



Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Graph Fourier transform based on ℓ_1 norm variation minimizationLihua Yang^{a,b}, Anna Qi^a, Chao Huang^{c,d}, Jianfeng Huang^{e,*}^a School of Mathematics, Sun Yat-sen University, Guangzhou 510275, China^b Guangdong Provincial Key Laboratory of Computational Science, Sun Yat-sen University, Guangzhou 510275, China^c College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China^d Shenzhen Key Laboratory of Advanced Machine Learning and Applications, Shenzhen University, Shenzhen 518060, China^e School of Financial Mathematics & Statistics, Guangdong University of Finance, Guangzhou 510521, China

ARTICLE INFO

Article history:

Received 17 June 2018

Received in revised form 7 October 2019

Accepted 9 April 2020

Available online xxxx

Communicated by Amit Singer

Keywords:

Graph signal processing

Graph Fourier transform

Signal variation

 ℓ_1 norm minimization

ABSTRACT

The definition of graph Fourier transform is a fundamental issue in graph signal processing. Conventional graph Fourier transform is defined through the eigenvectors of the graph Laplacian matrix, which minimize the ℓ_2 norm signal variation. In this paper, we propose a generalized definition of graph Fourier transform based on the ℓ_1 norm variation minimization. We obtain a necessary condition satisfied by the ℓ_1 Fourier basis, and provide a fast greedy algorithm to approximately construct the ℓ_1 Fourier basis. Numerical experiments are also provided to compare the differences between the Laplacian basis, ℓ_1 basis and greedy basis.

© 2020 Elsevier Inc. All rights reserved.

1. Introduction

1.1. Graph Fourier transform

In applications such as social, transportation, sensor and neural networks, high-dimensional data is usually defined on the vertices of weighted graphs [3]. To process signals on graphs, traditional theories and methods established on the Euclidean domain need to be extended to the graph setting. There are many works in this area in recent years, including spectral graph theory [1], Fourier transform for directed graphs [5,11], short-time Fourier transform on graphs [4], uncertainty principle [6], wavelets on graphs [7–10], graph sampling theory [14].

* Corresponding author.

E-mail addresses: mcsylh@mail.sysu.edu.cn (L. Yang), 1350561656@qq.com (A. Qi), hchao@szu.edu.cn (C. Huang), 47-072@gdudf.edu.cn (J. Huang).

<https://doi.org/10.1016/j.acha.2020.04.001>

1063-5203/© 2020 Elsevier Inc. All rights reserved.

The definition of graph Fourier transform plays a central role in graph signal processing. By graph Fourier transform, a graph signal is decomposed into different spectral components and thus can be analyzed in the Fourier domain. One of the most commonly used definition of graph Fourier transform is via the eigenvectors of the graph Laplacian matrix. Although this definition has solid theoretical foundation, it has some practical limitations. First, the Laplacian definition only applies to undirected graphs, since for directed graphs the Laplacian matrix is not symmetric and does not have an orthonormal system of eigenvectors. Second, the eigen-decomposition needs $O(N^3)$ computations which is rather expensive. Third, the entries of Laplacian eigenvectors are dense. But in applications such as classification and community detection, sparse representations are more favored. Therefore, it is tempting to find a generalized definition of graph Fourier transform without these disadvantages.

One basic requirement for the Fourier basis is that the basis vectors should represent a range of different oscillating frequencies. For a time-domain signal, the classical Fourier transform decomposes it into different frequency components. Likewise, in the graph setting, one expects the graph Fourier basis to have a similar property. Generally speaking, the magnitude of oscillation of a signal can be measured by its variation. In fact, the ℓ_2 norm variation of the Laplacian eigenvectors \mathbf{u}_k^{LP} is characterized by the corresponding eigenvalue λ_k . When the eigenvalues λ_k are arranged in ascending order, the ℓ_2 norm variation of the eigenvector \mathbf{u}_k^{LP} will be increasing with k , thus representing a range of frequencies from low to high. Moreover, the eigenvector \mathbf{u}_k^{LP} minimizes the ℓ_2 norm variation in the subspace orthogonal to the span of the previous $k-1$ eigenvectors. Therefore \mathbf{u}_k^{LP} can be equivalently defined via solving an ℓ_2 norm variation minimization problem.

Recently, S. Sardellitti et al. proposed a definition of directed graph Fourier basis as the set of N orthonormal vectors minimizing the graph directed variation, and proposed two algorithms (SOC and PAMAL) to solve the related optimization problem [11]. However, there is a lack of theoretic analysis of the proposed Fourier basis, and the complexity of the proposed algorithms are rather high. Different from Sardellitti's approach, we propose a definition of ℓ_1 Fourier basis based on iteratively solving a sequence of ℓ_1 norm variation minimization problems. We rigorously prove a necessary condition satisfied by the ℓ_1 Fourier basis. Furthermore, we provide a fast greedy algorithm to approximately construct the ℓ_1 Fourier basis. Numerical experiments show the greedy algorithm is effective, and the n -term approximation errors under the greedy basis decay rapidly for simulated and real-world signals. Experiments also show that the greedy basis is more suitable in representing piecewise constant signals than the Laplacian basis.

The rest of the paper is organized as follows. In Section 2, we discuss the relation between graph Fourier basis and signal variation, and propose the definition of ℓ_1 Fourier basis based on ℓ_1 norm variation minimization. In Section 3, we prove a necessary condition of ℓ_1 Fourier basis, showing that the components of the k th basis vector $\mathbf{u}_k^{\ell_1}$ have at most k different values. In Section 4, we provide a fast greedy algorithm to approximately construct the ℓ_1 basis. In Section 5, we present some numerical experiment results. Section 6 is a final conclusion.

1.2. Notations

In this paper we use the following notations. We use \top to denote the transpose of a vector or a matrix. For a vector $\mathbf{x} = [x_1, \dots, x_n]^\top \in \mathbb{R}^n$, $\|\mathbf{x}\|$ denotes its ℓ_2 norm, i.e., $\|\mathbf{x}\| = (\sum_{i=1}^n |x_i|^2)^{1/2}$. Denote by $\mathbb{B}(\mathbf{x}, \varepsilon) := \{\mathbf{x}' \mid \|\mathbf{x} - \mathbf{x}'\| < \varepsilon\}$ the open ball centered at \mathbf{x} with radius $\varepsilon > 0$. For a matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$, $\|\mathbf{M}\|$ denotes its operator norm, i.e., $\sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{M}\mathbf{x}\|}{\|\mathbf{x}\|}$; $\ker \mathbf{M}$ denotes its kernel, i.e., $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{M}\mathbf{x} = \mathbf{0}\}$; $\text{span } \mathbf{M}$ denotes its column space, i.e., $\{\mathbf{M}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$. If \mathcal{A} is a set of vectors, $\text{span } \mathcal{A}$ denotes the set of finite linear combinations of elements in \mathcal{A} , i.e., $\{\sum_{k=1}^n c_k \mathbf{a}_k \mid \mathbf{a}_k \in \mathcal{A}, c_k \in \mathbb{R}, n \in \mathbb{N}\}$. The cardinality of a set \mathcal{A} is denoted by $|\mathcal{A}|$. Let N be a positive integer, and $\mathcal{V} = \{1, \dots, N\}$. For any $\mathcal{A} \subset \mathcal{V}$, we use $\mathbf{1}_{\mathcal{A}} \in \mathbb{R}^N$ to denote the indicating vector of \mathcal{A} , i.e., $\mathbf{1}_{\mathcal{A}}(i) = 1$ if $i \in \mathcal{A}$ and $\mathbf{1}_{\mathcal{A}}(i) = 0$ otherwise. In particular, $\mathbf{1}_{\mathcal{V}}$ is also written as $\mathbf{1}$.

2. Graph Fourier basis and signal variation

In this section, we shall discuss the relationship between the graph Fourier basis and signal variation. Let us begin with the basic terminology of graph signal processing. Let $\mathcal{G} = (\mathcal{V}, \mathbf{W})$ be a connected, undirected, and weighted graph, where the vertices set $\mathcal{V} = \{1, 2, \dots, N\}$ and the weight matrix $\mathbf{W} = [w_{ij}] \in \mathbb{R}^{N \times N}$ satisfying $w_{ij} = w_{ji} \geq 0$. The degree of the i th vertex is defined as $d_i = \sum_{j=1}^N w_{ij}$, and the degree matrix $\mathbf{D} = \text{diag}(d_1, \dots, d_N)$. The combinatorial Laplacian matrix is defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$. Since \mathbf{L} is symmetric and positive semi-definite, it has eigenvalues $0 = \lambda_1 \leq \dots \leq \lambda_N$ and a corresponding set of orthonormal eigenvectors $\{\mathbf{u}_1^{\text{Lp}}, \dots, \mathbf{u}_N^{\text{Lp}}\}$. We call the orthonormal matrix $\mathbf{U}^{\text{Lp}} := [\mathbf{u}_1^{\text{Lp}}, \dots, \mathbf{u}_N^{\text{Lp}}] \in \mathbb{R}^{N \times N}$ the *Laplacian basis* of \mathcal{G} . A graph signal \mathbf{x} is a real-valued function defined on \mathcal{V} , and can be regarded as a vector in \mathbb{R}^N . The Fourier transform of \mathbf{x} under the Laplacian basis is defined as $(\mathbf{U}^{\text{Lp}})^\top \mathbf{x}$.

Note that the ℓ_2 norm variation of the Laplacian eigenvector \mathbf{u}_k^{Lp} is increasing with k . To see this, let $\mathbf{x} = [x_1, \dots, x_N]^\top \in \mathbb{R}^N$, then it can be proved that

$$\mathbf{x}^\top \mathbf{L} \mathbf{x} = \sum_{1 \leq i < j \leq N} w_{ij} |x_i - x_j|^2. \quad (1)$$

That means the quadratic form $\mathbf{x}^\top \mathbf{L} \mathbf{x}$ gives a measure of the ℓ_2 norm variation of \mathbf{x} . Since $(\mathbf{u}_k^{\text{Lp}})^\top \mathbf{L} \mathbf{u}_k^{\text{Lp}} = \lambda_k$ and λ_k is increasing, we have

$$(\mathbf{u}_1^{\text{Lp}})^\top \mathbf{L} \mathbf{u}_1^{\text{Lp}} \leq \dots \leq (\mathbf{u}_N^{\text{Lp}})^\top \mathbf{L} \mathbf{u}_N^{\text{Lp}},$$

i.e., the ℓ_2 norm variation of \mathbf{u}_k^{Lp} is increasing with k . Note that the variation reflects the magnitude of oscillation. Therefore, the Laplacian eigenvectors $\{\mathbf{u}_k^{\text{Lp}} \mid 1 \leq k \leq N\}$ represent a range of oscillating frequencies from low to high.

