On localisation and uncertainty measures on graphs

Lausanne - Spring 2012

Perraudin Nathanaël
Master's thesis, supervised by **Vandergheynst Pierre** and **Shuman David**

LTS2 - EPFL

"Maturity of mind is the capacity to endure uncertainty."

John Finley

Contents

Abstract

Due to the appearance of data on networks such as internet or Facebook, the number of applications of signal on weighted graph is increasing. Unfortunately, because of the irregular structure of this data, classical signal processing techniques are not applicable. In this paper, we examine the windowed graph Fourier transform (WGFT) and propose ambiguity functions A_g to analyze the spread of the window in the vertex-frequency plane. We then observe through examples that there is a trade-off between the vertex and frequency resolution. This matches our intuition form classical signal processing. Finally, we demonstrate an uncertainty principle for the spread of the ambiguity function: $\frac{\|A_g\|_1}{\|A_g\|_\infty} \geqslant \frac{1}{\mu^2}$. We verify with examples that this principle is sharp for the extreme values of μ and emphasize the difference between the generalized graph ambiguity function and the classical one. We finish with demonstration of some Young and Hausdorff-Young like inequalities for graphs.

1 Introduction

Data, like images, sound or text, usually come with structure. This allows us to process the information even if it is incomplete or there is background noise. For instance, we can understand a text even if there are some spelling mistakes. Moreover, we are able to correct them with the help of the structure of the sentences present in the text. Signal processing attempts to do the same within special data thanks to the help of computers or electronics. This has inspired a large number of solutions for filtering, compressing, and recovering signals. To process information, we usually represent it on regular Euclidean spaces or manifolds. However, many applications involve more topologically-complicated structures like networks. The number of such applications is growing everyday on account of the development of the internet and, particularly, social networks.

Many applications in signal processing are based on transformation methods. The idea is to observe the signal in a domain that provides the desired information. One famous transform is the windowed Fourier transform, also called short time Fourier transform or Gabor transform when used with a Gaussian window. It helps us to extract information in time/space and frequency simultaneously and is obtained by the scalar product of the signal with atoms generated by shifting and modulating a window. Therefore the choice of the window will greatly affect the result. For example, the transform is a tool to examine music. In a melody, a sound is produced to be at a precise moment with a specific spectral content. In this case, the transform provides a sort of "mathematical score." However, there is no magic behind this process and the cost of this transform is a trade-off between time and frequency resolution. This is explained by uncertainty principles like the well known Heisenberg uncertainty principle. In order to measure the uncertainty, we use an ambiguity function constructed by the scalar product of two atoms. This will measure the difference between them in time and frequency. By doing so, the decay of the ambiguity function becomes an indicator of the precision of the window. Lieb [1] has shown that the L^p-norm of the ambiguity function is bounded using the Young's inequality and the Hausdorff-Young inequality. This means that it will always spread, whatever the chosen window. The purpose of this research is to define and study an analogous function for graphs.

A graph is a structure made of nodes called vertices and links named edges. A weight corresponding to the distance between the vertices is associated with each edge. It is possible to construct a Laplacian operator $\mathcal{L}$ for graphs from these weights only. In the case of a connected undirected graph, the Laplacian's eigenvectors form a complete orthogonal basis which is used as a Fourier basis. As in the classical theory, the Fourier transform of a graph signal gives its frequency content. We would like to develop a tool that allows us to simultaneously analyze the signal in the spectral and vertex domains.

Studying signal processing on graphs is a burgeoning field. Recently, D. Hammond and P. Vandergheynst adapted a wavelet transform to the graph setting [2]. A windowed graph Fourier transform (WGFT) was also defined by D. Shuman et al [3], whereby, the convolution, shift, and modulation for signals on graphs are defined in order to generate the atoms used in the transform. They also provide examples that fit the intuition and promise the WGFT a good future. Even more so, as all these definitions need only the weights that can be seen as the distance between the edges, the obtained tool can be useful for analyzing a large range of data. Indeed, weighted undirected graphs provide an extremely flexible model for approximating data domains.

These definitions are all quite new and require further investigation. The task is difficult, because the shift and modulation behave quite strangely and do not preserve energy. As a consequence, the WGFT does not conserve energy. Even if certain frequencies are present at some vertices, the coefficients can be insignificant. As there are no uncertainty principles for graphs, we do not know the limit of the WGFT ability to extraction information. In addition, the choice of the window remains a problem because there are no tools to study its spread in the spectral and vertex domains.

This paper first focuses on the study of the WGFT and proposes a new definition, which, we hope, will conserve energy. Secondly, and this is the main interest of the paper, it presents some ambiguity functions for graphs which enable us to compute the resolution of a window in time and frequency. Thirdly, studying the ambiguity function will lead us to some uncertainty principles. Finally, the convolution product for graphs is studied. We demonstrated some Young and Hausdorff-Young like inequalities which allow a better comprehension of the shift operator.

The paper is organized as follows. In a first section, we give a classical overview of the windowed Fourier transform and the ambiguity function. We then define the convolution, shift, and modulation of the WGFT for graphs and propose a normalized windowed graph Fourier transform (NWGFT). Later, we define the ambiguity function and demonstrates some bounds on its norm, in order to reach an uncertainty principle. In a second section (Chapter 5 and 6), we present classical inequalities as follow: Hausdorff-Young, Young and Lieb. We then demonstrate some analogous bounds on the norm of the convolution product (Young) and on the Fourier Transform (Hausdorff-Young) in the graph setting.

2 Classical Overview

We shall start by giving an overview of some of the classical theory for continuous and discrete functions including the windowed Fourier transform (short time Fourier transform or Gabor when used with a Gaussian window), which gives localization in space and in frequency simultaneously, and the ambiguity function, which measures the spread of a given window in space and in frequency.

The following definitions for the Fourier transform $\hat{f}$ of a function $f \in L^2$ will be employed:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i \xi x} \, \mathrm{d}x,$$

so that the inverse transform is

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi) e^{-2\pi i \xi x} \, \mathrm{d}\xi$$

and Parseval's relation becomes

$$\langle f, g \rangle = \langle \hat{f}, \hat{g} \rangle. \tag{1}$$

The direct consequence of (1) is the conservation of energy between a function f and its Fourier transform $\hat{f}$:

$$\|f\|_2 = \|\hat{f}\|_2.$$

As functions on graphs are discrete, we also add the definition used for the Discrete Fourier Transform (DFT) of a signal $f \in \mathbb{R}^N$:

$$\hat{f}[\ell] = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} f[n] e^{-2\pi i \frac{\ell}{N} n},$$

and the inverse transform is

$$f[n] = \frac{1}{\sqrt{N}} \sum_{\ell=0}^{N-1} f[\ell] e^{2\pi i \frac{n}{N} \ell}.$$

Parseval's relation remains unchanged.

2.1 Classical windowed Fourier transform

The classical windowed Fourier transform (CWFT) simultaneously provides information in time/space and frequency. Suppose we have a sound and we would like to extract the frequency of it. One may simply compute the DFT and list the frequency. Then, the spectrum of the sound would be known exactly at the cost of the loss of all temporal information. The CWFT does not work so drastically. Using it, we are able to extract the frequency present at a certain time in the signal. The price to pay is an uncertainty, which can be explained by the Heisenberg uncertainty principle. In order to build the CWFT, we need a window g that is confined in both domains. Furthermore, two shift operators are required: one in time or space and the other in frequency. The shift in frequency is called modulation.

For any $f \in L^2$ and $u \in \mathbb{R}$, the translation operator T_u: $L^2 \to L^2$ is defined by:

$$(T_u f)(t) := f(t - u)$$

and for any $\xi \in \mathbb{R}$, the modulation operator M_ξ: $L^2 \to L^2$ is defined by:

$$(M_\xi f)(t) = f(t) e^{i \xi t}.$$

A windowed Fourier atom $g_{u,\xi}$ is constructed by translating a window $g(t)$ of unit norm $\|g\|_2 = 1$ centered at $t = 0$ in both space and frequency.

$$g_{u,\xi} := (M_\xi T_u g)(t) = g(t - u) e^{i \xi t}. \tag{2}$$

The shift will move the window to a certain point without modifying the frequency content. Only the phase will be affected. The modulation will shift the frequency content of the window to a certain point without changing the localization in space or time.

Now, the CWFT projects a function $f \in L^2$ on every space-frequency atom.

$$Sf(u,\xi) := \langle f, g_{u,\xi} \rangle = \int_{-\infty}^{\infty} f(t)\overline{g(t-u)}e^{-i\xi t}\,\mathrm{d}t \tag{3}$$

It can be interpreted as a Fourier transform of f at the frequency ξ, localized by the window $g(t-u)$ in the neighborhood of u. $Sf(u,\xi)$ is big if f contains approximately frequency ξ around u and vice-versa.

2.2 Ambiguity function

Ideally, we would like to be precise in both domains at the same time. This can unfortunately not be achieved. Nevertheless, the ambiguity function is a powerful tool to evaluate the uncertainty of the CWFT. The overlap of the atoms contains the information for which we are looking.

In order to measure the space-frequency overlap of two atoms $g_{u,\xi}$ and g_{u_0,ξ_0} we use a kernel $K(u_0,u,\xi_0,\xi) = \langle g_{u,\xi}, g_{u_0,\xi_0} \rangle$ that decays with $u_0 - u$ and $\xi_0 - \xi$ at a rate that depends on the energy concentration of g and its Fourier transform $\hat{g}$. Then with the change of variable $v = t - \frac{u+u0}{2}$ the scalar product between $g_{u,\xi}$ and g_{u_0,ξ_0} the kernel becomes:

$$K(u_0,u,\xi_0,\xi) = \langle g_{u,\xi}, g_{u_0,\xi_0} \rangle = exp\left(-\frac{i}{2}(\xi_0-\xi)(u+u_0)\right) A_g(u_0-u,\xi_0-\xi),$$

where

$$A_g(\tau,\gamma) = \int_{-\infty}^{\infty} g\left(t+\frac{\tau}{2}\right) g\left(t-\frac{\tau}{2}\right) e^{-i\gamma t}\,\mathrm{d}t$$

is called the ambiguity function of g. The decay of the ambiguity function measures the spread in space and in frequency of the window g. This decay in space/time and in frequency respectively characterizes the uncertainty of both domains of the window. We observe that the more precise the transform is in frequency, the less precise it is in the other domain and vice-versa. The optimum precision is achieved by a Gaussian window.

In fact, the ambiguity function is a particular case of the cross-ambiguity function:

$$A_{f,g}(\tau,\gamma) = \int_{-\infty}^{\infty} f\left(t+\frac{\tau}{2}\right) g\left(t-\frac{\tau}{2}\right) e^{-i\gamma t}\,\mathrm{d}t.$$

For the discrete case, we define the cross ambiguity function by:

$$A_g f(m,k) = \langle f, M_k T_m g \rangle = \sum_{n=1}^{N} f[n]\bar{g}[n-m]e^{-2\pi i k\frac{n}{N}} \tag{4}$$

It measures the space frequency overlap of f and a modulated shifted window g. Note that this definition of the cross ambiguity function is equivalent to a windowed Fourier transform!

3 Spectral Graph Definitions

In this section, we present the main definitions used in this work including the generalization of the modulation and the translation for graphs. This will lead us to a windowed graph Fourier transform(WGFT).

3.1 Spectral graph theory notation

We consider an undirected, connected, and weighted graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, W\}$, where $\mathcal{V}$ is a finite set of vertices (with $|\mathcal{V}| = N$), $\mathcal{E}$ is a finite set of edges, and W the weighted adjacency matrix [4]. We define a signal $f : \mathcal{V} \to \mathbb{R}^N$ as a function assigning one value to each vertex. It can be seen

as a vector of size N with the n^{th} component representing the signal value at the n^{th} vertex. The non-normalized graph Laplacian is defined by $\mathcal{L} = D - W$ where D is the diagonal degree matrix ($D_{ii} = \sum_{j=1}^{N} W_{ij}$).

As the graph is undirected, the adjacency matrix W and the Laplacian $\mathcal{L}$ are symmetric matrices. Since the weights are all real, the graph Laplacian possesses a complete set of real and orthonormal eigenvectors which will be denoted by $\{\chi_\ell\}_{\ell=0,1,...,N-1}$. The eigenvalues of $\mathcal{L}$ are real and positive. We order them as follows: $0 = \lambda_0 < \lambda_1 \leqslant \lambda_2 \leqslant ... \leqslant \lambda_{N-1} := \lambda_{max}$. In order to simplify the writing, we denote the square matrix composed of the N column eigenvectors: $\Phi = \{\chi_0, \chi_1 \cdots \chi_{N-1}\}$.

A Fourier transform of a graph signal f is defined by

$$\hat{f}(\ell) = \langle f, \chi_\ell \rangle = \sum_{n=1}^{N} \chi_\ell^*(n) f(n) = (\Phi^t f)(\ell), \tag{5}$$

where we use x^* to denote the transpose complex conjugate of the vector x. We write $\overline{f}$ the complex conjugate of the signal f. An inverse Fourier transform is given by

$$f(n) = \sum_{\ell=0}^{N-1} \hat{f}(\ell) \chi_\ell(n) = (\Phi \hat{f})(n).$$

One very interesting property is:

$$\mathcal{F}\{\mathcal{L}^k f\}(\ell) = \sum_{\ell=0}^{N-1} \lambda_\ell^k \hat{f}(\ell). \tag{6}$$

3.2 Relation between graph and classical theory

The definition (5) is analogous to the classical Fourier transform: $\hat{f}(w) = \langle f, e^{iwx} \rangle$. In fact, it has been shown that the eigenvectors of the Laplacian are discrete cosines for special edge configurations. Strang [5] has demonstrated the following relations.

The Discrete Fourier Transform (DFT) eigenvectors are the eigenfunctions for the Laplacian or a ring graph. In addition, the eigenvectors of the path graph can be chosen to be the same as column of the DCT-2 transform. Table 1 presents the eigenvectors and the eigenvalues for the DFT, real DFT (RDFT) and the DCT-2. In this paper, DFT will be used for the case of a ring graph with classical eigenvectors and RDFT for the same graph, but with real eigenvectors. When DCT-2 is used, we suppose a path graph with real eigenvectors.

Remark: In the classical case, presented by Strang [5], the order of the eigenvalues is not crescent: $\lambda_\ell = 2 - 2\cos(\frac{2\pi\ell}{N})$, for $\ell \in \{0, 1, ..., N-1\}$. The consequence of this order will be presented later.

For a connected graph, there is only one complete orthonormal base of real eigenvector. This solution is often used because algorithms that compute the eigenvectors usually return real eigenvector. As a consequence, it is relevant to also observe the real DFT (RDFT). This is characterized by another choice of eigenvectors for the ring graph. Observing differences between the DFT and the RDFT will explain some differences between the classical theory and its generalization.

3.3 A generalized convolution product for signals on graphs

For two signals $f, g \in L^2$, the classical convolution is defined as $h(t) = (f * g)(t) = \int_{\mathbb{R}} f(x) g(t - x) dx$. One nice property of the convolution is its correspondence with a product in the Fourier domain:

$$(f * g)(t) = \mathcal{F}^{-1}\left\{ \int_{\mathbb{R}} \hat{f}(w) \hat{g}(w) dw \right\}(t) = \int_{\mathbb{R}} \hat{f}(w) \hat{g}(w) \psi_w(t) dw,$$

with $\psi_w(k) = e^{2\pi i k t}$. The definition of the convolution product for signals on a graph is based on that property. This choice is not arbitrary. Indeed, this property is fundamental for signal filtering.

Classical name	Equivalent graph	Eigenvalues	Eigenvectors
DFT	Ring	$\lambda_\ell = 2 - 2\cos(\frac{\pi\ell}{N})$ when ℓ is even and $\lambda_\ell = 2 - 2\cos(\frac{\pi(\ell+1)}{N})$ when ℓ is odd.	$\chi_\ell(n) = \frac{1}{\sqrt{N}} \exp\left(\pi i \frac{\ell n}{N}\right)$ when ℓ is even and $\chi_\ell(n) = \frac{1}{\sqrt{N}} \exp\left(\pi i \frac{(N-\ell)n}{N}\right)$ when ℓ is odd.
RDFT	Ring	$\lambda_\ell = 2 - 2\cos(\frac{\pi\ell}{N})$ when ℓ is even and $\lambda_\ell = 2 - 2\cos(\frac{\pi(\ell+1)}{N})$ when ℓ is odd.	$\chi_\ell(n) = \sqrt{\frac{2}{N}} \cos(\frac{\pi\ell n}{N})$ when ℓ is even, $\chi_\ell(n) = \sqrt{\frac{2}{N}} \sin(\frac{\pi(\ell+1)n}{N})$ when ℓ is odd and $\chi_{N-1}(n) = \frac{1}{\sqrt{N}}(-1)^n$.
DCT-2	Path	$\lambda_\ell = 2 - 2\cos(\frac{\pi\ell}{N})$	$\chi_\ell(n) = \sqrt{\frac{2}{N}} \cos\left(\pi \frac{\ell(n-0.5)}{N}\right)$

Table 1: Eigenvalues and eigenvector for classical configuration. $n \in \{1, 2, ..., N\}$ and $\ell \in \{1, 2, ..., N-1\}$. Note that $\chi_0(n) = \frac{1}{\sqrt{N}}$.

