

Scalable M -Channel Critically Sampled Filter Banks for Graph Signals

Shuni Li, Yan Jin, 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 fast versions of the proposed transform that scale 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 using 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 reconstruction error resulting from the numerical approximations, and the ability of the proposed transform to compress piecewise-smooth graph signals. The proposed filter bank also yields a fast, approximate graph Fourier transform with a coarse resolution in the spectral domain.

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]–[9], diffusion-based designs such as [10], [11], spectral domain designs such as [12]–[14], windowed graph Fourier transforms [15], and generalized filter banks, the focus of this paper.

Shuni Li is with the Department of Statistics, University of California, Berkeley, CA 94720, USA (email: shuni_li@berkeley.edu). Yan Jin is with the Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (email: yjin1@mit.edu). David I Shuman is with the Department of Mathematics, Statistics, and Computer Science, Macalester College, St. Paul, MN 55105, USA (email: 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 II-C.

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.

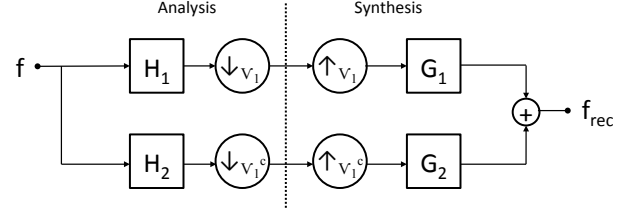


Fig. 1. Two channel critically sampled graph filter bank. Here, $\mathbf{H}_1$ is a lowpass graph spectral filter, and $\mathbf{H}_2$ is a highpass graph spectral filter.

The extension of the classical two channel critically sampled filter bank to the graph setting is first proposed in [16]. 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 sets 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, [17], [18] 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 [19] adds vertices and edges to the original graph to form an approximating bipartite graph. References [20], [21] 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 [22, Chapter 5.1]. These graphs have a circulant graph Laplacian and their eigenvectors are the columns of the discrete Fourier transform matrix. Any graph can be written as the sum of circulant graphs, and [23]–[25] 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 [26], [27] or pyramid transforms [28].

Our approach in this paper is to replace the synthesis filters with interpolation operators on each subband of the graph spectrum. While this idea is suggested independently in [29] for small graphs (say 5,000 or fewer vertices), we investigate it in more detail here, and extend it to large, sparse graphs.

We develop three variants of M -channel critically sampled filter banks (M -CSFB) for graph signals. The exact M -CSFB (initially presented in [30]) features dictionary atoms that are jointly localized in the vertex and graph spectral domains, and therefore can compactly represent graph signals

with localized singularities [12]. The fast M -CSFB and fast, signal-adapted M -CSFB transforms we propose here share the same general structure as the exact M -CSFB, but scale efficiently for large, sparse graphs; i.e., they do not require a full eigendecomposition of the graph Laplacian. Specific contributions of our work include:

- 1) A constructive method to partition the graph into uniqueness sets for any given partition of the spectrum. This is a key step in the exact M -CSFB (Section II).
- 2) A new filter bank design that is both adapted to an efficient estimate of the distribution of the graph Laplacian spectrum and more amenable to polynomial approximation (Section III).
- 3) A *scalable* method to sample and interpolate bandpass and highpass graph signals. While we leverage the recent flurry of work in sampling and reconstruction of graph signals [29]–[41], the prior literature focuses on methods that require a full eigendecomposition or assume the graph signals are smooth (lowpass). We use efficient convex optimization methods with a novel penalty term to perform the interpolation (Section III).
- 4) The idea to adapt both the non-uniform sampling weights and the number of samples allocated to each band to the actual signal being analyzed, in addition to the graph structure, because there is less benefit from taking samples in areas of the graph where the filtered signal does not have much energy. The fast, signal-adapted M -CSFB in Section IV is based on this concept.
- 5) Empirical explorations of the exact M -CSFB, fast M -CSFB, and signal-adapted fast M -CSFB transforms. In Section V, we investigate computation times, the reconstruction error resulting from the numerical approximations, tradeoffs involved in choosing the parameters, and applications such as compression and fast, approximate graph Fourier transforms.

II. M -CHANNEL CRITICALLY SAMPLED FILTER BANK

A. Notation

We consider graph signals $\mathbf{f} \in \mathbb{R}^N$ 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 $\mathcal{L}$ to be the unnormalized 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 $\mathcal{L} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^*$, where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues $\lambda_0, \lambda_1, \dots, \lambda_{N-1}$ of $\mathcal{L}$, and the columns $\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{N-1}$ of $\mathbf{U}$ are the associated eigenvectors of $\mathcal{L}$. The graph Fourier transform of a signal is $\hat{\mathbf{f}} = \mathbf{U}^* \mathbf{f}$, and $h(\mathcal{L})\mathbf{f} = \mathbf{U} h(\mathbf{\Lambda}) \mathbf{U}^* \mathbf{f}$ applies the filter $h : [0, \lambda_{\max}] \rightarrow \mathbb{R}$ to the graph signal $\mathbf{f}$. We let $\mathbf{U}_{\mathcal{R}}$ denote the submatrix formed by taking the columns of $\mathbf{U}$ associated with the Laplacian eigenvalues indexed by $\mathcal{R} \subseteq \{0, 1, \dots, N-1\}$, and $\mathbf{U}_{\mathcal{S}, \mathcal{R}}$ denote the submatrix formed by taking the rows of $\mathbf{U}_{\mathcal{R}}$ associated with the vertices indexed by the set $\mathcal{S} \subseteq \{1, 2, \dots, N\}$.

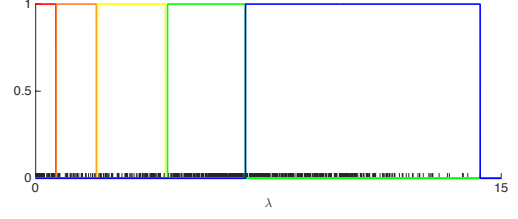


Fig. 2. Example ideal filter bank. The red, orange, yellow, green, and blue filters span 31, 31, 63, 125, and 250 graph Laplacian eigenvalues, respectively, on a 500 node sensor network with a maximum graph Laplacian eigenvalue of 14.3. The tick marks on the x-axis represent the locations of the graph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

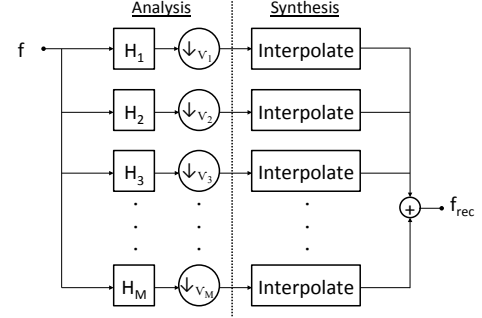


Fig. 3. The M -channel critically sampled filter bank architecture. The sets $\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_M$ form a partition of the set $\mathcal{V}$ of vertices, where each set $\mathcal{V}_m$ is a uniqueness set for graph signals supported on a different subband in the graph spectral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

B. Architecture

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

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

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

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

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

Definition 1 (Uniqueness set [31]). *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 ([33], [35]). *The set $\mathcal{S}$ of k vertices is a uniqueness set for $\text{col}(\mathbf{U}_{\mathcal{T}})$ if and only if the matrix whose columns are $\mathbf{u}_{\mathcal{T}_1}, \mathbf{u}_{\mathcal{T}_2}, \dots, \mathbf{u}_{\mathcal{T}_k}, \delta_{\mathcal{S}_1^c}, \delta_{\mathcal{S}_2^c}, \dots, \delta_{\mathcal{S}_{n-k}^c}$ is nonsingular, where $\mathbf{u}_{\mathcal{T}_j}$ is the j th column of $\mathbf{U}_{\mathcal{T}}$, and each $\delta_{\mathcal{S}_i^c}$ is a Kronecker delta centered on a vertex not included in $\mathcal{S}$.*

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 interpolate from the samples on $\mathcal{V}_m$ to $\text{col}(\mathbf{U}_{\mathcal{R}_m})$. Denoting the analysis coefficients (i.e., the filtered and downsampled signal) of the m^{th} branch by $\mathbf{y}_{\mathcal{V}_m}$, we have

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

If there is no error in the coefficients, then the reconstruction is perfect, because $\mathcal{V}_m$ is a uniqueness set for $\text{col}(\mathbf{U}_{\mathcal{R}_m})$, 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 with interpolation on the synthesis side. *Critically sampled* refers to the fact that the number of analysis coefficients is equal to the length of the original signal; that is, $\sum_{m=1}^M |\mathcal{V}_m| = N$.

C. 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.

1) $M = 2$ 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 $\mathcal{T} \subseteq \{0, 1, \dots, N-1\}$ denote a subset of the graph Laplacian eigenvalue indices, and let $\mathcal{T}^c = \{0, 1, \dots, N-1\} \setminus \mathcal{T}$. Then $\mathcal{S}^c$ is a uniqueness set for $\text{col}(\mathbf{U}_{\mathcal{T}^c})$ if and only if $\mathcal{S}$ is a uniqueness set for $\text{col}(\mathbf{U}_{\mathcal{T}})$.*

This fact follows from either the CS decomposition [42, Equation (32)] or the nullity theorem [43, Theorem 2.1]. We also provide a standalone proof in the Appendix 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. The Steinitz exchange lemma [44] 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 [35] to do so.

2) $M > 2$ 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 [45], and we later discovered the same method in [46], [47, Theorem 3.3].

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

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

Input $\mathbf{U}$, a partition $\{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_M\}$
 $\mathcal{S} \leftarrow \emptyset$
for $m = 1, 2, \dots, M$ **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. [47] 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 $\{\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_M\}$

Proof: First consider the case $M = 2$, and let $k = |\beta_1|$. Then by the generalized Laplace expansion [48],

$$\det(\mathbf{A}) = \sum_{\{\alpha_1 \subset \{1, 2, \dots, N\} : |\alpha_1| = k\}} \sigma_{\alpha_1, \beta_1} \det(\mathbf{A}_{\alpha_1, \beta_1}) \det(\mathbf{A}_{\alpha_1^c, \beta_2}), \quad (3)$$

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 (3) 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

$$\det(\mathbf{A}) = \sum_{\{\text{Partitions } \alpha \text{ of } \{1, 2, \dots, N\} : |\alpha_i| = |\beta_i| \forall i\}} \sigma_{\alpha} \prod_{i=1}^M \det(\mathbf{A}_{\alpha_i, \beta_i}), \quad (4)$$

where again $|\sigma_{\alpha}| = 1$, and one of the terms in the summation in (4) must be nonzero, yielding the desired partition. ■

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

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 [47], which gives an algorithm to find the desired row partition. We summarize this method in Algorithm

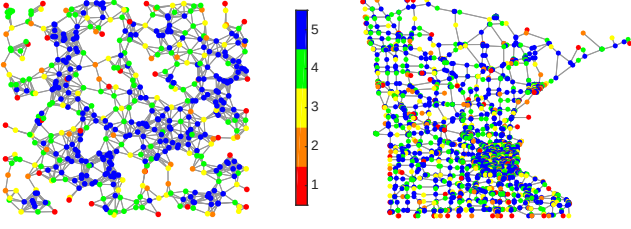


Fig. 4. Partitions of a 500 node random sensor network and the Minnesota road network [49] into uniqueness sets for five different spectral bands, with the indices increasing from lowpass bands (1) to highpass bands (5).

1, which takes in a partition $\{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_M\}$ of the spectral indices and constructs the partition $\{\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_M\}$ 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. The initial choices of γ_i in each loop play a significant role in the final partition. In the numerical experiments, we use the greedy algorithm in [35, Algorithm 1] to find an initial choice for γ_1 , permute the complement of γ_1 to the top, and then perform row reduction to find an initial choice for γ_2 .

