A Scalable M-Channel Critically Sampled Filter Bank for Graph Signals

Yan Jin, Shuni Li, and David I Shuman

***Abstract*—We investigate a scalable M-channel critically sampled filter bank for graph signals, where each of the M filters is supported on a different subband of the graph Laplacian spectrum. For analysis, the graph signal is filtered on each subband and downsampled on a corresponding set of vertices. However, the classical synthesis filters are replaced with interpolation operators. For small graphs, we use a full eigendecomposition of the graph Laplacian to partition the graph vertices such that the m^{th} set comprises a uniqueness set for signals supported on the m^{th} subband. The resulting transform is critically sampled, the dictionary atoms are orthogonal to those supported on different bands, and graph signals are perfectly reconstructable from their analysis coefficients. We also investigate a fast version of the proposed transform that scales efficiently for large, sparse graphs. Issues that arise in this context include designing the filter bank to be more amenable to polynomial approximation, estimating the number of samples required for each band, performing non-uniform random sampling for the filtered signals on each band, and efficient reconstruction methods. We empirically explore the joint vertex-frequency localization of the dictionary atoms, the sparsity of the analysis coefficients for different classes of signals, the ability of the proposed transform to compress piecewise-smooth graph signals, and the reconstruction error resulting from the numerical approximations.**

***Index Terms*—Graph signal processing, filter bank, non-uniform random sampling, interpolation, wavelet, compression**

I. Introduction

In graph signal processing [2], transforms and filter banks can help exploit structure in the data, in order, for example, to compress a graph signal, remove noise, or fill in missing information. Broad classes of recently proposed transforms include graph Fourier transforms, vertex domain designs such as [3], [4], top-down approaches such as [5], [6], [7], diffusion-based designs such as [8], [9], spectral domain designs such as [10], [11], windowed graph Fourier transforms [12], and generalized filter banks, the last of which we focus on in this paper. For further introduction to dictionary designs for graph signals, see [2].

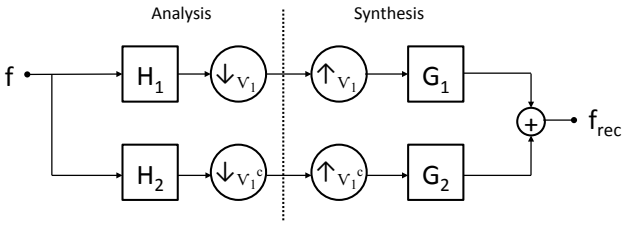

Fig. 1. Extension of the classical two channel critically sampled filter bank to the graph setting. Here, $\mathbf{H}_1$ is a lowpass graph spectral filter, and $\mathbf{H}_2$ is a highpass graph spectral filter.

The extension of the classical two channel critically sampled filter bank to the graph setting was first proposed in [13]. Fig. 1 shows the analysis and synthesis banks, where $\mathbf{H}_i$ and $\mathbf{G}_i$ are graph spectral filters [2], and the lowpass and highpass bands are downsampled on complementary sets of vertices. For a general weighted, undirected graph, it is not straightforward how to design the downsampling and the four graph spectral filters to ensure perfect reconstruction. One approach is to separate the graph into a union of subgraphs, each of which has some regular structure. For example, [14], [15] show that the normalized graph Laplacian eigenvectors of *bipartite* graphs have a spectral folding property that make it possible to design analysis and synthesis filters to guarantee perfect reconstruction. They take advantage of this property by decomposing the graph into bipartite graphs and constructing a multichannel, separable filter bank, while [16] adds vertices and edges to the original graph to form an approximating bipartite graph. References [17], [18] generalize this spectral folding property to M-block cyclic graphs, and leverage it to construct M-channel graph filter banks. Another class of regular structured graphs is *shift invariant* graphs [19, Chapter 5.1]. These graphs have a circulant graph Laplacian and their eigenvectors are the columns of the discrete Fourier

Yan Jin is with the Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (email: yjin1@mit.edu). Shuni Li and David I Shuman are with the Department of Mathematics, Statistics, and Computer Science, Macalester College, St. Paul, MN 55105, USA (email: {sli,dshuman1}@macalester.edu).

This research has been funded in part by a grant to Macalester College from the Howard Hughes Medical Institute through the Precollege and Undergraduate Science Education Program.

The authors would like to thank Federico Poloni for providing a proof to Proposition 2, and Andrew Bernoff for helpful discussions about the matrix partitioning problem discussed in Section III.

MATLAB code for all numerical experiments in this paper is available at http://www.macalester.edu/~dshuman1/publications.html. It leverages the open access GSPBox [1], into which it will soon be integrated.

transform matrix. Any graph can be written as the sum of circulant graphs, and [20], [21], [22] take advantage of this fact in designing critically sampled graph filter banks with perfect reconstruction. Another approach is to use architectures other than the critically sampled filter bank, such as lifting transforms [23], [24] or pyramid transforms [25].

Our approach in this paper is to replace the synthesis filters with interpolation operators on each subband of the graph spectrum. While this idea was initially suggested independently in [26], we investigate it in more detail here. Our construction leverages the recent flurry of work in sampling and reconstruction of graph signals [26]-[35]. The key property we use is that any signal whose graph Fourier transform has exactly k non-zero coefficients can be perfectly recovered from samples of that signal on k appropriately selected vertices (see, e.g., [26, Theorem 1] [35, Proposition 1]).

Add to intro here: scalable issues, outline, and contribution relative to conference paper.

II. M-Channel Critically Sampled Filter Bank

We consider graph signals $f \in \mathbb{R}^N$ residing on a weighted, undirected graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$, where $\mathcal{V}$ is the set of N vertices, $\mathcal{E}$ is the set of edges, and $\mathbf{W}$ is the weighted adjacency matrix. Throughout, we take $\boldsymbol{\mathcal{L}}$ to be the combinatorial graph Laplacian $\mathbf{D} - \mathbf{W}$, where $\mathbf{D}$ is the diagonal matrix of vertex degrees. However, our theory and proposed transform also apply to the normalized graph Laplacian $\mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}}$, or any other Hermitian operator. We can diagonalize the graph Laplacian as $\boldsymbol{\mathcal{L}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^*$, where $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues $\lambda_0, \lambda_1, \ldots, \lambda_{N-1}$ of $\boldsymbol{\mathcal{L}}$, and the columns $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{N-1}$ of $\mathbf{U}$ are the associated eigenvectors of $\boldsymbol{\mathcal{L}}$. The graph Fourier transform of a signal is $\hat{\mathbf{f}} = \mathbf{U}^*\mathbf{f}$, and $h(\boldsymbol{\mathcal{L}})\mathbf{f} = \mathbf{U}h(\boldsymbol{\Lambda})\mathbf{U}^*\mathbf{f}$ applies the filter $h : [0, \lambda_{\max}] \to \mathbb{R}$ to the graph signal $\mathbf{f}$. We use the notation $\mathbf{U}_{\mathcal{R}}$ to denote the submatrix formed by taking the columns of $\mathbf{U}$ associated with the Laplacian eigenvalues indexed by $\mathcal{R} \subseteq \{0, 1, \ldots, N-1\}$. Similarly, we use the notation $\mathbf{U}_{\mathcal{S},\mathcal{R}}$ to denote the submatrix formed by taking the rows of $\mathbf{U}_{\mathcal{R}}$ associated with the vertices indexed by the set $\mathcal{S} \subseteq \{1, 2, \ldots, N\}$.

We start by constructing an ideal filter bank of M graph spectral filters, where for a set of band endpoints $0 = \tau_0 < \tau_1 < \ldots < \tau_{M-1} \leq \tau_M$ (with $\tau_M > \lambda_{\max}$), the m^{th} filter is defined as

$$h_m(\lambda) = \begin{cases} 1, & \tau_{m-1} \leq \lambda < \tau_m \\ 0, & \text{otherwise} \end{cases}, \quad m = 1, 2, \ldots, M. \tag{1}$$

Fig. 2 shows an example of such an ideal filter bank. Note that for each $\ell \in \{0, 1, \ldots, N-1\}$, $h_m(\lambda_\ell) = 1$ for exactly one m. Equivalently, we are forming a partition $\{\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_M\}$ of $\{0, 1, \ldots, N-1\}$ and setting

$$h_m(\lambda_\ell) = \begin{cases} 1, & \text{if } \ell \in \mathcal{R}_m \\ 0, & \text{otherwise} \end{cases}, \quad m = 1, 2, \ldots, M.$$

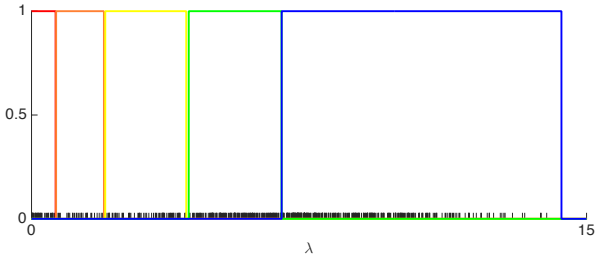

Fig. 2. Example ideal filter bank. The red, orange, yellow, green, and blue filters span 31, 31, 63, 125, and 250 graph Laplacian eigenvalues, respectively, on a 500 node sensor network with a maximum graph Laplacian eigenvalue of 14.3. The tick marks on the x-axis represent the locations of the graph Laplacian eigenvalues.

