Digraph Fourier Transform via Spectral Dispersion Minimization

Gonzalo Mateos

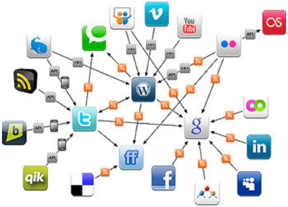
Dept. of Electrical and Computer Engineering
University of Rochester
gmateosb@ece.rochester.edu
http://www.ece.rochester.edu/~gmateosb/

Collaborators: R. Shafipour, A. Khodabakhsh, and E. Nikolova
Acknowledgment: NSF Awards CCF-1750428 and ECCS-1809356

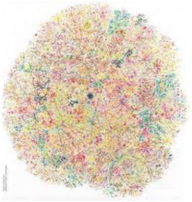
Holetown, Barbados, February 11, 2019

Network Science analytics

Online social media



Internet



Clean energy and grid analytics

- Network as graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$: encode pairwise relationships
- Desiderata: Process, analyze and learn from network data [Kolaczyk'09]
  - $\Rightarrow$ Use $\mathcal{G}$ to study graph signals, data associated with nodes in $\mathcal{V}$
- Ex: Opinion profile, buffer congestion levels, neural activity, epidemic

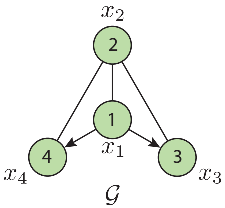
Graph signal processing and Fourier transform

- Directed graph (digraph) $\mathcal{G}$ with adjacency matrix $\mathbf{A}$

  $\Rightarrow A_{ij} =$ Edge weight from node i to node j

- Define a signal $\mathbf{x} \in \mathbb{R}^N$ on $\mathcal{V}$ $(N := |\mathcal{V}|)$

  $\Rightarrow x_i =$ Signal value at node i



- Associated with $\mathcal{G}$ is the underlying undirected graph $\mathcal{G}^u$

  $\Rightarrow$ Laplacian marix $\mathbf{L} = \mathbf{D} - \mathbf{A}^u$, eigenvectors $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_N]$

- Graph Signal Processing (GSP): exploit structure in $\mathbf{A}$ or $\mathbf{L}$ to process $\mathbf{x}$

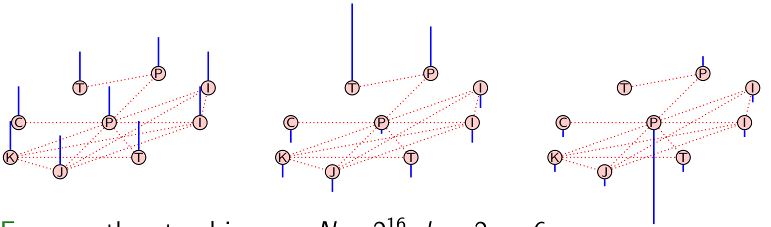
- Graph Fourier Transform (GFT): $\tilde{\mathbf{x}} = \mathbf{V}^T\mathbf{x}$ for undirected graphs

  $\Rightarrow$ Decompose $\mathbf{x}$ into different modes of variation

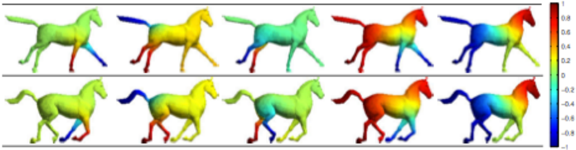
  $\Rightarrow$ Inverse (i)GFT $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$, eigenvectors as frequency atoms

Frequency modes of the Laplacian

- Let us plot some of the eigenvectors $\mathbf{v}_k$ of $\mathbf{L}$ (also graph signals)
- Ex: gene network, $N = 10$, $k = 1$, $k = 2$, $k = 9$

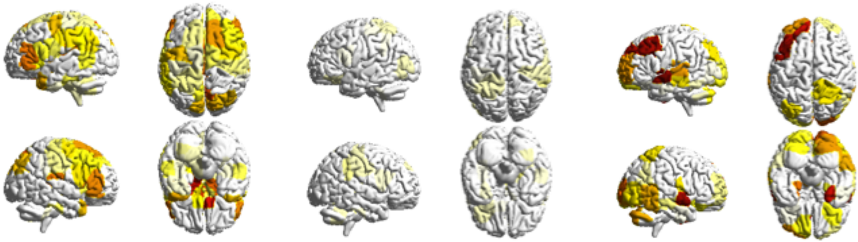


- Ex: smooth natural images, $N = 2^{16}$, $k = 2, \ldots, 6$



Frequency analysis of brain signals

- GFT of brain activity signals during a visual-motor learning task
  - $\Rightarrow$ Decomposed into low, medium and high frequency components



