N = u_{\text{max}}$. 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 [41] deals with an orthogonality constrained problem of the form

$$\min_{U \in \mathbb{R}^{N \times p}} \phi(U), \ \text{subject to} \ U^T U = I_p,$$

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

$$U_{k+1}(\tau) = \left( I_n + \frac{\tau}{2} B_k \right)^{-1} \left( I_n - \frac{\tau}{2} B_k \right) U_k,$$

where $B_k := G_k^T U_k - U_k G_k^T$ is a skew-symmetric ($B_k^T = -B_k$) projection of the gradient onto the constraint’s tangent space. Update rule (15) is known as the Cayley transform.
Algorithm 1 Directed Variation Maximization

1: Input: Adjacency matrix $A$ and parameter $\epsilon > 0$.
2: Initialize $k = 0$ and unit-norm $u_0 \in \mathbb{R}^N$ at random.
3: repeat
4:   Evaluate objective $\phi(u_k) := -DV(u_k)$ in (2).
5:   Compute gradient $g_k \in \mathbb{R}^N$ via (17).
6:   Form $B_k := g_k u_k^T - u_k g_k^T$.
7:   Select $\tau_k$ satisfying conditions (16a) and (16b).
8:   Update $u_{k+1}(\tau_k) = (I_N + \frac{\rho_1}{2}B_k)^{-1}(I_N - \frac{\rho_2}{2}B_k)u_k$.
9:   $k \leftarrow k + 1$.
10: until $\|u_k - u_{k-1}\| \leq \epsilon$
11: Return $u_{\text{max}} := u_k$ and $f_{\text{max}} := DV(u_{\text{max}})$.

which preserves orthogonality (i.e., $U_{k+1}^T U_{k+1} = I_p$), since $(I_u + \frac{\rho_1}{2}B_k)^{-1}$ and $I_u - \frac{\rho_2}{2}B_k$ commute. Other noteworthy properties of the update are: i) $U_{k+1}(0) = U_k$; ii) $U_{k+1}(\tau)_{\tau}$ is smooth as a function of the step size $\tau$; and iii) $\frac{d}{d\tau} U_{k+1}(0)$ is the projection of $-G_k$ into the tangent space of the Stiefel manifold at $U_k$.

Most importantly, iii) ensures that the update (15) is a descent path for a proper step size $\tau$, which can be obtained through a curvilinear search satisfying the Armijo-Wolfe conditions

$$\phi(U_{k+1}(\tau_k)) \leq \phi(U_{k+1}(0)) + \rho_1 \tau_k \phi'(U_{k+1}(0))$$

(16a)

$$\phi'(U_{k+1}(\tau_k)) \geq \rho_2 \phi'(U_{k+1}(0))$$

(16b)

where $0 < \rho_1 < \rho_2 < 0$ are two parameters [22]. One can show that if $\phi(U_{k+1}(\tau))$ is continuously differentiable and bounded below as is the case for problems (6) and (8), then there exists a $\tau_k$ satisfying (16a) and (16b). Moreover, the derivative of $\phi(U_{k+1}(\tau))$ at $\tau = 0$ is given by $\phi'(U_{k+1}(0)) = -1/2 \|B_k\|_F^2$; see [41] for additional details.

All in all, the iterations (15) are well defined and by implementing the aforementioned curvilinear search, [41] Theorem 2] asserts that the overall procedure converges to a stationary point of $\phi(U)$, while generating feasible points in the Stiefel manifold at every iteration.

A. Directed variation maximization

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

$$g_i = 2 (A_i^T [u - u_i 1_N]_+ - A_i [u_i 1_N - u]_+)$$

(17)

where $A_i$ denotes the $i$th column of the adjacency matrix $A$, and $A_i$ the $i$th row.