D. Transform properties

1) *Dictionary atoms:* Let $\mathbf{M}_m \in \mathbb{R}^{|\mathcal{V}_m| \times N}$ be the down-sampling 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 $\mathcal{F} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ by $\mathcal{F}\mathbf{f} = \Phi^\top \mathbf{f}$, where the resulting dictionary is of the form $\Phi := [h_1(\mathcal{L})\mathbf{M}_1^\top \mid h_2(\mathcal{L})\mathbf{M}_2^\top \mid \dots \mid h_M(\mathcal{L})\mathbf{M}_M^\top]$. While the transform is not orthogonal, each atom (column of Φ) is orthogonal to all atoms concentrated on other spectral bands. This is because the atoms are projections of Kronecker deltas onto the orthogonal subspaces spanned by the Laplacian eigenvectors of each band. That is, each atom is of the form $h_m(\mathcal{L})\delta_i$, where vertex i is in $\mathcal{V}_m$. If $m \neq m'$, then the inner product of two atoms from different bands is given by

$$\begin{aligned} \langle h_m(\mathcal{L})\delta_i, h_{m'}(\mathcal{L})\delta_{i'} \rangle \\ = \delta_i^\top \mathbf{U} h_m(\Lambda) \mathbf{U}^* \mathbf{U} h_{m'}(\Lambda) \mathbf{U}^* \delta_{i'} = 0, \end{aligned} \quad (5)$$

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.

2) *Signals that are sparsely represented by the M-CSFB transform:* Globally smooth signals trivially lead to sparse analysis coefficients because the coefficients are only nonzero for the first set(s) of vertices in the partition. More generally, signals that are concentrated in the graph Fourier domain have sparse M-CSFB analysis coefficients, because the coefficients for any channel whose filter does not overlap with the support of the signal in the graph Fourier domain are all equal to zero.

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

Next, we empirically examine the joint localization of the dictionary atoms in the vertex and graph spectral domains,

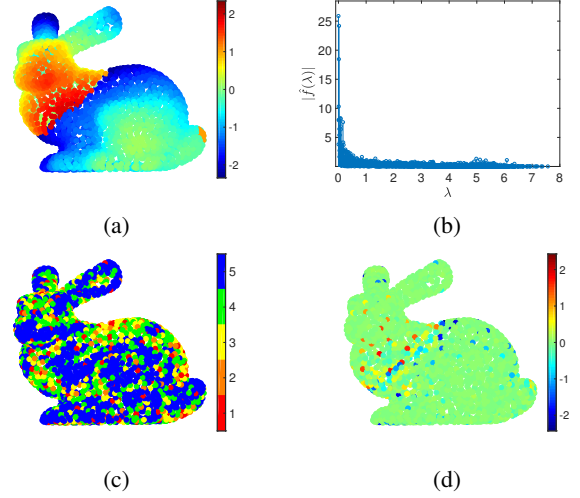


Fig. 5. (a)-(b) Piecewise smooth signal on the Stanford bunny graph [50] in the vertex and graph spectral domains, respectively. (c) Partition of the graph into uniqueness sets for five different spectral bands. (d) M -channel filter bank analysis coefficients of the signal shown in (a) and (b).

which is key for their ability to compactly represent localized phenomena (e.g., discontinuities, edges). On the Stanford bunny graph [50] with 2503 vertices, we partition the spectrum into five bands, and show the resulting partition into uniqueness sets in Fig. 5(c). The first row of Fig. 6 shows five example atoms whose energies are concentrated on different spectral bands. These atoms are also generally localized in the vertex domain, with the wavelets becoming more localized at higher scales, as expected. The second row of Fig. 6 shows the localization of the spectral content of all atoms in each band, with the averages represented by thick black lines.

We then apply the proposed transform to a piecewise-smooth graph signal $\mathbf{f}$ that is shown in the vertex domain in Fig. 5(a), and in the graph spectral domain in Fig. 5(b). The full set of analysis coefficients is shown in Fig. 5(d), and these are separated by band in the third row of Fig. 6. 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. 6 shows the interpolation of these coefficients onto the corresponding spectral bands. If we sum these reconstructions together, we recover the original signal in Fig. 5(a).

III. FAST M-CSFB TRANSFORM

In the numerical examples in the previous sections, we have computed a full eigendecomposition of the graph Laplacian and 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 this section, 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 $h_m(\mathcal{L})\mathbf{f}$, a function of sparse matrix times a vector, include approximating the function $h_m(\cdot)$ by a polynomial (e.g., via a truncated

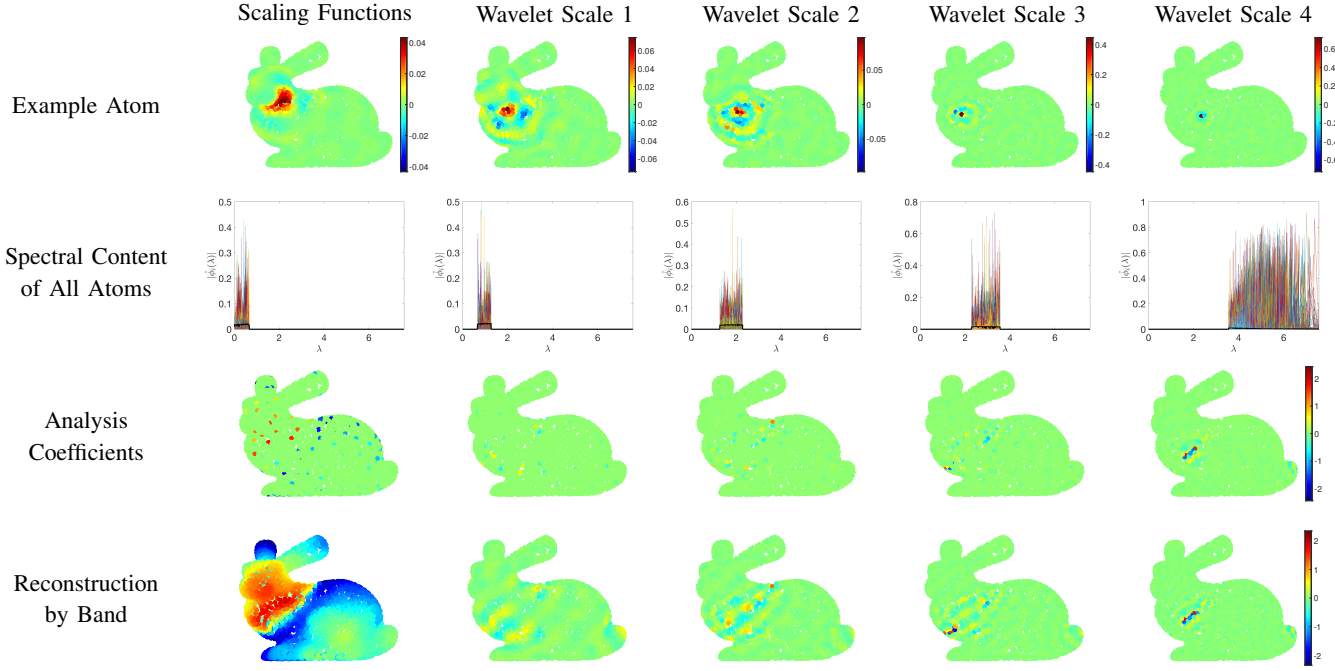


Fig. 6. M -channel filter bank example. The first row shows example atoms in the vertex domain. The second row shows all atoms in the spectral domain, with an average of the atoms in each band shown by the thick black lines. The third row shows the analysis coefficients of Fig. 5(d) by band, and the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

Chebyshev or Legendre expansion), approximating $h_m(\cdot)$ by a rational function, Krylov space methods (Lanczos in our case of a symmetric matrix $\mathcal{L}$), and the matrix version of the Cauchy integral theorem (see, e.g., [51]-[54] and references therein). The first three of these methods have been examined in graph signal processing settings [12], [36], [55]-[58]. Here, to approximate the analysis side filters, we focus on order K Chebyshev polynomial approximations of the form

$$\tilde{h}(\mathcal{L})\mathbf{f} := \sum_{k=0}^K \alpha_k \bar{T}_k(\mathcal{L})\mathbf{f}. \quad (6)$$

In (6), $\bar{T}_k(\cdot)$ are Chebyshev polynomials shifted to the interval $[0, \lambda_{\max}]$. Thus, $\bar{T}_0(\mathcal{L})\mathbf{f} = \mathbf{f}$, $\bar{T}_1(\mathcal{L})\mathbf{f} = \frac{2}{\lambda_{\max}}\mathcal{L}\mathbf{f} - \mathbf{f}$, and for $k \geq 2$, by the three term recurrence relation of Chebyshev polynomials, we have

$$\bar{T}_k(\mathcal{L})\mathbf{f} = \frac{4}{\lambda_{\max}} \left(\mathcal{L} - \frac{\lambda_{\max}}{2} \mathbf{I} \right) \bar{T}_{k-1}(\mathcal{L})\mathbf{f} - \bar{T}_{k-2}(\mathcal{L})\mathbf{f}. \quad (7)$$

The coefficients in (6) are often taken to be $\alpha_0 = \frac{1}{2}c_0$ and $\alpha_k = c_k$ for $k = 1, 2, \dots, K$, where $\{c_k\}_{k=0, \dots, K}$ are the truncated Chebyshev expansion coefficients

$$c_k := \langle h, \bar{T}_k \rangle = \frac{2}{\pi} \int_0^\pi \cos(k\phi) h\left(\frac{\lambda_{\max}}{2}(\cos(\phi) + 1)\right) d\phi. \quad (8)$$

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 (5). In an attempt to at least preserve *near* orthogonality across bands, we therefore use the Jackson-Chebyshev polynomial approximations from [59], [60] that damp the Gibbs oscillations

appearing in Chebyshev expansions. With the damping,

$$\alpha_0 = \frac{1}{2}c_0 \text{ and } \alpha_k = \gamma_{k,K}c_k \text{ for } k = 1, 2, \dots, K, \quad (9)$$

where, as presented in [59],

$$\gamma_{k,K} = \frac{(1 - \frac{k}{K+2}) \sin(\frac{\pi}{K+2}) \cos(\frac{k\pi}{K+2}) + \frac{1}{K+2} \cos(\frac{\pi}{K+2}) \sin(\frac{k\pi}{K+2})}{\sin(\frac{\pi}{K+2})}. \quad (10)$$

Fig. 7 shows the Jackson-Chebyshev polynomial approximations for ideal bandpass filters of two different graphs.¹

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:

$$\begin{aligned} \|\tilde{h}_m(\mathcal{L}) - h_m(\mathcal{L})\|_2 &= \max_{\ell \in \{0, 1, \dots, N-1\}} |\tilde{h}_m(\lambda_\ell) - h_m(\lambda_\ell)| \\ &\leq \max_{\lambda \in [0, \lambda_{\max}]} |\tilde{h}_m(\lambda) - h_m(\lambda)|. \end{aligned} \quad (11)$$

While approximation theory often aims to minimize the upper bound in (11), only the errors exactly at the graph Laplacian eigenvalues affect the overall approximation error $\|\tilde{h}_m(\mathcal{L}) - h_m(\mathcal{L})\|_2$. Since, as seen in Fig. 7, 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 $\{\tau_m\}_{m=1, \dots, M-1}$ 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

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

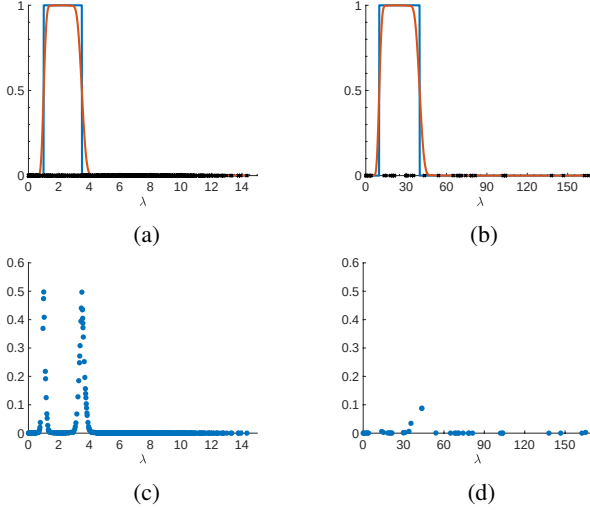


Fig. 7. Degree 80 Jackson-Chebyshev polynomial approximations for ideal bandpass filters on the (a) 500 vertex random sensor network of Fig. 4 and (b) Andrianov net25 graph from [61] with 9,520 vertices. In (c) and (d), we show the errors $|\tilde{h}(\lambda_\ell) - h(\lambda_\ell)|$ at each of the Laplacian eigenvalues of the corresponding graphs in (a) and (b).

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 to have the endpoints close to fewer eigenvalues of $\mathcal{L}$.

