APPROXIMATING SIGNALS SUPPORTED ON GRAPHS

Xiaofan Zhu, Michael Rabbat

Department of Electrical and Computer Engineering, McGill University 3480 University St, Montreal QC, Canada H3A 2A7

ABSTRACT

In this paper, we introduce the concept of smoothness for signals supported on the vertices of a graph. We provide theoretical explanations when and why the Laplacian eigenbasis can be regarded as a meaningful "Fourier" transform of such signals. Moreover, we analyze the desired properties of the underlying graphs for better compressibility of the signals. We verify our theoretical work by experiments on real world data.

Index Terms— Graph Laplacian, Fourier transform, smoothness, compressibility

1. INTRODUCTION

Signals on graphs are now common in various applications of different areas including biology, network monitoring and the smart grid. For example, in field estimation [1, 2], a huge number of wireless sensors are distributed randomly in a field to collect measurements, such as temperature or solar radiation, where the whole sensor network can be modeled as a random geometric graph. In computer graphics, the shape of a 3D object can be approximated by a regular graph, with its nodes containing the coordinate information [3, 4]. In the traditional realm of signal processing, we are interested in approximating a certain function by a simpler one and approximation theory has been well developed. So far the approximation theory has focused on 1D signals and 2D images while less work has considered signals on graphs. So a general question one might ask is: how can we approximate signals supported on graphs?

To address this problem, first let us delve into the traditional approximation theory. It is well known that the Fourier transform plays a core role in this area. Moreover, the idea that any arbitrary periodic function can be represented as a series of harmonically related sinusoids has a profound impact in mathematical analysis, physics, and engineering. In signal processing, it has been shown that a smooth signal can be well approximated by a small portion of its Fourier coefficients because of the compressibility. Thus our question becomes more specific: can we find a "Fourier transform" for signals on graphs? In academia, it has been believed for quite a while that the eigenbases of a Laplacian matrix can be deemed as the Fourier basis for its corresponding graph. In this paper, we denote it as the Graph Fourier Transform (GFT). Furthermore, there have already existed certain applications which utilize the GFT in data compression [3, 5], signal denoising [6] or compressed sensing [7]. However, none of them provides a detailed theoretical analysis on why the graph Laplacian eigenbases can be regarded as the Fourier transform of graphs. Nor do they discuss whether the Laplacian eigenvectors are meaningful basis vectors for all graphs.

In this paper, we address issues. We first generalize the concept of smooth signals and define a metric to measure the smoothness of a graph signal. Later, we derive certain properties of the GFT. Those properties imply that if the eigenvalues of the graph Laplacian roughly maintain an increasing trend, then the smooth signals on that graph are likely to be compressible.

The rest of this paper is organized as follows: In section 2, the concept of smooth signals on graphs is defined and we derive certain properties of the GFT. Based on those properties, we give some rules of thumb for generating graphs where the signal is compressible. In section 3, we conduct experiments on real world data to verify our theory about the GFT. Finally, a conclusion is made in section 4.

2. THE GRAPH FOURIER TRANSFORM

2.1. Properties of the Fourier Transform

The Fourier transform is a mathematical operation that decomposes a signal into its constituent frequencies. It plays an important role in signal processing. In this subsection, we will review some of the important properties of the Fourier transform.

Definition 1. For a continuous differentiable function f, the total variation is defined as $||f||_V = \int_{-\infty}^{+\infty} |f'(t)| dt$, where f'(t) is the derivative, and $||f||_V = \sum_n |f(n) - f(n-1)|$ for discrete signals. We say that f has bounded variation if $||f||_V < +\infty$.

Total variation measures the total amplitude of signal oscillations. It plays an important role in signal processing since it impacts the decaying behavior of its Fourier coefficients. The following proposition shows that the total variation affects the decaying upper bound of the Fourier coefficients.

Proposition 2.1 ([8]). If f(t) is differentiable and $\widehat{f(\omega)} = \int_{-\infty}^{+\infty} f(t)e^{-iwt}dt$ denotes its Fourier transform, then $|\widehat{f(\omega)}| \leq \frac{\|f\|_V}{\|f\|_V}$.