We define a generalized convolution product of signal $f, g \in \mathbb{R}$ on a graph by

$$(f * g)(n) := \sum_{\ell=0}^{N-1} \hat{f}(\ell)\hat{g}(\ell)\chi_\ell(n).$$

3.3.1 Properties

For $f, g, h \in \mathbb{R}$ and $\alpha, \beta \in \mathbb{R}$, the generalized convolution product satisfies the following properties:

1. **Convolution in the vertex domain is equivalent to the multiplication in the spectral domain:**
$$\widehat{f * g} = \hat{f} \cdot \hat{g}.$$
2. **Comutativity:**
$$(f * g) = (g * f).$$
3. **Linearity:**
$$f * (\alpha g + \beta h) = \alpha(f * g) + \beta(f * h).$$
4. **Associativity:**
$$f * (g * h) = (f * g) * h.$$
5. **Invariance to the graph Laplacian operator:** The Laplacian is the difference operator.
$$\mathcal{L}(f * g)(n) = \sum_{\ell=0}^{N-1} \lambda_\ell \hat{f}(\ell)\hat{g}(\ell)\chi_\ell(n) = (\mathcal{L}f * g)(n) = (f * \mathcal{L}g)(n).$$
6. **Multiplicative identity:** for the signal $g_0 \in \mathbb{R}^N$ defined by $g_0(n) = \sum_{\ell=0}^{N-1} \chi_\ell(n)$, we have:
$$f * g_0 = f.$$
We observe that $\hat{g}(\ell) = 1$, for all $\ell = 0, 1, ..., N-1$ and $\|g_0\|_2 = \sqrt{N}$.
7. **Inverse element** If $\hat{g}(\ell) \neq 0$, $\forall \ell \in \{0, 1, ..., N-1\}$, then the convolution product has an inverse element $g^{-1}(n) = \sum_{\ell=0}^{N-1} \frac{1}{\hat{g}(\ell)} \chi_\ell(n)$ so that
$$g * g^{-1} = g_0.$$

8. **Sum of the generalized convolution product:**

$$\sum_{n=1}^{N}(f*g)(n) = \frac{1}{\sqrt{N}}\left[\sum_{n=1}^{N} f(n)\right]\left[\sum_{n=1}^{N} g(n)\right] = \sqrt{N}\hat{f}(0)\hat{g}(0).$$

3.4 A generalized translation

The intuition of translation is obvious for functions in L^2. It is simply a shift. For graphs, there aren't such comparisons. In the classical case, the translation can be expressed with the convolution. By analogy, we define a generalized translation operator $T_i : \mathbb{R}^N \to \mathbb{R}^N$ defined on a graph $\mathcal{G}$ and any $i \in \{1, 2, ..., N\}$, using the convolution with a delta δ_i centered at vertex i:

$$(T_i f)(n) := \sqrt{N}(f * \delta_i)(n) = \sqrt{N}\sum_{\ell=0}^{N-1} \hat{f}(\ell)\chi_\ell^*(i)\chi_\ell(n)$$

Because of the factor $\sqrt{N}$, we do have the nice property:

$$\begin{aligned}
\sum_{n=1}^{N} T_i f(n) &= \sqrt{N}\sum_{n=1}^{N}\sum_{\ell=0}^{N-1} \hat{f}(\ell)\chi_\ell^*(i)\chi_\ell(n) \\
&= \sqrt{N}\sum_{\ell=0}^{N-1} \hat{f}(\ell)\chi_\ell^*(i)\sum_{n=1}^{N}\chi_\ell(n) \\
&= \sqrt{N}\sum_{\ell=0}^{N-1} \hat{f}(\ell)\chi_\ell^*(i)\sqrt{N}\delta_0(\ell) \\
&= N\hat{f}(0)\chi_0^*(i) \\
&= N\sum_{n=1}^{N} f(n)\chi_0^*(n)\frac{1}{\sqrt{N}} \\
&= \sqrt{N}\sum_{n=1}^{N} f(n)\frac{1}{\sqrt{N}} \\
&= \sum_{n=1}^{N} f(n)
\end{aligned}$$

Remark: In general, graphs are not shift-invariant. The shape of a signal is transformed by the shift. Even on a ring, depending on the choice of the eigenvectors, the translated signal does not look like the original one.

There is no easy intuitive way of understanding the generalized translation. It depends highly on the graph and of the choice on the eigenvectors. Nevertheless, we observe a nice property for the heat-kernel window $\hat{g}(\ell) = Ce^{-\lambda_\ell \tau}$. The shift T_i localizes the window around the vertex i (Figure 1).

Question: Are there other kernels that have the same localization property?

3.5 A generalized modulation

The classical modulation is a multiplication by an eigenvector. By analogy, we define, for $k \in \{0, 1, ..., N-1\}$, a generalized modulation operator $M_k : \mathbb{R}^N \to \mathbb{R}^N$ by

$$(M_k f)(n) := \sqrt{N} f(n)\chi_k(n).$$

We remark that M_0 is the identity operator. Indeed, $\chi_0(n) = \frac{1}{\sqrt{N}}$ for a connected graph. The classical modulation is equivalent to a shift in frequency:

$$\widehat{M_k(f)}(\omega) = \hat{f}(\omega - \xi), \forall \omega \in \mathbb{R}.$$

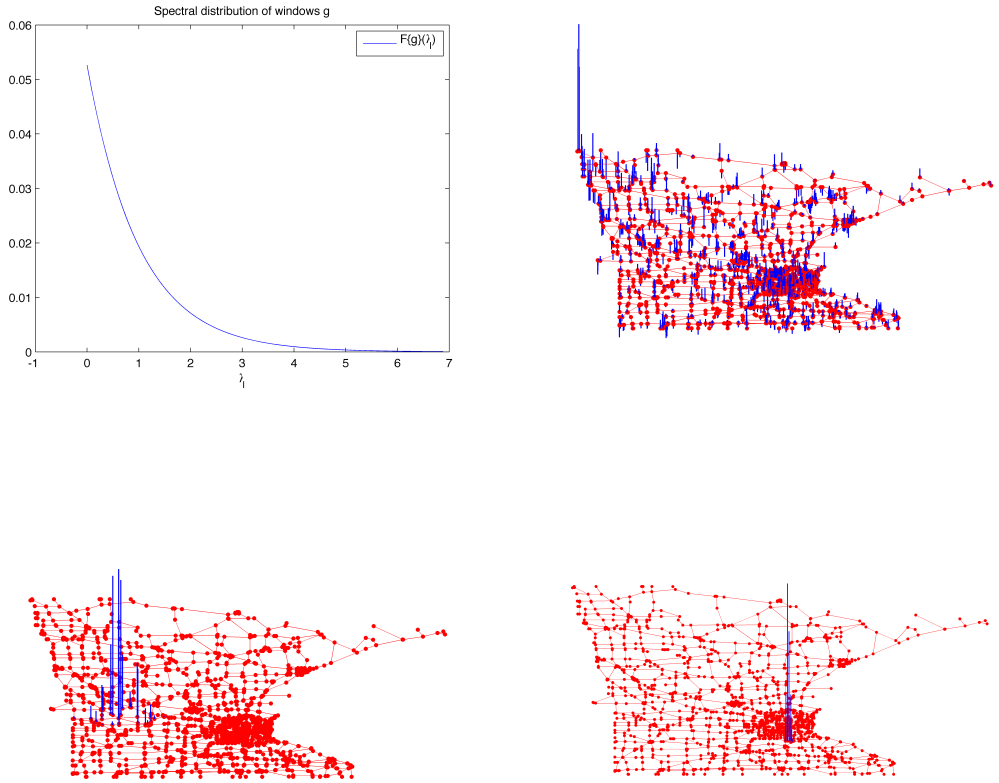

Figure 1: Localization property of the heat kernel on the Minnesota road graph. Top left: Spectral distribution of the kernel $\hat{g}(\ell) = Ce^{-\lambda_\ell \tau}$. τ is chosen equal to 1 and C is designed so that $\|g\|_2 = 1$. Top right: kernel in the vertex domain. Bottom left: kernel shifted on vertex 1024. Bottom right: kernel shifted on vertex 2024.

This property is not preserved for the generalized modulation. However, we do have the nice property that if $\hat{g}(l) = \delta_0(l)$, then

$$\widehat{M_k g(\ell)} = \delta_0(\ell - k).$$

This means that the DC component of any signal $f \in \mathbb{R}^N$ is mapped by the modulation M_k to $\hat{f}(0)\chi_k$. As shown on the Figure 2, the non DC component will spread around λ_k. This phenomenon is quantified in [3].

3.6 A windowed graph Fourier transform

As well as the generalized shift and modulation, the WGFT is presented by D. Shuman et al. in [3]. Similarly to the classical cases (2) and (3), we define for a window $g \in \mathbb{R}$, a windowed graph Fourier atom by:

$$g_{i,k}(n) := (M_k T_i g)(n) = \sqrt{N}\chi_k(n) \sum_{\ell=0}^{N-1} \hat{g}(\ell)\chi_\ell^*(i)\chi_\ell(n).$$

In addition, the WGFT of a signal $f \in \mathbb{R}^N$ is:

$$Sf(i,k) = \langle f, g_{i,k} \rangle = N \sum_{n=1}^{N} f(n)\chi_k^*(n) \sum_{\ell=0}^{N-1} \bar{\hat{g}}(\ell)\chi_\ell(i)\chi_\ell^*(n). \quad (7)$$

This way of defining the ambiguity function is simple and logical but it will not preserve the energy since the modulation and the translation are not energy invariant operators. This is

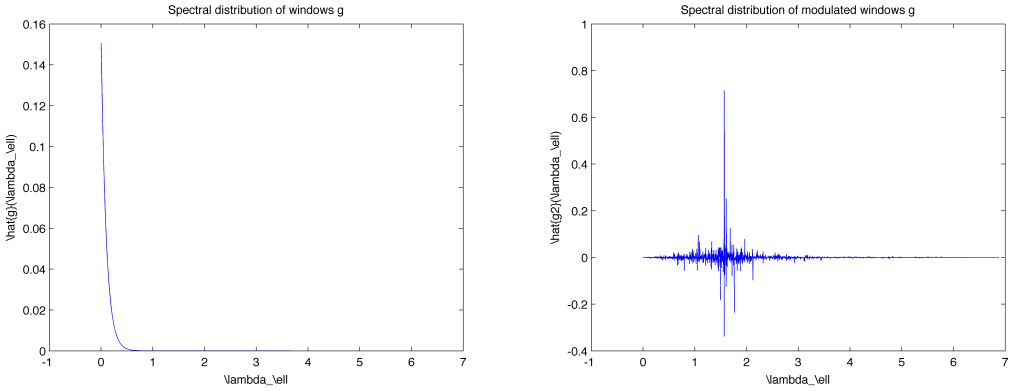

Figure 2: Spectral localization property of the heat kernel on the Minnesota road graph. Left: Spectral distribution of the kernel $\hat{g}(\ell) = Ce^{-\lambda_\ell \tau}$. τ is chosen equal to 10 and C is designed such that $\|g\|_2 = 1$. Right: Spectral distribution of the modulated kernel: $M_{1000}g$. Note that $\lambda_{1000} = 1.53$.

presented in Figure 3. In the ring case, there are border effects, but the norm seems quite well preserved by the modulation and the shift. However, in a random graph like the sensor network, this is not true anymore.

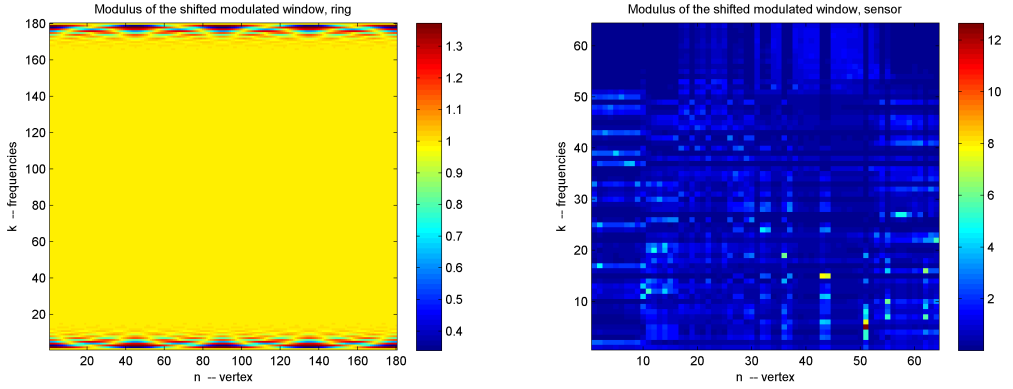

Figure 3: Norm of the shifted modulated windows $\hat{g}(\ell) = \exp(-\lambda_\ell \tau)$. The fact that from frequency 20 to 160, the shifted modulated window has an unitary norm is not true for any kernel with the ring graph. Left: Ring with RDFT eigenvectors, $\tau 50$. Right: Sensor network (see Figure 5), $\tau = 0.4$.

One may be tempted to solve this problem by normalizing the atoms. In that case, they are defined by:

$$g_{i,k}(n) := \frac{(M_k T_i g)(n)}{\|M_k T_i g\|_2}.$$

We have to be careful with this normalization. Indeed, the modulation and the translation of a signal can both make it vanish. In that case, the energy of the signal becomes 0 and the definition lacks sense. However if all the eigenvectors have no 0 values ($|\chi_\ell(n)| > 0, \forall n \in \{1, 2, ..., N\}, \ell \in \{0, 1, ..., N-1\}$) , then the normalized atoms are always well defined. The demonstration for the modulation is trivial and for the translation operator, it is a consequence of Theorem 21. In order to have no hypothesis on the graph or for numerical application, we can satisfy ourself with $\|M_k T_i g\|_2 > 0$, $\forall k, i$. We are now able to define a normalized windowed graph Fourier transform

(NWGFT) of a signal $f \in \mathbb{R}^N$:

$$Sf(i,k) = \langle f, g_{i,k} \rangle = \frac{\langle f, M_k T_i g \rangle}{\|M_k T_i g\|_2}.$$

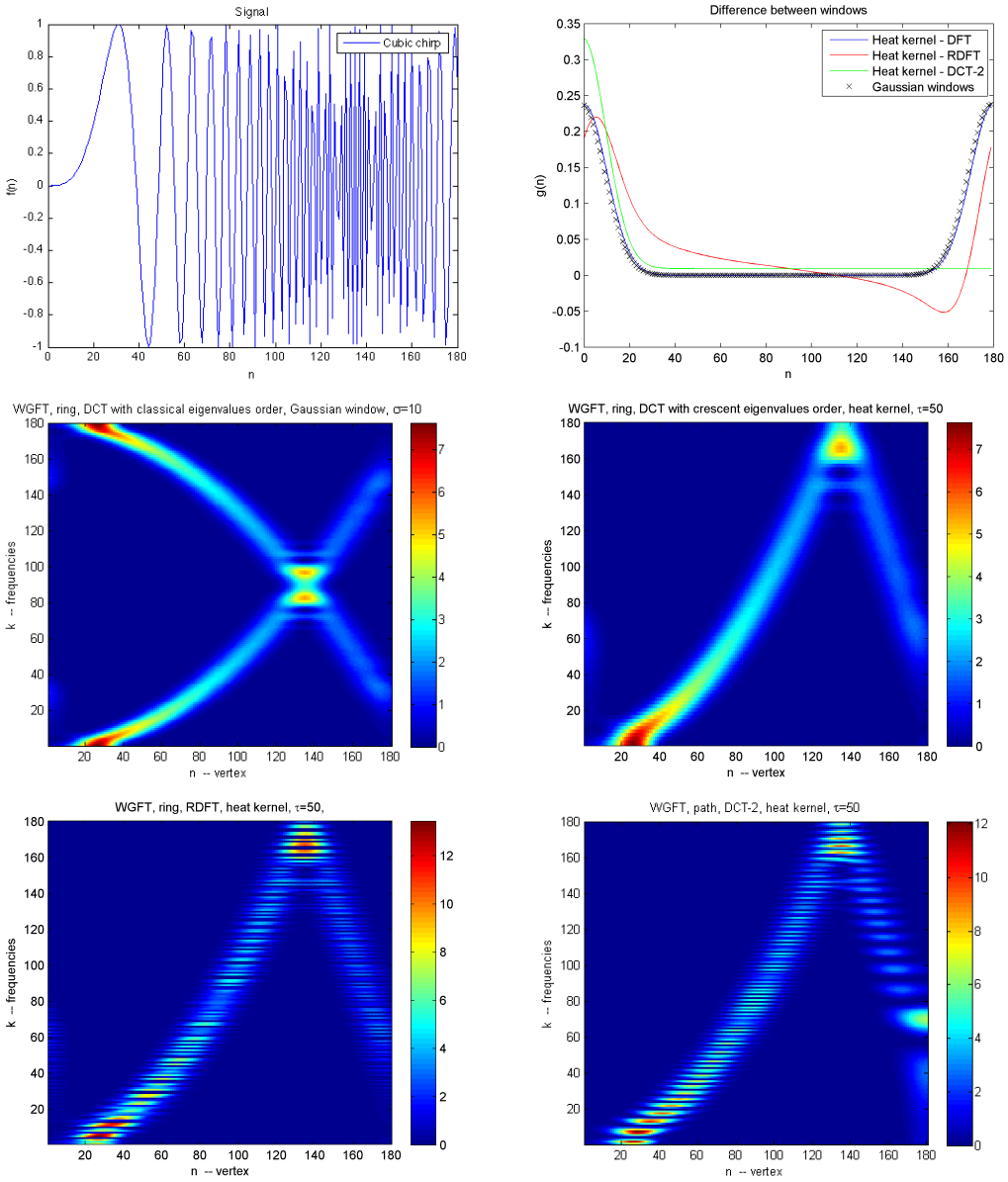

Figure 4: WGFT "spectrogram" example on the ring and path graphs. Top left: the analyzed signal is a cubic chirp: $f(n) = \sin(0.6\pi n^3)$. Top right: windows used for the different WGFTs. The shape of the heat-kernel in the vertex domain depends highly on the graph and on the choice of the eigenvectors. Middle left: classical windowed Fourier transform of f with classic eigenvalues order. Middle right: classical windowed Fourier transform of f with crescent eigenvalues order. Bottom left: WGFT of f using RDFT and the heat kernel. Bottom right: WGFT of f on a path graph using the heat kernel.

Figure 4 presents an example of the spectrogram of WGFT: $|Sf(i,k)|^2$. We analyze a cubic chirp $f(n) = \sin(0.6\pi n^3)$ using different normalized windows for the ring and the path graphs. The eigenvectors are given in table 1. The heat kernel window is: $\hat{g}(\ell) = Ce^{-\lambda_\ell \tau}$. τ is equal to 50

and C is designed such that $\|g\|_2 = 1$. The Gaussian window is a discrete normalized Gaussian of variance: $\sigma = 10$. We first observe that the choice of the eigenvector has an influence on the shape of the window in the vertex domain: there is a difference between the RDFT and the DFT. We also observe, that in the ring-DFT case, the heat kernel is equivalent to a Gaussian window. The effect of the order of the eigenvalues is also observed. With the classical order, we have the classical windowed Fourier transform with a symmetry in the middle of the spectrum because the signal f is real. Reordering the eigenvalues sort all frequencies in a crescent manner. We then use the RDFT-eigenvectors, and this affect the window and the result. However, despite the appearance of gaps in the WGFT, we are still able to recognize the chirp in frequency.

