

Multiscale network analysis through tail-greedy bottom-up approximation, with applications in neuroscience

Xinyu Kang

Department of Mathematics and Statistics
Boston University
Boston, MA 02215
Email: xkang@bu.edu

Piotr Fryzlewicz

Department of Statistics
London School of Economics
London, WC2A 2AE
Email: p.fryzlewicz@lse.ac.uk

Catherine Chu

Department of Neurology
Massachusetts General Hospital
Boston, MA 02114
Email: cjchu@mgh.harvard.edu

Mark Kramer

Department of Mathematics and Statistics
Boston University
Boston, MA 02215
Email: mak@math.bu.edu

Eric D. Kolaczyk

Department of Mathematics and Statistics
Boston University
Boston, MA 02215
Email: kolaczyk@bu.edu

Abstract—We propose the TGUH (Tail-Greedy Unbalanced Haar) transform for networks, which results in an orthonormal, adaptive decomposition of the network adjacency matrix into Haar-wavelet like components. The ‘tail-greediness’ of the algorithm – indicating multiple greedy steps are taken in a single pass through the data – enables both fast computation and consistent estimation of network signals. We focus on development of our multiscale network decomposition and a corresponding method for network signal denoising. Moreover, we establish consistency of our resulting denoising methodology, present numerical simulations illustrating compression, and illustrate through application to signals on diffusion tensor imaging (DTI) networks.

I. INTRODUCTION

Wavelet-related methods have been developed extensively for classical signal and image processing problems. In recent years, there has been a resurgence of interest in this area, within the emerging field of graph signal processing. Advancement in this area does not come without challenges. Not all tools in classical signal processing can be directly transferred to graph signal processing. Problems with which people are concerned include but are not limited to: How to effectively compress a signal on a network? How to identify and remove noise from a signal on a graph? See [21] for seminal work in graph signal processing. An early characterization of the multi-scale aspect of this field can be found in [22].

In this work, we present an algorithm that offers certain solutions to these problems. We work with an undirected connected graph $G^0 = (V^0, E^0)$, where $V^0 = \{v_1^0, \dots, v_N^0\}$ is the set of vertices and $E^0 = \{e_{ij}^0 | v_i^0 \in V^0, v_j^0 \in V^0, i \neq j\}$ is the set of edges (assumed to be without self-loops).

Our contribution is a new class of graph wavelets, based on the notion of ‘tail-greedy’ unbalanced Haar transformations. We also establish a consistency result for piecewise-constant

function estimation using thresholding procedures in the corresponding basis obtained from our TGUH.

A. Related work

Various previous studies have focused on either wavelet transforms *on* networks and/or wavelet transforms *of* networks. This highly active literature has already become too large to survey adequately here. The first category includes those transformations based on the graph Laplacian, relevant to analysis of signals over networks, e.g., see [6] and [14]. Earlier work of Crovella and Kolaczyk [7] extends the Mexican hat wavelet to unweighted graphs and uses it to analyze computer network traffic. Also related is work by Gavish, Nadler and Coifman [11], who develop various results on graphs that possess a hierarchical tree structure, and the work by Irion and Saito [15], where they compute graph basis dictionaries using graph Laplacian eigen transforms and generalized Haar-Walsh transforms. The second category focuses more exclusively on transforms of the graph itself. Examples include work by Murtagh [20], who develops an invertible wavelet transform based on hierarchical clustering using Ward’s criterion, and the work by Lee, Nadler and Wasserman [19], who propose an attribute-based construction of adaptive multi-scale hierarchical trees. Other methods of this type incorporate ideas from matrix compression and factorization. For example, see [18], where they propose a multi-scale way of doing matrix factorization, which effectively decomposes large matrices using a series of transformation matrices that capture structures at different scales. Other examples of relevant literature are cited strategically throughout this paper.

The tail-greedy unbalanced Haar transformation was originally proposed by Fryzlewicz [8] for the one-dimensional signal plus noise model. The algorithm results in a nonlinear

but conditionally orthonormal, multiscale decomposition of the data with respect to an adaptively chosen unbalanced Haar wavelet basis. Related work also includes the work by Fryzlewicz and Timmermans [10], where an adaptive Haar-type of transformation is used for image compression and denoising. Here we extend the TGUH method to networks and use it to show various results and applications.

The organization of this paper is as follows. In section II, we present the development of our TGUH transforms. In section III, we discuss graph signal denoising with the TGUH and propose a consistent method of estimation. In section IV, we illustrate the utility of our algorithm through simulation and application in computational neuroscience. In section V, we discuss potential extensions.

II. TGUH OF NETWORKS

Consider a graph $G^0 = (V^0, E^0)$, the connectivity of which we summarize by its adjacency matrix $\mathbf{W}^0 \in \mathbb{R}^{n \times n}$, where $w_{ij}^0 \geq 0$ is the (possibly non-binary) edge weight between vertex i and j , such that $w_{ij}^0 = 0$ indicates no edge.