- Brain: Complex system where regularity coexists with disorder [Sporns'11]
  - $\Rightarrow$ Signal energy mostly in the low and high frequencies
  - $\Rightarrow$ In brain regions alike to the visual and sensorimotor cortices

Our work in context

- Spectral analysis and filter design [Tremblay et al'17], [Isufi et al'16]
  $\Rightarrow$ GFT as a promising tool in neuroscience [Huang et al'16]
- Noteworthy GFT approaches
  - Jordan decomposition of $\mathbf{A}$ [Sandryhaila-Moura'14], [Deri-Moura'17]
  - Lovaśz extension of the graph cut size [Sardellitti et al'17]
  - Generalized variation operators and inner products [Girault et al'18]
- Dictionary learning (DL) for GSP
  - Joint topology- and online data-driven prediction [Forero et al'14]
  - Parametric dictionaries for graph signals [Thanou et al'14]
  - Dual graph-regularized DL [Yankelevsky-Elad'17]
- **Our contribution:** digraph (D)GFT (dictionary) design
  - Orthonormal basis signals (atoms) offer notions of frequency
  - Frequencies are distributed as even as possible in $[0, f_{\max}]$
  - Sparsely represents bandlimited graph signals

Signal variation on digraphs

- Total variation of signal $\mathbf{x}$ with respect to $\mathbf{L}$

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

$\Rightarrow$ Smoothness measure on the graph $\mathcal{G}^u$

- For Laplacian eigenvectors $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_N]$ $\Rightarrow$ $\text{TV}(\mathbf{v}_k) = \lambda_k$

$\Rightarrow$ Can view $0 = \lambda_1 < \cdots \leq \lambda_N$ as frequencies

- **Def:** Directed variation for signals over digraphs $([x]_+ = \max(0, x))$

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

$\Rightarrow$ Captures signal variation (flow) along directed edges

$\Rightarrow$ Consistent, since $\text{DV}(\mathbf{x}) \equiv \text{TV}(\mathbf{x})$ for undirected graphs

DGFT with spread frequeny components

- **Goal:** find N orthonormal basis vectors capturing different modes of DV
- Collect the desired basis signals in $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_N] \in \mathbb{R}^{N \times N}$

$$\text{DGFT:} \quad \tilde{\mathbf{x}} = \mathbf{U}^T \mathbf{x}$$

$\Rightarrow$ Atom $\mathbf{u}_k$ represents the kth frequency mode with $f_k := \text{DV}(\mathbf{u}_k)$

- Similar to the DFT, seek N *equidistributed* graph frequencies

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

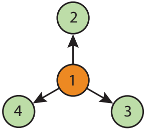
$\Rightarrow$ $f_{\max}$ is the maximum DV of a unit-norm graph signal on $\mathcal{G}$

- **Q**: Why spread frequencies?
  - Parsimonious representations of slowly-varying signals
  - Interpretability $\Rightarrow$ better capture low, medium, and high *frequencies*
  - Aid filter design in the graph spectral domain

Motivation for spread frequencies

Ex: Directed variation minimization [Sardellitti et al'17]

$$\min_{\mathbf{U}} \quad \sum_{i,j=1}^{N} A_{ij}[\mathbf{u}_i - \mathbf{u}_j]_+$$
$$\text{s.t.} \quad \mathbf{U}^T\mathbf{U} = \mathbf{I}$$



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

- $\mathbf{U}^*$ is the optimum basis where $a = \frac{1+\sqrt{5}}{4}$, $b = \frac{1-\sqrt{5}}{4}$, and $c = -0.5$
- All columns of $\mathbf{U}^*$ satisfy $\text{DV}(\mathbf{u}_k^*) = 0, \ k = 1, \ldots, 4$
  - $\Rightarrow$ Expansion $\mathbf{x} = \mathbf{U}^*\tilde{\mathbf{x}}$ fails to capture *different* modes of variation
- **Q:** Can we always find *equidistributed* frequencies in $[0, f_{\max}]$?

