Irregularity-Aware Graph Fourier Transforms

Benjamin Girault, Antonio Ortega *Fellow, IEEE* and Shrikanth S. Narayanan *Fellow, IEEE*

Abstract—In this paper, we present a novel generalization of the graph Fourier transform (GFT). Our approach is based on separately considering the definitions of signal energy and signal variation, leading to several possible orthonormal GFTs. Our approach includes traditional definitions of the GFT as special cases, while also leading to new GFT designs that are better at taking into account the irregular nature of the graph. As an illustration, in the context of sensor networks we use the Voronoi cell area of vertices in our GFT definition, showing that it leads to a more sensible definition of graph signal energy even when sampling is highly irregular.

Index Terms—Graph signal processing, graph fourier transform.

I. INTRODUCTION

With the ever growing deluge of data, graph signal processing has been proposed for numerous applications in recent years, thanks to its ability to study signals lying on irregular discrete structures. Examples include weather stations [1], taxi pickups [2] or bicycle rides [3], [4], people in a social network [5], or motion capture data [6]. While classical signal processing is typically built on top of regular sampling in time (1D Euclidean space), or space (2D Euclidean space), graphs can be applied in irregular domains, as well as for irregular sampling of Euclidean spaces [7]. Thus, graph signal processing can be used to process datasets while taking into consideration irregular relationships between the data points.

Successful use of graph signal processing methods for a given application requires identifying: i) the right graph structure, and ii) the right frequency representation of graph signals. The choice of graph structure has been studied in recent work on graph learning from data [8]–[12]. In this paper we focus on the second question, namely, given a graph structure, how to extend the classical Fourier transform definition, which relies on a regular structure, to a *graph Fourier transform*, which is linked to a discrete irregular structure.

State of the art methods derive the definition of the graph Fourier transform (GFT) from algebraic representations of the graph such as the adjacency matrix $\mathbf{A}$, whose entries are the weights of the edges connecting vertices. If i and j are two vertices connected through an edge weighted by

$w(ij)$, then $A_{ij} = w(ij)$. In the context of sensor networks, edges are often defined by selecting the K-nearest neighbors to each vertex, with weights given by a Gaussian kernel of the Euclidean distance. This is a setting shown to have good properties in the context of manifold sampling [13] when the number of samples is large. The adjacency matrix is then used to compute the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$ with $\mathbf{D}$ the diagonal degree matrix verifying $D_{ii} = d_i = \sum_j w(ij)$. Conventional methods in graph signal processing use the eigenvectors of either $\mathbf{A}$ [14] or $\mathbf{L}$ [15] as graph Fourier modes.

One motivation of this paper comes from physical sensing applications, where data collection is performed by recording on a discrete set of points, such as weather readings in a weather station network. In these applications the distribution of the sensors is irregular, and so it is not easy to map these measurements back onto a grid in order to use the classical signal processing toolbox. Importantly, in many of these applications, the specific arrangement of sensors is unrelated to the quantity measured. As an obvious example, the exact location and distribution of weather stations does not have any influence over the weather patterns in the region. Thus, it would be desirable to develop graph signal processing representations that i) have a meaningful interpretation in the context of the sensed data, and ii) are not sensitive to changes in graph structure, and in particular its degree of regularity.[1]

In order to motivate our problem more precisely on this sensor network example, consider a popular definition of GFT, based on the eigenvalues of $\mathbf{L}$, with variation for a graph signal $\boldsymbol{x}$ defined as:

$$\boldsymbol{x}^{\mathrm{H}}\mathbf{L}\boldsymbol{x} = \frac{1}{2}\sum_{ij} w(ij)\left|x_i - x_j\right|^2.$$

A signal $\boldsymbol{x}$ with high variation will have very different values (*i.e.*, large $|x_i - x_j|^2$) at nodes connected by edges with large weights $w(ij)$. This GFT definition is such that variation for constant signals is zero, since $\mathbf{L}\mathbf{1} = \mathbf{0}$. This is a nice property, as it matches definitions of frequency in continuous domain, *i.e.*, a constant signal corresponds to minimal variation and thus frequency zero. Note that this property is valid independently of the number and position of sensors in the environment, thus achieving our goal of limiting the impact of graph choice on signal analysis.

In contrast, consider the frequency representation associated to an impulse localized to one vertex in the graph. The impulse signal $\boldsymbol{\delta_i}$ has variation $\boldsymbol{\delta_i}^{\mathrm{H}}\mathbf{L}\boldsymbol{\delta_i} = d_i$, where d_i is the degree

---

[1] Note that studies of stationarity of graph signals, such as [1], are concerned with the variability across multiple observations of signals on the *same* graph. Instead here we focus on how the choice of different graphs (*i.e.*, different vertex/sensor positions placed in space) affects the spectral representation of the corresponding graph signals (*e.g.*, different discrete samples of the same continuous domain signal).

of vertex i. Impulses do not have equal variation, so that a highly localized phenomenon in continuous space (or a sensor anomaly affecting just one sensor) would produce different spectral signatures depending on the degree of the vertex where the measurement is localized. As a consequence, if the set of sensors changes, *e.g.*, because only a subset of sensors are active at any given time, the same impulse at the same vertex may have a significantly different spectral representation.

As an alternative, choosing the symmetric normalized Laplacian $\mathcal{L} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$ to define the GFT would lead to the opposite result: all impulses would have the same variation, but a constant signal would no longer correspond to the lowest frequency of the GFT. Note the importance of graph regularity in this trade-off: the more irregular the degree distribution, the more we deviate from desirable behavior for the impulses ($\mathbf{L}$) or for the lowest frequency ($\mathcal{L}$).

As further motivation, consider the definition of signal energy, which is typically its ℓ_2-norm in the literature: $E_{\boldsymbol{x}} = \|\boldsymbol{x}\|_2^2$ [14]–[16]. Assume that there is an area within the region being sensed where the energy of the continuous signal is higher. For this given continuous space signal, energy estimates through graph signal energy will depend significantly on the position of the sensors (*e.g.*, these estimates will be higher if there are more sensors where the continuous signal energy is higher). Thus, estimated energy will depend on the choice of sensor locations, with more significant differences the more irregular the sensor distribution is.

In this paper, we propose a novel approach to address the challenges associated with irregular graph structures by replacing the dot product and the ℓ_2-norm, with a different inner product and its induced norm. Note that here *irregularity* refers to how irregular the graph structure is, with vertices having local graph structures that can vary quite significantly. In particular, a graph can be regular in the sense of all vertices having equal degree, yet vertices not being homogeneous *w.r.t.* to some other respect. Different choices of inner product can be made for different applications. For any of these choices we show that we can compute a set of graph signals of increasing variation, and orthonormal with respect to the chosen inner product. These graph signals form the basis vectors for novel irregularity-aware graph Fourier transforms that are both theoretically and computationally efficient. This framework applies not only to our motivating example of a sensor network, but also to a wide range of applications where vertex irregularity needs to be taken into consideration.

In the rest of this paper, we first introduce our main contribution of an irregularity-aware graph Fourier transform using a given graph signal inner product and graph signal energy (Sec. II). We then explore the definition of graph filters in the context of this novel graph Fourier transform and show that they share many similarities with the classical definition of graph filters, but with more richness allowed by the choice of inner product (Sec. III). We then discuss specific choices of inner product, including some that correspond to GFTs known in the literature, as well novel some novel ones (Sec. IV). Finally, we present two applications of these irregularity-aware graph Fourier transforms: vertex clustering and analysis of

sensor network data (Sec. V).

II. GRAPH SIGNAL ENERGY: LEAVING THE DOT PRODUCT

One of the cornerstones of classical signal processing is the Fourier transform, and one of its essential properties is orthogonality. Indeed, this leads to the Generalized Parseval Theorem: The inner product in the time $\langle \boldsymbol{x}, \mathbf{y} \rangle$ and spectral $\langle \widehat{\boldsymbol{x}}, \widehat{\mathbf{y}} \rangle$ domains of two signals are equal. In this section, we propose a generalization to the definition of the GFT. Our key observation is to note that the choice of an inner product is a parameter in the definition of the GFT, and the usual dot product is not the only choice available for a sound definition of a GFT verifying Parseval's Theorem.

A. Graph Signal Processing Notations

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, w)$ be a *graph* with $\mathcal{V} = \{1, \ldots, N\}$ its *vertex* set,[2] $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ its *edge* set, and $w : \mathcal{E} \to \mathbb{R}_+$ the *weight* function of the edges (as studied in [17], these weights should describe the similarity between vertices). For simplicity, we will denote the edge from i to j as ij. A graph is said *undirected* if for any edge ij, ji is also an edge, and their weights are equal: $w(ij) = w(ji)$. A *self-loop* is an edge connecting a vertex to itself.

A graph is algebraically represented by its *adjacency matrix*[3] $\mathbf{A}$ verifying $A_{ij} = 0$ if ij is not an edge, and $A_{ij} = w(ij)$ otherwise.[4] We use also a normalization of the adjacency matrix by its eigenvalue of largest magnitude $\mu_{\max}$ such that $\mathbf{A}^{(\text{norm})} = \frac{1}{\mu_{\max}}\mathbf{A}$ [18]. We denote $d_i = \sum_j A_{ij}$ the *degree* of vertex i, and $\mathbf{D}$ the diagonal *degree matrix* having $\mathbf{d}$ as diagonal. When the graph is undirected, the *Laplacian matrix* is the matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$. Two normalizations of this matrix are frequently used in the literature: the *normalized Laplacian matrix* $\mathcal{L} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ and the *random walk Laplacian* $\mathbf{L}_{\text{RW}} = \mathbf{D}^{-1}\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{A}$.

In this graph signal processing literature, the goal is to study signals lying on the vertices of a graph. More precisely, a *graph signal* is defined as a function of the vertices $x : \mathcal{V} \to \mathbb{R}$ or $\mathbb{C}$. The vertices being indexed by $\{1, \ldots, N\}$, we represent this graph signal using the column vector $\boldsymbol{x} = [x_1, \ldots, x_N]^T$.

Following the literature, the GFT is defined as the mapping from a signal $\boldsymbol{x}$ to its spectral content $\widehat{\boldsymbol{x}}$ by orthogonal projection onto a set of *graph Fourier modes* (which are themselves graph signals) $\{\mathbf{u_0}, \ldots, \mathbf{u_{N-1}}\}$ and such that $x_i = \sum_l \widehat{x}(l)\,[\mathbf{u_l}]_i = [\mathbf{U}\widehat{\mathbf{x}}]_i$, where $\mathbf{U} = [\mathbf{u_0} \cdots \mathbf{u_{N-1}}]$. We denote $\mathbf{F}^{-1} = \mathbf{U}$ the *inverse GFT matrix* and $\mathbf{F} = \mathbf{U}^{-1}$ the *GFT matrix*.

Furthermore, we assume that there is an operator $\Delta : \mathbb{C}^{\mathcal{V}} \to \mathbb{R}_+$ quantifying how much variation a signal shows on the

---

[2] Although vertices are indexed, the analysis should be independent of the specific indexing chosen, and any other indexing shall lead to the same outcome for any vertex for a sound GSP setting.

[3] Throughout this paper, bold uppercase characters represent matrices and bold lowercase characters represent column vectors.

[4] Note that A_{ij} corresponds here to the edge from i to j, opposite of the convention used in [14]. This allows for better clarity at the expense of the interpretation of the filtering operation $\mathbf{A}\boldsymbol{x}$. Here, $(\mathbf{A}\boldsymbol{x})_i = \sum_j A_{ij}x_j$ is the diffusion along the edges in reverse direction, from j to i.

TABLE I. Hermitian positive semidefinite (HPSD) graph variation operators. Here GQV is the Graph Quadratic Variation (see Sec. IV-A3).

Name (M)	$\Delta(\boldsymbol{x})$
[15] Comb. Lapl. ($\mathbf{L}$)	$\frac{1}{2}\sum_{ij} w(ij)\lvert x_i - x_j\rvert^2$
[15] Norm. Lapl. ($\boldsymbol{\mathcal{L}}$)	$\frac{1}{2}\sum_{ij} w(ij)\left\lvert \frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}}\right\rvert^2$
[18] GQV (($\mathbf{I}-\mathbf{A}^{(\mathrm{norm})})^2$)	$\sum_i \lvert x_i - [\mathbf{A}^{(\mathrm{norm})}\boldsymbol{x}]_i\rvert^2$

TABLE II. Non-HPSD graph variation operators. Here, GTV is the Graph Total Variation (see Sec. IV-A3), and GDV is the graph directed variation (see Sec. II-F).

Name	$\Delta(\boldsymbol{x})$
[18] GTV	$\sum_i \lvert x_i - [\mathbf{A}^{(\mathrm{norm})}\boldsymbol{x}]_i\rvert$
[19] GDV	$\sum_{ij} w(ij)[x_i - x_j]_+$

graph. Δ typically depends on the graph chosen. Several examples are given on Tabs. I and II. The graph variation Δ is related to a frequency as variation increases when frequency increases. The *graph frequency* of the graph Fourier mode $\mathbf{u_l}$ is then defined as $\lambda_l = \Delta(\mathbf{u_l})$.[5]

In summary, the two ingredients we need to perform graph signal processing are i) the graph Fourier matrix $\mathbf{F}$ defining the projection on a basis of elementary signals (*i.e.*, the graph Fourier modes), and ii) the variation operator Δ defining the graph frequency of those elementary signals. Our goal in this paper is to propose a new method to choose the graph Fourier matrix $\mathbf{F}$, one driven by the choice of graph signal inner product. For clarity and conciseness, we focus on undirected graphs, but our definitions extend naturally to directed graphs. An in-depth study of directed graphs will be the subject of a future communication.

B. Norms and Inner Products

In the literature on graph signal processing, it is desirable for the GFT to orthogonal, *i.e.*, *the graph Fourier transform is an orthogonal projection on an orthonormal basis of graph Fourier modes*. Up to now, orthogonality has been defined using the dot product: $\langle \boldsymbol{x},\mathbf{y}\rangle = \mathbf{y}^{\mathrm{H}}\boldsymbol{x}$, with $.^{\mathrm{H}}$ the transpose conjugate operator. We propose here to relax this condition and explore the benefits of choosing an alternative inner product on graph signals to define orthogonality.