The next step, which we discuss in detail in the next section, is to partition the vertex set $\mathcal{V}$ into subsets $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_M$ such that $\mathcal{V}_m$ forms a uniqueness set for col $(\mathbf{U}_{\mathcal{R}_m})$.

Definition 1 (Uniqueness set [27])**.** *Let $\mathcal{P}$ be a subspace of $\mathbb{R}^n$. Then a subset $\mathcal{V}_s$ of the vertices $\mathcal{V}$ is a uniqueness set for $\mathcal{P}$ if and only if for all $\mathbf{f}, \mathbf{g} \in \mathcal{P}$, $\mathbf{f}_{\mathcal{V}_s} = \mathbf{g}_{\mathcal{V}_s}$ implies $\mathbf{f} = \mathbf{g}$. That is, if two signals in $\mathcal{P}$ have the same values on the vertices in the uniqueness set $\mathcal{V}_s$, then they must be the same signal.*

The following equivalent characterization of a uniqueness set is often useful.

Lemma 1 ([29], [31])**.** *The set $\mathcal{S}$ of k vertices is a uniqueness set for col$(\mathbf{U}_{\mathcal{T}})$ if and only if the matrix whose columns are $\mathbf{u}_{\mathcal{T}_1}, \mathbf{u}_{\mathcal{T}_2}, \ldots, \mathbf{u}_{\mathcal{T}_k}, \boldsymbol{\delta}_{\mathcal{S}_1^c}, \boldsymbol{\delta}_{\mathcal{S}_2^c}, \ldots, \boldsymbol{\delta}_{\mathcal{S}_{n-k}^c}$ is nonsingular, where each $\boldsymbol{\delta}_{\mathcal{S}_i^c}$ is a Kronecker delta centered on a vertex not included in S.*

The m^{th} channel of the analysis bank filters the graph signal by an ideal filter on subband $\mathcal{R}_m$, and downsamples the result onto the vertices in $\mathcal{V}_m$. For synthesis, we can interpolate from the samples on $\mathcal{V}_m$ to col $(\mathbf{U}_{\mathcal{R}_m})$. Namely, denoting the analysis coefficients of the m^{th} branch by $\mathbf{y}_{\mathcal{V}_m}$, we have

$$\mathbf{f}_{rec} = \sum_{m=1}^{M} \mathbf{U}_{\mathcal{R}_m} \mathbf{U}_{\mathcal{V}_m,\mathcal{R}_m}^{-1} \mathbf{y}_{\mathcal{V}_m}. \tag{2}$$

If there is no error in the coefficients, then the reconstruction is perfect, because $\mathcal{V}_m$ is a uniqueness set for col $(\mathbf{U}_{\mathcal{R}_m})$, ensuring $\mathbf{U}_{\mathcal{V}_m,\mathcal{R}_m}$ is full rank. Fig. 3 shows the architecture of the proposed M-channel critically sampled filter bank (M-CSFB) with interpolation on the

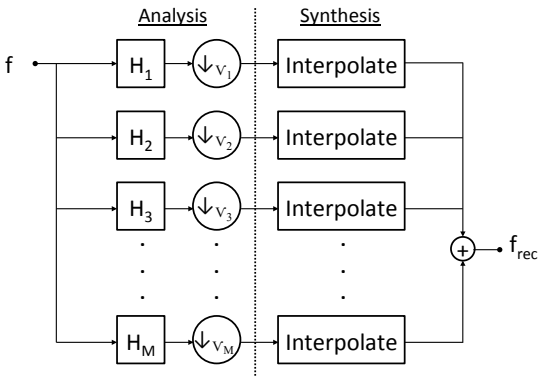

Fig. 3. The M-channel critically sampled filter bank architecture. The sets $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_M$ form a partition of the set $\mathcal{V}$ of vertices, where each set $\mathcal{V}_m$ is a uniqueness set for graph signals supported on a different subband in the graph spectral domain.

synthesis side.

III. Partitioning the Graph into Uniqueness Sets for Different Frequency Bands

In this section, we show how to partition the set of vertices into uniqueness sets for different subbands of the graph Laplacian eigenvectors. We start with the easier case of $M = 2$ and then examine the general case.

A. $M = 2$ channels

First we show that if a set of vertices is a uniqueness set for a set of signals contained in a band of spectral frequencies, then the complement set of vertices is a uniqueness set for the set of signals with no energy in that band of spectral frequencies.

Proposition 1. *On a graph $\mathcal{G}$ with N vertices, let $\mathcal{T} \subseteq \{0, 1, \ldots, N-1\}$ denote a subset of the graph Laplacian eigenvalue indices, and let $\mathcal{T}^c = \{0, 1, \ldots, N-1\} \setminus \mathcal{T}$. Then $\mathcal{S}^c$ is a uniqueness set for col$(\mathbf{U}_{\mathcal{T}^c})$ if and only if $\mathcal{S}$ is a uniqueness set for col$(\mathbf{U}_{\mathcal{T}})$.*

This fact follows from either the CS decomposition [36, Equation (32)]) or the nullity theorem [37, Theorem 2.1]. We also provide a standalone proof below that only requires that the space spanned by the first k columns of $\mathbf{U}$ is orthogonal to the space spanned by the last $N - k$ columns, not that $\mathbf{U}$ is an orthogonal matrix.

Proof of Proposition 1: We assume without loss of generality that $\mathcal{T} = \{0, 1, 2, \ldots, k-1\}$. Suppose first that the set $\mathcal{S}$ is a uniqueness set for col$(\mathbf{U}_{\mathcal{T}})$, but $\mathcal{S}^c$ is not a uniqueness set for col$(\mathbf{U}_{\mathcal{T}^c})$. Then by Lemma 1, the matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{u}_0 & \mathbf{u}_1 & \cdots & \mathbf{u}_{k-1} & \boldsymbol{\delta}_{\mathcal{S}_1^c} & \boldsymbol{\delta}_{\mathcal{S}_2^c} \cdots & \boldsymbol{\delta}_{\mathcal{S}_{N-k}^c} \end{bmatrix}$$

has full rank, and the matrix

$$\mathbf{B} = \begin{bmatrix} \mathbf{u}_k & \mathbf{u}_{k+1} & \cdots & \mathbf{u}_{N-1} & \boldsymbol{\delta}_{\mathcal{S}_1} & \boldsymbol{\delta}_{\mathcal{S}_2} \cdots & \boldsymbol{\delta}_{\mathcal{S}_k} \end{bmatrix}$$

is singular, implying

$$\text{span}(\mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_{N-1}, \boldsymbol{\delta}_{\mathcal{S}_1}, \boldsymbol{\delta}_{\mathcal{S}_2}, \ldots, \boldsymbol{\delta}_{\mathcal{S}_k}) \neq \mathbb{R}^N. \tag{3}$$

Since

$$\dim(\text{span}(\mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_{N-1})) = N - k$$

and

$$\dim(\text{span}(\boldsymbol{\delta}_{S_1}, \boldsymbol{\delta}_{S_2}, \ldots, \boldsymbol{\delta}_{\mathcal{S}_k})) = k,$$

equation (3) implies that there must exist a vector $\mathbf{x} \neq \mathbf{0}$ such that

$$\mathbf{x} \in \text{span}(\mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_{N-1}) \text{ and}$$
$$\mathbf{x} \in \text{span}(\boldsymbol{\delta}_{\mathcal{S}_1}, \boldsymbol{\delta}_{\mathcal{S}_2}, \ldots, \boldsymbol{\delta}_{\mathcal{S}_k}).$$

Yet, $\mathbf{x} \in \text{col}(\mathbf{U}_{\mathcal{T}^c})$ implies $\mathbf{x}$ is orthogonal to $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$, and, similarly, $\mathbf{x} \in \text{span}(\boldsymbol{\delta}_{\mathcal{S}_1}, \boldsymbol{\delta}_{\mathcal{S}_2}, \ldots, \boldsymbol{\delta}_{\mathcal{S}_k})$ implies $\mathbf{x}$ is orthogonal to $\boldsymbol{\delta}_{\mathcal{S}_1^c}, \boldsymbol{\delta}_{\mathcal{S}_2^c}, \ldots, \boldsymbol{\delta}_{\mathcal{S}_{N-k}^c}$. In matrix notation, we have $\mathbf{A}^\top \mathbf{x} = \mathbf{0}$, so $\mathbf{A}^\top$ has a non-trivial null space, and thus the square matrix $\mathbf{A}$ is not full rank and $\mathcal{S}$ is not a uniqueness set for col$(\mathbf{U}_{\mathcal{T}})$, a contradiction. We conclude that if $\mathbf{A}$ is full rank, then $\mathbf{B}$ must be full rank and $\mathcal{S}^c$ is a uniqueness set for col$(\mathbf{U}_{\mathcal{T}^c})$, completing the proof of sufficiency. Necessity follows from the same argument, with the roles of $\mathbf{A}$ and $\mathbf{B}$ interchanged. ■

The Steinitz exchange lemma [38] guarantees that we can find the uniqueness set $\mathcal{S}$ (and thus $\mathcal{S}^c$), and the graph signal processing literature contains methods such as Algorithm 1 of [31] to do so.