The TGUH is a bottom-up method. At each run, we select columns from the adjacency matrix, corresponding to linked pairs of nodes, and merge them by applying an orthonormal transformation to the columns. We define \mathbf{W}^ℓ to be the adjacency matrix, and V^ℓ to be the corresponding set of vertices, after the ℓ -th iteration. We denote by v_r^ℓ the r -th node in V^ℓ , and by N_r^ℓ and $N_{r'}^\ell$, the number of nodes in V^0 represented (through merging) in meta nodes $v_r^{\ell-1}$ and $v_{r'}^{\ell-1}$. Let $\rho \in (0, 1)$ be a constant, used to describe the proportion of pairs of nodes to merge at each run. More precisely, we merge $\lceil \rho |V^{\ell-1}| \rceil$ pairs of nodes at the ℓ -th iteration, where the expression $|\cdot|$ is the cardinality of the vertex set. The parameter ρ controls the speed and greediness of our method. When $\rho = 1/2$, the transform reduces to the non-adaptive (and therefore non-greedy) classical Haar transform; the degree of greediness increases as ρ decreases. In our applications, we use $\rho = 0.01$.

We now outline the procedure.

- 1) At the ℓ -th iteration, search for $\lceil \rho |V^{\ell-1}| \rceil$ pairs of columns for which the ℓ_2 norms of the detail coefficient vectors are the smallest. To be more precise, the algorithm proceeds as follows:

For each pair of columns corresponding to pairs of connected nodes $(v_r^{\ell-1}, v_{r'}^{\ell-1})$, construct a "detail" filter $(a_{(r,r')}^\ell, -b_{(r,r')}^\ell)$, where each filter is uniquely indexed by ℓ and the pair (r, r') . Here $a_{(r,r')}^\ell > 0$ and $b_{(r,r')}^\ell > 0$ are set in the following way:

- a) To produce a sparse representation of the initial input matrix, the algorithm needs to produce zero details over regions of constancy of the network, by which we mean nodes sharing identical neighborhood structure and weighting. Let $j = j(r)$ and $j' = j(r')$ denote the corresponding positions in the adjacency matrix and assume $j < j'$. We compute the details using

$$\mathbf{d}_{(r,r')}^\ell = a_{(r,r')}^\ell \mathbf{W}_{\cdot j}^{\ell-1} - b_{(r,r')}^\ell \mathbf{W}_{\cdot j'}^{\ell-1}.$$

- b) To force orthonormality of the transformation, we impose the following requirement:

$$a_{(r,r')}^\ell{}^2 + b_{(r,r')}^\ell{}^2 = 1$$

The two requirements above determine a unique filter. The detail coefficient vector is computed as

$$\mathbf{d}_{(r,r')}^\ell = \left[\mathbf{W}_{\cdot j}^{\ell-1}, \mathbf{W}_{\cdot j'}^{\ell-1} \right]_{N \times 2} \begin{bmatrix} a_{(r,r')}^\ell \\ -b_{(r,r')}^\ell \end{bmatrix}_{2 \times 1}.$$

- 2) Sort the norms of the detail coefficient vectors $\|\mathbf{d}_{(r,r')}^\ell\|_{\ell_2}$ in ascending order and extract $\lceil \rho |V^{\ell-1}| \rceil$ detail coefficient vectors corresponding to the smallest $\lceil \rho |V^{\ell-1}| \rceil$ elements in the sorted sequence. If any element of the sorted sequence uses nodes already used by any of the previous elements, it is discarded and the next candidate is considered. In the case where there are fewer than $\lceil \rho |V^{\ell-1}| \rceil$ detail coefficient vectors, extract all of them. No nodes will be merged more than once at iteration ℓ .
- 3) For each $\|\mathbf{d}_{(r,r')}^\ell\|_{\ell_2}$, use filter $(b_{(r,r')}^\ell, a_{(r,r')}^\ell)$, which is orthogonal to the previous filter used for computing the detail coefficient, to produce the corresponding merged columns:

$$\mathbf{W}_{\cdot j}^\ell \leftarrow \left[\mathbf{W}_{\cdot j}^{\ell-1}, \mathbf{W}_{\cdot j'}^{\ell-1} \right]_{N \times 2} \begin{bmatrix} b_{(r,r')}^\ell \\ a_{(r,r')}^\ell \end{bmatrix}_{2 \times 1}$$

$$\mathbf{W}_{\cdot j'}^\ell \leftarrow \mathbf{d}_{(r,r')}^\ell$$

where \leftarrow indicates replacement of the original rows/columns with the new one. Alternatively, these operations can be written as

$$\left[\mathbf{w}_{\cdot j}^\ell, \mathbf{w}_{\cdot j'}^\ell \right]_{N \times 2} = \left[\mathbf{w}_{\cdot j}^{\ell-1}, \mathbf{w}_{\cdot j'}^{\ell-1} \right]_{N \times 2} \begin{bmatrix} b_{(r,r')}^\ell & a_{(r,r')}^\ell \\ a_{(r,r')}^\ell & -b_{(r,r')}^\ell \end{bmatrix}_{2 \times 2}.$$

The transformation matrix is a rotation matrix.

