



HAL
open science

Approximate fast graph Fourier transforms via multi-layer sparse approximations

Luc Le Magoarou, Rémi Gribonval, Nicolas Tremblay

► **To cite this version:**

Luc Le Magoarou, Rémi Gribonval, Nicolas Tremblay. Approximate fast graph Fourier transforms via multi-layer sparse approximations. 2016. hal-01416110v2

HAL Id: hal-01416110

<https://inria.hal.science/hal-01416110v2>

Preprint submitted on 6 Apr 2017 (v2), last revised 16 Jun 2017 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Approximate fast graph Fourier transforms via multi-layer sparse approximations

Luc Le Magoarou, Rémi Gribonval, *Fellow, IEEE*, Nicolas Tremblay

Abstract—The Fast Fourier Transform (FFT) is an algorithm of paramount importance in signal processing as it allows to apply the Fourier transform in $\mathcal{O}(n \log n)$ instead of $\mathcal{O}(n^2)$ arithmetic operations. Graph Signal Processing (GSP) is a recent research domain that generalizes classical signal processing tools, such as the Fourier transform, to situations where the signal domain is given by any arbitrary graph instead of a regular grid. Today, there is no method to rapidly apply graph Fourier transforms. We propose in this paper a method to obtain approximate graph Fourier transforms that can be applied rapidly and stored efficiently. It is based on a greedy approximate diagonalization of the graph Laplacian matrix, carried out using a modified version of the famous Jacobi eigenvalues algorithm. The method is described and analyzed in detail, and then applied to both synthetic and real graphs, showing its potential.

Index Terms—Graph signal processing, Fast Fourier Transform, Greedy algorithms, Jacobi eigenvalues algorithm, Sensor networks.

I. INTRODUCTION

GRAPHS are powerful mathematical objects used to model pairwise relationships between elements of a set. Graph theory has been extensively developed since the eighteenth century, and has found a variety of applications, ranging from biology to computer science or linguistics [1].

Recently, methods have been developed to analyze and process signals defined over the vertices of a graph [2], [3], instead of over a regular grid, as is classically assumed in discrete signal processing. The starting point of graph signal processing is to define a Fourier transform, via an analogy with classical signal processing. Depending on the preferred analogy, there exists several definitions of graph Fourier transforms [2], [3]. Following [2], we choose in this paper to define the graph Fourier basis as the eigenvector matrix of the graph Laplacian operator \mathbf{L} (whose precise definition is given in Section II-A). In a graph with n vertices, $\mathbf{L} \in \mathbb{R}^{n \times n}$ and

$$\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \quad (1)$$

where $\mathbf{U} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix whose columns are the graph Fourier modes and $\mathbf{\Lambda} \in \mathbb{R}^{n \times n}$ is a non-negative diagonal matrix whose diagonal entries correspond to the squared graph frequencies.

The graph Fourier matrix \mathbf{U} being non-sparse in general, applying it costs $\mathcal{O}(n^2)$ arithmetic operations. In the classical signal processing case, the well-known Fast Fourier Transform

(FFT) [4] allows to apply the Fourier transform in only $\mathcal{O}(n \log n)$ arithmetic operations. The FFT is a fast linear algorithm [5], which implies that the classical Fourier matrix can be factorized into sparse factors, as discussed in [6]. Given that classical signal processing is equivalent to graph signal processing on the ring graph, it is natural to wonder if this kind of factorization can be generalized to other graphs.

We proposed in a previous work a method to obtain computationally efficient¹ approximations of matrices, based on multi-layer sparse factorizations [6]. The method amounts to approximate a matrix of interest $\mathbf{A} \in \mathbb{R}^{p \times q}$ with a product of sparse matrices, as

$$\mathbf{A} \approx \mathbf{S}_J \dots \mathbf{S}_1, \quad (2)$$

where the matrices $\mathbf{S}_1, \dots, \mathbf{S}_J$ are sparse, allowing for cheap storage and multiplication. We applied this method to the graph Fourier matrix \mathbf{U} of various graphs in [7], in order to get approximate Fast Graph Fourier Transforms (FGFT). The method showed good potential. However in the context of graph signal processing, this approach suffers from at least two limitations:

- (L1) It requires a full diagonalization of the graph Laplacian \mathbf{L} before it can be applied. Indeed, the method takes the graph Fourier matrix \mathbf{U} as input. Performing this diagonalization costs $\mathcal{O}(n^3)$ arithmetic operations, which is prohibitive if n is large.
- (L2) It provides non-orthogonal approximate FGFTs. Indeed, the details of the method make it difficult to get sparse *and* orthogonal factors $\mathbf{S}_1, \dots, \mathbf{S}_J$. This leads to approximate graph Fourier transforms that are not easily invertible, which can be a problem for applications where signal reconstruction is needed.