First, note that $\langle .,.\rangle_{\mathbf{Q}}$ is an inner product on graph signals, if and only if there exists a *Hermitian positive definite* (HPD) matrix $\mathbf{Q}$ such that:

$$\langle \boldsymbol{x},\mathbf{y}\rangle_{\mathbf{Q}} = \mathbf{y}^{\mathrm{H}}\mathbf{Q}\boldsymbol{x},$$

for any graph signals $\boldsymbol{x}$ and $\mathbf{y}$. We refer to this inner product as the $\mathbf{Q}$-*inner product*. With this notation, the standard dot product is the $\mathbf{I}$-inner product. Moreover, the $\mathbf{Q}$-inner product induces the $\mathbf{Q}$-norm:

$$\lVert \boldsymbol{x}\rVert_{\mathbf{Q}} = \sqrt{\langle \boldsymbol{x},\boldsymbol{x}\rangle_{\mathbf{Q}}} = \sqrt{\boldsymbol{x}^{\mathrm{H}}\mathbf{Q}\boldsymbol{x}}.$$

Therefore, an orthonormal set for the $\mathbf{Q}$-inner product is a set $\{\mathbf{u_l}\}_l$ verifying:

$$\mathbf{u_k}^{\mathrm{H}}\mathbf{Q}\mathbf{u_l} = \begin{cases} 0 & \text{if } k \neq l \\ 1 & \text{otherwise,} \end{cases}$$

i.e., $\mathbf{U}^{\mathrm{H}}\mathbf{Q}\mathbf{U} = \mathbf{I}$ with $\mathbf{U} = [\cdots \mathbf{u_l}\cdots]$.

Although any HPD matrix $\mathbf{Q}$ defines a proper inner product, we will be mostly focusing on diagonal matrices $\mathbf{Q} = \mathrm{diag}(q_1,\ldots,q_N)$. In that case, the squared $\mathbf{Q}$-norm of a graph

[5]We also used before $\lambda_l = \sqrt{\Delta(\mathbf{u_l})}$ when Δ is a quadratic variation [16]. In this paper, however, the difference between the two does not affect the results.

signal $\boldsymbol{x}$, *i.e.*, its energy, is a weighted sum of its squared components:

$$\lVert \boldsymbol{x}\rVert_{\mathbf{Q}}^2 = \sum_{i\in\mathcal{V}} q_i\lvert x_i\rvert^2.$$

Such a norm is simple but will be shown to yield interesting results in our motivating example of sensor networks (see Sec. V-B). Essentially, if q_i quantifies how important vertex q_i is on the graph, the energy above can be used to account for irregularity of the graph structure by correctly balancing vertex importance. Examples of diagonal matrix $\mathbf{Q}$ include $\mathbf{I}$ and $\mathbf{D}$, but more involved definitions can be chosen such as with the illustrating example of a ring in Sec. II-D or our novel approach based on Voronoi cells in Sec. IV-D.

Moreover, endowing the space of graph signals with the $\mathbf{Q}$-inner product yields the Hilbert space $\mathcal{H}_{\mathcal{G}}(\mathbf{Q}) = (\mathbb{C}^{\mathcal{V}},\langle .,.\rangle_{\mathbf{Q}})$ of graph signals of $\mathcal{G}$. This space is important for the next section where we generalize the graph Fourier transform.

Remark 1. *To simplify the proofs, we observe that many results from matrix algebra that rely on diagonalization in an orthonormal basis w.r.t. the dot product can actually be extended to the $\mathbf{Q}$-inner product using the following isometric operator on Hilbert spaces:*

$$\varphi:\ \begin{vmatrix} \mathcal{H}_{\mathcal{G}}(\mathbf{Q}) & \longrightarrow & \mathcal{H}_{\mathcal{G}}(\mathbf{I}) \\ \boldsymbol{x} & \longmapsto & \mathbf{Q}^{1/2}\boldsymbol{x} \end{vmatrix}.$$

Since $\mathbf{Q}$ is invertible, φ is invertible, and an orthonormal basis in one space is mapped to an orthonormal basis in the other space, i.e., $\{\mathbf{u_l}\}_l$ is an orthonormal basis of $\mathcal{H}_{\mathcal{G}}(\mathbf{Q})$ if and only if $\{\varphi(\mathbf{u_l})\}_l$ is an orthonormal basis of $\mathcal{H}_{\mathcal{G}}(\mathbf{I})$.

For example, if h is a linear operator of $\mathcal{H}_{\mathcal{G}}(\mathbf{Q})$ such that $h(\boldsymbol{x}) = \mathbf{H}\boldsymbol{x}$, then $\widetilde{h}\colon \mathbf{y} \mapsto \varphi(h(\varphi^{-1}(\mathbf{y}))) = \mathbf{Q}^{1/2}\mathbf{H}\mathbf{Q}^{-1/2}\mathbf{y}$ is an operator of $\mathcal{H}_{\mathcal{G}}(\mathbf{I})$. Also, the eigenvectors of $\mathbf{H}$ are orthonormal *w.r.t.* the $\mathbf{Q}$-inner product if and only if the eigenvectors of $\mathbf{Q}^{1/2}\mathbf{H}\mathbf{Q}^{-1/2}$ are orthonormal *w.r.t.* the dot product. This applies for example to $\mathbf{L}_{RW} = \mathbf{D}^{-1}\mathbf{L}$ the random walk Laplacian. Since $\boldsymbol{\mathcal{L}} = \mathbf{D}^{1/2}(\mathbf{D}^{-1}\mathbf{L})\mathbf{D}^{-1/2}$ has orthonormal eigenvectors *w.r.t.* the dot product, then the eigenvectors of $\mathbf{L}_{RW}$ are orthonormal *w.r.t.* the $\mathbf{D}$-inner product (see Sec. IV-B).

C. Contribution 1: Generalized Graph Fourier Modes

Definition 1 (Generalized Graph Fourier Modes). *Given the Hilbert space $\mathcal{H}_{\mathcal{G}}(\mathbf{Q})$ of graph signals and the variation operator Δ, the set of $(\Delta,\mathbf{Q})$-graph Fourier modes is defined as an orthonormal basis of graph signals $\{\mathbf{u_l}\}_l$ solution to the following sequence of minimization*

problems, for increasing $L \in \{0, \ldots, N-1\}$*:*

$$\min_{\mathbf{u_L}} \Delta(\mathbf{u_L}) \qquad subj.\ to \quad \mathbf{U_L^H Q U_L = I}, \qquad (1)$$

where $\mathbf{U_L} = [\mathbf{u_0} \ldots \mathbf{u_L}]$.

This definition is mathematically sound since the search space of each minimization problem is non empty. Note also that this definition relies on only two assumptions: i) $\mathbf{Q}$ is an inner product matrix, and ii) Δ maps graph signals to real values.[6]

Among the examples of graph variation examples given in Tabs. I and II, those of Tab. I share the important property of being quadratic forms:

Definition 2 (Hermitian Positive Semi-Definite Form)**.** *The variation operator* Δ *is a* Hermitian positive semi-definite (HPSD) form *if and only if* $\Delta(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{H}} \mathbf{M} \boldsymbol{x}$ *and* $\mathbf{M}$ *is a Hermitian positive semi-definite matrix.*

When Δ is an HPSD form it is then algebraically characterized by the matrix $\mathbf{M}$. In what follows we denote the graph variation operator as $\mathbf{M}$ whenever $\Delta(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{H}} \mathbf{M} \boldsymbol{x}$ is verified. Examples from the literature of graph variation operators that are HPSD forms are shown on Tab. I, and non-HPSD ones are shown on Tab. II.

The following theorem is of utmost importance as it relates the solution of (1) to a generalized eigenvalue problem when Δ is an HPSD form:

Theorem 1. *If* Δ *is an HPSD form with HPSD matrix* $\mathbf{M}$*, then* $\{\mathbf{u_l}\}_l$ *is a set of* $(\Delta, \mathbf{Q})$*-graph Fourier modes if and only if the graph signals* $\{\mathbf{u_l}\}_l$ *solve the following generalized eigenvalue problems for increasing eigenvalues* $\lambda_l = \Delta(\mathbf{u_l})$*:*

$$\mathbf{M u_l} = \lambda_l \mathbf{Q u_l},$$

with $\|\mathbf{u_l}\|_{\mathbf{Q}}^2 = 1$.

Proof. Observing that $\boldsymbol{x}^{\mathrm{H}} \mathbf{M} \boldsymbol{x} = \langle \mathbf{Q}^{-1} \mathbf{M} \boldsymbol{x}, \boldsymbol{x} \rangle_{\mathbf{Q}}$, we can show that $\mathbf{M}$ Hermitian is equivalent to $\mathbf{Q}^{-1}\mathbf{M}$ being a self-adjoint operator on the Hilbert space $\mathcal{H}_{\mathcal{G}}(\mathbf{Q})$. Therefore, using the spectral theorem, there exists an orthonormal basis $\{\mathbf{u_l}\}_l$ of $\mathcal{H}_{\mathcal{G}}(\mathbf{Q})$ of eigenvectors of $\mathbf{Q}^{-1}\mathbf{M}$, such that:
$$\Delta(\mathbf{u_l}) = \mathbf{u_l}^{\mathrm{H}} \mathbf{M u_l} = \mathbf{u_l}^{\mathrm{H}} \mathbf{Q Q}^{-1} \mathbf{M u_l} = \lambda_l \mathbf{u_l}^{\mathrm{H}} \mathbf{Q u_l} = \lambda_l. \ \square$$

The relation above between the variation of the graph Fourier modes and the eigenvalues of $\mathbf{M}$ also shows that, just as approaches of the literature based on the Laplacian [15] or on the adjacency matrix [14], the set of graph Fourier modes is not unique when there are multiple eigenvalues: If $\Delta(\mathbf{u_l}) = \Delta(\mathbf{u_k})$, then if $\mathbf{v_l} = \alpha \mathbf{u_l} + (1-\alpha)\mathbf{u_k}$ and $\mathbf{v_k} = (1-\alpha)\mathbf{u_l} + \alpha \mathbf{u_k}$, we have $\Delta(\mathbf{u_l}) = \Delta(\mathbf{v_l}) = \Delta(\mathbf{v_k})$ and $\langle \mathbf{v_l}, \mathbf{v_k} \rangle_{\mathbf{Q}} = 0$, for any $\alpha \in [0, 1]$. This remark will be important in Sec. III, as it is desirable for the processing of graph signals to be independent of the particular choice of $(\Delta, \mathbf{Q})$-graph Fourier modes.

D. Example with the Ring Graph: Influence of $\mathbf{Q}$

To give intuitions on the influence of $\mathbf{Q}$ on the $(\Delta, \mathbf{Q})$-graph Fourier modes, we study the example of a ring graph with the Laplacian matrix as graph variation operator matrix: $\mathbf{M} = \mathbf{L}$. In Fig. 1, we show the $(\mathbf{L}, \mathbf{Q})$-graph Fourier modes of a ring with 8 vertices for three choices of $\mathbf{Q}$:

- $\mathbf{Q} = \mathrm{diag}(0.1, 1, \ldots, 1)$ with a less important vertex 1,
- $\mathbf{Q} = \mathbf{I}$, *i.e.*, the combinatorial Laplacian GFT,
- $\mathbf{Q} = \mathrm{diag}(10, 1, \ldots, 1)$ with a more important vertex 1.

Several observations can be made. First of all, any $(\mathbf{L}, \mathbf{I})$-graph Fourier mode which is zero on vertex 1 (*i.e.*, $\mathbf{u_1}$, $\mathbf{u_3}$, $\mathbf{u_5}$) is also an $(\mathbf{L}, \mathbf{Q})$-graph Fourier mode with the same graph variation. Intuitively, this vertex having zero value means that it has no influence on the mode (here, $\mathbf{Q u_1} = \mathbf{u_1}$), hence its importance does not affect the mode.

For the remaining modes, we make several observations. First of all, we consider the spectral representation of a highly localized signal on vertex 1, such as $\boldsymbol{\delta_1}$ pictured in the last column of Fig. 1. While $\mathbf{Q} = \mathbf{I}$ involves many spectral components of roughly the same power, the other two cases $q_1 = 0.1$ and $q_1 = 10$ are distinctive. Indeed, in the first case, we have $\widehat{\boldsymbol{\delta_1}}(7)$ that is large (almost an order of magnitude larger), interpreted by a graph Fourier mode $[\mathbf{u_7}]$ that is close to our localized signal. On the other hand, for $q_1 = 10$ the two largest Fourier components of $\boldsymbol{\delta_1}$ are $\widehat{\boldsymbol{\delta_1}}(0)$ and $\widehat{\boldsymbol{\delta_1}}(1)$ with $[\mathbf{u_1}]$ being close to our localized signal. In other words, q_i shifted the impulse $\boldsymbol{\delta_1}$ to the higher spectrum when small (unimportant vertex i) or the lower spectrum when large (important vertex i).

These cases give intuitions on the impact of $\mathbf{Q}$, and ultimately on how to choose it.

E. Discussion on $\mathbf{M}$

When Δ is and HPSD from, we can rewrite the minimization of (1) as a generalized Rayleigh quotient minimization. Indeed, the Lth minimization problem is also given by:

$$\min_{\boldsymbol{x}: \mathbf{U_L^H Q} \boldsymbol{x} = \mathbf{0}} \Delta \left(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_{\mathbf{Q}}} \right),$$

which is exactly a Rayleigh quotient minimization when Δ is an HPSD form of HPSD matrix $\mathbf{M}$ since:

$$\Delta \left(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_{\mathbf{Q}}} \right) = \frac{\boldsymbol{x}^{\mathrm{H}} \mathbf{M} \boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}} \mathbf{Q} \boldsymbol{x}}.$$

For example, this allows the study of bandpass signals using spectral proxies as in [20].

Having an HPSD form for Δ also has two advantages, the first one being a simple solution to (1) as stated in Thm. 1. But more importantly, this solution is efficient[7] to compute through the generalized eigenvalue problem since both $\mathbf{Q}$ and $\mathbf{M}$ are Hermitian [21, Section 8.7]. Therefore, instead of inverting $\mathbf{Q}$ and computing eigenvalues and eigenvectors of the non-Hermitian matrix $\mathbf{Q}^{-1}\mathbf{M}$, the two matrices $\mathbf{M}$ and $\mathbf{Q}$ can be directly used to compute the $(\mathbf{M}, \mathbf{Q})$-graph Fourier modes using their Hermitian properties.