Furthermore, the Laplacian eigenvector \mathbf{u}_k^{Lp} minimizes the ℓ_2 norm variation in the subspace orthogonal to the span of the previous $k - 1$ eigenvectors [1], i.e.,

$$\begin{aligned} \mathbf{u}_k^{\text{Lp}} &= \arg \min_{\mathbf{x} \in \mathbb{R}^N} \mathbf{x}^\top \mathbf{L} \mathbf{x} \\ \text{s. t. } & [\mathbf{u}_1^{\text{Lp}}, \dots, \mathbf{u}_{k-1}^{\text{Lp}}]^\top \mathbf{x} = \mathbf{0}, \quad \|\mathbf{x}\| = 1. \end{aligned} \quad (2)$$

In fact, suppose $\mathbf{x} \in \mathbb{R}^N$ satisfy the constraints $[\mathbf{u}_1^{\text{Lp}}, \dots, \mathbf{u}_{k-1}^{\text{Lp}}]^\top \mathbf{x} = \mathbf{0}$ and $\|\mathbf{x}\| = 1$. Let the Fourier transform of \mathbf{x} be $\hat{\mathbf{x}} = (\mathbf{U}^{\text{Lp}})^\top \mathbf{x} = [\hat{x}_1, \dots, \hat{x}_N]^\top$. Then \mathbf{x} can be expressed as $\sum_{j=k}^N \hat{x}_j \mathbf{u}_j^{\text{Lp}}$, hence

$$\mathbf{x}^\top \mathbf{L} \mathbf{x} = \sum_{j=k}^N \lambda_j |\hat{x}_j|^2 \geq \lambda_k \sum_{j=k}^N |\hat{x}_j|^2 = \lambda_k = (\mathbf{u}_k^{\text{Lp}})^\top \mathbf{L} \mathbf{u}_k^{\text{Lp}}.$$

That means \mathbf{u}_k^{Lp} solves the ℓ_2 norm variation minimization problem (2) for $k = 2, \dots, N$.

It is natural to consider using the more general ℓ_p norm to measure the variation of a signal. In this paper, we are interested in the ℓ_1 norm variation defined as follows,

$$S(\mathbf{x}) := \sum_{1 \leq i < j \leq N} w_{ij} |x_i - x_j|. \quad (3)$$

By solving the ℓ_1 norm variation minimization problem, we can define a new type of Fourier basis.

Definition 1. Let $\mathbf{u}_1^{\ell_1} := \frac{1}{\sqrt{N}}\mathbf{1}$. For $k = 2, \dots, N$, define

$$\begin{aligned} \mathbf{u}_k^{\ell_1} &:= \arg \min_{\mathbf{x} \in \mathbb{R}^N} S(\mathbf{x}) \\ \text{s. t. } &[\mathbf{u}_1^{\ell_1}, \dots, \mathbf{u}_{k-1}^{\ell_1}]^\top \mathbf{x} = \mathbf{0}, \quad \|\mathbf{x}\| = 1. \end{aligned} \quad (4)$$

We call the orthonormal matrix $\mathbf{U}^{\ell_1} := [\mathbf{u}_1^{\ell_1}, \dots, \mathbf{u}_N^{\ell_1}] \in \mathbb{R}^{N \times N}$ an ℓ_1 Fourier basis of the graph \mathcal{G} .

Remark 1. The above definition can be generalized to the ℓ_p cases. Consider the ℓ_p norm variation $S_p(\mathbf{x}) := \left(\sum_{i < j} w_{ij} |x_i - x_j|^p\right)^{1/p}$, and define the ℓ_p Fourier basis $\mathbf{u}_k^{\ell_p}$ as

$$\begin{aligned} \mathbf{u}_k^{\ell_p} &:= \arg \min_{\mathbf{x} \in \mathbb{R}^N} S_p(\mathbf{x}) \\ \text{s. t. } &[\mathbf{u}_1^{\ell_p}, \dots, \mathbf{u}_{k-1}^{\ell_p}]^\top \mathbf{x} = \mathbf{0}, \quad \|\mathbf{x}\| = 1, \end{aligned} \quad (5)$$

for $k = 2, \dots, N$. When $p = 2$, the ℓ_p Fourier basis coincides with the Laplacian basis \mathbf{U}^{LP} .

Remark 2. It is not hard to see that the ℓ_p Fourier basis $\mathbf{u}_k^{\ell_p}$ defined in Equation (5) has increasing ℓ_p norm variation, i.e.,

$$S_p(\mathbf{u}_1^{\ell_p}) \leq S_p(\mathbf{u}_2^{\ell_p}) \leq \dots \leq S_p(\mathbf{u}_N^{\ell_p}). \quad (6)$$

That means the oscillating frequency of $\mathbf{u}_k^{\ell_p}$ increases from low to high as k grows, which is consistent with the usual comprehension of Fourier basis.

Remark 3. The definition of ℓ_1 Fourier basis can be also extended to directed graphs. All one needs is to replace the ℓ_1 norm variation $S(\mathbf{x})$ by its directed version

$$\tilde{S}(\mathbf{x}) := \sum_{1 \leq i, j \leq N} w_{ij} (x_i - x_j)_+, \quad (7)$$

where $(x_i - x_j)_+ = \max\{x_i - x_j, 0\}$ (more details can be found in [11]). Then one can similarly define the directed ℓ_1 Fourier basis as the solutions of the directed variation minimization problems. Without loss of generality, we only consider undirected graphs in this paper. Many theoretical results can be extended to the directed case without essential difficulties.

3. ℓ_1 Fourier basis

In the previous section, the ℓ_1 Fourier basis vectors are defined as the solutions of $N - 1$ optimization problems, which can be written in a more compact form:

$$\mathcal{P}_{\mathbf{U}} := \begin{cases} \min_{\mathbf{x} \in \mathbb{R}^N} & S(\mathbf{x}) \\ \text{s. t. } & \mathbf{U}^\top \mathbf{x} = \mathbf{0}, \quad \|\mathbf{x}\| = 1 \end{cases} \quad (8)$$

where \mathbf{U} is a matrix with its first column being $\frac{1}{\sqrt{N}}\mathbf{1}$. With this notation, the definition of ℓ_1 basis can be simplified to $\mathbf{u}_k^{\ell_1} := \arg \mathcal{P}_{[\mathbf{u}_1^{\ell_1}, \dots, \mathbf{u}_{k-1}^{\ell_1}]}$ for $k = 2, \dots, N$.

For convenience, we denote the feasible region of problem $\mathcal{P}_{\mathbf{U}}$ by $\mathcal{X}_{\mathbf{U}}$, i.e.,

$$\mathcal{X}_{\mathbf{U}} := \{\mathbf{x} \in \mathbb{R}^N \mid \mathbf{U}^\top \mathbf{x} = \mathbf{0}, \|\mathbf{x}\| = 1\}, \quad (9)$$

and denote the set of all local minimizers of problem $\mathcal{P}_{\mathbf{U}}$ by $\mathcal{X}_{\mathbf{U}}^{**}$, i.e.,

$$\mathcal{X}_{\mathbf{U}}^{**} := \{\mathbf{x} \in \mathcal{X}_{\mathbf{U}} \mid \exists \varepsilon > 0 \text{ such that } S(\mathbf{x}') \geq S(\mathbf{x}), \forall \mathbf{x}' \in \mathcal{X}_{\mathbf{U}} \cap \mathbb{B}(\mathbf{x}, \varepsilon)\}. \quad (10)$$

Obviously, if \mathbf{x} is a global minimizer of problem $\mathcal{P}_{\mathbf{U}}$, then $\mathbf{x} \in \mathcal{X}_{\mathbf{U}}^{**}$.

Due to the sphere constraint $\|\mathbf{x}\| = 1$, problem $\mathcal{P}_{\mathbf{U}}$ is a nonconvex and nonsmooth optimization problem. As far as we know, there are no general results about the global minimizer of such problems, and in many cases it is only possible to approach the local minimizer by iterative algorithms [12,13]. As the main result of this section, we shall prove a necessary condition satisfied by the local minimizer of problem $\mathcal{P}_{\mathbf{U}}$ (Theorem 4). The key ingredient is the concept of partition matrix introduced as follows.

Definition 2. Let $\mathbf{x} = [x_1, \dots, x_N]^T \in \mathbb{R}^N$. Suppose the components of \mathbf{x} have m different values which can be sorted as $x_{(1)} < x_{(2)} < \dots < x_{(m)}$. Let $\mathcal{A}_j := \{i \mid 1 \leq i \leq N, x_i = x_{(j)}\}$, $\mathbf{M} := [\mathbf{1}_{\mathcal{A}_1}, \dots, \mathbf{1}_{\mathcal{A}_m}] \in \mathbb{R}^{N \times m}$ and $\mathbf{a} := [x_{(1)}, \dots, x_{(m)}]^T \in \mathbb{R}^m$. Then $\mathbf{x} = \mathbf{M}\mathbf{a}$. We call \mathbf{M} the *partition matrix* of \mathbf{x} , and denote $\phi(\mathbf{x}) = \mathbf{M}$.

Through partition matrix, the ℓ_1 norm variation $S(\mathbf{x})$ can be locally simplified to a linear form. For any subsets $\mathcal{A}, \mathcal{B} \subset \{1, \dots, N\}$, denote $\mathbf{W}(\mathcal{A}, \mathcal{B}) := \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{B}} w_{ij}$. Then we have the following lemma.

Lemma 3. Suppose $\mathbf{x} \in \mathbb{R}^N$, $\phi(\mathbf{x}) = \mathbf{M} = [\mathbf{1}_{\mathcal{A}_1}, \dots, \mathbf{1}_{\mathcal{A}_m}]$, $\mathbf{x} = \mathbf{M}\mathbf{a}$ and $m \geq 2$. Then there exists $\varepsilon > 0$ and $\mathbf{f} = [f_1, \dots, f_m]^T \in \mathbb{R}^m$ such that

$$S(\mathbf{M}\mathbf{a}') = \mathbf{f}^T \mathbf{a}', \quad \forall \mathbf{a}' \in \mathbb{B}(\mathbf{a}, \varepsilon), \quad (11)$$

where

$$f_i := \sum_{j=1}^{i-1} \mathbf{W}(\mathcal{A}_i, \mathcal{A}_j) - \sum_{j=i+1}^m \mathbf{W}(\mathcal{A}_i, \mathcal{A}_j), \quad i = 1, \dots, m. \quad (12)$$

Proof. Suppose $\mathbf{a} = [a_1, \dots, a_m]^T$, then $a_1 < \dots < a_m$. Let $\mathbf{a}' = [a'_1, \dots, a'_m]^T$. When $\|\mathbf{a} - \mathbf{a}'\|$ is sufficiently small, it holds $a'_1 < \dots < a'_m$, i.e., there exists $\varepsilon > 0$ such that for any $\mathbf{a}' \in \mathbb{B}(\mathbf{a}, \varepsilon)$, \mathbf{M} is the partition matrix of $\mathbf{M}\mathbf{a}'$. Suppose $\mathbf{M}\mathbf{a}' = [x'_1, \dots, x'_N]^T$. Then

$$\begin{aligned} S(\mathbf{M}\mathbf{a}') &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N w_{ij} |x'_i - x'_j| \\ &= \frac{1}{2} \sum_{i=1}^m \sum_{p \in \mathcal{A}_i} \sum_{j=1}^m \sum_{q \in \mathcal{A}_j} w_{pq} |x'_p - x'_q| \\ &= \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m |a'_i - a'_j| \sum_{p \in \mathcal{A}_i} \sum_{q \in \mathcal{A}_j} w_{pq} \\ &= \sum_{1 \leq i < j \leq m} (a'_j - a'_i) \mathbf{W}(\mathcal{A}_i, \mathcal{A}_j) \\ &= \mathbf{f}^T \mathbf{a}'. \quad \square \end{aligned}$$

Based on Lemma 3, we can prove a necessary condition satisfied by the local minimizer of problem $\mathcal{P}_{\mathbf{U}}$.