- 4) Perform the corresponding row-wise rotation and symmetrize \mathbf{W}^ℓ .

$$\begin{bmatrix} \mathbf{w}_{\cdot j}^\ell \\ \mathbf{w}_{\cdot j'}^\ell \end{bmatrix}_{2 \times N} = \begin{bmatrix} b_{(r,r')}^\ell & a_{(r,r')}^\ell \\ a_{(r,r')}^\ell & -b_{(r,r')}^\ell \end{bmatrix}_{2 \times 2} \begin{bmatrix} \mathbf{w}_{\cdot j}^{\ell-1} \\ \mathbf{w}_{\cdot j'}^{\ell-1} \end{bmatrix}_{2 \times N}$$

- 5) Set $\ell \leftarrow \ell + 1$ and go back to step 1, unless the transform is completed.

A compact summary of the above description is provided as Algorithm 1. Code implementing this algorithm is available at github.com/KolaczykResearch/NetworkTGUH-Code/.

We make a few comments regarding the TGUH. The resulting transformation is non-linear, but it is orthonormal conditional on the order in which the detail coefficient vectors are selected. Given that the transform is conditionally orthonormal, it preserves the ℓ_2 energy of the adjacency matrix. Because small detail coefficients are selected at the beginning of the algorithm, most energy will be concentrated at coarser scales.

The algorithm can be viewed as a variation on agglomerative hierarchical clustering for community detection (e.g., [16, Ch 4.3.3.1]), using a column-wise ℓ_2 norm as our measure of so-called 'linkage', but with particular attention paid to the notions

Algorithm 1 TGUH transform of network

Input: Adjacency matrix: \mathbf{W}^0

- 1: **for** level $\ell = 1$ to L **do**
- 2: **for** each pair of connected nodes (r, r') **do**
- 3: **Compute** candidate ‘‘detail’’ coefficient vector norms $\|\mathbf{d}_{(r,r')}^\ell\|_2 = \|a_{(r,r')}^\ell \mathbf{W}_{\cdot j}^{\ell-1} - b_{(r,r')}^\ell \mathbf{W}_{\cdot j'}^{\ell-1}\|_2$.
- 4: **Store** $\|\mathbf{d}_{(r,r')}^\ell\|_2$.
- 5: **end for**
- 6: **Sort** $\|\mathbf{d}_{(r,r')}^\ell\|_2$ in ascending order.
- 7: **for** $i = 1$ to $\lceil \rho |V^\ell| \rceil$ **do**
- 8: **Select** the column/row corresponding to the smallest $\|\mathbf{d}_{(r,r')}^\ell\|_2$.
- 9: **Update** $W^{\ell-1}$ by

$$\begin{aligned} \begin{bmatrix} \mathbf{W}_{\cdot j}^\ell & \mathbf{W}_{\cdot j'}^\ell \end{bmatrix} &\leftarrow \begin{bmatrix} \mathbf{W}_{\cdot j}^{\ell-1} & \mathbf{W}_{\cdot j'}^{\ell-1} \end{bmatrix} \begin{bmatrix} b_{(r,r')}^\ell & a_{(r,r')}^\ell \\ a_{(r,r')}^\ell & -b_{(r,r')}^\ell \end{bmatrix} \\ \begin{bmatrix} \mathbf{W}_{j \cdot}^\ell \\ \mathbf{W}_{j' \cdot}^\ell \end{bmatrix} &\leftarrow \begin{bmatrix} b_{(r,r')}^\ell & a_{(r,r')}^\ell \\ a_{(r,r')}^\ell & -b_{(r,r')}^\ell \end{bmatrix} \begin{bmatrix} \mathbf{W}_{j \cdot}^{\ell-1} \\ \mathbf{W}_{j' \cdot}^{\ell-1} \end{bmatrix}. \end{aligned}$$

- 10: **end for**
 - 11: **end for**
-

of coarsening and detail, in the usual multiscale tradition. From this perspective, our TGUH is similar in spirit to the hierarchical clustering algorithm of Singh, Nowak, and Calderbank [23].

The term ‘‘tail-greedy’’ comes from the fact that in each run, we select from among the lower tail of the distribution of ‘‘details’’. ‘‘Tail-greedy’’ is not as greedy as standard greedy methods, as it may select more than one detail per run, which ensures the method terminates in at most $\mathcal{O}(\log N)$ levels.