In the theoretical analysis of approximation theory[8], we often consider a signal f to be square integrable over [0, 1], we can decompose a signal $f(t) = \sum_{m=-\infty}^{+\infty} |\langle f(u), e^{i2\pi m u} \rangle |e^{i2\pi m t}$ with $\langle f(u), e^{i2\pi m u} \rangle = \int_0^1 f(u) e^{-i2\pi m u} du$, the M-term Fourier approximation is $f_M = \sum_{|m| < M/2} |\langle f(u), e^{i2\pi m u} \rangle |e^{i2\pi m t}$.

Definition 2. M-term Linear Fourier Approximation Error: $\epsilon_l(M, f) = \sum_{|m| > M/2} |\langle f(u), e^{i2\pi m u} \rangle|^2$.

Linear approximation keeps the M lowest frequency components while discarding the rest.

Theorem 2.2 ([8]). If $||f||_V < +\infty$, then $\epsilon_l(M, f) = O(||f||_V M^{-1})$.

Theorem 2.3 ([8]). For any s > 1/2, if $\sum_{m=0}^{+\infty} |m|^{2s} |\langle f, g_m \rangle|^2 < +\infty$ where g_m is the mth vector from a certain orthogonal basis, then $\epsilon_l(M, f) \sim o(M^{-2s})$.

The theorems above describe the decay rate of Fourier coefficients and the behavior of linear approximation error. It is worth noting that Theorem 2.1 is consistent with the fact that a smooth signal is likely to be compressible in the Fourier domain. Theorem 2.2 shows that the linear approximation error is upper bounded by total variation and thus signals with small total variation will result in less linear approximation error. Theorem 2.3 highlights that the linear approximation error depends on the decay rate of $|\langle f, g_m \rangle|$. In the next several subsections, we show that all the three theorems above have similar versions for the GFT.

2.2. Towards the Graph Fourier Transform

Signals supported on graphs are fairly common in real applications. If given a graph G(V, E) and a signal f, we say that $f \in \mathbb{R}^V$ if the entries of f are supported on the vertices of G.

Since the topology of an underlying graph is crucial to the signals supported on it, we need some tools to analyze the graph topology. For an undirected, unweighted graph G = (V, E), which consists of a set of edges E and a set of vertices V, the adjacency matrix A of G is the $N \times N$ matrix with entries

$$A_{i,j} = \begin{cases} 1 & : & \text{if there is an edge between vertex } i \text{ and } j \\ 0 & : & \text{otherwise} \end{cases}$$

and N is the number of nodes. The degree of vertex *i*, denoted by d_i , is the number of all the edges incident to it. Let the degree matrix D have diagonal elements equal to the degrees, and zeros elsewhere. The non-normalized graph Laplacian is defined as: L = D - A.

An interesting fact has been noticed for a long time: that the 1-D ring and the 2-D grid are examples of circulant graphs and it is well known that the Discrete Fourier Transform(DFT) produce an eigenbases of all circulant matrices[9], i.e., the Laplacian eigenbasis of any circulant graph is exactly the DFT basis. This might be a starting point for some scholars to relate the Laplacian eigenbasis to the "Fourier" transform of graphs. Naturally, one might ask: Is it possible for graphs with more general structures to have similar properties of the Fourier transform? The following subsection considers this issue.

2.3. Properties of the Graph Fourier Transform

One vital concept closely related to the Fourier transform is the smoothness of signals since smooth signals have compressible Fourier coefficients; i.e., the sorted magnitude of Fourier coefficients exhibits a power law decay. Hence, we can keep a small portion of the large ones to approximate the signal while discarding all the others. Similarly, for our case, the very first notion we need to figure out is the smoothness of signals on graphs. Karni and Gotsman [3] argue the smoothness for spectral compression is when "the coordinates of a vertex are very close to the average coordinates of its neighbors." Their work is limited to coordinates on meshes while we care about more general graphs and signals. Accordingly, we extend this notion to "the value associated with a vertex is very close to that of its neighbors". More concretely, the following definition of 2-norm graph total variation describes the overall smoothness of a signal.