1) *Estimating the spectral density*: Lin et al. [62] provide an excellent overview of methods to approximate the *spectral density function* [63, Chapter 6] (also called the *Density of States* or *empirical spectral distribution* [64, Chapter 2.4]) of a matrix, which in our context for the graph Laplacian $\mathcal{L}$ 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 [65]–[67] described in [62] to estimate the *cumulative spectral density function* or *empirical spectral cumulative distribution*

$$P_\lambda(z) := \frac{1}{N} \sum_{\ell=0}^{N-1} \mathbb{1}_{\{\lambda_\ell \leq z\}}. \quad (12)$$

The procedure 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 [68] to estimate η_i , the number of eigenvalues less than or equal to ξ_i . Defining the Heaviside function $\Theta_{\xi_i}(\lambda) := \mathbb{1}_{\{\lambda \leq \xi_i\}}$, we have

$$\eta_i = \text{tr}(\Theta_{\xi_i}(\mathcal{L})) = \mathbb{E}[\mathbf{x}^\top \Theta_{\xi_i}(\mathcal{L}) \mathbf{x}] \quad (13)$$

$$\approx \frac{1}{J} \sum_{j=1}^J \mathbf{x}^{(j)\top} \Theta_{\xi_i}(\mathcal{L}) \mathbf{x}^{(j)} \quad (14)$$

$$\approx \frac{1}{J} \sum_{j=1}^J \mathbf{x}^{(j)\top} \tilde{\Theta}_{\xi_i}(\mathcal{L}) \mathbf{x}^{(j)}. \quad (15)$$

In (13), $\mathbf{x}$ is a random vector with each component having an independent and identical standard normal distribution. Each vector $\mathbf{x}^{(j)}$ in (14) is chosen according to this same distribution, and in our experiments, we take the default number of vectors to be $J = 30$. In (15), $\tilde{\Theta}_{\xi_i}$ is the Jackson-Chebyshev approximation to Θ_{ξ_i} discussed in Section III-A. If we place the J random vectors into the columns of an $N \times J$

Algorithm 2 Spectral density approximation

Input graph $\mathcal{G}$, estimate for $\lambda_{\max}$, degree K , number of random vectors J
Generate an $N \times J$ matrix $\mathbf{X}$ whose columns are i.i.d. standard normal random vectors
Compute $\{\tilde{T}_k(\mathcal{L})\mathbf{X}\}_{k=0,1,\dots,K}$ via (7)
Choose T linearly spaced points $\{\xi_i\}_{i=1,\dots,T}$ between 0 and $\lambda_{\max}$
for $i = 1, 2, \dots, T$ **do**
 Approximate η_i via (15) and (16)
end for
Estimate the spectral density function $\tilde{P}_\lambda$ by performing monotonic cubic interpolation on the set of points $\{\xi_i, \frac{\eta_i}{N}\}$
Output $\mathbf{X}$, $\{\tilde{T}_k(\mathcal{L})\mathbf{X}\}_{k=0,1,\dots,K}$, and $\tilde{P}_\lambda$

matrix $\mathbf{X}$, the computational cost of estimating the spectral distribution is dominated by computing

$$\tilde{\Theta}_{\xi_i}(\mathcal{L})\mathbf{X} = \sum_{k=0}^K \alpha_k \tilde{T}_k(\mathcal{L})\mathbf{X} \quad (16)$$

for each ξ_i . Yet, we only need to compute $\{\tilde{T}_k(\mathcal{L})\mathbf{X}\}_{k=0,1,\dots,K}$ 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 [13], 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 [69] on the series of points $\{(\xi_i, \frac{\eta_i}{N})\}_{i=1,2,\dots,T}$. We denote the result as $\tilde{P}_\lambda(\cdot)$. Algorithm 2 summarizes these computations.

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 $\tau_m = \frac{m}{M} \lambda_{\max}$. Fig. 8(a)–(b) show a spectrum-adapted, logarithmically spaced choice with $\tau_m = \tilde{P}_\lambda^{-1}(\frac{1}{2}^{M-m})$ for $m = 1, 2, \dots, M$, 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 $m = 1, 2, \dots, M-1$ and some $\Delta > 0$, we let the final endpoint be

$$\tau_m^* = \underset{\tau \in \mathcal{I}_m}{\operatorname{argmin}} \left\{ \frac{\tilde{P}_\lambda(\tau + \Delta) - \tilde{P}_\lambda(\tau - \Delta)}{2\Delta} \right\}, \quad (17)$$

where $\mathcal{I}_m$ is an interval around the initial choice of τ_m . Fig. 8(c) shows the objective function in (17), along with the initial band ends, search intervals, and adjusted band ends. Comparing Fig. 8(b) and Fig. 8(d), the band end adjustments lead to fewer eigenvalues falling close to the filter borders, reducing the error incurred by the polynomial approximation

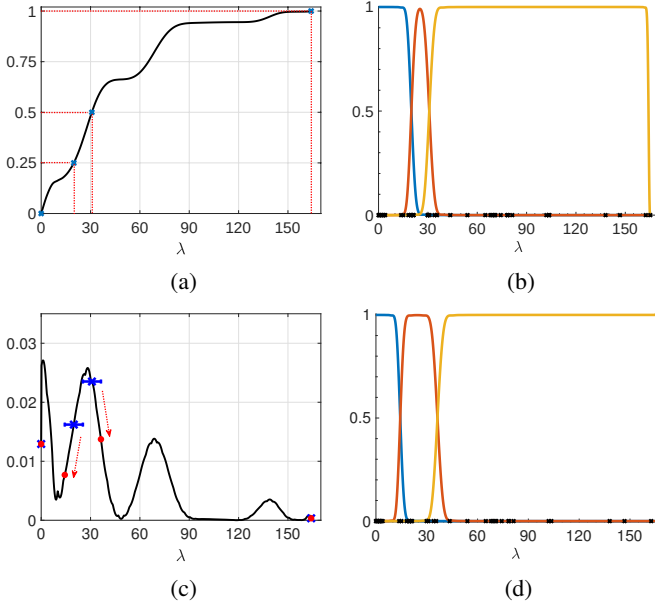


Fig. 8. (a) The approximate cumulative spectral density function, $\tilde{P}_\lambda(\cdot)$, for the net25 graph described in Fig. 7. The blue X marks correspond to the initial choice of band endpoints computed by taking the inverse of logarithmically spaced points on the vertical axis. (b) The degree 80 Jackson-Chebyshev approximations to the ideal filters defined by the initial choice of band ends from (a). (c) The objective function of (17) (a discrete approximation of the spectral density function $p_\lambda(\cdot)$) with $\Delta = .1$. 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.

process. Algorithm 3 summarizes the filter bank design in the case of spectrum-adapted and logarithmically spaced filters.

C. Non-uniform random sampling distribution

The partitioning of the vertices into uniqueness sets described in Algorithm 1 requires a full eigendecomposition of the graph Laplacian to compute the matrix $\mathbf{U}$. Two broad approaches to more efficient sampling have recently been investigated: greedy methods [29], [33], [38], [40] and random sampling methods [35], [36], [39], which have close connections to *leverage score* sampling in the statistics and numerical linear algebra literature (see, e.g., [70]–[72]). Reference [40] has a nice review of the computational complexities of the various greedy routines for identifying uniqueness sets. Most of these are designed specifically for lowpass signals.

We adapt the non-uniform random sampling method of [36], which scales more efficiently than greedy methods. For the m^{th} band, we identify the downsampling set $\mathcal{V}_m$ by sampling the vertices $\mathcal{V}$ without replacement according to a discrete probability distribution ω_m . To minimize the graph weighted coherence, it is ideal to take $\omega_m(i) \propto \|\mathbf{U}_{\mathcal{R}_m}^\top \delta_i\|_2^2$ [36]; however, we do not have access to $\mathbf{U}_{\mathcal{R}_m}$. Instead, we take

$$\omega_m(i) \propto \|(\tilde{h}_m(\mathcal{L})\mathbf{X})^\top \delta_i\|_2^2, \quad (18)$$

which [36] shows is an unbiased estimator of $\|\mathbf{U}_{\mathcal{R}_m}^\top \delta_i\|_2^2$ when $\mathbf{X}$ is the random matrix from (16). Since we already compute and store the series of matrices $\{\tilde{T}_k(\mathcal{L})\mathbf{X}\}$ for the spectral density estimation of Section III-B1, we just need to

Algorithm 3 Spectrum-adapted and logarithmically spaced filter bank design

Input Estimate for $\lambda_{\max}$, approximate spectral density $\tilde{P}_\lambda$, number of bands M , $\Delta > 0$, degree K

for $m = 1, 2, \dots, M$ **do**

 Compute the initial band end: $\tau_m = \tilde{P}_\lambda^{-1}(\frac{1}{2}^{M-m})$

end for

Set $\tau_0^* = \tau_0 = 0$, $\tau_M^* = \tau_M = \lambda_{\max}$,

for $m = 1, \dots, M-1$ **do**

 Set the search radius:

$$r = \min\left\{\frac{\tau_m - \tau_{m-1}}{2}, \frac{\tau_{m+1} - \tau_m}{2}\right\}$$

 Set the search interval: $\mathcal{I}_m = [\tau_m - r, \tau_m + r]$

 Update the band ends:

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

end for

for $m = 1, 2, \dots, M$ **do**

 Construct the ideal filter $h_m(\lambda)$ according to (1) using τ_{m-1}^* and τ_m^*

 Construct the polynomial filter approximation $\tilde{h}_m(\lambda)$ and the corresponding Jackson-Chebyshev polynomial coefficients $\alpha_{m,k}$ via (8)–(10)

end for

Output Degree K Jackson-Chebyshev polynomial filters $\{\tilde{h}_1(\lambda), \tilde{h}_2(\lambda), \dots, \tilde{h}_M(\lambda)\}$, and associated coefficients $\{\alpha_{m,k}\}_{m=1,2,\dots,M; k=0,1,\dots,K}$

compute the polynomial approximation coefficients $\{\alpha_{m,k}\}$ in (16) for $h_m(\lambda)$ in order to compute $\tilde{h}_m(\mathcal{L})\mathbf{X}$.

Intuitively, the sampling distribution approximates the energy of the selected eigenvectors concentrated on each vertex. In the extreme case that the selected eigenvectors are completely concentrated on a single vertex or small neighborhood of vertices, sampling signal values outside of this set provides no additional information, justifying the zero weight in the sampling distribution. For eigenvectors whose energy is equally spread across the graph, this results in uniform sampling. In particular, for any walk-regular graph, a class that includes vertex-transitive graphs, which in turn include shift-invariant graphs such as the cycle graph, $\|\mathbf{U}_{\mathcal{R}}^\top \delta_i\|_2^2$ is constant across vertices i for any choice of eigenvectors $\mathcal{R}$ [73, Corollary 3.2], resulting in uniform random sampling for all bands. As discussed in [36, Section 5.1.2], non-uniform sampling is particularly beneficial for bands with localized eigenvectors, which most commonly occur at the middle and upper ends of the spectrum. For the low end of the spectrum with smooth eigenvectors, the intuition is that it is easier to interpolate missing values in highly connected regions of the graph, and therefore there are slightly higher weights on the less connected vertices (e.g., near the boundaries in Fig. 11(g)).