---

[6]$\mathbf{Q}$ may depend on Δ, but no such relation is required in Def. 1. An example is given at the end of this section where $q_i = \Delta(\boldsymbol{\delta_i})$ is desirable.

[7]Efficiency comes here from the cost of computing generalized eigenvalues and eigenvectors compared to the when $\mathbf{M}$ is not HPSD or $\mathbf{Q}$ is not HPD.
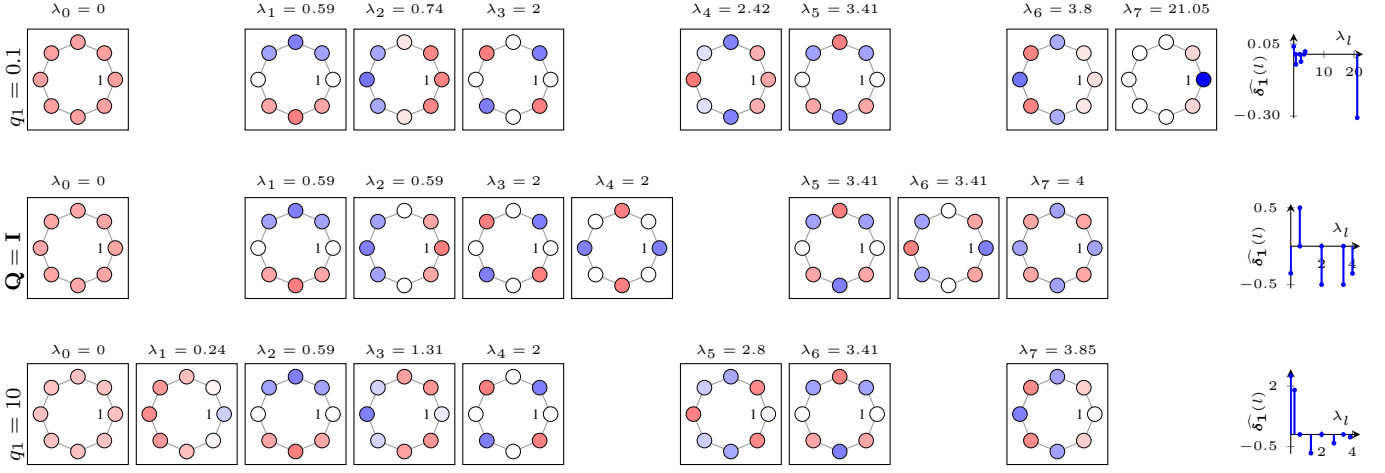
Fig. 1. $(\mathbf{L}, \mathbf{Q})$-graph Fourier modes of the ring graph with $N = 8$ vertices. $\mathbf{Q} = \mathrm{diag}(q_1, 1, \ldots, 1)$ with one choice of q_1 per row of the table, for increasing values of q_1. Modes are not aligned on their index, but when they correspond either exactly (3rd, 5th, and 8th column), or approximately, to highlight the impact of q_1 on the graph Fourier modes. Colors scale from deep blue for -1 to deep red for 1 and color being lighter as values get closer to 0 until white for 0. Note that $[\mathbf{u_7}]_1 = -3.08$ for $q_1 = 0.1$ is out of this range. The last column shows the graph Fourier transform $\widehat{\boldsymbol{\delta_1}}$ of the impulse $\boldsymbol{\delta_1}$.

In some applications, it may also be desirable to have $\mathbf{Q}$ depend on $\mathbf{M}$. In our motivating example the key variations of the constant signal $\mathbf{1}$ and of the impulses $\boldsymbol{\delta_i}$ are given by:

$$\Delta(\mathbf{1}) = \sum_{ij} M_{ij} \qquad \Delta(\boldsymbol{\delta_i}) = M_{ii}.$$

In the case of $\mathbf{1}$, variation should be zero, hence $\sum_{ij} M_{ij} = 0$. However, for $\boldsymbol{\delta_i}$, the variations above cannot be directly compared from one impulse to another as those impulses have different energy. Variations of the energy normalized signals yields:

$$\Delta\left(\frac{\boldsymbol{\delta_i}}{\|\boldsymbol{\delta_i}\|_{\mathbf{Q}}}\right) = \frac{M_{ii}}{Q_{ii}}.$$

We advocated for a constant (energy normalized) variation. In effect, this leads to a relation between $\mathbf{Q}$ and $\mathbf{M}$ given by constant M_{ii}/Q_{ii}, *i.e.*, the diagonals of $\mathbf{Q}$ and $\mathbf{M}$ should be equal, up to a multiplicative constant. For instance, choosing $\mathbf{M} = \mathbf{L}$, which leads to $\Delta(\mathbf{1}) = 0$, and $\Delta(\boldsymbol{\delta_i}) = d_i$, our requirement on $\mathbf{Q}$ leads to $q_i = d_i$, hence $\mathbf{Q} = \mathbf{D}$. This is the random walk Laplacian approach described in Sec. IV-B.

F. Relation to [19]

Finally, we look at the recent advances in defining the graph Fourier transform from the literature, and in particular to [19] which aims at defining an orthonormal set of graph Fourier modes for a directed graph with non-negative weights. After defining the *graph directed variation* as:

$$\mathrm{GDV}(\boldsymbol{x}) := \sum_{ij} w(ij) \left[x_i - x_j\right]_+,$$

where $[x]_+ = \max(0, x)$, the graph Fourier modes are computed as a set of orthonormal vectors *w.r.t.* the dot product that solve:

$$\min_{\{\mathbf{u_l}\}_l} \sum_l \mathrm{GDV}(\mathbf{u_l}) \qquad \text{subj. to} \quad \mathbf{U}^{\mathrm{H}}\mathbf{U} = \mathbf{I}. \qquad (2)$$

The straightforward generalization of this optimization problem to any graph variation operator Δ and $\mathbf{Q}$-inner product,

is then:

$$\min_{\{\mathbf{u_l}\}_l} \sum_l \Delta(\mathbf{u_l}) \qquad \text{subj. to} \quad \mathbf{U}^{\mathrm{H}}\mathbf{Q}\mathbf{U} = \mathbf{I}. \qquad (3)$$

Unfortunately, we cannot use this generalization when Δ is an HPSD form. Indeed, in that case, the sum in (3) is exactly $\mathrm{tr}(\mathbf{U}^{\mathrm{H}}\mathbf{M}\mathbf{U}) = \mathrm{tr}(\mathbf{Q}^{-1}\mathbf{M})$ for any matrix $\mathbf{U}$ verifying $\mathbf{U}^{\mathrm{H}}\mathbf{Q}\mathbf{U} = \mathbf{I}$. For example, $\mathbf{U} = \mathbf{Q}^{-1/2}$ solves (3), but under the assumption that $\mathbf{Q}$ is diagonal, this leads to a trivial diagonal GFT, modes localized on vertices of the graph.

Note that, for any graph variation, a solution to our proposed minimization problem in (1) is also a solution to the generalization in (3):

Property 1. *If $\{\mathbf{u_l}\}_l$ is a set of $(\Delta, \mathbf{Q})$-graph Fourier modes, then it is a solution to:*

$$\min_{\{\mathbf{u_l}\}_l} \sum_l \Delta(\mathbf{u_L}) \qquad \text{subj. to} \quad \mathbf{U}^{\mathrm{H}}\mathbf{Q}\mathbf{U} = \mathbf{I},$$

with $\mathbf{U} = [\mathbf{u_0} \ldots \mathbf{u_{N-1}}]$.

For example, the set of $(\mathrm{GDV}, \mathbf{I})$-graph Fourier modes is a solution to (2), hence a set of graph Fourier modes according to [19]. This property is important as it allows, when direct computation through a closed form solution of the graph Fourier modes is not possible, to approximate the $(\Delta, \mathbf{Q})$-graph Fourier modes by first using the techniques of [19] to obtain the $(\mathbf{Q}^{1/2}\, \Delta\, \mathbf{Q}^{-1/2}, \mathbf{I})$-graph Fourier modes and then using Remark 1.

Finally, another recent work uses a related optimization function to obtain graph Fourier modes [22]:

$$\min_{\{\mathbf{u_l}\}_l} \sum_l \left(f_l - \mathrm{GQDV}(\mathbf{u_l})\right)^2 \qquad \text{subj. to} \quad \mathbf{U}^{\mathrm{H}}\mathbf{U} = \mathbf{I},$$

with $f_l = \frac{l-1}{N-1} f_{\max}$ and the *graph quadratic directed variation* defined as:

$$\mathrm{GQDV}(\boldsymbol{x}) := \sum_{ij} w(ji) \left[x_i - x_j\right]_+^2.$$

Beyond the use of a squared directed difference $\left[x_i - x_j\right]_+^2$,

the goal of this optimization is to obtain evenly distributed graph frequencies, and not orthogonal graph Fourier modes of minimally increasing variation of [19] and of our contribution. Note that this alternative approach can implement the constraint $\mathbf{U}^{\mathrm{H}}\mathbf{Q}\mathbf{U} = \mathbf{I}$ to use an alternative definition of graph signal energy. This is however out of the scope of this paper.

G. Contribution 2: Generalized GFT

Given the $(\Delta, \mathbf{Q})$-graph Fourier modes in the previous section, and the definition of the inverse graph Fourier transform $\mathbf{F}^{-1} = \mathbf{U}$ found in the literature and recalled in Sec. II-A, we can now define the generalized graph Fourier transform. Note that Δ is not assumed to be an HPSD form anymore.

Definition 3 (Generalized Graph Fourier Transform). *Let $\mathbf{U}$ be the matrix of $(\Delta, \mathbf{Q})$-graph Fourier modes. The $(\Delta, \mathbf{Q})$-graph Fourier transform $((\Delta, \mathbf{Q})$-GFT) is then:*

$$\mathbf{F} = \mathbf{U}^{\mathrm{H}}\mathbf{Q},$$

and its inverse is:

$$\mathbf{F}^{-1} = \mathbf{U}.$$

The inverse in Def. 3 above is a proper inverse since $\mathbf{F}\mathbf{F}^{-1} = \mathbf{U}^{\mathrm{H}}\mathbf{Q}\mathbf{U} = \mathbf{I}$. In Sec. II-A, we introduced the graph Fourier transform as an orthogonal projection on the graph Fourier modes. Def. 3 is indeed such a projection since:

$$\widehat{x}(l) = [\mathbf{F}\boldsymbol{x}]_l = \left[\mathbf{U}^{\mathrm{H}}\mathbf{Q}\boldsymbol{x}\right]_l = \mathbf{u_l}^{\mathrm{H}}\mathbf{Q}\boldsymbol{x} = \langle \boldsymbol{x}, \mathbf{u_l}\rangle_{\mathbf{Q}}.$$

Another important remark on Def. 3 concerns the complexity of computing the graph Fourier matrix. Indeed, this matrix is the inverse of the inverse graph Fourier matrix $\mathbf{F}^{-1} = \mathbf{U}$ which is directly obtained using the graph Fourier modes. Computation of a matrix inverse is in general costly and subject to approximation errors. However, just as classical GFTs for undirected graphs, we can obtain this inverse without preforming an actual matrix inverse. Indeed, our GFT matrix is obtained through a simple matrix multiplication $\mathbf{U}^{\mathrm{H}}\mathbf{Q} = \mathbf{U}^{-1}$ that uses the orthonormal property of the graph Fourier modes basis. Additionally, when $\mathbf{Q}$ is diagonal, this matrix multiplication is extremely simple and easy to compute.

One property that is essential in the context of an orthonormal set of graph Fourier modes is Parseval's Identity where the energy in the vertex and spectral domains are equal:

Property 2 (Generalized Parseval's Theorem). *The $(\Delta, \mathbf{Q})$-GFT is an isometric operator from $\mathcal{H}_{\mathcal{G}}(\mathbf{Q})$ to $\mathcal{H}_{\widehat{\mathcal{G}}}(\mathbf{I})$. Mathematically:*

$$\langle \boldsymbol{x}, \mathbf{y}\rangle_{\mathbf{Q}} = \langle \widehat{\boldsymbol{x}}, \widehat{\mathbf{y}}\rangle_{\mathbf{I}}.$$

Proof. $\langle \boldsymbol{x}, \mathbf{y}\rangle_{\mathbf{Q}} = \mathbf{y}^{\mathrm{H}}\mathbf{Q}\boldsymbol{x} = \widehat{\mathbf{y}}^{\mathrm{H}}\left(\mathbf{F}^{-1}\right)^{\mathrm{H}}\mathbf{Q}\mathbf{F}^{-1}\widehat{\boldsymbol{x}} = \widehat{\mathbf{y}}^{\mathrm{H}}\widehat{\boldsymbol{x}}$. ☐

Finally, Prop. 2 bears similarities with Remark 1. Indeed, in both cases, there is an isometric map from the Hilbert space $\mathcal{H}_{\mathcal{G}}(\mathbf{Q})$ to either $\mathcal{H}_{\mathcal{G}}(\mathbf{I})$ or $\mathcal{H}_{\widehat{\mathcal{G}}}(\mathbf{I})$. However, in Prop. 2, the graph signal $\boldsymbol{x}$ is mapped to the spectral components $\mathbf{F}\boldsymbol{x}$ of $\boldsymbol{x}$ instead of $\mathbf{Q}^{1/2}\boldsymbol{x}$ such that both cases are distinct. Intuitively, φ and $\mathbf{Q}^{1/2}$ reshape the space of graph signals to account for irregularity, while the graph Fourier matrix $\mathbf{F}$ decomposes the graph signal into elementary graph Fourier modes of distinct graph variation.

III. GRAPH FILTERS

In this section we explore the concept of operator on graph signals, and more specifically, operators whose definition is intertwined with the graph Fourier transform. Such a relation enforces a spectral interpretation of the operator and ultimately ensures that the graph structure plays an important role in the output of the operator.