Theorem 4. If $\mathbf{x} \in \mathcal{X}_{\mathbf{U}}^{**}$ and $\phi(\mathbf{x}) = \mathbf{M}$, then

$$\dim \ker (\mathbf{U}^T \mathbf{M}) = 1. \quad (13)$$

Proof. The main idea is to transform problem $\mathcal{P}_{\mathbf{U}}$ to an easier one. Suppose $\mathbf{M} = [\mathbf{1}_{\mathcal{A}_1}, \dots, \mathbf{1}_{\mathcal{A}_m}]$ and $\mathbf{x} = \mathbf{M}\mathbf{a}$. By assumption of problem $\mathcal{P}_{\mathbf{U}}$, we have $\langle \mathbf{x}, \mathbf{1} \rangle = 0$ and $\|\mathbf{x}\| = 1$, therefore \mathbf{x} is a non-constant signal, i.e., $m \geq 2$. Since \mathbf{x} is a local minimizer of $\mathcal{P}_{\mathbf{U}}$, there exists $\varepsilon_1 > 0$ such that

$$\begin{aligned} \mathbf{x} &= \arg \min_{\mathbf{x}' \in \mathbb{R}^N} S(\mathbf{x}') \\ \text{s. t. } &\mathbf{U}^\top \mathbf{x}' = \mathbf{0}, \|\mathbf{x}'\| = 1, \mathbf{x}' \in \mathbb{B}(\mathbf{x}, \varepsilon_1). \end{aligned} \quad (14)$$

By Lemma 3, there exists $\varepsilon_2 > 0$ and $\mathbf{f} \in \mathbb{R}^m$ such that $S(\mathbf{x}') = \mathbf{f}^\top \mathbf{a}'$ for all $\mathbf{a}' \in \mathbb{B}(\mathbf{a}, \varepsilon_2)$ and $\mathbf{x}' = \mathbf{M}\mathbf{a}'$. Let $\varepsilon := \min\{\varepsilon_1/\|\mathbf{M}\|, \varepsilon_2\}$, then $\mathbf{a}' \in \mathbb{B}(\mathbf{a}, \varepsilon)$ implies $\mathbf{x}' \in \mathbb{B}(\mathbf{x}, \varepsilon_1)$ and $S(\mathbf{x}') = \mathbf{f}^\top \mathbf{a}'$. Let $\mathbf{\Lambda} := \mathbf{M}^\top \mathbf{M} = \text{diag}(|\mathcal{A}_1|, \dots, |\mathcal{A}_m|)$, then $\mathbf{a}'^\top \mathbf{\Lambda} \mathbf{a}' = 1$ implies $\|\mathbf{x}'\| = 1$, and $\mathbf{U}^\top \mathbf{M}\mathbf{a}' = \mathbf{0}$ implies $\mathbf{U}^\top \mathbf{x}' = \mathbf{0}$. Therefore

$$\begin{aligned} \mathbf{a} &= \arg \min_{\mathbf{a}' \in \mathbb{R}^m} \mathbf{f}^\top \mathbf{a}' \\ \text{s. t. } &\mathbf{U}^\top \mathbf{M}\mathbf{a}' = \mathbf{0}, \mathbf{a}'^\top \mathbf{\Lambda} \mathbf{a}' = 1, \mathbf{a}' \in \mathbb{B}(\mathbf{a}, \varepsilon). \end{aligned} \quad (15)$$

Suppose $\dim \ker(\mathbf{U}^\top \mathbf{M}) = l$ and $\mathbf{V} \in \mathbb{R}^{m \times l}$ is an orthonormal basis of $\ker(\mathbf{U}^\top \mathbf{M})$. Define $\mathbf{c} := \mathbf{V}^\top \mathbf{a}$, $\mathbf{g} := \mathbf{V}^\top \mathbf{f}$, $\mathbf{Q} := \mathbf{V}^\top \mathbf{\Lambda} \mathbf{V}$. Note that \mathbf{Q} is symmetric and positive definite. Then we have

$$\begin{aligned} \mathbf{c} &= \arg \min_{\mathbf{c}' \in \mathbb{R}^l} \mathbf{g}^\top \mathbf{c}' \\ \text{s. t. } &\mathbf{c}'^\top \mathbf{Q} \mathbf{c}' = 1, \mathbf{c}' \in \mathbb{B}(\mathbf{c}, \varepsilon). \end{aligned} \quad (16)$$

According to Lagrange multiplier theory (Dimitri P. Bertsekas [2], Proposition 3.1.1), there exists a unique Lagrange multiplier $\mu \in \mathbb{R}$ such that

$$\nabla[\mathbf{g}^\top \mathbf{c} + \mu(\mathbf{c}^\top \mathbf{Q} \mathbf{c} - 1)] = \mathbf{g} + 2\mu \mathbf{Q} \mathbf{c} = \mathbf{0} \quad (17)$$

and

$$\mathbf{y}^\top \nabla^2[\mathbf{g}^\top \mathbf{c} + \mu(\mathbf{c}^\top \mathbf{Q} \mathbf{c} - 1)] \mathbf{y} = 2\mu \mathbf{y}^\top \mathbf{Q} \mathbf{y} \geq 0, \quad \forall \mathbf{y} \in \{\mathbf{y} \mid \mathbf{y}^\top \mathbf{Q} \mathbf{c} = 0\}. \quad (18)$$

Therefore $\mathbf{g} = -2\mu \mathbf{Q} \mathbf{c}$. Since $0 < S(\mathbf{x}) = \mathbf{f}^\top \mathbf{a} = \mathbf{g}^\top \mathbf{c} = -2\mu \mathbf{c}^\top \mathbf{Q} \mathbf{c}$, we have $\mu < 0$. So condition (18) is equivalent to

$$\mathbf{y}^\top \mathbf{Q} \mathbf{y} = 0, \quad \forall \mathbf{y} \in \{\mathbf{y} \mid \mathbf{y}^\top \mathbf{Q} \mathbf{c} = 0\}. \quad (19)$$

i.e. $\mathbf{y} \perp \mathbf{Q} \mathbf{c}$ implies $\mathbf{y} = \mathbf{0}$. This means $\mathbf{Q} \mathbf{c}$ is a one-dimensional vector, i.e. $l = 1$. The proof is complete. \square

We remark that the condition (13) is not a sufficient condition. To get some intuition of this condition, let us consider the special case $k = 2$. When $k = 2$, we have $\mathbf{u}_1^{\ell_1} = \frac{1}{\sqrt{N}} \mathbf{1}$ and $\mathbf{U} = \mathbf{u}_1^{\ell_1}$, so the condition becomes $\dim \ker(\mathbf{1}^\top \mathbf{M}) = 1$. For any signal \mathbf{x} , it has representation $\mathbf{x} = \sum_{i=1}^m a_i \mathbf{1}_{\mathcal{A}_i} = \mathbf{M}\mathbf{a}$, where m is the number of different values of \mathbf{x} 's components. Suppose $m = 3$, i.e., $\mathbf{x} = a_1 \mathbf{1}_{\mathcal{A}_1} + a_2 \mathbf{1}_{\mathcal{A}_2} + a_3 \mathbf{1}_{\mathcal{A}_3}$ and $\mathbf{M} = [\mathbf{1}_{\mathcal{A}_1}, \mathbf{1}_{\mathcal{A}_2}, \mathbf{1}_{\mathcal{A}_3}]$. Then the subspace

$$\ker(\mathbf{1}^\top \mathbf{M}) = \{\tilde{\mathbf{a}} \in \mathbb{R}^3 \mid \mathbf{1}^\top \mathbf{M} \tilde{\mathbf{a}} = 0\} = \{\tilde{\mathbf{a}} \in \mathbb{R}^3 \mid |\mathcal{A}_1| \tilde{a}_1 + |\mathcal{A}_2| \tilde{a}_2 + |\mathcal{A}_3| \tilde{a}_3 = 0\}$$

has dimension 2. That means if \mathbf{x} has three different values, then it violates the necessary condition and is not a local minimizer. Therefore, $\mathbf{u}_2^{\ell_1}$, as the local minimizer of problem $\mathcal{P}_{\mathbf{U}}$ for $k = 2$, must be a two-valued signal. More generally, we can prove the following corollary.

Corollary 5. If $\mathbf{x} \in \mathcal{X}_{\mathbf{U}}^{**}$ and $\text{rank}(\mathbf{U}) = k - 1$, then \mathbf{x} have at most k different components.

Proof. Suppose $\phi(\mathbf{x}) = \mathbf{M} \in \mathbb{R}^{N \times m}$. By Theorem 4, $\dim \ker(\mathbf{U}^\top \mathbf{M}) = 1$. Since

$$\begin{aligned} k - 1 &= \text{rank}(\mathbf{U}) \\ &\geq \text{rank}(\mathbf{M}^\top \mathbf{U}) \\ &= \dim \text{span}(\mathbf{M}^\top \mathbf{U}) \\ &= m - \dim \ker(\mathbf{U}^\top \mathbf{M}) \\ &= m - 1, \end{aligned}$$

we have $m \leq k$. By definition of partition matrix, m is the number of different components of \mathbf{x} . The proof is complete. \square

Corollary 5 says that the k th ℓ_1 basis vector $\mathbf{u}_k^{\ell_1}$, as the global (hence local) minimizer of problem $\mathcal{P}_{[\mathbf{u}_1^{\ell_1}, \dots, \mathbf{u}_{k-1}^{\ell_1}]}$, is at most a k -valued signal. In particular, $\mathbf{u}_1^{\ell_1}$ is a constant signal and $\mathbf{u}_2^{\ell_1}$ is exactly a two-valued signal.

Another implication of condition (13) is the finiteness of the local minimizers set $\mathcal{X}_{\mathbf{U}}^{**}$. Denote the set of partition matrices of vectors in \mathbb{R}^N by

$$\mathcal{M} := \{\mathbf{M} \mid \exists \mathbf{x} \in \mathbb{R}^N \text{ such that } \mathbf{M} = \phi(\mathbf{x})\}. \quad (20)$$

For any $\mathbf{M} \in \mathcal{M}$, \mathbf{M} has at most N columns, and each entry of \mathbf{M} is either 0 or 1. Therefore \mathcal{M} is a finite set. Denote the set of partition matrices satisfying condition (13) by

$$\mathcal{M}_{\mathbf{U}}^* := \{\mathbf{M} \in \mathcal{M} \mid \dim \ker(\mathbf{U}^\top \mathbf{M}) = 1\}. \quad (21)$$

$\mathcal{M}_{\mathbf{U}}^*$ as a subset of \mathcal{M} is also finite.

By Theorem 4, if \mathbf{x} is a local minimizer of problem $\mathcal{P}_{\mathbf{U}}$, then its partition matrix belongs to $\mathcal{M}_{\mathbf{U}}^*$. Conversely, given a partition matrix $\mathbf{M} \in \mathcal{M}_{\mathbf{U}}^*$, we show that there are two \mathbf{x} 's in $\mathcal{X}_{\mathbf{U}}$ corresponding to partition matrix \mathbf{M} .