Definition 3. 2-norm Graph Total Variation: Given a signal $f \in \mathbb{R}^V$, $||f||_G = (f^T L f)^{1/2} = (\sum_{i \sim j} w_{ij} (f(i) - f(j))^2)^{1/2}$, where $i \sim j$ means there exists an edge between node i and node j.

The 2-norm graph total variation describes the smoothness of a signal defined on graph vertices. The smaller graph total variation a

signal has, the smoother the signal is on the graph. Zhu et al. [10] also mention that $f^T L f$ measures the smoothness of f on the graph.

Definition 4. We say that $f \in \mathbb{R}^V$ has a bounded variation if we can find a positive $C \ll \lambda_{N-1}$ such that $||f||_G^2 \leq C||f||^2$, where λ_{N-1} is the largest Laplacian eigenvalue of the underlying graph. Bounded variation can also be defined for graphs with an infinite number of nodes: if $||f||_G < +\infty$, then f has a bounded variation.

Although there exists no infinitely large graphs in real world, discussing the properties of such graphs can provide certain implications about the behavior of large graphs. Actually, the bounded variation for infinitely large graphs implies $\sum_{i=0}^{+\infty} \lambda_i |\widehat{f(\lambda_i)}|^2 < +\infty$, which gives $\lim_{i\to\infty} \lambda_i |\widehat{f(\lambda_i)}|^2 = 0$. Hence, a graph signal with bounded variation doesn't necessarily have decaying GFT coefficients. For example, if we consider a complete graph here, $|\widehat{f(\lambda_i)}| \to 0$ since $\lambda_i \to +\infty$, where $i = 1, 2, \cdots$, i.e., signals only contains DC component can be considered smooth for complete graphs.

Now let us define the linear and non-linear approximation error for the GFT, they are similar to those of the Fourier transform.

Definition 5. M-term Linear Approximation Error: $\epsilon_l(M, f) = \sum_{i=M}^{N-1} |\widehat{f(\lambda_i)}|^2$, where $\widehat{f(\lambda_i)} = \langle f, u_i \rangle$ denotes the GFT coefficient of signal f, where u_i is the *i*th eigenvector of the Laplacian matrix of graph G.

Definition 6. M-term Non-linear Approximation(Best M-term Approximation) Error: $\epsilon_n(M, f) = \sum_{i \notin \Omega} |\widehat{f(\lambda_i)}|^2$, where Ω corresponds to the set of indices of the M largest graph Fourier coefficients in magnitude.

The following theorems describe the properties of the graph Fourier.

Theorem 2.4. Given a signal $f \in \mathbb{R}^V$ on vertices of a graph G(V, E), let λ_i denote the *i*th eigenvalue of the Laplacian matrix L and $\widehat{f(\lambda_i)}$ denotes the *i*th GFT coefficient of the signal f. Then, $|\widehat{f(\lambda_i)}| \leq \frac{\|f\|_G}{\sqrt{\lambda_i}}$.

Proof Sketch: It is straightforward to see that $\lambda_i |\widehat{f(\lambda_i)}|^2 \leq \sum_{i=0}^{N-1} \lambda_i |\widehat{f(\lambda_i)}|^2 = f^T (\sum_{i=0}^{N-1} \lambda_i u_i u_i^T) f = f^T L f = ||f||_G^2$, where u_i is the *i*th eigenvector of the Laplacian matrix L. \Box

Compared with Proposition 2.1, Theorem 2.4 implies the eigenvalues of the graph Laplacian plays the same role as "frequencies" in traditional signal processing. $\lambda_0, \dots, \lambda_{N-1}$ correspond to the graph Fourier coefficients from the lower frequencies to the higher frequencies. Accordingly, the eigenvectors of the Laplacian are actually the "frequency" components of a graph. The next theorem discusses the bound for linear approximation error.