B. $M > 2$ channels

The issue with using the methods of Proposition 1 for the case of $M > 2$ is that while the submatrix $\mathbf{U}_{\mathcal{S}^c, \mathcal{T}^c}$ is nonsingular, it is not necessarily orthogonal, and so we cannot proceed with an inductive argument. The following proposition and corollary circumvent this issue by only using the nonsingularity of the original matrix. The proof of the following proposition is due to Federico Poloni [39], and we later discovered the same method in [40], [41, Theorem 3.3].

Proposition 2. *Let $\mathbf{A}$ be an $N \times N$ nonsingular matrix, and $\beta = \{\beta_1, \beta_2, \ldots, \beta_M\}$ be a partition of $\{1, 2, \ldots, N\}$. Then there exists another partition $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_M\}$ of $\{1, 2, \ldots, N\}$ with $|\alpha_i| = |\beta_i|$ for all i such that the M square submatrices $\mathbf{A}_{\alpha_i, \beta_i}$ are all nonsingular.*

Proof: First consider the case $M = 2$, and let $k = |\beta_1|$. Then by the generalized Laplace expansion [42],

$$\det(\mathbf{A}) = \sum_{\{\alpha_1 \subset \{1,2,\ldots,N\} : |\alpha_1| = k\}} \sigma_{\alpha_1, \beta_1} \det(\mathbf{A}_{\alpha_1, \beta_1}) \det(\mathbf{A}_{\alpha_1^c, \beta_2}), \tag{4}$$

where the sign $\sigma_{\alpha_1,\beta_1}$ of the permutation determined by α_1 and β_1 is equal to 1 or -1. Since $\det(\mathbf{A}) \neq 0$, one of the terms in the summation of (4) must be nonzero, ensuring a choice of α_1 such that the submatrices $\mathbf{A}_{\alpha_1,\beta_1}$ and $\mathbf{A}_{\alpha_1^c,\beta_2}$ are nonsingular. We can choose $\{\alpha_1, \alpha_1^c\}$ as the desired partition. For $M > 2$, by induction, we have

$$\det(\mathbf{A}) = \sum_{\{\text{Partitions } \alpha \text{ of } \{1,2,\ldots,N\}:|\alpha_i|=|\beta_i| \ \forall i\}} \sigma_\alpha \prod_{i=1}^{M} \det(\mathbf{A}_{\alpha_i,\beta_i}), \quad (5)$$

where again σ_α is always 1 or -1 and one of the terms in the summation in (5) must be nonzero, yielding the desired partition. ∎

Corollary 1. *For any partition $\{\mathcal{R}_1, \mathcal{R}_2, \ldots \mathcal{R}_M\}$ of the graph Laplacian eigenvalue indices $\{0, 1, \ldots, N-1\}$ into M subsets, there exists a partition $\{\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_M\}$ of the graph vertices into M subsets such that for every $m \in \{1, 2, \ldots, M\}$, $|\mathcal{V}_m| = |\mathcal{R}_m|$ and $\mathcal{V}_m$ is a uniqueness set for col $(\mathbf{U}_{\mathcal{R}_m})$.*

Proof: By Proposition 2, we can find a partition such that $\mathbf{U}_{\mathcal{V}_m,\mathcal{R}_m}$ is nonsingular for all m. Let $\mathbf{E}_m$ be the matrix formed by joining the k_m columns of $\mathbf{U}$ indexed by $\mathcal{R}_m$ with $N - k_m$ Kronecker deltas centered on all vertices not included in $\mathcal{V}_m$. By Lemma 1, it suffices to show that the matrices $\mathbf{E}_m$ are all nonsingular. Yet, for all m, we have

$$|\det(\mathbf{E}_m)| = |\det(\mathbf{U}_{\mathcal{V}_m,\mathcal{R}_m})| \neq 0.$$

∎

Corollary 1 ensures the existence of the desired partition, and the proof of Proposition 2 suggests that we can find it inductively. However, given a partition of the columns of $\mathbf{A}$ into two sets $\mathcal{T}$ and $\mathcal{T}^c$, Proposition 2 does not provide a constructive method to partition the rows of $\mathbf{A}$ into two sets $\mathcal{S}$ and $\mathcal{S}^c$ such that the submatrices $\mathbf{A}_{\mathcal{S},\mathcal{T}}$ and $\mathbf{A}_{\mathcal{S}^c,\mathcal{T}^c}$ are nonsingular. This problem is studied in the more general framework of matroid theory in [41], which gives an algorithm to find the desired row partition into two sets. We summarize this method in Algorithm 1, which takes in a partition $\{\mathcal{R}_1, \mathcal{R}_2, \ldots \mathcal{R}_M\}$ of the spectral indices and constructs the partition $\{\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_M\}$ of the vertices. In Fig. 4, we show two examples of the resulting partitions.

Remark 1. *While Algorithm 1 always finds a partition into uniqueness sets, such a partition is usually not unique, and the initial choices of γ_i in each loop play a significant role in the final partition. In the numerical experiments in the next section, we use the greedy algorithm in [31, Algorithm 1] to find an initial choice for γ_1, and use row reduction after permuting the complement of γ_1 to the top to find an initial choice for γ_2.*

Algorithm 1 Partition the vertices into uniqueness sets for each frequency band

Input $\mathbf{U}$, a partition $\{\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_M\}$
$\mathcal{S} \leftarrow \emptyset$
for $m = 1, 2, \ldots, M$ **do**
  Find sets $\gamma_1, \gamma_2 \subset \mathcal{S}^c$ s.t. $\mathbf{U}_{\gamma_1,\mathcal{R}_m}$ and $\mathbf{U}_{\gamma_2,\mathcal{R}_{m+1:M}}$ are nonsingular
  while $\gamma_1 \cap \gamma_2 \neq \emptyset$ **do**
    Find a chain of pivots from an element $y \in \mathcal{S}^c \setminus (\gamma_1 \cup \gamma_2)$ to an element $z \in \gamma_1 \cap \gamma_2$ (c.f. [41] for details)
    Update γ_1 and γ_2 by carrying out a series of exchanges resulting with y and z each appearing in exactly one of γ_1 or γ_2
  end while
  $\mathcal{V}_m \leftarrow \gamma_1$
  $\mathcal{S} \leftarrow \mathcal{S} \cup \gamma_1$
end for
Output the partition $\{\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_M\}$

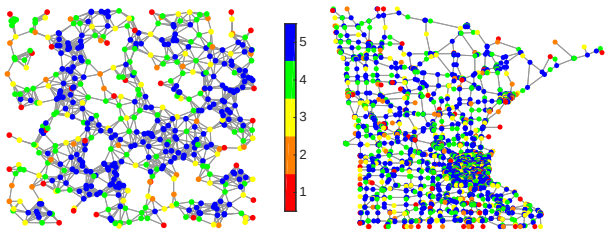

Fig. 4. Partitions of a 500 node random sensor network and the Minnesota road network [43] into uniqueness sets for five different spectral bands, with the indices increasing from lowpass bands (1) to highpass bands (5).

IV. Transform Properties

We now briefly examine some properties of the proposed transform.

A. Dictionary atoms

Let $\mathbf{M}_m \in \mathbb{R}^{|\mathcal{V}_m| \times N}$ be the downsampling matrix for the m^{th} channel. That is, $\mathbf{M}_m(i,j) = 1$ if vertex j is the i^{th} element of $\mathcal{V}_m$, and 0 otherwise. The proposed transform is a linear mapping $\mathcal{F} : \mathbb{R}^N \to \mathbb{R}^N$ by $\mathcal{F}\mathbf{f} = \mathbf{T}\mathbf{f}$, where

$$\mathbf{T} = \begin{bmatrix} \mathbf{M}_1 h_1(\boldsymbol{\mathcal{L}}) \\ \mathbf{M}_2 h_2(\boldsymbol{\mathcal{L}}) \\ \vdots \\ \mathbf{M}_M h_M(\boldsymbol{\mathcal{L}}) \end{bmatrix} \in \mathbb{R}^{N \times N}.$$

The atoms of the resulting dictionary are given by the columns of $\mathbf{T}^\top$. While the transform is not orthogonal, each atom is orthogonal to all atoms concentrated on other spectral bands. Informally, this is because the

atoms are projections of Kronecker deltas onto the orthogonal subspaces spanned by the Laplacian eigenvectors of each band. More formally, each atom is of the form $h_m(\boldsymbol{\mathcal{L}})\boldsymbol{\delta}_i$, where vertex i is in $\mathcal{V}_m$. If $m \neq m'$, then the inner product of the two atoms from different bands is given by

$$\begin{aligned}&\langle h_m(\boldsymbol{\mathcal{L}})\boldsymbol{\delta}_i, h_{m'}(\boldsymbol{\mathcal{L}})\boldsymbol{\delta}_{i'}\rangle \\ &\quad = \boldsymbol{\delta}_i^\top \mathbf{U} h_m(\boldsymbol{\Lambda})\mathbf{U}^*\mathbf{U} h_{m'}(\boldsymbol{\Lambda})\mathbf{U}^*\boldsymbol{\delta}_{i'} = 0, \end{aligned} \quad (6)$$

since $\mathbf{U}^*\mathbf{U} = \mathbf{I}$ and $h_m(\lambda)h_{m'}(\lambda) = 0$ for all λ by design. Note also that the wavelet atoms at all scales ($m > 1$) have mean zero, as they have no energy at eigenvalue zero.