The algorithm starts from a random unit-norm vector and then via (15) 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 $f_{\text{max}} = DV(u_{\text{max}})$ exactly if the number of initializations is chosen large enough; see Section VI. While finding $f_{\text{max}}$ is of interest in its own right, our focus next is on using the obtained $u_{\text{max}}$ to formulate and solve the spectral dispersion minimization problem (9).

B. Spectral dispersion minimization

As the second and final step, here we develop an algorithm to find the orthonormal basis $U$ that minimizes the spectral dispersion (7). To cast the optimization problem (9) in the form of (13) and apply the previously outlined feasible method, we penalize the objective $\delta(U)$ with a measure of the constraint violations to obtain

$$\min_U \phi(U) := \delta(U) + \lambda \frac{1}{2} (\|u_1 - u_{\text{min}}\|^2 + \|u_N - u_{\text{max}}\|^2)$$

subject to $U^T U = I_N$, (18)

where $\lambda > 0$ is chosen large enough to ensure $u_1 = u_{\text{min}}$ and $u_N = u_{\text{max}}$. The resulting iterations are tabulated under Algorithm 2 where the gradient matrix $G := \nabla \phi(U) \in \mathbb{R}^{N \times N}$ has columns given by

$$g_1 = [DV(u_1) - DV(u_2)] g(u_1) + \lambda (u_1 - u_{\text{min}})$$

$$g_i = [DV(u_i) - DV(u_{i-1})] - [DV(u_{i-1})] g(u_i), i \in [2, N - 1]$$

$$g_N = [DV(u_{N-1}) - DV(u_N)] g(u_N) + \lambda (u_N - u_{\text{max}})$$

(19)

where the entries of $g$ are specified in (17). Once more, it is convenient to run the algorithm multiple times and retain the least disperse DGFT basis $U$.

While provably convergent to a stationary point of (18), Algorithm 2 does not offer guarantees on the global optimality of the solution $U$. Still, numerical tests in Section VI corroborate the effectiveness and robustness of the proposed optimization strategy. The computational complexity of Algorithm 2 is $O(N^3)$ 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 bases using the eigenvectors of $G^T$’s Laplacian matrix.

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 $G^n$ and use the eigenvectors of its Laplacian matrix to construct a disperse set of frequencies. The reason for using the eigenvectors of $L$ 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 $G$) by flipping the sign of each Laplacian eigenvector; and iii) they can be used to approximate $f_{\text{max}}$ within a factor of $1/2$ as asserted in Proposition 2. Unlike the undirected case, the directed variation of any eigenvector $u_i$ will in general be different from the variation of $-u_i$, so we can pick the one that we desire without compromising the orthogonality constraint.

Fixing $f_1 = 0$ and $f_N \leq f_{\text{max}}$ from (13), in lieu of (8) we will henceforth construct a disperse set of frequencies by using the eigenvectors of $L$. Let $f_i := \text{DV}(u_i)$ and $\tilde{f}_i := \text{DV}(-u_i)$, where $u_i$ is the $i$th eigenvector of $L$. Define the set of all candidate frequencies as $F := \{f_i, \tilde{f}_i : 1 \leq i < N\}$. The goal is to select $N - 2$ frequencies from $F$ such that together with $\{0, f_{\text{max}}\}$ they form our well-spread Fourier frequencies. To preserve orthonormality, we would select exactly one from each pair $\{f_i, \tilde{f}_i\}$. 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 2-C, we define a spectral dispersion (set) function that measures how much spread the corresponding frequencies are over $[0, f_{\text{max}}]$. For a candidate frequency set $S \subseteq F$, let $s_1 \leq s_2 \leq ... \leq s_m$ be the elements of $S$ in non-decreasing order, where $m = |S|$. Then we define the dispersion of $S$ as