D. Number of samples

One option to ensure critical sampling is to choose the number of samples for each band according to the initial filter

Algorithm 4 Construct the downsampling sets

Input graph $\mathcal{G}$, $\mathbf{X}$, $\{\bar{T}_k(\mathcal{L})\mathbf{X}\}_{k=0,1,\dots,K}$, Jackson-Chebyshev coefficients $\{\alpha_{m,k}\}_{m=1,2,\dots,M;k=0,1,\dots,K}$ for the polynomial filters $\{h_1(\lambda), h_2(\lambda), \dots, h_M(\lambda)\}$, signal $\mathbf{f}$ (optional)

for $m = 1, 2, \dots, M$ **do**

 Compute $\tilde{h}_m(\mathcal{L})\mathbf{X} = \sum_{k=0}^K \alpha_{m,k} \bar{T}_k(\mathcal{L})\mathbf{X}$

 Set the weight for each vertex $i \in \mathcal{V}$:

$$\omega_m(i) = \|(\tilde{h}_m(\mathcal{L})\mathbf{X})^\top \delta_i\|_2^2$$

if signal-adapted weights **then**

 Compute $\tilde{h}_m(\mathcal{L})\mathbf{f}$ via (6) with the same $\{\alpha_{m,k}\}$

 Adapt the weights:

$$\omega_m(i) = \omega_m(i) \cdot \log(1 + |(\tilde{h}_m(\mathcal{L})\mathbf{f})(i)|)$$

end if

 Normalize the weights: $\omega_m(i) = \frac{\omega_m(i)}{\sum_{i=1}^N \omega_m(i)}$

 Set the initial number of samples based on (15):

$$n_m = \frac{1}{J} \text{Trace}(\mathbf{X}^\top \tilde{h}_m(\mathcal{L})\mathbf{X})$$

end for

if signal-adapted number of samples **then**

for $m = 1, 2, \dots, M$ **do**

 Set $n_m = n_m \cdot \log(1 + \|\tilde{h}_m(\mathcal{L})\mathbf{f}\|)$

end for

end if

Compute total initial number of samples: $N_0 = \sum_m n_m$

for $m = 1, 2, \dots, M$ **do**

 Normalize the number of total samples:

$$n_m = \text{round}\left(\frac{n_m}{N_0} N_T\right),$$

where N_T is the target number of samples (e.g., $N_T = N$ for critical sampling)

end for

Adjust to meet target number of samples:

if $\sum_m n_m > N_T$ **then**

 Set $n_M = n_M - (\sum_m n_m - N_T)$

else if $\sum_m n_m < N_T$ **then**

 Set $n_1 = n_1 + (N_T - \sum_m n_m)$

end if

for $m = 1, 2, \dots, M$ **do**

 Choose the downsampling set $\mathcal{V}_m$ by randomly sampling n_m vertices according to the distribution ω_m

end for

Output downsampling sets $\{\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_M\}$, sampling distributions $\{\omega_1, \omega_2, \dots, \omega_M\}$

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 III-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 $N \cdot (\tilde{P}_\lambda(\tau_m) - \tilde{P}_\lambda(\tau_{m-1}))$.

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 (15), except using the bandpass filter $h_m(\lambda)$ from (1). We already compute $\tilde{h}_m(\mathcal{L})\mathbf{X}$ to calculate the sampling distribution in (18). We can substitute the columns $\tilde{h}_m(\mathcal{L})\mathbf{x}^{(j)}$ of this matrix into (15) 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 [59].

We make small adjustments to ensure the total number of samples is equal to some target N_T . In our experiments, we take $N_T = N$ to ensure critical sampling. Our default is to add samples to the lowest band if the normalized total is below N_T , and remove samples from the highest band if the normalized total is above N_T . Algorithm 4 summarizes the proposed method to choose the downsampling sets.

Note that $\dim(\text{col}(\tilde{h}_m(\mathcal{L}))) \geq \dim(\text{col}(h_m(\mathcal{L})))$, 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 $\text{col}(\tilde{h}_m(\mathcal{L}))$, we need more samples than the number of eigenvalues in the support of $h_m(\cdot)$. In Section V, we explore how the reconstruction error is reduced as we increase the number of samples in each band.

E. Interpolation

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

$$\mathbf{f}_{m, \text{rec}} = \mathbf{U}_{\mathcal{R}_m} (\mathbf{U}_{\mathcal{V}_m, \mathcal{R}_m}^\top \mathbf{U}_{\mathcal{V}_m, \mathcal{R}_m})^{-1} \mathbf{U}_{\mathcal{V}_m, \mathcal{R}_m}^\top \mathbf{y}_{\mathcal{V}_m} \quad (19)$$

also requires $\mathbf{U}_{\mathcal{R}_m}$. One option explored in [74], [75] is to leverage $\{\bar{T}_k(\mathcal{L})\mathbf{X}\}$ 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 [36] to efficiently reconstruct lowpass signals is to relax the optimization problem

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

to

$$\min_{\mathbf{z} \in \mathbb{R}^N} \left\{ \mathbf{z}^\top \varphi_m(\mathcal{L}) \mathbf{z} + \kappa \|\Omega_{m, \mathcal{V}_m}^{-\frac{1}{2}} (\mathbf{M}_m \mathbf{z} - \mathbf{y}_{\mathcal{V}_m})\|_2^2 \right\}, \quad (20)$$

where $\Omega_{m, \mathcal{V}_m}$ 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, and $\kappa > 0$ is a parameter to trade off the two optimization objectives. The regularization term $\mathbf{z}^\top \varphi_m(\mathcal{L}) \mathbf{z}$ in (20) penalizes reconstructions with support outside of the desired spectral band. For lowpass signals, Puy et al. [36] take the penalty function $\varphi_m(\lambda)$ to be a nonnegative, nondecreasing polynomial, such as λ^l , with 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 $\varphi_m(\mathcal{L})$

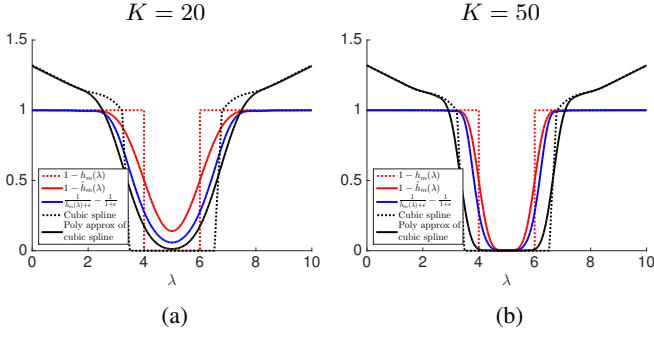


Fig. 9. Example penalty filters φ_m for the regularization term in (20), with $\epsilon = \frac{\sqrt{5}-1}{2}$. Here, $\tilde{h}_m$ is a Jackson-Chebyshev polynomial approximation of h_m of degree 20 and 50 in (a) and (b), respectively.

is positive semi-definite and the optimization problem (20) 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) the polynomial filter $\varphi_m(\lambda) = 1 - \tilde{h}(\lambda)$; (ii) the rational filter $\varphi_m(\lambda) = \frac{1}{\tilde{h}(\lambda) + \epsilon} - \frac{1}{1 + \epsilon}$; and (iii) a polynomial approximation of a penalty function constructed as a piecewise cubic spline, an approach explored in [76]. See Fig. 9 for example graphs of these penalty functions.

From the first-order optimality conditions, the solution to (20) is the solution to the linear system of equations

$$\left(\kappa M_m^\top \Omega_{m, \mathcal{V}_m}^{-1} M_m + \varphi_m(\mathcal{L}) \right) \mathbf{z} = \kappa M_m^\top \Omega_{m, \mathcal{V}_m}^{-1} \mathbf{y}_{\mathcal{V}_m}, \quad (21)$$

which can be solved, for example, with the preconditioned conjugate gradient method. For the preconditioner, we use a diagonal matrix whose i^{th} element is equal to 1 if $i \notin \mathcal{V}_m$ and $1 + \frac{\kappa}{\omega_m(i)}$ if $i \in \mathcal{V}_m$, which serves as an approximation to the matrix $\kappa M_m^\top \Omega_{m, \mathcal{V}_m}^{-1} M_m + \varphi_m(\mathcal{L})$ in (21).

F. Summary and properties of the fast M-CSFB transform

In summary, as shown in the flow chart in Fig. 10, the set up for the fast M-CSFB consists of approximating the spectral density of the graph Laplacian, designing the filter bank via Algorithm 3, and choosing the downsampling sets via Algorithm 4. To analyze a signal, we apply each of the M Jackson-Chebyshev polynomial filters output from Algorithm 3 to the signal, and then downsample on the corresponding set of vertices output from Algorithm 4. To synthesize a signal from its transform coefficients, we solve (21) for each band and sum the results. The complexity of the set up is dominated by the computation of $\{\tilde{T}_k(\mathcal{L})\mathbf{X}\}_{k=0,1,\dots,K}$ via (7), which has computational complexity $\mathcal{O}(JK|\mathcal{E}|)$. The computational complexity of the analysis is $\mathcal{O}(K|\mathcal{E}|)$. So, if N is large, the number of random vectors J and degree of polynomial approximation K are small compared to N , and the graph is sparse ($|\mathcal{E}|$ is roughly a small constant times N), the setup and analysis scale linearly with the number of vertices. If each φ_m is taken to be an order K polynomial and the conjugate gradient is run for at most I iterations, the bottleneck computation of the synthesis has complexity $\mathcal{O}(IMK|\mathcal{E}|)$. In practice, the required number of iterations and corresponding computation time depend on the conditioning of the matrix on the left-hand side of (21) and the choice of preconditioner.

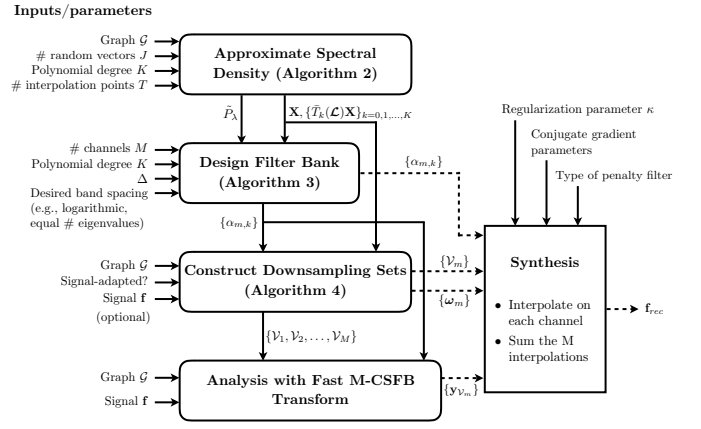


Fig. 10. Flow chart of the set up, analysis, and synthesis for the fast M-CSFB transform. The only differences for the signal-adapted transform are in the construction of the downsampling sets.

When the degree of approximation K is small, the energies of the fast, approximate transform atoms, which are of the form $\tilde{h}_m(\mathcal{L})\delta_i$, may be slightly more spread in the spectral domain than those of the atoms of the form $h_m(\mathcal{L})\delta_i$, due to the polynomial approximation of the ideal filter. However, the energy is also guaranteed to be supported completely within a radius of K hops from the center vertex i [12], [15]. Thus, we have better control over the spread in the vertex domain, which, as we saw in the scale 3 wavelet atom in Fig. 6, may be larger with the ideal filters.

IV. SIGNAL-ADAPTED FAST M-CSFB TRANSFORM

Just as it is helpful for interpolation to sample more signal values at vertices where the energy of the selected eigenvectors is concentrated, it is also helpful to sample more values where the energy of the filtered signals is concentrated. This motivates three adaptations to the fast M-CSFB transform.

First, we subtract the mean of the signal (i.e., let $\mathbf{f} = \mathbf{f} - \frac{1}{N}\mathbf{f}^\top \mathbf{1}$) before sending it into the filter bank, and then add this constant back to every vertex when summing the interpolations from the M channels. To ensure critical sampling, we only allow $N - 1$ total samples in addition to this mean.

Second, we adapt the sampling weights by setting $\omega_m(i) = \omega_m(i) \cdot \log(1 + |(\tilde{h}_m(\mathcal{L})\mathbf{f})(i)|)$. Thus, if a filtered signal on a given band is concentrated on a certain region of the graph, the sampling weights are concentrated on the intersection of that region and the set of vertices where the energies of the selected eigenvectors are concentrated.

Third, beyond the distribution of samples within each band, we need to decide how many samples to allocate to each band. In the exact computation (small graph) case, allocating the samples according to the number of eigenvalues contained in the disjoint bands ensures perfect reconstruction. However, with approximate computations, it is beneficial to the overall reconstruction error to do a better job of interpolation on the bands whose filtered signals have the most energy. We set the initial number of samples by multiplying the estimate of the number of eigenvalues in the band with $\log(1 + \|\tilde{h}_m(\mathcal{L})\mathbf{f}\|)$, as shown in Algorithm 4. In the extreme case of a filtered signal with no energy, this choice leads to zero measurements and a reconstruction of the all zero vector.