As a second example, we are going to consider a sensor network with the heat kernel. We separate the graph in two parts as show n Figure 5, and we create a signal f with the composition of two eigenvectors. On the blue bar vertices (1-15), we put the signal χ_{50} and on the other (16-64) the signal χ_{10}. The WGFT and the NWGFT are presented in Figure 6. We observe the trade-offs between the resolution in the frequency and vertex domains. When τ gets bigger, the window becomes sharper in frequency and the precision of the spectrogram increases in frequency and decreases in space.

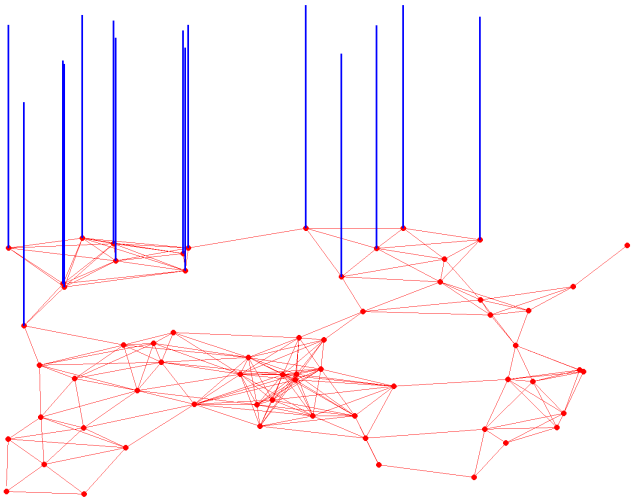

Figure 5: Sensor network separated in two parts by the blue bars: 64 vertices including only 15 bars.

4 The Ambiguity Function

With the windowed graph Fourier transform, we are able to define an ambiguity function by taking the scalar product between an original window and its shifted modulated transform.

The non-normalized ambiguity function for a window g is defined as

$$A_g(i,k) := \langle g, g_{i,k} \rangle = N \sum_{n=1}^{N} g(n) \chi_k^*(n) \sum_{\ell=0}^{N-1} \overline{\hat{g}}(\ell) \chi_\ell(i) \chi_\ell^*(n). \tag{8}$$

We notice that this definition is equivalent to the WGFT of the window g given in (7). In this section, we demonstrate some theorems for the WGFT and others specific to the ambiguity function.

One may be tempted to normalize the ambiguity function. The definition (8) becomes:

$$A'_g(i,k) := \langle g, \frac{M_k T_i g}{\|M_k T_i g\|_2} \rangle \tag{9}$$

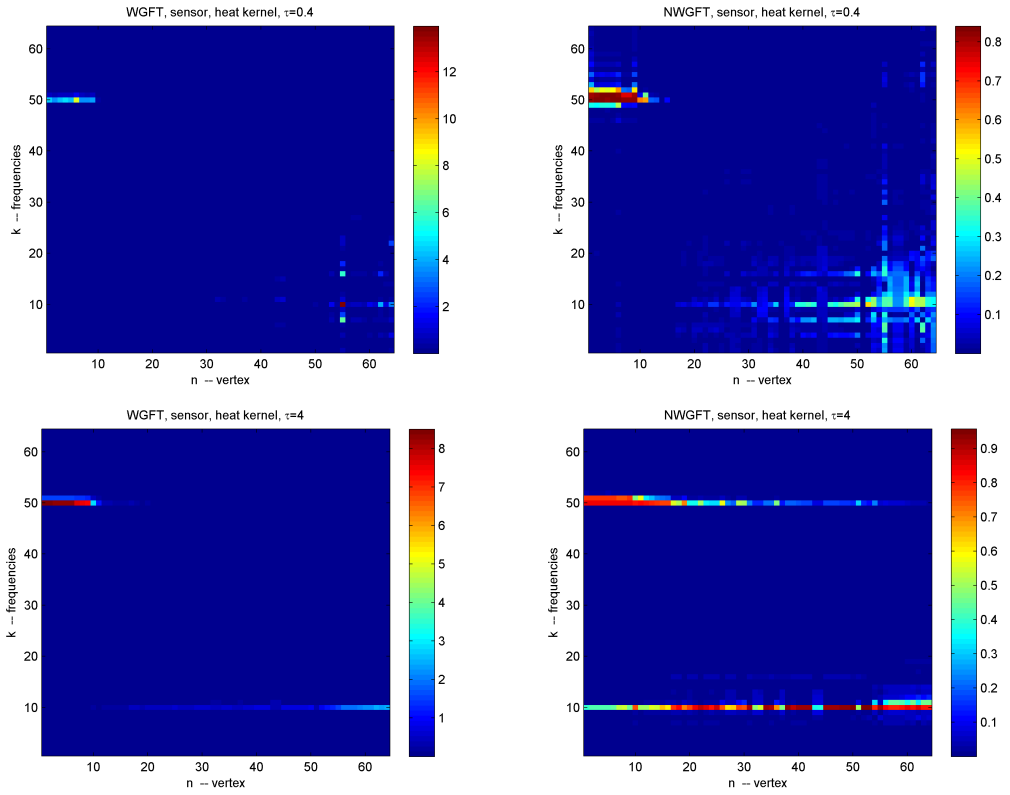

Figure 6: Spectrograms $|Sf|^2$. Top left: WGFT of f with $\tau = 0.4$. Top right: NWGFT of f with $\tau = 0.4$. Bottom left: WGFT of f with $\tau = 4$. Bottom right: NWGFT of f with $\tau = 4$.

We need to be careful with this definition. Like for the WGFT, the generalized modulation and translation can in special cases make the signal disappear (all coefficients equal to zero). We require that $|\chi_\ell(n)| > 0, \forall n \in \{1, 2, ..., N\}, \ell \in \{0, 1, ..., N-1\}$ to assure the existence of the normalized ambiguity function. This hypothesis can be exchanged with $\|M_k T_i g\|_2 > 0$, $\forall i, k$.

Definitions (8) and (9) are well suited for a kernel g that is localized in both vertex and spectral domains. If it is not, one may also define the ambiguity function in a different way:

$$A_g(i, i_0, k, k_0) := \langle g_{i,k}, g_{i_0,k_0} \rangle \tag{10}$$

In that case, the cross-ambiguity function would be:

$$A_{g,f}(i, i_0, k, k_0) := \langle g_{i,k}, f_{i_0,k_0} \rangle.$$

Again, we define a second normalized ambiguity function as:

$$A'_g(i, i_0, k, k_0) := \langle \frac{M_{k_0} T_{i_0} g}{\|M_{k_0} T_{i_0} g\|_2}, \frac{M_k T_i g}{\|M_k T_i g\|_2} \rangle. \tag{11}$$

In order to evaluate the spread of the ambiguity function $A(i, k)$ with $i \in \{1, 2, ..., N\}$ and $k \in \{0, 1, ..., N-1\}$, we will use the "entrywise" p-norm ($p > 0$) defined by:

$$\|A\|_p := \left(\sum_i \sum_k |A(i, k)|^p \right)^{\frac{1}{p}}.$$

If we use the definitions with four parameters ((10),(11)), we fix i_0 and k_0.

Demonstrations will be done mostly for (8) and (10) because the norm of the shifted modulated window is still hard to handle.

As expected, we observe that the spreading of the ambiguity function in space and in frequency characterizes the spread of the window. In our first example (Figure 7) we observe the ambiguity functions (8) and (10) for a ring graph with RDFT eigenvectors. We notice that using (8) the biggest coefficients are localized around node 1 and frequency χ_0. Moreover, the trade-off between frequency and vertex resolution seems to be observed with the variation of τ. Increasing τ allows the resolution in the vertex domain to grow, but diminishes the resolution in frequency.

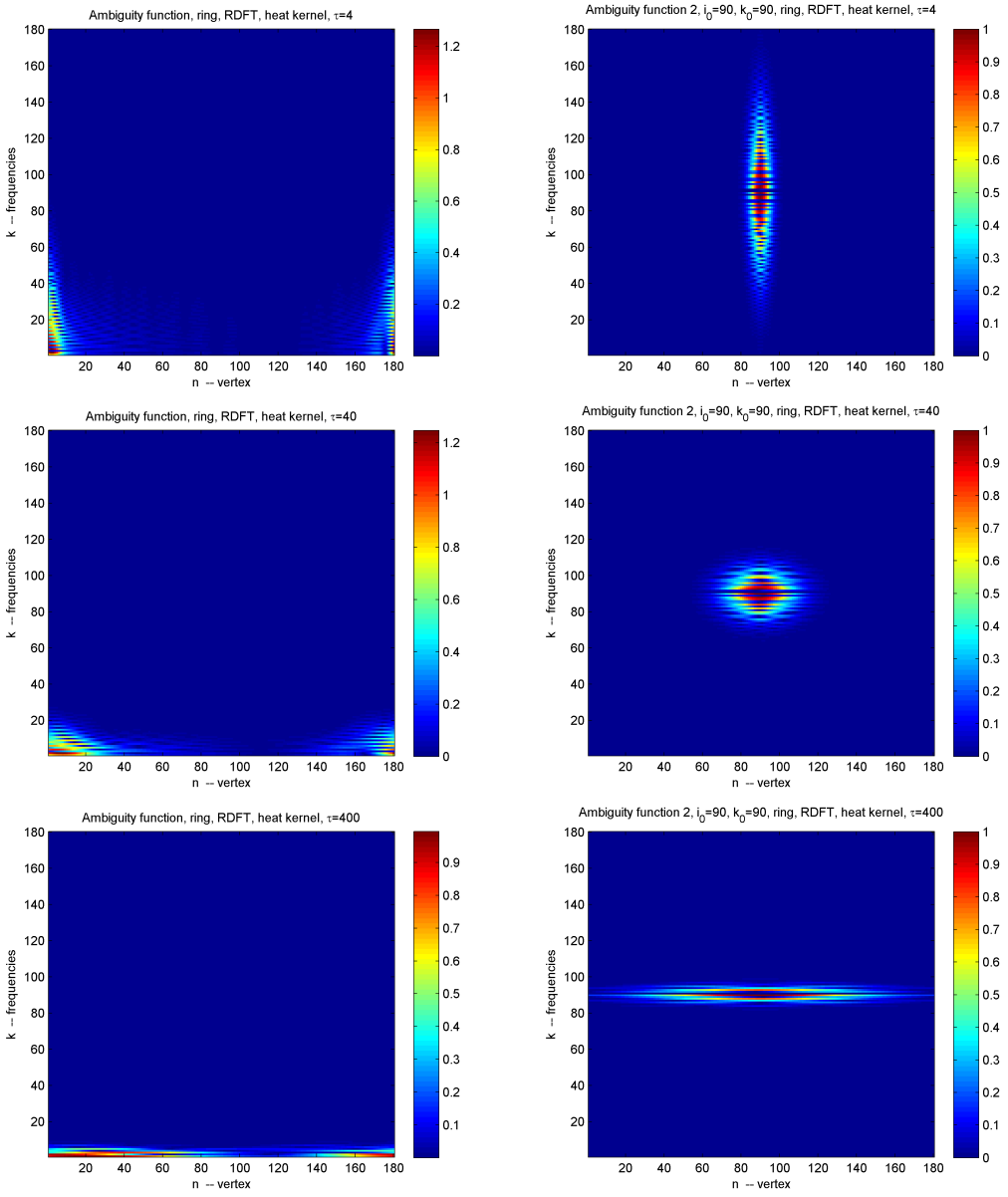

Figure 7: Square norm of ambiguity function for the ring graph with RDFT eigenvectors (Table 1). The heat kernel is chosen ($\hat{g}(\ell) = Ce^{-\lambda_\ell \tau}$, C is such that $\|g\|_2 = 1$). Top left: definition (8) with $\tau = 4$. Top right: definition (10) with $\tau = 4$, $i_0 = 90$ and $k_0 = 90$. Middle left: definition (8) with $\tau = 40$. Middle right: definition (10) with $\tau = 40$, $i_0 = 90$ and $k_0 = 90$. Bottom left: definition (8) with $\tau = 400$. Bottom right: definition (10) with $\tau = 400$, $i_0 = 90$ and $k_0 = 90$.

In the second example we first observe the differences between the four different definitions in a sensor graph (Figure 8). We can see that definitions (10) and (11) localize the ambiguity function "around" i_0 and k_0. The closer the vertices are to vertex i_0 (on the graph, not in the figure) the bigger the coefficients are. This localization property is due to the heat kernel. With it, the shift

T_i will move the window around the vertex i. However, definitions (8) and (9) lack localization in the vertex domain. In this example, the kernel g seems to be placed around vertex 25, but this is only by chance. There is no localization guarantee for the unshifted heat kernel. One may be tempted to choose a kernel that is localized in both domains. In that case, it is uncertain that the shift T_i will move the window around vertex i.

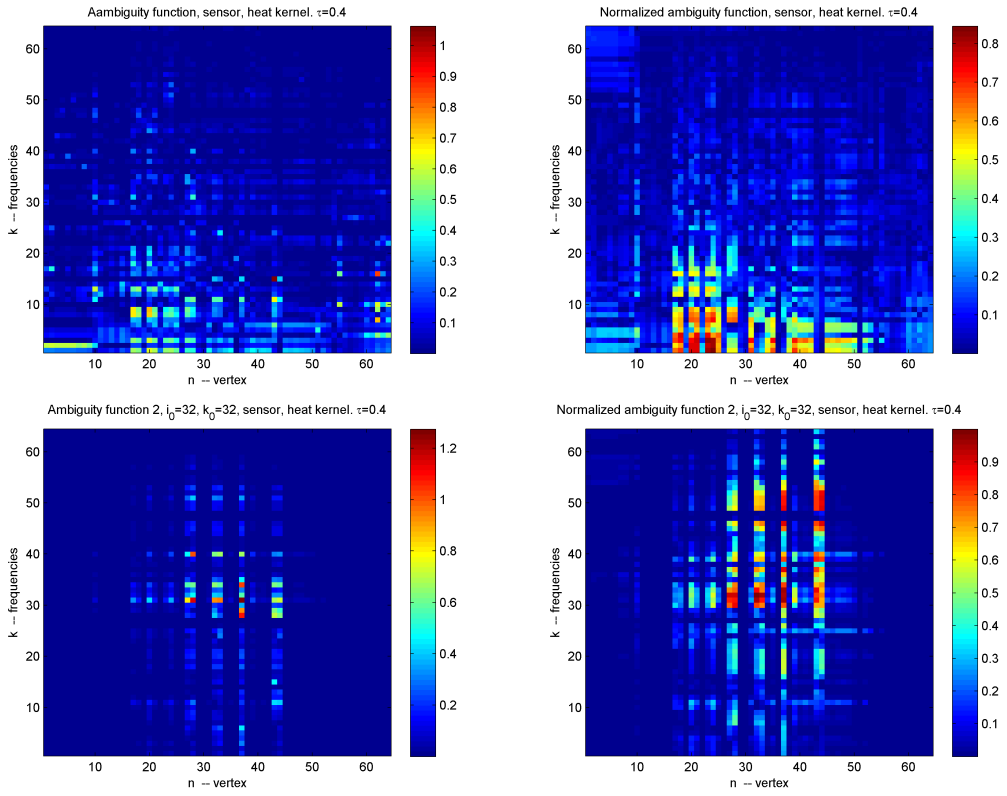

Figure 8: Square norm of ambiguity function for the sensor graph presented in Figure 5. The heat kernel ($\hat{g}(\ell) = Ce^{-\lambda_\ell \tau}$ is analyzed, C is chosen such that $\|g\|_2 = 1$ with $\tau = 0.4$. Top left: definition (8). Top right: definition (10), $i_0 = 32$ and $k_0 = 32$. Bottom left: definition (9). Bottom right: definition (11), $i_0 = 32$ and $k_0 = 32$.

Computing the norm of the ambiguity function is a way of evaluating the spread of a window. Values for the four different definitions are given in Table 2. Note that $\|A_g\|_2 \leqslant \|A_g\|_1 \leqslant N\|A_g\|_2$ and $\|A_g\|_\infty \leqslant \|A_g\|_2 \leqslant N\|A_g\|_\infty$. The case that is most sparse is $\|A_g\|_1 = \|A_g\|_2 = \|A_g\|_\infty$, which means that all the energy is concentrated into one single value. The least sparse case is $\|A_g\|_1 = N\|A_g\|_2 = N^2\|A_g\|_\infty$, which means that the energy is equally distributed along all values. As a consequence, we want those norms to be close to each other in order to have a sparse ambiguity function. If the ambiguity function is sparse then the window is confined in both vertex and spectral domains. Computing the value $\frac{\|A_g\|_1}{\|A_g\|_\infty}$ gives an idea of the sparsity. Numerically, the normalized definitions give less sparse ambiguity functions than unormalized definitions. This can be checked in Figure 8.

In Figure 9, we observe the effect of τ on the ambiguity function. This parameter allows us to control the tradeoff between the spread in the vertex and spectral domains.

The spread of the ambiguity function has been observed experimentally. It seems, there is a trade-off between the frequency and the vertex resolution. Now, we will try to express this trade-off as an uncertainty principle as follows: $\|A_g\|_1 \geqslant a$ and $\|A_g\|_\infty \leqslant b$ which implies that:

$$\frac{\|A_g\|_1}{\|A_g\|_\infty} \geqslant \frac{a}{b} > 1.$$

Definition	$\|A_g\|_1$	$\|A_g\|_2$	$\|A_g\|_\infty$	$\frac{\|A_g\|_1}{\|A_g\|_\infty} \leqslant N^2 = 4096$
$A_g(i,k) = \langle g, M_k T_i g \rangle$	211.6573	6.8311	1.0635	199.0196
$A'_g(i,i_0,k,k_0) = \langle M_k T_i g, M_{k_0} T_{i_0} g \rangle$	78.4283	5.0307	1.2737	61.5727
$A_g(i,k) = \langle g, \frac{M_k T_i g}{\|M_k T_i g\|_2} \rangle$	341.4633	9.3751	0.8450	404.0986
$A'_g(i,i_0,k,k_0) = \langle \frac{M_{k_0} T_{i_0} g}{\|M_{k_0} T_{i_0} g\|_2}, \frac{M_k T_i g}{\|M_k T_i g\|_2} \rangle$	224.7007	9.6587	1	224.7007

Table 2: "Entrywise" norms of the different definitions of the ambiguity function presented in Figure 8. $N = 64$.