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

(20)

where $s_0 = 0$ and $s_{m+1} = \tilde{f}_{\text{max}}$ [cf. (7)]. It can be verified that $\delta(S)$ is a monotone non-increasing function, which means that for any sets $S_1 \subseteq S_2$, we have $\delta(S_1) \geq \delta(S_2).$ For a fixed value of $m$, one can show that $\delta(S)$ is minimized when the $s_i$’s form an arithmetic sequence, consistent with our design goal in (5). Hence, we seek to minimize $\delta(S)$ 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 $S$ be a finite ground set. A set function $f : 2^S \rightarrow \mathbb{R}$ is submodular if:

$$f(T_1 \cup \{e\}) - f(T_1) \geq f(T_2 \cup \{e\}) - f(T_2),$$

(21)

for all subsets $T_1 \subseteq T_2 \subseteq S$ and any element $e \in S \setminus T_2$. 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 $T_2$, compared to adding the same element to a subset of $T_2$ like $T_1$. The diminishing returns property arises in many science and engineering applications including facility location, sensor placement, and feature selection [21], 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 $\delta : 2^F \rightarrow \mathbb{R}$ defined in (20) is a supermodular function.

Proof: Consider two subsets $S_1, S_2$ such that $S_1 \subseteq S_2 \subseteq F$, and a single element $e \in F \setminus S_2$. Let $s_1^R$ and $s_1^L$ be the largest value smaller than $e$ and the smallest value greater than $e$ in $S_1 \cup \{0, f_{\text{max}}\}$, respectively (i.e., $e$ breaks the gap between $s_1^R$ and $s_1^L$). Similarly, let $s_2^R$ and $s_2^L$ be defined for $S_2$. Since $S_1 \subseteq S_2$, 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

$$\delta(S_1 \cup \{e\}) - \delta(S_1) = (s_1^R - e)^2 + (e - s_1^L)^2 - (s_1^R - s_1^L)^2$$

$$= -2(s_1^R - e)(e - s_1^L)$$

$$\leq -2(s_2^R - e)(e - s_2^L)$$

$$= \delta(S_2 \cup \{e\}) - \delta(S_2).$$

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

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

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

- Hereditary Property: If $T \in \mathcal{I}$, then $T' \in \mathcal{I}$ for all $T' \subseteq T$.

- Augmentation Property: If $T_1, T_2 \in \mathcal{I}$ and $|T_1| < |T_2|$, then there exists $e \in T_2 \setminus T_1$ such that $T_1 \cup \{e\} \in \mathcal{I}$.

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 bases 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 $S$ is a set of (not necessarily independent) vectors, then the linearly independent subsets of $S$ 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 [29]) Let $S$ denote a finite set and let $S_1, ..., S_M$ be a partition of $S$, i.e. a collection of disjoint sets such that $S_1 \cup ... \cup S_M = S$. Let $d_1, ..., d_M$ be a collection of non-negative integers. Define a set $\mathcal{I}$ by $A \in \mathcal{I}$ if $|A \cap S_i| \leq d_i$ for all $i = 1, ..., M$. Then, $\mathcal{M} = (S, \mathcal{I})$ is called the partition matroid.
Algorithm 3 Greedy Spectral Dispersion Minimization

1: Input: Set of possible frequencies \( \mathcal{F} \).
2: Initialize \( S = \emptyset \).
3: repeat
4: \( e \leftarrow \arg\max_{f \in \mathcal{F}} \left\{ \delta(S \cup \{f\}) - \delta(S) \right\} \).
5: \( S \leftarrow S \cup \{e\} \).
6: Delete from \( \mathcal{F} \) the pair \( \{f, f_i\} \) that \( e \) belongs to.
7: until \( \mathcal{F} = \emptyset \)

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 \( B \) in (22) is the set of all bases of \( \mathcal{M} \).