	Sensor Network $N = 500$ $ \mathcal{E} = 2,050$			Bunny $N = 2,503$ $ \mathcal{E} = 13,726$			Andrianov net25 Graph $N = 9,520$ $ \mathcal{E} = 195,841$			Community Graph $N = 25,000$ $ \mathcal{E} = 480,459$			Temperatures $N = 469,404$ $ \mathcal{E} = 1,865,415$		
	Anal. Time	Synth. Time	Rec. NMSE	Anal. Time	Synth. Time	Rec. NMSE	Anal. Time	Synth. Time	Rec. NMSE	Anal. Time	Synth. Time	Rec. NMSE	Anal. Time	Synth. Time	Rec. NMSE
Graph Fourier Transform	0.1	0.01	5.4e-30	9.8	0.02	2.5e-29	295.7	0.08	1.4e-28	8544.8	0.6	4.5e-28	NA	NA	NA
Exact M -CSFB	2.2	0.06	7.8e-30	380.4	0.1	7.8e-23	NA	NA	NA	NA	NA	NA	NA	NA	NA
Diffusion Wavelets [10]	8.5	0.03	1.2e-30	313.9	0.02	1.2e-29	14354	0.3	1.0e-26	NA	NA	NA	NA	NA	NA
Graph-QMF [17]	0.6	0.1	5.4e-8	4.9	3.4	3.2e-8	38.4	21.0	3.3e-9	1062.7	978.0	6.0e-8	NA	NA	NA
Fast M -CSFB (Scenario A: faster)	0.6	0.5	6.8e-2	0.8	0.9	8.2e-2	2.3	3.1	1.6e-1	2.8	12.4	2.2e-1	55.1	94.5	1.4e-2
Fast M -CSFB (Scenario B: more accurate)	0.7	1.0	9.2e-2	0.9	3.7	3.3e-2	1.4	12.1	1.4e-1	4.4	71.7	1.5e-1	91.6	874.3	7.0e-3
Signal-Adapted Fast M -CSFB (Scenario A: faster)	0.7	0.5	3.8e-2	0.8	0.9	3.4e-2	0.8	2.2	6.7e-2	2.8	9.9	1.2e-1	47.6	98.4	1.7e-3
Signal-Adapted Fast M -CSFB (Scenario B: more accurate)	0.7	1.1	2.4e-2	0.9	3.6	1.2e-2	1.3	9.7	7.7e-2	4.4	71.1	7.9e-2	81.2	976.0	6.6e-4

TABLE I
COMPARISON OF COMPUTATION TIMES (SECONDS) AND RECONSTRUCTION ERRORS

Note that when analyzing a single signal, these adaptations do not add significantly to the computational complexity of the transform. However, if we are repeating the transform on many different signals residing on the same graph, we do need to rerun the random selection of vertices for each signal.

V. NUMERICAL EXPERIMENTS

A. Scalability

The one indisputable advantage of the proposed fast M -CSFB transforms over other critically sampled transforms for graph signals with jointly localized atoms is their scalability to sparse graphs with a large number of vertices. In Table I, we compare the computation times of the proposed transform to those of the exact graph Fourier transform (i.e., a full diagonalization of $\mathcal{L}$); diffusion wavelets [10] with five scales (one scaling and four wavelets) and a precision of $\epsilon = 1e-4$; and a graph quadrature mirror filter (QMF) bank [17] with polynomial approximation order of $K = 50$. For the fast M -CSFB (original and signal adapted versions), we consider two scenarios. Scenario A is faster, but less accurate, with $K = 25$, a conjugate gradient (CG) tolerance of $1e-8$, and a maximum of 100 CG iterations. Scenario B is slower, but more accurate, with $K = 50$, a CG tolerance of $1e-10$, and a maximum of 250 iterations. For all fast M -CSFB cases, we let $M = 5$, $J = 30$ and $\kappa = 1$, and subtract out the mean of the signal before applying the filter bank. For larger graphs, calculations such as a full diagonalization are either not possible due to memory limits or would take in excess of a day to compute. We denote these by NA. Note that the graph-QMF transform is slower due to two additional bottlenecks. First, it requires a graph coloring via algorithms with complexities of $\mathcal{O}(N^3)$ or $\mathcal{O}(N^4)$ [77]. Second, whereas for the M -CSFB we can reuse the single sequence of vectors $\{\tilde{T}_k(\mathcal{L})\mathbf{f}\}$ in computing the filtered signal $\tilde{h}_m(\mathcal{L})\mathbf{f}$ on each channel, the iterative nature of the graph-QMF filter bank leads to different input signals at each level, resulting in a complexity increase from $\mathcal{O}(K|\mathcal{E}|)$

(assuming M is on the order of the average degree of the graph) to $\mathcal{O}(K|\mathcal{E}|(2^L - 1))$, where L is the number of bipartite subgraphs in the graph-QMF transform. The second bottleneck is more significant in the computation times for smaller graphs, while the first becomes prohibitive for extremely large graphs.

We apply all transforms to the signals shown above for the sensor network and bunny graph, to Gaussian random vectors with independent entries for the net25 and community (100 communities) graphs, and to a temperature signal discussed in detail in the next subsection. Note that in all of these examples, we are simply performing analysis followed by synthesis, without doing any compression or other adjustments to the analysis coefficients (compression examples are included in Section V-D). Of particular note in Table I is that the fast M -CSFB analysis times are under a minute for a graph with almost a half of a million vertices and 2 million edges.

B. Illustrative Example: Temperatures

The last column of Table I is based on the average temperatures for March 2018, taken from the Gridded 5km GHCN-Daily Temperature and Precipitation Dataset (nClimGrid) [78], [79] of the United States National Oceanic and Atmospheric Administration (NOAA). The measurements are on a grid with spacing of $\frac{1}{24}$ of a degree for both latitude and longitude. We form an unweighted graph by connecting each measurement location to its eight neighbors on the grid, if they contain measurements, as shown in Fig. 11(a)-(b). We eliminate isolated vertices and small components (e.g., islands), yielding a connected graph with 469,404 vertices (measurement locations). Fig. 11(c)-(e) show the estimated spectral distribution, the approximate filter bank with five bands and degree 50 polynomial approximations, and the temperatures.

An intuitive explanation of the distribution for the second band, shown in Fig. 11(g), is that we want to sample with higher probability near the edges, as there are fewer neighbors from whom to interpolate the local average value. The 27,021 vertices selected for $\mathcal{V}_2$, shown in Fig. 11(i), are spread across

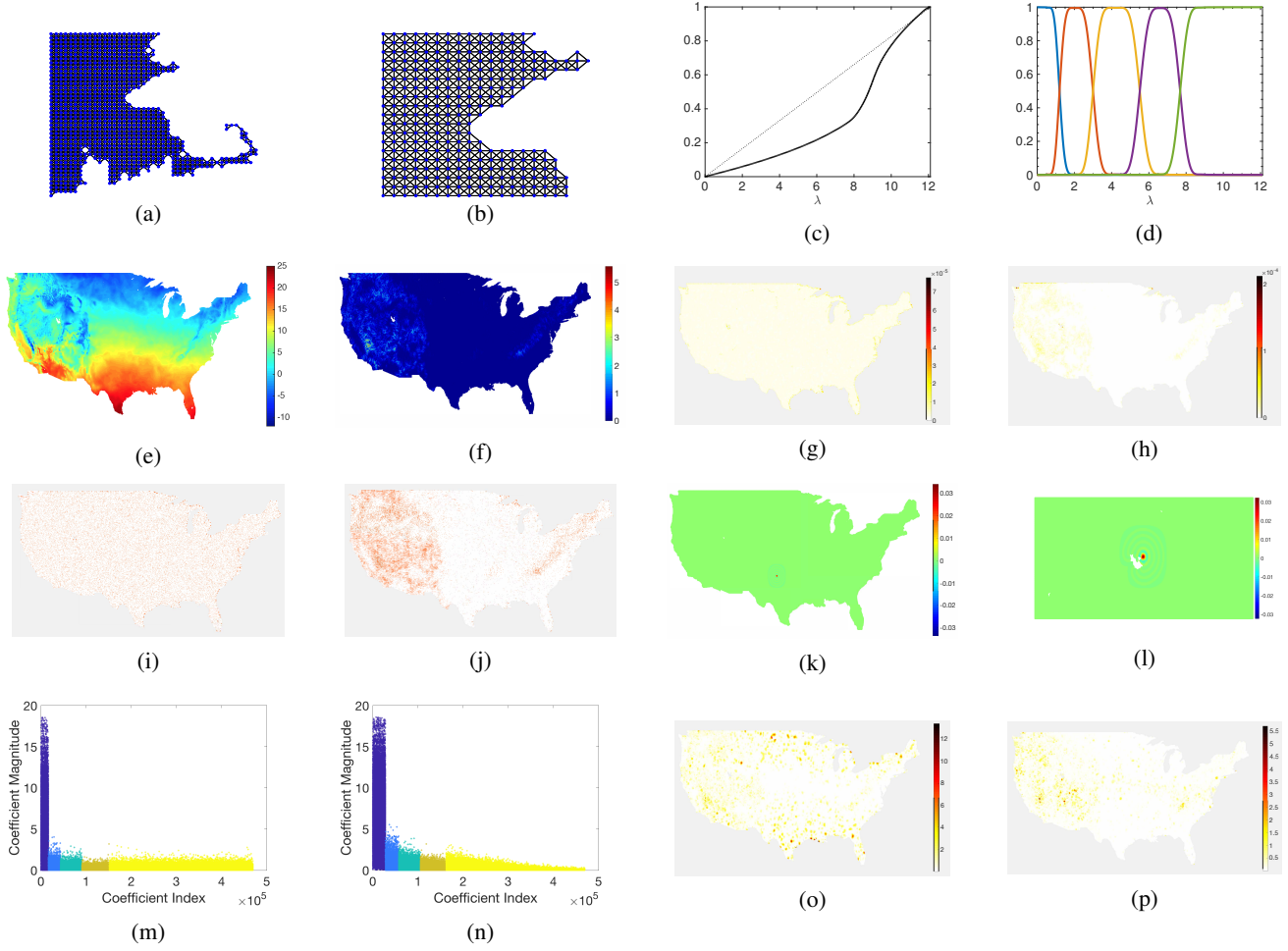


Fig. 11. Fast M -channel filter bank example. (a)-(b) The eight-neighbor local graph structure for the Eastern Massachusetts and Boston regions. (c) Estimate of the cumulative spectral distribution function of the graph. (d) Approximate filter bank with five bands and degree 50 polynomial approximations. (e) Average temperatures for March 2018, measured at 469,404 locations. (f) Magnitudes of the filtered signal on the second band, $|h_2(\mathcal{L})\mathbf{f}|$. (g)-(h) Weights of the non-uniform sampling distribution for the second band, without and with adaptation to the filtered signal in (f). (i) The 27,021 vertices randomly selected according to the distribution in (g) to be included in $\mathcal{V}_2$. (j) The 29,516 vertices randomly selected according to the signal-adapted distribution in (h) to be included in $\mathcal{V}_2$. (k)-(l) Example scaling functions, with the latter zoomed in to see the effect of the missing measurements over Great Salt Lake. (m)-(n) Magnitudes of all 469,404 fast M-CSFB analysis coefficients for the signal in (e), colored by band, for the non-adapted and signal-adapted transforms, respectively. (o)-(p) Absolute values of the differences between the reconstructions and the original signal, for the non-adapted and adapted transforms, respectively.

the graph, with a higher density near the boundaries. The corresponding non-uniform sampling distribution and realization for the second band in the signal-adapted case are shown in Fig. 11(h) and Fig. 11(j). The non-zero analysis coefficients of the bandpass filters tend to coincide with topographical changes, as shown Fig. 11(f). Accordingly, the signal-adapted transform selects more samples from these regions.

Two scaling functions for the fast M-CSFB are shown in Fig. 11(k)-(l), with the image for the latter one zoomed in near the Great Salt Lake. Because $K = 50$, the atoms are localized within 50 hops of the center vertex. When the center is in a region where each vertex has eight neighbors, the atom is symmetric, resembling the scaling function of a 2D-wavelet transform. When the center is near an edge, like the one shown in Fig. 11(l), the atom adapts to the shape of the underlying graph. While the signal adaptation changes the distribution of center locations, it does not change the shape of the atoms; that is, if a vertex i is chosen as a center location for a given scale in both the non-adapted and signal-adapted versions of the transform, the corresponding atoms are identical.

The magnitudes of the analysis coefficients, shown in Fig. 11(m)-(n), decay quickly, as the signal is generally smooth with some discontinuities that tend to coincide with topographical changes. The signal-adapted transform allocates more samples to the bands on the lower end of the spectrum (57,539 total for the first two bands, as opposed to 43,384 in the non-adapted case). Fig. 11(o)-(p) show the reconstruction errors for the two versions of the fast M -CSFB transform, with the same parameters used in Scenario B above. The MSE of the signal-adapted transform is 0.0479, as opposed to 0.4994 without the signal adaptation. The first driver of this reduction is the allocation of additional samples to the first band, where 0.4586 of the 0.4994 MSE is incurred when not adapted to the signal. Additionally, more of this band's samples are taken in the upper and lower thirds of the country, where the energy of the filtered signal is concentrated.

