Localized Fourier Analysis for Graph Signal Processing

Basile de Loynes, Fabien Navarro, Baptiste Olivier

Localized Fourier Analysis for Graph Signal Processing

Basile de Loynes*, Fabien Navarro†, Baptiste Olivier‡

June 10, 2020

Abstract

We propose a new point of view in the study of Fourier analysis on graphs, taking advantage of localization in the Fourier domain. For a signal f on vertices of a weighted graph $\mathcal{G}$ with Laplacian matrix $\mathcal{L}$, standard Fourier analysis of f relies on the study of functions $g(\mathcal{L})f$ for some filters g on $I_{\mathcal{L}}$, the smallest interval containing the Laplacian spectrum $\mathrm{sp}(\mathcal{L}) \subset I_{\mathcal{L}}$. We show that for carefully chosen partitions $I_{\mathcal{L}} = \sqcup_{1\leq k\leq K} I_k$ ($I_k \subset I_{\mathcal{L}}$), there are many advantages in understanding the collection $(g(\mathcal{L}_{I_k})f)_{1\leq k\leq K}$ instead of $g(\mathcal{L})f$ directly, where $\mathcal{L}_I$ is the projected matrix $P_I(\mathcal{L})\mathcal{L}$. First, the partition provides a convenient modelling for the study of theoretical properties of Fourier analysis and allows for new results in graph signal analysis (*e.g.* noise level estimation, Fourier support approximation). We extend the study of spectral graph wavelets to wavelets localized in the Fourier domain, called LocLets, and we show that well-known frames can be written in terms of LocLets. From a practical perspective, we highlight the interest of the proposed localized Fourier analysis through many experiments that show significant improvements in two different tasks on large graphs, noise level estimation and signal denoising. Moreover, efficient strategies permit to compute sequence $(g(\mathcal{L}_{I_k})f)_{1\leq k\leq K}$ with the same time complexity as for the computation of $g(\mathcal{L})f$.

Keywords: Nonparametric regression; Multiscale statistics; Variance estimation; Concentration inequalities; Graph signal processing; Spectral graph theory; Graph laplacian, Harmonic analysis on graphs

1 Introduction

Graphs provide a generic representation for modelling and processing data that reside on complex domains such as transportation or social networks. Numerous works combining both concepts from algebraic and spectral graphs with those from harmonic analysis (see for example [5, 6, 2] and references therein) have allowed to generalize fundamental notions from signal processing to the context of graphs thus giving rise to Graph Signal Processing (GSP). For an introduction to this emerging field and a review of recent developments and results see [26] and [21]. In general, two types of problems can be distinguished according to whether the underlying graph is known or unknown. The first case corresponds to the setup of a sampled signal at certain irregularly spaced points (intersections of a transportation network, nodes in a computer network, ...). In the second case, a graph is constructed from the data itself, it is generally interpreted as a noisy realization of one or several distributions supported by a submanifold of the Euclidean space. In this latter context, the theoretical submanifold is somehow approximated using standard methods such as k-NN, ε-graph and their Gaussian weighted versions. In any of these cases, the

*Basile de Loynes
ENSAI, France, E-mail: basile.deloynes@ensai.fr
†Fabien Navarro
CREST, ENSAI, France, E-mail: fabien.navarro@ensai.fr
‡Baptiste Olivier
Orange Labs, France. E-mail: baptiste.olivier@orange.com

framework is actually similar: it consists of a graph (given by the application or by the data) and signals are real-valued functions defined on the vertices of the graph.

Notions of graph Fourier analysis for signals on graphs were introduced and studied over the past several years [26, 23, 27, 24]. The graph Fourier basis is given by the eigenbasis $(\chi_\ell)_\ell$ of the Laplacian matrix $\mathcal{L}$. The Graph Fourier Transform (GFT) consists in representing a signal f in the Fourier basis $(\langle f, \chi_\ell \rangle)_\ell$, and by analogy with the standard case the eigenvalues of $\mathcal{L}$ play the role of *frequencies*. From this definition, it follows that many filtering techniques are written in terms of vectors $g(\mathcal{L})f$, for some filter functions g which act on the spectrum of $\mathcal{L}$ (scaling, selecting, ...). Fourier analysis on graphs has been successfully applied to many different fields such as stationary signals on graphs [22], graph signal energy study [14], convolutional neural networks on graphs [10].

Graph wavelets are an important application of graph Fourier analysis, and several definitions of graph wavelets were proposed [7, 6, 13, 18, 31, 15]. When performing Fourier analysis of a signal, there is no guarantee that localization of a signal in the frequency domain (a.k.a Fourier domain) implies localization in the graph domain. This phenomenon is illustrated by the fact that the eigenvectors corresponding to the upper part of Laplacian spectrum tend to be more oscillating than those from the bottom of the spectrum (see for example [32, Fig. 1.6, p. 28] for an illustration). To overcome this problem, [16] developed a fairly general construction of a frame enjoying the usual properties of standard wavelets: each vector of the frame is defined as a function $g(s\mathcal{L})\delta_m$ (where δ_m is a signal with zero values at every vertex except m) and is localized both in the graph domain and the spectral domain at fine scale s. The transform associated with this frame is named Spectral Graph Wavelet Transform (SGWT), and it was used in numerous subsequent works [30, 1, 15].

Signals which are sparse in the Fourier domain form an important class of graph signals. Indeed, there is a tight relationship between sparsity in the Fourier domain and the notion of regularity of a signal f on the vertices of a graph $\mathcal{G}$ which comes from the Laplacian matrix $\mathcal{L}$ of $\mathcal{G}$. Intuitively, a smooth signal will not vary much between two vertices that are close in the graph. This regularity property can be read in the Fourier domain: a very smooth signal will be correctly represented in the Fourier domain with a small number of eigenvectors associated with the lower spectral values; on the contrary, non-smooth signals (*i.e.* highly oscillating) are represented with eigenvectors corresponding to the upper part of the spectrum. Both the types of signal are said *frequency sparse*.

In this paper, we propose to exploit localization in the Fourier domain to improve graph Fourier analysis. More precisely, we consider vectors of the form $g(\mathcal{L}_{I_k})f$ instead of vectors $g(\mathcal{L})f$ in graph Fourier analysis, where $\mathcal{L}_{I_k}$ is defined as the matrix $\mathcal{L}P_{I_k}(\mathcal{L})$ and $P_{I_k}(\mathcal{L})$ denotes the projection onto the eigenspaces whose eigenvalue is contained in subset I_k. Localized Fourier analysis is motivated by problems and properties defined on strict subsets of the spectrum $\mathrm{sp}(\mathcal{L})$ (*e.g.* any problem defined in terms of frequency sparse graph signals). As a central application fo Fourier localization, we introduce the Fourier localized counterpart of SGWT, that we call *LocLets* for *Loc*alized graph wave*lets*. We prove that various frame constructions can be written in terms of LocLets, hence benefiting from all the advantages of localization discussed in this paper.

Defining $I_{\mathcal{L}}$ as the smallest interval containing the entire spectrum $\mathrm{sp}(\mathcal{L})$, the local Fourier analysis consists in choosing a suitable partition $I_{\mathcal{L}} = \sqcup_k I_k$ into subintervals on which standard Fourier analysis is performed. Such an analysis on disjoint intervals naturally benefits from several interesting properties. In particular, when f is modeled by a Gaussian random vector with independent entries, the disjointness of subintervals preserves these properties in the sense that random variables $(g(\mathcal{L}_{I_k})f)_k$ are still Gaussian and independent. This simple observation has important consequences to study the graph problem at stake. In this work, it allows us to propose some noise level estimator from the random variables sequence $(g(\mathcal{L}_{I_k})f)_k$, and to provide a theoretical analysis of the denoising problem. Disjointness of subsets $(I_k)_k$ also provides simple strategies to parallelize Fourier analysis computations.

We also consider the general problem given by a noisy signal on a graph $\widetilde{f} = f + \xi$, where ξ is some random Gaussian vector with noise level σ. We provide results for two important tasks: the estimation of σ when the latter is unknown, and the denoising of noisy signal $\widetilde{f}$ in order to recover signal f. We show

that for frequency sparse signals, localization allows to adapt to the unknown Fourier support of signal f. Theoretical guarantees and practical experiments show that localized Fourier analysis can improve state-of-the-art denoising techniques, not only in precision of the estimator $\widehat{f}$ of f, but also in time computations.

We provide an efficient method to choose a partition $I_{\mathcal{L}} = \sqcup_k I_k$ for the Fourier localized vectors $g(\mathcal{L}_{I_k})f$ to be sufficiently informative. Using well-known techniques for efficient graph Fourier analysis (a.k.a Chebyshev filters approximations), we propose scalable methods to perform localized Fourier analysis with no computational overhead over standard fast Fourier graph analysis.

The paper is structured as follows. Section 2 presents the relevant notions and techniques necessary to perform localized Fourier analysis. In Section 2.2, we introduce LocLets, the Fourier localized extension of SGWT. Section 3 is devoted to the study of the denoising problem for signals on graphs. The section provides results about noise level estimation, and signal denoising. Additional properties of LocLets, such as computational aspects and relationships with known wavelet transforms, are further developed in Section 4. In Section 5, we analyze the experiments made to support the interesting properties of localized Fourier analysis highlighted in this paper. Finally, the proofs are gathered in Section 7.

2 Localized Fourier analysis for graph signals

In this section, we introduce the central notion studied in this paper: localization of graph Fourier analysis. First, we recall the relevant notions of graph Fourier analysis. Then we introduce LocLets, an important application of Fourier localization to SGWT.

2.1 Functional calculus and Fourier analysis for graph signals

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected weighted graph with $\mathcal{V}$ the set of vertices, $\mathcal{E}$ the set of edges, $n = |\mathcal{V}|$ the number of nodes (the *size* of $\mathcal{G}$), and $(W_{ij})_{i,j\leq n}$ the weights on edges. Let us introduce the diagonal degree matrix whose diagonal coefficients are given by $D_{ii} = \sum_{1\leq j\leq n} W_{ij}$ for $1 \leq i \leq n$. The resulting non-normalized Laplacian matrix $\mathcal{L}$ of graph $\mathcal{G}$ is defined as $\mathcal{L} = D - W$. The n non-negative eigenvalues of $\mathcal{L}$, counted without multiplicity, are denoted by $\lambda_1, \ldots, \lambda_n$ in the decreasing order. In the sequel, $\mathrm{sp}(\mathcal{L})$ stands for the spectrum of $\mathcal{L}$. The corresponding eigenvectors are denoted $\chi_1, \ldots, \chi_n$.

Given a graph $\mathcal{G}$, the GFT of a real-valued function f defined on the vertices of $\mathcal{G}$ is nothing but the representation of f in the orthonormal basis of eigenvectors of $\mathcal{L}$. Namely, for a signal $f : \mathcal{G} \to \mathbb{R}$, the ℓ-th Fourier coefficient of f, denoted $\widehat{f}(\ell)$, is given by $\widehat{f}(\ell) = \langle f, \chi_\ell \rangle$. The Fourier support $\mathrm{supp}(\widehat{f})$ of signal f is the set of indices ℓ such that $\widehat{f}(\ell) \neq 0$. We will see in Section 2.2 that graph wavelets can be defined in a similar manner.

Functional calculus is a powerful technique to study matrices which is the heart of GSP. For a function g defined on some domain D_g, $\mathrm{sp}(\mathcal{L}) \subset D_g$, functional calculus reads as

$$g(\mathcal{L}) = \sum_{1\leq \ell \leq n} g(\lambda_\ell) \langle \chi_\ell, \cdot \rangle \chi_\ell.$$

Interpreting the eigenvalues λ_ℓ, $\ell = 1, \ldots, n$, as the fundamental frequencies associated with a graph, the linear map $g(\mathcal{L})$ is generally seen as a filter operator in terms of signal analysis.

Also, spectral projections of matrix $\mathcal{L}$ can be explicited with the help of functional calculus setting $g = 1_I$. More precisely, for any subset $I \subset I_{\mathcal{L}}$, consider the map $P_I(\mathcal{L})$ given by:

$$P_I(\mathcal{L}) = \sum_{1\leq \ell \leq n} 1_I(\lambda_\ell) \langle \chi_\ell, \cdot \rangle \chi_l = \sum_{\ell:\ \lambda_\ell \in I} \langle \chi_\ell, \cdot \rangle \chi_\ell.$$

Then, $P_I(\mathcal{L})$ is nothing but the spectral projection on the linear subspace spanned by the eigevectors associated with the eigenvalues belonging to I. In the sequel, $n_I = |I \cap \mathrm{sp}(\mathcal{L})|$ will stand for the number of eigenvalues contained in subset $I \cap \mathrm{sp}(\mathcal{L})$.

Spectral projections are a practical tool to focus on some part of the spectrum $\mathrm{sp}(\mathcal{L})$. More precisely, let $I_{\mathcal{L}} = \sqcup_{1\leq k\leq K} I_k$ be a partition of interval $I_{\mathcal{L}}$ into disjoint subsets $(I_k)_k$. Since intervals I_k are disjoints, functional analysis of $\mathcal{L}$ reduces to that of its projections $\mathcal{L}_{I_k} = \mathcal{L}P_{I_k}(\mathcal{L})$ in the sense of the identity:

$$g(\mathcal{L}) = \sum_{1\leq k\leq K} g(\mathcal{L}_{I_k}).$$

In this paper, one will study the extent to which Fourier analysis on large graphs is improved when considering *local* Fourier analysis on each subset I_k instead of *global* Fourier analysis on $I_{\mathcal{L}}$.