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

B. Greedy algorithm for DGFT bases 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 \( \Omega(\log \log N) \), we create a submodular function \( \delta(S) \) and use the algorithms for submodular maximization to find a set of disperse bases \( \mathcal{U} \). In particular, we define

\[
\delta(S) := \tilde{f}_\text{max}^2 - \delta(S),
\]

which is a non-negative (increasing) submodular function, because \( \delta(\emptyset) = f_\text{max}^2 \) is an upper bound for \( \delta(S) \). There are several results for maximizing submodular functions under matroid constraints for both the non-monotone \( [16] \) 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 \( \delta(S) \). 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 \( S^* \) be the solution of problem (22) and \( S^a \) be the output of the greedy Algorithm 3. Then,

\[
\delta(S^a) \geq \frac{1}{2} \times \delta(S^*).
\]

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 \( \mathcal{L} \). Second, the set \( \mathcal{F} \) is formed by calculating the directed variation for each eigenvector \( u_i \) and its negative \( -u_i \), \( i = 2, \ldots, N - 1 \). 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 bases comprising \( \mathcal{U} \).

VI. Numerical Results

Here we carry out computer simulations on three graphs to assess the performance of the algorithms developed to construct a DGFT with spread frequency components. We also compare these bases 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. 4 and compare them with the GFT put forth in [28] that relies on an augmented Lagrangian optimization method termed PAMAL, as well as with the eigenbasis of the directed Laplacian in [5]. To define said directed Laplacian \( \mathcal{L}_d \), 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 \( \mathbf{\Pi} = \text{diag}(\pi) \) be the diagonal matrix with the stationary distribution \( \pi \) of the random walk on the diagonal. Using these definitions, the directed Laplacian in [5] is given by \( \mathbf{L}_d := \mathbf{\Pi} - (\mathbf{\Pi} \mathbf{P} + \mathbf{P}^T \mathbf{\Pi})/2 \).

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. 5, 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. (18) and (22)]. Finally, Fig. 6 shows the frequency components obtained via Algorithm 3 (similar results are obtained for Algorithm 2 and these are omitted here to avoid repetition). Each subplot depicts one
Fig. 5. Comparison of directed variations (i.e., graph frequencies) for the synthetic digraph in Fig. 4 and different GFT methods: eigenvectors of the combinatorial Laplacian matrix $L_k$ introduced in [5]; augmented Lagrangian method (PAMAL) in [28]; 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.

Fig. 6. DGFT bases obtained using the greedy Algorithm 3 for the synthetic digraph in Fig. 4 along with their respective directed variation values (frequencies). Notice how the frequency components associated with lower frequencies are roughly constant over node clusters.

basis vector (column) of the resulting DGFT matrix $U_k$ along with its corresponding directed variation values defined in (2). It is apparent that the first bases exhibit less variability than the higher frequency components. Moreover, the greedy approach constructs the basis starting from Laplacian eigenvectors of $G^n$, and one can see that lower frequency components have the additional desired property of being roughly constant over network clusters; see also the design in [28].

We also use the Monte-Carlo method to study the convergence properties of our algorithms. In Fig. 7 (top) we show the evolution of iterates for the feasible method in (41), when used to find the maximum directed variation (i.e., $f_{\text{max}}$) for the same 15-node graph in Fig. 4. We do so for 100 different (random) initializations and report the median as well as the first and third quartiles versus the number of iterations. We observe that all the realizations converge [41 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 (6). It is worth mentioning that after about 10 iterations, the exact value of $f_{\text{max}}$ is achieved by a quarter of the realizations (and this improves to half of the realizations...
with about 30 iterations). Similarly, Fig. 7 (bottom) shows the median, first, and third quartiles of the dispersion function iterates $\delta(U_k)$, when minimized using Algorithm 2. Again, 100 different Monte-Carlo simulations are considered and we observe that all of them 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 [11, 25] 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 DGFT bases are well distributed in the graph spectral domain, Fig. 8 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. 8 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 [12]. 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-called 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 $x \in \mathbb{R}^{48}$ shown in Fig. 9 [7]. 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.