The complexity of the TGUH transform is nearly linear in the number of edges in the graph. For a graph of size N , the number of nodes remaining after ℓ iterations is at most $(1 - \rho)^\ell N$. Solving for the smallest ℓ such that $(1 - \rho)^\ell N < 1$ yields that $\ell > \frac{\log N}{\log(1-\rho)^{-1}}$. At each step, we need to compute details for every edge and sort, which requires up to $\mathcal{O}(|E| \log |E|)$ operations. Accordingly, the complexity of the overall TGUH scales like $\mathcal{O}(\frac{\log N}{\log(1-\rho)^{-1}} \times |E| \log |E|)$. Note that for sparse graphs, where the number of edges is of the same order as the number of vertices, $|E| \sim N$, the complexity scales like $\mathcal{O}(\frac{N \log^2 N}{\log(1-\rho)^{-1}})$.

Note that the TGUH can be expressed in a series of matrix multiplications:

$$\mathbf{W}^L = F^L \dots F^1 \mathbf{W}^0 F^{1\top} \dots F^{L\top}.$$

Using our tail-greedy algorithm, the collection of the column spaces of the F ’s corresponds to that of an unbalanced Haar type of basis and the adjacency matrix is effectively decomposed in a bottom-up fashion. From this perspective, our TGUH transform is a special case of the multi-resolution matrix factorization (MMF) method of Kondor, Teneva and Garg [18], which compresses matrices efficiently through the use of a sequence of sparse orthogonal transforms. Specifically, our TGUH constitutes a 2nd order Jacobi MMF in the language

of that paper. We note, however, that the problem of denoising a graph signal is not considered in [18].

III. DENOISING GRAPH SIGNALS USING TGUH

We have introduced a Haar-like basis for a network G^0 . Now consider a signal f on that network. If the signal varies in a way that is somehow ‘consistent with’ the network structure that drives the network-adaptive steps of our TGUH algorithm, then we should have good signal compression when transforming f through the resulting TGUH orthobasis. Estimation techniques that can exploit this compression should then prove effective for denoising a signal observed with noise. In this section we explore the use of TGUH bases for denoising graph signals.

We adopt the standard signal plus noise model, $g(v) = f(v) + \epsilon(v)$, for $v \in V^0$. Here $g(v)$ is the observed signal, $f(v)$ is an unknown true signal, and $\epsilon(v)$ is an independent and identically distributed $N(0, 1)$ noise. We assume that f is ‘piece-wise constant’ in the sense that the number of edges $e_{ij}^0 \in E^0$ for which $f(i) \neq f(j)$ is no more than some constant $K > 0$. See [3], for example, for related notions of ‘piece-wise smooth’ functions based on vertex subsets.

Traditional multiresolution analysis constructs a sequence of function spaces $\{U_\ell\}$ of increasingly finer scale, by recursively dividing each U_ℓ into a coarser part $U_{\ell+1}$ and its orthogonal complement $W_{\ell+1}$. The latter are the wavelet subspaces. The original space U_0 can thus be decomposed as

$$U_0 = \bigoplus_{\ell=1}^L W_\ell \bigoplus U_L.$$

Decompositions of a function $f \in U_0$ follow accordingly.

The TGUH wavelet transform up to level L can be expressed as

$$f(v) = \sum_{\ell=1}^L \sum_{r=1}^{k(\ell)} \mu_r^\ell \psi_r^\ell(v) + \sum_{r=1}^{k(L)} \gamma_r \phi_r^L(v),$$

where $\mu_r^\ell = \langle f, \psi_r^\ell \rangle$ are the wavelet coefficients with respect to the Haar basis functions ψ_r^ℓ with support on the nodes set V_r^ℓ . We estimate f by estimating each μ_r^ℓ and then taking the inverse transformation.

Define empirical coefficients

$$\alpha_r^\ell = \sum_v g(v) \psi_r^\ell(v).$$

Suppose that the two sets $V_m^{\ell'-1}$ and $V_{m'}^{\ell'-1}$ contain the nodes that merge into the meta node $v_r^{\ell'}$ at the next level, that is $V_r^{\ell'} = V_m^{\ell'-1} \cup V_{m'}^{\ell'-1}$. We define the estimator

$$\hat{\mu}_r^{\ell'} = \alpha_r^{\ell'} \mathbb{I} \left\{ \exists V_{r'}^{\ell'} \subseteq V_r^{\ell'} \mid |\alpha_{r'}^{\ell'}| > \lambda(\ell', r') \right\}, \quad (1)$$

where

$$\lambda(\ell', r') = \sqrt{2 \log N} \left\{ \frac{\sqrt{|V_m^{\ell'-1}|} + \sqrt{|V_{m'}^{\ell'-1}|}}{\sqrt{|V_m^{\ell'-1}|} + |V_{m'}^{\ell'-1}|} \right\}. \quad (2)$$

In this case, we estimate $\mu_r^{\ell'}$ by zero if $\alpha_r^{\ell'}$ and all of its children coefficients fall below the threshold. The advantage of using this

type of threshold is that it allows us to construct an unbiased estimator \hat{f} of f , in the sense that within each constant regime, our estimator is the sample mean of the observed signal within that constant section.

For any estimator \tilde{f} of f , we denote the squared empirical L_2 risk as $R = \frac{1}{N} \sum_v (\tilde{f}(v) - f(v))^2$. We then have the following result characterizing the performance of our estimator \hat{f} .