Theorem 6. If $\mathbf{M} \in \mathcal{M}_{\mathbf{U}}^*$ and $\mathbf{x}, \mathbf{x}' \in \mathcal{X}_{\mathbf{U}} \cap \text{span } \mathbf{M}$, then $\mathbf{x} = \pm \mathbf{x}'$.

Proof. Since $\mathbf{x}, \mathbf{x}' \in \mathcal{X}_{\mathbf{U}} \cap \text{span } \mathbf{M}$, there exist \mathbf{a}, \mathbf{a}' such that $\mathbf{x} = \mathbf{M}\mathbf{a}$ and $\mathbf{x}' = \mathbf{M}\mathbf{a}'$. Then $\mathbf{U}^\top \mathbf{M}\mathbf{a} = \mathbf{U}^\top \mathbf{x} = \mathbf{0}$ and $\mathbf{U}^\top \mathbf{M}\mathbf{a}' = \mathbf{U}^\top \mathbf{x}' = \mathbf{0}$, i.e., $\mathbf{a}, \mathbf{a}' \in \ker(\mathbf{U}^\top \mathbf{M})$. Since $\dim \ker(\mathbf{U}^\top \mathbf{M}) = 1$ and $\mathbf{a}, \mathbf{a}' \neq \mathbf{0}$, there exists $t \in \mathbb{R}$ such that $\mathbf{a} = t\mathbf{a}'$, hence $\mathbf{x} = t\mathbf{x}'$. From $\|\mathbf{x}\| = \|\mathbf{x}'\| = 1$, we have $t = \pm 1$. The proof is complete. \square

Define

$$\mathcal{X}_{\mathbf{U}}^* := \bigcup_{\mathbf{M} \in \mathcal{M}_{\mathbf{U}}^*} (\mathcal{X}_{\mathbf{U}} \cap \text{span } \mathbf{M}). \quad (22)$$

By Theorem 6, for any $\mathbf{M} \in \mathcal{M}_{\mathbf{U}}^*$, $\mathcal{X}_{\mathbf{U}} \cap \text{span } \mathbf{M}$ has only two elements, which differ by a sign. Therefore $|\mathcal{X}_{\mathbf{U}}^*| \leq \sum_{\mathbf{M} \in \mathcal{M}_{\mathbf{U}}^*} |\mathcal{X}_{\mathbf{U}} \cap \text{span } \mathbf{M}| = 2|\mathcal{M}_{\mathbf{U}}^*| < \infty$, i.e., $\mathcal{X}_{\mathbf{U}}^*$ is a finite set.

The local minimizers set $\mathcal{X}_{\mathbf{U}}^{**}$ is a subset of $\mathcal{X}_{\mathbf{U}}^*$. In fact, if $\mathbf{x} \in \mathcal{X}_{\mathbf{U}}^{**}$ and $\phi(\mathbf{x}) = \mathbf{M}$, then $\mathbf{M} \in \mathcal{M}_{\mathbf{U}}^*$ and $\mathbf{x} \in \mathcal{X}_{\mathbf{U}} \cap \text{span } \mathbf{M}$, hence $\mathbf{x} \in \mathcal{X}_{\mathbf{U}}^*$. It follows that $\mathcal{X}_{\mathbf{U}}^{**}$ is also a finite set, i.e., each local minimizer is isolated and the total number of local minimizers is finite. Fig. 1 shows the relations between these sets and

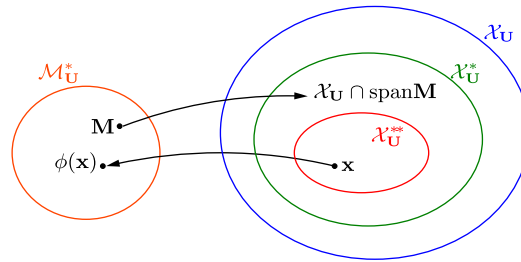
Fig. 1. Relations between \mathcal{M}_U^* , \mathcal{X}_U^{**} , \mathcal{X}_U^* and \mathcal{X}_U .

Table 1

Enumeration of \mathbf{x} in \mathcal{X}_U^* for $N = 4$, $\mathbf{U} = \frac{1}{\sqrt{N}}\mathbf{1}$.

$\mathbf{M} \in \mathcal{M}_U^*$	$\pm \mathbf{x} \in \mathcal{X}_U \cap \text{span } \mathbf{M}$	$S(\mathbf{x})$
$[\mathbf{1}_{\{1\}}, \mathbf{1}_{\{2,3,4\}}]$	$\pm \frac{1}{2\sqrt{3}}[-3, 1, 1, 1]^\top$	$\frac{2}{\sqrt{3}}(w_{12} + w_{13} + w_{14})$
$[\mathbf{1}_{\{2\}}, \mathbf{1}_{\{1,3,4\}}]$	$\pm \frac{1}{2\sqrt{3}}[1, -3, 1, 1]^\top$	$\frac{2}{\sqrt{3}}(w_{12} + w_{23} + w_{24})$
$[\mathbf{1}_{\{3\}}, \mathbf{1}_{\{1,2,4\}}]$	$\pm \frac{1}{2\sqrt{3}}[1, 1, -3, 1]^\top$	$\frac{2}{\sqrt{3}}(w_{13} + w_{23} + w_{34})$
$[\mathbf{1}_{\{4\}}, \mathbf{1}_{\{1,2,3\}}]$	$\pm \frac{1}{2\sqrt{3}}[1, 1, 1, -3]^\top$	$\frac{2}{\sqrt{3}}(w_{14} + w_{24} + w_{34})$
$[\mathbf{1}_{\{1,2\}}, \mathbf{1}_{\{3,4\}}]$	$\pm \frac{1}{2}[-1, -1, 1, 1]^\top$	$w_{13} + w_{14} + w_{23} + w_{24}$
$[\mathbf{1}_{\{1,3\}}, \mathbf{1}_{\{2,4\}}]$	$\pm \frac{1}{2}[-1, 1, -1, 1]^\top$	$w_{12} + w_{14} + w_{23} + w_{34}$
$[\mathbf{1}_{\{1,4\}}, \mathbf{1}_{\{2,3\}}]$	$\pm \frac{1}{2}[-1, 1, 1, -1]^\top$	$w_{12} + w_{13} + w_{23} + w_{24}$

definitions. Here \mathcal{X}_U^* resembles the concept of critical points, which contains but not equals the set of local minimizers.

Since \mathcal{X}_U^* is finite, a way to find the global minimizer of problem \mathcal{P}_U is to compute $S(\mathbf{x})$ for all \mathbf{x} 's in \mathcal{X}_U^* and pick out the smallest value. Table 1 shows a special case for $N = 4$, $\mathbf{U} = \frac{1}{\sqrt{N}}\mathbf{1}$. Through this method of enumeration, the continuous problem \mathcal{P}_U is equivalent to a discrete problem in which the variable \mathbf{x} belongs to a finite set \mathcal{X}_U^* . However, as far as we know, this discrete problem has no effective algorithm. Since the size of \mathcal{X}_U^* grows exponentially with N , the method of enumeration fails for large N . In the next section, we will give a fast greedy algorithm to approximately construct the ℓ_1 Fourier basis when N is large.

4. Greedy algorithm for ℓ_1 Fourier basis

In this section, we provide a fast greedy algorithm to approximately construct the ℓ_1 Fourier basis. Note that the partition matrix of the ℓ_1 basis vector $\mathbf{u}_k^{\ell_1}$ defines a partition of the vertices set \mathcal{V} . When the variation of $\mathbf{u}_k^{\ell_1}$ increases with k , the corresponding partition evolves from coarser to finer scales. On the contrary, given a sequence of partitions varying across different scales, one might be able to construct an orthonormal basis close to the ℓ_1 basis. Motivated by this idea, we propose a greedy algorithm, based on a partition sequence $\{\tau_k\}$ created by iteratively grouping the vertices. In each step, we pick out the two groups of vertices that maximize a mutual weight function F and combine them in a new group. Repeating the process, we get a partition sequence $\{\tau_k\}$ varying from finer to coarser scales. Then based on $\{\tau_k\}$ we define a sequence of subspaces \mathcal{V}_k of \mathbb{R}^N . By using a similar idea of multi-resolution analysis, we obtain an orthonormal basis $\tilde{\mathbf{U}}$.

4.1. Greedy partition sequence

Suppose $\mathcal{A}, \mathcal{B} \subset \mathcal{V}$ are two groups of vertices, and $F(\mathcal{A}, \mathcal{B})$ is a function that measures the mutual weights between \mathcal{A} and \mathcal{B} . Some examples of $F(\mathcal{A}, \mathcal{B})$ are $\mathbf{W}(\mathcal{A}, \mathcal{B})$, $\frac{\mathbf{W}(\mathcal{A}, \mathcal{B})}{|\mathcal{A}| \cdot |\mathcal{B}|}$ and $\frac{\mathbf{W}(\mathcal{A}, \mathcal{B})}{\max\{|\mathcal{A}|, |\mathcal{B}|\}}$. We define a partition sequence $\{\tau_k\}$ on the vertices set \mathcal{V} as follows.

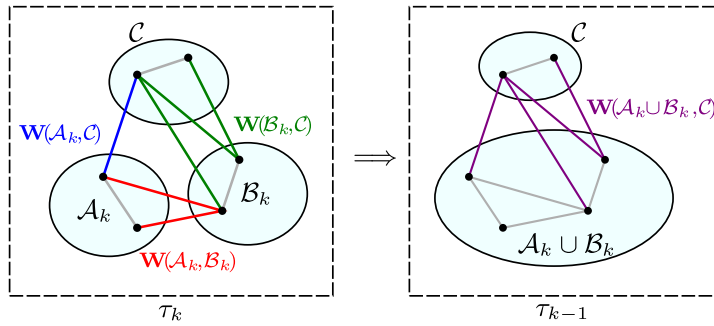


Fig. 2. In step k , we combine \mathcal{A}_k and \mathcal{B}_k of τ_k to get τ_{k-1} .

Definition 7. Let $\tau_N := \{\{1\}, \{2\}, \dots, \{N\}\}$. For $k = N, N-1, \dots, 2$, define

$$\mathcal{A}_k, \mathcal{B}_k := \arg \max_{\mathcal{A}, \mathcal{B} \in \tau_k} F(\mathcal{A}, \mathcal{B}), \quad (23)$$

$$\tau_{k-1} := \{\mathcal{A}_k \cup \mathcal{B}_k\} \cup \{\mathcal{C} \in \tau_k \mid \mathcal{C} \neq \mathcal{A}_k, \mathcal{C} \neq \mathcal{B}_k\}. \quad (24)$$

We call $\{\tau_k\}_{k=1}^N$ a *greedy partition sequence*.

Definition 7 actually represents a vertices grouping process. At the beginning, the finest partition τ_N has N groups, each group having one vertex. To get the next partition τ_{N-1} , we identify $\mathcal{A}_N, \mathcal{B}_N$ as the two groups having the largest mutual weight. Then we combine \mathcal{A}_N and \mathcal{B}_N to get a new group $\mathcal{A}_N \cup \mathcal{B}_N$, and together with the other groups in τ_N we form a new partition τ_{N-1} . This operation repeats for $N-1$ times. At the end, we get the coarsest partition $\tau_1 = \{\{1, 2, \dots, N\}\}$, with all the vertices belonging to a single group. It is easy to see that τ_k has k groups, i.e., $|\tau_k| = k$. See Fig. 2 for an illustration.