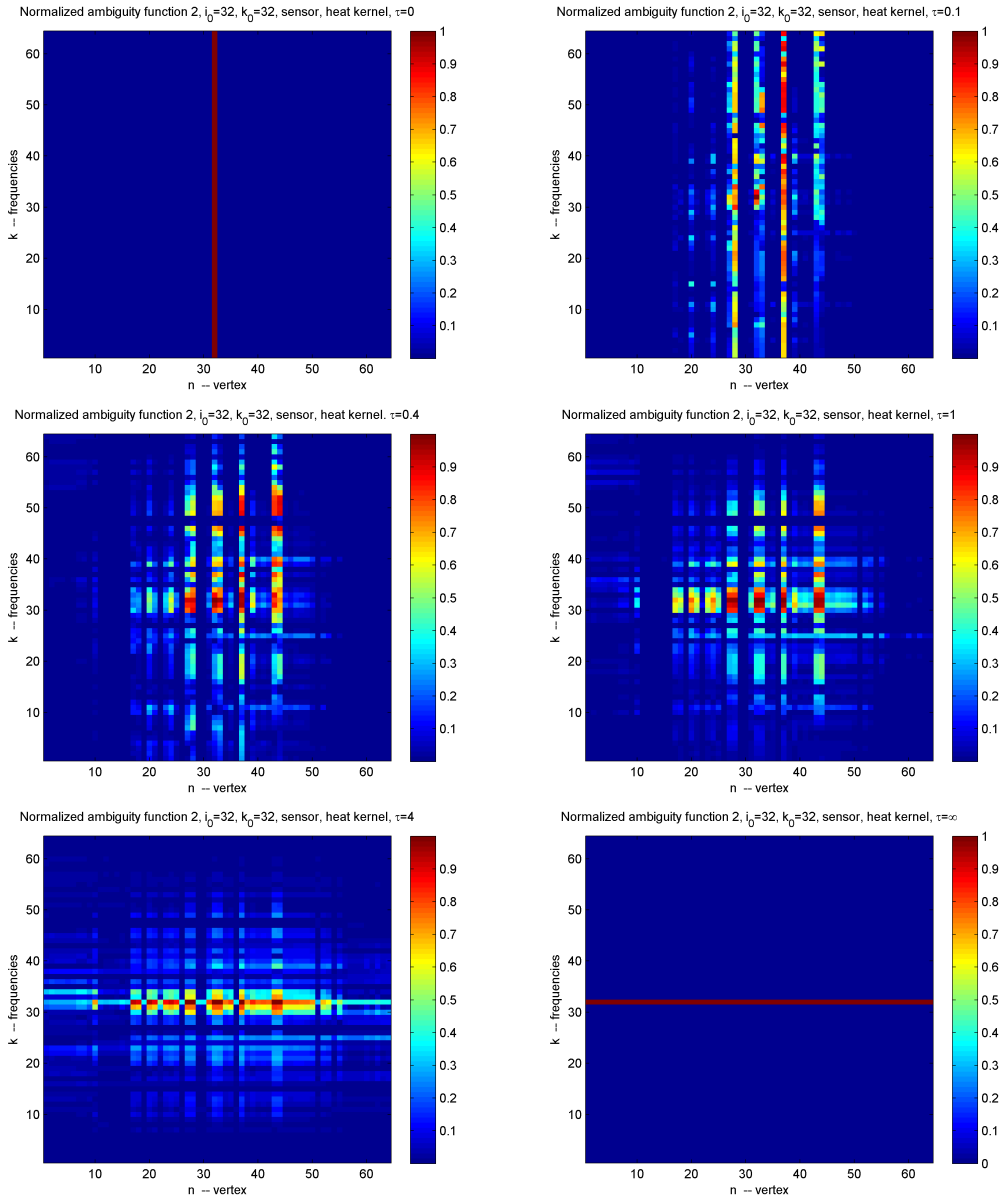

Figure 9: Square norm of ambiguity function (11) for the sensor graph presented in Figure 5. The heat kernel ($\hat{g}(\ell) = Ce^{-\lambda_\ell \tau}$, C is chosen such that $\|g\|_2 = 1$) is analyzed with respect to τ. We set $i_0 = 32$ and $k_0 = 32$. Top left: $\tau = 0$. Top right: $\tau = 0.1$. Middle left: $\tau = 0.4$. Middle right: $\tau = 1$. Bottom left: $\tau = 4$. Bottom right: $\tau = \infty$.

4.1 A bound for the 1-norm of the ambiguity function

By analogy to the classical case, the 1-norm of the ambiguity function cannot take arbitrarily small values for a normalized window g. In this section, we demonstrate the lower bound N for the 1-norm of the ambiguity function under some conditions linked to the kernel. The heat kernel, which is often used with graphs because of its good localization and computation properties, will satisfy the required hypotheses.

Question: Is the ambiguity function's lower 1-norm bound N true for every graph and every kernel?

Lemma 1. *If* $|\hat{g}(0)| \geqslant |\hat{g}(l)| \geqslant 0$ *for* $l = 1, 2, ...N-1$, *then*

$$|\hat{g}(0)|\|\hat{g}\|_1 \geqslant \|g\|_2^2.$$

Proof.

$$|\hat{g}(0)| \sum_{k=0}^{N-1} |\hat{g}(k)| \geqslant \sum_{k=0}^{N-1} |\hat{g}|^2(k) = \|\hat{g}\|_2^2 = \|g\|_2^2.$$

□

Theorem 1. *For* $g \in \mathbb{R}^N$, *satisfying* $\|g\|_2 = 1$ *and* $|\hat{g}(0)| \geqslant |\hat{g}(l)| \geqslant 0$ *for* $l = 1, 2, ...N-1$, *we have*

$$\|A_g\|_1 = \sum_{i=1}^{N} \sum_{k=0}^{N-1} |A_g(i,k)| \geqslant N,$$

with an equality if $g(n) = \frac{1}{\sqrt{N}}$.

Proof.

$$\begin{aligned}
\|A_g\|_1 &= \sum_{i=1}^{N} \sum_{k=0}^{N-1} |A_g(i,k)| \\
&\geqslant \sum_{k=0}^{N-1} \left| \sum_{i=1}^{N} A_g(i,k) \right| \\
&= N \sum_{k=0}^{N-1} \left| \sum_{n=1}^{N} \sum_{\ell=0}^{N-1} g(n) \chi_k^*(n) \overline{\hat{g}}(\ell) \chi_\ell^*(n) \sum_{i=1}^{N} \chi_\ell(i) \right| \\
&= N \sum_{k=0}^{N-1} \left| \sum_{n=1}^{N} \sum_{\ell=0}^{N-1} g(n) \chi_k^*(n) \overline{\hat{g}}(\ell) \chi_\ell^*(n) \delta_0(\ell) \sqrt{N} \right| \\
&= N^{\frac{3}{2}} \sum_{k=0}^{N-1} \left| \sum_{n=1}^{N} g(n) \chi_k(n) \overline{\hat{g}}(0) \chi_0^*(n) \right| \\
&= N^{\frac{3}{2}} |\hat{g}(0)| \sum_{k=0}^{N-1} \left| \hat{g}(k) \frac{1}{\sqrt{N}} \right| \\
&= N |\hat{g}(0)| \|\hat{g}\|_1 \\
&\geqslant N, \qquad (12)
\end{aligned}$$

where (12) comes from Lemma 1. This proves the first statement of the theorem.

To prove the equality for $g(n) = \frac{1}{\sqrt{N}}$, we simply compute the L_1 norm of A_g with this special window. Note that the Fourier transform of g is $\hat{g}(\ell) = \delta_0(\ell)$.

$$\begin{aligned}
\|A_g\|_1 &= \sum_{i=1}^{N}\sum_{k=0}^{N-1}|A_g(i,k)| \\
&= \sum_{i=1}^{N}\sum_{k=0}^{N-1}\left|N\sum_{n=1}^{N}\frac{1}{\sqrt{N}}\chi_k^*(n)\sum_{\ell=0}^{N-1}\delta_0(\ell)\chi_\ell(i)\chi_\ell^*(n)\right| \\
&= \sqrt{N}\sum_{i=1}^{N}\sum_{k=0}^{N-1}\left|\sum_{n=1}^{N}\chi_k^*(n)\chi_0(i)\chi_0^*(n)\right| \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{k=0}^{N-1}\left|\sum_{n=1}^{N}\chi_k^*(n)\right| \\
&= \sqrt{N}\sum_{k=0}^{N-1}\left|\sqrt{N}\delta_0(k)\right| \\
&= N.
\end{aligned}$$

□

This last theorem gives us a sharp inequality for decreasing Fourier kernels. Consequently, it is true for the heat-kernel.

Corollary 1. *For $f, g \in \mathbb{R}^N$, we have*

$$\|A_g f\|_1 = \sum_{i=1}^{N}\sum_{k=0}^{N-1}|A_g f(i,k)| \geqslant N|\hat{g}(0)|\|\hat{f}\|_1.$$

The proof is not given because it is very similar to the previous theorem.

Theorem 2. *For $f, a, b \in \mathbb{R}^N$ such that $\|f\|_2 = 1$ and $f = \Phi a = \Psi b$, with Φ, Ψ two $N \times N$ matrices made of N orthonormal vectors ϕ_i, ψ_j $(i, j = 1, ..., N)$, we have:*

$$\|a\|_1\|b\|_1 \geqslant \frac{1}{\mu}.$$

Proof.

$$\begin{aligned}
1 &= |f^t f| \\
&= |\Phi^t a^t \Psi b| \\
&= \left|\sum_{i=1}^{N}\sum_{j=1}^{N} a_i\langle\phi_i, \psi_j\rangle b_j\right| \\
&\leqslant \sum_{i=1}^{N}\sum_{j=1}^{N}|a_i\langle\phi_i, \psi_j\rangle b_j| \\
&\leqslant \mu\sum_{i=1}^{N}|a_i|\sum_{j=1}^{N}|b_j| \\
&= \mu\|a\|_1\|b\|_1
\end{aligned}$$

Dividing the previous inequality by μ leads to the desired result.

□

This proof can be found in [6]. Applying Theorem 2 to the Dirac and the Fourier basis gives for $f \in \mathbb{R}^N$:

$$\|f\|_1\|\hat{f}\|_1 \geqslant \frac{\|f\|_2^2}{\mu} \tag{13}$$

This last equation is an uncertainty principles for the Fourier bases. For the classical case (the ring graph), we obtain a minimum coherence of $\frac{1}{\sqrt{N}}$. The uncertainty becomes weaker as the coherence increases.

Using Corollary 1 and (13) leads to

$$\|A_g f\|_1 = \sum_{i=1}^{N} \sum_{k=0}^{N-1} |A_g f(i,k)| \geqslant N|\hat{g}(0)| \frac{\|f\|_2^2}{\mu \|f\|_1} = \sqrt{N}| \sum_{n=1}^{N} g(n)| \frac{\|f\|_2^2}{\mu \|f\|_1},$$

which is another inequality for the ambiguity function. This bound has the advantage to be defined with f and g in the vertex domain. We observe that the greater the 1-norm of f is, the weaker the smaller the bound is.

The choice of the eigenvectors influences the coherence μ and has an impact on the ambiguity function. However, if the kernel is purely defined in the Fourier domain, then the choice of the eigenvector does not influence the bound given in Corollary 1 . Figure 10 illustrates the difference between the DFT and RDFT [5], which is due to the DFT having a lower coherence than the RDFT.

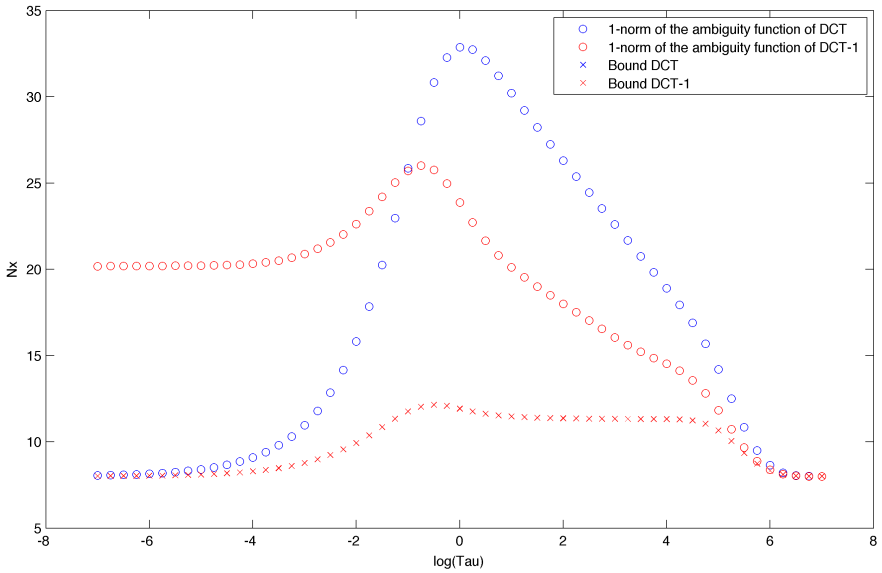

Figure 10: Ambiguity function value and bound for the ring case. The chosen window is $\hat{g}(\ell) = \hat{f}(\ell) = Ce^{-\lambda_\ell \tau}$ where C is chosen such that $\|f\|_2 = 1$. The bound $\sqrt{N}|\hat{g}(0)|\|\hat{g}\|_1$ for DFT and RDFT are the same. This is because the way we construct the Fourier transform of the window is independent from the choice eigenvectors.

4.2 A bound for the sup-norm of the ambiguity function

Theorem 3. *For a window $g \in \mathbb{R}^N$, we have*

$$\|A_g\|_\infty \leqslant \mu^2 N \|g\|_2^2$$

Proof.

$$\begin{aligned} \|A_g\|_\infty &= \max_{i,k} |\langle g, M_k T_i g \rangle| \\ &\leqslant \max_{i,k} \|g\|_2 \|M_k T_i g\|_2 \qquad (14) \\ &\leqslant \max_{i,k} \|g\|_2^2 \|T_i\|_2 \|M_k\|_2 \qquad (15) \\ &\leqslant \mu^2 N \|g\|_2^2, \qquad (16) \end{aligned}$$

where (14) follows from the Cauchy-Schwarz inequality and, (15) and (16)follows from (51) and (50). □

4.3 An uncertainty principle

Using Theorems 1 and 3 for a normalized heat kernel leads to an uncertainty principle theorem.

Theorem 4. *For a window $g \in \mathbb{R}^N$, satisfying $\|g\|_2 = 1$ and $|\hat{g}(0)| \geqslant |\hat{g}(l)| \geqslant 0$ for $l = 1, 2, ...N-1$, we have*

$$\frac{\|A_g\|_1}{\|A_g\|_\infty} \geqslant \frac{1}{\mu^2} > 1. \tag{17}$$

We notice that if we choose classical eigenvectors(DFT), then $\mu = \frac{1}{\sqrt{N}}$, and we obtain the same result as in the classical case (26). However Theorem 1 requires a hypothesis on the kernel. In our simulation, this theorem was always true even when the hypothesis was not satisfied. Demonstrating this relation without any hypotheses would lead to a more general uncertainty principle. The uncertainty given by Theorem 4 depends highly on μ. If the coherence is low, the uncertainty is strong. On the other hand, if it increases, the uncertainty diminishes. This is logical because the bases get closer and we can be precise in both domains simultaneously.

This suggests a new problem: finding the basis with the smallest coherence possible for a general graph.

Question: How can we found the Fourier basis with the smallest possible coherence for a general graph?

Extreme example: Let us suppose that we have a coherence equal to one. (It will be shown later that this case is not possible for a connected graph, for which the coherence is always strictly smaller than one. However, it can be arbitrary close to one.) There are i_0, ℓ_0 such that $\chi_{\ell_0}(n_0) = 1$; $\chi_\ell(n_0) = 0, \forall \ell \neq \ell_0$ and $\chi_{\ell_0}(n) = 0, \forall n \neq n_0$. Suppose that we chose $g(n) = \delta_{n_0}(n)$ as window. The ambiguity function becomes:

$$A_f(i,k) = \begin{cases} N & \text{if } i = n_0 \text{ and } k = \ell_0, \\ 0 & \text{otherwise.} \end{cases}$$

This means that $\|A_f\|_\infty = \|A_f\|_1 = N$, which implies no uncertainty any more. It's logical because we can be precise in both domains at the same time. This example shows that Theorem 4 is sharp for the extreme case $\mu = 1$.