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. 10. The first four bases are smooth as expected. The last bases 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 Ua, 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. To corroborate this, we aim to recover the temperature signal from noisy measurements $y = x + n$, where the additive noise $n$ is a zero-mean, Gaussian random vector with covariance matrix $10I_N$. To that end, we use a low-pass graph filter with frequency response $h = [h_1, \ldots, h_N]^T$, where $h_i = 1 \{i \leq w\}$ 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

$$\hat{x} = \text{Udiag}(\hat{h})\hat{y} = \text{Udiag}(\hat{h})U^Ty.$$  

Note that the filter $H := \text{Udiag}(\hat{h})U^T$ will in general not be expressible as a polynomial of the graph’s adjacency matrix $A$ (or some other graph shift operator [24]), 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 $H$ have been studied in [30].

Fig. 11-(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. Fig. 11-(b) shows a realization of the noisy graph signal $y$ superimposed with the denoised temperature profile $\hat{x}$ obtained using (24) with $w = 3$, and the original signal $x$. Filter design and the choice of $w$ is beyond the scope of this paper, but the average recovery error $e_f = ||\hat{x} - x||/||x||$, over 1000 Monte-Carlo simulations of independent noise, is approximately 12% and Fig. 11-(b) shows $\hat{x}$ closely approximates $x$. Moreover, we compute the relative recovery error with and without low-pass filtering as $e_f$ and $e = ||n||/||x||$, respectively. Fig. 11-(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 (and justifies the choice $w = 3$). To assess the importance of the network model, we reverse the orientation of all edges (N-S) and repeat the whole denoising experiment. As expected, Fig. 11-(c) shows that the performance degrades since the S-N digraph in Fig. 9 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 5.

For additional comparisons we also considered the following state-of-the-art approaches: i) PAMAL method in [28]; ii) filter design using adjacency matrix [27]; and iii) the directed Laplacian in [36]. The PAMAL approach fails to converge to orthonormal bases within a reasonable time. We also apply the filter design in [27] Section V] to both adjacency matrix and directed Laplacian defined as $D_{in} = A$, where $D_{in}$ is the diagonal in-degree matrix. In the constructed digraph, the adjacency matrix has only three different eigenvalues that limits the degree of freedom in the filter design. Furthermore, the directed Laplacian of [36] fails to provide a recovered signal with smaller error than $y$. Not only is the Jordan decomposition required in this method numerically unstable, but also the resulting polynomial filter gives rise to large reconstruction errors (even exceeding the original error without filtering).

**VII. CONCLUSIONS**

We considered the problem of finding an orthonormal set of graph Fourier bases 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 bases. 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 ($f_{\text{max}}$) 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

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

**Proposition 3** The unit-norm basis vector $\mathbf{u}_{\text{max}}$ defined in (6) is orthogonal to the constant vector $\mathbf{u}_{\text{min}} := \frac{1}{\sqrt{N}} \mathbf{1}_N$.

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

Without loss of generality assume that $\mathbf{u}_{\text{max}}^T \mathbf{1}_N = \epsilon > 0$, and define $\bar{\mathbf{u}} := \mathbf{u}_{\text{max}} - \frac{\epsilon}{N} \mathbf{1}_N$. First, note that $\text{DV}(\mathbf{u}_{\text{max}}) =$
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 \((u_{N-1}, u_N)\) 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. \(\blacksquare\)

**C. Proof of Proposition 1-2**

There should be at least one edge \((i, j)\) in the cycle for which \(u_i - u_j\), otherwise we obtain a closed loop of strict inequalities among consecutive \(u_i\) 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. \(\blacksquare\)

**D. Proof of Proposition 1-3**

Let \(V^+\) and \(V^-\) be the two node partitions of the unidirectional bipartite graph, where the edges are constrained to go from \(V^+\) to \(V^-\). First, we show that in the optimal solution maximizing the directed variation we must have \(u_i \geq 0\) for all \(i \in V^+\), and \(u_i \leq 0\) for all \(i \in V^-\). Otherwise, assume that there exists some node \(j \in V^-\) with \(u_j > 0\). Then we can improve the directed variation by setting \(u_j = 0\), because \(j\) has only incoming edges and decreasing \(u_j\) 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 \(V^+\) has negative value.