2.2 LocLets: a localized version of SGWT

This section introduces an important application of the localized graph Fourier analysis, namely the notion of localized SGWT.

2.2.1 Construction of a SGWT

Let $f : \mathcal{G} \to \mathbb{R}$ be a signal on the graph $\mathcal{G}$. Let $\varphi, \psi : \mathbb{R} \to \mathbb{R}$ be respectively the scaling and kernel functions (a.k.a. *father* and *mother* wavelet functions), and let $s_j > 0$, $1 \leq j \leq J$, be some scale values. The discrete SGWT is defined in [16] as follows:

$$\mathcal{W}f = (\varphi(\mathcal{L})f^T, \psi(s_1\mathcal{L})f^T, \ldots, \psi(s_J\mathcal{L})f^T)^T.$$

The adjoint matrix $\mathcal{W}^*$ of $\mathcal{W}$ is:

$$\mathcal{W}^*(\eta_0^T, \eta_1^T, \ldots, \eta_J^T)^T = \varphi(\mathcal{L})\eta_0 + \sum_{j=1}^{J} \psi(s_j\mathcal{L})\eta_j. \tag{1}$$

We also recall from [16] that a discrete transform reconstruction formula using SGWT coefficients $(c_{j,m})_{\substack{0\leq j\leq J\\ 1\leq m\leq n}}$ is obtained by the formula

$$(\mathcal{W}^*\mathcal{W})^{-1}\mathcal{W}^*(c_{j,m})_{j,m},$$

where $(\mathcal{W}^*\mathcal{W})^{-1}$ stands for a pseudo-inverse of matrix $\mathcal{W}^*\mathcal{W}$.

2.2.2 Definition of LocLets

Spectral graph wavelet functions are given by $(\varphi(\mathcal{L})\delta_m, \psi(s_j\mathcal{L})\delta_m)_{1\leq j\leq J, 1\leq m\leq n}$. We define a LocLet function to be the projection of a graph wavelet function onto a subset of eigenspaces of $\mathcal{L}$.

Definition 1. *Let $(\varphi(\mathcal{L})\delta_m, \psi(s_j\mathcal{L})\delta_m)_{1\leq j\leq J, 1\leq m\leq n}$ be the family functions induced by a SGWT. Then, for any subset $I \subset I_{\mathcal{L}}$, $1 \leq j \leq J$ and $1 \leq m \leq n$, set:*

$$\varphi_{m,I} = \varphi(\mathcal{L}_I)\delta_m = \sum_{\ell:\lambda_\ell\in I} \varphi(\lambda_\ell)\widehat{\delta}_m(\ell)\chi_\ell$$

$$\psi_{j,m,I} = \psi(s_j\mathcal{L}_I)\delta_m = \sum_{\ell:\lambda_\ell\in I} \psi(s_j\lambda_\ell)\widehat{\delta}_m(\ell)\chi_\ell.$$

The functions $(\varphi_{m,I}, \psi_{j,m,I})_{1\leq j\leq J, 1\leq m\leq n}$ are called Localized waveLets functions (LocLets). The functions $\varphi_{m,I}, \psi_{j,m,I}$ are said to be localized at I.

Let $I_{\mathcal{L}} = \sqcup_{1\leq k\leq K} I_k$ be some partition. Then, the localized SGWT transfom of f with respect to partition $(I_k)_{1\leq k\leq K}$, denoted by $\mathcal{W}^{(I_k)_k}f$, is defined as the family $\mathcal{W}^{(I_k)_k}f = (\mathcal{W}^{I_k}f)_k$ where

$$\mathcal{W}^{I_k}f = (\varphi(\mathcal{L}_{I_k})f^T, \psi(s_1\mathcal{L}_{I_k})f^T, \ldots)^T, \quad 1 \leq k \leq K.$$

Similarly to Equation (1), the adjoint transform is given by

$$\mathcal{W}^{I_k *}(\eta_0^T, \eta_1^T, \dots, \eta_J^T)^T = \varphi(\mathcal{L}_{I_k})\eta_0 + \sum_{j=1}^{J} \psi(s_j \mathcal{L}_{I_k})\eta_j, \quad 1 \leq k \leq K.$$

As already observed, localized SGWT of a signal f contains more precise information about signal f than its standard SGWT. The latter can easily be obtained from the former since subsets $(I_k)_k$ are pairwise disjoint and formula $g(s\mathcal{L}) = \sum_{1\leq k\leq K} g(s\mathcal{L}_{I_k})$ holds for all filter g, and in particular for $g = \varphi$ or $g = \psi$. When the partition $I_{\mathcal{L}} = \sqcup_{1\leq k\leq K} I_k$ is carefully chosen, we show that the SGWT localization provides interesting features such as independence of random variables in denoising modelling, or considerable improvements in denoising tasks.

Remark 2. *A different* localization *property is studied in [16]. The latter refers to the vanishing of coefficients $\langle \psi_{j,m}, \delta_{m'} \rangle$ when the geodesic distance in the graph between vertices m and m' is sufficiently large (see [16, Theorem 5.5]). This notion is a property observable in the graph domain (through impulse functions δ_m and $\delta_{m'}$), and is different from the notion of Fourier localization discussed in the current paper. We refer to the localization property in [16] as* graph domain localization.

3 Local Fourier analysis and graph functions denoising

The denoising problem is stated as follows: given an observed noisy signal $\widetilde{f}$ of the form $\widetilde{f} = f + \xi$ where ξ is a n-dimensional Gaussian vector distributed as $\mathcal{N}(0, \sigma^2 \mathrm{Id})$, provide an estimator of the *a priori* unknown signal f.

This section shows how localized Fourier analysis helps in estimating the noise level σ when it is unknown, and in recovering the original signal f when the latter is sparse in the Fourier domain. In what follows, we will focus on random variables of the form $\|P_{I_k}\widetilde{f}_{I_k}\|_2$ where $\widetilde{f}$ is the noisy signal and I_k is a subset in the partition $I_{\mathcal{L}} = \sqcup_k I_k$. To keep the notations light, n_k, f_k, ξ_k and $\widetilde{f}_k$ will stand for n_{I_k}, $P_{I_k} f$, $P_{I_k}\xi$ and $P_{I_k}\widetilde{f}$ respectively. In addition, the cumulative distribution function of a random variable X will be denoted by Φ_X.

3.1 Noise level estimation for frequency sparse signals

Since in real application the noise level σ remains unknown in general, new estimators $\widehat{\sigma}$ based on localization properties in the spectrum are introduced in the sequel.

3.1.1 Noise level estimation from projections along $\mathrm{sp}(\mathcal{L})$

For any filter g defined on $I_{\mathcal{L}}$ and any subset $I \subset I_{\mathcal{L}}$, simple computation gives rises to

$$\mathbb{E}(\widetilde{f}^T g(\mathcal{L}_I)\widetilde{f}) = f^T g(\mathcal{L}_I) f + \sigma^2 \mathrm{Tr}(g(\mathcal{L}_I)). \tag{2}$$

Since both $\widetilde{f}^T g(\mathcal{L}_I)\widetilde{f}$ and $\mathrm{Tr}(g(\mathcal{L}_I))$ are known, Equation (2) suggests building estimators from the expression $\frac{\widetilde{f}^T g(\mathcal{L}_I)\widetilde{f}}{\mathrm{Tr}(g(\mathcal{L}_I))}$. In [9], the noise level is estimated by $\frac{\widetilde{f}^T \mathcal{L} \widetilde{f}}{\mathrm{Tr}(\mathcal{L})}$ which can be seen as the graph analog of the Von Neumann estimator from [33]. The main drawback of this estimator is its bias.

Theoretically, without any assumption on the signal f, the bias term $\frac{f^T g(\mathcal{L}_I) f}{\mathrm{Tr}(g(\mathcal{L}_I))}$ is minimized when $g = 1_{\{\lambda_{\ell^*}\}}$ where $\ell^* = \mathrm{argmin}\{|\widehat{f}(\ell)| : \lambda_\ell \in \mathrm{sp}(\mathcal{L})\}$. The computation of such filters would require the complete reduction of $\mathcal{L}$ which does not scale well with the size of the graph. Instead, these ideal filters will be approximated by filters of the form $g = 1_{I_k}$, for I_k a subset in the partition $I_{\mathcal{L}} = \sqcup_k I_k$. It is worth noting that with $k^* = \mathrm{argmin}_k \|f_k\|_2$, the function $g^* = 1_{I_{k^*}}$ achieves the minimal bias of the estimator among all filters of the form $g = \sum_k \alpha_k 1_{I_k}$.

Discarding some intervals I_k with $n_k = 0$, it can be assumed without loss of generality that $n_k \neq 0$ for all $1 \leq k \leq K$. Also observe that the random variable $\|\widetilde{f}_k\|_2^2$ can be decomposed as follows

$$\|\widetilde{f}_k\|_2^2 = \|f_k\|_2^2 + \|\xi_k\|_2^2 + 2\langle f_k, \xi_k\rangle \tag{3}$$

where $\frac{\|\xi_k\|_2^2}{\sigma^2}$ and $\frac{\langle f_k,\xi_k\rangle}{\sigma}$ are random variables distributed as $\chi^2(n_k)$ and $\mathcal{N}(0, \|f_k\|_2^2)$ respectively.

Proposition 3. *Let $(c_k)_{1\leq k\leq K}$ be the sequence of non-negative random variables defined, for all $k = 1,\dots,K$, by $c_k = \|\widetilde{f}_k\|_2^2/n_k$. Then,*

1. *the random variables $c_1,\dots,c_K$ are independent;*
2. *for all k, k' such that $f_k = f_{k'}$, c_k and $c_{k'}$ are identically distributed if and only if $n_k = n_{k'}$;*
3. *for k such that $f_k = 0$, c_k is distributed as $\frac{\sigma^2}{n_k}\Gamma_{n_k}$ where $\Gamma_{n_k} \sim \chi^2(n_k)$.*

3.1.2 The case of frequency-sparse signals

When the signal f is sparse in the Fourier domain, the condition $f_k = 0$ is met for most of the intervals $I_k \subset I_{\mathcal{L}}$. Let us define $I_f = \sqcup_{k:I_k\cap \mathrm{supp}\widehat{f}\neq\emptyset} I_k$ to be the union of subsets I_k intersecting the Fourier support $\mathrm{supp}(\widehat{f})$ of f. Also, denote by $\overline{I_f} = I_L\backslash I_f$ its complement set. In order to take advantage of Fourier sparsity, let us introduce the quantities $\widehat{\sigma}_{\mathrm{mean}}$ and $\widehat{\sigma}_{\mathrm{med}}$ as follows:

$$\widehat{\sigma}_{\mathrm{mean}}(c)^2 = \frac{1}{|\{k : I_k \subset \overline{I_f}\}|}\sum_{k:I_k\subset\overline{I_f}} c_k \quad \text{and} \quad \widehat{\sigma}_{\mathrm{med}}(c)^2 = \mathrm{median}_{k:I_k\subset\overline{I_f}}(c_k). \tag{4}$$

The following concentration inequalities show that $\widehat{\sigma}_{\mathrm{mean}}$ and $\widehat{\sigma}_{\mathrm{med}}$ are natural estimators of the noise level σ.