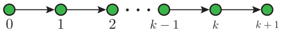
- Finding $f_{\max}$ is in general challenging

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

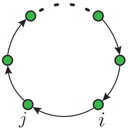
- Let $\mathbf{v}_N$ be the dominant eigenvector of $\mathbf{L}$

  $\Rightarrow$ Can 1/2-approximate $f_{\max}$ with $\tilde{\mathbf{u}}_{\max} = \underset{\mathbf{v}\in\{\mathbf{v}_N, -\mathbf{v}_N\}}{\operatorname{argmax}}\ \mathrm{DV}(\mathbf{v})$

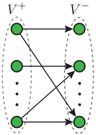
- $f_{\max}$ can be obtained analytically for particular graph families



$f_{\max} = 2\ \max_{i,j} A_{ij}$



$f_{\max} = 2\ \max_{i,j} A_{ij}$



$f_{\max} = \lambda_{\max}$

Challenges: Equidistributed frequencies

- *Equidistributed* $f_k = \frac{k-1}{N-1} f_{\max}$ may not be feasible. Ex: Undirected $\mathcal{G}^u$

$$f^u_{max} = \lambda_{\max} \quad \text{and} \quad \sum_{k=1}^{N} f_k = \sum_{k=1}^{N} \text{TV}(\mathbf{u}_k) = \text{trace}(\mathbf{L})$$

- **Idea:** Set $\mathbf{u}_1 = \mathbf{u}_{\min} := \frac{1}{\sqrt{N}} \mathbf{1}_N$ and $\mathbf{u}_N = \mathbf{u}_{\max}$ and minimize

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

$\Rightarrow$ $\delta(\mathbf{U})$ is the *spectral dispersion function*

$\Rightarrow$ Minimized when *free* DV values form an arithmetic sequence

Spectral dispersion minimization

- Non-convex optimization problem for finding spread frequencies

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

$$\text{subject to} \quad \mathbf{U}^T\mathbf{U} = \mathbf{I}$$

$$\mathbf{u}_1 = \mathbf{u}_{\min}$$

$$\mathbf{u}_N = \mathbf{u}_{\max}$$

  - Orthogonality-constrained minimization of smooth $\delta(\mathbf{U})$
  - Feasible since $\mathbf{u}_{\max} \perp \mathbf{u}_{\min}$

- Feasible method in the Stiefel manifold to design the DGFT:
  - (i) Obtain $f_{\max}$ (and $\mathbf{u}_{\max}$) by minimizing $-\mathrm{DV}(\mathbf{u})$ over $\{\mathbf{u} \mid \mathbf{u}^T\mathbf{u} = 1\}$
  - (ii) Find the orthonormal basis $\mathbf{U}$ with minimum spectral dispersion

Feasible method in the Stiefel manifold

- Rewrite the problem of finding orthonormal basis as

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

- Let $\mathbf{U}_k$ be a feasible point at iteration k and the gradient $\mathbf{G}_k = \nabla\phi(\mathbf{U}_k)$
  $\Rightarrow$ Skew-symmetric matrix $\mathbf{B}_k := \mathbf{G}_k{\mathbf{U}_k}^T - \mathbf{U}_k{\mathbf{G}_k}^T$

- Follow the update rule $\mathbf{U}_{k+1}(\tau) = \left(\mathbf{I} + \frac{\tau}{2}\mathbf{B}_k\right)^{-1}\left(\mathbf{I} - \frac{\tau}{2}\mathbf{B}_k\right)\mathbf{U}_k$
  - Cayley transform preserves orthogonality (i.e., ${\mathbf{U}_{k+1}}^T\mathbf{U}_{k+1} = \mathbf{I}$)
  - Is a descent path for a proper step size τ

Theorem (Wen-Yin'13) The procedure converges to a stationary point of smooth $\phi(\mathbf{U})$, while generating feasible points at every iteration

Algorithm

Input: Adjacency matrix $\mathbf{A}$, parameters $\lambda > 0$ and $\epsilon > 0$
Find $\mathbf{u}_{\text{max}}$ by a similar feasible method and set $\mathbf{u}_{\text{min}} = \frac{1}{\sqrt{N}}\mathbf{1}_N$
Initialize $k = 0$ and orthonormal $\mathbf{U}_0 \in \mathbb{R}^{N \times N}$ at random
repeat
    Compute gradient $\mathbf{G}_k = \nabla\phi(\mathbf{U}_k) \in \mathbb{R}^{N \times N}$
    Form $\mathbf{B}_k = \mathbf{G}_k{\mathbf{U}_k}^T - \mathbf{U}_k{\mathbf{G}_k}^T$
    Select τ_k satisfying Armijo-Wolfe conditions
    Update $\mathbf{U}_{k+1}(\tau_k) = (\mathbf{I} + \frac{\tau_k}{2}\mathbf{B}_k)^{-1}(\mathbf{I} - \frac{\tau_k}{2}\mathbf{B}_k)\mathbf{U}_k$
    $k \leftarrow k + 1$
until $\|\mathbf{U}_k - \mathbf{U}_{k-1}\|_F \leq \epsilon$
Return $\hat{\mathbf{U}} = \mathbf{U}_k$

- Overall run-time is $\mathcal{O}(N^3)$ per iteration

Additional details in arXiv:1804.03000 [eess.SP]

- **Q:** Can we make the DGFT design data-adaptive?