Theorem 5. *For a connected graph, the coherence is always strictly smaller than one: $\mu < 1$.*

Proof. Suppose $\mu = 1$. There are i_0, ℓ_0 such that $\chi_{\ell_0}(n_0) = 1$; $\chi_\ell(n_0) = 0, \forall \ell \neq \ell_0$ and $\chi_{\ell_0}(n) = 0, \forall n \neq n_0$. The eigenvectors equation is

$$L\chi_{\ell_0} = \lambda_{\ell_0}$$

and implies

$$L(i, \ell_0) = \begin{cases} 1 & \text{if } i = \ell_0, \\ 0 & \text{otherwise.} \end{cases}$$

This is in contradiction with the definition of the Laplacian of a connected graph. This implies that $\mu \neq 1$. □

The sensor graph has a coherence close to one, as $\mu = 0.9186$. In Figure 11, we show that the ambiguity function can become really sparse with certain windows. However this sparsity is mostly due to the norm of the shifted modulated windows and not to the orthogonality of the atoms! Even if the uncertainty is minimized in this example, we observe that the spectrogram is not what we are expecting. The component χ_{10} is completely lost, and more importantly, we

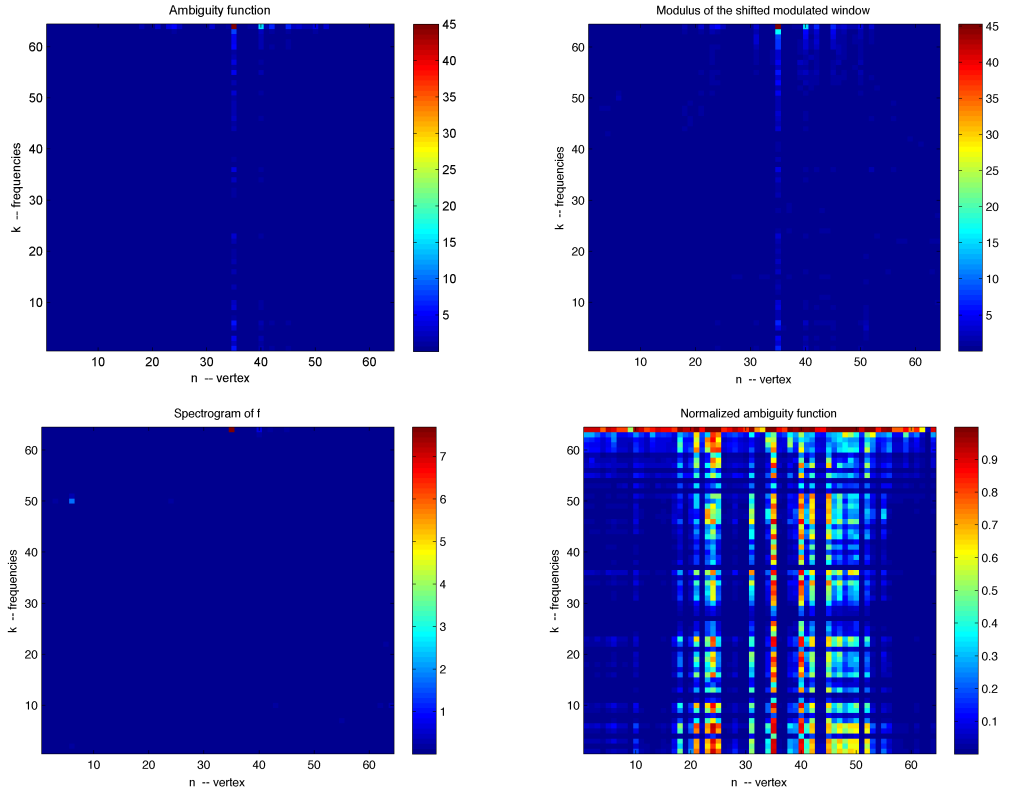

Figure 11: Top left: Ambiguity function for the sensor graph using the window $g(n) = \delta_{32}(n)$. We observe that $\|A_f\|_1 = 402.8812$ and $\|A_f\|_\infty = 45.0407$. Note that the maximum eigenvector value is given by $\mu = |\chi_{63}(32)| = 0.9186$. Top right: Norm of the shifted modulated window g. Bottom left: Spectrogram of f. The construction of f is presented in Section 3.6. Spectrogram with a suitable window are shown in Figure 6. Bottom right: Normalized ambiguity function of g.

.

observe coefficients at χ_{63} that were not in the signal. Finally, the normalized ambiguity function shows that this window is not a good choice for the NWGFT.

In the classical case, we usually want the window to be well localized in both domains. Gaussians are well suited for this the requirement and they optimize the uncertainty. However, for graphs, a window minimizing the uncertainty is not necessarily a good window because the shift and the modulation are not energy-invariant. As a consequence, the norm of a certain atom can be very small and some component of the WGFT cannot be detected.

Question: Does the NWGFT solve this problem?

5 Classical inequalities

At this point we start a new section that is almost independent of the work discussed so far. We describe the usual Fourier inequalities (Hausdorf-Young, Young, and Lieb) and explain their main applications. These expressions will provide an analog that will later inspire the generalization of these inequalities to the graph setting. There are two main interest in studying these inequalities. First they allow a better understanding of the Fourier transform and the convolution product. Second, they are used to demonstrate an uncertainty principle on the ambiguity function. Unfortunately, the proof of this uncertainty principle, first published by Lieb [1], is easily not generalizable to signals on graphs.

First we define some general constants that will be used for sharp inequalities. Let $0 < p \leqslant \infty$ and define q by $\frac{1}{q} + \frac{1}{p} = 1$, then D_p is defined to be

$$D_p = \left(\frac{p^{\frac{1}{p}}}{q^{\frac{1}{q}}} \right)^{\frac{1}{2}}$$

for $p \neq 1$ or ∞, while $D_1 = D_\infty = 1$.

For $1 < p < \infty$ and $1 \leqslant a, b \leqslant \infty$ with $\frac{1}{a} + \frac{1}{b} = 1$, we define $H(p, a, b) \geqslant 0$ by:

$$H(p, a, b)^2 = abp^{-2}|p-2|^{2-p}|p-a|^{-1+\frac{p}{a}}|p-b|^{-1+\frac{p}{b}},$$

with the convention that $0^0 \equiv 1$ and $H(1, 1, \infty) = H(1, \infty, 1) = 1$.

We also define $K(p, a, b) \geqslant 0$ by:

$$K(p, a, b)^2 = p^{-2}2^{2-p}a^{\frac{p}{a}}b^{\frac{p}{b}}$$

and $K(1, 1, \infty) = \sqrt{2}$. More details about those constants can be found in [1]

The classical p-norm of a function $f \in L^p$ is defined

$$\|f\|_p := \left(\int_{\mathbb{R}} |f(x)|^p \, \mathrm{d}x \right)^{\frac{1}{p}}.$$

Similarly, for a signal $f \in \mathbb{R}^N$, we define:

$$\|f\|_p := \left(\sum_{n=1}^{N} |f(n)|^p \right)^{\frac{1}{p}}.$$

5.1 Hausdorff-Young inequality

The Haudorff-Young inequality bound sthe norm of the Fourier transform of a function. This inequality uses the spread of a function in one domain(spectral or spacial) to bound the concentration in the other. The main application is to demonstrate lower and upper bounds for the ambiguity function.

Theorem 6. *Let $p' \geqslant 2$ and $\frac{1}{p'} = 1 - \frac{1}{p}$. If $f \in L^{p'}$, then $\hat{f} \in L^p$ and*

$$\|\hat{f}\|_p \leqslant D_{p'} \|f\|_{p'}.$$

Conversely, let $1 \leqslant p' \leqslant 2$ and assume $\hat{f} \in L^p$, then $f \in L^{p'}$ with $\frac{1}{q} = 1 - \frac{1}{p}$ and

$$\|\hat{f}\|_p \geqslant D_{p'} \|f\|_{p'}.$$

Proofs can be found in [7] and [8]. It is also demonstrated that the constants are sharp since equalizing are achieved by Gaussians.

We can build analogous inequalities for discrete signals. The proof is more straightforward, but the constants obtained do not give sharp inequalities.

Theorem 7. *Let $p' \geqslant 2$ and $\frac{1}{p'} = 1 - \frac{1}{p}$. If $f \in \mathbb{R}^N$, then*

$$\|\hat{f}\|_p \leqslant N^{\frac{1}{p}-\frac{1}{2}} \|f\|_{p'}.$$

Conversely, let $1 \leqslant p' \leqslant 2$, and assume $f \in \mathbb{R}^N$, then

$$\|\hat{f}\|_p \geqslant N^{\frac{1}{2}-\frac{1}{p}} \|f\|_{p'},$$

with $\frac{1}{p'} = 1 - \frac{1}{p}$.

The proof is not given here. Indeed, it is a particular case of the Hausdorff-Young inequality for graphs given in Theorem 13.

5.2 Young inequality

The next inequality is a sharp Young Inequality. It bounds the convolution product of two functions. For continuous signals f, g, the convolution product is defined as

$$(f * g)(t) = \int f(t-s)g(s)\,\mathrm{d}s,$$

and for two discrete signals $f, g \in \mathbb{R}^N$, it is defined as

$$(f * g)[n] = \sum_{k=1}^{N} f(n-k)g(k).$$

We use the definition of the circular convolution, which implies that signals are assumed to be periodic.

This product is important in Fourier analysis because convolution in the spatial domain corresponds to the multiplication in the spectral domain and vice-versa:

$$(f * g)(t) = \mathcal{F}^{-1}\{\hat{f} \cdot \hat{g}\}(t),$$

$$(\hat{f} * \hat{g})(\xi) = \mathcal{F}\{f \cdot g\}(\xi).$$

This property is fundamental in signal processing as it allows us to filter a signal using the convolution product.

Theorem 8. *Let $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ with $1 \leqslant p \leqslant \infty$, $1 \leqslant q \leqslant \infty$ and $1 \leqslant r \leqslant \infty$. Then, when $f \in L^p$ and $g \in L^q$, $f * g \in L^r$ and*

$$\|f * g\|_r \leqslant \frac{D_p D_q}{D_r} \|f\|_p \|g\|_q$$

Conversely, let $0 < p \leqslant 1$, $0 < q \leqslant 1$ and $0 < r \leqslant 1$ and assume $\hat{f} \in L^p$, then $f \in L^q$ with $\frac{1}{q} = 1 - \frac{1}{p}$ and

$$\|f * g\|_r \geqslant \frac{D_p D_q}{D_r} \|f\|_p \|g\|_q$$

Proofs were published by Beckner, Brascamp and Lieb [8] [7]. They also prove that equality are achieved when f and g were Gaussian.

The demonstrations for the strict inequality are not trivial. Thus, we take inspiration from a different demonstration that leads to a less strict inequality:

$$\|f * g\|_r \leqslant \|f\|_p \|g\|_q.$$

The demonstration uses the Riez-Thorin interpolation theorem and the following inequalities:

$$\|f * g\|_1 \leqslant \|f\|_1 \|g\|_1 \tag{18}$$

$$\|f * g\|_\infty \leqslant \|f\|_\infty \|g\|_1 \tag{19}$$

$$\|f * g\|_\infty \leqslant \|f\|_p \|g\|_{p'} \tag{20}$$

where $\frac{1}{p} + \frac{1}{p'} = 1$.

Theorem 9. *Riesz–Thorin interpolation theorem (p. 169 [9]) : Assume T is a bounded linear operator from L^{p_1} to L^{p_2}*

$$\|Tf\|_{p_2} \leqslant M_p \|f\|_{p_1}$$

and at the same time from L^{q_1} to L^{q_2}

$$\|Tf\|_{q_2} \leqslant M_q \|f\|_{q_1}.$$

Then it is also a bounded operator from L^{r_1} to L^{r_2}

$$\|Tf\|_{r_2} \leqslant M_r \|f\|_{r_1}.$$

with

$$\frac{1}{r_1} = \frac{t}{p_1} + \frac{1-t}{q_1}, \qquad \frac{1}{r_2} = \frac{t}{p_2} + \frac{1-t}{q_2},$$
$$M_r = M_p^t M_q^{1-t}$$

and t is any number between 0 and 1.

A proof of the interpolation theorem can be found in (p. 169 [9]).

Theorem 10. *For $f \in L^p$, $f \in L^q$ two functions and $1 \leqslant p, q, r \leqslant \infty$ so that $1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, we have*

$$\|f * g\|_r \leqslant \|f\|_p \|g\|_q. \tag{21}$$

Proof. For a function $f \in L^p$, we define an operator T_g by $(T_g f)(n) = (f * g)(n)$. Using (18) and (19), we observe that this operator is bounded from L^1 to L^1 by $\|g\|_1$ and from L^∞ to L^∞ by $\|g\|_1$. Thus, we can apply the Riez-Thorin theorem to this operator to get

$$\|f * g\|_p \leqslant \|f\|_p \|g\|_1.$$

Then, for a function $g \in L^q$, we define another operator T_f by $(T_f g)(n) = (f * g)(n)$. With the previous inequality and (20), we observe that this new operator is bounded from L^1 to L^p by $\|f\|_p$ and from $L^{p'}$ to L^∞ by $\|f\|_p$, where $\frac{1}{p} + \frac{1}{p'} = 1$. Again the Riez-Thorin theorem is applied and leads to the desired result.

$$\|f * g\|_r \leqslant \|f\|_p \|g\|_q,$$

where $1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. □

For the graph generalization we will try to follow the idea of Theorem 10. To do so, we need to find the inequalities analogous to (18), (19), and (20).

5.3 Lieb inequality

The Lieb inequality provides a lower bound on the ambiguity function.

Theorem 11. *Let $p > 2$ and assume that $f \in L^a$, $g \in L^b$ with $\frac{1}{a} + \frac{1}{b} = 1$ and with $\frac{p}{p-1} \leqslant a, b \leqslant p$. Then*

$$\|A_{f,g}\|_p^p = \int_{\mathbb{R}} \int_{\mathbb{R}} |A_{f,g}(\tau, \gamma)|^p \, \mathrm{d}\tau \, \mathrm{d}\gamma \leqslant H(p, a, b) \|f\|_a^p \|g\|_b^p. \tag{22}$$

Conversely, let $1 \leqslant p < 2$ and assume that $f \in L^a$, $g \in L^b$ with $\frac{1}{a} + \frac{1}{b} = 1$ and with $p \leqslant a, b \leqslant \frac{p}{p-1}$. Then

$$\|A_{f,g}\|_p^p = \int_{\mathbb{R}} \int_{\mathbb{R}} |A_{f,g}(\tau, \gamma)|^p \, \mathrm{d}\tau \, \mathrm{d}\gamma \geqslant K(p, a, b) \|f\|_a^p \|g\|_b^p. \tag{23}$$

The demonstration can be found in [1]. This last theorem gives fundamental information about the ambiguity function. For instance, if we take $f = g$ a real window and $a = b = 2$, (23) and (22) tell us that the L^1-norm of the ambiguity function cannot be arbitrarily small and the and the L^∞-norm cannot be arbitrarily large. As a result, the spread in space and in frequency of a window can not be arbitrarily small.

Inspired by Lieb [1], Feichtinger sketches a proof of a bound on the discrete ambiguity function [10].

Theorem 12. *For two signals $f, g \in \mathbb{R}^N$, if $2 < p < \infty$, $\frac{1}{a} + \frac{1}{b} = 1$, $\frac{p}{p-1} \leqslant a, b \leqslant p$ and $\frac{1}{p} + \frac{1}{p'} = 1$ then*

$$\|A_g f\|_p^p = \sum_{u=1}^{N} \sum_{k=0}^{N-1} |A_g f(u,k)|^p \leqslant N^{p-\frac{p}{p'}} \|f\|_a^p \|g\|_b^p.$$

Conversely, for two signals $f, g \in \mathbb{R}^N$, if $1 \leqslant p \leqslant 2$, $\frac{1}{a} + \frac{1}{b} = 1$, $p \leqslant a, b \leqslant \frac{p}{p-1}$ and $\frac{1}{p} + \frac{1}{p'} = 1$, then

$$\|A_g f\|_p^p = \sum_{u=1}^{N} \sum_{k=0}^{N-1} |A_g f(u,k)|^p \geqslant N^{p-\frac{p}{p'}} \|f\|_a^p \|g\|_b^p.$$

Proof. Suppose $p > 2$, then $\frac{p}{q} > 1$ and we have

$$\begin{aligned}
\|A_g f\|_p^p &= \sum_{u=1}^{N} \sum_{k=0}^{N-1} |A_g f(u,k)|^p \\
&= \sum_{u=1}^{N} \sum_{k=0}^{N-1} \left| \sum_{n=1}^{N} f[n] \overline{g}[u-n] e^{-2\pi i k \frac{n}{N}} \right|^p \\
&= N^{\frac{p}{2}} \sum_{u=1}^{N} \sum_{k=0}^{N-1} |\mathcal{F}\{f[\cdot] g[u-\cdot]\}(k)|^p \\
&= N^{\frac{p}{2}} \sum_{u=1}^{N} \|\mathcal{F}\{f[\cdot] g[u-\cdot]\}\|_p^p \\
&\leqslant N^{\frac{p}{2}} \sum_{u=1}^{N} N^{\frac{p}{p'} - \frac{p}{2}} \|f[\cdot] g[u-\cdot]\|_{p'}^p \qquad (24) \\
&= N^{p-\frac{p}{p'}} \sum_{u=1}^{N} \left(\sum_{n=1}^{N} |f[n] g[u-n]|^{p'} \right)^{\frac{p}{p'}} \\
&= N^{p-\frac{p}{p'}} \sum_{u=1}^{N} \left(\sum_{n=1}^{N} |f^{p'}[n]| |g^{p'}[u-n]| \right)^{\frac{p}{p'}} \\
&= N^{p-\frac{p}{p'}} \sum_{u=1}^{N} \left((|f^{p'}| * |g^{p'}|)(u) \right)^{\frac{p}{p'}} \\
&= N^{p-\frac{p}{p'}} \left\| |f^{p'}| * |g^{p'}| \right\|_{\frac{p}{p'}}^{\frac{p}{p'}} \\
&\leqslant N^{p-\frac{p}{p'}} \|f^{p'}\|_{\alpha}^{\frac{p}{p'}} \|g^{p'}\|_{\beta}^{\frac{p}{p'}} \qquad (25) \\
&= N^{p-\frac{p}{p'}} \|f\|_a^p \|g\|_b^p,
\end{aligned}$$

where (24) comes from the Hausdorff-Young inequality given in Theorem 7 and (25) results from the Young inequality given in Theorem 10. □

We use the L^1-norm as an indicator for sparsity. The smaller it is, the more sparse is the signal. Let's compare two different normalized signals in $\mathbb{R}^N$: $f_1[n] = \delta_1[n]$ and $f_2[n] = \frac{1}{\sqrt{N}}, \forall n \in \{1, 2, ..., N\}$. These two signals have the same energy, but a very different L^1-norm: $\|f_1\|_1 = 1$ and $\|f_2\|_1 = \sqrt{N}$. We observe that f_1, which is a sparse signal, has a lower L^1-norm than f_2, which is not sparse at all. In fact, this is the consequence of a more general concept. The lower the order of a norm is, the more its minimization will increase the sparsity.