Proposition 4. *Let $K_f = |\{k : I_k \subset \overline{I_f}\}|$, $n_0 = \min\{n_k : k, I_k \subset \overline{I_f}\}$, $n_\infty = \max\{n_k : k, I_k \subset \overline{I_f}\}$, $V_f = 2\sigma^4\sum_{k:I_k\subset\overline{I_f}} 1/n_k$ and $B_f = 2\sigma^2/n_0$. Then the following concentration inequalities hold:*

1. *for all $t \geq 0$,*

$$\mathbb{P}\left(\widehat{\sigma}_{\mathrm{mean}}(c)^2 - \sigma^2 \geq t\right) \leq \exp\left(-\frac{K_f^2 t^2}{V_f(1 + B_f + \sqrt{1 + \frac{2B_f K_f t}{V_f}})}\right),$$

and for all $0 \leq t \leq \sigma^2$,

$$\mathbb{P}\left(\widehat{\sigma}_{\mathrm{mean}}(c)^2 - \sigma^2 \leq -t\right) \leq \exp\left(-\frac{K_f^2 t^2}{2V_f}\right);$$

2. *for all $t \geq 0$, with $\beta = n_0/n_\infty$,*

$$\mathbb{P}\left(\widehat{\sigma}_{\mathrm{med}}^2 \geq \beta^{-1}\sigma^2 + 2\sigma^2\beta^{-1}t\right) \leq \exp\left(\frac{K_f}{2}\ln\left[4p^+(t)(1-p^+(t))\right]\right),$$

and for all $0 \leq t \leq 1$ such that $p^-(t) \leq 1/2$,

$$\mathbb{P}\left(\widehat{\sigma}_{\mathrm{med}}^2 \leq \beta\sigma^2 - \sigma^2\beta t\right) \leq \exp\left(\frac{K_f}{2}\ln\left[4p^-(t)(1-p^-(t))\right]\right),$$

where

$$p^+(t) = \mathbb{P}(\Gamma_{n_\infty} \geq n_\infty + 2n_\infty t) \quad \textit{and} \quad p^-(t) = \mathbb{P}(\Gamma_{n_0} \leq n_0 - n_0 t).$$

Obviously, the Fourier support $\mathrm{supp}(\widehat{f})$ and the subset $\overline{I_f}$ remain generally unknown in applications and have to be approximated. Let us recall that the main issue for estimating σ comes from the bias term $\frac{\|f_k\|_2^2}{n_k}$ in Equation (2), and in particular when the value σ^2 is negligible compared to $\frac{\|f_k\|_2^2}{n_k}$. Therefore, a suitable candidate to approximate $\overline{I_f}$ will be some subset $\overline{J_f} \subset I_{\mathcal{L}}$ for which the impact of larger values $\frac{\|f_k\|_2^2}{n_k}$ is minimized. This is made clear by Proposition 5 below. The latter involves the following concentration bounds for Gaussian random variables: for all $0 < \alpha < 1$

$$\mathbb{P}(|\langle f_k, \xi_k\rangle| \geq t_{\alpha,\sigma}\|f_k\|_2) \leq \alpha \quad \text{where} \quad t_{\alpha,\sigma} = \sigma \times \sqrt{-2\ln\Big(\frac{\alpha}{4}\Big)}. \tag{5}$$

Proposition 5. *Let $0 < \alpha < 1$. Let $t_{\alpha,\sigma}$ be defined by Equation (5). Assume that $f_\ell = 0$ and that the following inequality holds:*

$$\frac{\|f_k\|_2^2 + 2t_{\alpha,\sigma}\|f_k\|_2}{\sigma^2} \geq \Phi^{-1}_{\frac{n_k}{n_\ell}\Gamma_{n_\ell}-\Gamma_{n_k}}\left(1 - \frac{3\alpha}{2}\right).$$

Then, the quantities

$$b_k = \frac{\|\xi_k\|_2^2 + \|f_k\|_2^2 + 2\langle \xi_k, f_k\rangle}{n_k} \quad \textit{and} \quad b_\ell = \frac{\|\xi_\ell\|_2^2}{n_\ell}$$

satisfy $\mathbb{P}(b_k \geq b_\ell) \geq 1 - \alpha$.

By invariance under permutations, one may assume without loss of generality that the values c_k are ordered in the decreasing order. Proposition 5 quantifies the fact that the highest values of c_k correspond most likely to the indices k for which $f_k \neq 0$. Consequently, setting $\overline{J_f}(r) = \sqcup_{k\in\{r,r+1,\ldots K-r\}} I_k$ for all $1 \leq r \leq \frac{K}{2}$, the estimators introduced in Equation (4) may be rewritten replacing the unknown subset $\overline{I_f}$ by its known approximation $\overline{J_f}(r)$. So we define the estimators

$$\widehat{\sigma}^r_{\mathrm{mean}}(c)^2 = \frac{1}{|\{k : I_k \subset \overline{J_f}(r)\}|}\sum_{k:I_k\subset\overline{J_f}(r)} c_k \quad \text{and} \quad \widehat{\sigma}^r_{\mathrm{med}}(c)^2 = \mathrm{med}_{k:I_k\subset\overline{J_f}(r)}(c_k).$$

It is worth noting that from the symmetry of the subset $\overline{J}_f(r)$, it follows that the value $\widehat{\sigma}^r_{\mathrm{med}}$ actually does not depend on parameter r, and one will write $\widehat{\sigma}_{\mathrm{med}}$ in place of $\widehat{\sigma}^r_{\mathrm{med}}$.

3.2 Denoising Frequency Sparse Signals

Let us begin with a result that illustrates that localized Fourier analysis in $I_{\mathcal{L}}$ provides strong benefits in noise reduction tasks when the underlying is frequency sparse.

Proposition 6. *Assume $f = f_I$ for some subset $I \subset I_{\mathcal{L}}$. Then*

$$\mathbb{E}\left[\left\|f - \widetilde{f}_I\right\|_2^2\right] = \mathbb{E}\left[\left\|f - \widetilde{f}\right\|_2^2\right] - \sigma^2\big|\overline{I} \cap \mathrm{sp}(\mathcal{L})\big|.$$

In particular, denoising of $\widetilde{f}$ boils down to denoising of $\widetilde{f}_I = f_I + \xi_I$.

While Proposition 6 asserts a trivial denoising solution in the Fourier domain, *i.e.* simply destroying the projection $\widetilde{f}_{\overline{I}} = \xi_{\overline{I}}$, this approach is no longer that immediate when considering the graph domain observations since the Fourier support of f is unknown in practice and needs to be estimated. Based on the χ^2-statistics, Algorithm 1 is designed for this purpose. To the best of our knowledge, previous works that proposed method for Fourier support recovery for graph noisy signals [25] involve the complete eigendecomposition of matrix $\mathcal{L}$. The methodology suggested below makes use of projectors on eigenspaces which can be approximating with Chebyshev polynomials as detailed in the next Section 4.

Algorithm 1: Support approximation in the Fourier domain for noisy signal

Data: noisy signal $\widetilde{f}$, a subdivision $I_1, I_2, \ldots, I_K$, estimated $n_k = |I_k \cap \mathrm{sp}L|$, $k = 1, \ldots, K$, threshold $\alpha \in (0,1)$

Result: $\widetilde{f}_I = P_I(\mathcal{L})\widetilde{f}$, where I is an approximation of the Fourier support of $\widetilde{f}$

1 for $k = 1, \ldots, K$

2 Compute $\|\widetilde{f}_k\|_2^2 = \|P_{I_k}(\mathcal{L})\widetilde{f}\|_2^2$;

3 Compute

$$p_k = \mathbb{P}(\sigma^2\Gamma_{n_k} > \|\widetilde{f}_k\|_2^2) \quad \text{and} \quad \Gamma_{n_k} \sim \chi^2(n_k);$$

4 Compute $\widetilde{f}_I = \sum_{k:\ p_k \le \alpha} P_{I_k}\widetilde{f}$.

Heuristically, if I contains the support of the Fourier transform of f, on the complementary subset $\overline{I}$ we only observe pure white Gaussian noise so that $\|P_{\overline{I}}\widetilde{f}\|_2^2 = \|\widetilde{f}_{\overline{I}}\|_2^2$ is distributed as $\sigma^2\chi^2(n_I)$ with $n_I = |\overline{I} \cap \mathrm{sp}(L)|$. On the other hand, on I the square of the Euclidean norm of a non-centered Gaussian vector is observed. Consequently, the quantity $\mathbb{P}\left(\chi^2(n_I) > \sigma^{-2}\|P_I\widetilde{f}\|_2^2\right)$ is typically very close to zero whereas $\mathbb{P}\left(\chi^2(n-n_I) > \sigma^{-2}\|P_{\overline{I}}\widetilde{f}\|_2^2\right)$ remains away from 0. To put it in a nutshell, sliding a window along the spectrum of $\mathcal{L}$, Algorithm 1 performs a series of χ^2-test.

With the objective to provide theoretical guarantees that χ^2-tests approach $\mathrm{supp}(\widehat{f})$ correctly, it is important to turn the condition on the p_k-value into a condition involving only the values $\|f_k\|_2$ and σ. The next lemma shows that for sufficiently large values of the ratio $\frac{\|f_k\|_2}{\sigma}$, the inequality $p_k \le \alpha$ holds so that the corresponding components $\mathrm{supp}(\widehat{f}_k)$ of the Fourier domain are legitimately included in the support estimate I.

Lemma 7. *Let $0 < \alpha < 1$ and let $\Gamma_{n_k}, \Gamma'_{n_k}$ be two* i.i.d *$\chi^2(n_k)$ random variables. Assume that:*

$$\frac{\|f_k\|_2}{\sigma}\left(\frac{\|f_k\|_2}{\sigma} - 2\frac{t_{\alpha/2,\sigma}}{\sigma}\right) \ge \Phi^{-1}_{\Gamma_{n_k}-\Gamma'_{n_k}}\left(1-\frac{\alpha}{2}\right),$$

where $t_{\alpha,\sigma}$ is defined by Equation (5)*. Then $p_k \le \alpha$.*

In contrast to Lemma 7, the lemma below states that condition $p_k > \alpha$ holds for sufficiently small values of ratio $\sigma^{-1}\|f_k\|_2$.

Lemma 8. *Let $0 < \alpha < 1$ and let Γ_{n_k} be a $\chi^2(n_k)$ random variable. For $0 < \beta < 1$, set $t_{\beta,k} = \sigma^2\Phi^{-1}_{\Gamma_{n_k}}(1-\beta)$. Assume that*

$$\left(\frac{\|f_k\|_2 + \sqrt{t_{\beta,k}}}{\sigma}\right)^2 < \Phi^{-1}_{\Gamma_{n_k}}\left(1-\frac{\alpha}{1-\beta}\right).$$

Then $p_k > \alpha$.

Compared to Proposition 6, the result below quantifies the error resulting by approximating the support running Algorithm 1. Note that the requirement to have a constant sequence $(n_k)_k$ is used for statement clarity but similar assertions hold for the case $n_k \neq n_{k'}$.

Proposition 9. *Set $f_I = \sum_{k:p_k\le\alpha} P_{I_k}f$. Assume that $n_k = n_1$ for all $1 \le k \le K$. Then,*

1. *the Fourier support approximation ℓ_2-error satisfies*

$$\|f - f_I\|_2^2 \le |\{k, I_k \subset I_f, p_k > \alpha\}|\left(t_{\alpha/2,\sigma} + \sqrt{t^2_{\alpha/2,\sigma} + \left(\sigma\Phi^{-1}_{\Gamma_{n_1}-\Gamma'_{n_1}}\left(1-\frac{\alpha}{2}\right)\right)^2}\right)^2. \tag{6}$$

2. the Noise ℓ_2-error on Fourier support:

$$\mathbb{E}\|f_I - \widetilde{f}_I\|_2^2 = |\{k, p_k \leq \alpha\}| n_1 \sigma^2. \tag{7}$$

Lemma 8 asserts that the set $\{k, p_k > \alpha\}$ is small when most of the values $\|f_k\|_2$ are large enough compared to noise level σ for $I_k \cap \mathrm{supp}\widehat{f} \neq \emptyset$. In such a case, Fourier support approximation ℓ_2-error is small. Regarding the noise ℓ_2-error, the inclusion $\{k, p_k \leq \alpha\} \subset \{k, I_k \cap \mathrm{supp}\widehat{f} \neq \emptyset\}$ holds by Lemma 8. Moreover, Lemma 7 asserts that the set $\{k, p_k \leq \alpha\}$ contains the entire set $\{k, I_k \cap \mathrm{supp}\widehat{f} \neq \emptyset\}$ for sufficiently large values of $\sigma^{-1}\|f_k\|_2$ when $I_k \cap \mathrm{supp}\widehat{f} \neq \emptyset$. For such favorable situations, the noise ℓ_2-error is exactly $n_1\sigma|\{k, I_k \subset I_f\}|$, the amount of noise on the extended support I_f.

Algorithm 2: LocLets thresholding estimation procedure

Data: $\widetilde{f}$, α, $(I_k)_{k=1,\dots,K}$, estimated $n_k = |I_k \cap \mathrm{sp}(\mathcal{L})|$, thresholds t_1, t_2
Result: estimator $\widehat{f}$ of signal f

1 Apply Algorithm 1 with $\widetilde{f}$, α, $(I_k)_{k=1,\dots K}$, estimated n_k; it outputs $\widetilde{f}_I$ and $\widetilde{f}_{\overline{I}}$;
2 Apply soft-thresholding with threshold t_1 to $\mathcal{W}^I\widetilde{f}$ and t_2 to $\mathcal{W}^{\overline{I}}\widetilde{f}$;
3 Apply the inverse LocLet transform to the soft-thresholded coefficients to obtain $\widehat{f}_I, \widehat{f}_{\widetilde{I}}$;
4 Compute the estimator $\widehat{f} = \widehat{f}_I + \widehat{f}_{\overline{I}}$;

The second step gives an estimate of the original signal using a thresholding procedure on each element $\widetilde{f}_I$ and $\widetilde{f}_{\overline{I}}$. On the one hand, the methodology developed in [15] is prohibitive in terms of time and space complexity as soon as the underlying graphs become moderately large. On the other hand, the fast SGWT remains an approximating procedure. If a signal happens to be very frequency-sparse, then an even more optimal strategy is possible: first, the support I in the frequency domain is approximated with the help of Algorithm 1; then, the procedure of [15] is applied to $P_I f$ (the low-rank part) and LocLets on $P_{\overline{I}}(\mathcal{L})f$. This idea is made precisely in Algorithm 3.