C. Parameter Choices and Tradeoffs

For the next set of numerical experiments, we apply a 4-band filter bank to the piecewise smooth signal from Fig.

5(a), and consider the outputs of the first (lowpass) and third (bandpass) filters, shown in the the top two rows of Fig. 12. We start by examining two key design choices for random sampling and interpolation of these two filtered signals: the sampling distribution and the number of samples. Unless otherwise noted, we use the settings of Scenario B above.

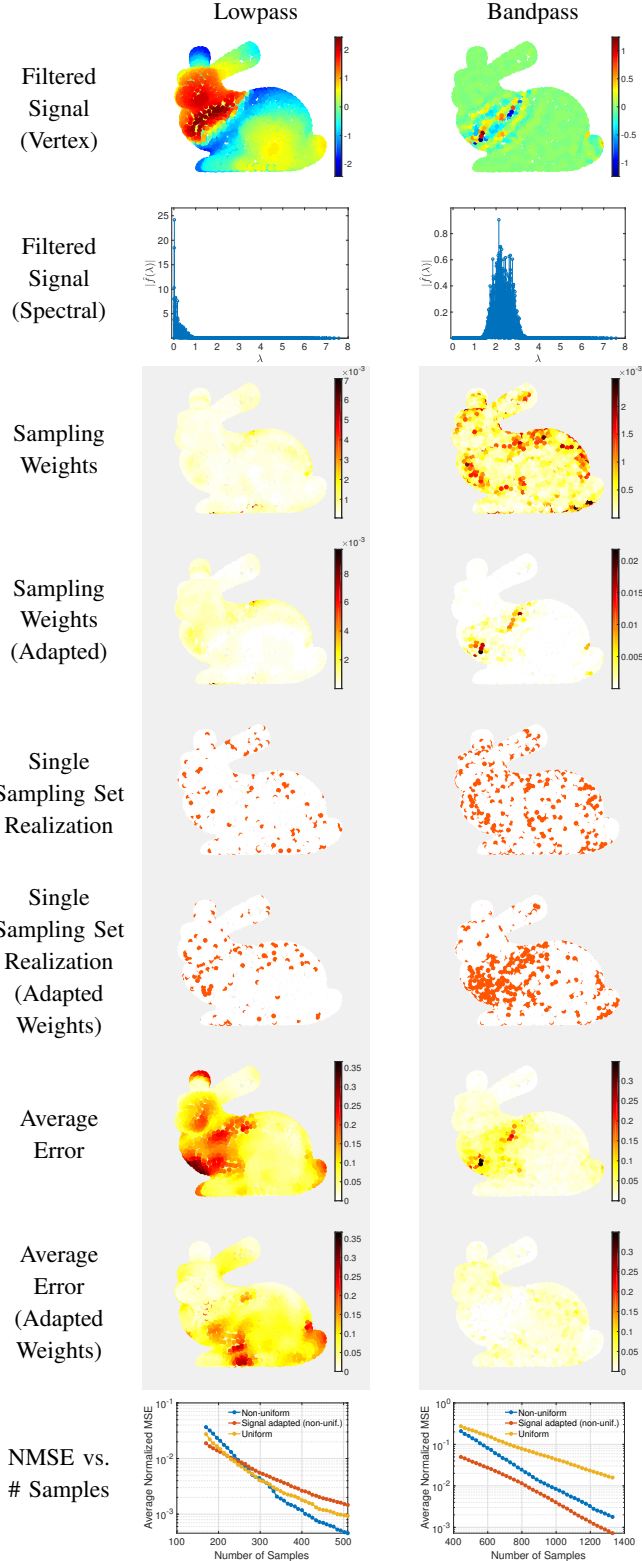


Fig. 12. Tradeoff between reconstruction error and the number of samples, with and without adapting the sampling weights to the signal.

1) *Sampling distributions:* In the third through eighth rows of Fig. 12, we examine the difference between using a non-uniform sampling distribution that is only adapted to the graph and a non-uniform sampling distribution that is adapted to both the graph and the signal, as discussed in Section IV. Because the energy of the lowpass signal is fairly evenly distributed across the bunny, the two sampling distributions are not that different for the first band, with the largest differences in the lower, rear region of the bunny. The energy of the bandpass signal is heavily concentrated around the midsection and tail of the bunny, the locations of the discontinuities in the original signal. The signal-adapted sampling distribution therefore places heavier weights in those areas. The seventh and eighth rows in Fig. 12 show the absolute values of the reconstruction errors, averaged over 50 trials of the random sampling, when the number of random samples is equal to the estimated number of eigenvalues in the specified band (170 in the lowpass case and 443 in the bandpass case). The benefit of the additional samples near the midsection for the bandpass channel is evident, as the average error is lower in this area.

Both non-uniform sampling distributions consistently outperform uniform sampling in our experiments. The graph Laplacian eigenvectors associated with lower eigenvalues tend to be less localized, resulting in non-uniform sampling weights that are closer to uniform weights. Subsequently, there is less benefit from performing the non-uniform sampling on the first band (c.f., bottom row of Fig. 12), consistent with the prior literature on sampling and interpolation of graph signals, such as [39]. The benefit of non-uniform sampling is greater for bands that include more eigenvectors whose energy is concentrated in certain regions of the graph.

2) *Number of samples vs. reconstruction error:* Note that the polynomial approximated filters have wider supports as compared to the ideal filters. By performing critical sampling based on the estimated supports of the ideal filters, we may not have enough samples to reconstruct signals from the (wider) filtered subspace. To get a better reconstruction, we could include more samples for each band. In the bottom row of Fig. 12, we explore the tradeoff between the number of samples and the reconstruction error, increasing the number of samples for each band to three times the estimated number of eigenvalues.

3) *Polynomial approximation order:* To examine the effect of the polynomial order K , we plot in Fig. 13 the reconstruction NMSEs for two channels of the fast M -CSFB transform applied to the piecewise smooth bunny signal with the parameters of Scenario B, averaged over 50 trials each for uniform sampling, non-uniform sampling, and non-uniform sampling adapted to the signal as well as the graph. One key takeaway is that the polynomial degree plays a more important role when the number of bands is larger, as each filter is narrower, and thus more difficult to approximate by lower order polynomials.

4) *Allocation of samples:* In addition to adapting the non-uniform sampling distributions, the signal-adapted fast M -CSFB transform redistributes the N samples between the channels, allocating more samples to bands where more of the signal's energy resides. Again for a 4-band M -CSFB transform of the piecewise smooth bunny graph signal, Table II breaks

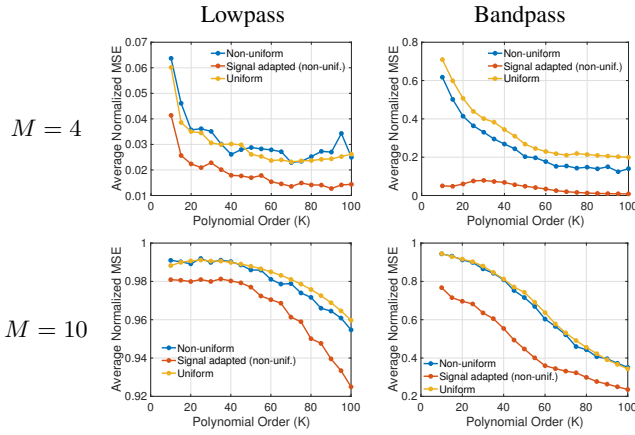


Fig. 13. The role of the polynomial order K in the average reconstruction error for two different channels of the fast M -CSFB transform.

down the improvement in NMSE (averaged over 50 trials). We see that both adaptations reduce the reconstruction error. Adapting the sampling distributions and allocation of samples to the signal also reduces the NMSE in each of the five examples and two scenarios shown in Table I.

	No Signal Adaptation	Sampling Distributions Adapted	Sampling Distributions & Allocations Adapted
Scenario A: faster	0.0399	0.0218	0.0106
Scenario B: more accurate	0.0318	0.0144	0.0052

TABLE II
AVERAGE NORMALIZED MEAN SQUARE RECONSTRUCTION ERROR FOR BUNNY SIGNAL WITH 4-BAND FAST M -CSFB TRANSFORM VARIANTS

D. Compression Examples

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

$$\underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{f} - \Phi \mathbf{x}\|_2^2 \quad \text{subject to } \|\mathbf{x}\|_0 \leq T, \quad (22)$$

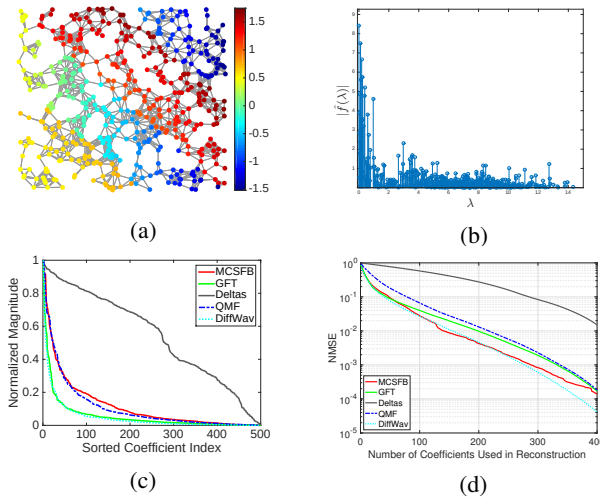


Fig. 14. Compression example. (a)-(b) Piecewise-smooth signal from [28, Fig. 11] in the vertex and graph spectral domains. (c) The normalized sorted magnitudes of the transform coefficients for the proposed M -CSFB (exact version), the graph Fourier transform, the basis of Kronecker deltas, the quadrature mirror filterbank [17], and the diffusion wavelet transform [10]. (d) The reconstruction errors as a function of the sparsity threshold T in (22).

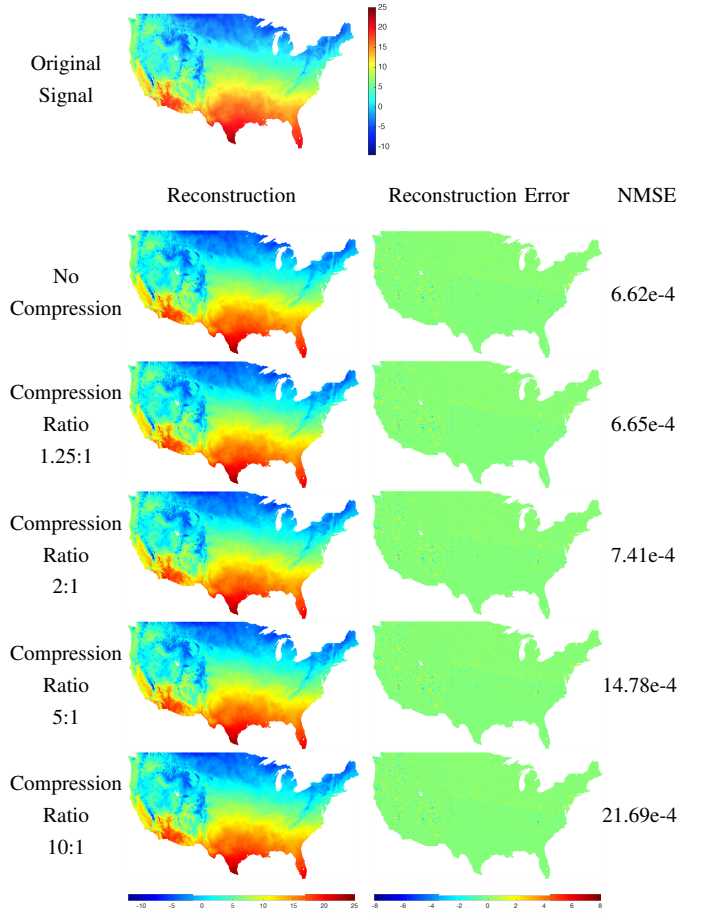


Fig. 15. Compression of the average temperature signal from Section V-B with a signal-adapted fast M -CSFB transform with $M = 5$ bands. From top to bottom, we keep 100%, 80%, 50%, 20%, and 10% of the analysis coefficients, setting the rest to 0 before synthesis. In all cases, we keep all 28,022 scaling coefficients (the first band, about 6% of the overall coefficients), and then use the remaining budget to store the wavelet coefficients with the largest magnitudes.

where T is a predefined sparsity level. After normalizing the atoms of various critically-sampled dictionaries, we use the greedy orthogonal matching pursuit (OMP) algorithm [80], [81] to approximately solve (22). We show the normalized mean square reconstruction errors (NMSE) $\frac{\|\mathbf{f}_{rec} - \mathbf{f}\|_2^2}{\|\mathbf{f}\|_2^2}$ in Fig. 14(d). Note that for a fair comparison, we use the exact computation of $\mathbf{U}$ in the design of all four dictionaries that utilize it. For the M -CSFB, Fig. 4 shows the partition into uniqueness sets, and Fig. 2 shows the filter bank.