B. Signals that are sparsely represented by the M-CSFB transform

- Concentrated in the graph Fourier domain
- If spectral coefficients decay, transform coefficients also decay
- Characterization of joint localization? Localization in vertex domain guaranteed when polynomial filters used
- We plan to more formally characterize the relationships between the decay of the analysis coefficients under the proposed transform and different properties of graph signals, as well as the underlying graph structure. Globally smooth signals trivially lead to sparse analysis coefficients because the coefficients are only nonzero for first set of vertices in the partition. The same line of reasoning applies more generally to signals that are sparse in the graph spectral domain. More interesting is a theory for the coefficient decay for piecewise-smooth graph signals such as the ones shown in Fig. 6(a) and Fig. 7(a).

V. Illustrative Examples I: Exact Calculations

We can represent the proposed filter bank as an $N \times N$ dictionary matrix $\boldsymbol{\Phi}$ that maps graph signals to their filter bank analysis coefficients.

A. Joint vertex-frequency localization of atoms and example analysis coefficients

We start by empirically showing that the dictionary atoms (the columns of $\boldsymbol{\Phi}$) are jointly localized in the vertex and graph spectral domains. On the Stanford bunny graph [44] with 2503 vertices, we partition the spectrum into five bands, and show the resulting partition into uniqueness sets in Fig. 6(c). The first row of Fig. 5 shows five different example atoms whose energy is concentrated on different spectral bands. We see that these atoms are generally localized in the vertex domain as well. Some atoms such as the one shown at wavelet scale 3 are more spread in the vertex domain, possibly as a result of using ideal filters in the filter bank, an issue we revisit in Section VIII. The second row of Fig. 5 shows the spectral content of all atoms in each band, with the average for each represented by a thick black line. As expected, the energies of the atoms are localized in the spectral domain as well.

Next we apply the proposed transform to a piecewise-smooth graph signal f that is shown in the vertex domain in Fig. 6(a), and in the graph spectral domain in Fig. 6(b). The full set of analysis coefficients is shown in Fig. 6(d), and these are separated by band in the third row of Fig. 5. We see that with the exception of the lowpass channel, the coefficients are clustered around the two main discontinuities (around the midsection and tail of the bunny). The bottom row of Fig. 5 shows the interpolation of these coefficients onto the corresponding spectral bands, and if we sum these reconstructions together, we recover exactly the original signal in Fig. 6(a).

B. Sparse approximation

Next, we compress a piecewise-smooth graph signal $\mathbf{f}$ via the sparse coding optimization

$$\underset{\mathbf{x}}{\operatorname{argmin}} \, ||\mathbf{f} - \boldsymbol{\Phi}\mathbf{x}||_2^2 \quad \text{subject to } ||\mathbf{x}||_0 \leq T, \quad (7)$$

where T is a predefined sparsity level. After first normalizing the atoms of various critically-sampled dictionaries, we use the greedy orthogonal matching pursuit (OMP) algorithm [45], [46] to approximately solve (7). For the M-channel filter bank, the partition into uniqueness sets is shown in Fig. 4, and the filter bank is shown in Fig. 2. We show the reconstruction errors in Fig. 7(d).

VI. Polynomial Filter Bank Design

In the numerical examples in the previous sections, we have computed a full eigendecomposition of the graph Laplacian and subsequently used it for all three of the filtering, sampling, and interpolation operations; however, such an eigendecomposition does not scale well with the size of the graph as it requires $\mathcal{O}(N^3)$ operations with naive methods. In these next three sections, we develop a fast approximate version of the proposed transform that scales more efficiently for large, sparse graphs.

A. Approximation by polynomial filters

Fast, approximate methods for computing $h_m(\boldsymbol{\mathcal{L}})\mathbf{f}$ a function of sparse matrix times a vector, include approximating the function $h_m(\cdot)$ by a polynomial (e.g., via a truncated Chebyshev or Legendre expansion), approximating $h_m(\cdot)$ by a rational function, Krylov space

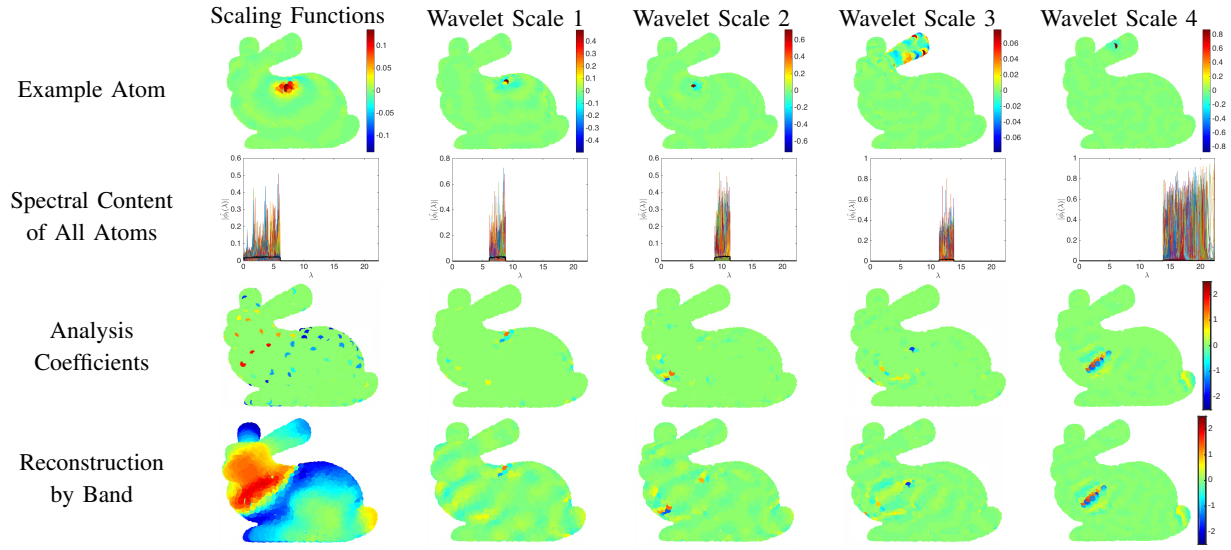

Fig. 5. M-channel filter bank example. The first row shows example atoms in the vertex domain. The second row shows all atoms in the spectral domain, with an average of the atoms in each band shown by the thick black lines. The third row shows the analysis coefficients of Fig. 6(d) by band, and the last row is the interpolation by band from those coefficients.

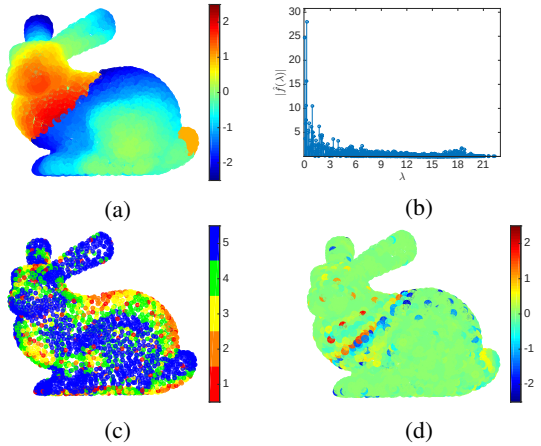

Fig. 6. (a)-(b) Piecewise smooth signal on the Stanford bunny graph [44] in the vertex and graph spectral domains, respectively. (c) Partition of the graph into uniqueness sets for five different spectral bands. (d) M-channel filter bank analysis coefficients of the signal shown in (a) and (b).