We propose in this paper another method, that does not suffer from these limitations, to obtain approximate FGFTs. In order to overcome (L1), we consider directly the Laplacian matrix \mathbf{L} as input. To overcome (L2), we look for factors $\mathbf{S}_1, \dots, \mathbf{S}_J$ constrained to be in a set of sparse and orthogonal matrices built with Givens rotations [8] (as explained in detail in section III). The proposed method amounts to an approximate diagonalization of the Laplacian matrix \mathbf{L} , as

$$\mathbf{L} \approx \mathbf{S}_1 \dots \mathbf{S}_J \hat{\mathbf{A}} \mathbf{S}_J^T \dots \mathbf{S}_1^T, \quad (3)$$

where the matrices $\mathbf{S}_1, \dots, \mathbf{S}_J$ are both sparse *and* orthogonal. The product $\hat{\mathbf{U}} = \mathbf{S}_1 \dots \mathbf{S}_J$ constitutes an efficient approximate graph Fourier matrix and $\hat{\mathbf{A}}$ is a diagonal

¹An $m \times n$ matrix is “efficient” if it is associated with a linear algorithm involving strictly less than mn scalar multiplications.

Luc Le Magoarou (luc.le-magoarou@inria.fr) and Rémi Gribonval (remi.gribonval@inria.fr) are both with Inria, Rennes, France, PANAMA team. Nicolas Tremblay is with CNRS, GIPSA-lab, Grenoble, France. This work was supported in part by the European Research Council, PLEASE project (ERC-StG- 2011-277906).

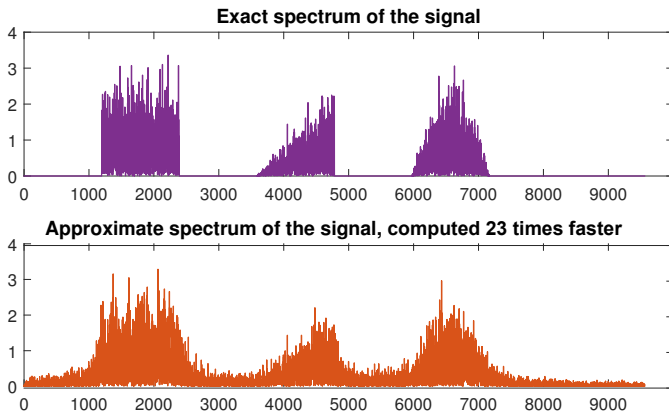


Fig. 1. Example of an approximate spectrum computed with our method on a real-world sensor graph. Details in Section V.

matrix whose diagonal entries are approximations of the squared graph frequencies.

Contributions. Given a graph Laplacian matrix, the main objective of this paper is to find an approximate graph Fourier matrix $\hat{\mathbf{U}}$ that both i) approximately diagonalizes the Laplacian and ii) is computationally efficient. The proposed method is a greedy strategy that amounts to truncating and slightly modifying the famous Jacobi eigenvalue algorithm [9]. Note that it could in principle be applied to any symmetric matrix (covariance matrix, normalized Laplacian, etc.), but the focus of the present paper is on the graph combinatorial Laplacian. Indeed, while a general symmetric matrix has no reason to be well approximated with a limited number of Givens factors, graph Laplacians are shown empirically to have this interesting property. The proposed method is compared experimentally with the direct approximation method of [7] on various graphs, showing that it is much faster while obtaining good results. In fact, we obtain approximate FGFTs of complexity $\mathcal{O}(n \log n)$ exhibiting constant error for growing n . Moreover, we briefly investigate the links between the graph structure and the quality of diagonalization. Also, a discussion on possible use cases of the method is undertaken and further experimental validations on both synthetic and real-world sensor networks are conducted, showing an interesting compromise between computational complexity and accuracy. Figure 1 illustrates one such experiment further developed in Section V. Finally, the method is applied to graph signal filtering, and compared to the usual method based on polynomial approximations. We show that when the filter function is not well approximated by a polynomial (such as low-pass filters with a steep cut-off), our method performs at least as well, and sometimes outperforms the polynomial approximation method.