In Fig. 15, we compress the average temperature signal from Section V-B, which has 469,404 values. We use a signal-adapted fast M -CSFB transform with $M = 5$ bands and the same parameter settings as Scenario B above. Fig. 15 captures the tradeoff between the normalized mean square reconstruction error and the compression ratio. Even using only 10% of the coefficients, it is difficult to visually identify errors between the reconstruction and the original signal.

E. Fast Approximate Graph Fourier Transform

If the graph Laplacian eigenvalues are distinct, using the exact M -CSFB transform of Section II with $M = N$ and the filter endpoints $\tau_0 = 0$, $\tau_M = \lambda_{\max} + 1$, and $\tau_m = \frac{\lambda_{m-1} + \lambda_m}{2}$

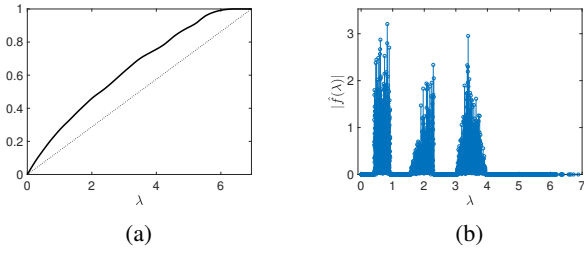


Fig. 16. (a) Estimate of the cumulative spectral distribution of the Minnesota road network shown in Fig. 4. (b) Synthetic signal in the graph spectral domain of the Minnesota graph, generated with the method from [82].

for $m = 1, 2, \dots, N - 1$ yields exactly the graph Fourier transform. The proposed fast M -CSFB transform can therefore be used as a fast approximate graph Fourier transform, with a coarser resolution in the spectral domain. Namely, we use the approximation

$$\hat{f}_{approx}(\lambda) = \frac{1}{\sqrt{\eta_{m_\lambda}}} \|\tilde{h}_{m_\lambda}(\mathcal{L})\mathbf{f}\|_2, \quad (23)$$

where m_λ is the index of the band containing λ and η_{m_λ} is the approximate number of eigenvalues contained in the band containing λ . This approximation is motivated by the fact that $\|h_m(\mathcal{L})\mathbf{f}\|_2^2 = \sum_{\{\lambda: h_m(\lambda)=1\}} |\hat{f}(\lambda)|^2$, by Parseval's equality.

In [82]–[84], Le Magoarou, Gribonval, and Tremblay propose approximate fast graph Fourier transform methods that either exactly compute $\mathbf{U}$ or approximate it by a product of sparse and orthogonal matrices. These methods reduce the complexity of applying the approximate graph Fourier transform from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$; however, they still incur the significant upfront computational cost to compute $\mathbf{U}$ or approximate it (e.g., on the Minnesota graph, the parallel truncated Jacobi method takes approximately one hour, compared to the five seconds required to perform an exact eigendecomposition). The fast M -CSFB method (23), on the other hand, provides a coarser approximation to the graph Fourier transform, but scales to much larger graphs.

As a first example, we generate a synthetic signal in the graph Fourier domain of the Minnesota road network using the method of [82, Fig. 1]. Fig. 16(a) shows the estimated cumulative spectral distribution of the Minnesota graph, and Fig. 16(b) shows the synthetic signal in the spectral domain. In Fig. 17, we apply the fast M -CSFB approximation (23), using four different choices of Jackson-Chebyshev polynomial filter banks, each with $M = 20$ filters. The first chooses the bands to have equal length; the second chooses the bands to have an approximately equal number of eigenvalues; the third shifts the ends of the second slightly according to the procedure outlined in the second for loop of Algorithm 3; and the fourth is the same as the third, except with a polynomial approximation order of $K = 250$, as compared to $K = 80$ for the first three filter banks. Quantitatively, the normalized mean square errors between the estimates and the actual $\hat{\mathbf{f}}$ are 2.33e-04, 1.64e-04, 1.68e-04, and 1.83e-04, respectively. If we let $M = 50$ and $K = 250$ for the shifted filters, the NMSE drops to 1.59e-04. Qualitatively, the shifted filters seem to better identify the support of the signal, but perform worse on the magnitudes.

In identifying the support of the signal in the spectral do-

main, there is a tradeoff between resolution and computational cost. The resolution of the approximation is controlled by the number of spectral bands M . As we add more bands for finer resolution in the spectral domain, however, the filters become narrower (less smooth). We therefore need higher degree polynomials to accurately approximate the filters, which in turn slows down the computations and may result in atoms whose energy is more spread in the vertex domain. We have seen experimentally that choosing M in the 20-50 range and K to be 4-5 times M results in reasonable approximations.

In Fig. 18, we apply the same approximate graph Fourier transform (23) to the average temperature signal from Fig. 11(e), with $M = 20$, $K = 80$ and the shifted filters. Because we cannot compute the exact eigenvalues in this case, we show the approximation as a continuous function of λ on the interval $[0, \lambda_{\max}]$. A coarse approximation like this can confirm that the signal's energy is concentrated on the low end of the spectrum. We are not aware of any other methods to approximate the graph Fourier transform of a signal on a graph of this size.

VI. CONCLUSION AND EXTENSIONS

We have proposed the first critically sampled spectral filter banks for graph signals that scale to graphs with hundreds of thousands to millions of vertices. The fast M -CSFB transform approximately projects a graph signal onto different bands of the graph Laplacian spectrum. To improve computational efficiency, we leverage the computation of $\{\tilde{T}_k(\mathcal{L})\mathbf{X}\}$ in multiple ways: to estimate the spectral density for the design of the filter bank, to estimate the number of samples for each band, and to estimate the non-uniform sampling distributions.

The key idea behind the filter bank design is to choose the end points of each band to be in less dense regions of the spectrum so that the resulting filters are more amenable to polynomial approximation. Adapting the non-uniform sampling distribution and allocation of the samples across the bands to the specific signal being analyzed improves the accuracy of the synthesis process without adding to the computational complexity of the setup and analysis steps.

On one hand, the proposed transform can be seen as a fast approximation of the graph Fourier transform with a coarser resolution in the spectral domain, as discussed in Section V-E. On the other hand, the atoms of the proposed transform can also be viewed as a subset of the atoms of a spectral graph wavelet transform [12], albeit with a different set of filters. Both transforms yield atoms of the form $\tilde{h}_m(\mathcal{L})\delta_i$, but the spectral graph wavelet transform includes every vertex i as a center vertex for every scale m .

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 supported on a smaller spectral intervals.

Another complementary direction for future investigation is the specific form of the filters. Within the same construction,

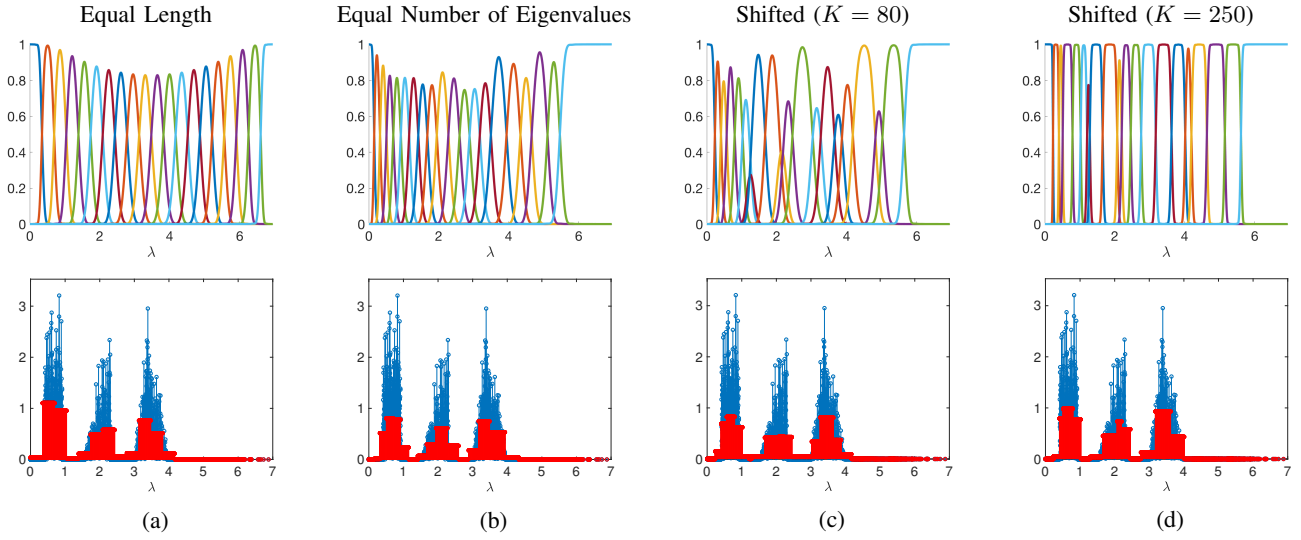


Fig. 17. Fast approximate graph Fourier transform example. The bottom row shows the approximations of the graph Fourier transform of the synthetic signal in Fig. 16(b), based on (23), using the fast M -CSFB transform with four different choices of 20-channel filter banks. The spectrum-adapted filter bank in (b) results in the closest approximation in terms of mean square error, but the filters based on shifted endpoints in (c) and (d) appear to more accurately identify the support of the signal. Ignoring the spectral distribution and taking the filters to have equal length in (a) yields the worst approximation.

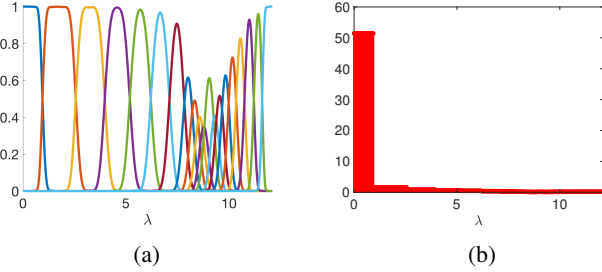


Fig. 18. Approximate graph Fourier transform of the average temperature signal from Fig. 11(e). (a) The 20 spectrum-adapted filters used in the fast M -CSFB. (b) The approximate graph Fourier transform confirms this is a smooth signal with its energy concentrated on the low end of the spectrum. we could use types of filters other than the Jackson-Chebyshev filters (e.g., [85], [86]), or adapt the filters to the energy distribution of the signal or an ensemble of signals [14].