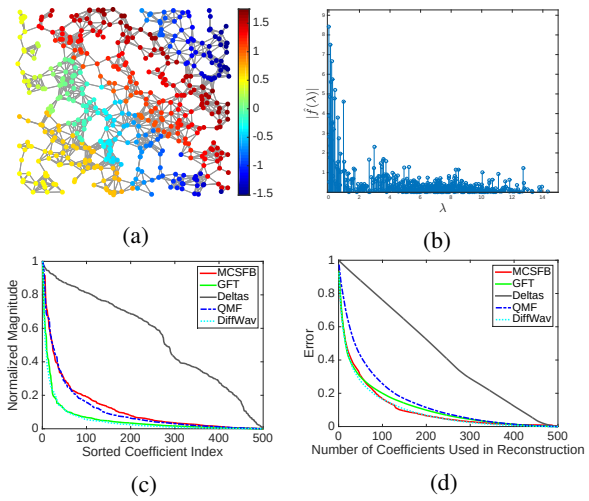

Fig. 7. Compression example. (a)-(b) Piecewise-smooth signal from [25, Fig. 11] in the vertex and graph spectral domains. (c) The normalized sorted magnitudes of the transform coefficients for the proposed M-channel critically sampled filter bank, the graph Fourier transform, the basis of Kronecker deltas, the quadrature mirror filterbank [14], and the diffusion wavelet transform [8]. (d) The reconstruction errors $\frac{\left|\left|\mathbf{f}_{\text{reconstruction}}-\mathbf{f}\right|\right|_2}{||\mathbf{f}||_2}$, as a function of the sparsity threshold T in (7).

methods (Lanczos in our case of a symmetric matrix $\boldsymbol{\mathcal{L}}$), the matrix version of the Cauchy integral theorem (see, e.g., [47]-[50] and references therein). The first three of these methods have been examined in graph signal processing settings [10], [32], [51]-[54]. While all of these could be used to approximate the filters in the analysis step of the proposed filter bank, we focus on order K Chebyshev polynomial approximations of the form

$$\tilde{h}(\boldsymbol{\mathcal{L}})\mathbf{f} := \sum_{k=0}^{K} \alpha_k \bar{T}_k(\boldsymbol{\mathcal{L}})\mathbf{f}. \tag{8}$$

In (8), $\bar{T}_k(\cdot)$ are Chebyshev polynomials shifted to the interval $[0, \lambda_{\max}]$. Thus, $\bar{T}_0(\boldsymbol{\mathcal{L}})\mathbf{f} = \mathbf{f}$, $\bar{T}_1(\boldsymbol{\mathcal{L}})\mathbf{f} = \frac{2}{\lambda_{\max}}\boldsymbol{\mathcal{L}}\mathbf{f} - \mathbf{f}$, and for $k \geq 2$, by the three term recurrence relation of Chebyshev polynomials, we have

$$\bar{T}_k(\boldsymbol{\mathcal{L}})\mathbf{f} = \frac{4}{\lambda_{\max}}\left(\boldsymbol{\mathcal{L}} - \frac{\lambda_{\max}}{2}\mathbf{I}\right)\bar{T}_{k-1}(\boldsymbol{\mathcal{L}})\mathbf{f} - \bar{T}_{k-2}(\boldsymbol{\mathcal{L}})\mathbf{f}.$$

The coefficients in (8) are often taken to be $\alpha_0 = \frac{1}{2}c_0$

and $\alpha_k = c_k$ for $k = 1, 2, \ldots, K$, where $\{c_k\}_{k=0,\ldots,K}$ are the truncated Chebyshev expansion coefficients

$$\begin{aligned} c_k &:= \langle h, \bar{T}_k \rangle \\ &= \frac{2}{\pi} \int_0^{\pi} \cos(k\phi)\ h\Big(\frac{\lambda_{\max}}{2}\big(\cos(\phi)+1\big)\Big)\ d\phi. \end{aligned}$$

However, the oscillations that arise in Chebyshev polynomial approximations of bandpass filters may result in larger values of $\tilde{h}_m(\lambda)\tilde{h}_{m'}(\lambda)$, even when the ideal filters $h_m(\cdot)$ and $h_{m'}(\cdot)$ have supports that do not come close to overlapping. This negates the orthogonality of the atoms across bands shown in (6). In an attempt to at least preserve *near* orthogonality across bands, we therefore use the Jackson-Chebyshev polynomial approximations from [55], [56] that damp the Gibbs oscillations appearing in Chebyshev expansions. With the damping, $\alpha_0 = \frac{1}{2}c_0$ and $\alpha_k = \gamma_{k,K} c_k$ for $k = 1, 2, \ldots, K$, where, as presented in [55],

$$\begin{aligned} \gamma_{k,K} = &\frac{\left(1 - \frac{k}{K+2}\right)\sin\left(\frac{\pi}{K+2}\right)\cos\left(\frac{k\pi}{K+2}\right)}{\sin\left(\frac{\pi}{K+2}\right)} \\ &+ \frac{\frac{1}{K+2}\cos\left(\frac{\pi}{K+2}\right)\sin\left(\frac{k\pi}{K+2}\right)}{\sin\left(\frac{\pi}{K+2}\right)}. \end{aligned}$$

Figure 8 shows the Jackson-Chebyshev polynomial approximations for ideal bandpass filters of two different graphs.[1]

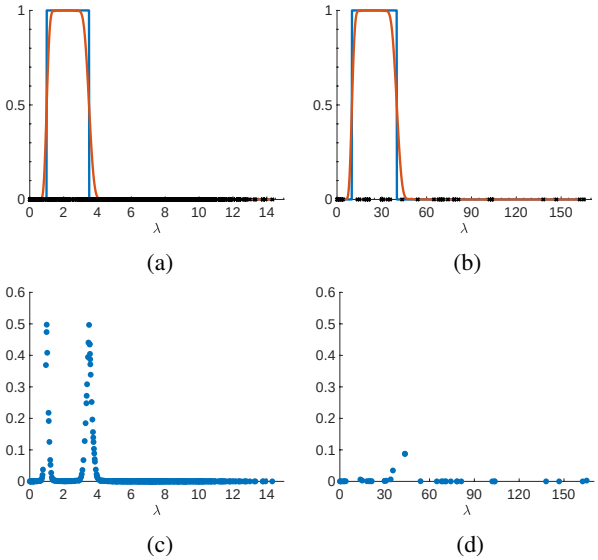

Fig. 8. Degree 80 Jackson-Chebyshev polynomial approximations for ideal bandpass filters on the (a) 500 vertex random sensor network of Figure 4 and (b) Andrianov net25 graph from [57] with 9,520 vertices. In (c) and (d), we show the errors $|\tilde{h}(\lambda_\ell) - h(\lambda_\ell)|$ at each of the Laplacian eigenvalues of the corresponding graphs in (a) and (b).

B. Filter bank design

We can quantify the worst case error introduced when approximating $h_m(\cdot)$ by an approximant $\tilde{h}_m(\cdot)$ as follows:

$$\begin{aligned} &||\tilde{h}_m(\boldsymbol{\mathcal{L}}) - h_m(\boldsymbol{\mathcal{L}})||_2 \\ &\quad = \max_{\ell \in \{0,1,\ldots,N-1\}} |\tilde{h}_m(\lambda_\ell) - h_m(\lambda_\ell)| \\ &\quad \leq \max_{\lambda \in [0, \lambda_{\max}]} |\tilde{h}_m(\lambda) - h_m(\lambda)|. \end{aligned} \tag{9}$$

While approximation theory often aims to minimize the upper bound in (9), only the errors exactly at the graph Laplacian eigenvalues affect the overall approximation error $||\tilde{h}_m(\boldsymbol{\mathcal{L}}) - h_m(\boldsymbol{\mathcal{L}})||_2$. Since, as seen in Figure 8, the errors of the Jackson-Chebyshev polynomial approximation are concentrated around the discontinuities of $h_m(\cdot)$, a guiding principle when designing the filter bank to be more amenable to fast approximation is to choose the endpoints $\{\tau_m\}_{m=1,\ldots,M-1}$ of the bandpass filters to be in gaps in the graph Laplacian spectrum. Unfortunately, we do not have access to the exact graph Laplacian eigenvalues (the reason for introducing this approximation in the first place is that they are too expensive to compute for large graphs); however, we can efficiently estimate the density of the spectrum in order to design the filters in a way that aims to have the endpoints close to fewer eigenvalues of $\boldsymbol{\mathcal{L}}$.

1) Estimating the spectral density: Lin et al. [58] provide an excellent overview of methods to approximate the *spectral density function* [59, Chapter 6]) (also called the *Density of States* or *empirical spectral distribution* [60, Chapter 2.4]) of a matrix, which in our context for the graph Laplacian $\boldsymbol{\mathcal{L}}$ is the probability measure

$$p_\lambda(s) := \frac{1}{N} \sum_{\ell=0}^{N-1} \mathbb{1}_{\{\lambda_\ell = s\}}.$$

Here, we use a variant of the Kernel Polynomial Method [61]-[63] described in [58] to estimate the *cumulative spectral density function* or *empirical spectral cumulative distribution*

$$P_\lambda(z) := \frac{1}{N} \sum_{\ell=0}^{N-1} \mathbb{1}_{\{\lambda_\ell \leq z\}}. \tag{10}$$

The procedure, outlined in Algorithm XXX, starts by estimating $\lambda_{\max}$, for example via the power iteration. Then for each of T linearly spaced points ξ_i between 0 and $\lambda_{\max}$, we use Hutchinson's stochastic trace estimator [64] to estimate η_i, the number of eigenvalues less than or equal to ξ_i. Defining the Heaviside function $\Theta_{\xi_i}(\lambda) :=$

[1]For the net25 graph, we have removed the self loops and added a single edge to connect the two connected components.