Theorem 2.5. Consider a graph G with a signal $f \in \mathbb{R}^V$ on it. If f has a bounded variation, then for adequately large M:

$$\epsilon_l(M, f) \le \|f\|_G^2 \lambda_M^{-1}$$

Proof Sketch: Notice that $\sum_{i=M}^{N-1} \lambda_i |\widehat{f(\lambda_i)}|^2 \leq \sum_{i=0}^{N-1} \lambda_i |\widehat{f(\lambda_i)}| = ||f||_G^2$ and $\epsilon_l(M, f) = \sum_{i=M}^{N-1} |\widehat{f(\lambda_i)}|^2$, we can relax the above conditions and consider the optimization problem:

$$\max \sum_{i=M}^{N-1} x_i^2 \ s.t. \ \sum_{i=M}^{N-1} \lambda_i x_i^2 \le \|f\|_G^2 \tag{1}$$

By solving this problem, we obtain its solution $x_M^* = \|f\|_G^2 \lambda_M^{-1}$ and $x_i^* = 0$ for all $i = M + 1, \dots, N - 1$. Thus, $\sum_{i=M}^{N-1} (x_i^*)^2$ is clearly an upper bound for $\epsilon_l(M, f)$. Since $\|f\|^2$ also upper bounds $\epsilon_l(M, f), \epsilon_l(M, f) \leq \min\{\|f\|_G^2 \lambda_M^{-1}\}$. Due to the bounded variation condition, we have $\|f\|_G^2 \lambda_M^{-1} \leq \frac{C}{\lambda_M} \|f\|^2$. Since $C \ll \lambda_N$, we can always find $\lambda_M > C$ for adequately large M such that $\|f\|_G^2 \lambda_M^{-1} < \|f\|^2$. \Box

This statement corresponds to Theorem 2.2 for Fourier transform. From Theorem 2.5, the upper bound of the linear approximation error is related to both the eigenvalue and the graph total variation. It implies that if the eigenvalues keep strictly increasing, the linear approximation error will have the decaying property. Moreover, the linear approximation error is also affected by the graph total variation $||f||_G$; i.e., bounded $||f||_G$ results in smaller linear approximation error. This is consistent with our intuition that a smoother signal tends to be better linear-approximated.

Lemma 2.6. Consider a signal $f \in \mathbb{R}^V$ on a connected graph:

$$\sum_{i=0}^{N-1} \lambda_i |\widehat{f(\lambda_i)}|^2 \le \sum_{M=0}^{N-1} \lambda_M \epsilon_l(M, f) \le \sum_{i=0}^{N-1} i\lambda_i |\widehat{f(\lambda_i)}|^2.$$

If we consider a graph G with infinite number of nodes, then

$$\sum_{i=0}^{+\infty} \lambda_i |\widehat{f(\lambda_i)}|^2 \le \sum_{M=0}^{+\infty} \lambda_M \epsilon_l(M, f) \le \sum_{i=0}^{+\infty} i\lambda_i |\widehat{f(\lambda_i)}|^2$$

Proof. Notice the fact that $\sum_{M=0}^{N-1} \lambda_M \sum_{i=M}^{N-1} |\widehat{f(\lambda_i)}|^2 = \sum_{M=0}^{N-1} |\widehat{f(\lambda_i)}|^2$

 $\frac{\sum_{i=0}^{N-1} |\widehat{f(\lambda_i)}|^2 (\sum_{M=0}^i \lambda_M)}{\text{ which immediately gives the inequality of left hand. Moreover, since } \lambda_n \leq \lambda_m \text{ for all } n \leq m \text{ , we obtain the upper bound.}$

Theorem 2.7. Given a graph G with infinite nodes, if

 $\sum_{i=0}^{+\infty} i\lambda_i |\widehat{f(\lambda_i)}|^2 < +\infty$, then the *M* term linear approximation error:

$$\epsilon_l(M, f) = o(\frac{1}{M\lambda_{M/2}}).$$

Proof. From the second statement of Lemma 2.6, we notice that

$$\epsilon_l(M, f) \sum_{m=M/2}^{M-1} \lambda_m \le \sum_{m=M/2}^{M-1} \lambda_m \epsilon_l(m, f)$$
(2)

$$\leq \sum_{m=M/2}^{+\infty} \lambda_m \epsilon_l(m, f) \tag{3}$$

$$\leq \sum_{i=0}^{+\infty} i\lambda_i |\widehat{f(\lambda_i)}|^2.$$
(4)