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

\[
\max_{\mathbf{u}} \quad TV(\mathbf{u}) = \mathbf{u}^T \mathbf{L} \mathbf{u}
\]

subject to \(\mathbf{u}^T \mathbf{u} = 1\)

(25)

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

**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 \(u_1, u_2, \ldots, u_N > 0\) be the dipath of length \(N - 1\), with directed edges going from \(u_i\) to \(u_{i+1}\), \(i = 1, \ldots, N - 1\) (for convenience here we use the signal value assigned to a node \(u_i\) and the node index \(i\), interchangeably). For the base case of \(N = 2\), we have to maximize \(A_{12}(u_1 - u_2)^2\) subject to \(u_2 \geq u_1\) and \(u_1^2 + u_2^2 = 1\). The solution to this optimization problem is \(u_1 = \sqrt{2}/2\) and \(u_2 = -\sqrt{2}/2\) which evaluates to \(A_{12}(u_1 - u_2)^2 = 2A_{12}\). 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 the \(N\) edges as well. If \(u_N \leq u_{N-1}\) in the optimal solution for \(N\) edges, then for the last edge \([u_N - u_{N+1}]_+ = 0\) 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 \(u_N = \sqrt{2}/2\) and \(u_{N+1} = -\sqrt{2}/2\) would improve the optimal solution and violates the assumption of \(u_N \leq u_{N+1}\). Therefore we assume \(u_N > u_{N+1}\). In this case, we claim that \(u_N\) should also be greater than or equal to \(u_{N-1}\). If not, we have \(u_{N-1} > u_N > u_{N+1}\), which cannot be an optimal solution. To see this, we can swap the value of \(u_N\) with either \(u_{N-1}\) or \(u_{N+1}\) and improve the directed variation because either \(A_{(N-1)N}(u_{N-1} - u_{N+1})^2\) or \(A_{N(N+1)}(u_N - u_{N+1})^2\) is greater than \(A_{N(N+1)}(u_{N-1} - u_{N+1})^2\). Finally, if \(u_N > u_{N+1}\) and \(u_N \geq u_{N-1}\), then the edge \((u_{N-1}, u_N)\) does not contribute to the directed variation and the path is divided into two sections. Since both \(\|\mathbf{u}\|^2\) and \(\text{DV}(\mathbf{u})\) 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

\[
\text{DV}(\tilde{\mathbf{u}}), \text{since adding (subtracting) a constant to (from) all coordinates will not change the directed variation. Second,}
\]

\[
\left\|\tilde{\mathbf{u}}\right\|^2 = u_{\max}^T u_{\max} - \frac{2\epsilon}{N} u_{\max}^T 1_N + \left(\frac{\epsilon}{N}\right)^2 1_N^T 1_N = 1 - \frac{\epsilon^2}{N}.
\]

Therefore, we have a new vector \(\tilde{\mathbf{u}}\) with the same directed variation but smaller norm. Now we can scale this vector as \(\alpha \tilde{\mathbf{u}}\) (with \(\alpha > 1\)) to obtain a normalized vector with \(\text{DV}(\alpha \tilde{\mathbf{u}}) = \alpha^2 \text{DV}(\tilde{\mathbf{u}})\), which improves upon \(u_{\max}\). \(\blacksquare\)
Lemma 3 For an undirected bipartite graph $\mathcal{G} = (\mathcal{V}_1, \mathcal{V}_2, A)$ with Laplacian matrix $L$, let $u \in \mathbb{R}^N$ be the dominant eigenvector of $L$ (corresponding to $\lambda_{\text{max}}$). Then $u$ 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 $u$ maximizes $TV(u) = u^T L u = \sum_{i,j=1,j>1}^N A_{ij} (u_i - u_j)^2$.