4.2. Greedy basis

The greedy partition sequence $\{\tau_k\}_{k=1}^N$ defined above yields a sequence of subspaces

$$\mathcal{V}_k := \text{span} \{\mathbf{1}_{\mathcal{A}} \mid \mathcal{A} \in \tau_k\}, \quad k = 1, \dots, N, \quad (25)$$

which satisfy the relations

$$\text{span } \mathbf{1} = \mathcal{V}_1 \subset \mathcal{V}_2 \subset \dots \subset \mathcal{V}_N = \mathbb{R}^N. \quad (26)$$

Denote the orthogonal complement of \mathcal{V}_{k-1} in \mathcal{V}_k by $\mathcal{V}_k \ominus \mathcal{V}_{k-1}$. By Definition 7, the partition τ_{k-1} is obtained by combining two groups \mathcal{A}_k and \mathcal{B}_k in τ_k . Suppose $\tau_k = \{\mathcal{A}_k, \mathcal{B}_k, \mathcal{C}_1, \dots, \mathcal{C}_{k-2}\}$ and $\tau_{k-1} = \{\mathcal{A}_k \cup \mathcal{B}_k, \mathcal{C}_1, \dots, \mathcal{C}_{k-2}\}$. Let $\mathbf{x} \in \mathcal{V}_k \ominus \mathcal{V}_{k-1}$. Then \mathbf{x} can be written in the form $a\mathbf{1}_{\mathcal{A}_k} + b\mathbf{1}_{\mathcal{B}_k} + \sum_i c_i \mathbf{1}_{\mathcal{C}_i}$. From $\langle \mathbf{x}, \mathbf{1}_{\mathcal{C}_i} \rangle = c_i |\mathcal{C}_i| = 0$, we get $c_i = 0, \forall i = 1, \dots, k-2$. Since

$$\langle \mathbf{x}, \mathbf{1}_{\mathcal{A}_k \cup \mathcal{B}_k} \rangle = a|\mathcal{A}_k| + b|\mathcal{B}_k| = 0,$$

there exists $t \in \mathbb{R}$ such that $a = t|\mathcal{B}_k|, b = -t|\mathcal{A}_k|$. By requiring $\|\mathbf{x}\| = 1$, we get $t = \frac{\pm 1}{\sqrt{|\mathcal{A}_k||\mathcal{B}_k|(|\mathcal{A}_k| + |\mathcal{B}_k|)}}$. We summarize these results in the following theorem.

Theorem 8. Suppose $\mathcal{A}_k, \mathcal{B}_k$ are defined as in Definition 7. Let $\tilde{\mathbf{u}}_1 := \frac{1}{\sqrt{N}}\mathbf{1}$,

$$\tilde{\mathbf{u}}_k := a_k \mathbf{1}_{\mathcal{A}_k} + b_k \mathbf{1}_{\mathcal{B}_k}, \quad k = 2, \dots, N, \quad (27)$$

Table 2
An example of greedy basis.

k	τ_k	\mathcal{A}_k	\mathcal{B}_k	$\tilde{\mathbf{u}}_k$
5	$\{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}$	$\{1\}$	$\{3\}$	$\frac{1}{\sqrt{2}}[-1, 0, 1, 0, 0]^\top$
4	$\{\{1, 3\}, \{2\}, \{4\}, \{5\}\}$	$\{2\}$	$\{5\}$	$\frac{1}{\sqrt{2}}[0, -1, 0, 0, 1]^\top$
3	$\{\{1, 3\}, \{2, 5\}, \{4\}\}$	$\{1, 3\}$	$\{4\}$	$\frac{1}{\sqrt{6}}[-1, 0, -1, 2, 0]^\top$
2	$\{\{1, 3, 4\}, \{2, 5\}\}$	$\{1, 3, 4\}$	$\{2, 5\}$	$\frac{1}{\sqrt{30}}[-2, 3, -2, -2, 3]^\top$
1	$\{\{1, 2, 3, 4, 5\}\}$			$\frac{1}{\sqrt{5}}[1, 1, 1, 1, 1]^\top$

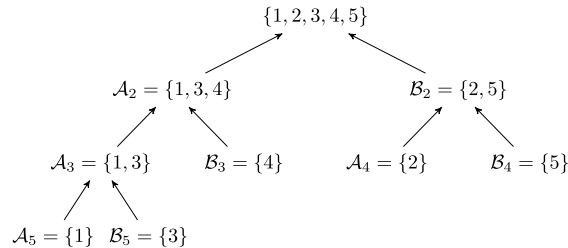


Fig. 3. The binary tree formed by \mathcal{A}_k 's and \mathcal{B}_k 's in the above example.

where

$$a_k := -t_k|\mathcal{B}_k|, \quad b_k := t_k|\mathcal{A}_k|, \quad t_k := \frac{1}{\sqrt{|\mathcal{A}_k||\mathcal{B}_k|(|\mathcal{A}_k| + |\mathcal{B}_k|)}}. \quad (28)$$

Then $\tilde{\mathbf{U}} = [\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_N]$ is orthonormal. We call $\tilde{\mathbf{U}}$ the greedy basis of the graph \mathcal{G} , which is also denoted by \mathbf{U}^{Gr} .

Table 2 shows a concrete example of the greedy basis $\tilde{\mathbf{U}}$ given a partition sequence $\{\tau_k\}$, where the number of vertices $N = 5$. Note that \mathcal{A}_k 's and \mathcal{B}_k 's form a binary tree (Fig. 3).

An interesting question is whether the greedy basis vector $\tilde{\mathbf{u}}_k$ minimizes the ℓ_1 norm variation. We show in the following theorem that the partition matrix \mathbf{M} of $\tilde{\mathbf{u}}_k$ satisfies the necessary condition (13) in Theorem 4.

Theorem 9. Let

$$\tilde{\mathbf{U}}_k := [\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_{k-1}], \quad k = 2, \dots, N, \quad (29)$$

where $\tilde{\mathbf{u}}_k$ is defined in Theorem 8. Let $\mathbf{M} = \phi(\tilde{\mathbf{u}}_k)$. Then $\dim \ker(\tilde{\mathbf{U}}_k^\top \mathbf{M}) = 1$.

Proof. Suppose $\tau_k = \{\mathcal{A}_k, \mathcal{B}_k, \mathcal{C}_1, \dots, \mathcal{C}_{k-2}\}$. Since $\tilde{\mathbf{u}}_k = a_k \mathbf{1}_{\mathcal{A}_k} + b_k \mathbf{1}_{\mathcal{B}_k}$, we have $\mathbf{M} = [\mathbf{1}_{\mathcal{A}_k}, \mathbf{1}_{\mathcal{B}_k}]$ where $\mathcal{C} = \cup_{i=1}^{k-2} \mathcal{C}_i$. Suppose $\mathbf{y} = [a, c, b]^\top \in \ker(\tilde{\mathbf{U}}_k^\top \mathbf{M})$ and $\mathbf{x} = \mathbf{M}\mathbf{y}$. Then $\tilde{\mathbf{U}}_k^\top \mathbf{x} = \tilde{\mathbf{U}}_k^\top \mathbf{M}\mathbf{y} = \mathbf{0}$, i.e., $\mathbf{x} \perp \text{span } \tilde{\mathbf{U}}_k$. Since $\text{span } \tilde{\mathbf{U}}_k = \mathcal{V}_{k-1}$, we have $\mathbf{x} \perp \mathcal{V}_{k-1}$. Because $\mathbf{x} = \mathbf{M}\mathbf{y} \in \text{span}\{\mathbf{1}_{\mathcal{A}} \mid \mathcal{A} \in \tau_k\} = \mathcal{V}_k$, we have $\mathbf{x} \in \mathcal{V}_k \ominus \mathcal{V}_{k-1}$. Since $\dim(\mathcal{V}_k \ominus \mathcal{V}_{k-1}) = 1$ and $\tilde{\mathbf{u}}_k \in \mathcal{V}_k \ominus \mathcal{V}_{k-1}$, there exists $t \in \mathbb{R}$ such that $\mathbf{x} = t\tilde{\mathbf{u}}_k$, i.e., $a\mathbf{1}_{\mathcal{A}_k} + b\mathbf{1}_{\mathcal{B}_k} + c\mathbf{1}_{\mathcal{C}} = ta_k\mathbf{1}_{\mathcal{A}_k} + tb_k\mathbf{1}_{\mathcal{B}_k}$. Hence $a = ta_k$, $b = tb_k$, $c = 0$, i.e., $\mathbf{y} = t[a_k, 0, b_k]^\top$, therefore $\ker(\tilde{\mathbf{U}}_k^\top \mathbf{M}) = \text{span}[a_k, 0, b_k]^\top$ and $\dim \ker(\tilde{\mathbf{U}}_k^\top \mathbf{M}) = 1$. \square

In Theorem 9, \mathbf{M} and $\tilde{\mathbf{U}}_k$ satisfy the condition (13), i.e., $\mathbf{M} \in \mathcal{M}_{\tilde{\mathbf{U}}_k}^*$. Since $\tilde{\mathbf{u}}_k \in \mathcal{X}_{\tilde{\mathbf{U}}_k} \cap \text{span } \mathbf{M}$, we have $\tilde{\mathbf{u}}_k \in \mathcal{X}_{\tilde{\mathbf{U}}_k}^*$, i.e., $\tilde{\mathbf{u}}_k$ can be seen as a “critical point” of problem $\mathcal{P}_{\tilde{\mathbf{U}}_k}^*$. Even though $\tilde{\mathbf{u}}_k$ may not be a local minimizer, the greedy basis $\tilde{\mathbf{U}}$ still provides a rather good approximation to the ℓ_1 basis, which will be demonstrated in the numerical experiments afterwards.



Fig. 4. Two extreme cases of \mathbb{T} : Perfect binary tree (left) and extremely unbalanced binary tree (right).

4.3. Computational complexity of the greedy basis

As mentioned before, obtaining the Laplacian basis takes $O(N^3)$ operations which is rather expensive. There are works in literature focusing on fast approximate graph Fourier transforms. In [18], K. Lu et al. show that sparse matrix factorization can lead to fast implementation of GFT. Based on this result, they learn from data a graph that has a fast GFT. The number of multiplication needed for fast GFT is $N(N_1 + N_2)$ where $N = N_1 N_2$, which arrives at minimum $2N^{3/2}$ if $N_1 = N_2$. In [19], Luc Le Magoarou et al. show empirically that the graph Laplacian for several popular graph families can be well approximated with a limited number of Givens factors, i.e. $\mathbf{L} \approx \mathbf{S}_1 \cdots \mathbf{S}_J \mathbf{A} \mathbf{S}_J^\top \cdots \mathbf{S}_1^\top$ for $J = O(N \log N)$. By truncating and modifying the Jacobi eigenvalue algorithm, they obtain approximate fast GFT of complexity $O(N \log N)$ exhibiting constant error for growing N , while the complexity of obtaining the approximate basis $\tilde{\mathbf{U}}$ is $O(N^2 \log N)$. In the following, we will discuss the computational complexity and storage requirement of the greedy basis $\tilde{\mathbf{U}}$.