The first inequality holds due to the fact $\epsilon_l(M, f) \leq \epsilon_l(m, f)$ for all $m \leq M$. Since $\sum_{i=0}^{+\infty} i\lambda_i |\widehat{f(\lambda_i)}|^2 < +\infty$, we have $\sum_{m=M/2}^{+\infty} \lambda_m \epsilon_l(m, f) < +\infty$. Thus,

$$\lim_{M \to \infty} \sum_{m=M/2}^{+\infty} \lambda_m \epsilon_l(m, f) = 0.$$
 (5)

Moreover, it is clear that $\frac{M}{2}\lambda_{M/2} \leq \sum_{m=M/2}^{M-1} \lambda_m$, Accordingly, Eq.2, Eq.3, along with Eq.5 implies that

$$\lim_{M \to \infty} M \lambda_{M/2} \epsilon_l(f, M) = 0.$$

Theorem 2.7 along with Lemma 2.6 describe the behavior of the linear approximation error of graphs with infinite number of nodes when its eigenvalues are strictly increasing. The condition $\sum_{i=0}^{+\infty} i\lambda_i |\widehat{f(\lambda_i)}|^2 < +\infty$ implies $|\widehat{f(\lambda_i)}|^2 = o(\frac{1}{i\lambda_i})$, which is stronger than the bounded variation condition. Then, similar decaying rate of $o(\frac{1}{M\lambda_{M/2}})$ is guaranteed for linear approximation error.

The above theorems provide us with some implications about what signals on what graph are likely to be compressible on the corresponding graph Fourier domain. To sum up, there are two main principles: First, from the perspective of signals, we need a smooth signal on the underlying graph, i.e., $||f||_G$ is small since it controls the upper bound of linear approximation error. Second, from the perspective of the underlying graphs, the Laplacian eigenvalue of the graph must have an increasing trend roughly in order for the graph Fourier coefficients to decay.

2.4. Obtaining Compressible Signal by Proper Graph Construction

Given a signal $f \in \mathbb{R}^N$, what graph leads to a GFT basis of best compression for f? The properties of the GFT provide us with certain implications of this question. First, each entry in f can be regarded as a node allocated one value. From Theorem 2.4 and its corollary, we desire smooth signals on graphs; i.e., $||f||_G$ should be kept small enough. One possible solution to this problem is to exploit ϵ -graph or K-nearest-neighbor(KNN) graphs. An ϵ -graph is generated by connecting the nodes whose distance is smaller than ϵ and a KNN graph is constructed by connecting each nodes' K nearest neighbors. More concretely, we construct the graph by putting an edge between the nodes which are likely to share similar values so that $(\sum_{i \sim j} w_{ij}(f(i) - f(j))^2)^{1/2}$ is kept small. The next question is how do we choose K or ϵ . From the perspective of $||f||_G$, we prefer the parameters to be small since fewer edges will result in smaller $||f||_G$. Moreover, we should avoid constructing a complete graph. Hence, the parameters shouldn't be too large. On the other hand, if the value of K or ϵ is selected to be too small, the connectivity of the graph will be weak and the eigenvalues corresponding to low frequencies might be equal to or close to 0. Such behavior contradicts the increasing trend of eigenvalues that we desire. Thus, the graph we construct should at least be a connected one.

For most of the time, we might not be able to know the prior information about the exact distribution of the signal f but we can construct the graph based on other information. For example, in field estimation, it is fairly reasonable to assume the value of nodes is highly correlated to its location, i.e., the nodes who are close to each other geographically are likely to have similar readings. Hence, we can build the graph based on the location information.

3. EXPERIMENT RESULTS

3.1. Experiment Setup

In this section, we investigate the performance of the GFT on the data from California Irrigation Management Information System (CIMIS) [11]. This dataset is generated by weather stations around the state of California. We use the solar radiation data for one day which contains 135 readings from different weather stations. We utilize KNN graphs based on the geological information of weather station to build its GFT basis.