Algorithm 3: LocLets support approximation, and low-rank Parseval Frame thresholding procedure

Data: $\widetilde{f}$, α, $(I_k)_{k=1,\dots,K}$, estimated $n_k = |I_k \cap \mathrm{sp}(\mathcal{L})|$, thresholds t_1, t_2
Result: estimator $\widehat{f}$ of signal f

1 Apply Algorithm 1 with $\widetilde{f}$, α, $(I_k)_{k=1,\dots K}$, estimated n_k; it outputs $\widetilde{f}_I$ and $\widetilde{f}_{\overline{I}}$;
2 Compute Parseval Frame for $\mathcal{L}_I$;
3 Apply Parseval Frame thresholding with threshold t_1 to $\widetilde{f}_I$; it outputs $\widehat{f}_I$;
4 Apply soft-thresholding with threshold t_2 to $\mathcal{W}^{\overline{I}}\widetilde{f}$;
5 Apply the inverse LocLet transform to the soft-thresholded coefficients to obtain $\widehat{f}_{\widetilde{I}}$;
6 Compute the estimator $\widehat{f} = \widehat{f}_I + \widehat{f}_{\overline{I}}$;

Estimator $\widehat{f}$ produced in Algorithm 3 satisfies a tighter oracle bound inequality than the one given in [15, Theorem 3]. This theoretical guarantee is widely supported by our experiments described in Section 5. Following notations from [15, Equation (21)], we denote by $OB(f_I)$ the oracle bound obtained from an *oracle* estimator of f_I from a noisy $\widetilde{f}_I$ exploiting some knowledge about the unknown signal f_I. We refer to [15] for precise details.

Theorem 10. *Let $I, \widehat{f}$ be respectively the support approximation and the estimator of f obtained from Algorithms 1 and 3 with threshold value $t_2 = 0$. Then we have*

$$\mathbb{E}\|f - \widehat{f}\|_2^2 \leq \mathbb{E}\|f - f_I\|_2^2 + (2\log(n_I) + 1)(\sigma^2 + OB(f_I)).$$

The right-hand side in the inequality of Theorem 10 has a more explicit expression in terms of α, σ using Proposition 9. Up to the error made by approximating the support with Algorithm 1, the ℓ_2-risk is essentially bounded by the ℓ_2-risk of the Parseval frame procedure from [15] on the low-rank projection f_I of f, that is

$$\mathbb{E}\|f - \widehat{f}\|_2^2 \lesssim (2\log(n_I) + 1)(\sigma^2 + OB(f_I)).$$

To conclude, Theorem 10 provides a theoretical guarantee that the support approximation improves the denoising performances obtained from [15].

4 Properties of LocLets

In this section, we highlight important properties for the application of Fourier localization in practice. First we discuss computational analysis, and methods to apply our techniques to large graphs. Then we study the relationships of LocLets with well-known graph wavelet constructions.

4.1 Fast LocLet Transform and Computational Analysis

In the case of large graphs, GSP requires a special care for being efficient since functional calculus relies *a priori* on the complete reduction of the Laplacian. Actually, several efficient methods were designed to retrieve only partial information from the eigendecomposition as matrix reduction techniques (see for instance [20, 30]) or polynomial approximations [16, 28, 11]. In this paper, the widely adopted latter approach with Chebyshev polynomials approximation is preferred and briefly recalled below (we refer the reader to [28, Section III.C.] for a brief but more detailed description of Chebyshev approximation).

4.1.1 Chebyshev approximations

Roughly speaking, the idea is to approximate the function g with its Chebyshev expansion g_N at order N. More precisely, the Chebyshev polynomials of the first kind $(T_i)_{i\geq 0}$ are defined from the second order recursion

$$T_0(x) = 1, \quad T_1(x) = x, \quad |x| \leq 1, \quad \text{and} \quad T_i(x) = xT_{i-1}(x) - T_{i-2}(x),$$

for $i \geq 2$. Then, the matrix $\mathcal{L}$ is *normalized* as $\widetilde{\mathcal{L}} = \frac{2}{\lambda_1}\mathcal{L} - I_n$ so that $\mathrm{sp}(\widetilde{\mathcal{L}}) \subset [-1, 1]$. This gives rise to some function $\widetilde{g} : [-1, 1] \to \mathbb{R}$ with the property $g(\mathcal{L}) = \widetilde{g}(\widetilde{\mathcal{L}})$. In fact, $\widetilde{g}(x) = g(\frac{\lambda_1}{2}(x + 1))$ for all $x \in [-1, 1]$. Then $g(\mathcal{L})$ has the following truncated Chebyshev expansion $g(\mathcal{L}) \approx g_N(\mathcal{L})$:

$$g_N(\mathcal{L}) = \sum_{0\leq i\leq N} a_i(\widetilde{g})T_i(\widetilde{\mathcal{L}}),$$

where N is the maximal degree of polynomials T_i used in the expansion, and $a_i(\widetilde{g})$ is the i-th coefficient in the N-th order Chebyshev expansion of function $\widetilde{g}$. Following [16], for any filter g on $\mathrm{sp}(\mathcal{L})$ and any signal f on graph $\mathcal{G}$, the approximation $g_N(\mathcal{L})$ provides a vector value close to $g(\mathcal{L})f$ with time complexity $O(|\mathcal{E}|N)$.

The object presented in the sequel involves in particular the spectral projection $P_I(\mathcal{L})f$ of a signal f for any subset $I \subset I_{\mathcal{L}}$ which can be derived from the Chebyshev expansion of the indicator function $g = 1_I$. This observation actually appears in several recent works [11, 12]. More importantly for our study, this efficient estimation is part of the Hutchinson stochastic trace estimator technique [17], providing us with an effective method to estimate $n_I = \mathrm{Tr}(\mathcal{L}_\mathrm{I})$. Finally, the present paper focuses on the computation of a sequence $g(\mathcal{L}_{I_k})_{1\leq k\leq K}$ (or its vector counterpart $g(\mathcal{L}_{I_k})f$) instead of a single $g(\mathcal{L})$ (resp. $g(\mathcal{L})f$). While a naive estimation would suggest that the computational complexity is then multiplied by a factor K compared to the complexity of the computation of $g(\mathcal{L})$, we argue in the following that there is in fact no significant computational overhead.

4.1.2 Sharing Chebyshev polynomial matrices among filters

Let us assume that it is needed to compute the estimated values of $g_k(\mathcal{L})f$ for a given signal f for several filters g_k, $k = 0, \dots, K$. Then the following two-step strategy can be adopted: (1) pre-compute Chebyshev expansions $\widetilde{g}_k(x) \approx \widetilde{g}_{k,N}(x) = \sum_{0\le i\le N} a_i(\widetilde{g}_k)T_i(x)$ for all $k = 0, \dots, K$; independently, compute Chebyshev approximation vectors $T_i(\widetilde{\mathcal{L}})f$ for all $0 \le i \le N$; (2) combine the previous results to compute the Chebyshev approximation $g_{k,N}(\mathcal{L})f$ of $g_k(\mathcal{L})f$:

$$g_{k,N}(\mathcal{L})f = \sum_{0\le i\le N} a_i(\widetilde{g}_k)T_i(\widetilde{\mathcal{L}})f.$$

The complexity of the first step is dominated by the N matrix-vector multiplications required to obtain $T_i(\widetilde{\mathcal{L}})f$. So the first step has complexity $O(|\mathcal{E}|N)$. The second step adds N weighted matrices $a_i(\widetilde{g}_k)T_i(\widetilde{\mathcal{L}})$ together, which is an operation of complexity $O(Nn^2)$ at most. As an important matter of fact, the overall complexity for this procedure is bounded by $O(|\mathcal{E}|N + Nn^2)$, which is independent of the number of filters g_k, and the same as for the computation of $g(\mathcal{L})$.

Sharing matrices among filters has several examples of applications in the current paper:

1. Computation of $g(\mathcal{L}_{I_k})f$ for all $1 \le k \le K$: the equation $g(\mathcal{L}_{I_k})f = g(1_{I_k}(\mathcal{L})\mathcal{L})f$ holds so that we can consider filters $g_k(x) = g(1_{I_k}(x)x)$.

2. Computation of $g(s\mathcal{L})f$ for several scale values s: consider filters of the form $g_s(x) = g(sx)$.

3. Computation of n_{I_k} for all $1 \le k \le K$: Hutchinson's stochastic estimation computes averages of $f_i^T P_{I_k}(\mathcal{L})f_i$ for some random vectors f_i ($i \le n_H$) whose computational complexity is dominated by the approximation of vectors $P_{I_k}f_i$. Considering filters $g_k(x) = 1_{I_k}(x)$, and sharing random vectors $(f_i)_i$ among all approximations of n_k, we end up with a complexity of $O(n_H N|\mathcal{E}|)$, independent of value K.

In particular, Algorithm 1 has complexity $O(n_H N|\mathcal{E}| + Nn^2)$. Indeed, its efficiency is calibrated on the computations of sequences $(\|\widetilde{f}_k\|_2)_{1\le k\le K}$ and $(n_k)_{1\le k\le K}$ whose computational analysis was discussed previously. It is worth observing that values n_k do not depend on signal f and should be estimated only once in the case where several signals $\widetilde{f}_1, \widetilde{f}_2, \dots$ are to be denoised.

4.1.3 Optimizing storage of LocLets coefficients

The storage of wavelet coefficients $(\mathcal{W}^{I_k}f)_{1\le k\le K}$ requires *a priori* K times the storage cost associated with the original transform $\mathcal{W}f$. When matrix reduction techniques are used to compute wavelets transform [30], one may reduce the storage consumption of the localized SGWT by suitably choosing the *impulse* functions $(\delta_m)_m$. For instance, assume that for each subset I_k a Lanczos basis $(v_m^k)_m$ of the subspace spanned by $\{\chi_\ell, \ell \in I_k\}$ is given. Then the size of sequences $(v_m^k)_m$ and $(v_m^k)_{m,k}$ are respectively of order $O(|I_k \cap \mathrm{sp}(\mathcal{L})|) = O(n_k)$ and $O(n)$. Thus, with impulse functions $(v_m^k)_n$ in place of δ_m for transform $\mathcal{W}^{I_k}$, the storage requirements of localized transform $(\mathcal{W}^{I_k}f)_{1\le k\le K}$ and the original one $\mathcal{W}f$ are of the same order $O(Jn)$.

4.2 Connections with Well-know Frames

A family $\mathfrak{F} = \{r_i\}_{i\in I}$ of vectors of $\mathbb{R}^{\mathcal{V}}$ is a frame if there exist $A, B > 0$ satisfying for all $f \in \mathbb{R}^{\mathcal{V}}$

$$A\|f\|_2^2 \le \sum_{i\in I} |\langle f, r_i\rangle|^2 \le B\|f\|_2^2.$$

A frame is said to be *tight* if $A = B$. This section gives two examples of frames introduced in the literature which can be realized as a LocLets representation and thus benefit from the advantages given by localization in the spectrum.

4.2.1 Parseval frames

Parseval frames are powerful representations to design wavelets with nice reconstruction properties [18, 15]. In this section, we investigate the extent to which Parseval frames can be obtained from some LocLet representation. We show that for a particular choice of partition $I_{\mathcal{L}} = \sqcup_k I_k$, there exist frames which are Parseval frames and composed only of LocLets functions.

A finite collection $(\psi_j)_{j=0,\dots,J}$ is a finite partition of unity on the compact $[0, \lambda_1]$ if

$$\psi_j : [0, \lambda_1] \to [0, 1] \quad \text{for all} \quad j \leq J \quad \text{and} \quad \forall \lambda \in [0, \lambda_1], \quad \sum_{j=0}^{J} \psi_j(\lambda) = 1. \tag{8}$$

Given a finite partition of unity $(\psi_j)_{j=0,\dots,J}$, the Parseval identity implies that the following set of vectors is a tight frame:

$$\mathfrak{F} = \left\{ \sqrt{\psi_j}(\mathcal{L})\delta_i, \quad j = 0, \dots, J, \quad i \in V \right\}.$$

Some constructions of partition of unity involve functions $(\psi_j)_j$ that have almost pairwise disjoint supports *i.e.* $\operatorname{supp}(\psi_j) \cap \operatorname{supp}(\psi_{j'}) = \emptyset$ as soon as $|j - j'| > 1$. For such partition of unity, set $I_0 = \operatorname{supp}(\psi_0)$, $I_J = I_0 = \operatorname{supp}(\psi_J)$ and $I_j = \operatorname{supp}(\psi_j) \cap \operatorname{supp}(\psi_{j+1})$ for all $1 \leq j \leq J-1$. Then, the sequence $(I_j)_{0 \leq j \leq J}$ defines a finite partition of $[0, \lambda_1]$, $[0, \lambda_1] = \sqcup_{0 \leq j \leq J} I_j$, such that:

$$\psi_0 1_{I_0} = 1_{I_0}, \quad (\psi_j + \psi_{j+1}) 1_{I_j} = 1_{I_j}, \quad 0 < j < J, \quad \text{and} \quad \psi_J 1_{I_J} = 1_{I_J}. \tag{9}$$

An alternative tight frame can be constructed using a LocLet representation as shown in the following proposition.