Let's apply Theorem 12 with $p = 1$. We obtain:

$$\|A_g f\|_1 \geqslant N \|f\|_2 \|g\|_2.$$

Moreover, it is trivial to get:

$$\|A_g f\|_\infty = \max_{u,k} |\langle g, M_k T_u f \rangle| \leqslant \|g\|_2 \|M_k T_u f\|_2 = \|f\|_2 \|g\|_2.$$

We can express the ratio of the two last expression to get an uncertainty principle:

$$\frac{\|A_g f\|_1}{\|A_g f\|_\infty} \geqslant N. \tag{26}$$

6 Graph Inequalities

Next to the classical study, in this section, we shall demonstrate analogous inequalities for graphs. Those inequalities allow a better understanding of the Fourier transform. They also permit us to bound some operators like the shift.

We first need to define some constants. We will use the mixed norm of a matrix, which is defined as:

$$\|H\|_{p,q} = \left[\sum_i \left(\sum_j |h_{i,j}|^p\right)^{q/p}\right]^{1/q} = \max_{f\in\mathbb{R}^N} \frac{\|Hf\|_q}{\|f\|_p}$$

The coherence between the Dirac and the Fourier bases is:

$$\mu := \|\Phi\|_{\infty\infty} = \max_{\substack{\ell\in\{0,1,\dots N-1\},\\ n\in\{1,2,\dots N\}}} |\chi_\ell(n)|.$$

We also define:

$$\mu_k = \max_{n\in\{1,2,\dots N\}} |\chi_\ell(n)|$$

and

$$\tilde{\mu}_i = \max_{\ell\in\{0,1,\dots N-1\}} |\chi_\ell(n)|.$$

6.1 The Hausdorff-Young inequality

As signals on graphs are vectors, the definition of the p-norm of f is $\|f\|_p = (\sum_{n=1}^N |f(n)|^p)^{\frac{1}{p}}$. The following proof is an extension of the classical proof using the Riez-Thorin interpolation theorem (p. 174 [9]).

Theorem 13. *For $f \in \mathbb{R}^N$ a graph signal, $1 \leqslant p \leqslant 2$ and μ the coherence of the Fourier and the Dirac bases, we have*

$$\|\hat{f}\|_q \leqslant \mu^{1-\frac{2}{q}} \|f\|_p,$$

for

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Proof. First, using the Parseval identity (true for signals on graphs [2]), we have

$$\|f\|_2^2 = \|\hat{f}\|_2^2.$$

which implies

$$\|\hat{f}\|_2 \leqslant \|f\|_2. \tag{27}$$

Secondly, as all eigenvectors are normalized, $|\chi_\ell(n)| \leqslant \mu$ for all n, ℓ and thus

$$\|\hat{f}\|_\infty = \max_\ell |\hat{f}(\ell)| = \max_\ell \left|\sum_{n=1}^N \chi_\ell^*(n) f(n)\right| \leqslant \mu \sum_{n=1}^N |f(n)| = \mu\|f\|_1 \tag{28}$$

Thirdly, the graph Fourier transform $\hat{f} = \Phi f$, is a linear operator. Moreover, it is bounded from L^2 to L^2 (27) and from L^1 to L^∞ (28). Applying the Riez-Thorin theorem with $p_1 = 2$, $p_2 = 2$, $q_1 = 1$, $q_2 = \infty$, $M_p = 1$, $M_q = \mu$ leads to the desired result:

$$\|\hat{f}\|_q \leqslant \mu^{1-\frac{2}{q}} \|f\|_p,$$

for $\frac{1}{p} + \frac{1}{q} = 1$. □

Theorem 14. *For $f \in \mathbb{R}^N$ a graph signal, $2 \leqslant p \leqslant \infty$, we have*

$$\|\hat{f}\|_q \geqslant \mu^{\frac{2}{q}-1}\|f\|_p,$$

for

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Proof. Firstly, using the Parseval identity (true for signals on graphs [2]), we have

$$\|f\|_2^2 = \|\hat{f}\|_2^2.$$

which implies

$$\|\hat{f}\|_2 \geqslant \|f\|_2. \tag{29}$$

Secondly, as all eigenvectors are normalized, $|\chi_\ell(n)| \leqslant \mu$ for all n, ℓ and thus

$$\frac{1}{\mu}\|f\|_\infty = \frac{1}{\mu}\max_n |f(n)| = \frac{1}{\mu}\max_n \left|\sum_{\ell=0}^{N-1} \chi_\ell(n)\hat{f}(\ell)\right| \leqslant \sum_{\ell=0}^{N-1} |\hat{f}(\ell)| = \|\hat{f}\|_1 \tag{30}$$

Thirdly, the inverse graph Fourier transform $f = \Phi^t\hat{f}$, is a linear operator. Moreover, it is bounded from L^2 to L^2 (29) and from L^1 to L^∞ (30). Applying the Riez-Thorin theorem with $p_1 = 2$, $p_2 = 2$, $q_1 = 1$, $q_2 = \infty$, $M_p = 1$, $M_q = \mu$ leads to the desired result:

$$\|\hat{f}\|_q \geqslant \mu^{\frac{2}{q}-1}\|f\|_p,$$

for $\frac{1}{p} + \frac{1}{q} = 1$. □

For the DFT, the coherence between the Dirac and Fourier bases is minimal and equal to $\frac{1}{\sqrt{N}}$. Hence, the classical discrete Hausdorff-Young inequality given in Theorem 7 becomes a particular case of Theorems 13 and 14.

6.2 Graph Fourier bound

As the Fourier transform of a signal is a projection, we are able to bound it. For $p \geqslant 1$, we define the constant C_p as:

$$C_p := N^{|\frac{1}{2}-\frac{1}{p}|}.$$

Remark: $C_p \in [1, \sqrt{N}]$, $C_1 = C_\infty = \sqrt{N}$ and $C_2 = 1$.

Lemma 2. *For $f \in \mathbb{R}^N$ and $2 \leqslant p \leqslant \infty$, we have*

$$\|f\|_2 \leqslant C_p\|f\|_p$$

Proof. The sup-norm satisfies $\|p\|_\infty \leqslant N^{\frac{1}{2}}\|p\|_2$. We can apply the Riez-Thorin theorem to the identity operator. It is bounded from L^2 to L^2 by 1 and from L^2 to L^∞ by $N^{\frac{1}{2}}$. Doing so leads to the desired result. □

Lemma 3. *Conversely, for $f \in \mathbb{R}^N$ and $1 \leqslant p \leqslant 2$, we have*

$$\|f\|_2 \geqslant C_p\|f\|_p$$

Proof. The 1-norm satisfies $\|p\|_1 \geqslant N^{\frac{1}{2}}\|p\|_2$. We can apply the Riez-Thorin theorem to the identity operator. It is bounded from L^2 to L^2 by 1, and from L^2 to L^1 by $N^{\frac{1}{2}}$. Doing so leads to the desired result. □

Theorem 15. *For $f \in \mathbb{R}$, $C_p = N^{|\frac{1}{2}-\frac{1}{p}|}$ and $p \geqslant 1$,*

$$\|f\|_p \leqslant C_p \|\hat{f}\|_p$$

and conversely

$$\|\hat{f}\|_p \leqslant C_p \|f\|_p.$$

Proof. Suppose $p > 2$, then

$$\begin{aligned} \|f\|_p &\leqslant \|f\|_2 \\ &= \|\hat{f}\|_2 \qquad (31) \\ &\leqslant C_p \|\hat{f}\|_p, \qquad (32) \end{aligned}$$

where (31) follows from the Parseval identity and (32) from Lemma 2. Similarly, for $1 \leqslant p \leqslant 2$, we have

$$\begin{aligned} \|f\|_p &\leqslant C_p \|f\|_2 \\ &= C_p \|\hat{f}\|_2 \\ &\leqslant C_p \|\hat{f}\|_p. \end{aligned}$$

The converse follows from a similar proof. □

Theorem 15 yields:

$$\frac{1}{C_p}\|f\|_p \leqslant \|\hat{f}\|_p \leqslant C_p \|f\|_p.$$

Remark: This theorem gives us a first loose bound on the 1-norm of the convolution product.

Corollary 2. *For two signals $f, g \in \mathbb{R}^N$, we have:*

$$\|f * g\|_1 \leqslant C_1 C_p C_q \|f\|_p \|g\|_q$$

where $\frac{1}{p} + \frac{1}{q} = 1$.

Proof.

$$\begin{aligned} \|f * g\|_1 &\leqslant C_1 \|\hat{f} \cdot \hat{g}\|_1 \qquad (33) \\ &\leqslant C_1 \|\hat{f}\|_p \|\hat{g}\|_q \qquad (34) \\ &\leqslant C_1 C_p C_q \|f\|_p \|g\|_q \qquad (35) \end{aligned}$$

where $\frac{1}{p} + \frac{1}{q} = 1$. Equation (34) is obtained by applying Hölder's inequality; (33) and (35) are the results of Theorem 15. □

Using the same proof with a generalized version of Holder's Inequality, we obtain a loose bound on the r-norm of the convolution product.

Corollary 3. *For two signals $f, g \in \mathbb{R}^N$, we have:*

$$\|f * g\|_r \leqslant C_r C_p C_q \|f\|_p \|g\|_q$$

where $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$.

6.3 Young's Fourier inequality

We already found a Young's like inequality in the previous section. However, the bound seems to be quite loose. In order to find a sharper inequality, we will now bound the convolution product with the norm in the Fourier domain.

The structure is, once again, similar to the traditional Young inequality.

Lemma 4. *The convolution product satisfies* $\|f * g\|_1 \leqslant \sqrt{N}\|\hat{f}\|_p\|\hat{g}\|_q$ *where* $\frac{1}{p} + \frac{1}{q} = 1$

Proof.

$$\begin{aligned}
\|f * g\|_1 &= \sum_{n=1}^{N} \left| \sum_{\ell=0}^{N-1} \hat{f}(\ell)\hat{g}(\ell)\chi_\ell(n) \right| \\
&\leqslant \sum_{\ell=0}^{N-1} |\hat{f}(\ell)\hat{g}(\ell)| \sum_{n=1}^{N} |\chi_\ell(n)| \\
&\leqslant \sqrt{N} \sum_{\ell=0}^{N-1} |\hat{f}(\ell)\hat{g}(\ell)| \\
&\leqslant \sqrt{N}\|\hat{f}\|_p\|\hat{g}\|_q \qquad (36)
\end{aligned}$$

Equation (36) follows from Hölder's inequality.

□

Even if this inequality seems to be loose, it is not. Experimentally, we have checked that in a particular case $\|f * g\|_1$ grows in the same way as the bound (See Figure 12). This means, we will not be able to do much better.

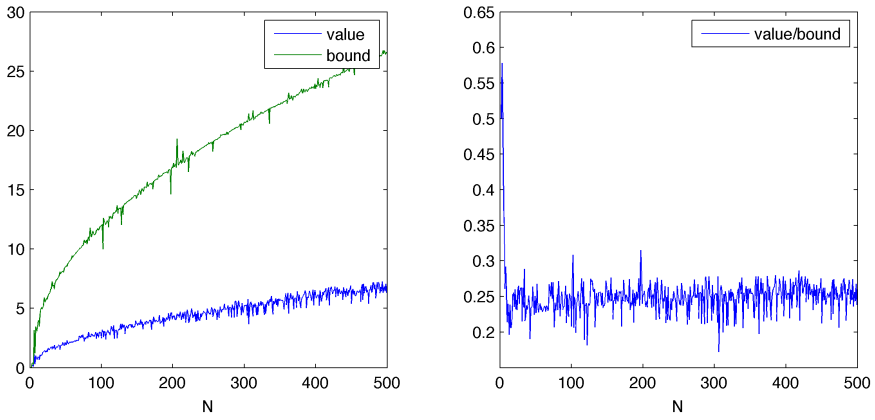

Figure 12: On the left the 1-norm of the convolution product $\|f * g\|_1$ is shown with is computed bound $\sqrt{N}\|\hat{f}\|_1\|\hat{g}\|_\infty$. The two signals f, g are discrete normalised Gaussians with $\sigma_f = 5, \sigma_g = 3, \mu_f = 4, \mu_g = 10$. The x-axis is composed of N linearly spaced values between -50 and 50. The graph is a ring of N points. On the right, we observe the ratio between the convolution product and the bound. It seems to be independent from N.

Remark: In particular, we have

$$\|f * g\|_1 \leqslant \sqrt{N}\|\hat{f}\|_\infty\|\hat{g}\|_1 \leqslant \sqrt{N}\|\hat{f}\|_1\|\hat{g}\|_1. \qquad (37)$$

Lemma 5. *The convolution product satisfies* $\|f * g\|_\infty \leqslant \mu\|\hat{f}\|_p\|\hat{g}\|_q$ *where* $\frac{1}{p} + \frac{1}{q} = 1$.

Proof.

$$\begin{aligned} \|f * g\|_\infty &= \max_n \left| \sum_{\ell=0}^{N-1} \hat{f}(\ell)\hat{g}(\ell)\chi_\ell(n) \right| \\ &\leqslant \mu \sum_{\ell=0}^{N-1} \left| \hat{f}(\ell)\hat{g}(\ell) \right| \\ &\leqslant \mu \|\hat{f}\|_p \|\hat{g}\|_q \end{aligned} \tag{38}$$

Equation (38) follows from Hölder's inequality. □

Remark: In particular, we have:

$$\|f * g\|_\infty \leqslant \mu \|\hat{f}\|_\infty \|\hat{g}\|_1. \tag{39}$$

Remark: If we choose $p = q = 2$, we get the following results:

$$\|f * g\|_\infty \leqslant \mu \|\hat{f}\|_2 \|\hat{g}\|_2 = \mu \|f\|_2 \|g\|_2. \tag{40}$$

Note that in the classical case, it is not possible to obtain such small bounds. The classical Young's inequality leads to:

$$\|f * g\|_\infty \leqslant \|f\|_2 \|g\|_2. \tag{41}$$

For special graphs, μ is proportional to $\frac{1}{\sqrt{N}}$. For instance, ring graphs have a coherence of $\sqrt{\frac{1}{N}}$ (DFT) or $\sqrt{\frac{2}{N}}$ (RDFT) depending the choice of the eigenvectors. This means that the convolution on graphs can act quite differently than traditional convolution. Moreover, for the ring configuration, where the eigenvectors are discrete cosines [5] and which seems to be the closer to the classical case, the convolution acts differently for the graphs than for the classical case. An illustration is given in Figure 13. In fact, this due to the definition of the convolution. To be consistent with the definition of the discrete circular convolution, we need to multiply the graph convolution product by $\sqrt{N}$.

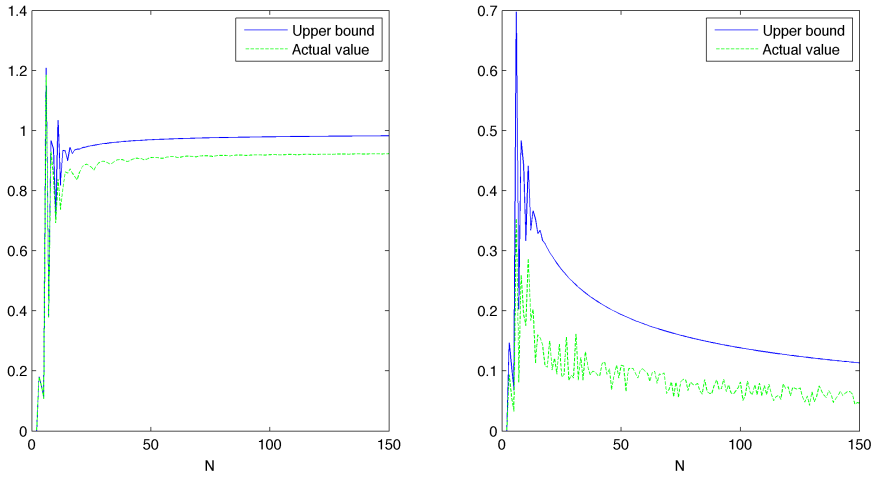

Figure 13: This figure shows the sup-norm of the convolution product (traditional on the left and graph(RDFT) on the right) of two discrete normalized Gaussians ($\sigma_f = 5, \sigma_g = 3, \mu_f = 4, \mu_g = 10$). The x-axis is composed of N linearly spaced values between -50 and 50 included. The graph is a ring of N points. The ploted bounds are given in equation (40) and (41).

Theorem 16. *For $f, g \in \mathbb{R}^N$ two graph signals and $p, q, r \in \mathbb{R}$ such that $1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, we have*

$$\|f * g\|_r \leqslant N^{\frac{1}{2r}} \mu^{p-\frac{1}{r}} \|\hat{f}\|_p \|\hat{g}\|_q. \tag{42}$$

Proof. For a graph signal $f \in \mathbb{R}^N$, we define an operator $T_g : \mathbb{R}^N \to \mathbb{R}^N$ by $(T_g \hat{f})(n) = (f * g)(n)$. Using Lemma 37 and (39), we observe that this operator is bounded from L^1 to L^1 by $\|\hat{g}\|_1 \sqrt{N}$ and from L^∞ to L^∞ by $\|\hat{g}\|_1 \mu$. Thus, we can apply the Riez-Thorin theorem to this operator and we get:

$$\|f * g\|_p \leqslant \mu^{\frac{1}{p'}} N^{\frac{1}{2p}} \|\hat{f}\|_p \|\hat{g}\|_1.$$

Then, for a graph signal $g \in \mathbb{R}^N$, we define another operator $T_f : \mathbb{R}^N \to \mathbb{R}^N$ by $(T_f \hat{g})(n) = (f * g)(n)$. With the previous inequality and Lemma 5, we observe that this new operator is bounded from L^1 to L^p by $\mu^{\frac{1}{p'}} N^{\frac{1}{2p}} \|\hat{f}\|_p$ and from $L^{p'}$ to L^∞ by $\mu \|\hat{f}\|_p$, where $\frac{1}{p} + \frac{1}{p'} = 1$. Again the Riez-Thorin theorem leads to the desired result.

$$\|f * g\|_r \leqslant N^{\frac{1}{2r}} \mu^{p-\frac{1}{r}} \|\hat{f}\|_p \|\hat{g}\|_q,$$

where $1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. □

This previous theorem yields an inequality similar to Young's; however, we can obtain a sharper inequality using Lemma 4 instead of (37).