Since $\tilde{\mathbf{u}}_k = a_k \mathbf{1}_{\mathcal{A}_k} + b_k \mathbf{1}_{\mathcal{B}_k}$, the computation and storage of $\tilde{\mathbf{u}}_k$ amounts to those of \mathcal{A}_k 's and \mathcal{B}_k 's. In the $(k-1)$ th step of the greedy algorithm, to find $\mathcal{A}_{k-1}, \mathcal{B}_{k-1} = \arg \max_{\mathcal{A}, \mathcal{B} \in \tau_{k-1}} F(\mathcal{A}, \mathcal{B})$, we need to compare the mutual weight $F(\mathcal{A}, \mathcal{B})$ for all $\mathcal{A}, \mathcal{B} \in \tau_{k-1}$, where $\tau_{k-1} = \{\mathcal{A}_k \cup \mathcal{B}_k, \mathcal{C}_1, \dots, \mathcal{C}_{k-2}\}$. Since $F(\mathcal{C}_i, \mathcal{C}_j)$ for $1 \leq i < j \leq k-2$ are available from the previous steps of calculation, it only needs to compute the new mutual weight $F(\mathcal{A}_k \cup \mathcal{B}_k, \mathcal{C}_i)$ for $i = 1, \dots, k-2$, which further amounts to the computation of $\mathbf{W}(\mathcal{A}_k \cup \mathcal{B}_k, \mathcal{C}_i) = \mathbf{W}(\mathcal{A}_k, \mathcal{C}_i) + \mathbf{W}(\mathcal{B}_k, \mathcal{C}_i)$ for all i . Therefore finding \mathcal{A}_{k-1} and \mathcal{B}_{k-1} needs $O(N)$ arithmetic operations, and the computation of $\{\mathcal{A}_k, \mathcal{B}_k\}_{k=2}^N$ needs $O(N^2)$ operations.

Let us consider the storage of the sets $\{\mathcal{A}_k, \mathcal{B}_k\}_{k=2}^N$, which forms a binary tree \mathbb{T} . The root node of \mathbb{T} is $\{1, \dots, N\}$, and every \mathcal{A}_k or \mathcal{B}_k either is a leaf node or has exactly two children \mathcal{A}_j and \mathcal{B}_j for some $j > k$. The storage of \mathbb{T} is proportional to its size $\|\mathbb{T}\| := \sum_{k=2}^N (|\mathcal{A}_k| + |\mathcal{B}_k|)$. Consider two extreme cases of \mathbb{T} : (i) $|\mathcal{A}_k| = |\mathcal{B}_k|$ for all k ; (ii) $|\mathcal{B}_k| = 1$ for all k . See Fig. 4 for an illustration. For simplicity, suppose N is a power of 2. In the first case, \mathbb{T} is a perfect binary tree and $\|\mathbb{T}\| = N \log_2 N$. In the second case, \mathbb{T} is extremely unbalanced and $\|\mathbb{T}\| = \frac{N^2 + N - 2}{2}$. Experiments show that the binary tree generated from the greedy algorithm is often intermediate between these two extreme cases, so the storage of \mathbb{T} would be generally between $O(N \log N)$ and $O(N^2)$, depending on how balanced the tree is. We remark that when the graph is a Path or Ring (assuming N is a power of 2 and all edge weights are equal), the greedy algorithm yields a perfect binary tree, and the greedy basis is exactly the classical Haar wavelet basis (see the Path case in Table 3).

Let us consider the expansion of a signal \mathbf{x} with the greedy basis. The k th coefficient

$$\langle \mathbf{x}, \tilde{\mathbf{u}}_k \rangle = \langle \mathbf{x}, a_k \mathbf{1}_{\mathcal{A}_k} + b_k \mathbf{1}_{\mathcal{B}_k} \rangle = a_k \langle \mathbf{x}, \mathbf{1}_{\mathcal{A}_k} \rangle + b_k \langle \mathbf{x}, \mathbf{1}_{\mathcal{B}_k} \rangle. \quad (30)$$

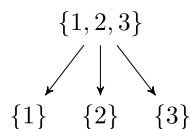
If $|\mathcal{A}_k| = 1$, then $\langle \mathbf{x}, \mathbf{1}_{\mathcal{A}_k} \rangle = x_i$ where $i \in \mathcal{A}_k$. Otherwise $\mathcal{A}_k = \mathcal{A}_j \cup \mathcal{B}_j$ for some $j > k$. Provided that $\langle \mathbf{x}, \mathbf{1}_{\mathcal{A}_j} \rangle$ and $\langle \mathbf{x}, \mathbf{1}_{\mathcal{B}_j} \rangle$ are available from the previous calculations, computing $\langle \mathbf{x}, \mathbf{1}_{\mathcal{A}_k} \rangle = \langle \mathbf{x}, \mathbf{1}_{\mathcal{A}_j} \rangle + \langle \mathbf{x}, \mathbf{1}_{\mathcal{B}_j} \rangle$ needs only one addition. It is similar for computing $\langle \mathbf{x}, \mathbf{1}_{\mathcal{B}_k} \rangle$. Therefore the computation of $\langle \mathbf{x}, \tilde{\mathbf{u}}_k \rangle$ for all k

needs $O(N)$ operations. Note that computing the Fourier transform of a signal under the Laplacian basis needs $O(N^2)$ operations.

4.4. Comparison with related works

Constructing orthonormal basis based on a multi-resolution analysis (MRA) on a partition tree is not a new idea. In the following we compare our greedy basis with several related works in literature.

In Gavish's paper [7], the authors propose a harmonic analysis framework on the high-dimensional data which is represented by a hierarchical tree of increasing refined partitions. The tree representation of data naturally induces a multi-resolution analysis with an associated Haar-like wavelet basis. They assume the hierarchical tree is given in advance. Not like we are using a binary partition tree, Gavish's tree is not limited to binary, i.e., a parent node could have three or more children. In that case, the basis is generally not unique, unless more constraints are posed. To see this, consider the following partition tree. The level-2 subspace $\mathcal{V}^{(2)} = \text{span}\{\mathbf{1}_{\{1\}}, \mathbf{1}_{\{2\}}, \mathbf{1}_{\{3\}}\}$,



the level-1 subspace $\mathcal{V}^{(1)} = \text{span}\{\mathbf{1}_{\{1,2,3\}}\}$, and the orthogonal complement $\mathcal{W} = \mathcal{V}^{(2)} \ominus \mathcal{V}^{(1)}$ has dimension 2. Therefore the basis chosen from \mathcal{W} is not unique. However, if both methods are using identical binary partition tree, then the corresponding orthonormal basis of ours and Gavish's would be the same. As theoretical contribution, Gavish et al. establish relations between the decay of transform coefficients and function smoothness.

In Chen's paper [15], the authors use the similar idea of multi-resolution analysis and construct the binary tree by iteratively partitioning the graph into two disjoint connected subgraphs. Since they promote bisection, their binary tree is balanced and has depth around $1 + \log_2 N$. The main difference between our method and theirs is the construction process of the binary tree. They use a coarse-to-fine approach, while we use a fine-to-coarse approach. If the two methods yield the same partition tree, then the corresponding orthonormal bases would be the same. Chen et al. also estimate the relation between the zero norm of the transform coefficients and signal variation, and apply the proposed signal dictionaries in approximation and localization tasks.

In Tremblay's paper [16,17], the authors define a filter-bank based on a partition of graph into connected subgraphs. The filtering operation is defined through the local Fourier basis of each subgraph. Therefore the orthonormal basis thus obtained is essentially the union of all local Fourier bases, which is quite different from the multi-resolution analysis approach.











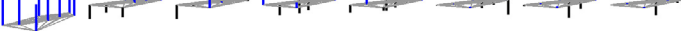

In summary, the multi-resolution analysis approach we use to construct the greedy basis is similar to Gavish or Chen. Once a partition tree is defined on the graph, an MRA can be carried out to produce an orthonormal basis. The main theoretical contribution of our work is that we show the orthonormal basis derived from an MRA satisfies a condition that is also satisfied by the minimizer of an ℓ_1 norm variation problem.

5. Numerical experiments

In this section, we conduct four experiments to compare the performances of Laplacian basis, ℓ_1 basis and greedy basis. The Matlab codes are available at <https://github.com/Jianfeng-Huang-at-github/GFT>.

Table 3

Comparison of Laplacian basis \mathbf{U}^{LP} , ℓ_1 basis \mathbf{U}^{ℓ_1} and greedy basis \mathbf{U}^{Gr} on Path, Comet and Sensor network. These figures are generated with the help of GSPBOX [20].

Path 	<div> \mathbf{u}_1^{LP} \mathbf{u}_2^{LP} \mathbf{u}_3^{LP} \mathbf{u}_4^{LP} \mathbf{u}_5^{LP} \mathbf{u}_6^{LP} \mathbf{u}_7^{LP} \mathbf{u}_8^{LP} </div>  <div> $\mathbf{u}_1^{\ell_1}$ $\mathbf{u}_2^{\ell_1}$ $\mathbf{u}_3^{\ell_1}$ $\mathbf{u}_4^{\ell_1}$ $\mathbf{u}_5^{\ell_1}$ $\mathbf{u}_6^{\ell_1}$ $\mathbf{u}_7^{\ell_1}$ $\mathbf{u}_8^{\ell_1}$ </div>  <div> \mathbf{u}_1^{Gr} \mathbf{u}_2^{Gr} \mathbf{u}_3^{Gr} \mathbf{u}_4^{Gr} \mathbf{u}_5^{Gr} \mathbf{u}_6^{Gr} \mathbf{u}_7^{Gr} \mathbf{u}_8^{Gr} </div> 
Comet 	<div> \mathbf{u}_1^{LP} \mathbf{u}_2^{LP} \mathbf{u}_3^{LP} \mathbf{u}_4^{LP} \mathbf{u}_5^{LP} \mathbf{u}_6^{LP} \mathbf{u}_7^{LP} \mathbf{u}_8^{LP} </div>  <div> $\mathbf{u}_1^{\ell_1}$ $\mathbf{u}_2^{\ell_1}$ $\mathbf{u}_3^{\ell_1}$ $\mathbf{u}_4^{\ell_1}$ $\mathbf{u}_5^{\ell_1}$ $\mathbf{u}_6^{\ell_1}$ $\mathbf{u}_7^{\ell_1}$ $\mathbf{u}_8^{\ell_1}$ </div>  <div> \mathbf{u}_1^{Gr} \mathbf{u}_2^{Gr} \mathbf{u}_3^{Gr} \mathbf{u}_4^{Gr} \mathbf{u}_5^{Gr} \mathbf{u}_6^{Gr} \mathbf{u}_7^{Gr} \mathbf{u}_8^{Gr} </div> 
Sensor 	<div> \mathbf{u}_1^{LP} \mathbf{u}_2^{LP} \mathbf{u}_3^{LP} \mathbf{u}_4^{LP} \mathbf{u}_5^{LP} \mathbf{u}_6^{LP} \mathbf{u}_7^{LP} \mathbf{u}_8^{LP} </div>  <div> $\mathbf{u}_1^{\ell_1}$ $\mathbf{u}_2^{\ell_1}$ $\mathbf{u}_3^{\ell_1}$ $\mathbf{u}_4^{\ell_1}$ $\mathbf{u}_5^{\ell_1}$ $\mathbf{u}_6^{\ell_1}$ $\mathbf{u}_7^{\ell_1}$ $\mathbf{u}_8^{\ell_1}$ </div>  <div> \mathbf{u}_1^{Gr} \mathbf{u}_2^{Gr} \mathbf{u}_3^{Gr} \mathbf{u}_4^{Gr} \mathbf{u}_5^{Gr} \mathbf{u}_6^{Gr} \mathbf{u}_7^{Gr} \mathbf{u}_8^{Gr} </div> 

5.1. Greedy basis on toy graphs

To get a more intuitive understanding of the ℓ_1 basis and greedy basis, we consider three types of simple graphs: Path, Comet and Sensor network. These graphs are generated using GSPBOX [20]. In Table 3 we show the Laplacian basis \mathbf{U}^{LP} (top), the ℓ_1 basis \mathbf{U}^{ℓ_1} (middle) and the greedy basis $\mathbf{U}^{\text{Gr}} = \tilde{\mathbf{U}}$ (bottom) for each type of graph. The mutual weight function we are using here and afterwards is $F(\mathcal{A}, \mathcal{B}) = \mathbf{W}(\mathcal{A}, \mathcal{B}) / (|\mathcal{A}| \cdot |\mathcal{B}|)$. The ℓ_1 basis is calculated using the enumeration method mentioned in Section 3. We remark that the ℓ_1 basis is not necessarily unique for a graph, because the ℓ_1 norm variation minimization problem may have non-unique global minimizers, especially when the weights are integers.