Proposition 11. *Assume Equations* (8) *and* (9) *hold and set, for all* $1 \leq k \leq J$, $\varphi_{n,k} = \sqrt{\psi_0}(\mathcal{L}_{I_k})\delta_n$, $\psi_{1,n,k} = \sqrt{\psi_k}(\mathcal{L}_{I_k})\delta_n$ *and* $\psi_{2,n,k} = \sqrt{\psi_k}(\mathcal{L}_{I_{k+1}})\delta_n$ *for all* $1 \leq k \leq J$. *Then* $(\varphi_{n,k}, \psi_{j,n,k})_{1 \leq j \leq 2,\ 1 \leq m \leq n,\ 1 \leq k \leq J}$ *is a tight frame.*

The resulting tight frame of Proposition 11 is actually frame of LocLets if additionally the functions ψ_j is of the form $\psi_j = \psi_1(s_j .)$ for some scale parameter s_j, $1 \leq j \leq J$. This is typically the case for the frames introduced in [18, 15]. In these papers, the partition of unity is defined as follows: let $\omega : \mathbb{R}^+ \to [0, 1]$ be some function with support in $[0, 1]$, satisfying $\omega \equiv 1$ on $[0, b^{-1}]$ and set $\psi_0(\cdot) = \omega(\cdot)$ and for $j = 1, \dots, J$

$$\psi_j(\cdot) = \omega(b^{-j}\cdot) - \omega(b^{-j+1}\cdot) \quad \text{with} \quad J = \left\lfloor \frac{\log \lambda_1}{\log b} \right\rfloor + 2.$$

In particular, the functions ψ_k have supports in intervals $J_k = [b^{k-2}, b^k]$. Thus, one may define disjoint intervals $(I_k)_k$ as follows: $I_k = [b^{k-1}, b^k]$. We have $J_k = I_k \cup I_{k+1}$, so that Equations (9) hold whereas the scaling property $\psi_j = \psi_1(b^{-1}\cdot)$ is straightforward. By Proposition 11, the set of vectors

$$\left\{ \sqrt{\psi_0}(\mathcal{L}_{I_k})\delta_n, \sqrt{\psi_1}(s_k\mathcal{L}_{I_k})\delta_n, \sqrt{\psi_1}(s_k\mathcal{L}_{I_{k+1}})\delta_n, \quad n, k \right\}$$

is a tight frame of LocLets. Observe that the transform $(\mathcal{W}^{I_k})_{I_k}$each component $\mathcal{W}^{I_k}$ of the LocLet transform $(\mathcal{W}^{I_k})_{I_k}$ only admit two scale parameters s_k, s_{k-1}.

4.2.2 Spectrum-adapted tight frames

Let us consider another family of tight frames tailored to the distribution of the Laplacian $\mathcal{L}$ eigenvalues proposed in [1]. As shown below, these frames can be written in terms of a warped version of LocLets, and up to some approximation, in terms of (non-warped) LocLets. First, let us briefly recall the construction from [1].

The notion of *warped* SGWT is introduced in [1] to adapt the kernel to the spectral distribution. Given a warping function $\omega : I_{\mathcal{L}} \to \mathbb{R}$, the warped SGWT is defined as:

$$\mathcal{W}^{\omega} f = (\varphi(\omega(\mathcal{L}))f^T, \psi(s_1\omega(\mathcal{L}))f^T, \ldots, \psi(s_J\omega(\mathcal{L}))f^T)^T.$$

As for our spectral localization, the objective of warping is to take benefits from the distribution of $\mathrm{sp}(\mathcal{L})$ along interval $I_{\mathcal{L}}$. While the two techniques show similarities (*e.g.* estimation of $\mathrm{sp}(\mathcal{L})$ distribution), they are meant to answer different problems: warped SGWT is a technique to adapt the whole spectrum to some task (*e.g.* producing a tight frame), whereas localized SGWT is designed to answer problems related to localized subsets in the spectrum (*e.g.* denoising a frequency sparse signal). Here we show that the advantages of both LocLets and warped SGWT are obtained when the two methods are combined in a warped LocLet representation.

Let ω be some warping function on $I_{\mathcal{L}}$ chosen in the form $\omega(\cdot) = \log(C\omega_0(\cdot))$ where ω_0 stands for the cumulative spectral distribution of $\mathcal{L}$ and C is some normalization constant as shown in [1]. Then, let $\gamma > 0$ be an upper bound on $\mathrm{sp}(\mathcal{L})$ and let R, J be two integers such that $2 \leq R \leq J$. Setting $\omega_{\gamma,J,R} = \frac{\gamma}{J+1+R}$, Corollary 2 in [1] asserts that the family $(g_{m,j})_{m,j}$ of functions defined below is a tight frame

$$g_{m,j} = \sum_{\ell} \widehat{g}_j(\lambda_\ell)\widehat{\delta}_m(\ell)\chi_\ell, \tag{10}$$

where functions $\widehat{g}_j$ arise from some kernel $\widehat{g}$ as

$$\widehat{g}_j(\lambda) = \widehat{g}(\omega(\lambda) - j\omega_{\gamma,J,R}) = \widehat{g}\left(\log \frac{C\omega(\lambda)}{e^{j\omega_{\gamma,J,R}}}\right).$$

Typically in [1], the kernel $\widehat{g}$ takes the form

$$\widehat{g}(\lambda) = \left[\sum_{0\leq j\leq J} a_j \cos\left(2\pi j \cos\left(\frac{\lambda}{R\omega_{\gamma,J,R}} + \frac{1}{2}\right)\right)\right] 1_{[-R\omega_{\gamma,J,R},0]}(\lambda).$$

for some sequence $(a_j)_j$ satisfying $\sum_j (-1)^j a_j = 0$.

The following proposition states that Equation (10) admits an alternative form involving only (warped) LocLets functions.

Proposition 12. *Setting $\psi(\lambda) = \widehat{g}(\log(C\lambda))$ for $\lambda > 0$, consider the family of warped LocLets defined for all $0 \leq k \leq R-1$, $1 \leq m \leq n$ and $1 \leq j \leq J$ by*

$$\psi_{j,m,I_k} = \sum_{\ell\in I_k} \psi(s_j\omega_0(\lambda_\ell))\widehat{\delta}_m(\ell)\chi_\ell \quad \textit{with} \quad I_k = \left[\frac{e^{(k-R)\omega_{\gamma,J,R}}}{C}, \frac{e^{(k-R+1)\omega_{\gamma,J,R}}}{C}\right].$$

Then, the following identity holds for all $j = 1, \ldots, J$ and all $m = 1, \ldots, n$

$$g_{m,j} = \sum_{1\leq k\leq R-1} \psi_{j,m,I_k}.$$

5 Experiments on some large matrices suite

This section details experiments made on large graphs to validate the Fourier localization techniques introduced in that paper. After describing the experimental settings, we describe the outcomes of several experiments showing strong advantages in the use of Fourier localization in practice.

5.1 Choice of spectral partition $I_{\mathcal{L}} = \sqcup_k I_k$

In order to keep the problem combinatorially tractable, it is necessary to reduce the choice of possible partitions of $I_{\mathcal{L}}$ into subintervals I_k. That is why, the partitions considered in the sequel are regular in the sense that all intervals have the same length λ_1/K for some integer $K \geq 1$. Thereafter, the parameter K is chosen so that the eigenvalues are distributed as evenly as possible in each interval I_k. Without prior information, it is indeed natural not to favor one part of the spectrum over another. Most importantly, in the view of the concentration property of the median around the noise level σ^2 of Proposition 4, it is essential to keep the parameter β as close to one as possible.

In order to implement the ideas above, it is necessary to estimate the spectral measure of $\mathcal{L}$ which can be described by the so-called spectral density function:

$$\varphi_{\mathcal{L}}(\lambda) = \frac{1}{n}\sum_{\ell=1}^{n} \delta(\lambda - \lambda_\ell) \quad \text{for all} \quad \lambda \in I_{\mathcal{L}}.$$

There are several techniques for such an approximation among which the *Kernel Polynomial Method* (see, *e.g.* [29, 34]). The latter approximates the spectral density $\varphi_{\mathcal{L}}$ with the help of matrix Chebyshev expansion $(\varphi_{\mathcal{L}}^N)_N$ (see [19] for a detailed presentation).

Now, let $(I_k)_{1\leq k\leq K}$ be some regular partition of $I_{\mathcal{L}}$ and $(n_k)_{1\leq k\leq K}$ be the corresponding numbers of eigenvalues in each I_k. Choosing the parameter $K \geq 1$ so that the entropy defined by

$$E(K) = -\sum_{1\leq k\leq K} \frac{n_k}{n} \log\left(\frac{n_k}{n}\right)$$

is maximal ensures that the eigenvalues are as equally distributed in each interval as possible. In application, the Kernel Polynomial Method provides an approximation n_k^N of n_k and the corresponding empirical entropy $E_N(K)$ is used as a proxy for the theoretical one.

Empirically, the entropy increases logarithmically and then stabilizes from a certain elbow value K_{elbow} as illustrated in Figure 1. This elbow value is displayed in dashed lines in Figure 1. In the experiments, we choose this value K_{elbow} motivated by two reasons. First, as the intervals become shorter it is more difficult to obtain a uniform repartition of the eigenvalues into those intervals. The second reason is related to the quality of the estimate n_k^N of n_k as the sample size decreases. To illustrate this fact, we consider the Mean Relative Error (MRE) defined by

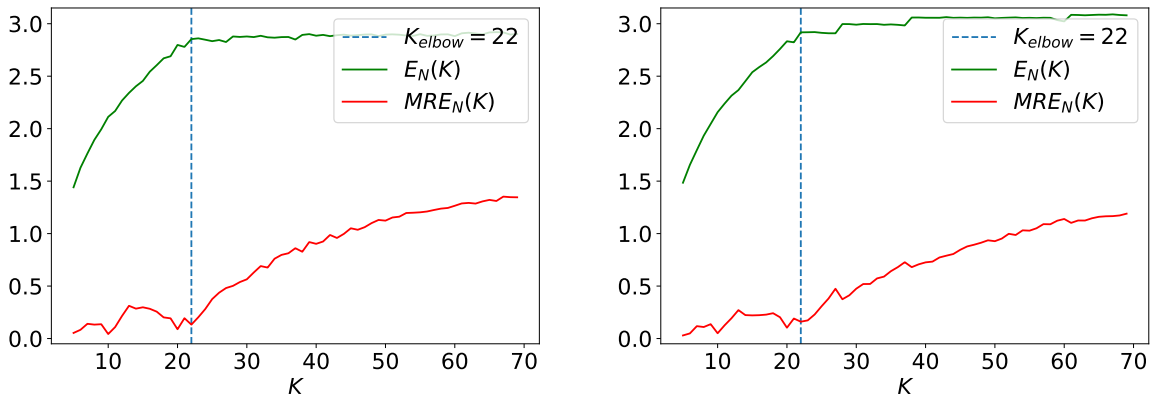
$$\text{MRE}_N(K) = \frac{\sum_{1\leq k\leq K} |n_k - n_k^N|}{n}.$$

As highlighted by Figure 1, the empirical entropy actually stabilizes when the Chebyshev approximation, in terms of MRE, is no longer sharp enough.

5.2 The experimental settings

Following [12], we propose to validate our techniques on an extended suite of large matrices extracted from the Sparse Matrix Collection in [8]. Most of these matrices have an interpretation as the Laplacian matrix of a large graph. We define matrix $\mathcal{L}$ from the following matrices of the suite:*si2* ($n = 769$), *minnesota* ($n = 2642$), *cage9* ($n = 3534$), *saylr4* ($n = 3564$) and *net25* ($n = 9520$). We extend this graph collection with the well-studied *swissroll* graph Laplacian matrix ($n = 1000$).

We sample randomly signals whose support are sparse in the Fourier domain. We will use the notation f_{i-j} for normalized signals supported on a sub-interval of $I_{\mathcal{L}}$ containing exactly the eigenvalues $\lambda_i, \lambda_{i+1}, \ldots, \lambda_j$. As an example, f_{n-n} is a constant signal while f_{1-2} is a highly non-smooth signal supported on the eigenspaces of large eigenvalues λ_1, λ_2. For experiments, the signals were calculated from the knowledge of $\text{sp}(\mathcal{L})$, and relevant projections of random functions on the graph.



(a) *Si2* graph (b) *minnesota* graph

Figure 1: Variations of $E_N(K)$ and $\mathrm{MRE}_N(K)$ with parameter K.

We have compared the performances of Algorithms 2 and 3 against the thresholding procedure described in [15]. As the denoising method in [15] requires the computation of the whole spectral decomposition of the Laplacian, it does not scale to large graphs. We stress here that we provide a fair comparison with [15], only in terms of denoising performance, and with no computational considerations. Moreover, we choose for LocLets to use the most naive thresholding procedure by considering a global and scale independent threshold level.

For all the experiments below, the SGWT and LocLets are built upon the scale and kernel functions giving rise to the Parseval frame of [15], whose construction is recalled in Section 4.2.1. More precisely, set respectively $\varphi = \sqrt{\zeta_0}$ and $\psi = \sqrt{\zeta_1}$ for the scale and kernel functions with $\zeta_0(x) = \omega(x)$, and $\zeta_1(x) = \omega(b^{-1}x) - \omega(x)$, where we choose $b = 2$ and ω is piecewise linear, vanishes on $[1, \infty)$ and is constant equal to one on $(-\infty, b]$. The scales are of the form $s_j = b^{-j+1}$ for $j = 1, \dots, J$ where J is chosen similarly to [15].