Theorem III.1. *Let \hat{f} be an estimator of f obtained through the inverse TGUH transform of the estimated coefficients $\hat{\mu}_r^\ell$ in (1), based on the thresholding function $\lambda(\ell', r')$ in (2). Suppose $K = o(N/\log^2 N)$. Then we have that the risk $R(\hat{f}, f)$ is of order $\mathcal{O}\left(\frac{K \log^2 N}{N \log(1-\rho)^{-1}}\right)$ on the set*

$$\mathcal{A} = \left\{ \forall r, \ell, \quad |V_r^\ell|^{-1/2} \left| \sum_{v \in V_r^\ell} \epsilon(v) \right| \leq \sqrt{2 \log N} \right\},$$

where $\mathbb{P}(\mathcal{A}) \rightarrow 1$ as $N \rightarrow \infty$.

Theorem III.1 says that the estimator \hat{f} is an L_2 -consistent estimator of the signal f . The key driver behind the result is the property of ‘‘tail-greediness’’, by which multiple pairs of nodes are merged at each level ℓ . L_2 consistency cannot be guaranteed if the algorithm is greedy, where only one pair of nodes from the tail distribution is merged. Proof of Theorem III.1 can be found in the appendix.

We note that the assumption of piecewise constant f is presumably stronger than needed here, and can likely be relaxed to the case of functions of a certain Hölder smoothness. For example, the type of necessary intermediate approximation theoretic results for Hölder smooth functions follows for our TGUH basis immediately from [5].

IV. APPLICATIONS

A. Simulations

We use simulation to establish a simple proof of concept regarding compression by TGUH. Specifically, we look at the use of our TGUH transform to compress a ‘barbell’ type of network, i.e., a network with two fully connected components joined by a single link. We generated such a barbell with 10 nodes in each fully connected component and applied the TGUH transform described in Algorithm 1. All detail coefficients are zero except that resulting from the last step of the algorithm, where the two connected components merge together, which demonstrates that the TGUH transform is able to capture well the structure of the barbell network. The resulting compression curve is shown in Figure 1.

To demonstrate the robustness of this result, we simulated 100 such barbells perturbed with both Type I and Type II errors, i.e., declaring non-edges edges and vice versa. Here we set $P(\text{Type I error}) = 0.01$ and $P(\text{Type II error}) = (1 - \text{Den})/\text{Den} \times P(\text{Type I error})$, where Den is the density of adjacency matrix of the noise-free barbell. Under this setting, the expected density of the ‘noisy’ network is the same as the original one. Figure 1 shows the average compression curve resulting from applying the TGUH to these noisy networks.

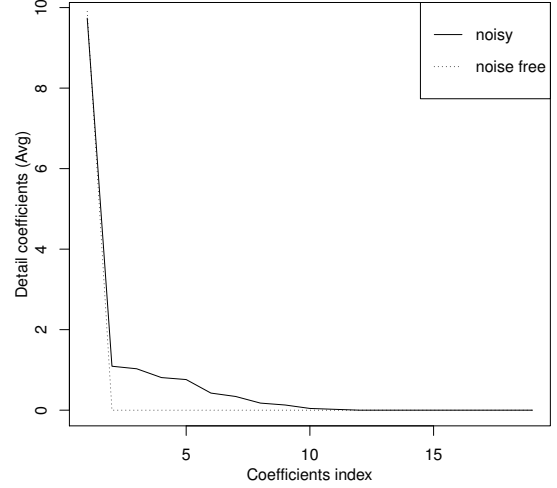


Fig. 1: Compression curve for barbell network (dotted) and average compression curve for noisy barbell network (solid) using TGUH.

We see that this curve is qualitatively quite similar to that for the noise-free version of these networks.

B. EEG data on a DTI network

We now provide an application of using the TGUH to denoise EEG signals over a DTI network. Understanding the relationship between brain anatomical connectivity and brain dynamics remains an active research area [4] [17]. In general, different frequency bands have been associated with different brain function [2], and different spatial organization over the brain’s surface, but how brain anatomical connectivity relates to brain rhythms remains incompletely understood.

As an example application of the method developed here, we consider two types of data collected from a human subject. Brain anatomical connectivity was computed from high resolution diffusion tensor imaging data using previously described methods [4]. Briefly, 324 regions of interest (ROI) at the gray-white matter border were selected using the topology of a recursively subdivided icosahedron fitted in the subject’s cortical surface inflated to a sphere [9] [12]. Quantitative bidirectional white matter connectivity between each ROI pair was computed using Protrackx2 software [1] where 500 streamlines were sampled per voxel within each ROI. A connectivity index was then computed for each ROI pair as the proportion of successful streamlines connected between the ROI pair over the total number of streamlines sampled.