A. Fundamental Matrix of the GFT

Before defining graph filters, we need to introduce what we call the *fundamental matrix of the $(\Delta, \mathbf{Q})$-GFT* given its graph Fourier matrix $\mathbf{F}$ and the diagonal matrix of graph frequencies $\boldsymbol{\Lambda}$:

$$\mathbf{Z} := \mathbf{F}^{-1}\boldsymbol{\Lambda}\mathbf{F} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{H}}\mathbf{Q}.$$

Although Δ does not appear in the definition above, $\mathbf{Z}$ does depend on Δ through the graph Fourier matrix $\mathbf{F}$. Some authors use the term *shift* for this matrix when $\mathbf{Q} = \mathbf{I}$. However, the literature uses very often the Laplacian matrix [15] as the fundamental matrix, and as a close equivalent to a second order differential operator [13], it does not qualify as the equivalent of a shift operator. Therefore, we choose the more neutral term of *fundamental matrix*. Yet, our definition is a generalization of the literature where such a matrix is always diagonalizable in the graph Fourier basis, with graph frequencies as eigenvalues. Tab. III shows several classical choices of fundamental matrices depending on $\mathbf{Q}$ and Δ.

Further assuming that Δ is an HPSD form, we have $\mathbf{Z} = \mathbf{Q}^{-1}\mathbf{M}$. As we will see in Sec. IV, this is consistent with the graph signal processing literature. Noticeably, this matrix is also uniquely defined under these conditions, even though the graph Fourier matrix may not be. Moreover, complexity of computing this matrix is negligible when $\mathbf{Q}$ is diagonal. A diagonal matrix $\mathbf{Q}$ also leads to $\mathbf{Z}$ having the same sparsity as $\mathbf{M}$ since $\mathbf{M}$ and $\mathbf{Z}$ differ only for non-zero elements of $\mathbf{M}$. Algorithms whose efficiency is driven by the sparsity of $\mathbf{M}$ are generalized without loss of efficiency to $\mathbf{Q}^{-1}\mathbf{M} = \mathbf{Z}$. This will be the case for the examples of this paper.

B. Definitions of Graph Filters

We recall here the classical definitions of graph filters and describe how we straightforwardly generalize them using the fundamental matrix $\mathbf{Z}$ of the GFT.

First of all, given a graph filter $\mathbf{H}$, we denote by $\widehat{\mathbf{H}} = \mathbf{F}\mathbf{H}\mathbf{F}^{-1}$ the same graph filter in the spectral domain such that $\widehat{\mathbf{H}\boldsymbol{x}} = \widehat{\mathbf{H}}\widehat{\boldsymbol{x}}$. This notation allows to properly study the spectral response of a given graph filter. We can now state the three definitions of graph filters.

The first one directly extends the invariance through time shifting. This is the approach followed by [14], replacing the adjacency matrix by the fundamental matrix of the graph:

Definition 4 (Graph Filters by Invariance). **H** *is a* graph filter *if and only if it commutes with the fundamental matrix* **Z** *of the GFT:*

$$\mathbf{HZ} = \mathbf{ZH}.$$

The second definition extends the convolution theorem for temporal signals, where convolution in the time domain is equivalent to pointwise multiplication in the spectral domain [17, Sec. 3.2][8]. The following definition is identical to those in the literature, simply replacing existing choices of GFT with one of our proposed GFTs:

Definition 5 (Graph Filters by Convolution Theorem). **H** *is a graph filter if and only if there exists a graph signal* **h** *such that:*

$$\widehat{\mathbf{Hx}}(l) = \widehat{h}(l)\widehat{x}(l).$$

The third definition is also a consequence of the convolution theorem, but with a different interpretation. Indeed, in the time domain, the Fourier transform $\widehat{s}$ of a signal s is a function of the (signed) frequency. Given that there is a finite number of graph frequencies, any function of the graph frequency can be written as a polynomial of the graph frequency (through polynomial interpolation). We obtain the following definition using the fundamental matrix instead [23], [24]:

Definition 6 (Graph Filters by Polynomials). **H** *is a* graph filter *if and only if it is a polynomial in the fundamental matrix* **Z** *of the GFT:*

$$\mathbf{H} = \sum_k h_k \mathbf{Z}^k.$$

Interestingly, these definitions are equivalent for a majority of graphs where no two graph frequencies are equal. However, in the converse case of a graph with two equal graph frequencies $\lambda_k = \lambda_l$, these definitions differ. Indeed, Def. 6 implies that $\widehat{h}(k) = \widehat{h}(l)$, while $\widehat{h}(k)$ and $\widehat{h}(l)$ are arbitrary according to Def. 5. Also, a graph filter according to Def. 4 does not necessarily verify $\widehat{\mathbf{H}}$ diagonal, since $\widehat{\mathbf{H}}u_l = H_{k,l}u_k + H_{l,l}u_l$ is a valid graph filter, even with $H_{k,l} \neq 0$. Choosing one definition over another is an application-dependent choice driven by the meaning of two graph Fourier modes of equal graph frequencies. For example, if these modes are highly related, then Def. 6 is a natural choice with equal spectral response of the filter, whereas in the opposite case of unrelated modes, Def. 5 allows to account for more flexibility in the design of the filter.

C. Mean Square Error

Mean Square Error (MSE) is classically used to study the error made by a filter when attempting to recover a signal from a noisy input. More precisely, given an observation $\mathbf{y} = \boldsymbol{x} + \mathbf{n}$

of a signal $\boldsymbol{x}$ with additive random noise $\mathbf{n}$ (with zero mean), MSE is defined as:

$$\mathrm{MSE}_{\mathbf{H}}(\boldsymbol{x}) := \mathbb{E}\left[\|\mathbf{Hy} - \boldsymbol{x}\|_2^2\right]$$
$$= \sum_i \mathbb{E}\left[([\mathbf{Hy}]_i - x_i)^2\right]. \quad (4)$$

In other words, this is the mean energy of the error made by the filter. However, the definition of energy used above is the dot product. As stated before, this energy does not account for irregularity of the structure.

In the general case of an irregular graph, we defined in Sec. II-B graph signal energy using the **Q**-norm. We define in this section the energy of an error by the **Q**-norm of that error, thus generalizing mean square error into the **Q**-MSE:

$$\mathrm{MSE}_{\mathbf{H}}^{(\mathbf{Q})}(\boldsymbol{x}) := \mathbb{E}\left[\|\mathbf{Hy} - \boldsymbol{x}\|_{\mathbf{Q}}^2\right].$$

Using the zero mean assumption, this yields the classical bias/variance formula:

$$\mathrm{MSE}_{\mathbf{H}}^{(\mathbf{Q})}(\boldsymbol{x}) := \underbrace{\|\mathbf{Hx} - \boldsymbol{x}\|_{\mathbf{Q}}^2}_{\text{bias term}} + \underbrace{\mathbb{E}\left[\|\mathbf{Hn}\|_{\mathbf{Q}}^2\right]}_{\text{noise variance term}} \quad (5)$$

The definition of graph Fourier transform introduced in Def. 3 is then a natural fit to study this MSE thanks to Parseval's identity (Prop. 2):

$$\mathrm{MSE}_{\mathbf{H}}^{(\mathbf{Q})}(\boldsymbol{x}) = \left\|\left(\widehat{\mathbf{H}} - \mathbf{I}\right)\widehat{\boldsymbol{x}}\right\|_2^2 + \mathbb{E}\left[\left\|\widehat{\mathbf{H}}\widehat{\mathbf{n}}\right\|_2^2\right].$$

Since $\widehat{\mathbf{H}}$ is diagonal (or block diagonal if using Def. 4), studying the bias and the variance from a spectral point of view is much simpler. Recalling the meaning of the graph Fourier modes in terms of graph variation, this also allows to quantify the bias and variance for slowly varying components of the signal (lower spectrum) to quickly varying components (higher spectrum) giving an intuitive interpretation to MSE and the bias/variance trade off across graph frequencies.

To have a better intuition on how using the **Q**-norm allows to account for irreguarity, we now assume that **Q** is diagonal. Let $\boldsymbol{\epsilon} = \mathbf{Hn}$ be the filtered noise. The noise variance term above becomes:

$$\mathbb{E}\left[\|\boldsymbol{\epsilon}\|_{\mathbf{Q}}^2\right] = \sum_i q_i \mathbb{E}\left[|\epsilon_i|^2\right]$$

In other words, those vertices with higher q_i have a higher impact on the overall noise energy. Using $\mathbb{E}\left[\|\boldsymbol{\epsilon}\|_{\mathbf{Q}}^2\right] = \mathrm{tr}\left(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{H}}\mathbf{Q}\right)$, we also have:

$$\mathbb{E}\left[\|\mathbf{Hn}\|_{\mathbf{Q}}^2\right] = \mathrm{tr}\left(\widehat{\mathbf{H}}^{\mathrm{H}}\widehat{\mathbf{H}}\mathbf{F}\boldsymbol{\Sigma}_{\mathbf{n}}\mathbf{Q}\mathbf{F}^{-1}\right),$$

where $\boldsymbol{\Sigma}_{\mathbf{n}} = \mathbb{E}\left[\mathbf{nn}^{\mathrm{H}}\right]$ is the covariance matrix of the noise. We introduce a definition of **Q**-white noise with a tractable power spectrum:

Definition 7 (Q-White Noise). *The graph signal noise* **n** *is said* **Q**-*White Noise (***Q**-*WN) if and only if it is centered* $\mathbb{E}[\mathbf{n}] = \mathbf{0}$ *and its covariance matrix verifies* $\boldsymbol{\Sigma}_{\mathbf{n}} = \sigma^2 \mathbf{Q}^{-1}$, *for some* $\sigma \geq 0$.

There are two important observations to make on this definition. In the vertex domain, if we assume a diagonal **Q**, this noise has higher power on vertices with smaller q_i:

---

[8]In [23], the authors also define convolutions as pointwise multiplications in the spectral domain. However, the notation used for the GFT is unclear since $\widehat{x}(\lambda_l)$ (instead of $\widehat{x}(l)$) can be interpreted as the graph signal $\boldsymbol{x}$ having equal spectral components for equal graph frequencies. Such a requirement actually falls within the setting of Def. 6.

$\mathbb{E}\big[|n_i|^2\big] = \frac{\sigma^2}{q_i}$. *Assuming $\mathbf{Q}$-WN is equivalent to assuming less noise on vertices with higher q_i.* Therefore, this definition of noise can account for the irregularity of the graph structure through $\mathbf{Q}$.

Second, the importance of this definition is best seen in the spectral domain. Indeed, the spectral covariance matrix verifies:

$$\mathbf{\Sigma}_{\widehat{\mathbf{n}}} = \mathbf{F}\mathbf{\Sigma}_{\mathbf{n}}\mathbf{F}^{\mathrm{H}} = \sigma^2\mathbf{I}.$$

In other words, a $\mathbf{Q}$-WN has a flat power spectrum. Note that this is true independently of the variation operator Δ chosen to define the $(\Delta, \mathbf{Q})$-GFT matrix $\mathbf{F}$, hence the name of $\mathbf{Q}$-WN (and not $(\Delta, \mathbf{Q})$-WN).

The noise variance term in (5) under the assumption of a $\mathbf{Q}$-WN $\mathbf{n}$ becomes:

$$\mathbb{E}\left[\|\mathbf{H}\mathbf{n}\|_{\mathbf{Q}}^2\right] = \sigma^2\,\mathrm{tr}\left(\widehat{\mathbf{H}}^{\mathrm{H}}\widehat{\mathbf{H}}\right).$$

In other words, it is completely characterized by the spectral response of the filter $\mathbf{H}$. Note that using Def. 4, $\widehat{\mathbf{H}}$ is not necessarily Hermitian. Using Def. 5 or Def. 6, $\widehat{\mathbf{H}}$ is diagonal and the noise variance term simplifies to:

$$\mathbb{E}\left[\|\mathbf{H}\mathbf{n}\|_{\mathbf{Q}}^2\right] = \sigma^2\sum_l\left|\widehat{h}(l)\right|^2.$$

IV. CHOOSING THE GRAPH SIGNAL ENERGY MATRIX $\mathbf{Q}$

The goal of this section is twofold. First, we look at the literature on the graph Fourier transform to show how closely it relates to our definition, including the classical Fourier transform for temporal signals. This is summarized in Tab. III. The second goal is to give examples of graph Fourier transforms that use the newly introduced degree of freedom allowed by the introduction of the $\mathbf{Q}$-inner product.

A. The Dot Product: $\mathbf{Q} = \mathbf{I}$

1) Temporal Signals: This case derives from classical digital signal processing [25], with $x[k]$ a periodic temporal signal of period T and sampled with sampling frequency N/T (N samples per period). This sampling corresponds to a ring graph with N vertices. In this case, energy is classically defined as a scaled ℓ_2-norm of the vector $[x[0], \ldots, x[N-1]]^T$, *i.e.*, $E_{\boldsymbol{x}} = \frac{T}{N}\boldsymbol{x}^{\mathrm{H}}\mathbf{I}\boldsymbol{x}$. DFT modes are eigenvectors of the continuous Laplacian operator $\Delta_T\boldsymbol{x} = \frac{d^2\boldsymbol{x}}{dt^2}$, thus corresponding to the variation $\Delta(\boldsymbol{x}) = \langle\Delta_T\boldsymbol{x}, \boldsymbol{x}\rangle = \boldsymbol{x}^{\mathrm{H}}\Delta_T\boldsymbol{x}$. Finally, DFT modes are orthogonal *w.r.t.* the dot product.

2) Combinatorial and Normalized Laplacians: This classical case relies on the computations of the eigenvectors of the combinatorial Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$ or of the normalized Laplacian $\boldsymbol{\mathcal{L}} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$ to define the graph Fourier modes [15]. These graph Fourier transforms are exactly the $(\mathbf{L}, \mathbf{I})$-GFT and the $(\boldsymbol{\mathcal{L}}, \mathbf{I})$-GFT.