The remainder of the paper is organized as follows. The problem is formulated in Section II, and an approach to solve it is given in Section III. The proposed method is then evaluated, compared to the direct approximation method and its links with the graph structure are briefly investigated in Section IV. Then, an application to sensor networks is

proposed in Section V. Finally, an application to graph signal filtering is presented in Section VI.

II. PROBLEM FORMULATION

In this section, we give a concrete formulation of the main problem addressed in this paper. We first set up the notations and conventions used, before presenting the objective in detail. We then discuss the advantages in terms of computational complexity expected from the method and end with a presentation of the related work.

A. Notations and conventions

General notations. Matrices are denoted by bold upper-case letters: \mathbf{A} ; vectors by bold lower-case letters: \mathbf{a} ; the i th column of a matrix \mathbf{A} by: \mathbf{a}_i ; its entry located at the i th row and j th column by: a_{ij} . Sets are denoted by calligraphic symbols: \mathcal{A} , and we denote by $\delta_{\mathcal{A}}$ the indicator function of the set \mathcal{A} in the optimization sense ($\delta_{\mathcal{A}}(x) = 0$ if $x \in \mathcal{A}$, $\delta_{\mathcal{A}}(x) = +\infty$ otherwise). The standard vectorization operator is denoted $\text{vec}(\cdot)$. The ℓ_0 -pseudonorm is denoted $\|\cdot\|_0$ (it counts the number of non-zero entries), $\|\cdot\|_F$ denotes the Frobenius norm, and $\|\cdot\|_2$ the spectral norm. By abuse of notations, $\|\mathbf{A}\|_0 = \|\text{vec}(\mathbf{A})\|_0$. \mathbf{Id} denotes the identity matrix.

Graph Laplacian. We consider in this paper undirected weighted graphs, denoted $\mathcal{G} \triangleq \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$, where \mathcal{V} represents the set of vertices (otherwise called nodes), $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges, and \mathbf{W} is the weighted adjacency matrix of the graph. We denote $n \triangleq |\mathcal{V}|$ the total number of vertices and the adjacency matrix $\mathbf{W} \in \mathbb{R}^{n \times n}$ is symmetric and such that $w_{ij} = w_{ji}$ is non-zero only if $(i, j) \in \mathcal{E}$ and represents the strength of the connection between nodes i and j . We define the degree matrix $\mathbf{D} \in \mathbb{R}^{n \times n}$ as a diagonal matrix with $\forall i, d_{ii} \triangleq \sum_{j=1}^n w_{ij}$, and the combinatorial Laplacian matrix $\mathbf{L} \triangleq \mathbf{D} - \mathbf{W}$ (we only consider this type of Laplacian matrix in this paper, and hereafter simply call it the Laplacian).

B. Objective

Our goal is to approximately diagonalize the Laplacian \mathbf{L} with an efficient approximate eigenvector matrix $\hat{\mathbf{U}} = \mathbf{S}_1 \dots \mathbf{S}_J$, where the factors $\mathbf{S}_1, \dots, \mathbf{S}_J \in \mathbb{R}^{n \times n}$ are sparse and orthogonal. Using the Frobenius norm to measure the quality of approximation, this objective can be stated as the following optimization problem:

$$\begin{aligned} \underset{\hat{\mathbf{A}}, \mathbf{S}_1, \dots, \mathbf{S}_J}{\text{minimize}} \quad & \left\| \mathbf{L} - \mathbf{S}_1 \dots \mathbf{S}_J \hat{\mathbf{A}} \mathbf{S}_J^T \dots \mathbf{S}_1^T \right\|_F^2 \\ & + \sum_{j=1}^J \delta_{\mathcal{S}}(\mathbf{S}_j) + \delta_{\mathcal{D}}(\hat{\mathbf{A}}), \end{aligned} \quad (\text{DP})$$

where \mathcal{D} is a set of diagonal matrices and \mathcal{S} is a set of sparse and orthogonal matrices. The set of sparse matrices $\{\mathbf{A} \in \mathbb{R}^{n \times n}, \|\mathbf{A}\|_0 \leq s\}$ and the set of orthogonal matrices $\{\mathbf{A} \in \mathbb{R}^{n \times n}, \mathbf{A}^T \mathbf{A} = \mathbf{Id}\}$ are both easy to project onto, but their intersection is not. This is the reason why approaches using projected gradient descent similar to what is proposed in [6] cannot be used to solve this problem. Instead, we propose a greedy strategy using Givens rotations that is explained in detail in Section III.