Theorem 17. *For $f, g \in \mathbb{R}^N$ two graph signals and $p, q, r, r' \in \mathbb{R}$ such that $1 = \frac{1}{p} + \frac{1}{q}$ and $1 = \frac{1}{r} + \frac{1}{r'}$, we have*

$$\|f * g\|_r \leqslant N^{\frac{1}{2r}} \mu^{\frac{1}{r'}} \|\hat{f}\|_p \|\hat{g}\|_q. \tag{43}$$

Proof. The inequality follows directly from the application of the Riez-Thorin theorem to Lemmas 4 and 5. □

6.4 The new inequality

The previous Young inequality was in the Fourier domain. In order to get a bound in the vertex domain, we will use the Parseval equality.

Using Lemma 4, we find:

$$\|f * g\|_1 \leqslant \sqrt{N} \|\hat{f}\|_2 \|\hat{g}\|_2 = \sqrt{N} \|f\|_2 \|g\|_2. \tag{44}$$

Using Lemma 5, we have:

$$\|f * g\|_\infty \leqslant \mu \|\hat{f}\|_2 \|\hat{g}\|_2 = \mu \|f\|_2 \|g\|_2 \tag{45}$$

Lemma 6. *The convolution product satisfies $\|f * g\|_2 \leqslant \|f\|_2 \|g\|_2$.*

Proof.

$$\begin{aligned} \|f * g\|_2 &= \|\widehat{f * g}\|_2 \\ &= \|\hat{f} \cdot \hat{g}\|_2 \\ &\leqslant \|\hat{f}\|_2 \|\hat{g}\|_2 \qquad (46) \\ &= \|f\|_2 \|g\|_2, \end{aligned}$$

where (46) follows from the Cauchy-Schwartz inequality. □

Lemma 7. *The convolution product satisfies $\|f * g\|_2 \leqslant \mu \|f\|_2 \|g\|_1$.*

Proof.

$$\begin{aligned} \|f * g\|_2 &= \|\widehat{f * g}\|_2 \\ &= \|\hat{f} \cdot \hat{g}\|_2 \\ &\leqslant \|\hat{f}\|_2 \|\hat{g}\|_\infty \\ &\leqslant \mu \|f\|_2 \|g\|_1, \qquad (47) \end{aligned}$$

where (47) follows from Theorem 13 (Hausdorff-Young). □

Theorem 18. *For $f, g \in \mathbb{R}^N$ two graph signals and $r \in \mathbb{R}$, we have*

$$\|f * g\|_r \leqslant N^{\frac{1}{r}-\frac{1}{2}} \|f\|_2 \|g\|_2 \tag{48}$$

for $r \in [1, 2]$, and

$$\|f * g\|_r \leqslant \mu^{1-\frac{2}{r}} \|f\|_2 \|g\|_2 \tag{49}$$

for $r \in [2, \infty]$.

Proof. For a graph signal $f \in \mathbb{R}^N$, we define an operator $T_g : \mathbb{R}^N \to \mathbb{R}^N$ by $(T_g f)(n) = (f * g)(n)$. Using (44) and Lemma 6, we observe that this operator is bounded from L^2 to L^1 by $\|g\|_2\sqrt{N}$ and from L^2 to L^2 by $\|g\|_2$. Thus, we can apply the Riez-Thorin theorem to this operator and we get (48).

Then, for a graph signal $g \in \mathbb{R}^N$, we define another operator $T_f : \mathbb{R}^N \to \mathbb{R}^N$ by $(T_f g)(n) = (f * g)(n)$. With Lemma 6 and (45), we observe that this new operator is bounded from L^2 to L^2 by $\|g\|_2$ and from L^2 to L^∞ by $\mu\|f\|_2$. Again the Riez-Thorin theorem leads to the desired result (49). □

Remark: If $\mu < N^{-\frac{1}{4}}$ (ring graph for instance), the following inequality is sharper than Theorem 18.

$$\|f * g\|_r \leqslant N^{\frac{1}{2r}} \mu^{\frac{1}{r'}} \|f\|_2 \|g\|_2.$$

It is the result of the application of the Parseval equality to Theorem 17. In order to get a bound on the p-norm, we can apply the following Lemma. Using Lemma 2 and Theorem 18, we have the following results:

	$1 \leqslant p \leqslant 2$	$p > 2$
$1 \leqslant r \leqslant 2$	$\|f * g\|_r \leqslant N^{\frac{1}{r}-\frac{1}{2}} \|f\|_p \|g\|_2$	$\|f * g\|_r \leqslant N^{\frac{1}{r}-\frac{1}{p}} \|f\|_p \|g\|_2$
$r > 2$	$\|f * g\|_r \leqslant \mu^{1-\frac{2}{r}} \|f\|_p \|g\|_2$	$\|f * g\|_r \leqslant N^{\frac{1}{2}-\frac{1}{p}} \mu^{1-\frac{2}{r}} \|f\|_p \|g\|_2$

6.5 Consequences of the inequalities

Those inequalities can be used to bound the translation operator for instance. Applying Lemma 7 leads to:

$$\|T_i\|_2 = \sup_{g \in \mathbb{R}^N} \frac{\|T_i g\|_2}{\|g\|_2} = \sup_{g \in \mathbb{R}^N} \frac{\sqrt{N}\|g * \delta_i\|_2}{\|g\|_2} \leqslant \sqrt{N}\mu$$

Furthermore translating the neutral element $g_0(n) = \sum_{\ell=0}^{N-1} \chi_\ell(n)$ gives a lower bound:

$$\|T_i\|_2 = \sup_{g \in \mathbb{R}^N} \frac{\|T_i g\|_2}{\|g\|_2} = \sup_{g \in \mathbb{R}^N} \frac{\sqrt{N}\|g * \delta_i\|_2}{\|g\|_2} \geqslant \frac{\sqrt{N}\|g_0 * \delta_i\|_2}{\|g_0\|_2} = \frac{\sqrt{N}\|\delta_i\|_2}{\sqrt{N}} = 1.$$

Thus we have:

$$1 \leqslant \|T_i\|_2 \leqslant \sqrt{N}\mu \tag{50}$$

Similarly, for the modulation, we have:

$$\|M_k\|_2 = \sup_{g \in \mathbb{R}^N} \frac{\|M_k g\|_2}{\|g\|_2} = \sup_{g \in \mathbb{R}^N} \frac{\sqrt{N}\|g \cdot \chi_k\|_2}{\|g\|_2} \leqslant \sqrt{N}\mu.$$

For a special $g_s(n) = 1$, the bound becomes

$$\|M_k\|_2 = \sup_{g \in \mathbb{R}^N} \frac{\|M_k g\|_2}{\|g\|_2} = \sup_{g \in \mathbb{R}^N} \frac{\sqrt{N}\|g \cdot \chi_k\|_2}{\|g\|_2} \geqslant \frac{\sqrt{N}\|g_s \cdot \chi_k\|_2}{\|g_s\|_2} = \frac{\sqrt{N}\|\chi_k\|_2}{\sqrt{N}} = 1$$

and

$$1 \leqslant \|M_k\|_2 \leqslant \sqrt{N}\mu. \tag{51}$$

We observe that (50) and (51) become tight, when $\mu = \frac{1}{\sqrt{N}}$ which is the case of the DFT.

6.6 Summary Tables

6.6.1 Inequalities

Name	Classical	Graph
Parseval equality	$\langle f, g\rangle = \langle \hat{f}, \hat{g}\rangle$	$\langle f, g\rangle = \langle \hat{f}, \hat{g}\rangle$
Hausdorff-Young inequality	$\|\hat{f}\|_{p'} \leqslant D_p \|f\|_p$, $1 \leqslant p \leqslant 2$	$\|\hat{f}\|_{p'} \leqslant \mu^{1-\frac{2}{p'}} \|f\|_p$, $1 \leqslant p \leqslant 2$
Hausdorff-Young discrete	$\|\hat{f}\|_{p'} \leqslant N^{\frac{1}{p'}-\frac{1}{2}} \|f\|_p, 1 \leqslant p \leqslant 2$	
Hausdorff-Young converse	$\|\hat{f}\|_{p'} \geqslant D_p \|f\|_p$, $p \geqslant 2$	$\|\hat{f}\|_q \geqslant \mu^{\frac{2}{p'}-1} \|f\|_p, p \geqslant 2$
Hausdorff-Young discrete 2	$\|\hat{f}\|_{p'} \geqslant N^{\frac{1}{2}-\frac{1}{p'}} \|f\|_p, p \geqslant 2$	
Young's inequality	$\|f * g\|_r \leqslant (D_p D_q D_{r'}) \|f\|_p \|g\|_q$	$\|f * g\|_r \leqslant \mu N \|f\|_p \|g\|_q$
		$\|f * g\|_r \leqslant N^{\frac{1}{2r}} \mu^{p-\frac{1}{r}} \|\hat{f}\|_p \|\hat{g}\|_q$
Fourier bound		$\|\hat{f}\|_p \leqslant C_p \|f\|_p$
Lemma 4	$\|f * g\|_1 \leqslant \|f\|_1 \|g\|_1$	$\|f * g\|_1 \leqslant \sqrt{N} \|\hat{f}\|_p \|\hat{g}\|_{p'}$
Lemma 5	$\|f * g\|_\infty \leqslant \|f\|_p \|g\|_{p'}$	$\|f * g\|_\infty \leqslant \mu \|\hat{f}\|_p \|\hat{g}\|_{p'}$

In the table, $p, p', q, r > 0$ satisfy $\frac{1}{p} + \frac{1}{p'} = 1$ and $1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. All of the classical inequalities are sharp.

Constant name	Value	Range
D_p	$\left(\frac{p^{\frac{1}{p}}}{p'^{\frac{1}{p'}}}\right)^{\frac{1}{2}}$	$[0, 1]$
$H(p, a, b)^2$	$abp^{-2}\|p-2\|^{2-p}\|p-a\|^{-1+\frac{p}{a}}\|p-b\|^{-1+\frac{p}{b}}$	$[0, 1]$
$K(p, a, b)^2$	$p^{-2} 2^{2-p} a^{\frac{p}{a}} b^{\frac{p}{b}}$	$[0, 1]$
μ	$\max_{n,\ell}\{\|\chi_\ell(n)\|\}$	$[\frac{1}{\sqrt{N}}, 1]$
C_p	$N^{\|\frac{1}{2}-\frac{1}{p}\|}$	$[1, \sqrt{N}]$

Here are some graph inequalities without classical analogy:

Inequality	Remark
$\|f\|_p \leqslant C_p \|\hat{f}\|_p$	$p \geqslant 1$
$\|\hat{f}\|_p \leqslant C_p \|f\|_p$	$p \geqslant 1$
$\|f * g\|_r \leqslant C_r C_p C_q \|f\|_p \|g\|_q$	$\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$
$\|f * g\|_2 \leqslant \|f\|_2 \|g\|_2$	
$\|f * g\|_1 \leqslant \sqrt{N} \|f\|_2 \|g\|_2$	
$\|f * g\|_\infty \leqslant \mu \|f\|_2 \|g\|_2$	
$\|f * g\|_r \leqslant N^{\frac{1}{2r}} \mu^{\frac{1}{r'}} \|f\|_2 \|g\|_2$	$\frac{1}{r} + \frac{1}{r'} = 1$

		$1 \leqslant p \leqslant 2$	$p > 2$
$1 \leqslant r \leqslant 2$	$\|f * g\|_r \leqslant N^{\frac{1}{r}-\frac{1}{2}} \|f\|_2 \|g\|_2$	$\|f * g\|_r \leqslant N^{\frac{1}{r}-\frac{1}{2}} \|f\|_p \|g\|_2$	$\|f * g\|_r \leqslant N^{\frac{1}{r}-\frac{1}{p}} \|f\|_p \|g\|_2$
$r > 2$	$\|f * g\|_r \leqslant \mu^{1-\frac{2}{r}} \|f\|_2 \|g\|_2$	$\|f * g\|_r \leqslant \mu^{1-\frac{2}{r}} \|f\|_p \|g\|_2$	$\|f * g\|_r \leqslant N^{\frac{1}{2}-\frac{1}{p}} \mu^{1-\frac{2}{r}} \|f\|_p \|g\|_2$

And bound on operators:

$$1 \leqslant \|T_i\|_2 \leqslant \sqrt{N}\mu$$

$$1 \leqslant \|M_k\|_2 \leqslant \sqrt{N}\mu$$

6.6.2 Uncertainty principles

Name	Classical	Graph
Lieb theorem $1 \leqslant p \leqslant 2$	$\|A_g f\|_p \geqslant H(p,a,a')^{\frac{1}{p}} \|g\|_a \|f\|_{a'}$	$\|A_g f\|_p \geqslant N\mu^{1-\frac{2}{q}} (\min_s \|T_s g\|_q) \|f\|_q$
Lieb discrete $1 \leqslant p \leqslant 2$	$\|A_g f\|_p \geqslant N^{1-\frac{1}{p'}} \|f\|_a \|g\|_{a'}$	
Lieb theorem $p \geqslant 2$	$\|A_g f\|_p \geqslant K(p,a,a')^{\frac{1}{p}} \|g\|_a \|f\|_{a'}$	
Lieb discrete $p \geqslant 2$	$\|A_g f\|_p \leqslant N^{1-\frac{1}{p'}} \|f\|_a \|g\|_{a'}$	
One norm bound	$\|A_g f\|_1 \geqslant N\|f\|_2 \|\|g\|_2$	$\|A_g f\|_1 \geqslant N\|\hat{g}(0)\| \|\hat{f}\|_1$
Sup-norm bound	$\|A_g f\|_\infty \leqslant \|f\|_2 \|\|g\|_2$	$\|A_g f\|_\infty \leqslant \mu^2 N\|f\|_2 \|\|g\|_2$
Uncetainty principle	$\frac{\|A_g f\|_1}{\|A_g f\|_\infty} \geqslant N$	$\frac{\|A_g f\|_1}{\|A_g f\|_\infty} \geqslant \frac{1}{\mu^2}$, for decreasing kernel.

A decreasing kernel satisfies $|\hat{g}(0)| \geqslant |\hat{g}(l)| \geqslant 0$ for $l = 1, 2, ...N-1$.

7 Conclusion

This work first focused on a generalized convolution product for graphs. After a general exploration of its principal properties, we demonstrated some Young-like theorems. A Hausdorff-Young's like inequality was also demonstrated. We then defined a modulation (multiplication with an eigenvector) and a translation (convolution with a Dirac) operator. With these, we were able to construct a windowed graph Fourier transform (WGFT) which allows simultaneous analysis of a signal on graphs in the vertex and spectral domains. This construction was already made [3], but we suggested normalizing the atoms in order to obtain energy consistency in the transform. This led to the normalized windowed graph Fourier transform (NWGFT). Afterwards, to better understand the WGFT and the NWGFT, we defined several ambiguity functions. Of these we observed that the way in which they spread in the spectral/vertex domains was correlated with the localization of the window in those same domains. Finally, we demonstrated some inequalities on the ambiguity function leading to an uncertainty principle useful for the WGFT:

$$\frac{\|A_g\|_1}{\|A_g\|_\infty} \geqslant \frac{1}{\mu^2}.$$

Unfortunately, even if this seems true for all kernels, it has only been demonstrated for those satisfying some conditions. However, the heat kernel, one natural choice for the window, satisfies these conditions. The most important consideration is that the smaller the coherence, the bigger the uncertainty. This is the logical consequence of the bases being closer if the coherence is bigger. In this case, we can be precise in both domains at the same time.

Much more work needs to be done on this topic. Theoretically, the uncertainty principle is true only for specific kernels (including the heat kernel). We observed that the 1-norm of the ambiguity function was always greater than N, but were able to demonstrate this only when the Fourier transform of the kernel was decreasing. Furthermore, the bound of the ambiguity function sup-norm becomes loose when the coherence increases. This phenomenon was observed but is still not well understood. Studying it could lead to a sharper bound and thus a better uncertainty principle. Finally, the work has been carried out only for the non-normalized definitions. Doing the same for the normalized ambiguity function would give an uncertainty principle for the NWGFT. This research should be undertaken carefully since the NWGFT requires a hypothesis on eigenvectors to exist. On the practical side, finding kernels with suitable properties linked to the translation and modulation is still an open question. The heat kernel behaves well with translation as it is located after the application of the operator. Questions remain as to the existence of other windows with the same properties and kernels with similar behavior for the modulation. Fast algorithms that do not require an explicit calculation of the eigenvalues and of the eigenvectors exist for the translation operator [2]. Finding a similar technique for the modulation could lead to an efficient implementation of the WGFT.

The uncertainty principle obtained leads to a new question which is: How can we find the Fourier basis with the minimum coherence?

8 Thanks

Dans un premier temps, je tiens particulièrement à remercier David Shuman. Il s'attelé aux difficiles tâches d'essayer de m'apprendre à écrire un rapport en anglais, de structurer mes pensées, et, enfin, de me comprendre lorsque je n'étais pas clair. Il a su tout au long de ce travail m'aiguiller de façon à ce que je ne m'égare pas trop et il a investi un temps considérable à la correction de ce rapport. Dans un second temps, Pierre Vandergheysnt a lui aussi droit à ses quelques lignes personnelles. Il a eu l'audace, à raison je l'espère, de me lancer dans un sujet complexe mais fort intéressant qui, en fin de compte, m'apprit bien plus que ce qui se trouve dans ce rapport. Pierre fut capable, surement sans le savoir, d'apaiser chez moi cette maladie qui ronge tout chercheur : la peur de ne pas réussir à franchir les obstacles. C'est grâce à vous que je suis satisfait de conclure ce travail éternellement inachevé par cette petite note personnelle dans une langue que je chéris beaucoup. A tout bientôt !

9 Appendix

9.1 The heat kernel

The shifted heat kernel is well localized in the vertex domain. Moreover, the unshifted kernel is well localized in the spectral domain. As a consequence, it seems to be a good kernel for the WGFT or the NWFT. We will demonstrate here some of its properties.

The heat kernel is defined in the vertex domain by:

$$\hat{g}(\ell) = Ce^{-\lambda_\ell \tau}$$

where C is a normalization constant. In vertex domain, it is not necessary localized (when not shifted). We will first prove that the shifted kernel is always positive in the vertex domain.