In what follows, 'PF' stands for Parseval Frame and refers to the estimator of [15]; the estimators implemented by Algorithm 2 and Algorithm 3 are referred to as 'LLet' and 'LLet+PF' respectively. The notation '$\mathrm{SNR}_{\mathrm{in}}$' refers to the trivial model releasing the noisy signal $\widetilde{f}$, corresponding to the classical input noise level measurement, and serves as a worst-case baseline for other models. Below, the latter methodology is shown to outperform all the others for very frequency-sparse signals. It is also worth recalling that 'LLet+PF' benefits from the dimension reduction property of LocLets. More precisely, whereas the whole eigendecomposition of $\mathcal{L}$ is required to apply 'PF', for Parseval frame denoising in the context of 'LLet+PF', only a low-rank spectral decomposition is needed, namely the decomposition of $\mathcal{L}_I$ for I the estimate of $\mathrm{supp}\widehat{f}$.

For all our experiments, we set $\alpha = 0.001$ for Algorithm 1. For the denoising experiment, we compute the best SNR result r_D over a large grid of values (t_1, t_2), and for each denoising method D with $D \in \{$'$\mathrm{SNR}_{\mathrm{in}}$','PF','LLet','LLet+PF'$\}$. Then, we calculate two metrics: the maximum M_D and average value μ_D of the values r_D over 10 random perturbations of the signal f. We recall that a good quality in denoising is reflected by a large value of the SNR metric.

5.3 Analysis of our experiments

5.3.1 Noise level estimation

We have evaluated the performances of estimators $\widehat{\sigma}^r_{\mathrm{mean}}$ and $\widehat{\sigma}_{\mathrm{med}}$ in the estimation of the unknown noise level σ from 10 realizations of the noisy signal $\widetilde{f} = f + \xi$ for a given noise level σ. Figure 2 (resp. Figure 3) shows the best performances of each estimator on the *minnesota* (resp. *net25*) graph for the non-regular but frequency sparse signal $f_{1392-1343}$ (resp. $f = f_{4971-5020}$), when parameter K ranges in

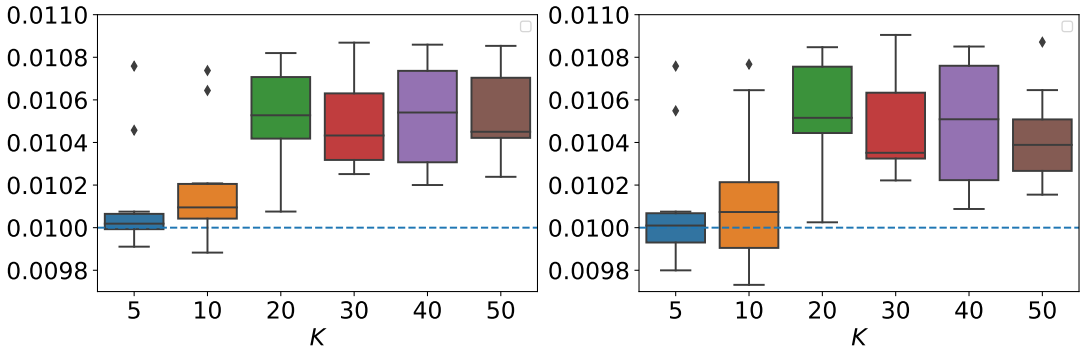


Figure 2: Performances of estimators $\widehat{\sigma}_{\text{mean}}$ (left) and $\widehat{\sigma}_{\text{med}}$ (right) for *minnesota* graph, signal $f_{1392-1343}$ and $\sigma = 0.01$.

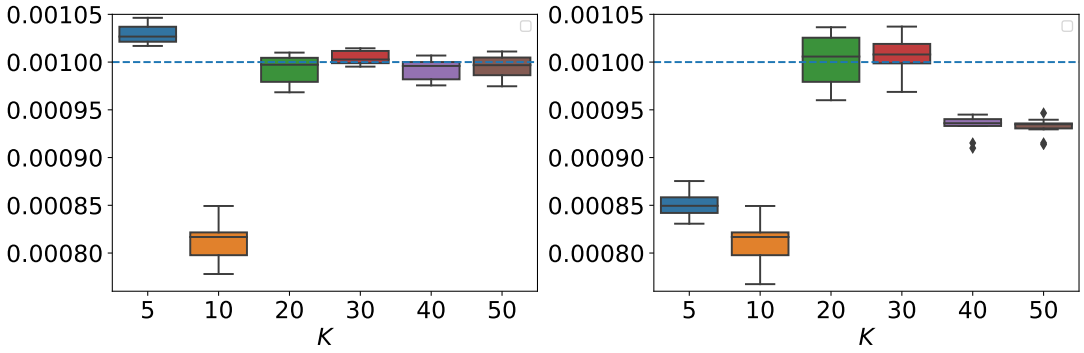


Figure 3: Performances of estimators $\widehat{\sigma}^r_{\text{mean}}$ (left) and $\widehat{\sigma}^r_{\text{med}}$ (right) for *net25* graph, signal $f_{4971-5020}$ and $\sigma = 0.001$.

$\{5, 10, 20, 30, 40, 50\}$ and for level of noise $\sigma = 0.01$ (resp. $\sigma = 0.001$).

Figure 2 illustrates that both estimators $\widehat{\sigma}^r_{\text{mean}}$ and $\widehat{\sigma}_{\text{med}}$ can provide good estimates of σ. Best performances are obtained for values of parameter K below the elbow value $K_{elbow}(minnesota) = 22$ introduced in Section 5.1. We observe that performances drop considerably if almost no localization is used (for instance, for parameter values $K = 1$ or $K = 2$, $\widehat{\sigma} \sim 0.021$ in the experiment of Figure 2, far from the performances for $K \geq 5$ for estimating $\sigma = 0.01$).

Figure 3 shows that localization is necessary, namely $K \geq 10$ or even $K \geq 20$, in order to reach the best performances for the large *net25* graph. Contrary to experiments for the *minnesota* graph, estimators $\widehat{\sigma}^r_{\text{mean}}$ and $\widehat{\sigma}_{\text{med}}$ underestimate the value of σ. Also, best values of K range between 10 and 30 for *net25* graph, compared to best values $K = 5$ and $K = 10$ for *minnesota* graph (see Figure 2). This illustrates the idea that noise level estimation strongly depends on the underlying graph structure. As a consequence, a parameter K selection has to take graph and signal information into account to be relevant. Interestingly, the elbow values $K_{elbow}(minnesota) = 22$ and $K_{elbow}(net25) = 22$ provide performances which are not optimal, but close to the best possible ones.

In Figure 4, performances for various values of parameter r are displayed for a fixed parameter $K = K_{elbow}(minnesota)$. While it is true that $\widehat{\sigma}^r_{\text{mean}}$ can perform better than $\widehat{\sigma}^r_{\text{med}}$, it happens only for very specific values of r, which a priori depend on the signal regularity. Without any further parameter selection, these observations suggest using the most robust estimator $\widehat{\sigma}_{\text{med}}$ in practice.

5.3.2 Sparse signal denoising

As a first denoising experiment, we have compared the performances of 'LLet', 'PF' and 'LLet+PF' for a fixed value $K = K_{elbow}$ given by the rule of thumb described in Section 5.1. For each matrix in the Extended Matrices Suite, we have experimented the denoising task on two frequency-sparse signals, one

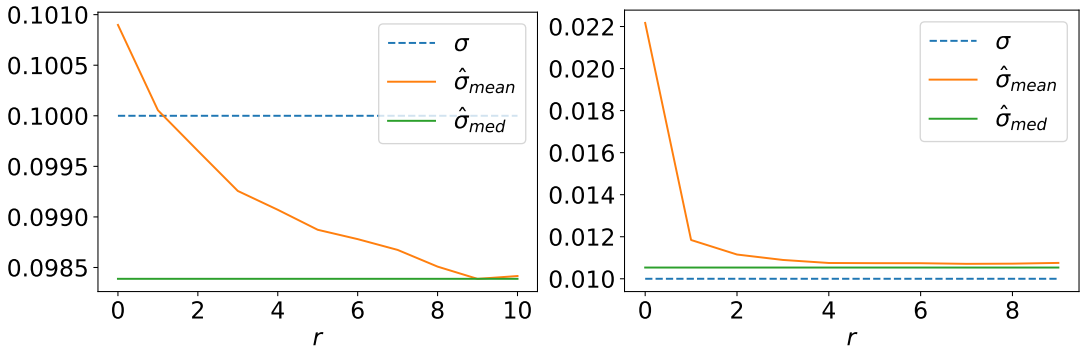


Figure 4: Dependence on parameter r for *minnesota* graph, signal $f_{1392-1343}$, $K = 22$ and $\sigma = 0.1$ (left), $\sigma = 0.01$ (right).

regular and the other non-regular. Several values of noise level σ were used, corresponding to values of $\mathrm{SNR}_{\mathrm{in}}$ ranging in $[4, 18]$. Results from experiments are displayed in Tables 5.3.2 and 2. The first obvious observation is that 'LLet+PF' performs better than its competitors in almost all situations. The gain is sometimes considerable since we observed a gap of 5dB in μ_D-metric between 'LLet+PF' and its closest concurrent 'PF' in some cases, and up to 7dB in M_D-metric. These experiments confirm the theoretical guarantees obtained in Theorem 10. The benefits of localization are reduced for graph *net25*: Table 2 shows that the more conservative choice $K = 5$ is better than $K = 25$. It appears that for *net25*, the spectrum $\mathrm{sp}(\mathcal{L})$ is localized at a small number of distinct eigenvalues, hence diminishing the advantages of localizing with our methods.

Table 1: SNR performance for Swissroll ($n = 1000$, $K = 22$).

signal	σ	$\mathrm{SNR}_{\mathrm{in}}$	M_{PF}	$\mathrm{M}_{\mathrm{LLet}}$	$\mathrm{M}_{\mathrm{LLet+PF}}$	μ_{PF}	μ_{LLet}	$\mu_{\mathrm{LLet+PF}}$
$f_{951-1000}$	0.005	16.195	17.557	20.580	20.528	17.361	20.035	19.974
$f_{501-550}$	0.005	15.859	18.298	8.244	20.821	18.044	8.140	20.245
$f_{951-1000}$	0.01	10.267	12.183	15.564	15.701	10.433	13.652	13.760
$f_{501-550}$	0.01	10.178	13.121	7.879	16.204	11.165	7.646	14.518
$f_{951-1000}$	0.015	6.763	9.430	12.661	13.129	8.961	12.127	12.388
$f_{501-550}$	0.015	6.362	9.898	7.611	14.159	9.540	7.481	13.398

Another interesting observation is that 'LLet' may outperform 'PF' in some specific signal and noise level configurations, as shown in Table 5.3.2. This is a very favorable result for localized Fourier analysis, since 'LLet' appears to be a technique which is more accurate and more efficient as well compared to 'PF' in some situations. However in many cases, 'LLet' performances drop down compared to the more stable thresholding techniques 'PF' and 'LLet+PF', which use thresholds adapted to the wavelet basis.

We also provide experimental results to understand the extent to which our results depend on the partition size parameter K. A few remarks are suggested by Figure 5:

- The best performances are not obtained for the elbow value K_{elbow}, suggesting searching for a more task-adapted size of partition K.
- Good performances persist for values of K much larger than K_{elbow}, and in particular for regular signals.
- For large values of K, there is a severe drop in performances. As explained before, the error

Table 2: SNR performances for denoising task.