$\mathbb{1}_{\{\lambda \le \xi_i\}}$, we have

$$\eta_i = \mathrm{tr}\Big(\Theta_{\xi_i}(\boldsymbol{\mathcal{L}})\Big) = \mathbb{E}[\mathbf{x}^\top \Theta_{\xi_i}(\boldsymbol{\mathcal{L}})\mathbf{x}] \tag{11}$$

$$\approx \frac{1}{J}\sum_{j=1}^{J} \mathbf{x}^{(j)\top} \Theta_{\xi_i}(\boldsymbol{\mathcal{L}})\mathbf{x}^{(j)} \tag{12}$$

$$\approx \frac{1}{J}\sum_{j=1}^{J} \mathbf{x}^{(j)\top} \tilde{\Theta}_{\xi_i}(\boldsymbol{\mathcal{L}})\mathbf{x}^{(j)}. \tag{13}$$

In (11), $\mathbf{x}$ is a random vector with each component having an independent and identical standard normal distribution. Each vector $\mathbf{x}^{(j)}$ in (12) is chosen according to this same distribution, and in our experiments, we take the default number of vectors to be $J = 30$. In (13), $\tilde{\Theta}_{\xi_i}$ is the Jackson-Chebyshev approximation to Θ_{ξ_i} discussed in Section VI-A, with a default polynomial order of $K = 30$. If we place the J random vectors into the columns of an $N \times J$ matrix $\mathbf{X}$, the computational cost of estimating the spectral distribution is dominated by computing

$$\tilde{\Theta}_{\xi_i}(\boldsymbol{\mathcal{L}})\mathbf{X} = \sum_{k=0}^{K} \alpha_k \bar{T}_k(\boldsymbol{\mathcal{L}})\mathbf{X} \tag{14}$$

for each ξ_i. Yet, we only need to compute $\{\bar{T}_k(\boldsymbol{\mathcal{L}})\mathbf{X}\}_{k=0,1,\ldots,K}$ recursively once, as this sequence can be reused for each ξ_i, with different choices of the α_k's. Therefore, the overall computational cost is $\mathcal{O}(KJ|\mathcal{E}|)$.

As in [11], once we compute the eigenvalue count estimates $\{\eta_i\}$, we approximate the empirical spectral cumulative distribution $P_\lambda(\cdot)$ by performing monotonic piecewise cubic interpolation [65] on the series of points $\left\{\left(\xi_i, \frac{\eta_i}{N}\right)\right\}_{i=1,2,\ldots,T}$. We denote the result as $\tilde{P}_\lambda(\cdot)$.

2) Choosing initial band ends: When selecting the band ends $\{\tau_m\}$ for each of the M ideal filters, we consider two factors: spectrum-adaptation and spacing. In our implementation, the filter bank can either be adapted to the spectral distribution or just to the support of the spectrum $[0, \lambda_{\max}]$, and it can be either evenly or logarithmically spaced (four options in all). For example, if the filter bank is only adapted to the support of the spectrum and is evenly spaced, then $\tau_m = \frac{m}{M}\lambda_{\max}$. Figure 9(a) shows a spectrum-adapted, logarithmically spaced choice with $\tau_m = \tilde{P}_\lambda^{-1}\left(\frac{1}{2}^{M-m}\right)$ for $m = 1, 2, \ldots, M$, such that approximately half of the eigenvalues are in the highest band, a quarter in the next highest band, and so forth.

3) Adjusting the band ends: In order to make the filters more amenable to approximation, we then adjust the initial choice of band endpoints so that they lie in lower density regions of the spectrum. Specifically, for each $m = 1, 2, \ldots, M-1$ and some choice of $\Delta > 0$,

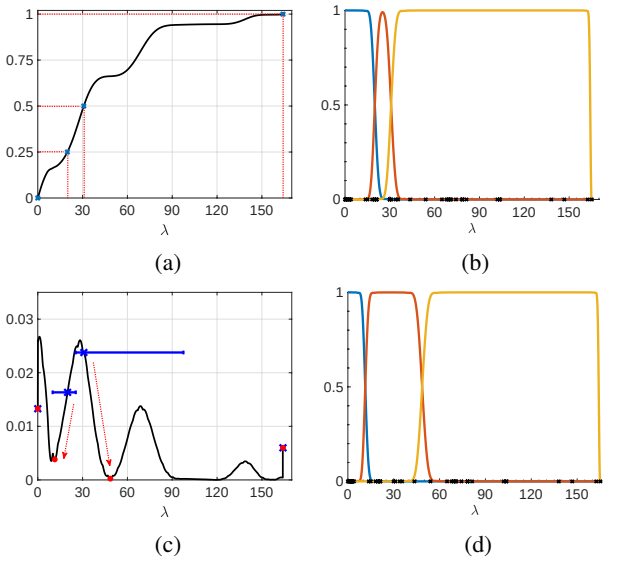

Fig. 9. (a) The approximate cumulative spectral density function, $\tilde{P}_\lambda(\cdot)$, for the net25 graph described in Figure 8. The blue X marks correspond to the initial choice of band endpoints computed by taking the inverse of logarithmically spaced points on the vertical axis. (b) The degree 80 Jackson-Chebyshev approximations to the ideal filters defined by the initial choice of band ends from (a). (c) The objective function of (15) (a discrete approximation of the spectral density function $p_\lambda(\cdot)$) with $\Delta = .001$. The blue horizontal lines correspond to the search intervals $\mathcal{I}_1$ and $\mathcal{I}_2$, and the red circles represent the adjusted band ends $\{\tau_m\}_{m=0,1,2,3}$. (d) The degree 80 Jackson-Chebyshev approximations to the ideal filters defined by the adjusted choice of band ends. Note that the errors between the approximate filters and ideal bandpass filters are concentrated in regions with fewer eigenvalues (gaps in the spectrum).

we let the final endpoint be

$$\tau_m^* = \underset{\tau \in \mathcal{I}_m}{\mathrm{argmin}} \left\{ \frac{\tilde{P}_\lambda(\tau + \Delta) - \tilde{P}_\lambda(\tau - \Delta)}{2\Delta} \right\}, \tag{15}$$

where $\mathcal{I}_m$ is an interval around the initial choice of τ_m. Figure 9(c) shows the objective function in (15), along with the initial band ends, search intervals, and adjusted band ends. Comparing Figure 9(b) and Figure 9(d), we see that adjusting the band ends leads to fewer eigenvalues falling close to the borders of the filters, reducing the error incurred by the polynomial approximation process.

VII. Non-Uniform Random Sampling and Reconstruction

Both the partitioning of the vertices into uniqueness sets described in Algorithm 1 and the synthesis via interpolation in (2) require a full eigendecomposition of the graph Laplacian to compute the matrix $\mathbf{U}$. In this section, we examine more efficient sampling and reconstruction methods that do not require the full eigendecomposition.

A. Sampling distribution

- Fast sampling and reconstruction. Two main approaches to sampling: greedy and random
- Most work on lowpass/smooth signals
- Review of non-uniform random sampling literature

Reference [35] has a nice review of the computational complexities of the various routines for identifying uniqueness sets. The random sampling method proposed in [32] does not require the full eigendecomposition of the graph Laplacian and therefore scales significantly better with the size of the graph. We are currently examining whether there is a way to extend the approach of [32] to non-uniformly randomly sample in a manner that leads with high probability to uniqueness sets for higher bands of the graph Laplacian spectrum.

- Note that the benefits of non-uniform sampling come when eigenvectors are more localized (c.f., Section 5.1.2 of Gilles? paper); this is especially important for high pass filters
- Link to literature on column subset selection (statistical leverage)
- Note that for any walk regular graph, optimal sampling distribution is uniform; give some examples (cycle, path, grid, vertex transitive graphs, etc.); cite godsil paper on symmetry and eigenvectors

B. Number of samples

To yield a critically sampled transform, one option is to choose the number of samples for each band to be in correspondence with the initial filter bank design. For example, if the filter bank is designed to be adapted to the spectrum with logarithmic spacing, we can choose $\frac{N}{2}$ samples for the highest band, $\frac{N}{4}$ for the next highest, and so forth. However, the adjustments we make in Section VI-B3 affect the number of eigenvalues contained in each band. Since we have an estimate of the cumulative spectral distribution, one approximation for the number of samples in the adjusted m^{th} band is to round $N * (\tilde{P}_\lambda(\tau_m) - \tilde{P}_\lambda(\tau_{m-1}))$. As a band end τ_m may fall at a point where $\tilde{P}_\lambda$ has been interpolated via cubic functions, another option is to estimate the number of eigenvalues between τ_{m-1} and τ_m, once again with the stochastic trace estimator in (13), except using the bandpass filter $h_m(\lambda)$ from (1). Since we have already computed and stored the series of matrices $\bar{T}_k(\boldsymbol{\mathcal{L}})\mathbf{X}$, we just need to compute the polynomial approximation coefficients $\{\alpha_k\}$ in (14) for $h_m(\lambda)$, and substitute the resulting $\tilde{h}_m(\boldsymbol{\mathcal{L}})\mathbf{x}^{(j)}$ values into (13) for an estimate of the number of eigenvalues in the m^{th} band. An added benefit of this extra step is that the thresholds $\{\tau_m\}$ are chosen to be in areas of low spectral density, which improves the accuracy of the eigenvalue count estimate [55]. If critical sampling is desired, we can make small adjustments to ensure the total number of samples is equal to N. Our default is to add samples to the lowest band if the estimated total is lower than N, and remove samples from the highest band if the estimated total is higher than N.