Theorem 19. *For a window $\hat{g}(\ell) = Ce^{-\lambda_\ell \tau}$ with $\tau \geqslant 0$ and $C > 0$,*

$$(T_i g)(n) \geqslant 0, \forall n \in \{1, 2, ..., N\}, i \in \{1, 2, ..., N\}$$

Proof. First, we notice that:

$$\begin{aligned}
(T_i g) &= \sum_{\ell=0}^{N-1} e^{-\lambda_\ell \tau} x_\ell^*(i) \chi_\ell \\
&= \sum_{\ell=0}^{N-1} \sum_{k=0}^{\infty} \frac{(-\lambda_\ell \tau)^k}{k!} x_\ell^*(i) \chi_\ell \\
&= \sum_{k=0}^{\infty} \sum_{\ell=0}^{N-1} \frac{(-\lambda_\ell \tau)^k}{k!} x_\ell^*(i) \chi_\ell \\
&= \sum_{k=0}^{\infty} \frac{(-\mathcal{L}\tau)^k}{k!} \delta_i \qquad (52) \\
&= e^{-\mathcal{L}\tau} \delta_i,
\end{aligned}$$

where (52) follows from (6). Yet as $\mathcal{L} = D - W$ and $DW = WD$, we have

$$e^{-\mathcal{L}\tau} = e^{-(D-W)\tau} = e^{-D\tau} e^{W\tau}$$

All elements in $e^{-D\tau}$ are positive because $-D\tau$ is a diagonal matrix. All element in $e^{W\tau}$ are positive because $W\tau$ has only positive or null entry. As a consequence $e^{-\mathcal{L}\tau}$ has only positive element and $(T_i g)(n) \geqslant 0$. □

9.2 A bound for the p-norm of the ambiguity function

Theorem 20. *For $g, f \in \mathbb{R}^N$ and $1 \leqslant p \leqslant 2$ we have*

$$\|A_g f\|_p = \sum_{i=1}^{N} \left(\sum_{k=0}^{N-1} |A_g(i,k)|^p \right)^{\frac{1}{p}} \geqslant \sqrt{N} \mu^{\frac{2}{q}-1} (\min_s \|T_s g\|_q) \|f\|_q,$$

where $\frac{1}{p} + \frac{1}{q} = 1$.

Proof.

$$\begin{aligned}
\|A_g f\|_p^p &= \sum_{i=1}^{N}\sum_{k=0}^{N-1} |A_g(i,k)|^p \\
&= \sum_{i=1}^{N}\sum_{k=0}^{N-1} |\langle f, M_k T_i g\rangle|^p \\
&= \sqrt{N^p}\sum_{i=1}^{N}\sum_{k=0}^{N-1} |\mathcal{F}(f\cdot T_i^* g)(k)|^p \\
&= \sqrt{N^p}\sum_{i=1}^{N} \|\mathcal{F}(f\cdot T_i^* g)\|_p^p \\
&\geqslant \sqrt{N^p}\mu^{\frac{2p}{q}-p}\sum_{i=1}^{N} \|f\cdot T_i^* g\|_q^p \qquad (53)\\
&= \sqrt{N^p}\mu^{\frac{2p}{q}-p}\sum_{i=1}^{N}\left(\sum_{n=1}^{N} |f(n)T_i^* g(n)|^q\right)^{\frac{p}{q}} \\
&\geqslant \sqrt{N^p}\mu^{\frac{2p}{q}-p}\left(\sum_{i=1}^{N}\sum_{n=1}^{N} |f(n)T_i^* g(n)|^q\right)^{\frac{p}{q}} \qquad (54)\\
&= \sqrt{N^p}\mu^{\frac{2p}{q}-p}\left(\sum_{i=1}^{N}\sum_{n=1}^{N} |f(n)|^q\,|T_n^* g(i)|^q\right)^{\frac{p}{q}} \\
&= \sqrt{N^p}\mu^{\frac{2p}{q}-p}\left(\sum_{n=1}^{N} |f(n)|^q\,\|T_n g\|_q^q\right)^{\frac{p}{q}} \\
&\geqslant \sqrt{N^p}\mu^{\frac{2p}{q}-p}\left(\min_s \|T_s g\|_q^q\right)^{\frac{p}{q}}\left(\sum_{n=1}^{N} |f(n)|^q\right)^{\frac{p}{q}} \\
&= \sqrt{N^p}\mu^{\frac{2p}{q}-p}(\min_s \|T_s g\|_q)^p \|f\|_q^p. \qquad (55)
\end{aligned}$$

Equation (54) follows from $\frac{p}{q} \leqslant 1$, and (53) is the application of the converse of the Hausdorff-Young Inequality given by Corollary 14. Taking the p-root of (55) leads to the desired result.

□

This last theorem suggests an uncertainty principle. Indeed, if the q-norm of f is fixed, the p-norm of the ambiguity function $A_g f$ cannot be arbitrarily small. However, to be sure of that, we have to show that $(\min_s \|T_s g\|_p)$ is always above zero.

Theorem 21. *For $g \in \mathbb{R}^N$, such that $\exists\ \ell \mid \hat{g}(\ell)\chi_\ell(s) \neq 0$ and $0 < p < \infty$, we have $\|T_s g\|_p > 0$.*

Proof. As the Fourier transform is an orthonormal projection, $\|g\|_p > 0$ implies $\|\hat{g}\|_p > 0$.

Let's suppose that $\|T_s g\|_p = 0$. We then have

$$\begin{aligned}
\|T_s g\|_p^p &= \sqrt{N}\sum_{n=1}^{N}\left|\sum_{\ell=0}^{N-1} \hat{g}(\ell)\chi_\ell^*(s)\chi_\ell(n)\right|^p \\
&= \sqrt{N}\sum_{n=1}^{N}\left|\langle \hat{g}(\cdot)\chi_{(\cdot)}^*(s), \chi_{(\cdot)}^*(n)\rangle\right|^p = 0.
\end{aligned}$$

This implies that $n = 1, 2, ..., N$, $\hat{g}(\cdot)\chi_{(\cdot)}^*(s) \perp \chi_{(\cdot)}^*(n)$. As $\chi_{(\cdot)}^*(n)$ forms a complete orthonormal basis, this is true only if $\hat{g}(\cdot)\chi_{(\cdot)}^*(s) = 0$ which is in contradiction with the hypothesis $\exists\ \ell \mid \hat{g}(\ell)\chi_\ell(s) \neq 0$. □

This last theorem gives us a condition on the window g in order to have a lower bound greater than zero on the ambiguity function. To satisfy this condition we can choose a window where $\hat{g}(0) \neq 0$.

9.3 Estimate of $|\hat{g}(0)|$

Let's suppose that we use a special kernel $\hat{k}(\lambda_l)$ (defined in the graph Fourier domain) for the ambiguity function. The window is defined as $\hat{g}(\lambda_l) = C\hat{k}(\lambda_l)$ with $C \in \mathbb{R}^N$ such that $\|\hat{g}\|_2 = 1$. Knowing the eigenvalues of the graph, $\hat{g}(0)$ becomes computable with:

$$\hat{g}(\lambda_0) = C\hat{k}(\lambda_0) = \frac{\hat{k}(\lambda_0)}{\|\hat{k}\|_2} = \frac{\hat{k}(\lambda_0)}{\left(\sum_{\ell=0}^{N-1} |k(\lambda_\ell)|^2\right)^{\frac{1}{2}}}.$$

For instance, the heat-kernel ($\hat{k}(\lambda_\ell) = e^{-\lambda_\ell \tau}$) produces a $\hat{g}(0) \in [\frac{1}{\sqrt{(N)}}, 1]$ depending on the graph eigenvalues and on τ.

Unfortunately, the computation cost of the eigenvalue is large for large graphs. For special types of graphs, we know the probability density function of the non-zero eigenvalues: $p(\lambda_l)$. In that case, we can compute an estimation of $\hat{g}(0)$ without needing all of the eigenvalues.

To make writing easier, we will note $\Lambda = [\lambda_0, \lambda_1, \lambda_2, ..., \lambda_{N-1}]$. As the graph is connected, the vector Λ has $N-1$ degrees of freedom ($\lambda_0 = 0$). The expected norm of a kernel is defined as:

$$\mathbb{E}\|k\|_2 = \int_{\Lambda \in \mathbb{R}^N} p(\Lambda)\|k(\Lambda)\|_2 \mathrm{d}\Lambda = C^{-1} \tag{56}$$

Example: Suppose we are working on a random d-regular graph (d=degree). McKay [11] has proved that the spectral distribution of the adjacency matrix converges when $N \to \infty$ (In our example, we assume that this hypothesis is satisfied if $N >> d$) to:

$$f_d(x) = \begin{cases} \frac{d\sqrt{4(d-1)-x^2}}{2\pi(d^2-x^2)} & \text{if } -2\sqrt{d-1} \leqslant x \leqslant 2\sqrt{(d-1)}, \\ 0 & \text{otherwise.} \end{cases}$$

An illustration of this distribution is given in Figure 14.

As the graph is regular, the Laplacian and the weight matrix have the same eigenvectors.

$$\begin{aligned} \lambda_i \chi_i &= \mathcal{L}\chi_i \\ &= (D - W)\chi_i \\ &= D\chi_i - W\chi_i \\ &= d\chi_i - W\chi_i \\ &\Leftrightarrow \\ (d - \lambda_i)\chi_i &= W\chi_i \end{aligned}$$

If all weights are the same and equal to one, then the adjacency matrix is the same as the weight matrix. In that case, we know the probability density function of the Laplacian eigenvalues. We are able to evaluate (56) numerically. The distribution does not include the 0 eigenvalue. This is to be taken into account in the numerical calculation. The results are summarized in Table 3.

d\N	10		100		1000	
3	0.8994	*0.8007*	0.3552	*0.3465*	0.1169	*0.1161*
5	0.9966	*0.9694*	0.8002	*0.7712*	0.3601	*0.3566*
9	1	*0.9998*	0.9990	*0.9984*	0.9849	*0.9840*

Table 3: Value of $\hat{g}(0)$ for the heat kernel with $\tau = 1$. The value in italic is the estimation. We observe that the estimation is always lower than real value.

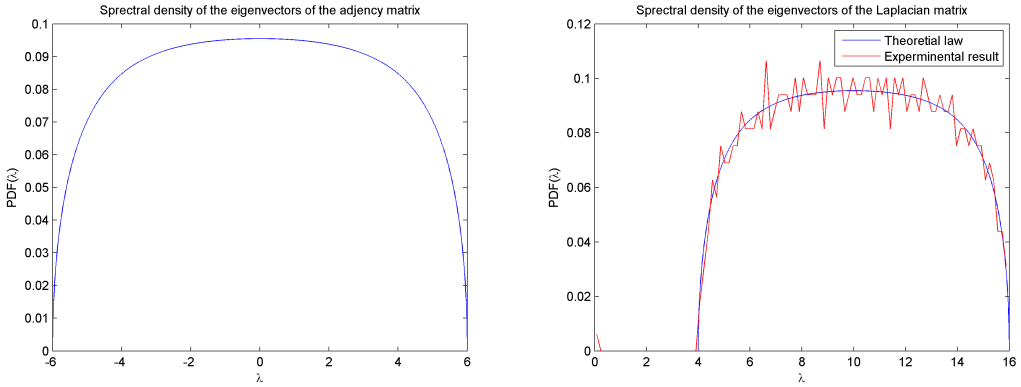

Figure 14: Left: McKay spectral distribution of the adjacency matrix for a 10-regular random graph. Right: spectral distribution of the Laplacian matrix. The experimental result is the histogram of the normalized eigenvalues in 100 bins for a graph with 1000 vertices.

9.4 Young's inequality

In order to prove Young's inequality for signals on graphs, we first need some lemmas.

Lemma 8. *The magnitude of the convolution product of two diracs is bounded by μ.*

$$|(\delta_m * \delta_u)(n)| = \left| \sum_{\ell=0}^{N-1} \chi_l^*(m)\chi_l^*(u)\chi_l(n) \right| \leqslant \mu_n \leqslant \mu$$

for $m, n, u \in [1, ..., N]$.

Proof.

$$\begin{aligned} |(\delta_m * \delta_u)(n)| &= \left| \sum_{\ell=0}^{N-1} \chi_l^*(m)\chi_l^*(u)\chi_l(n) \right| \\ &\leqslant \sum_{\ell=0}^{N-1} |\chi_l^*(m)\chi_l^*(u)|\,|\chi_l(n)| \\ &\leqslant \mu_n \sum_{\ell=0}^{N-1} |\chi_l^*(m)\chi_l^*(u)| \\ &\leqslant \mu_n \|\chi_l^*\|_2 \|\chi_l^*\|_2 \qquad (57) \\ &\leqslant \mu_n \\ &\leqslant \mu \end{aligned}$$

where (57) follows from Hölder inequality. □

Corollary 4. *Similarly, we obtain:*

$$\left| \sum_{m}^{N} \chi_l(m)\chi_{l'}(m)\chi_{l''}^*(m) \right| \leqslant \tilde{\mu}_{l''} \leqslant \mu$$

for $l, l', l'' \in [0, ..., N-1]$.

Lemma 9. *For $f, g \in \mathbb{R}^N$ two signals on graphs, the*

$$\|f * g\|_1 \leqslant N\mu \|f\|_1 \|g\|_1$$

Proof.

$$\begin{aligned}
\|f*g\|_1 &= \sum_{n=1}^{N}\left|\sum_{\ell=0}^{N-1}\hat{f}(\ell)\hat{g}(\ell)\chi_\ell(n)\right| \\
&= \sum_{n=1}^{N}\left|\sum_{\ell=0}^{N-1}\sum_{n'=1}^{N}\sum_{n''=1}^{N} f(n')g(n'')\chi_\ell^*(n')\chi_\ell^*(n'')\chi_\ell(n)\right| \\
&\leqslant \sum_{n=1}^{N}\sum_{n'=1}^{N}|f(n')|\sum_{n''=1}^{N}|g(n'')|\left|\sum_{\ell=0}^{N-1}\chi_\ell^*(n')||\chi_\ell^*(n'')||\chi_\ell(n)\right| \\
&\leqslant N\mu\|f\|_1\|g\|_1 \qquad (58)\\
& \qquad (59)
\end{aligned}$$

□

Lemma 10. *For $f, g \in \mathbb{R}^N$ two signals on graphs, the convolution product satisfies*

$$\|f*g\|_\infty \leqslant \mu C_p C_q \|f\|_p \|g\|_q,$$

where $\frac{1}{p}+\frac{1}{q}=1$.

Proof.

$$\begin{aligned}
\|f*g\|_\infty &= \max_n\left|\sum_{\ell=0}^{N-1}\hat{f}(\ell)\hat{g}(\ell)\chi_\ell(n)\right| \\
&\leqslant \mu\sum_{\ell=0}^{N-1}\left|\hat{f}(\ell)\hat{g}(\ell)\right| \\
&\leqslant \mu\|\hat{f}\|_p\|\hat{g}\|_q \qquad (60)\\
&\leqslant \mu C_p C_q\|f\|_p\|g\|_q \qquad (61)
\end{aligned}$$

Equation (61) follows from theorem 15 and (60) from Hölder's inequality. □

Remark: In particular, we have

$$\|f*g\|_\infty \leqslant \mu N\|f\|_\infty\|g\|_1. \qquad (62)$$

Theorem 22. *For $f, g \in \mathbb{R}^N$ two graph signals and $p, q, r \in \mathbb{R}$ such that $1+\frac{1}{r}=\frac{1}{p}+\frac{1}{q}$, we have*

$$\|f*g\|_r \leqslant N\mu\|f\|_p\|g\|_q. \qquad (63)$$

Proof. For a graph signal $f \in \mathbb{R}^N$, we define an operator $T_g : \mathbb{R}^N \to \mathbb{R}^N$ by $(T_g f)(n) = (f*g)(n)$. Using Lemma 9 and (62), we observe that this operator is bounded from L^1 to L^1 by $N\mu\|g\|_1$ and from L^∞ to L^∞ by $N\mu\|g\|_1$. Thus, we can apply the Riez-Thorin theorem to this operator and we get

$$\|f*g\|_p \leqslant N\mu\|f\|_p\|g\|_1.$$

Then, for a graph signal $g \in \mathbb{R}^N$, we define another operator operator $T_f : \mathbb{R}^N \to \mathbb{R}^N$ by $(T_f g)(n) = (f*g)(n)$. With the previous inequality and Lemma 10, we observe that this new operator is bounded from L^1 to L^p by $N\mu\|f\|_p$ and from $L^{p'}$ to L^∞ by $N\mu\|f\|_p$, where $\frac{1}{p}+\frac{1}{p'}=1$. Again the Riez-Thorin Theorem leads to the desired result:

$$\|f*g\|_r \leqslant N\mu\|f\|_p\|g\|_q,$$

where $1+\frac{1}{r}=\frac{1}{p}+\frac{1}{q}$. □

References

[1] E. H. Lieb, "Integral bounds for radar ambiguity functions and Wigner distributions," *Journal of Mathematical Physics*, vol. 31, pp. 594–599, 1990.

[2] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, Mar. 2011.

[3] D. I Shuman, B. Ricaud, and P. Vandergheynst, "A windowed graph Fourier tranform," in *Proc. IEEE Stat. Signal Process. Wkshp.*, Ann Arbor, MI, Aug. 2012.

[4] F. R. K. Chung, *Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92)*. American Mathematical Society, Feb. 1997.

[5] G. Strang, "The discrete cosine transform," *SIAM Review*, vol. 41, no. 1, pp. 135–147, 1999.

[6] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, 3rd ed. Academic Press, 2008.

[7] H. J. Brascamp and E. H. Lieb, "Best constants in Young's inequality, its converse, and its generalization to more than three functions," *Advances in Mathematics*, vol. 20, no. 2, pp. 151 – 173, 1976.

[8] W. Beckner, "Inequalities in Fourier analysis," *Annals of Mathematics*, vol. 102, no. 1, pp. 159–182, 1975.

[9] M. A. Pinksy, *Introduction to Fourier Analysis and Wavelets*. Pacific Gorve, CA, 2002, vol. 102.

[10] H. G. Feichtinger, D. Onchis-Moaca, B. Ricaud, B. Torresani, and C. Wiesmeyr, "A method for optimizing the ambiguity function concentration," *EUSPICO*, Aug 2012.

[11] B. D. McKay, "The expected eigenvalue distribution of a large regular graph." *Linear Algebra Appl*, vol. 40, pp. 203–216, 1981.