C. Relative Complexity Gain

An efficient approximate Fourier matrix $\hat{\mathbf{U}} = \mathbf{S}_1 \dots \mathbf{S}_J$ is beneficial in terms of complexity provided its *relative complexity* with respect to the exact Fourier matrix \mathbf{U} is small. In order to quantify this statement, we introduce a measure of the relative complexity in the following definition.

Definition II.1. *The Relative Complexity Gain (abbreviated RCG) is the ratio between the number of non-zero entries in the Fourier matrix \mathbf{U} and the total number of non-zero entries in the multi-layer sparse factorization of the efficient approximate Fourier matrix $\hat{\mathbf{U}} = \mathbf{S}_1 \dots \mathbf{S}_J$:*

$$RCG \triangleq \frac{\|\mathbf{U}\|_0}{\sum_{j=1}^J \|\mathbf{S}_j\|_0}. \quad (4)$$

The relative complexity gain is a good measure of the *theoretical* benefit of using the efficient approximation $\hat{\mathbf{U}} = \mathbf{S}_1 \dots \mathbf{S}_J$ in place of the exact Fourier matrix \mathbf{U} , in several aspects of its manipulation, such as storage and multiplication (see [6] for a detailed explanation on this). For example, having $RCG = 10$ means that the multiplication of any vector $\mathbf{x} \in \mathbb{R}^n$ by the efficient approximation or its transpose will require ten times less scalar multiplications than a multiplication by \mathbf{U} . It also means that storing the approximate FGFT will require roughly ten times less memory than the original Fourier transform, as will be demonstrated on the DC road graph which Fourier matrix requires 700 MB while its approximate FGFT only requires 60MB. However, this does not necessarily mean that the multiplication will be ten times faster *in practice*, as we will see in the experimental part of this paper (Sections IV to VI).

D. Related work

Computationally efficient approximations of operators have been studied quite extensively in various domains over the last years. Applications of such methods range from dictionary learning [10], [11], [12], [13], to Non-negative Matrix Factorization (NMF) [14] or compression of integral operators arising from discretized continuous kernels [15], [16], [17], [18]. For a brief description of these various methods in a unified way, see [6].

We focus here on efficient *orthogonal* approximation of operators. In this context, the idea to use a multi-layer product of sparse and orthogonal matrices (otherwise called elementary rotations) in order to make complexity savings has been investigated in several approaches over the last few years. Very recently, it has been proposed [19] to learn an efficient orthogonal dictionary being the product of a low number of Householder reflectors [20] (another type of sparse and orthogonal matrices, different from Givens rotations).

Several approaches have also considered diagonalizing symmetric matrices with a product of sparse and orthogonal matrices. The first one comes from statistics [21], and introduces a method to design an efficient orthogonal transform to decorrelate some data, based on an empirical similarity matrix that is to be approximately diagonalized with a low number of Givens rotations [8] whose supports are organized as a

tree. Another approach [22] aims at estimating the covariance matrix of some data, under the constraint that its eigenvector matrix is the product of a low number of Givens rotations, with a maximum likelihood formulation involving determinants. In the context of video compression, it has also been proposed [23], [24] to optimize the coding gain by an approximate diagonalization of the covariance matrix with Givens rotations, with a criterion involving logarithms.

More related to graphs and to the method we propose in this paper, two recent approaches called Multiresolution Matrix Factorization (MMF) [25] and parallel Multiresolution Matrix Factorization (pMMF) [26] propose to design a wavelet transform on graph via a greedy factorization of the graph Laplacian with the help of general elementary rotations being constrained to comply with certain multiresolution axioms. For the pMMF approach, the factorization algorithm is parallelized using a clustering algorithm that cuts down the matrix to factorize into several pieces, and is applied several times along the factorization algorithm. MMF and pMMF have been used for various tasks such as matrix compression and preconditioning.