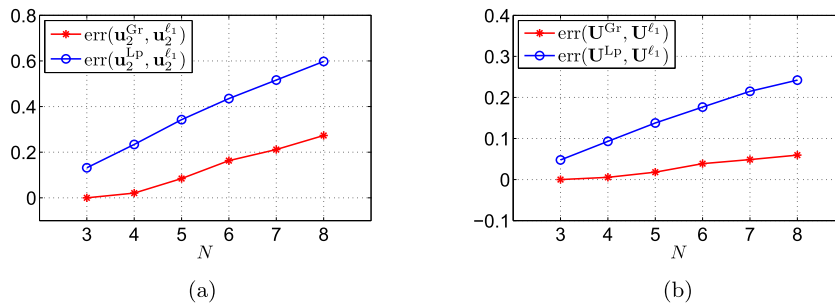


Fig. 5. Comparison of variation between different bases. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

In the Path graph case, one can clearly see that \mathbf{u}_k^{LP} varies continuously across the vertices, while the $\mathbf{u}_k^{\ell_1}$ and \mathbf{u}_k^{Gr} tend to be constant on adjacent vertices. Moreover, \mathbf{u}_k^{LP} is usually supported on the whole vertices domain, while $\mathbf{u}_k^{\ell_1}$ and \mathbf{u}_k^{Gr} tend to have smaller support size as k increases. The other two cases of Comet and Sensor network are similar. In a word, the ℓ_1 basis and greedy basis are more sparse than the Laplacian basis.

5.2. Error between the greedy basis and ℓ_1 basis

In this experiment, we aim to examine the difference between the greedy basis \mathbf{U}^{Gr} and the ℓ_1 basis \mathbf{U}^{ℓ_1} . Since \mathbf{u}_1^{Gr} and $\mathbf{u}_1^{\ell_1}$ are equal, we begin the comparison from \mathbf{u}_2^{Gr} and $\mathbf{u}_2^{\ell_1}$. Denote the relative error of their ℓ_1 norm variations by

$$\text{err}(\mathbf{u}_2^{\text{Gr}}, \mathbf{u}_2^{\ell_1}) := \frac{S(\mathbf{u}_2^{\text{Gr}}) - S(\mathbf{u}_2^{\ell_1})}{S(\mathbf{u}_2^{\ell_1})}.$$

In Fig. 5(a) the red line plots the average of $\text{err}(\mathbf{u}_2^{\text{Gr}}, \mathbf{u}_2^{\ell_1})$ for 100 random graphs. Each of these graphs is generated by N random points p_i chosen uniformly in the unit square $[0, 1]^2$, and the weights are defined by $w_{ij} := \exp(-\|p_i - p_j\|^2/\sigma^2)$ for some parameter σ . For the sake of completeness, we also plot the relative error $\text{err}(\mathbf{u}_2^{\text{LP}}, \mathbf{u}_2^{\ell_1})$ in the blue line, where \mathbf{u}_2^{LP} is the second Laplacian eigenvector.

We also compare the sum of variations of the two bases. Denote

$$S(\mathbf{U}^{\ell_1}) := \sum_{k=1}^N S(\mathbf{u}_k^{\ell_1}), \quad S(\mathbf{U}^{\text{Gr}}) := \sum_{k=1}^N S(\mathbf{u}_k^{\text{Gr}})$$

and the relative error

$$\text{err}(\mathbf{U}^{\text{Gr}}, \mathbf{U}^{\ell_1}) := \frac{S(\mathbf{U}^{\text{Gr}}) - S(\mathbf{U}^{\ell_1})}{S(\mathbf{U}^{\ell_1})}.$$

The average of $\text{err}(\mathbf{U}^{\text{Gr}}, \mathbf{U}^{\ell_1})$ for 100 random graphs is plotted in Fig. 5(b) by the red line. The relative error $\text{err}(\mathbf{U}^{\text{LP}}, \mathbf{U}^{\ell_1})$ between $S(\mathbf{U}^{\text{LP}})$ and $S(\mathbf{U}^{\ell_1})$ is also plotted for comparison (blue line in Fig. 5(b)).

5.3. n -term approximation

A nice property of the classical Fourier transform is that the Fourier coefficients usually have a fast decay for most real-world signals. That means one can drop the high frequency coefficients without losing much information, which serves as the foundation of various signal compression methods. In the following, we will examine this property for the greedy basis \mathbf{U}^{Gr} , and compare it to the Laplacian basis \mathbf{U}^{LP} .

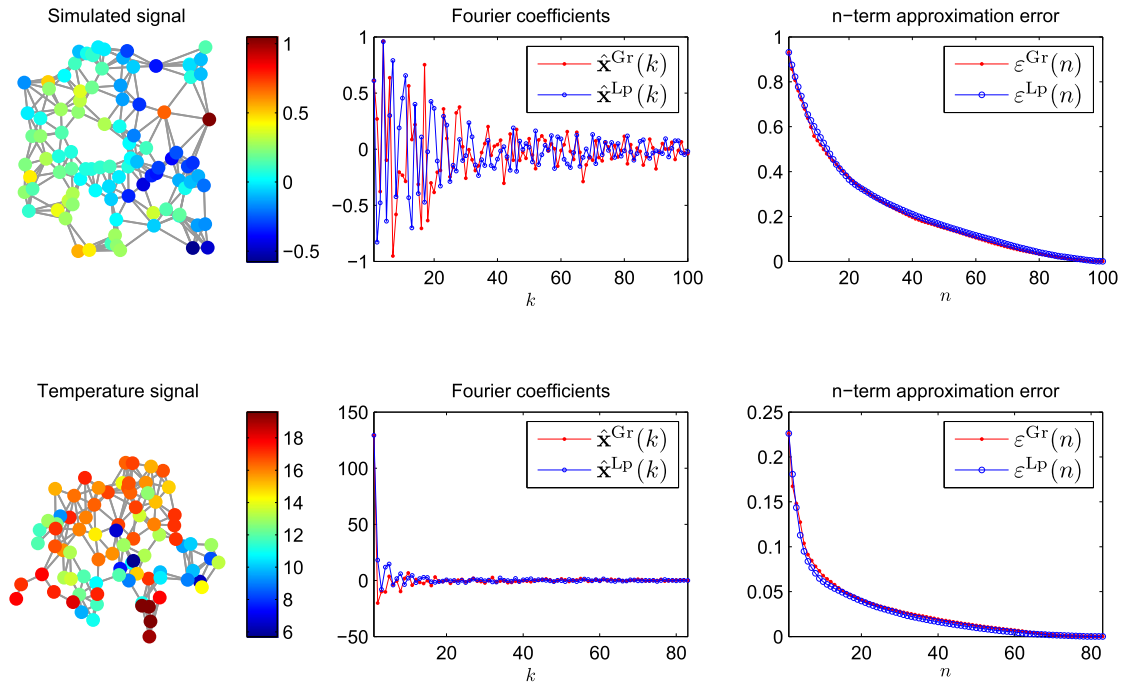


Fig. 6. Comparison of n -term approximation errors under the greedy basis (red) and Laplacian basis (blue).

Given a signal \mathbf{x} , let the Fourier transform under the Laplacian basis be denoted by $\hat{\mathbf{x}}^{\text{LP}}$, and the Fourier transform under the greedy basis be denoted by $\hat{\mathbf{x}}^{\text{Gr}}$. Suppose we use the largest n terms of coefficients to approximate \mathbf{x} . Namely we sort the absolute values of the coefficients in descending order, say $|\hat{\mathbf{x}}^{\text{Gr}}(k_1)| \geq \dots \geq |\hat{\mathbf{x}}^{\text{Gr}}(k_N)|$, for the greedy basis. Then we define the n -term approximation

$$\mathbf{y}_n^{\text{Gr}} := \sum_{i=1}^n \hat{\mathbf{x}}^{\text{Gr}}(k_i) \mathbf{u}_{k_i}^{\text{Gr}}$$

and the approximation error

$$\varepsilon^{\text{Gr}}(n) := \frac{\|\mathbf{x} - \mathbf{y}_n^{\text{Gr}}\|}{\|\mathbf{x}\|} = \frac{(\sum_{i=n+1}^N |\hat{\mathbf{x}}^{\text{Gr}}(k_i)|^2)^{1/2}}{\|\hat{\mathbf{x}}^{\text{Gr}}\|}.$$

For the Laplacian basis, the n -term approximation \mathbf{y}_n^{LP} and the approximation error $\varepsilon^{\text{LP}}(n)$ are defined in a similar way.

The experiment is performed to two types of signal. The first is a simulated signal defined through its Laplacian coefficient $\hat{\mathbf{x}}^{\text{LP}}$ as follows,

$$\hat{\mathbf{x}}^{\text{LP}}(k) := \frac{1}{1 + \mu \lambda_k} \times \text{rand}(),$$

where μ is a constant, λ_k is the Laplacian eigenvalue, and $\text{rand}()$ is a random number uniformly distributed on $[-1, 1]$. The first row of Fig. 6 plots the simulated signal (left), its Fourier coefficients under the two bases (middle) and the corresponding approximation errors (right). The second example is a real-world signal: the average June temperature of Switzerland during 1981-2010 [21]. See the second row of Fig. 6 for the results. It can be seen that for either simulated or real-world signal, both types of Fourier transform lead to a fast decay of approximation error, and the rates of decay are almost the same.