Dynamic brain electrical activity was collected from the same ROIs using the MNE-C and Freesurfer software packages [9] [13] and previously reported methods [4]. Briefly, EEG data during stage 2 sleep was recorded with a 70 channel EEG cap and electrode positions collected using a 3D digitizer. Anatomical cortical surfaces of the brain were reconstructed

using T1-weighted MRI data and a forward solution was calculated using a three-layer boundary element model consisting of the inner skull, outer skull and outer skin surfaces. The digitized EEG electrode coordinates were coregistered to the reconstructed surface using the nasion and auricular points. The inverse operator was computed from the forward model and the resultant current estimates at each ROI calculated. Ten seconds of artifact free data were selected for analysis. From the electrical source estimates at each ROI, we computed the power spectrum using the multitaper method (bandwidth 1 Hz, 9 tapers). We then compute the average power in the theta band (5-8 Hz), alpha band (8-12 Hz), beta band (12-20 Hz), and gamma band (20-40 Hz).

The DTI network contains 324 nodes and 1487 edges. Power in each spectral band was compressed using the TGUH bases on the DTI network. The resulting compression curves appear in Figure 2. In all four bands, half of the signals were captured using the leading 50 basis functions.

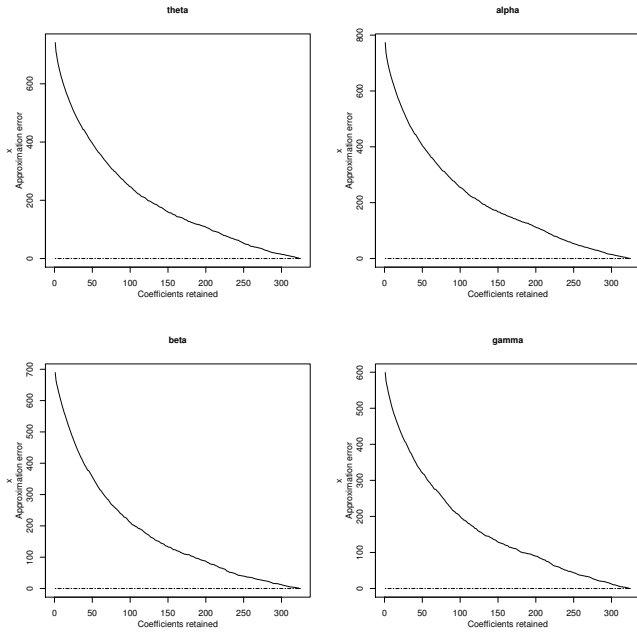


Fig. 2: Compressed spectral bands of the TGUH bases

To illustrate TGUH denoising, we only show the result of denoising the alpha band signal, in Figure 3. There the size of the nodes indicates the strength of the signal. Using the theoretical threshold suggested by Theorem III.1, signal on only 6 nodes in the occipital cortex remain; the rest has been eliminated. The intuition is that the cluster of electrodes with increased alpha power likely represents the posterior dominant rhythm in this area.

V. DISCUSSION & FUTURE RESEARCH

In this paper, we develop a new class of graph wavelets, based on adaptive unbalanced Haar transformations, and

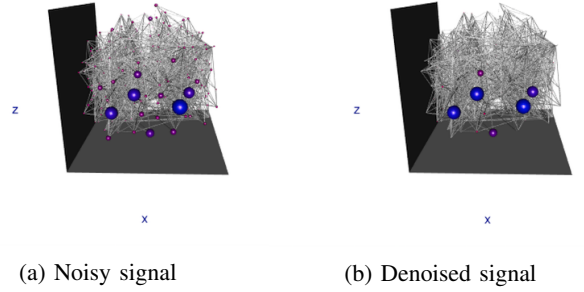


Fig. 3: (a) alpha band of the DTI network; (b) denoised alpha band of the DTI network.

establish consistency for estimating appropriate functions over a network. Theoretical analysis, simulation, and real data application in the context of computational neuroscience suggest the promise of this class. Our algorithm is especially useful in the case where the signal behaves differently at different scales and these scales correspond to analogous variations in the underlying network topology.

Even with the current advancement in this now-highly-active space, the interaction of network topology, basis and signal is still an under-explored area. In future work, denoising of the network itself seems a natural extension here, although theoretical analysis even in toy cases is challenging. Connections to graph coarsening and visualization are possible as well.

VI. ACKNOWLEDGEMENTS

We would like to thank Lauren M Ostrowski and Daniel Y Song for collecting and cleaning the DTI data. This work was supported in part by EPSRC grant no. EP/L014246/1, AFOSR award 12RSL042, and NIH award 1R01NS095369-01.