3) Graph Shift: This is a more complex case where the (generalized) eigenvectors of the adjacency matrix are used as graph Fourier modes [14]. When the graph is undirected, it can be shown that this case corresponds to the $((\mathbf{I} - \mathbf{A}^{(\mathrm{norm})})^2, \mathbf{I})$-GFT [18], with $\Delta(\boldsymbol{x}) = \|\boldsymbol{x} - \mathbf{A}^{(\mathrm{norm})}\boldsymbol{x}\|_2^2 = \mathrm{GQV}(\boldsymbol{x})$, the *Graph Quadratic Variation*. An alternative definition of smoothness based on the ℓ_1-norm of $\boldsymbol{x} - \mathbf{A}^{(\mathrm{norm})}\boldsymbol{x}$ leading

to the *Graph Total Variation* $\Delta(\boldsymbol{x}) = \|\boldsymbol{x} - \mathbf{A}^{(\mathrm{norm})}\boldsymbol{x}\|_1 = \mathrm{GTV}(\boldsymbol{x})$ is also investigated in [18]. However, this norm leads to a different graph Fourier transform when used to solve (1), since an ℓ_1-norm promotes sparsity (smaller support in graph Fourier modes) while an ℓ_2-norm promotes smoothness. Finally, when the graph is directed, the resulting GFT of [14] has no guarantee that its graph Fourier modes are orthogonal. Note that the same $\Delta(\boldsymbol{x}) = \|\boldsymbol{x} - \mathbf{A}^{(\mathrm{norm})}\boldsymbol{x}\|_2^2$ can be used in the directed case, leading to $\mathbf{M} = (\mathbf{I} - \mathbf{A}^{(\mathrm{norm})})^{\mathrm{H}}(\mathbf{I} - \mathbf{A}^{(\mathrm{norm})})$. Solving (1) is then equivalent to computing the SVD of $(\mathbf{I} - \mathbf{A}^{(\mathrm{norm})})$ and using its right singular vectors as $(\mathbf{M}, \mathbf{I})$-graph Fourier modes. However, these differ from the eigenvectors of $(\mathbf{I} - \mathbf{A}^{(\mathrm{norm})})$ used in [14] since those are orthogonal.

4) Directed Variation Approach ($\Delta = $ GDV): We presented in Sec. II-F a recently proposed GFT [19]. This approach minimizes the sum of directed variation of an orthogonal set of graph signals (see (3)). As explained in Sec. II-F, the $(\Delta, \mathbf{I})$-GFT is a directed GFT as described in [19].

B. The Degree-Norm (Random Walk Laplacian): $\mathbf{Q} = \mathbf{D}$

The random walk Laplacian, $\mathbf{L}_{\mathrm{RW}} = \mathbf{D}^{-1}\mathbf{L}$, is not widely used in the graph signal processing literature, given its lack of symmetry, so that its eigenvectors are not necessarily orthogonal. Therefore, the graph Fourier matrix $\mathbf{F}$ is not unitary, hence naively computing it through the matrix inverse $\mathbf{F} = \mathbf{U}^{-1}$ is not efficient. Yet, this normalization is successfully used in clustering in a similar manner to the combinatorial and normalized Laplacians [26]. Noticeably, it can be leveraged to compute an optimal embedding of the vertices into a lower dimensionnal space, using the first few eigenvectors [27]. In graph signal processing we can cite [28], [29] as examples of applications of the random walk Laplacian. Another example, in image processing, is [30] where the authors use the random walk Laplacian as smoothness prior for soft decoding of JPEG images through $x^{\mathrm{H}}\mathbf{L}\mathbf{D}^{-1}\mathbf{L}x = \|\mathbf{L}_{\mathrm{RW}}x\|_{\mathbf{D}}^2$.

Our framework actually allows for better insights on this case from a graph signal processing perspective[9]. Indeed, considering the inner product with $\mathbf{Q} = \mathbf{D}$, we obtain the $(\mathbf{L}, \mathbf{D})$-GFT, whose fundamental matrix is $\mathbf{Z} = \mathbf{D}^{-1}\mathbf{L} = \mathbf{L}_{\mathrm{RW}}$. In [30], this leads to minimizing $\|\mathbf{L}_{\mathrm{RW}}x\|_{\mathbf{D}}^2 = \|\mathbf{Z}x\|_{\mathbf{D}}^2$ which is equivalent to minimizing the energy in the higher spectrum since $\mathbf{Z}$ is a high pass filter. This GFT is orthonormal *w.r.t.* the $\mathbf{D}$-inner product, leading to a graph signal energy definition based on the *degree-norm*: $E_{\boldsymbol{x}} = \|\boldsymbol{x}\|_{\mathbf{D}}^2 = \boldsymbol{x}^{\mathrm{H}}\mathbf{D}\boldsymbol{x}$. As stated in Remark 1, this normalization is related to the normalized Laplacian through the relation $\mathbf{L}_{\mathrm{RW}} = \mathbf{D}^{-1/2}\boldsymbol{\mathcal{L}}\mathbf{D}^{1/2}$, such that $\mathbf{L}_{\mathrm{RW}}$ and $\boldsymbol{\mathcal{L}}$ share the same eigenvalues and their eigenvectors are related: If $\boldsymbol{x}$ is an eigenvector of $\mathbf{L}_{\mathrm{RW}}$, $\varphi(\boldsymbol{x})$ is an eigenvector of $\boldsymbol{\mathcal{L}}$.

By combining the variation of the combinatorial Laplacian with the normalization of the normalized Laplacian, the random walk Laplacian achieves properties that are desirable for

---

[9]The property that the eigenvectors of the random walk Laplacian are orthogonal *w.r.t.* the $\mathbf{Q}$-norm is well known in the literature. We are however the first to make the connection with a properly defined graph Fourier transform.

TABLE III. State of the art of the graph Fourier transform.

Ref	Name	Directed	Weights	Δ	Q	Orthon.	$\Delta(\mathbf{1})=0$	$\Delta(\boldsymbol{\delta_i})=cst$	Z
[15]	Comb. Lapl.	✗	Non-neg.	$\mathbf{L=D-A}$	I	✓	✓	✗	L
[15]	Norm. Lapl.	✗	Non-neg.	$\boldsymbol{\mathcal{L}}=\mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}}$	I	✓	✗	✓	$\boldsymbol{\mathcal{L}}$
[14]	Graph Shift	✗ ✓	Complex	$(\mathbf{I-A}^{\text{norm}})^{\text{H}}\times$ $(\mathbf{I-A}^{\text{norm}})$	I	✓ ✗	✗	✗	A
[19]	Graph Cut	✓	Non-neg.	GDV	I	Approx.	✓	✗	n.a.
	RW Lapl.	✗	Non-neg.	$\mathbf{L_{RW}=D^{-1}L}$ L	I D	✗ ✓	✓	✓	$\mathbf{L_{RW}}$

a sensor network: $\Delta(\boldsymbol{\delta_i}/\|\boldsymbol{\delta_i}\|_{\mathbf{D}})=\frac{d_i}{d_i}=1$ since $\|\boldsymbol{\delta_i}\|_{\mathbf{D}}=d_i$, and $\mathbf{L1}=\mathbf{0}=0\cdot\mathbf{D1}$ i.e., the graph frequencies are normalized, the constant signals have zero variation, and all impulses have the same variation, while having different energies[10].

This case is actually justified in the context of manifold sampling in the extension [13] of [7]. The authors show that under some conditions on the weights of the edges of the similarity graph, the random walk Laplacian is essentially the continuous Laplacian, without additive or multiplicative term, even if the probability density function of the samples is not uniform.

C. Bilateral Filters: $\mathbf{Q=I+D}$

Bilateral filters are used in image processing to denoise images while retaining clarity on edges in the image [31]:

$$y_i=\frac{1}{1+d_i}\left(x_i+\sum_j w(ij)x_j\right)$$

with

$$w(ij)=\exp\left(-\frac{\|p_i-p_j\|^2}{2\sigma_{\text{d}}^2}\right)\exp\left(-\frac{\|I(i)-I(j)\|^2}{2\sigma_{\text{i}}^2}\right)$$

where p_i is the position of pixel i, $I(i)$ is its intensity, and σ_{d} and σ_{i} are parameters. Intuitively, weights are smaller when pixels are either far (first Gaussian kernel) or of different intensities (second Gaussian kernel). This second case corresponds to image edges. We can rewrite this filtering operation in matrix form as:

$$\mathbf{y}=(\mathbf{I+D})^{-1}(\mathbf{I+A})\boldsymbol{x}=\mathbf{I}-(\mathbf{I+D})^{-1}\mathbf{L}\boldsymbol{x}.$$

In other words, using $\mathbf{Z}=(\mathbf{I+D})^{-1}\mathbf{L}$, we obtain that bilateral filtering is the polynomial graph filter $\mathbf{I-Z}$. This corresponds exactly to a graph filter with the $(\mathbf{L,I+D})$-GFT.

Moreover, given a noisy observation $\mathbf{y}=\boldsymbol{x}+\mathbf{n}$ of the noiseless image $\boldsymbol{x}$, noise on the output of this filter is given by $(\mathbf{I-Z})\mathbf{y}-\boldsymbol{x}=\mathbf{Z}\boldsymbol{x}+(\mathbf{I-Z})\mathbf{n}$. This noise can be studied using the $(\mathbf{I+D})$-MSE introduced in Sec. III-C. Indeed, we can experimentally observe that pixels do not contribute equally to the overall error, with pixels on the edges (lower degree) being less filtered than pixels in smooth regions (higher degree). experimentally, we observe that $(\mathbf{I-Z})\mathbf{n}$ is an $(\mathbf{I+D})$-WN, thus validating the use of the $(\mathbf{I+D})$-MSE. This approach of quantifying noise is coherent with human perception of noise as the human eye is more sensitive to small changes in smooth

areas. We will develop the study of bilateral filters with respect to our setting in a future communication.

Finally, note that the expression of bilateral filtering shown here is different than in [29]:

$$y_i=\frac{1}{d_i}\sum_j w(ij)x_j.$$

This expression is equivalent to considering the filter $\mathbf{I-Z}$ with the $(\mathbf{L,D})$-GFT, and is not the original approach of [31].

D. Voronoi Cell Areas: $\mathbf{Q=C}$

In the motivating example of a sensor network, we wished to find a graph Fourier transform that is not biased by the particular sampling being performed. Considering the energy of the signal being measured, this energy should not vary with the sampling. In the continuous domain, we observe that energy of a continuous signal $\widetilde{s}$ is defined as:

$$E_s:=\int|\widetilde{s}(x)|^2 dx.$$

In the definition above, dx acts as an elementary volume, and the integral can be interpreted as a sum over elementary volumes of the typical value $|\widetilde{s}(x)|$ within that volume times the size of the volume dx. The discrete version of this integral is therefore:

$$E_s\approx\sum_i|\widetilde{s}(x_i)|^2\operatorname{vol}(i)=\sum_i|s_i|^2\operatorname{vol}(i),$$

with $\mathbf{s}$ the sampled signal on the points $\{x_i\}_i$.

Given a particular sampling, the question is then what is a good value of $\operatorname{vol}(i)$? A simple and intuitive approximation is to approximate this volume with the volume of the subspace of points whose closest sample is x_i. This subspace is exactly the Voronoi cell of i, and the volume is therefore the area of the cell in 2D. Let c_i be this area, and $\mathbf{C}=\operatorname{diag}(c_1,\ldots,c_N)$. We obtain:

$$E_s\approx\sum_i|s_i|^2 c_i=\mathbf{s}^{\text{H}}\mathbf{Cs}=\|\mathbf{s}\|_{\mathbf{C}}^2.$$

Intuitively, this corresponds to interpolating the sampled signal into a piecewise constant signal for which the signal values are equal within each Voronoi cell, and then computing the continuous energy of this interpolated signal. Other interpolation schemes could be considered. For example, if we assume a weighted linear interpolation scheme, then we obtain the interpolated signal $\widetilde{\mathbf{s}}(\boldsymbol{x})=\sum_i f(\boldsymbol{x},\boldsymbol{x_i})s_i$ and its energy is:

$$E_{\widetilde{\mathbf{s}}}=\int|\widetilde{\mathbf{s}}(\boldsymbol{x})|^2 d\boldsymbol{x}=\sum_{i,j}\left(\int f(\boldsymbol{x},\boldsymbol{x_i})^{\text{H}}f(\boldsymbol{x},\boldsymbol{x_j})d\boldsymbol{x}\right)s_i^{\text{H}}s_j,$$

[10]Impulse energies show here the irregularity of the graph structure, and as such are naturally not constant.
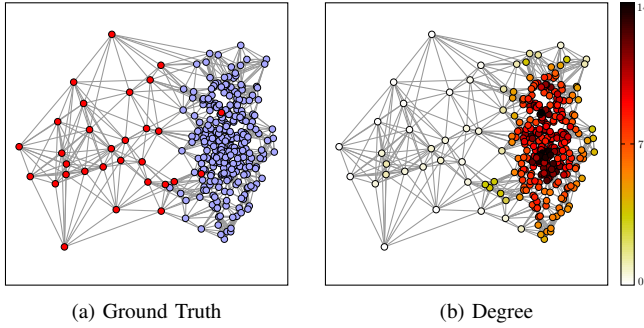
(a) Ground Truth   (b) Degree

Fig. 2. Two cluster dataset input with a sparse cluster (left) and a dense cluster (right). (a) Ground truth clusters. (b) Vertex degrees from the 10-nearest neighbor graph Gaussian weights.

and we have $E_{\widetilde{\mathbf{s}}} = \|\mathbf{s}\|_{\mathbf{Q}}^2$ with $q_{ij} = \int f(\mathbf{x}, \mathbf{x_i})^{\mathrm{H}} f(\mathbf{x}, \mathbf{x_j}) d\mathbf{x}$. As soon as there is one location $\mathbf{x}$ whose interpolation involves two samples i,j, then this interpolation scheme corresponds a non-diagonal matrix $\mathbf{Q}$. However, as we will show in Sec. V-B, the approach based on the Voronoi cell areas gives already good results compared to the state of the art of graph signal processing for sensor networks.

V. EXPERIMENTS

The choice of a graph Fourier transform, by selecting a graph variation operator Δ and a matrix $\mathbf{Q}$ such as those discussed in Sec. IV, is application dependent. In particular, meaningful properties for the graph Fourier transform can be drastically different. For example, in a clustering context, the practitioner is interested in extracting the structure, *i.e.*, in identifying tight groups of vertices. This is in contrast to sensor networks where the particular arrangement of sensors should have as little influence over the results as possible, *i.e.*, tight groups of stations should not be a source of bias (*e.g.*, weather stations).