By virtue of Lemma 3 either $u$ or $-u$ is feasible in $\mathbb{R}^N$ and attains the optimum objective value $\lambda_{\text{max}}$.

REFERENCES

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, Optimization Algorithms on Matrix Manifolds. Princeton University Press, 2009.

[2] G. Calinescu, C. Chekuri, M. Pál, and J. Vondrak, “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, J. V. Arma, A. Sandryhaila, and J. Kovačević, “Discrete signal processing on graphs: Sampling theory,” IEEE Trans. Signal Process., vol. 63, no. 24, pp. 6510–6523, 2015.

[5] F. Chung, “Laplacians and the Cheeger inequality for directed graphs,” Annals of 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. on Signal and Information Process., Washington, DC, Dec. 7-9, 2016, pp. 1275–1279.

[7] ——, “Spectral projector-based graph Fourier transforms,” IEEE J. Sel. Topics Signal Process., vol. 11, no. 6, pp. 785–795, 2017.

[8] M. L. Fisher, G. L. Nemhauser, and L. A. Wolsey, “An analysis of approximations for maximizing submodular set functions—I,” SIAM J. Comput., vol. 10, no. 7, pp. 133–152, 1981.

[9] ——, “Discrete Fourier transform for directed graphs,” in Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Process., Calgary, Alberta, Canada, Apr. 15-20, 2018.

[10] D. I. Shuman, K. Weinberger, and mom, “Network topology inference from non-stationary graph signals,” in Proc. Int. Conf. Acoustics, Speech, Signal Process., New Orleans, USA, Mar. 5-9, 2017, pp. 5870–5874.

[11] D. I. Shuman, K. Weinberger, and mom, “Network topology inference from non-stationary graph signals,” in Proc. Int. Conf. Acoustics, Speech, Signal Process., New Orleans, USA, Mar. 5-9, 2017, pp. 5870–5874.

[12] D. I. Shuman, K. Weinberger, and mom, “Network topology inference from non-stationary graph signals,” in Proc. Int. Conf. Acoustics, Speech, Signal Process., New Orleans, USA, Mar. 5-9, 2017, pp. 5870–5874.

[13] D. I. Shuman, K. Weinberger, and mom, “Network topology inference from non-stationary graph signals,” in Proc. Int. Conf. Acoustics, Speech, Signal Process., New Orleans, USA, Mar. 5-9, 2017, pp. 5870–5874.

[14] D. I. Shuman, K. Weinberger, and mom, “Network topology inference from non-stationary graph signals,” in Proc. Int. Conf. Acoustics, Speech, Signal Process., New Orleans, USA, Mar. 5-9, 2017, pp. 5870–5874.

[15] D. I. Shuman, K. Weinberger, and mom, “Network topology inference from non-stationary graph signals,” in Proc. Int. Conf. Acoustics, Speech, Signal Process., New Orleans, USA, Mar. 5-9, 2017, pp. 5870–5874.

[16] D. I. Shuman, K. Weinberger, and mom, “Network topology inference from non-stationary graph signals,” in Proc. Int. Conf. Acoustics, Speech, Signal Process., New Orleans, USA, Mar. 5-9, 2017, pp. 5870–5874.

[17] D. I. Shuman, K. Weinberger, and mom, “Network topology inference from non-stationary graph signals,” in Proc. Int. Conf. Acoustics, Speech, Signal Process., New Orleans, USA, Mar. 5-9, 2017, pp. 5870–5874.

[18] D. I. Shuman, K. Weinberger, and mom, “Network topology inference from non-stationary graph signals,” in Proc. Int. Conf. Acoustics, Speech, Signal Process., New Orleans, USA, Mar. 5-9, 2017, pp. 5870–5874.