VII. PROOFS

A. Proof of Theorem III.1

Proof. The fact that $\mathbb{P}(\mathcal{A}) \rightarrow 1$ as $N \rightarrow \infty$ follows from Lemma 1 of [24]. We begin by defining two sets S_0^ℓ and S_1^ℓ with $S_1^\ell = \{1 \leq r \leq k(\ell) : \text{the support of } \psi_r^\ell \text{ crosses multiple regions of constancy at level } \ell\}$ and $S_0^\ell = \{1, \dots, k(\ell)\} \setminus S_1^\ell$.

$$\begin{aligned}
R(\hat{f}, f) &= \frac{1}{N} \sum_v (\hat{f}(v) - f(v))^2 \\
&= \frac{1}{N} \sum_{\ell=1}^L \sum_{r=1}^{k(\ell)} \left(\alpha_r^{\ell, \mathbb{I}} \{ \exists V_{r'}^{\ell'} \subseteq V_r^\ell \mid |\alpha_{r'}^{\ell'}| > \lambda(\ell', r') \} - \mu_r^\ell \right)^2 \\
&\quad + \frac{1}{N} (\alpha_0^0 - \mu_0^0)^2 \\
&= \frac{1}{N} \sum_{\ell=1}^L \left(\sum_{r \in S_0^\ell} + \sum_{r \in S_1^\ell} \right) \left(\alpha_r^{\ell, \mathbb{I}} \{ \exists V_{r'}^{\ell'} \subseteq V_r^\ell \mid |\alpha_{r'}^{\ell'}| > \lambda(\ell', r') \} - \mu_r^\ell \right)^2 \\
&\quad + \frac{1}{N} (\alpha_0^0 - \mu_0^0)^2 \\
&\leq \frac{1}{N} \sum_{\ell=1}^L \left(\sum_{r \in S_0^\ell} + \sum_{r \in S_1^\ell} \right) \left(\alpha_r^{\ell, \mathbb{I}} \{ \exists V_{r'}^{\ell'} \subseteq V_r^\ell \mid |\alpha_{r'}^{\ell'}| > \lambda(\ell', r') \} - \mu_r^\ell \right)^2 \\
&\quad + \frac{2}{N} \log N
\end{aligned}$$

By Lemma VII.1, we have that on the set S_0^ℓ , $|\alpha_r^\ell| \leq \sqrt{2 \log N} \left\{ \frac{\sqrt{|V_m^{\ell-1}|} + \sqrt{|V_{m'}^{\ell-1}|}}{\sqrt{|V_m^{\ell-1}| + |V_{m'}^{\ell-1}|}} \right\}$. We then have

$$R(\hat{f}, f) \leq \frac{1}{N} \sum_{\ell=1}^L \sum_{r \in S_1^\ell} \left(\alpha_r^\ell \mathbb{I} \left\{ \exists V_{r'}^\ell \subseteq V_r^\ell \mid |\alpha_{r'}^\ell| > \lambda(\ell', r') \right\} - \mu_r^\ell \right)^2 + \frac{2}{N} \log N.$$

Denote by \mathcal{E} the event

$$\left\{ \exists V_{r'}^\ell \subseteq V_r^\ell \mid |\alpha_{r'}^\ell| > \sqrt{2 \log N} \left\{ \frac{\sqrt{|V_m^{\ell-1}|} + \sqrt{|V_{m'}^{\ell-1}|}}{\sqrt{|V_m^{\ell-1}| + |V_{m'}^{\ell-1}|}} \right\} \right\}.$$

We compute

$$\begin{aligned} (\alpha_r^\ell \mathbb{I}(\mathcal{E}) - \mu_r^\ell)^2 &= (\alpha_r^\ell \mathbb{I}(\mathcal{E}) - \alpha_r^\ell + \alpha_r^\ell - \mu_r^\ell)^2 \\ &\leq (\alpha_r^\ell)^2 \mathbb{I}(\neg \mathcal{E}) + (\alpha_r^\ell - \mu_r^\ell)^2 + 2|\alpha_r^\ell \mathbb{I}(\neg \mathcal{E})| |\alpha_r^\ell - \mu_r^\ell| \\ &\leq \lambda^2 + 2\lambda \sqrt{2 \log N} + 2 \log N \\ &\leq (6 + 4\sqrt{2}) \log N. \end{aligned}$$

Note that the level L associated with the TGUH transformation is bounded, i.e., $L \leq \log N / \log(1 - \rho)^{-1}$. Combining this with the fact that $|S_\ell^1| \leq K$, and the assumption that $K = o(N / \log^2 N)$, we have that $R(\hat{f}, f)$ is of order $\mathcal{O}\left(\frac{K \log^2 N}{N \log(1 - \rho)^{-1}}\right)$. \square

Lemma VII.1. Let $S_0^\ell = \{1 \leq m \leq k(\ell) : \mu_r^\ell = 0\}$. On \mathcal{A} , for $\ell = 1, \dots, L$, $k \in S_0^\ell$, we have

$$|\alpha_r^\ell| \leq \sqrt{2 \log N} \left\{ \frac{\sqrt{|V_m^{\ell-1}|} + \sqrt{|V_{m'}^{\ell-1}|}}{\sqrt{|V_m^{\ell-1}| + |V_{m'}^{\ell-1}|}} \right\}.$$