For this reason, there is no unique way of studying how good a particular graph Fourier transform is. In this section, we use these two applications, *i.e.*, clustering and sensor networks, to show how our framework can be leveraged to achieve their application-specific goals.

All experiments were performed using the `GraSP` toolbox for `Matlab` [32].

A. Clustering

In this section, we study the problem of clustering defined as grouping similar objects into groups of object that are similar while being dissimilar to objects in other groups. In a graph setting, we are interested in grouping vertices of a graph such that there are many edges within each group, while groups are connected by few edges. Our goal is not to provide a new approach to perform clustering, but rather use this problem as a showcase for how using a well defined matrix $\mathbf{Q}$ can help achieve the goals of a target application.

In the context of clustering, spectral clustering extracts groups using the spectral properties of the graph [26]. More precisely, using C-means (k-means with $k = C$) on the first C graph Fourier modes yields interesting partitions. In

[26], the author interprets this approach using graph cuts, for $(\mathbf{L}, \mathbf{I})$-GFT (combinatorial Laplacian), $(\mathcal{L}, \mathbf{I})$-GFT (normalized Laplacian) and $(\mathbf{L}, \mathbf{D})$-GFT (random walk Laplacian). We extend this interpretation here to any variation operator Δ and any diagonal innner product matrix $\mathbf{Q}$.

For each cluster c, let $\mathcal{V}_c \subset \mathcal{V}$ be the subset of its vertices. Then the set $\{\mathcal{V}_c\}_c$ of all these subsets is a partition of the set of vertices $\mathcal{V}$. Let the normalized cut of $\{\mathcal{V}_c\}_c$ associated to the $(\Delta, \mathbf{Q})$-GFT be:

$$(\Delta, \mathbf{Q})\text{-Ncut}\left(\{\mathcal{V}_c\}_c\right) := \sum_c \Delta\left(\mathbf{h}_c^{(\mathbf{Q})}\right)$$

with $\mathbf{h}_c^{(\mathbf{Q})}$ the normalized indicator function of cluster $\mathcal{V}_c$ verifying:

$$\mathbf{h}_c^{(\mathbf{Q})} := \frac{1}{\|\mathbf{1}_{\mathcal{V}_c}\|_{\mathbf{Q}}} \mathbf{1}_{\mathcal{V}_c}$$

and $\mathbf{1}_{\mathcal{V}_c}$ the indicator function of cluster c. This extends the normalized cut interpretation of [26]. Using the notation $\mathbf{H} = [\mathbf{h}_1^{(\mathbf{Q})} \cdots \mathbf{h}_C^{(\mathbf{Q})}]^T$, and since $\mathbf{Q}$ is diagonal, we obtain the orthonormality property $\mathbf{H}^{\mathrm{H}} \mathbf{Q} \mathbf{H} = \mathbf{I}$. Finding a partition of vertices minimizing the normalized cut $(\Delta, \mathbf{Q})$-Ncut $(\{\mathcal{V}_c\}_c)$ is therefore equivalent to finding orthonormal normalized indicator functions of minimal variation.

Spectral clustering is performed by first relaxing the constraint that the graph signals $\mathbf{h}_c^{(\mathbf{Q})}$ are normalized indicator functions. Using Prop. 1, the first C $(\Delta, \mathbf{Q})$-graph Fourier modes are solutions to this relaxed optimization problem. The final partition is then obtained through C-means on the spectral features, where the feature vector of vertex i is $\left[[\mathbf{u}_0]_i, \ldots, [\mathbf{u}_{C-1}]_i\right]^T$.

The accuracy of spectral clustering is therefore driven by the choice of variation operator Δ and inner product matrix $\mathbf{Q}$. To illustrate this, and the statement that choosing $\mathbf{Q}$ and Δ is application-dependent, we look at several choices in the context of a 2-class clustering problem with skewed clusters. Combined with an irregular distribution of inputs within each cluster, we are in the context of irregularity where our framework thrives through its flexibility.

The resulting graph is shown on Fig. 2, with a sparse cluster on the left (30 inputs), and a dense one on the right (300 inputs). Each cluster is defined by a 2D Gaussian distribution, with an overlapping support, and samples are drawn from these two distributions. We already observe that inputs are irregularly distributed, and some inputs are almost isolated. These almost isolated inputs are important as they represent outliers. We will see that correctly choosing $\mathbf{Q}$ alleviates the influence of those outliers. The graph of inputs is built in a

TABLE IV. F_1 score of the sparse (left) cluster on Fig. 2 for the clustering results of Fig. 3.

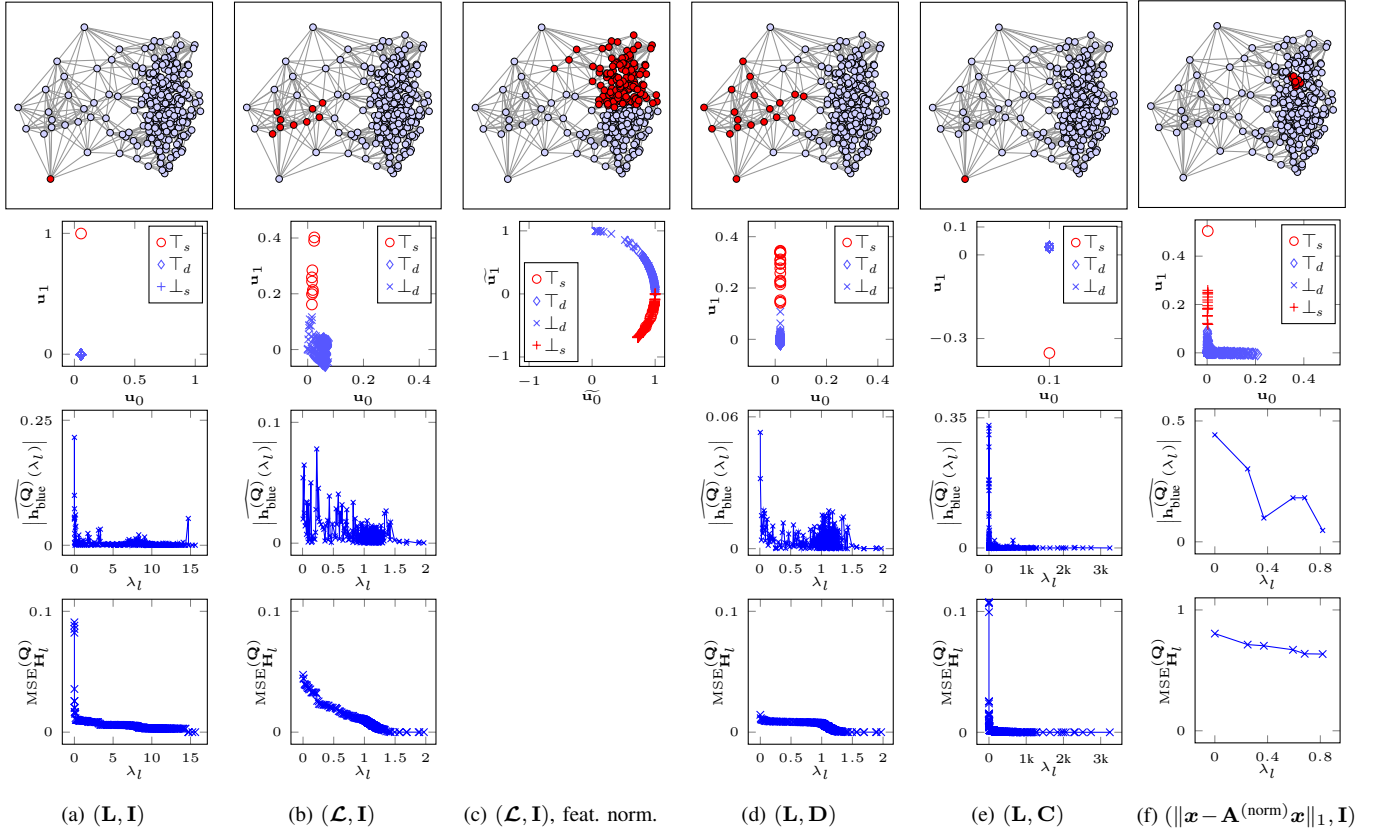| Approach	Accuracy	Sparse F_1 score
$(\mathbf{L}, \mathbf{I})$	91.21%	6.45%
$(\mathcal{L}, \mathbf{I})$	93.64%	46.15%
$(\mathcal{L}, \mathbf{I})$, feat. norm.	46.67%	4.35%
$(\mathbf{L}, \mathbf{D})$	**96.67%**	**77.55%**
$(\mathbf{L}, \mathbf{C})$	91.21%	6.45%
$(\|\boldsymbol{x} - \mathbf{A}^{(\mathrm{norm})}\boldsymbol{x}\|_1, \mathbf{I})$	86.06%	4.17%

Fig. 3. Several spectral clustering results on the graph of Fig. 2. The smallest cluster is in dark red, and the largest cluster is in light blue. The second row of plots shows the input in the feature space, *i.e.*, to each vertex i is associated a point $\left[[\mathbf{u_0}]_i, [\mathbf{u_1}]_i\right]^T$ in the spectral feature space. The set $\top_s$ (resp. $\top_d$) corresponds to the vertices in the sparse (resp. dense) cluster and correctly clustered. The set $\perp_s$ (resp. $\perp_d$) corresponds to vertices incorrectly clustered in the sparse (resp. dense) cluster. Colors in both rows of figures match for consistency. The third row shows the $(\mathbf{M}, \mathbf{Q})$-GFT of the corresponding normalized indicator function $\mathbf{h}_{\text{blue}}^{(\mathbf{Q})}$ (the first Fourier component is not shown as it is quite large, except for (f)). The fourth row shows the $\mathbf{Q}$-MSE of the ideal low-pass filter $\mathbf{H}_l$ with varying cut-off frequency λ_l applied to this normalized indicator function. Generalized GFT: (a): $(\mathbf{L}, \mathbf{I})$, (b) $(\mathcal{L}, \mathbf{I})$, (c) $(\mathcal{L}, \mathbf{I})$ with normalized feature vectors, (d) $(\mathbf{L}, \mathbf{D})$, (e) $(\mathbf{L}, \mathbf{C})$, and (f) $(\Delta, \mathbf{I})$ with $\Delta(\boldsymbol{x}) = \|\boldsymbol{x} - \mathbf{A}^{(\text{norm})}\boldsymbol{x}\|_1$.

K-nearest neighbor fashion ($K = 10$) with weights chosen using a Gaussian kernel of the Euclidean distance ($\sigma = 0.4$).

Fig. 3 shows the results of spectral clustering on this graph, using several GFTs (first two rows), the analysis of $\mathbf{h}_{\text{blue}}^{(\mathbf{Q})}$ in the spectral domain (third row), and an example of $\mathbf{Q}$-MSE of several ideal low pass filters on the same indicator function. First of all, the *Random Walk Laplacian* case based on the $(\mathbf{L}, \mathbf{D})$-GFT gives the best results, consistent with [26] where the author advocates this approach. [26, Prop. 5] gives an intuition on why this case works best using a random walk interpretation: If a starting vertex is chosen at random from the stationary distribution of the random walk on the graph, then the $(\mathbf{L}, \mathbf{D})$-Ncut is exactly the probability to jump from one cluster to another. Minimizing this cut, is then minimizing transitions between clusters. In the spectral domain, we see that the indicator function has a lot of energy in the first few spectral components, with a rapid decay afterwards. We notice also, a slight increase of energy around graph frequency 1 due to the red vertices of the sparse (red) cluster which are in the middle of the dense (blue) cluster (see Fig. 2(a)). As expected, the $\mathbf{Q}$-MSE decreases when the cut-off frequency increases. But interestingly, if we use the classical $\mathbf{I}$-MSE, it is not decreasing with the cut-off frequency because of the use of the $\mathbf{Q}$-norm instead of the ℓ_2-norm (see Fig. 4(a).

The combinatorial Laplacian case, *i.e.*, using the $(\mathbf{L}, \mathbf{I})$-GFT, suffers from the presence of outliers. Indeed, if i is such an outlier, then having it in a separate cluster, *i.e.*, $\mathcal{V}_1 = \{i\}$ and $\mathcal{V}_2 = \mathcal{V}\setminus\{i\}$ for $C = 2$, leads to $\Delta(\mathbf{h}_1^{(\mathbf{I})}) = \Delta(\mathbf{h}_2^{(\mathbf{I})}) = d_i$. Since i is isolated, it has extremely low degree, and the resulting cut is small, making it a good clustering according to the $(\mathbf{L}, \mathbf{I})$-Ncut. Not accounting for irregularity of the degree in the definition of the normalized indicator function $\mathbf{h}_c^{(\mathbf{I})}$ leads therefore to poor clustering in the presence of outliers. A similar behavior is observed for the $(\mathbf{L}, \mathbf{C})$-GFT with a normalization by the Voronoi cell area that is large for isolated vertices, hence an even smaller normalized cut. In the spectral domain, more Fourier components of the normalized indicator function are large, especially around the graph frequency 0. This results in a $\mathbf{Q}$-MSE that is large for our ideal low-pass filters. In other words, more lowpass components are required to correctly approximate the indicator function using a low-pass signal.