Note that due to the polynomial approximation, the dimension of $\text{col}(\tilde{h}_m(\boldsymbol{\mathcal{L}}))$ is at least as large as the dimension of $\text{col}(h_m(\boldsymbol{\mathcal{L}}))$, with the difference depending on the number of Laplacian eigenvalues just outside the end points of $h_m(\cdot)$ and the degree of approximation used for $\tilde{h}_m(\cdot)$. Therefore, we expect that to perfectly reconstruct signals in $\text{col}(\tilde{h}_m(\boldsymbol{\mathcal{L}}))$, we need more samples than the number of eigenvalues in the support of $h_m(\cdot)$. In Section VIII, we explore how the reconstruction error is reduced as we increase the number of samples in each band. The tradeoff is the loss of critical sampling, as the total number of samples extends beyond N.

C. Interpolation

The exact interpolation (2) requires the eigenvector matrix $\mathbf{U}$, and in case $U_{\mathcal{V}_m,\mathcal{R}_m}$ is not full rank, the standard least squares reconstruction for the m^{th} channel

$$\mathbf{f}_{m,rec} = \mathbf{U}_{\mathcal{R}_m}(\mathbf{U}_{\mathcal{V}_m,\mathcal{R}_m}^\top \mathbf{U}_{\mathcal{V}_m,\mathcal{R}_m})^{-1}\mathbf{U}_{\mathcal{V}_m,\mathcal{R}_m}^\top \mathbf{y}_{\mathcal{V}_m} \tag{16}$$

also requires $\mathbf{U}_{\mathcal{R}_m}$. One option explored in [66], [67] is to leverage $\{\bar{T}_k(\boldsymbol{\mathcal{L}})\mathbf{X}\}$ again to approximate the column space of $\mathbf{U}_{\mathcal{R}_m}$ by filtering at least $|\mathcal{R}_m|$ standard normal random vectors with the filter $\tilde{h}_m(\cdot)$, possibly followed by orthonormalization via QR factorization.

A second approach suggested in [32] to efficiently reconstruct lowpass signals is to relax the optimization problem

$$\min_{\mathbf{z}\in\text{col}(\mathbf{U}_{\mathcal{R}_m})} ||\boldsymbol{\Omega}_{m,\mathcal{V}_m}^{-\frac{1}{2}}(\mathbf{M}_\mathbf{m}\mathbf{z} - \mathbf{y}_{\mathcal{V}_m})||_2$$

to

$$\min_{\mathbf{z}\in\mathbb{R}^N} ||\boldsymbol{\Omega}_{m,\mathcal{V}_m}^{-\frac{1}{2}}(\mathbf{M}_\mathbf{m}\mathbf{z} - \mathbf{y}_{\mathcal{V}_m})||_2 + \mathbf{z}^\top\varphi_m(\boldsymbol{\mathcal{L}})\mathbf{z}, \tag{17}$$

where $\boldsymbol{\Omega}_{m,\mathcal{V}_m}$ is a $|\mathcal{V}_m| \times |\mathcal{V}_m|$ diagonal matrix with the m^{th} channel sampling weights of $\mathcal{V}_m$ along the diagonal. The regularization term $\mathbf{z}^\top\varphi_m(\boldsymbol{\mathcal{L}})\mathbf{z}$ in (17) penalizes reconstructions with support outside of the desired spectral band. For lowpass signals, Puy et al. [32] take the penalty function $\varphi_m(\lambda)$ to be a nonnegative, nondecreasing polynomial, such as $\kappa\lambda^l$, with $\kappa > 0$ and l a positive integer. For more general classes of signals (i.e., the midpass and highpass signals output from the higher bands of the proposed filter bank), it is important to keep the nonnegativity property, in order to ensure that $\varphi_m(\boldsymbol{\mathcal{L}})$ is positive semi-definite and the optimization problem (17) is convex. However, we can drop the nondecreasing requirement, and instead choose penalty

functions concentrated outside the m^{th} spectral band. Options we explore include (i) $\varphi_m(\lambda) = \kappa\left[1 - \tilde{h}(\lambda)\right]$; (ii) $\varphi_m(\lambda) = \frac{1}{\tilde{h}(\lambda)+\epsilon} - \frac{1}{1+\epsilon}$ Check nonnegativity; and (iii) Spline approach of Saad. See Figure XXX.

From the first-order optimality conditions, the solution to (17) is the solution to the linear system of equations

$$\left(M_m^\top \mathbf{\Omega}_{m,\mathcal{V}_m}^{-1} M_m + \varphi_m(\boldsymbol{\mathcal{L}})\right)\mathbf{z} = M_m^\top \mathbf{\Omega}_{m,\mathcal{V}_m}^{-1} \mathbf{y}_{\mathcal{V}_m}, \tag{18}$$

which can be solved, for example, with the preconditioned conjugate gradient method. Add more details? Preconditioner? How to compute $\varphi_m(\boldsymbol{\mathcal{L}})\mathbf{z}$.

[Say something about which option we choose and why, and point to empirical comparison in the next section.] The analysis and synthesis operations for the proposed scalable M-channel filter bank are summarized in Algorithms XXX and ZZZ, respectively.

- Theoretical analysis of reconstruction error
- Reconstruction robustness to noisy or thresholded filter bank coefficients (see, e.g., [35, Section III.B] for more on reconstruction stability)

VIII. Illustrative Examples II: Approximate Calculations

- Reconstruction of bandpass signal. Tradeoff between number of samples and reconstruction error. Show different methods?
- Repeat bunny compression example. Much worse?
- Make sure to have some very large examples and show computation times

IX. Conclusion and Extensions

Insert key findings here

- Leveraged the computation of $\{\bar{T}_k(\boldsymbol{\mathcal{L}})\mathbf{X}\}$ in multiple ways
- Can be seen as a coarse, fast GFT, maybe compare to that literature

As with the classical wavelet construction, it is possible to iterate the filter bank on the output from the lowpass channel. This could be beneficial, for example, in the case that we want to visualize the graph signal at different resolutions on a sequence of coarser and coarser graphs. An interesting question for future work is how iterating the filter bank with fewer channels at each step compares to a single filter bank with more channels and each filter supported on a smaller region of the spectrum.

References

[1] N. Perraudin, J. Paratte, D. I Shuman, V. Kalofolias, P. Vandergheynst, and D. K. Hammond, "GSPBOX: A toolbox for signal processing on graphs," https://lts2.epfl.ch/gsp/.

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

[3] M. Crovella and E. Kolaczyk, "Graph wavelets for spatial traffic analysis," in *Proc. IEEE INFOCOM*, vol. 3, Mar. 2003, pp. 1848–1857.

[4] W. Wang and K. Ramchandran, "Random multiresolution representations for arbitrary sensor network graphs," in *Proc. IEEE Int. Conf. Acc., Speech, and Signal Process.*, vol. 4, May 2006, pp. 161–164.

[5] A. D. Szlam, M. Maggioni, R. R. Coifman, and J. C. Bremer, Jr., "Diffusion-driven multiscale analysis on manifolds and graphs: top-down and bottom-up constructions," in *Proc. SPIE Wavelets*, vol. 5914, Aug. 2005, pp. 445–455.

[6] M. Gavish, B. Nadler, and R. R. Coifman, "Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning," in *Proc. Int. Conf. Mach. Learn.*, Haifa, Israel, Jun. 2010, pp. 367–374.

[7] J. Irion and N. Saito, "Hierarchical graph Laplacian eigen transforms," *JSIAM Letters*, vol. 6, pp. 21–24, 2014.

[8] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 53–94, 2006.

[9] M. Maggioni, J. C. Bremer, R. R. Coifman, and A. D. Szlam, "Biorthogonal diffusion wavelets for multiscale representations on manifolds and graphs," in *Proc. SPIE Wavelet XI*, vol. 5914, Sep. 2005.

[10] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, Mar. 2011.

[11] D. I Shuman, C. Wiesmeyr, N. Holighaus, and P. Vandergheynst, "Spectrum-adapted tight graph wavelet and vertex-frequency frames," *IEEE Trans. Signal Process.*, vol. 63, no. 16, pp. 4223–4235, Aug. 2015.