We will compare the performance of compressed sensing [12, 13], linear approximation and non-linear approximation on

this dataset. For compressed sensing, we randomly select a number of the readings from the sensor nodes as the measurement and use LASSO in graph Fourier basis as the decoding algorithm. All the experiments are repeated 50 times and the average values are reported.



Fig. 1. Performance of Compressed Sensing, linear approximation and non-linear approximation.



Fig. 2. Performance of Compressed Sensing with different graph Fourier basis. M is the number of measurements. X axis shows the number of neighbors we use to formulate a symmetric KNN graph.

3.2. Results

Fig 1 illustrates the performance of CS, linear approximation and non-linear approximation with increasing compression rate when we set the parameter K = 7 to construct the underlying graph. The compression ratio is defined as $\frac{M}{N}$, where M is the number of measurements and N is the dimension of signal. Distortion is calculated with Mean Square Error(MSE). It is well implicated that the nonlinear approximation outperforms the other two methods, while linear approximation performs a little bit better than Compressed Sensing. All of the results in this figure show a rapid decaying behavior of MSE, which verifies that we can obtain a GFT basis where the original signal is compressible by proper graph construction.

Fig 2 describes explicitly how the connectivity of a graph affects the performance of compressed sensing. The result agrees with our earlier discussion about the choice of parameter K. Given a constant compression rate, the best performance of Compressed Sensing appears when K is in the range 5–10. When K is smaller than 5, the graph is unconnected with high probability. In this case, we have multiple zero eigenvalues. When K become larger than 30, the graph approximate the complete graph, which also gives a poor compressibility.

4. CONCLUSION

This paper analyzes a concept of the GFT. To the best of our knowledge, this is the first work to address why we can compress signals on the graph Fourier domain and what conditions the graph and signals should hold. We define the smoothness of signals supported on graphs and show its impact on the linear approximation. The GFT extends the conventional approximation theory to signals on graphs. We believe it has a lot of potential applications in the realm of sensor networks, computer graphics and compressed sensing.

5. REFERENCES

- W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proceedings of the 5th international conference on Information processing in sensor networks*. ACM, 2006, pp. 134–142.
- [2] W. Bajwa, A. Sayeed, and R. Nowak, "Matched source-channel communication for field estimation in wireless sensor networks," in *Proceedings of the 4th international symposium on Information processing in sensor networks*. IEEE Press, 2005, pp. 44–es.
- [3] Z. Karni and C. Gotsman, "Spectral compression of mesh geometry," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 2000, pp. 279–286.
- [4] M. Ben-Chen and C. Gotsman, "On the optimality of spectral compression of mesh data," ACM Transactions on Graphics (TOG), vol. 24, no. 1, pp. 60–80, 2005.
- [5] D. Ustebay, R. Castro, and M. Rabbat, "Efficient decentralized approximation via selective gossip," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 805–816, 2011.
- [6] D.I. Shuman, P. Vandergheynst, and P. Frossard, "Chebyshev polynomial approximation for distributed signal processing," *Arxiv preprint* arXiv:1105.1891, 2011.
- [7] X. Zhu and M. Rabbat, "Graph spectral compressed sensing for sensor networks," in *Proc. IEEE ICASSP*, Kyoto, Japan, 2012.
- [8] S.G. Mallat, A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way, Academic Press, 2008.
- [9] R.M. Gray, *Toeplitz and circulant matrices: A review*, Information Systems Laboratory, Stanford University, 1971.
- [10] X. Zhu, J. Kandola, J. Lafferty, and Z. Ghahramani, "Graph kernels by spectral transforms," in O. Chapelle, B. Schoelkopf, and A. Zien, editors, Semi- Supervised Learning. 2006, pp. 277–291, MIT Press.
- [11] "Cimis data," [Online]. Available: http://www.cimis.water.ca.gov, 2010.
- [12] D.L. Donoho, "Compressed sensing," *IEEE Transactions on Informa*tion Theory, vol. 52, no. 4, pp. 1289–1306, 2006.
- [13] E.J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.