Proof. Denote the two sub-regions which merge into V_r^ℓ in the next level as $V_m^{\ell-1}$ and $V_{m'}^{\ell-1}$. On \mathcal{A} , for $\ell = 1, \dots, L$, $k \in S_0^\ell$, we have

$$\begin{aligned} |\alpha_r^\ell| &= \left| \left\{ \frac{|V_{m'}^{\ell-1}|}{|V_r^\ell|} \right\}^{1/2} \frac{\sum_{v \in V_{m'}^{\ell-1}} \epsilon(v)}{\sqrt{|V_m^{\ell-1}|}} - \left\{ \frac{|V_m^{\ell-1}|}{|V_r^\ell|} \right\}^{1/2} \frac{\sum_{v \in V_m^{\ell-1}} \epsilon(v)}{\sqrt{|V_{m'}^{\ell-1}|}} \right| \\ &\leq \sqrt{2 \log N} \left(\left\{ \frac{|V_{m'}^{\ell-1}|}{|V_r^\ell|} \right\}^{1/2} + \left\{ \frac{|V_m^{\ell-1}|}{|V_r^\ell|} \right\}^{1/2} \right) \\ &= \sqrt{2 \log N} \left\{ \frac{|V_m^{\ell-1}|^{1/2} + |V_{m'}^{\ell-1}|^{1/2}}{|V_r^\ell|^{1/2}} \right\}. \end{aligned}$$

\square

REFERENCES

- [1] T. E. Behrens, H. J. Berg, S. Jbabdi, M. F. Rushworth, and M. W. Woolrich, "Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?" *Neuroimage*, vol. 34, no. 1, pp. 144–155, 2007.
- [2] G. Buzsáki, "Rhythms of the Brain". Oxford University Press, 2006.
- [3] S. Chen, R. Varma, A. Singh, J. Kovačević, "Representations of piecewise smooth signals on graphs", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 6370–6374, 2016.
- [4] C. J. Chu, N. Tanaka, J. Diaz, B. L. Edlow, O. Wu, M. Hämäläinen, S. Stufflebeam, S. S. Cash, and M. A. Kramer, "Eeg functional connectivity is partially predicted by underlying white matter connectivity," *Neuroimage*, vol. 108, pp. 23–33, 2015.
- [5] R. Coifman and W. Leeb, "Earth mover's distance and equivalent metrics for spaces with hierarchical partition trees," 2013.
- [6] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 53–94, 2006.
- [7] M. Crovella and E. Kolaczyk, "Graph wavelets for spatial traffic analysis," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 3. IEEE, pp. 1848–1857.
- [8] P. Fryzlewicz, "Tail-greedy bottom-up data decompositions and fast multiple change-point detection," *Annals of Statistics*, to appear, 2017.
- [9] B. Fischl, "Freesurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.
- [10] P. Fryzlewicz and C. Timmermans, "Shah: Shape-adaptive haar wavelets for image processing," *Journal of Computational and Graphical Statistics*, vol. 25, no. 3, pp. 879–898, 2016.
- [11] 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 *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 367–374.
- [12] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen *et al.*, "Meg and eeg data analysis with mne-python," *Frontiers in neuroscience*, vol. 7, 2013.
- [13] M. S. Hamalainen and J. Sarvas, "Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data," *IEEE transactions on biomedical engineering*, vol. 36, no. 2, pp. 165–171, 1989.
- [14] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.
- [15] J. Iriou and N. Saito, "Efficient approximation and denoising of graph signals using the multiscale basis dictionaries," *IEEE Transactions on Signal and Information Processing over Networks*, 2017.
- [16] E.D. Kolaczyk, "Statistical Analysis of Network Data: Methods and Models." Springer New York, 2009.
- [17] N. J. Kopell, H. J. Gritton, M. A. Whittington, and M. A. Kramer, "Beyond the connectome: the dynome," *Neuron*, vol. 83, no. 6, pp. 1319–1328, 2014.
- [18] R. Kondor, N. Teneva, and V. Garg, "Multiresolution matrix factorization," in *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, 2014, pp. 1620–1628.
- [19] A. B. Lee, B. Nadler, and L. Wasserman, "Treelets: an adaptive multi-scale basis for sparse unordered data," *The Annals of Applied Statistics*, pp. 435–471, 2008.
- [20] F. Murtagh, "The haar wavelet transform of a dendrogram," *Journal of Classification*, vol. 24, no. 1, pp. 3–32, 2007.
- [21] A. Sandryhaila and J.M.F. Moura, "Discrete signal processing on graphs," *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1644–1656, 2013.
- [22] 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 Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.
- [23] A. Singh, R. Nowak, and R. Calderbank, "Detecting weak but hierarchically-structured patterns in networks", *Proceedings of the Thirtieth International Conference on Artificial Intelligence and Statistics (AISTATS)*, 749–756, 2010.
- [24] Y-C. Yao, "Estimating the number of change-points via Schwarz' criterion," *Statistics and Probability Letters*, vol. 6, no. 3, pp. 181–189, 1988.