matrix	signal	σ	$\mathrm{SNR}_{\mathrm{in}}$	M_{PF}	$\mathrm{M}_{\mathrm{LLet+PF}}$	μ_{PF}	$\mu_{\mathrm{LLet+PF}}$
Si2	$f_{720-769}$	0.005	17.104	22.344	26.973	21.849	25.170
($n = 762$, $K = 22$)	$f_{370-419}$	0.005	16.820	18.034	21.778	17.813	20.673
	$f_{720-769}$	0.01	11.408	16.821	22.501	16.572	20.558
	$f_{370-419}$	0.01	11.175	12.444	19.740	12.151	17.121
	$f_{720-769}$	0.02	4.826	11.476	15.695	11.104	14.711
	$f_{370-419}$	0.02	5.047	7.354	13.542	6.925	12.677
Minnesota	$f_{2593-2642}$	0.004	13.599	17.839	20.717	17.672	20.035
($n = 2642$, $K = 22$)	$f_{1343-1392}$	0.004	13.741	15.999	20.388	15.822	19.234
	$f_{2593-2642}$	0.005	11.681	16.086	19.342	15.830	18.417
	$f_{1343-1392}$	0.005	11.916	14.459	18.392	14.298	18.029
	$f_{2593-2642}$	0.01	5.911	10.875	14.143	10.605	13.556
	$f_{1343-1392}$	0.01	5.843	9.660	11.762	9.409	10.952
Cage9	$f_{3485-3534}$	0.003	15.016	20.477	9.700	20.216	9.664
($n = 3534$, $K = 22$)	$f_{1785-1834}$	0.003	15.014	15.876	16.945	15.798	16.799
	$f_{3485-3534}$	0.005	10.503	17.290	18.410	16.772	18.185
	$f_{1785-1834}$	0.005	10.507	12.118	13.032	12.035	12.898
	$f_{3485-3534}$	0.009	5.423	13.002	13.264	12.763	12.898
	$f_{1785-1834}$	0.009	5.395	8.329	10.468	8.168	9.827
Saylr4	$f_{3515-3564}$	0.003	14.871	23.108	24.516	23.040	24.117
($n = 3564$, $K = 22$)	$f_{2015-2064}$	0.003	15.069	21.365	23.903	21.010	23.412
	$f_{3515-3564}$	0.005	10.478	19.135	20.966	18.943	20.268
	$f_{2015-2064}$	0.005	10.635	17.016	19.610	16.662	18.732
	$f_{3515-3564}$	0.009	5.420	15.277	17.070	14.791	16.567
	$f_{2015-2064}$	0.009	5.480	12.055	14.802	11.830	14.031
Net25	$f_{9471-9520}$	0.006	4.682	5.171	5.319	5.094	5.205
($n = 9520$, $K = 5$)	$f_{4971-5020}$	0.006	4.577	5.811	6.035	5.714	5.933
	$f_{9471-9520}$	0.007	3.282	5.287	5.416	5.127	5.251
	$f_{4971-5020}$	0.007	3.406	5.267	5.451	5.104	5.266
	$f_{9471-9520}$	0.008	2.208	5.171	5.268	4.928	5.034
	$f_{4971-5020}$	0.008	2.153	4.450	4.594	4.319	4.471
Net25	$f_{9471-9520}$	0.006	4.495	5.319	6.004	5.190	5.834
($n = 9520$, $K = 25$)	$f_{4971-5020}$	0.006	4.574	5.849	4.752	5.714	4.579
	$f_{9471-9520}$	0.007	3.307	5.264	5.765	4.951	5.447
	$f_{4971-5020}$	0.007	3.256	5.132	4.182	5.014	4.102
	$f_{9471-9520}$	0.008	2.111	4.913	5.249	4.756	5.128
	$f_{4971-5020}$	0.008	2.150	4.345	3.670	4.251	3.554

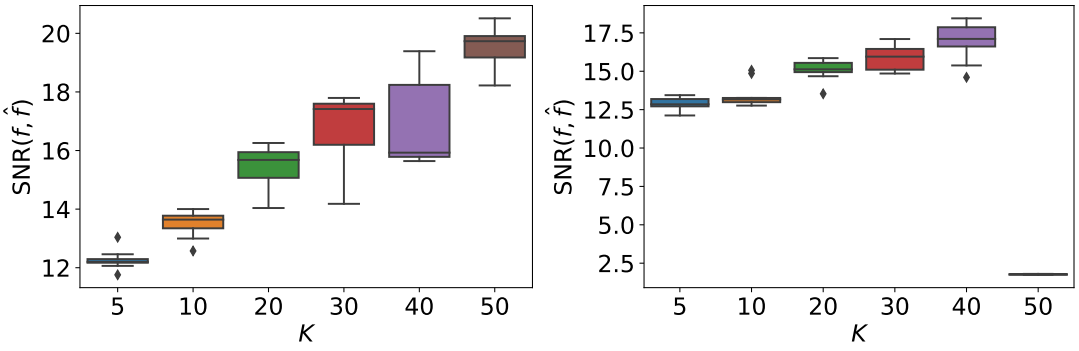


Figure 5: SNR performance depending on parameter K for *minnesota* graph, $\sigma = 0.01$, a regular signal $f_{2593-2642}$ (left) and a non-regular signal $f_{1343-1392}$ (right), over 10 realizations of noise.

generated by Chebyshev's approximation grows with the number of intervals in the partition, which makes the approximation of the support more difficult.

6 Conclusion and future works

We have introduced a novel technique to perform efficiently graph Fourier analysis. This technique uses functional calculus to perform Fourier analysis on different subsets of the graph Laplacian spectrum. In this paper, we have demonstrated that localization in the spectrum provides interesting improvements in theoretical results for some graph signal analysis tasks. New estimators of the noise level were introduced taking advantage of the convenient modelling of the denoising problem given by localization, and for which concentration results were proved. Localization allows also to study theoretically the denoising procedure with wavelets, and fits with the design of many well-known techniques (*e.g.* tight frames for graph analysis). Through many experiments, we have validated that localization techniques introduced in this paper improve on state-of-the-art methods for several standard tasks.

Although we provide a rule of thumb to choose a partition $I_{\mathcal{L}} = \sqcup_{1\leq k\leq K} I_k$ for which denoising results show good performances, experiments suggest that our *elbow rule* is not optimal in most cases. There is certainly an interesting topic in searching for a suitable partition $I_{\mathcal{L}} = \sqcup_{1\leq k\leq K} I_k$ that would be more adapted to a specific task (*e.g.* denoising). To extend the current work, it would also be interesting to consider other common tasks in GSP, such as de-convolution or in-painting.

7 Proofs

Proof of Proposition 3. 1. We have $\tilde{f}_k = f_k + \xi_k$, where $\xi_k = \sum_{\ell:\lambda_\ell\in I_k} \widehat{\xi}(\ell)\chi_\ell$. Random variables $(\widehat{\xi}(\ell))_\ell$ are all distributed as $\mathcal{N}(0,\sigma^2)$ and independent by orthogonality of the eigenbasis $(\chi_\ell)_\ell$. In particular for $k \neq k'$, vectors ξ_k and $\xi_{k'}$ are independent as expressions involving variables $\widehat{\xi}(\ell)$ over disjoint subsets I_k and $I_{k'}$. Thus the random variables $(c_k)_{1\leq k\leq K}$ are also independent.

2. When $n_k = n_{k'}$, ξ_k and $\xi_{k'}$ are identically distributed and the result follows from Equality (3).
When $n_k \neq n_{k'}$, we have $\mathbb{E}(c_k) \neq \mathbb{E}(c_{k'})$ as the following equality holds for all $1 \leq k \leq K$:

$$\mathbb{E}(c_k) = \frac{\|f_k\|_2^2}{n_k} + \sigma^2.$$

3. Since $(\widehat{\xi}(\ell))_\ell$ are independent normal variables $\mathcal{N}(0,\sigma^2)$, the statement is clear from the expression $c_k = \frac{1}{n_k}\sum_{\ell:\lambda_\ell\in I_k} |\widehat{\xi}(\ell)|^2$.

□

The following lemma is useful for the proof of Proposition 4.

Lemma 13. *Let $Z \sim \mathcal{B}(n,p)$ for some parameters $n \geq 1$ and $p \leq 1/2$. Then*

$$\mathbf{P}(Z \geq \lceil n/2 \rceil) \leq \exp\left(\frac{n}{2}\ln(4p(1-p))\right).$$

Proof. A simple consequence of [4, Theorem 1] implies that for all $n \geq 1$ and all $a \geq p$

$$\mathbf{P}(Z \geq na) \leq \left[\frac{(1-p)^{1-a}}{1-a}\left(\frac{1-a}{a}p\right)^a\right]^n.$$

Now the result follows since $na = \lceil n/2 \rceil$ implies $\frac{1}{2} \leq a \leq \frac{1}{2} + \frac{1}{n}$ so that

$$\frac{(1-p)^{1-a}}{1-a}\left(\frac{1-a}{a}p\right)^a \leq \frac{(1-a)^{a-1}}{a^a}\sqrt{p(1-p)} \leq \sqrt{4p(1-p)}.$$

□

Proof of Proposition 4. 1. For k such that $I_k \subset \overline{I_f}$, we have $c_k = \frac{\sigma^2}{n_k}\Gamma_{n_k}$, which follows the $\Gamma(\frac{n_k}{2}, \frac{2\sigma^2}{n_k})$ distribution. Then concentration inequalities for $\widehat{\sigma}_{\text{mean}}(c)^2$ are a direct consequence of Theorem 2.57 in [3], applied with $a_k = \frac{n_k}{2}$ and $b_k = \frac{2\sigma^2}{n_k}$.

2. For all $k = 1, \ldots, K_f$, we define

$$\gamma_k^- = \Phi^{-1}_{\Gamma_{n_0}} \circ \Phi_{\Gamma_{n_k}}\left(\frac{n_k}{\sigma^2}c_k\right) \quad \text{and} \quad \gamma_k^+ = \Phi^{-1}_{\Gamma_{n_\infty}} \circ \Phi_{\Gamma_{n_k}}\left(\frac{n_k}{\sigma^2}c_k\right).$$

As a matter of fact, $(\gamma^-)_{k=1,\ldots,K_f}$ and $(\gamma_k^+)_{k=1,\ldots,K_f}$ are two sequences of *i.i.d.* random variables with $\gamma_1^- \sim \chi^2(n_0)$ and $\gamma_1^+ \sim \chi^2(n_\infty)$ such that

$$\forall k = 1, \ldots, K_f, \quad \gamma_k^- \leq \frac{n_k}{\sigma^2}c_k \leq \gamma_k^+ \quad \text{almost surely.}$$

Then, for all $t > 0$,

$$\begin{aligned}\mathbb{P}\left(\widehat{\sigma}^2_{\text{med}} \geq \beta^{-1}\sigma^2 + 2\sigma^2\beta^{-1}t\right) &= \mathbb{P}\left(\sum_{k=1}^{K_f} 1_{\{c_k \geq \beta^{-1}\sigma^2 + 2\sigma^2\beta^{-1}t\}} \geq \left\lceil \frac{K_f}{2} \right\rceil\right) \\ &\leq \mathbb{P}\left(\sum_{k=1}^{K_f} 1_{\left\{\gamma_k^+ \geq n_\infty + 2n_\infty t\right\}} \geq \left\lceil \frac{K_f}{2} \right\rceil\right). \qquad (11)\end{aligned}$$

Similarly, for all $t \in (0,1)$,

$$\mathbb{P}\left(\widehat{\sigma}^2_{\text{med}} \leq \beta\sigma^2 - \sigma^2\beta t\right) \leq \mathbb{P}\left(\sum_{k=1}^{K_f} 1_{\left\{\gamma_k^- \leq n_0 - n_0 t\right\}} \geq \left\lceil \frac{K_f}{2} \right\rceil\right). \qquad (12)$$

To conclude, apply Lemma 13 to Inequalities (11) and (12) to obtain our result.

□

Proof of Proposition 5. The concentration bound of Equation (5) implies that

$$\begin{aligned}\mathbb{P}(b_k \geq b_\ell) = \mathbb{P}(n_k b_k \geq n_k b_\ell) &= \mathbb{P}\left(\sigma^{-2}\left(\frac{n_k}{n_\ell}\|\xi_\ell\|_2^2 - \|\xi_k\|_2^2\right) \leq \sigma^{-2}\left(\|f_k\|_2^2 + 2\langle f_k, \xi_k\rangle\right)\right) \\ &\geq \frac{\alpha}{2} + \mathbb{P}\left(\sigma^{-2}\left(\frac{n_k}{n_\ell}\|\xi_\ell\|_2^2 - \|\xi_k\|_2^2\right) \leq \sigma^{-2}\left(\|f_k\|_2^2 + 2t_{\alpha,\sigma}\|f_k\|_2\right)\right).\end{aligned}$$

Since subsets I_k and I_ℓ are disjoint, random variables $\|\xi_k\|_2^2$ and $\|\xi_\ell\|_2^2$ are independent. Thus, $\frac{n_k}{n_\ell}\|\xi_\ell\|_2^2 - \|\xi_k\|_2^2$ is distributed as $\frac{n_k}{n_\ell}\Gamma_{n_\ell} - \Gamma_{n_k}$ where Γ_{n_k} and Γ_{n_ℓ} are independent random variables with $\Gamma_{n_k} \sim \chi^2(n_k)$ and $\Gamma_{n_\ell} \sim \chi^2(n_\ell)$. Therefore, the statement of Proposition 5 follows. □

Proof of Proposition 6. First, the following equalities hold:

$$f - \widetilde{f} = f - \widetilde{f}_I + \widetilde{f}_I - \widetilde{f} = (f - \widetilde{f})_I + \widetilde{f}_{\overline{I}}.$$

As $(f - \widetilde{f})_I$ and $\widetilde{f}_{\overline{I}}$ are orthogonal vectors, it follows that

$$\left\|f - \widetilde{f}\right\|_2^2 = \left\|f - \widetilde{f}_I\right\|_2^2 + \left\|\widetilde{f}_{\overline{I}}\right\|_2^2.$$

It remains to notice that $\mathbb{E}(||\widetilde{f}_{\overline{I}}||^2) = \sigma^2|\overline{I} \cap \mathrm{sp}(\mathcal{L})|$. □