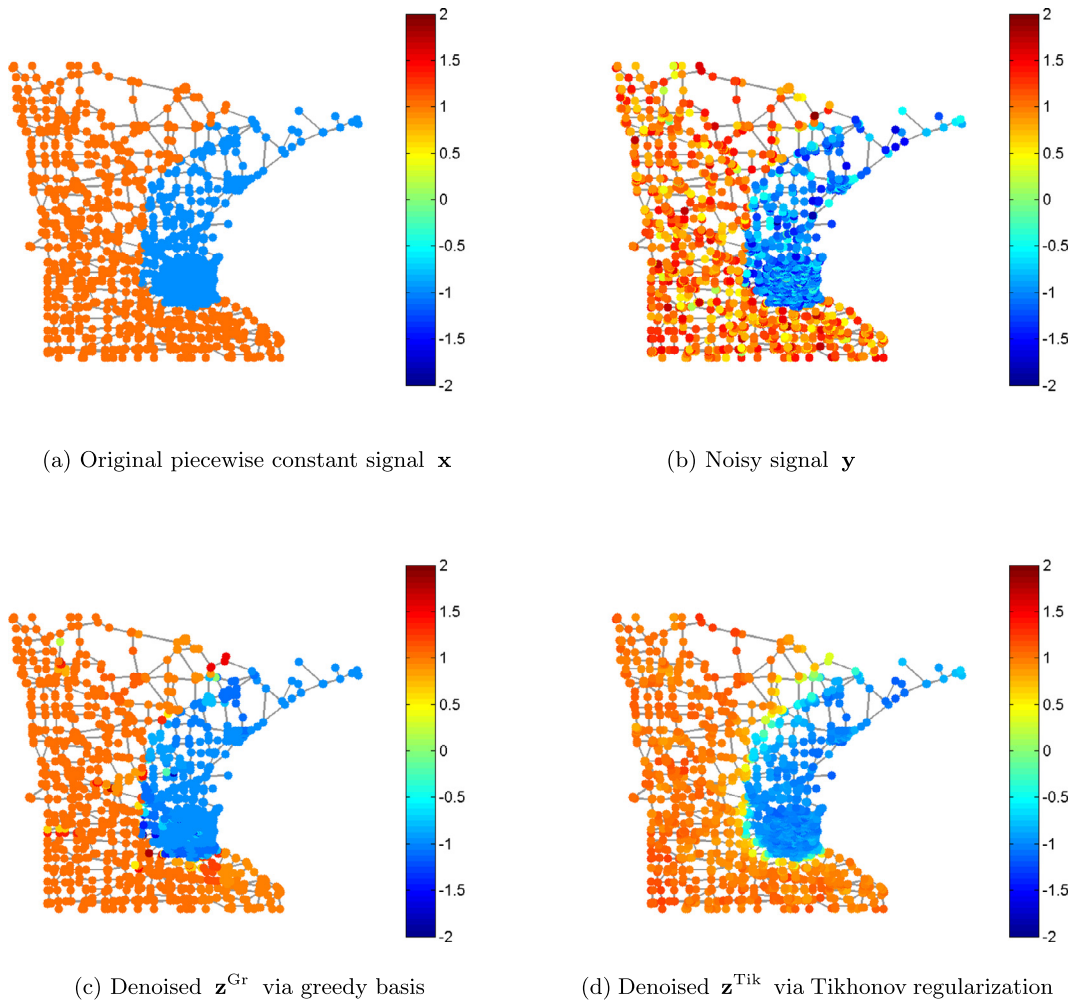


Fig. 7. Comparison of two methods in denoising a piecewise constant signal.

5.4. Denoising

Since the greedy basis tends to be locally constant, it is expected to have advantages over the Laplacian basis in representing piecewise constant signals. To show the potential of the greedy basis in this regard, we conduct the following denoising experiment.

The original signal \mathbf{x} is a piecewise constant signal on the Minnesota graph, as shown in Fig. 7 (a). We add 10 dB white Gaussian noise to get a noisy signal \mathbf{y} , as shown in Fig. 7 (b).

Two denoising methods are compared here. The first method is the n -term greedy coefficients approximation. Let $\hat{\mathbf{y}}^{\text{Gr}}$ be the coefficients of \mathbf{y} under the greedy basis. We use 3% largest greedy coefficients to reconstruct a denoised signal

$$\mathbf{z}^{\text{Gr}} := \sum_{i=1}^{N'} \hat{\mathbf{y}}^{\text{Gr}}(k_i) \mathbf{u}_{k_i}^{\text{Gr}},$$

where $N' = \lfloor 0.03N \rfloor$ and $\hat{\mathbf{y}}^{\text{Gr}}(k_i)$ is the sorted greedy coefficient. The second method is the classical Tikhonov regularization [3]. Let $\hat{\mathbf{y}}^{\text{LP}}$ be the coefficients of \mathbf{y} under the Laplacian basis. The Tikhonov denoised signal is defined as

$$\mathbf{z}^{\text{Tik}} := \sum_{k=1}^N \frac{\hat{\mathbf{y}}^{\text{LP}}(k)}{1 + \gamma \lambda_k} \mathbf{u}_k^{\text{LP}},$$

where $\gamma > 0$ and λ_k is the Laplacian eigenvalue. The two denoised signals are shown respectively in Fig. 7 (c) and (d). It can be seen that \mathbf{z}^{Gr} preserves the discontinuous edge pretty well, while \mathbf{z}^{Tik} introduces some blurring over the edge.

6. Conclusion

In this paper we propose a generalized definition of ℓ_1 Fourier basis of a graph as the solutions of a sequence of ℓ_1 norm variation minimization problems. We obtain a necessary condition satisfied by the local minimizer of problem $\mathcal{P}_{\mathbf{U}}$, which implies the number of different components of $\mathbf{u}_k^{\ell_1}$ is at most k . Furthermore, we show that $\mathcal{P}_{\mathbf{U}}$ has only finitely many isolated local minimizers, contained in a finite set $\mathcal{X}_{\mathbf{U}}^*$, so it is possible to enumerate $\mathcal{X}_{\mathbf{U}}^*$ to find the global minimizer of $\mathcal{P}_{\mathbf{U}}$ when N is small. For large N , we give a fast greedy algorithm to approximately construct the ℓ_1 basis, based on a greedy partition sequence created by grouping the vertices according to their mutual weights. Numerical experiments show that the greedy basis provides a good approximation to the ℓ_1 basis. Also, the approximation errors under the two bases (greedy basis and Laplacian basis) have nearly the same rate of decay for simulated and real-world signals. A denoising experiment shows that the greedy basis is more suitable to represent piecewise constant signals than Laplacian basis. For future directions, one may consider the general ℓ_p norm variation minimization problem and the corresponding ℓ_p Fourier basis. Since there are rich results connecting the Laplacian basis to the graph structure, it is an important and meaningful problem to explore the relationship between the ℓ_1 basis and the graph structure.

Acknowledgments

The authors would like to thank the anonymous reviewers and the Editor for their insightful comments, which help enhance the quality of the paper significantly. This work is supported by National Natural Science Foundation of China (Nos. 11601532, 11771458, 11501377).

References

- [1] Fan R.K. Chung, *Spectral Graph Theory*, No. 92, American Mathematical Society, 1997.
- [2] Dimitri P. Bertsekas, *Nonlinear Programming*, 2nd edition, Athena Scientific, 1999.
- [3] David I. Shuman, Sunil K. Narang, Pascal Frossard, Antonio Ortega, Pierre Vandergheynst, The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains, *IEEE Signal Process. Mag.* 30 (3) (2013) 83–98, <https://doi.org/10.1109/MSP.2012.2235192>.
- [4] David I. Shuman, Benjamin Ricaud, Pierre Vandergheynst, Vertex-frequency analysis on graphs, *Appl. Comput. Harmon. Anal.* 40 (2) (2016) 260–291, <https://doi.org/10.1016/j.acha.2015.02.005>.
- [5] Aliaksei Sandryhaila, José M.F. Moura, Discrete signal processing on graphs, *IEEE Trans. Signal Process.* 61 (7) (2013) 1644–1656, <https://doi.org/10.1109/TSP.2013.2238935>.
- [6] Ameiya Agaskar, Yue M. Lu, A spectral graph uncertainty principle, *IEEE Trans. Inf. Theory* 59 (7) (2013) 4338–4356, <https://doi.org/10.1109/TIT.2013.2252233>.
- [7] Matan Gavish, Boaz Nadler, Ronald R. Coifman, Multiscale wavelets on trees, graphs and high dimensional data: theory and applications to semi supervised learning, in: *ICML*, 2010.
- [8] David K. Hammond, Pierre Vandergheynst, Remi Gribonval, Wavelets on graphs via spectral graph theory, *Appl. Comput. Harmon. Anal.* 30 (2) (2011) 129–150, <https://doi.org/10.1016/j.acha.2010.04.005>.
- [9] Xu Chen, Xiuyuan Cheng, Stéphane Mallat, Unsupervised deep haar scattering on graphs, *Adv. Neural Inf. Process. Syst.* 27 (2014).
- [10] Bin Dong, Sparse representation on graphs by tight wavelet frames and applications, *Appl. Comput. Harmon. Anal.* 42 (3) (2017) 452–479, <https://doi.org/10.1016/j.acha.2015.09.005>.
- [11] Stefania Sardellitti, Sergio Barbarossa, Paolo Di Lorenzo, On the graph Fourier transform for directed graphs, *IEEE J. Sel. Top. Signal Process.* 11 (6) (2017) 796–811, <https://doi.org/10.1109/JSTSP.2017.2726979>.
- [12] Xavier Bresson, Thomas Laurent, David Uminsky, James V. Brecht, Convergence and energy landscape for Cheeger cut clustering, *Adv. Neural Inf. Process. Syst.* 25 (2012).

- [13] Rongjie Lai, Stanley Osher, A splitting method for orthogonality constrained problems, *J. Sci. Comput.* 58 (2) (2014) 431–449, <https://doi.org/10.1007/s10915-013-9740-x>.
- [14] Siheng Chen, Rohan Varma, Aliaksei Sandryhaila, Jelena Kovačević, Discrete signal processing on graphs: sampling theory, *IEEE Trans. Signal Process.* 63 (24) (2015) 6510–6523, <https://doi.org/10.1109/TSP.2015.2469645>.
- [15] Siheng Chen, Aarti Singh, Jelena Kovačević, Multiresolution representations for piecewise-smooth signals on graphs, *ArXiv e-prints* arXiv:1803.02944, <https://arxiv.org/abs/1803.02944>.
- [16] Nicolas Tremblay, Pierre Borgnat, Subgraph-based filterbanks for graph signals, *IEEE Trans. Signal Process.* 64 (15) (2016) 3827–3840, <https://doi.org/10.1109/TSP.2016.2544747>.
- [17] Nicolas Tremblay, Paulo Goncalves, Pierre Borgnat, Design of graph filters and filterbanks, in: *Cooperative and Graph Signal Processing*, Academic Press, 2018, pp. 299–324.
- [18] Keng-Shih Lu, Antonio Ortega, A graph Laplacian matrix learning method for fast implementation of graph Fourier transform, in: *2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, 2017, pp. 1677–1681.
- [19] Luc Le Magoarou, Rémi Gribonval, Nicolas Tremblay, Approximate fast graph Fourier transforms via multilayer sparse approximations, *IEEE Trans. Signal Inf. Process. Netw.* 4 (2) (2018) 407–420, <https://doi.org/10.1109/TSIPN.2017.2710619>.
- [20] Nathanaël Perraudin, Johan Paratte, David Shuman, Lionel Martin, Vassilis Kalofolias, Pierre Vandergheynst, David K. Hammond, GSPBOX: a toolbox for signal processing on graphs, *ArXiv e-prints* arXiv:1408.5781, <http://arxiv.org/abs/1408.5781>.
- [21] http://www.meteoswiss.admin.ch/product/input/climate-data/normwerte-pro-messgroesse/np8110/nvrep_np8110_tre200m0_e.txt.