[12] D. I Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Appl. Comput. Harmon. Anal.*, vol. 40, no. 2, pp. 260–291, Mar. 2016.

[13] S. K. Narang and A. Ortega, "Local two-channel critically sampled filter-banks on graphs," in *Proc. Int. Conf. Image Process.*, Hong Kong, Sep. 2010, pp. 333–336.

[14] ——, "Perfect reconstruction two-channel wavelet filter-banks for graph structured data," *IEEE. Trans. Signal Process.*, vol. 60, no. 6, pp. 2786–2799, Jun. 2012.

[15] ——, "Compact support biorthogonal wavelet filterbanks for arbitrary undirected graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4673–4685, Oct. 2013.

[16] A. Sakiyama and Y. Tanaka, "Oversampled graph Laplacian matrix for graph filter banks," *IEEE Trans. Signal Process.*, vol. 62, no. 24, pp. 6425–6437, Dec. 2014.

[17] O. Teke and P. P. Vaidyanathan, "Graph filter banks with M-channels, maximal decimation, and perfect reconstruction," in *Proc. IEEE Int. Conf. Acc., Speech, and Signal Process.*, Mar. 2016, pp. 4089–4093.

[18] ——, "Extending classical multirate signal processing theory to graphs – Part II: M-channel filter banks," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 423–437, Jan. 2017.

[19] L. J. Grady and J. R. Polimeni, *Discrete Calculus*. Springer, 2010.

[20] V. N. Ekambaram, G. C. Fanti, B. Ayazifar, and K. Ramchandran, "Circulant structures and graph signal processing," in *Proc. Int. Conf. Image Process.*, Melbourne, Australia, Sep. 2013.

[21] V. N. Ekambaram, G. Fanti, B. Ayazifar, and K. Ramchandran, "Critically-sampled perfect-reconstruction spline-wavelet filter-banks for graph signals," in *Proc. IEEE Glob. Conf. Signal and Inform. Process.*, Dec. 2013, pp. 475–478.

[22] M. S. Kotzagiannidis and P. L. Dragotti, "The graph FRI-framework- spline wavelet theory and sampling on circulant graphs," in *Proc. IEEE Int. Conf. Acc., Speech, and Signal Process.*, Shanghai, China, Mar. 2016.

[23] M. Jansen, G. P. Nason, and B. W. Silverman, "Multiscale methods for data on graphs and irregular multidimensional situations," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 71, no. 1, pp. 97–125, 2009.

[24] S. K. Narang and A. Ortega, "Lifting based wavelet transforms on graphs," in *Proc. APSIPA ASC*, Sapporo, Japan, Oct. 2009, pp. 441–444.

[25] D. I Shuman, M. J. Faraji, and P. Vandergheynst, "A multiscale pyramid transform for graph signals," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 2119–2134, Apr. 2016.

[26] 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.

[27] I. Pesenson, "Sampling in Paley-Wiener spaces on combinatorial graphs," *Trans. Amer. Math. Soc*, vol. 360, no. 10, pp. 5603–5627, 2008.

[28] S. K. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, May 2013, pp. 5445–5449.

[29] A. Anis, A. Gadde, and A. Ortega, "Towards a sampling theorem for signals on arbitrary graphs," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, May 2014, pp. 3864–3868.

[30] A. Gadde and A. Ortega, "A probabilistic interpretation of sampling theory of graph signals," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, Apr. 2015, pp. 3257–3261.

[31] H. Shomorony and A. S. Avestimehr, "Sampling large data on graphs," in *Proc. IEEE Glob. Conf. Signal and Inform. Process.*, Dec. 2014, pp. 933–936.

[32] G. Puy, N.Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *Appl. Comput. Harmon. Anal.*, vol. 44, no. 2, pp. 446–475, Mar. 2018.

[33] 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.

[34] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4845–4860, Sep. 2016.

[35] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3775–3789, Jul. 2016.

[36] C. Paige and M. Wei, "History and generality of the CS decomposition," *Linear Algebra Appl.*, vol. 208-209, pp. 303–326, Sep. 1994.

[37] G. Strang and T. Nguyen, "The interplay of ranks of submatrices," *SIAM Review*, vol. 46, no. 4, pp. 637–646, 2004.

[38] Wikipedia, https://en.wikipedia.org/wiki/Steinitz_exchange_lemma.

[39] F. Poloni, "Partitioning an orthogonal matrix into full rank square submatrices," MathOverflow, http://mathoverflow.net/q/243990.

[40] C. Greene, "A multiple exchange property for bases," *Proc. AMS*, vol. 39, no. 1, pp. 45–50, Jun. 1973.

[41] C. Greene and T. L. Magnanti, "Some abstract pivot algorithms," *SIAM J. Appl. Math.*, vol. 29, no. 3, pp. 530–539, Nov. 1975.

[42] Wikipedia, https://en.wikipedia.org/wiki/Laplace_expansion.

[43] D. Gleich, "The MatlabBGL Matlab library," http://www.cs.purdue.edu/homes/dgleich/packages/matlab_bgl/index.html.

[44] Stanford University Computer Graphics Laboratory, "The Stanford 3D Scanning Repository," http://graphics.stanford.edu/data/3Dscanrep/.

[45] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[46] M. Elad, *Sparse and Redundant Representations*. Springer, 2010.

[47] N. J. Higham, *Functions of Matrices*. Society for Industrial and Applied Mathematics, 2008.

[48] P. I. Davies and N. J. Higham, "Computing $f(a)b$ for matrix functions f," in *QCD and Numerical Analysis III*. Springer, 2005, pp. 15–24.

[49] A. Frommer and V. Simoncini, "Matrix functions," in *Model Order Reduction: Theory, Research Aspects and Applications*. Springer, 2008, pp. 275–303.

[50] C. Moler and C. Van Loan, "Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later," *SIAM Review*, vol. 45, no. 1, pp. 3–49, 2003.

[51] D. I Shuman, P. Vandergheynst, D. Kressner, and P. Frossard, "Distributed signal processing via Chebyshev polynomial approximation," *Trans. Signal Inf. Process. Netw.*, 2018, in press.

[52] A. Šušnjara, N. Perraudin, D. Kressner, and P. Vandergheynst, "Accelerated filtering on graphs using Lanczos method," *arXiv ePrints*, 2015. [Online]. Available: https://arxiv.org/abs/1509.04537

[53] X. Shi, H. Feng, M. Zhai, T. Yang, and B. Hu, "Infinite impulse response graph filters in wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1113–1117, Aug. 2015.

[54] A. Loukas, A. Simonetto, and G. Leus, "Distributed autoregressive moving average graph filters," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1931–1935, Nov. 2015.

[55] E. Di Napoli, E. Polizzi, and Y. Saad, "Efficient estimation of eigenvalue counts in an interval," *Numer. Linear Algebra Appl.*, vol. 23, no. 4, pp. 674–692, Aug. 2016.

[56] G. Puy and P. Pérez, "Structured sampling and fast reconstruction of smooth graph signals," *arXiv e-Prints*, 2017. [Online]. Available: http://arxiv.org/abs/1705.02202

[57] T. A. Davis and Y. Hu, "The University of Florida sparse matrix collection," *ACM Trans. Math. Softw.*, vol. 38, no. 1, pp. 1:1–1:25, 2011.

[58] L. Lin, Y. Saad, and C. Yang, "Approximating spectral densities of large matrices," *SIAM Review*, vol. 58, no. 1, pp. 34–65, 2016.

[59] P. Van Mieghem, *Graph Spectra for Complex Networks*. Cambridge University Press, 2011.

[60] T. Tao, *Topics in Random Matrix Theory*. American Mathematical Society, 2012.

[61] R. Silver and H. Röder, "Densities of states of mega-dimensional Hamiltonian matrices," *Int. J. Mod. Phys. C*, vol. 5, no. 4, pp. 735–753, 1994.

[62] R. Silver, H. Röder, A. Voter, and J. Kress, "Kernel polynomial approximations for densities of states and spectral functions," *J. Comput. Phys.*, vol. 124, no. 1, pp. 115–130, 1996.

[63] L.-W. Wang, "Calculating the density of states and optical-absorption spectra of large quantum systems by the plane-wave moments method," *Phy. Rev. B*, vol. 49, no. 15, p. 10154, 1994.

[64] M. F. Hutchinson, "A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines," *Commun. Stat. Simul. Comput.*, vol. 18, no. 3, pp. 1059–1076, 1989.

[65] F. N. Fritsch and R. E. Carlson, "Monotone piecewise cubic interpolation," *SIAM J. Numer. Anal.*, vol. 17, no. 2, pp. 238–246, Apr. 1980.

[66] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Review*, vol. 53, no. 2, pp. 217–288, 2011.

[67] J. Paratte and L. Martin, "Fast eigenspace approximation using random signals," *arXiv e-Prints*, 2016. [Online]. Available: http://arxiv.org/abs/1611.00938