Proof of Lemma 7. By Equation (3) and the concentration bound of Equation (5), it follows that

$$\begin{aligned}p_k &= \mathbb{P}(\sigma^2\Gamma_{n_k} > \|\xi_k\|_2^2 + \|f_k\|_2^2 + 2\langle f_k, \xi_k\rangle) \\ &\leq \frac{\alpha}{2} + \mathbb{P}\left(\sigma^2\Gamma_{n_k} > \|\xi_k\|_2^2 + \|f_k\|_2^2 - 2\|f_k\|_2 t_{\alpha/2,\sigma}\right) \\ &= \frac{\alpha}{2} + \mathbb{P}\left(\sigma^2(\Gamma_{n_k} - \Gamma'_{n_k}) > \|f_k\|_2^2 - 2\|f_k\|_2 t_{\alpha/2,\sigma}\right) \\ &= \frac{\alpha}{2} + 1 - \Phi_{\Gamma_{n_k} - \Gamma'_{n_k}}(\theta(f_k, \alpha, \sigma)),\end{aligned}$$

where $\theta(f_k, \alpha, \sigma) = \sigma^{-2}(\|f_k\|_2 - 2t_{\alpha/2,\sigma})\|f_k\|_2$. Consequently, $1 - \Phi_{\Gamma_{n_k} - \Gamma'_{n_k}}(\theta(f_k, \alpha, \sigma)) \leq \alpha/2$ and $p_k \leq \alpha$. □

Proof of Lemma 8. Using an estimate on the $\chi^2(n_k)$ tail distribution and independence of Γ_{n_k} and $\Gamma'_{n_k} = \sigma^{-2}\|\xi_k\|_2^2$, it follows

$$\begin{aligned}p_k &\geq \mathbb{P}\left(\sigma^2\Gamma_{n_k} > \|\xi_k\|_2^2 + \|f_k\|_2^2 + 2\langle \xi_k, f_k\rangle, \|\xi_k\|_2^2 \leq t_{\beta,k}\right) \\ &\geq \mathbb{P}\left(\sigma^2\Gamma_{n_k} > \|\xi_k\|_2^2 + \|f_k\|_2^2 + 2\|\xi_k\|_2\|f_k\|_2, \|\xi_k\|_2^2 \leq t_{\beta,k}\right) \\ &= \mathbb{P}\left(\sigma^2\Gamma_{n_k} > (\|f_k\|_2 + \sqrt{t_{\beta,k}})^2, \|\xi_k\|_2^2 \leq t_{\beta,k}\right) \\ &= \mathbb{P}\left(\sigma^2\Gamma_{n_k} > (\|f_k\|_2 + \sqrt{t_{\beta,k}})^2, \sigma^2\Gamma'_{n_k} \leq t_{\beta,k}\right) \\ &\geq \mathbb{P}\left(\sigma^2\Gamma_{n_k} > (\|f_k\|_2 + \sqrt{t_{\beta,k}})^2\right)(1 - \beta) \\ &\geq \frac{\alpha}{1-\beta} \times (1 - \beta) = \alpha.\end{aligned}$$

□

Proof of Proposition 9. 1. To prove Inequality (6), first observe that $f = \sum_{k:I_k \subset I_f} f_k$ so that

$$f - f_I = \sum_{k:I_k \subset I_f} f_k - \sum_{k:p_k \leq \alpha} f_k.$$

The summands which are not present in both terms are exactly those satisfying either $I_k \subset I_f$ and $p_k > \alpha$ or $I_k \cap I_f = \emptyset$ and $p_k \leq \alpha$. Noting that $f_k = 0$ when $I_k \cap I_f = \emptyset$, it comes

$$\|f - f_I\|_2^2 = \sum_{k: I_k \subset I_f, p_k > \alpha} \|f_k\|_2^2.$$

Applying Lemma 7 for all indices $1 \leq k \leq K$ satisfying $p_k > \alpha$, one deduce

$$\|f_k\|_2 < t_{\alpha/2,\sigma} + \sqrt{t_{\alpha/2,\sigma}^2 + \left(\sigma \Phi^{-1}_{\Gamma_{n_k} - \Gamma'_{n_k}}(1 - \frac{\alpha}{2})\right)^2}.$$

from which, since $n_k = n_1$ for all k, Inequality (6) follows.

2. Since $\sigma^{-2}||\xi_k||_2^2$ is distributed as a $\chi^2(n_1)$ random variable, the second Inequality (7) follows

$$\mathbb{E}\|f_I - \widetilde{f}_I\|_2^2 = \sum_{k: p_k \leq \alpha} \mathbb{E}\|\xi_k\|_2^2 = |\{k, p_k \leq \alpha\}| n_1 \sigma^2.$$

□

Proof of Theorem 10. Since threshold value is $t_2 = 0$ on $\overline{I}$, $\widehat{f} = \widehat{f}_I$. Then, clearly $f_{\overline{I}}(f_I - \widehat{f}_I) = 0$ almost surely so that

$$\mathbb{E}\|f - \widehat{f}\|_2^2 = \mathbb{E}\|f - \widehat{f}_I\|_2^2 = \mathbb{E}\|f - f_I + f_I - \widehat{f}_I\|_2^2 = \mathbb{E}\|f - f_I\|_2^2 + \|f_I - \widehat{f}_I\|_2^2.$$

Applying Theorem 3 from [15] to $\mathbb{E}\|f_I - \widehat{f}_I\|_2^2$ yields our statement. □

Proof of Proposition 11. Recalling that, for any function g defined on $\mathrm{sp}(\mathcal{L})$ and any subset $I \subset I_{\mathcal{L}}$,

$$\sum_n |\langle \sqrt{g}(\mathcal{L}_I)\delta_n, f\rangle|^2 = \|\sqrt{g}(\mathcal{L}_I) f\|_2^2 = \langle g(\mathcal{L}_I) f, f\rangle$$

it follows by Equations (8) and (9).

$$\begin{aligned}\sum_{n,k} |\langle \varphi_{n,k}, f\rangle|^2 + |\langle \psi_{1,n,k}, f\rangle|^2 + |\langle \psi_{2,n,k}, f\rangle|^2 \\ = \sum_k \langle \psi_0(\mathcal{L}_{I_k}) f, f\rangle + \langle \psi_k(\mathcal{L}_{I_k}) f, f\rangle + \langle \psi_k(\mathcal{L}_{I_{k+1}}) f, f\rangle \\ = \langle \psi_0(\mathcal{L}) f, f\rangle + \sum_k \langle \psi_k(\mathcal{L}) f, f\rangle = \|f\|_2^2\end{aligned}$$

□

Proof of Proposition 12. Remarking that $\widehat{g}_j(\lambda) = \psi(s_j \omega_0(\lambda))$ with $s_j = e^{-j\omega_{\gamma,J,R}}$, Equation (10) implies that

$$g_{m,j} = \sum_l \psi(s_j \omega_0(\lambda_l)) \widehat{\delta}_m(l) \chi_l. \tag{13}$$

Setting $J_j = [C^{-1} e^{(j-R)\omega_{\gamma,J,R}}, C^{-1} e^{j\omega_{\gamma,J,R}}]$ and recalling that $\mathrm{supp}(\widehat{g}) = [-R\omega_{\gamma,J,R}, 0]$, it follows that $\lambda \in \mathrm{supp}(\widehat{g}_j)$ if and only if $\omega_0(\lambda) \in J_j$ if and only if $s_j \omega_0(\lambda) \in J_0$. Moreover, $J_0 = \sqcup_k I_k$ and $J_j = s_j^{-1} J_0$ yield $J_j = \sqcup_k s_j^{-1} I_k$ with $s_j^{-1} I_k \cap s_{j'}^{-1} I_{k'}$ excepted when $j = j'$ ad $k = k'$. Consequently, Equation (13) can be reformulated as

$$g_{m,j} = \sum_{1 \leq k \leq R-1} \sum_{\ell \in I_k} \psi(s_j \omega_0(\lambda_\ell)) \widehat{\delta}_m(\ell) \chi_\ell.$$

□

References

[1] Hamid Behjat, Ulrike Richter, Dimitri Van De Ville, and Leif Sörnmo. Signal-adapted tight frames on graphs. *IEEE Trans. Signal Process.*, 64(22):6017–6029, 2016.

[2] Mikhail Belkin and Partha Niyogi. Towards a theoretical foundation for laplacian-based manifold methods. *J. Comput. Syst. Sci.*, 74(8):1289–1308, 2008.

[3] Bernard Bercu, Bernard Delyon, and Emmanuel Rio. *Concentration inequalities for sums and martingales*. SpringerBriefs in Mathematics. Springer, Cham, 2015.

[4] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statistics*, 23:493–507, 1952.

[5] Fan RK Chung and Fan Chung Graham. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997.

[6] Ronald R Coifman and Mauro Maggioni. Diffusion wavelets. *Appl. Comput. Harmon. Anal.*, 21(1):53–94, 2006.

[7] Mark Crovella and Eric Kolaczyk. Graph wavelets for spatial traffic analysis. In *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428)*, volume 3, pages 1848–1857. IEEE, 2003.

[8] Timothy A. Davis and Yifan Hu. The University of Florida sparse matrix collection. *ACM Trans. Math. Software*, 38(1):Art. 1, 25, 2011.

[9] Basile de Loynes, Fabien Navarro, and Baptiste Olivier. Data-driven thresholding in denoising with spectral graph wavelet transform. *arXiv preprint arXiv:1906.01882*, 2019.

[10] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in neural information processing systems*, pages 3844–3852, 2016.

[11] Edoardo Di Napoli, Eric Polizzi, and Yousef Saad. Efficient estimation of eigenvalue counts in an interval. *Numer. Linear Algebra Appl.*, 23(4):674–692, 2016.

[12] Li Fan, David I Shuman, Shashanka Ubaru, and Yousef Saad. Spectrum-adapted polynomial approximation for matrix functions. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4933–4937. IEEE, 2019.

[13] Matan Gavish, Boaz Nadler, and Ronald 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 International Conference on Machine Learning*, pages 367–374, 2010.

[14] Benjamin Girault, Antonio Ortega, and Shrikanth S Narayanan. Irregularity-aware graph fourier transforms. *IEEE Trans. Signal Process.*, 66(21):5746–5761, 2018.

[15] Franziska Göbel, Gilles Blanchard, and Ulrike von Luxburg. Construction of tight frames on graphs and application to denoising. In *Handbook of Big Data Analytics*, pages 503–522. Springer, 2018.

[16] David K Hammond, Pierre Vandergheynst, and Rémi Gribonval. Wavelets on graphs via spectral graph theory. *Appl. Comput. Harmon. Anal.*, 30(2):129–150, 2011.

[17] M. F. Hutchinson. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Comm. Statist. Simulation Comput.*, 19(2):433–450, 1990.

[18] Nora Leonardi and Dimitri Van De Ville. Tight wavelet frames on multislice graphs. *IEEE Trans. Signal Process.*, 61(13):3357–3367, 2013.

[19] Lin Lin, Yousef Saad, and Chao Yang. Approximating spectral densities of large matrices. *SIAM review*, 58(1):34–65, 2016.

[20] Sridhar Mahadevan. Fast spectral learning using lanczos eigenspace projections. In *Proc. AAAI Conference on Artificial Intelligence, 2008*, pages 1472–1475, 2008.

[21] Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura, and Pierre Vandergheynst. Graph signal processing: Overview, challenges, and applications. *Proceedings of the IEEE*, 106(5):808–828, 2018.

[22] Nathanaël Perraudin and Pierre Vandergheynst. Stationary signal processing on graphs. *IEEE Trans. Signal Process.*, 65(13):3462–3477, 2017.

[23] Aliaksei Sandryhaila and Jose MF Moura. Discrete signal processing on graphs: Frequency analysis. *IEEE Trans. Signal Process.*, 62(12):3042–3054, 2014.

[24] Stefania Sardellitti, Sergio Barbarossa, and Paolo Di Lorenzo. On the graph fourier transform for directed graphs. *IEEE J. Sel. Top. Signal Process.*, 11(6):796–811, 2017.

[25] Santiago Segarra, Antonio G Marques, Geert Leus, and Alejandro Ribeiro. Aggregation sampling of graph signals in the presence of noise. In *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 101–104. IEEE, 2015.

[26] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and 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):83–98, 2013.

[27] David I Shuman, Benjamin Ricaud, and Pierre Vandergheynst. Vertex-frequency analysis on graphs. *Appl. Comput. Harmon. Anal.*, 40(2):260–291, 2016.

[28] David I Shuman, Pierre Vandergheynst, and Pascal Frossard. Chebyshev polynomial approximation for distributed signal processing. In *2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS)*, pages 1–8. IEEE, 2011.

[29] RN Silver and H Röder. Densities of states of mega-dimensional hamiltonian matrices. *Int. J. Mod. Phys. C*, 5(04):735–753, 1994.

[30] Ana Susnjara, Nathanael Perraudin, Daniel Kressner, and Pierre Vandergheynst. Accelerated filtering on graphs using lanczos method. *arXiv preprint arXiv:1509.04537*, 2015.

[31] Yuichi Tanaka and Akie Sakiyama. m-channel oversampled graph filter banks. *IEEE Trans. Signal Process.*, 62(14):3578–3590, 2014.

[32] Nicolas Tremblay. *Networks and signal : signal processing tools for network analysis*. Theses, Ecole normale supérieure de lyon - ENS LYON, October 2014.

[33] John von Neumann. Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statistics*, 12:367–395, 1941.

[34] Lin-Wang Wang. Calculating the density of states and optical-absorption spectra of large quantum systems by the plane-wave moments method. *Phys. Rev. B*, 49(15):10154, 1994.