Spectral dispersion and sparsity minimization

- *Sparsify* a set of bandlimited signals $\mathbf{X} \in \mathbb{R}^{N \times P} \rightarrow$ Minimize $||\mathbf{U}^T\mathbf{X}||_1$
- **Problem:** given $\mathcal{G}$ and $\mathbf{X}$, find sparsifying DGFT with spread frequencies

$$\min_{\mathbf{U}} \quad \Psi(\mathbf{U}) := \sum_{i=1}^{N-1} [\text{DV}(\mathbf{u}_{i+1}) - \text{DV}(\mathbf{u}_i)]^2 + \mu||\mathbf{U}^T\mathbf{X}||_1$$

$$\text{subject to} \quad \mathbf{U}^T\mathbf{U} = \mathbf{I}$$

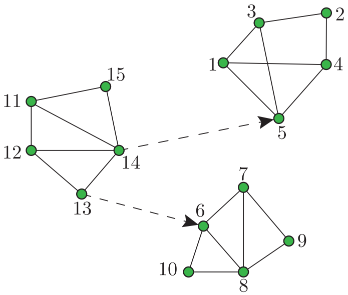
$$\mathbf{u}_1 = \mathbf{u}_{\text{min}}$$

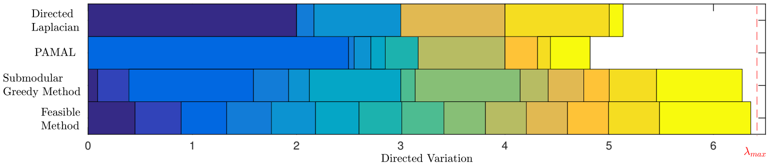
$$\mathbf{u}_N = \mathbf{u}_{\text{max}}$$

  - Non-convex, orthogonality-constrained minimization
  - Non-differentiable $\Psi(\mathbf{U})$
- Variable-splitting solver:
  (i) Obtain f_{max} (and $\mathbf{u}_{\text{max}}$) by minimizing $-\text{DV}(\mathbf{u})$ over $\{\mathbf{u} \mid \mathbf{u}^T\mathbf{u} = 1\}$
  (ii) Replace $\mathbf{U}^T\mathbf{X}$ with an auxiliary variable $\mathbf{Y} \in \mathbb{R}^{N \times P}$, enforce $\mathbf{Y} = \mathbf{U}^T\mathbf{X}$
  (iii) Alternate between feasible method and soft thresholding

Numerical test: Synthetic graph

- Compute **U** and directed variations using
  - Directed Laplacian eigenvectors [Chung'05]
  - PAMAL method [Sardellitti et al'17]
  - Greedy heuristic
  - Spectral dispersion minimization





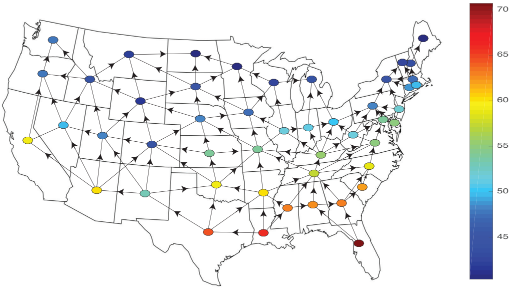
- Rescale DV values to $[0, 1]$ and calculate *spectral dispersion* $\delta(\mathbf{U})$

  $\Rightarrow$ 0.256, 0.301, 0.118, and **0.076** respectively

  $\Rightarrow$ Confirms the proposed methods yield a better frequency spread

Numerical test: US average temperatures

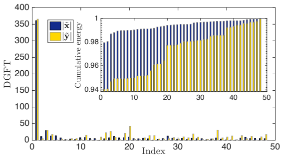
- Consider the graph of the $N = 48$ contiguous United States
  - $\Rightarrow$ Connect two states if they share a border
  - $\Rightarrow$ Set arc directions from lower to higher latitudes

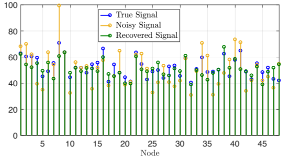


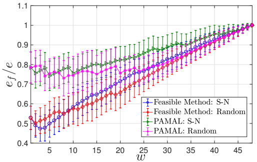
- Graph signal $\mathbf{x}$ $\rightarrow$ Average annual *temperature* of each state