Next, is the normalized Laplacian case, *i.e.*, using the $(\mathcal{L}, \mathbf{I})$-GFT. Just as the Random Walk Laplacian case, singleton clusters with outliers are not associated to small cuts since $(\mathcal{L}, \mathbf{I})$-Ncut$(\{i\}, \mathcal{V}\setminus\{i\}) = (\mathbf{L}, \mathbf{D})$-Ncut$(\{i\}, \mathcal{V}\setminus\{i\})$. However, a careful study of the graph Fourier modes reveals the weakness of this case. Indeed, using the well known relation
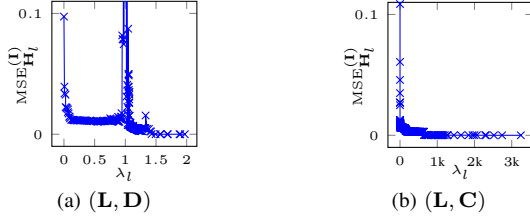
Fig. 4. **I**-MSE of the ideal low-pass filter $\mathbf{H}_l$ applied to the normalized indicator function $\mathbf{h}_{\text{blue}}^{(\mathbf{Q})}$ corresponding to (a) Fig. 3(d) and (b) to Fig. 3(e).

between the $(\mathcal{L}, \mathbf{I})$-graph Fourier modes $\{\mathbf{u_l}\}_l$ and the $(\mathbf{L}, \mathbf{D})$-graph Fourier modes $\{\mathbf{v_l}\}_l$ with $\varphi(\mathbf{v_l}) = \mathbf{u_l} = \mathbf{D}^{1/2}\mathbf{v_l}$, this case deviates from the Random Walk Laplacian by introducing a scaling of graph Fourier modes by the degree matrix. Differences are then noticeable in the spectral feature space (bottom plots in Fig. 3). Indeed, the graph Fourier modes verify $[\mathbf{u_l}]_i = \sqrt{d_i}[\mathbf{v_l}]_i$, such that low degree vertices, *e.g.*, outliers and vertices in the boundary of a cluster, have spectral features of smaller magnitude: $\left\| [[\mathbf{u_0}]_i, [\mathbf{u_1}]_i]^T \right\| \ll \left\| [[\mathbf{v_0}]_i, [\mathbf{v_1}]_i]^T \right\|$. In other words, spectral features of low degree vertices are moved closer to the origin of the feature space, and closer to each other. C-means cluster those vertices together, resulting in the poor separation of clusters we see on Fig. 3(b). In the spectral domain, we clearly see the interest of using the random walk Laplacian approach: the normalized indicator functions are not lowpass. This is further seen on the $\mathbf{Q}$-MSE that decreases continuously with the cutoff frequency without the fast decay we wish for.

The spectral clustering literature is actually not using the raw spectral features to run C-means, but normalize those features by their ℓ_2-norm prior to running C-means [33]. This is actually equivalent to the projection of the $(\mathbf{L}, \mathbf{D})$-graph Fourier modes $\{\mathbf{v_l}\}_l$ on the unit circle:

$$[\widetilde{\mathbf{u_0}}]_i = \cos(\theta_i) \quad [\widetilde{\mathbf{u_1}}]_i = \sin(\theta_i) \quad \theta_i = \operatorname{atan}\left(\|\mathbf{1}\|_{\mathbf{D}}[\mathbf{v_1}]_i\right).$$

We can use this relation to characterize how spectral features of vertices are modified compared to the $(\mathbf{L}, \mathbf{D})$-GFT approach, and how C-means behaves. Indeed, separation of features (*i.e.*, how far the spectral features of two vertices are) is modified by the transformation above. Looking at the norm of the gradient of these spectral features yields:

$$\left\| \overrightarrow{\operatorname{grad}} \begin{bmatrix} [\widetilde{\mathbf{u_0}}]_i \\ [\widetilde{\mathbf{u_1}}]_i \end{bmatrix} \right\|_2 = \frac{\|\mathbf{1}\|_{\mathbf{D}}}{1 + (\|\mathbf{1}\|_{\mathbf{D}}[\mathbf{v_1}]_i)^2}$$

Therefore, the gradient is smaller for higher values of $|[\mathbf{v_1}]_i|$: This spectral feature normalization brings features closer for higher values of $|[\mathbf{v_1}]_i|$. As features of the dense and sparse clusters are identified by the magnitude of $|[\mathbf{v_1}]_i|$, this normalization actually brings features of the two clusters closer. This yields less separation in the feature space and poor clustering output from C-means (see Fig. 3(c)).

Finally, we performed the same specral clustering approach using the $(\Delta, \mathbf{I})$-GFT with $\Delta(x) = \|\boldsymbol{x} - \mathbf{A}^{(\text{norm})}\boldsymbol{x}\|_1$ the graph total variation. Results on Fig. 3(f) show that this approach actually identifies the stronger small clusters rather than larger clusters. More importantly, in this case, the spectral features of both clusters are not separated, such that C-means cannot

separate clusters based on the spectral features. Furthermore, it turns out that the clusters given by these spectral features are not the best with an $(\text{GTV}, \mathbf{I})$-Ncut of 8.21, while those given by the combinatorial Laplacian yields a $(\text{GTV}, \mathbf{I})$-Ncut of 6.61. In other words, the relaxation we made that consider arbitrary $\mathbf{h}_c^{(\mathbf{I})}$ instead of normalized indicator functions leads to large errors. The spectral domain analysis suffers from the difficulty to compute efficiently the graph Fourier modes, since the graph variation is not HPSD. However, the first few components of the normalized indicator function show that this signal is now low pass, with a lot of energy in these components. $\mathbf{Q}$-MSE is not sharply decreasing either and stays very high.

B. Sensor Networks

Here we explore our motivating example of a sensor network, where the goal is the opposite of that in the one above: clusters of vertices should not bias the results. More precisely, we are interested in studying signals measured by sensor networks where the location of the sensors is unrelated to the quantity measured, *e.g.*, temperature readings in a weather station network. In this section, we show that we can obtain a good definition of inner product achieving this goal.

Let $\mathcal{G}^{(\text{U})}$ be a graph whose vertices are uniformly sampled on a 2D square plane, and $\mathcal{G}^{(\text{NU})}$ a similar graph obtained with a non-uniform sampling distribution (see Fig. 5(e) where the areas of higher density are indicated). Edges of both graphs are again selected using K-nearest neighbors, with $K = 10$. Weights are chosen as a Gaussian kernel of the Euclidean distance $w_{ij} = \exp\left(-\mathrm{d}(i, j)^2/(2\sigma^2)\right)$, with $\sigma = 0.3$ (empirically chosen to have a good spread of weights). Example graphs with 500 sampled vertices are shown in Figs. 5(a) and 5(e). In this experiment, we are interested in studying how we can remove the influence of the particular arrangement of sensors such as to obtain the same results from one graph realization to another. To that end, we generate 500 graph realizations, each with 1000 vertices, to compute empirical statistics.

The signals we consider are pure sine waves, which allows us to control the energies and frequencies of the signals we use. We experiment with several signals for each frequency and report the worst case statistics over those signals of equal (continuous) frequency. This is done by choosing several values of phase shift for each frequency.

Let $\widetilde{s_\nu}(x, y; \varphi) = \cos(2\pi\nu x_i + \varphi)$ be a continuous horizontal Fourier mode of frequency ν phase shifted by φ. For a given graph $\mathcal{G}$ generated with any of the two schemes above, we define $s_\nu(i; \varphi, \mathcal{G}) = \widetilde{s_\nu}(x_i, y_i; \varphi)$ the sampled signal on the graph. Its energy *w.r.t.* the $\mathbf{Q}$-norm is then:

$$E_\nu^{(\mathbf{Q})}(\varphi; \mathcal{G}) := \|s_\nu(\_; \varphi, \mathcal{G})\|_{\mathbf{Q}}^2.$$

Note that this energy depends on the graph if $\mathbf{Q}$ does, *i.e.*, we have $\mathbf{Q}(\mathcal{G})$ which may vary from one graph to another such as $\mathbf{Q}(\mathcal{G}) = \mathbf{D}(\mathcal{G})$. We use the shorter notation $\mathbf{Q}$ to keep notations shorter.

To get better insights on the influence of the sensor sampling, we are interested in statistical quantities of the graph
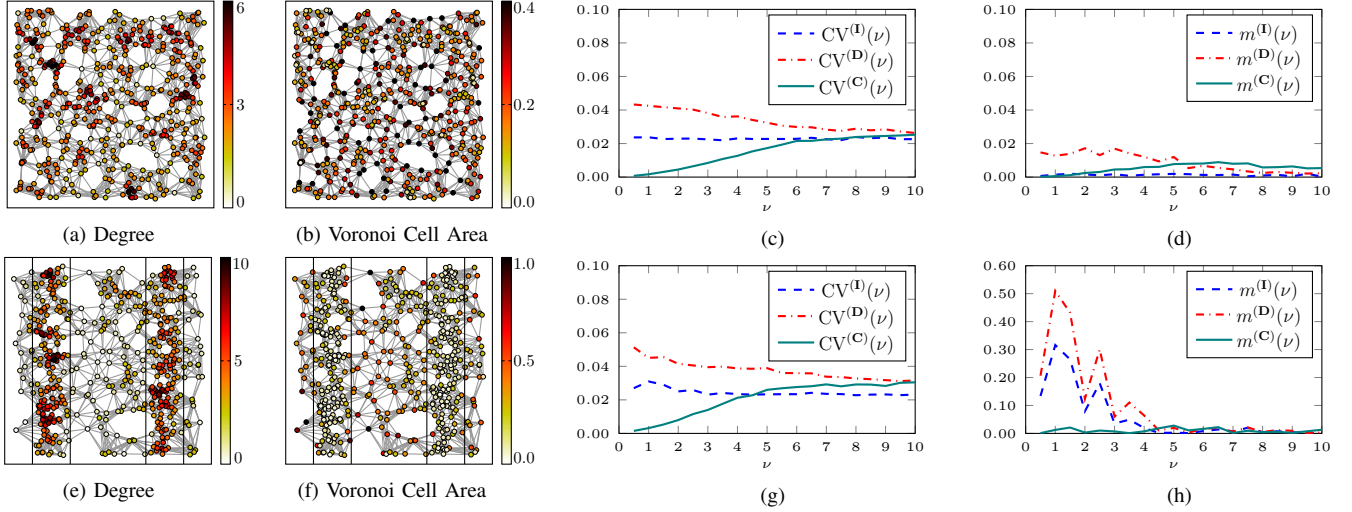
Fig. 5. Study of the graph energy of single cosine continuous signal $\widetilde{s_\nu}(x, y; \varphi) = \cos(2\pi\nu x_i + \varphi)$ sampled uniformly (a)-(d) and non-uniformly (e)-(h) as $s_\nu(i; \varphi) = \widetilde{s_\nu}(x_i, y_i; \varphi)$. Example sampling with 500 samples (vertices) showing vertex degree ((a) and (e)), and Voronoi cell area ((b) and (f)) with colors. (c)-(d) and (g)-(h) 1000 samples (vertices), with results averaged over 500 sampling (graph) realizations. (c) and (g) coefficient of variation $\text{CV}^{(\mathbf{Q})}(\nu)$ of the signal energy depending on the continuous frequency ν. (d) and (h) maximum absolute deviation $m^{(\mathbf{Q})}(\nu)$ of the normalized mean signal energy depending on the continuous frequency ν.

signal energy. First, we study the *empirical mean*:

$$\mu^{(\mathbf{Q})}(\nu; \varphi) := \left\langle E_\nu^{(\mathbf{Q})}(\varphi; \mathcal{G}) \right\rangle_{\mathcal{G}} = \frac{1}{N_\mathcal{G}} \sum_{n=1}^{N_\mathcal{G}} E_\nu^{(\mathbf{Q})}(\varphi; \mathcal{G}_n),$$

with $N_\mathcal{G}$ the number of sampling realizations. It is interesting to study how the mean varies depending on φ and ν in order to check whether the graph signal energy remains constant, given that the energy of the continuous signals being sampled is constant. The first thing we observe is that the averaging over φ, *i.e.*, over many signals of equal continuous frequency, yields the same average mean for all continuous frequencies. However, $\mu^{(\mathbf{Q})}(\nu; \varphi)$ does depend on φ. To show this, we use the *maximum absolute deviation* of the normalized mean $\bar{\mu}^{(\mathbf{Q})}(\nu; \varphi) := \mu^{(\mathbf{Q})}(\nu; \varphi)/\langle\mu^{(\mathbf{Q})}(\nu; \varphi)\rangle_\varphi$:

$$m^{(\mathbf{Q})}(\nu) := \max_\varphi \left| 1 - \bar{\mu}^{(\mathbf{Q})}(\nu; \varphi) \right|.$$

Use of the normalized mean is necessary to be able to compare the various choices of $\mathbf{Q}$ since they can yield quite different average mean $\langle\mu^{(\mathbf{Q})}(\nu; \varphi)\rangle_\varphi$.

However, the mean is only characterizing the bias of the graph signal energy approximation to the continuous signal energy. We also need to characterize the variance of this estimator. To that end, we consider the *empirical standard deviation* of the graph signal energy:

$$\sigma^{(\mathbf{Q})}(\nu; \varphi) := \sqrt{\frac{1}{N_\mathcal{G} - 1} \sum_{n=1}^{N_\mathcal{G}} \left(E_\nu^{(\mathbf{Q})}(\varphi; \mathcal{G}_n) - \mu^{(\mathbf{Q})}(\nu; \varphi) \right)^2}.$$

This quantity shows how much the actual sampling being performed influences the signal energy estimator. Since the mean energy is influenced by the choice of $\mathbf{Q}$, we need to normalize the standard deviation to be able to compare results between various choices of $\mathbf{Q}$. This yields the *coefficient of variation*, and we report its maximum over all signals of equal

(continuous) frequency:

$$\text{CV}^{(\mathbf{Q})}(\nu) := \max_\varphi \frac{\sigma^{(\mathbf{Q})}(\nu; \varphi)}{\mu^{(\mathbf{Q})}(\nu; \varphi)},$$

from which we can study the variance of the graph signal energy estimator depending on the continuous frequency.

We experiment here with three choices for $\mathbf{Q}$. The first one a standard GFT from the literature, corresponding to the ℓ_2-norm: $\mathbf{Q} = \mathbf{I}$. The second one corresponds to the random-walk Laplacian with $\mathbf{Q} = \mathbf{D}$. The third one is our novel approach based on the Voronoi cell area inner product $\mathbf{Q} = \mathbf{C}$ (see Sec. IV-D).

Results for $\mathcal{G}^{(\text{U})}$ and $\mathcal{G}^{(\text{NU})}$ are given in Fig. 5. The coefficient of variation shows here the strong advantage of using the Voronoi cell area inner product: $\text{CV}^{(\mathbf{C})}(\nu)$ is very small in the lower continuous spectrum, and almost zero (see Figs. 5(c) and 5(g)). In other words, for signals that do not vary too quickly compared to the sampling density, *the $\mathbf{C}$-norm gives an estimated energy with small variance, removing the influence of the particular arrangement of sensors*. This is true for both the uniform distribution and the non-uniform distribution. Both the dot product $\mathbf{Q} = \mathbf{I}$ and the degree norm $\mathbf{Q} = \mathbf{D}$ have larger variance here.