VII. APPENDIX

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

$$\mathbf{A} = [\mathbf{u}_0 \quad \mathbf{u}_1 \quad \dots \quad \mathbf{u}_{k-1} \quad \delta_{\mathcal{S}_1^c} \quad \delta_{\mathcal{S}_2^c} \quad \dots \quad \delta_{\mathcal{S}_{N-k}^c}]$$

has full rank, and the matrix

$$\mathbf{B} = [\mathbf{u}_k \quad \mathbf{u}_{k+1} \quad \dots \quad \mathbf{u}_{N-1} \quad \delta_{\mathcal{S}_1} \quad \delta_{\mathcal{S}_2} \quad \dots \quad \delta_{\mathcal{S}_k}]$$

is singular, implying

$$\text{span}(\mathbf{u}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_{N-1}, \delta_{\mathcal{S}_1}, \delta_{\mathcal{S}_2}, \dots, \delta_{\mathcal{S}_k}) \neq \mathbb{R}^N. \quad (24)$$

Since $\dim(\text{span}(\mathbf{u}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_{N-1})) = N - k$ and $\dim(\text{span}(\delta_{\mathcal{S}_1}, \delta_{\mathcal{S}_2}, \dots, \delta_{\mathcal{S}_k})) = k$, equation (24) implies that there must exist a vector $\mathbf{x} \neq \mathbf{0}$ such that $\mathbf{x} \in \text{span}(\mathbf{u}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_{N-1})$ and $\mathbf{x} \in \text{span}(\delta_{\mathcal{S}_1}, \delta_{\mathcal{S}_2}, \dots, \delta_{\mathcal{S}_k})$. Yet, $\mathbf{x} \in \text{col}(\mathbf{U}_{\mathcal{T}^c})$ implies $\mathbf{x}$ is orthogonal to $\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{k-1}$, and, similarly, $\mathbf{x} \in \text{span}(\delta_{\mathcal{S}_1}, \delta_{\mathcal{S}_2}, \dots, \delta_{\mathcal{S}_k})$ implies $\mathbf{x}$ is orthogonal to $\delta_{\mathcal{S}_1^c}, \delta_{\mathcal{S}_2^c}, \dots, \delta_{\mathcal{S}_{N-k}^c}$. 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 $\text{col}(\mathbf{U}_{\mathcal{T}})$, 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 $\text{col}(\mathbf{U}_{\mathcal{T}^c})$, completing the proof of sufficiency. Necessity follows from the same argument, with the roles of $\mathbf{A}$ and $\mathbf{B}$ interchanged. ■

REFERENCES

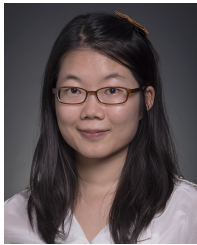
- [1] N. Perraudin, J. Paratte, D. I Shuman, V. Kalofolias, P. Vandergheynst, and D. K. Hammond, "GSPBOX: A toolbox for signal processing on graphs," <https://lts2.epfl.ch/gsp/>.
- [2] D. I Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on 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.*, Jun. 2010, pp. 367–374.
- [7] J. Iriou and N. Saito, "Hierarchical graph Laplacian eigen transforms," *JSIAM Letters*, vol. 6, pp. 21–24, 2014.
- [8] N. Tremblay and P. Borgnat, "Subgraph-based filterbanks for graph signals," *IEEE Trans. Signal Process.*, vol. 64, no. 15, pp. 3827–3840, Mar. 2016.
- [9] S. Chen, A. Singh, and J. Kovačević, "Multiresolution representations for piecewise-smooth signals on graphs," *arXiv e-Prints*, 2018. [Online]. Available: <http://arxiv.org/abs/1803.02944>
- [10] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 53–94, 2006.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] H. Behjat, U. Richter, D. Van De Ville, and L. Sörnmo, "Signal-adapted tight frames on graphs," *IEEE Trans. Signal Process.*, vol. 64, no. 22, pp. 6017–6029, Nov. 2016.
- [15] 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.
- [16] S. K. Narang and A. Ortega, "Local two-channel critically sampled filter-banks on graphs," in *Proc. Int. Conf. Image Process.*, Sep. 2010, pp. 333–336.

- [17] —, “Perfect reconstruction two-channel wavelet filter-banks for graph structured data,” *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2786–2799, Jun. 2012.
- [18] —, “Compact support biorthogonal wavelet filterbanks for arbitrary undirected graphs,” *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4673–4685, Oct. 2013.
- [19] 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.
- [20] 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.
- [21] —, “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.
- [22] L. J. Grady and J. R. Polimeni, *Discrete Calculus*. Springer, 2010.
- [23] 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.
- [24] V. N. Ekambaram, G. C. Fanti, B. Ayazifar, and K. Ramchandran, “Critically-sampled perfect-reconstruction spline-wavelet filterbanks for graph signals,” in *Proc. IEEE Glob. Conf. Signal and Inform. Process.*, Dec. 2013, pp. 475–478.
- [25] 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.
- [26] 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.
- [27] S. K. Narang and A. Ortega, “Lifting based wavelet transforms on graphs,” in *Proc. APSIPA ASC*, Sapporo, Japan, Oct. 2009, pp. 441–444.
- [28] 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.
- [29] 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.
- [30] Y. Jin and D. I. Shuman, “An M-channel critically sampled filter bank for graph signals,” in *Proc. IEEE Int. Conf. Acc., Speech, and Signal Process.*, Mar. 2017, pp. 3909–3913.
- [31] I. Pesenson, “Sampling in Paley-Wiener spaces on combinatorial graphs,” *Trans. Amer. Math. Soc.*, vol. 360, no. 10, pp. 5603–5627, 2008.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] I. Shomorony and A. S. Avestimehr, “Sampling large data on graphs,” in *Proc. IEEE Glob. Conf. Signal and Inform. Process.*, Dec. 2014, pp. 933–936.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] S. Chen, R. Varma, A. Singh, and J. Kovačević, “Signal recovery on graphs: Fundamental limits of sampling strategies,” *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 4, pp. 539–554, Dec. 2016.
- [40] 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.
- [41] P. Di Lorenzo, S. Barbarossa, and P. Banelli, “Sampling and recovery of graph signals,” *arXiv e-Prints*, 2017. [Online]. Available: <http://arxiv.org/abs/1712.09310>
- [42] C. Paige and M. Wei, “History and generality of the CS decomposition,” *Linear Algebra Appl.*, vol. 208–209, pp. 303–326, Sep. 1994.
- [43] G. Strang and T. Nguyen, “The interplay of ranks of submatrices,” *SIAM Review*, vol. 46, no. 4, pp. 637–646, 2004.
- [44] Wikipedia, https://en.wikipedia.org/wiki/Steinitz_exchange_lemma.
- [45] F. Poloni, “Partitioning an orthogonal matrix into full rank square submatrices,” MathOverflow, <http://mathoverflow.net/q/243990>.
- [46] C. Greene, “A multiple exchange property for bases,” *Proc. AMS*, vol. 39, no. 1, pp. 45–50, Jun. 1973.
- [47] C. Greene and T. L. Magnanti, “Some abstract pivot algorithms,” *SIAM J. Appl. Math.*, vol. 29, no. 3, pp. 530–539, Nov. 1975.
- [48] Wikipedia, https://en.wikipedia.org/wiki/Laplace_expansion.
- [49] D. Gleich, “The MatlabBGL Matlab library,” http://www.cs.purdue.edu/homes/dgleich/packages/matlab_bgl/index.html.
- [50] Stanford University Computer Graphics Laboratory, “The Stanford 3D Scanning Repository,” <http://graphics.stanford.edu/data/3Dscanrep/>.
- [51] N. J. Higham, *Functions of Matrices*. Society for Industrial and Applied Mathematics, 2008.
- [52] 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.
- [53] A. Frommer and V. Simoncini, “Matrix functions,” in *Model Order Reduction: Theory, Research Aspects and Applications*. Springer, 2008, pp. 275–303.
- [54] 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.
- [55] D. I. Shuman, P. Vandergheynst, D. Kressner, and P. Frossard, “Distributed signal processing via Chebyshev polynomial approximation,” *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 4, pp. 736–751, Dec. 2018.
- [56] 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>
- [57] 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.
- [58] 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.
- [59] 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.
- [60] G. Puy and P. Pérez, “Structured sampling and fast reconstruction of smooth graph signals,” *Inf. Inference*, vol. 7, no. 4, pp. 657–688, 2018.
- [61] 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.
- [62] L. Lin, Y. Saad, and C. Yang, “Approximating spectral densities of large matrices,” *SIAM Review*, vol. 58, no. 1, pp. 34–65, 2016.
- [63] P. Van Mieghem, *Graph Spectra for Complex Networks*. Cambridge University Press, 2011.
- [64] T. Tao, *Topics in Random Matrix Theory*. American Mathematical Society, 2012.
- [65] 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.
- [66] 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.
- [67] L.-W. Wang, “Calculating the density of states and optical-absorption spectra of large quantum systems by the plane-wave moments method,” *Phys. Rev. B*, vol. 49, no. 15, p. 10154, 1994.
- [68] 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.
- [69] F. N. Fritsch and R. E. Carlson, “Monotone piecewise cubic interpolation,” *SIAM J. Numer. Anal.*, vol. 17, no. 2, pp. 238–246, Apr. 1980.
- [70] P. Drineas, M. Magdon-Ismael, M. W. Mahoney, and D. P. Woodruff, “Fast approximation of matrix coherence and statistical leverage,” *J. Mach. Learn. Res.*, vol. 13, no. Dec., pp. 3475–3506, 2012.
- [71] M. W. Mahoney and P. Drineas, “CUR matrix decompositions for improved data analysis,” *Proc. Natl. Acad. Sci.*, vol. 106, no. 3, pp. 697–702, 2009.
- [72] M. W. Mahoney, “Randomized algorithms for matrices and data,” *Foundations and Trends in Machine Learning*, vol. 3, no. 2, pp. 123–224, 2011.
- [73] A. Chan and C. D. Godsil, “Symmetry and eigenvectors,” in *Graph symmetry*. Springer, 1997, pp. 75–106.
- [74] 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.
- [75] J. Paratte and L. Martin, “Fast eigenspace approximation using random signals,” *arXiv e-Prints*, 2016. [Online]. Available: <http://arxiv.org/abs/1611.00938>
- [76] J. Chen, M. Anitescu, and Y. Saad, “Computing $f(A)b$ via least squares polynomial approximations,” *SIAM J. Sci. Comp.*, vol. 33, no. 1, pp. 195–222, Feb. 2011.
- [77] W. Klotz, “Graph coloring algorithms,” *Mathematik-Bericht*, vol. 5, pp. 1–9, 2002.
- [78] R. Vose, S. Applequist, M. Squires, I. Durre, M. Menne, C. Williams, C. Fenimore, K. Gleason, and D. Arndt, “Gridded 5km GHCN-daily temperature and precipitation dataset (nCLIMGRID), version 1,” *Maximum temperature, minimum temperature, average temperature, and precipitation*. NOAA National Centers for Environmental Information, 2014.
- [79] —, “Improved historical temperature and precipitation time series for U.S. climate divisions,” *J. Appl. Meteorol. Climatol.*, vol. 53, no. 5, pp. 1232–1251, May 2014.
- [80] J. A. Tropp, “Greed is good: Algorithmic results for sparse approximation,” *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [81] M. Elad, *Sparse and Redundant Representations*. Springer, 2010.
- [82] L. Le Magoarou, R. Gribonval, and N. Tremblay, “Approximate fast graph Fourier transforms via multilayer sparse approximations,” *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 2, pp. 407–420, Jun. 2018.
- [83] L. Le Magoarou and R. Gribonval, “Are there approximate fast Fourier transforms on graphs?” in *Proc. IEEE Int. Conf. Acc., Speech, and Signal Process.*, Mar. 2016, pp. 4811–4815.

- [84] —, “Flexible multilayer sparse approximations of matrices and applications,” *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 688–700, Jun. 2016.
- [85] D. B. Tay and Z. Lin, “Design of near orthogonal graph filter banks,” *IEEE Signal Process. Lett.*, vol. 22, no. 6, pp. 701–704, Jun. 2015.
- [86] J. Liu, E. Isufi, and G. Leus, “Filter design for autoregressive moving average graph filters,” *IEEE Trans. Signal Inf. Process. Netw.*, vol. 5, no. 1, pp. 47–60, 2019.



Shuni Li received the B.A. degree in Mathematics and Computer Science from Macalester College, St Paul, Minnesota in 2018. She is currently a Ph.D. student in Statistics at University of California, Berkeley. She is interested in machine learning and optimization, with applications in computational biology.



Yan Jin received the B.A. degree in mathematics and B.A. degree in sociology from Macalester College, in 2016. She is currently a Ph.D. student in the Social and Engineering Systems program at the Institute for Data, Systems and Society, Massachusetts Institute of Technology, Cambridge, MA. Her research interests include algorithms on graphs, algorithmic mechanism design, causal inference, and graph signal processing. She is a member of the IEEE Signal Processing Society (2017-) and the American Mathematical Society (2016-).



David I. Shuman received the B.A. degree in economics and the M.S. degree in engineering-economic systems and operations research from Stanford University, Stanford, CA, in 2001 and the M.S. degree in electrical engineering: systems, the M.S. degree in applied mathematics, and the Ph.D. degree in electrical engineering: systems from the University of Michigan, Ann Arbor, in 2006, 2009, and 2010, respectively.

He is currently an Associate Professor in the Department of Mathematics, Statistics, and Computer Science, Macalester College, St. Paul, Minnesota, which he joined in January 2014. From 2010 to 2013, he was a Postdoctoral Researcher at the Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. His research interests include signal processing on graphs, computational harmonic analysis, and stochastic scheduling and resource allocation problems.

Dr. Shuman has served as an Associate Editor for the IEEE Transactions on Signal and Information Processing Over Networks (2019-) and the IEEE Signal Processing Letters (2017-2019), and on the Technical Program Committee for the IEEE Global Conference on Signal and Information Processing (2015-2018). He received the 2016 IEEE Signal Processing Magazine Best Paper Award, and was a 2014-2015 Project NExT Fellow of the Mathematical Association of America.