Numerical test: Denoising US temperatures

- Noisy signal $\mathbf{y} = \mathbf{x} + \mathbf{n}$, with $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, 10 \times \mathbf{I}_N)$
- Define low-pass filter $\tilde{\mathbf{H}} = \text{diag}(\tilde{\mathbf{h}})$, where $\tilde{h}_i = \mathbb{I}\{i \leq w\}$ (for $w = 3$)
- Recover signal via filtering $\hat{\mathbf{x}} = \mathbf{U}\tilde{\mathbf{H}}\tilde{\mathbf{y}} = \mathbf{U}\tilde{\mathbf{H}}\mathbf{U}^T\mathbf{y}$
  - $\Rightarrow$ Compute recovery error $e_f = \frac{\|\hat{\mathbf{x}} - \mathbf{x}\|}{\|\mathbf{x}\|} \approx 12\%$
  - $\Rightarrow$ Pick random edge orientations and repeat the experiment







- DGFT basis offers a parsimonious (i.e., bandlimited) signal representation
  - $\Rightarrow$ Adequate network model improves the denoising performance

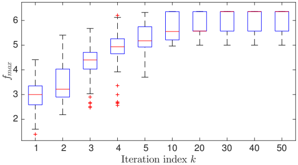
Numerical test: Convergence behavior

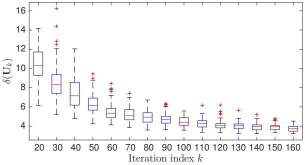
- Average monthly temperature over ~ 60 years for each state

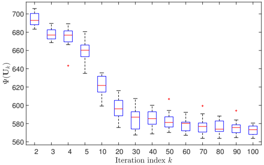
  $\Rightarrow$ Training signals $\mathbf{X} \in \mathbb{R}^{48 \times 12}$

- Monte-Carlo simulations to study the convergence behavior

  $\Rightarrow$ Plot $f_{\max}$, $\delta(\mathbf{U})$, and $\boldsymbol{\Psi}(\mathbf{U}) = \delta(\mathbf{U}) + \mu ||\mathbf{U}^T \mathbf{X}||_1$

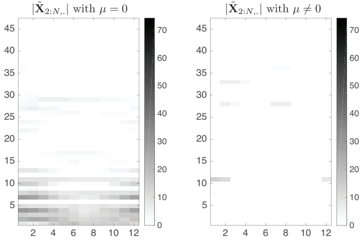




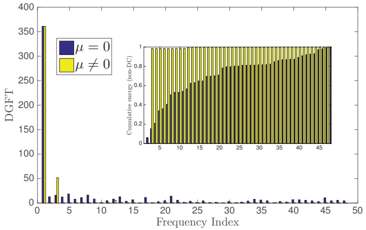


- Convergence is apparent, with limited variability on the solution

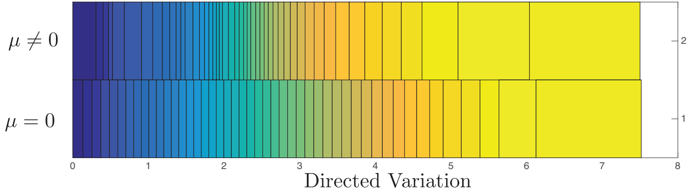
- Heat maps of the trained $\tilde{\mathbf{X}}$



- Spectral representation of test signal



- Distribution of all the frequencies



- **Tradeoff:** spectral dispersion for a sparser representation
  - Still attain well dispersed frequencies

Closing remarks

- Measure of directed variation to capture the notion of frequency on $\mathcal{G}$
- Find an orthonormal set of Fourier basis vectors for signals on digraphs
  - Span a maximal frequency range $[0, f_{\max}]$
  - Frequency modes are as evenly distributed as possible
- Two-step DGFT basis design via a feasible method over Stiefel manifold

  i) Find the maximum directed variation $f_{\max}$ over the unit sphere

  ii) Minimize a smooth spectral dispersion criterion over $[0, f_{\max}]$

  $\Rightarrow$ Provable convergence guarantees to a stationary point
- Ongoing work and future directions
  - Complexity of finding the maximum frequency $f_{\max}$ on a digraph?

    $\Rightarrow$ If NP-hard, what is the best approximation ratio
  - Optimality gap between the local and global optimal dispersions?
  - Scalable and fast(er) digraph Fourier transform?