Finally, the maximum absolute deviation on Figs. 5(d) and 5(h) shows again the advantage of the $\mathbf{C}$-norm. Indeed, considering several signals of equal continuous frequencies, these should have equal average energy. While this is the case for both the dot product and the $\mathbf{C}$-norm in the lower spectrum for a uniform sampling, using a non-uniform sampling yields a strong deviation for the dot product. The $\mathbf{C}$-norm appears again as a good approximation to the continuous energy with small deviation between signals of equal frequencies, and between samplings.

Remark 2. *In [13], [34], the authors advocate for the use of the random walk Laplacian as a good approximation for*

the continuous (Laplace-Beltrami) operator of a manifold. We showed here that this case is not ideal when it comes to variance of the signal energy estimator, but in those communications, the authors actually normalize the weights of the graph using the degree matrix prior to computing the random walk Laplacian, thus working on a different graph. If $\widetilde{\mathbf{A}}$ is the adjacency prior its normalization, and $\widetilde{\mathbf{D}}$ the associated degree matrix, then $\mathbf{A} = \widetilde{\mathbf{D}}^{-1}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-1}$. This normalization is important when we look at the degree. Indeed, without it, we see on Fig. 5(a)-(b) and 5(e)-(f) that degree and Voronoi cell area evolve in opposed directions: large degrees correspond to small areas. Pre-normalizing by the degree corrects this. Using this normalization leads to better results with respect to the energy (omitted here to save space), however, Voronoi cells still yield the best results.

These results show that it is possible to better account for irregularity in the definition of the energy of a graph signal: The Voronoi cell area norm yields very good results when it comes to analysing a signal lying in a Euclidean space (or on a manifold) independently of the sampling performed.

VI. Conclusion

We showed that it is possible to define an orthonormal graph Fourier transform with respect to any inner product. This allows in particular to finely account for an irregular graph structure. Defining the associated graph filters is then straightforward using the fundamental matrix of the graph. Under some conditions, this led to a fundamental matrix that is easy to compute and efficient to use. We also showed that the literature on graph signal processing can be interpreted with this graph Fourier transform, often with the dot product as inner product on graph signals. Finally, we showed that we are able to obtain promising results for sensor networks using the Voronoi cell areas inner product.

This work calls for many extensions, and giving a complete list of them would be too long for this conclusion. At this time, we are working on the sensor network case to obtain graph signal energies with even less variance, especially once the signals are filtered. We are also working on studying bilateral filters, and image smoothers in general [35], and extending them to graph signals with this new setting. We explore also an extension of the study of random graph signals and stationarity with respect to this new graph Fourier transform. Finally, many communications on graph signal processing can be re-interpreted using alternative inner product that will give more intuitions on the impact of irregularity.

Acknowledgement

Appendix

A. Low Pass Filtering of an Indicator Function

We now show an example to compare the quality of different low pass approximations to the normalized cluster indicator function, under different GFT definitions. This is motivated by the fact that spectral clustering techniques make use of the low frequencies of the graph spectrum. Fig. 6 shows these approximations in the vertex domain. Note that these correspond to the **Q**-MSE plots of Fig. 3.

Comparing results for $l = 1$ and $l = 5$ we note that for both $\mathbf{L}$ and $\mathcal{L}$, the low pass approximation is good, and significantly better than that achieved based on the GFT corresponding to graph total variation. However, the $(\mathbf{L}, \mathbf{I})$-GFT suffers from isolated vertices ($l = 1$ case on Fig. 6), and the approximation needs more spectral components for an output closer to the indicator function ($l = 5$ case), while the $(\mathcal{L}, \mathbf{I})$-GFT is clearly biased with the degree (smaller amplitudes on the vertices of the cluster boundaries).

This confirms that the indicator function approximation (and thus spectral clustering performance) is better for the $(\mathbf{L}, \mathbf{D})$-GFT, which does not have the bias towards isolated vertices shown by the $(\mathbf{L}, \mathbf{I})$-GFT (impulses on those isolated vertices are associated to lowpass signal, while they are associated to the graph frequency 1 for the $(\mathbf{L}, \mathbf{D})$-GFT), and without the bias of the first Fourier mode towards vertex degrees of the $(\mathcal{L}, \mathbf{I})$-GFT.

References

[1] Benjamin Girault, "Stationary Graph Signals using an Isometric Graph Translation," in *Signal Processing Conference (EUSIPCO), 2015 Proceedings of the 23rd European*. IEEE, 2015.

[2] Siheng Chen, Yaoqing Yang, Christos Faloutsos, and Jelena Kovačević, "Monitoring Manhattan's Traffic at 5 Intersections?," in *2016 IEEE Global Conference on Signal and Information Processing*, Washington D.C., USA, Dec 2016.

[3] Ronan Hamon, Pierre Borgnat, Patrick Flandrin, and Céline Robardet, "Networks as signals, with an application to a bike sharing system," in *2013 IEEE Global Conference on Signal and Information Processing*, Dec 2013, pp. 611–614.

[4] Ronan Hamon, *Analysis of temporal networks using signal processing methods : Application to the bike-sharing system in Lyon*, Theses, Ecole normale supérieure de lyon - ENS LYON, Sept. 2015.

[5] Benjamin Girault, Paulo Gonçalves, and Éric Fleury, "Graphe de contacts et ondelettes : étude d'une diffusion bactérienne," in *Proceedings of Gretsi 2013*, Sept. 2013.

[6] Jiun-Yu Kao, Antonio Ortega, and Shrikanth S Narayanan, "Graph-based approach for motion capture data representation and analysis," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2061–2065.

[7] Mikhail Belkin and Partha Niyogi, "Towards a theoretical foundation for Laplacian-based manifold methods," *Journal of Computer and System Sciences*, vol. 74, no. 8, pp. 1289–1308, 2008.

[8] Samuel I Daitch, Jonathan A Kelner, and Daniel A Spielman, "Fitting a graph to vector data," in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 201–208.

[9] Xiaowen Dong, Dorina Thanou, Pascal Frossard, and Pierre Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6160–6173, 2016.

[10] Hilmi E. Egilmez, Eduardo Pavez, and Antonio Ortega, "Graph Learning from Data under Laplacian and Structural Constraints," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 825–841, Sept 2017.

[11] Bastien Pasdeloup, Vincent Gripon, Grégoire Mercier, Dominique Pastor, and Michael G Rabbat, "Characterization and inference of graph diffusion processes from observations of stationary signals," *IEEE Transactions on Signal and Information Processing over Networks*, 2017.

[12] Jonathan Mei and José MF Moura, "Signal processing on graphs: Causal modeling of unstructured data," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2077–2092, 2017.
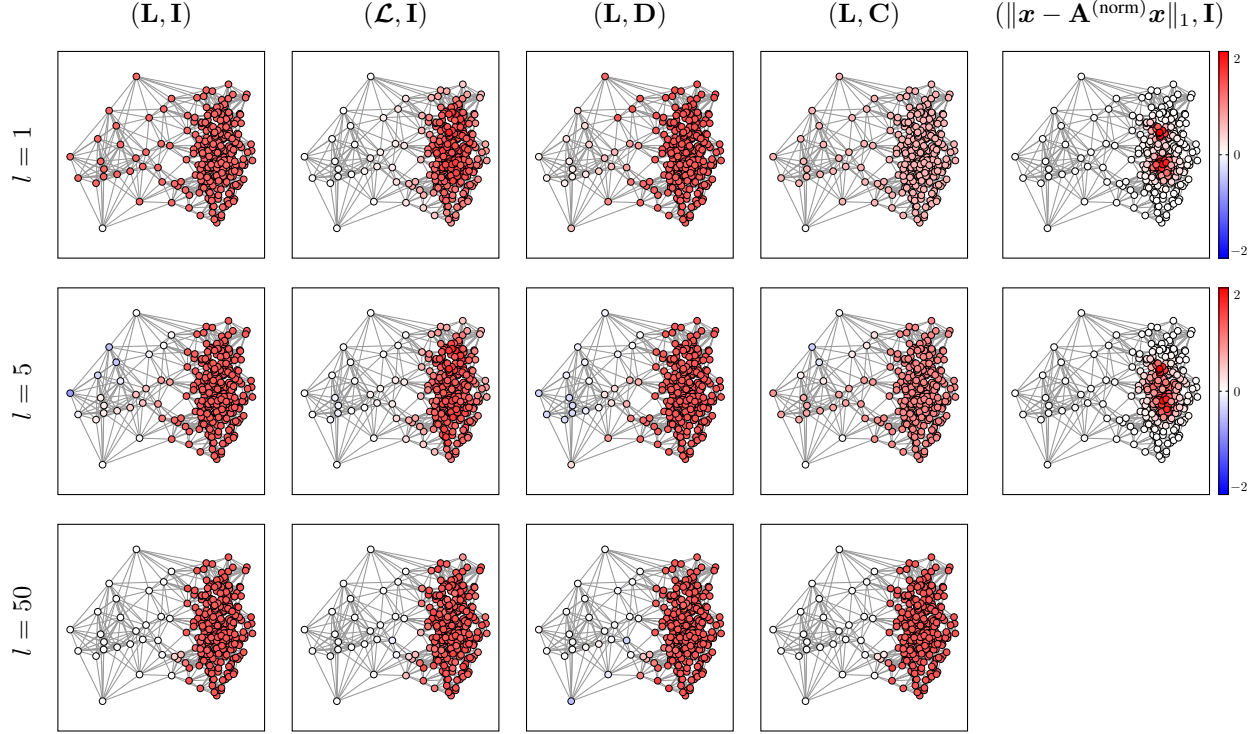
Fig. 6. Output of the ideal low pass filter $\mathbf{H}_l$ applied to the normalized indicator function $\mathbf{h}_{\text{blue}}^{(\mathbf{Q})}$. $\mathbf{H}_l$ keeps only the first $l+1$ spectral components (*i.e.*, up to the graph frequency λ_l). Note that $\mathbf{H}_l$ depends on the $(\Delta, \mathbf{Q})$-GFT.

[13] Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg, "Graph Laplacians and their Convergence on Random Neighborhood Graphs.," *Journal of Machine Learning Research*, vol. 8, pp. 1325–1368, 2007.

[14] Aliaksei Sandryhaila and José M. F. Moura, "Discrete Signal Processing on Graphs," *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1644–1656, 2013.

[15] David I. Shuman, Sunil K. Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst, "The Emerging Field of Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains.," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[16] Benjamin Girault, Paulo Gonçalves, and Éric Fleury, "Translation on Graphs: An Isometric Shift Operator," *Signal Processing Letters, IEEE*, vol. 22, no. 12, pp. 2416–2420, Dec 2015.

[17] Benjamin Girault, *Signal Processing on Graphs - Contributions to an Emerging Field*, Phd thesis, Ecole normale supérieure de lyon - ENS LYON, Dec. 2015.

[18] Aliaksei Sandryhaila and José M. F. Moura, "Discrete Signal Processing on Graphs: Frequency Analysis," *Signal Processing, IEEE Transactions on*, vol. 62, no. 12, pp. 3042–3054, June 2014.

[19] Stefania Sardellitti, Sergio Barbarossa, and Paolo Di Lorenzo, "On the Graph Fourier Transform for Directed Graphs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 796–811, 2017.

[20] Aamir Anis, Akshay Gadde, and Antonio Ortega, "Efficient Sampling Set Selection for Bandlimited Graph Signals Using Graph Spectral Proxies.," *IEEE Trans. Signal Processing*, vol. 64, no. 14, pp. 3775–3789, 2016.

[21] Gene H. Golub and Charles F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, 3rd edition, 1996.

[22] Rasoul Shafipour, Ali Khodabakhsh, Gonzalo Mateos, and Evdokia Nikolova, "A Digraph Fourier Transform With Spread Frequency Components," May 2017.

[23] David I. Shuman, Benjamin Ricaud, and Pierre Vandergheynst, "Vertex-frequency analysis on graphs," *Applied and Computational Harmonic Analysis*, vol. 40, no. 2, pp. 260–291, 2016.

[24] David K. Hammond, Pierre Vandergheynst, and Rémi Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[25] Alan V. Oppenheim and Ronald W. Schafer, *Discrete-time Signal Processing*, Pearson, 3rd edition, 2013.

[26] Ulrike von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[27] Mikhail Belkin and Partha Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[28] Sunil K Narang and Antonio Ortega, "Compact Support Biorthogonal Wavelet Filterbanks for Arbitrary Undirected Graphs," *IEEE Transactions on Signal Processing*, vol. 61, no. 19, pp. 4673–4685, 2013.

[29] Akshay Gadde, Sunil K. Narang, and Antonio Ortega, "Bilateral filter: Graph spectral interpretation and extensions," in *2013 IEEE International Conference on Image Processing*, Sept 2013, pp. 1222–1226.

[30] Xianming Liu, Gene Cheung, Xiaolin Wu, and Debin Zhao, "Random Walk Graph Laplacian-Based Smoothness Prior for Soft Decoding of JPEG Images," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 509–524, 2017.

[31] Carlo Tomasi and Roberto Manduchi, "Bilateral Filtering for Gray and Color Images," in *6th International Conference on Computer Vision '98, Proceedings of*, Washington, DC, USA, 1998, p. 839.

[32] Benjamin Girault, Shrikanth S. Narayanan, Antonio Ortega, Paulo Gonçalves, and Eric Fleury, "Grasp: A matlab toolbox for graph signal processing.," in *ICASSP*. 2017, pp. 6574–6575, IEEE.

[33] Andrew Y Ng, Michael I Jordan, and Yair Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in neural information processing systems*, MIT Press, Ed., Vancouver, Canada, 2002, pp. 849–856.

[34] Ronald R. Coifman and Stéphane Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 5–30, 2006.

[35] Peyman Milanfar, "A Tour of Modern Image Filtering: New Insights and Methods, Both Practical and Theoretical.," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 106–128, 2013.