A complementary approach to the design of fast graph Fourier transforms is presented in [27], where certain structured graphs are viewed as the product of two (or more) subgraphs (the whole graph being called a *product graph*). The product structure can be used to reduce the Fourier transform complexity. The two approaches are perfectly compatible, since the Fourier transform on the product graph can be further accelerated by computing approximate FGFTs for the Laplacian matrix of one (or more) of the subgraphs. For example if a graph represents the variations along time of a signal defined on a sensor network, it can be seen as the product of a sensor graph and a time series. The Laplacian matrix of the sensor graph can be approximately diagonalized, while the product structure can still be leveraged to decouple it from the time variations and accelerate the overall Fourier transform.

The approach we propose in this paper amounts to an approximate diagonalization with a product of sparse and orthogonal matrices. However it differs from [21], [22] since it departs from the probabilistic formulation that induces different update rules, and is applied to graph Laplacians rather than covariances or similarity matrices. On the other hand, it is closer to the MMF approaches [25], [26], but the main objectives and targeted applications differ. Indeed, we are only interested here in having an efficient linear transformation that approximately diagonalizes the graph Laplacian, so that we do not impose constraints during the approximate diagonalization to achieve some form of multiresolution. In that sense, it can be seen as simpler.

In summary, and as will be made clearer in the next section, the method we propose here can be seen as an approximate diagonalization using a *modified (in particular, truncated) version of the Jacobi eigenvalues algorithm* [9], [28], where the truncation allows for both a faster computation of the approximate Fourier transform and a faster multiplication by this approximate transform, as demonstrated numerically in Section IV. Actually, the novelty and genuine contributions

D. Parallel truncated Jacobi algorithm

The Relative Complexity Gain RCG, as defined in Section II, is only a theoretical measure of complexity gain, it corresponds to a ratio between the number of scalar multiplication required to compute the product with the true Fourier matrix \mathbf{U} and the number of scalar multiplications required to compute the product with its efficient approximation $\hat{\mathbf{U}} = \mathbf{S}_1 \dots \mathbf{S}_J$. However the actual time gain, that is observed when using the efficient approximation to multiply vectors, is related but not directly proportional to the number of required scalar multiplications. It depends indeed on other factors, such as the implementation of the matrix/vector product, involving or not parallelization. For example, in MATLAB, the product of a vector \mathbf{x} by the dense Fourier matrix \mathbf{U} can be faster than the product with the efficient approximation $\hat{\mathbf{U}} = \mathbf{S}_1 \dots \mathbf{S}_J$ even if the RCG is large. This is because the dense matrix/vector product in MATLAB involves parallelism whereas the product with the efficient approximation is done sequentially ($\mathbf{y} \leftarrow \mathbf{S}_J \mathbf{x}$, $\mathbf{y} \leftarrow \mathbf{S}_{J-1} \mathbf{y}$, \dots , $\mathbf{y} \leftarrow \mathbf{S}_1 \mathbf{y}$). Since some of the Givens rotations could be applied at the same time in parallel (as soon as their supports are disjoint), this leaves room for improvement.

In order to overcome this limitation, we propose to slightly modify the algorithm of Figure 2 to choose at each step of the approximate diagonalization not only one Givens rotation, but $\frac{n}{2}$ Givens rotations that can be applied in parallel². More formally, for an approximation with J Givens rotations, this corresponds to choosing $K \triangleq \lceil \frac{2J}{n} \rceil$ factors, where each of the sparse factors \mathbf{S}_k belongs to the set \mathcal{R}_P of matrices made of $\frac{n}{2}$ Givens rotations with mutually disjoint supports. Elements of this set are matrices of the form

$$\mathbf{S} = \mathbf{P} \begin{pmatrix} \mathbf{R}_1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{R}_{\frac{n}{2}} \end{pmatrix} \mathbf{P}^T,$$

where \mathbf{P} is a permutation matrix and $\forall i \in \{1, \dots, \frac{n}{2}\}$ we have

$$\mathbf{R}_i = \begin{pmatrix} \cos \theta_i & -\sin \theta_i \\ \sin \theta_i & \cos \theta_i \end{pmatrix}.$$

The algorithm corresponding to this modified version of the approximate diagonalization is given in Figure 4. In practice, the choice for the supports of the parallel Givens rotations at the k th step is done in a similar way as in the algorithm of Figure 2 (by sequentially choosing the greatest entry of \mathbf{L}_k), except that it is necessary to make sure that the support of each Givens rotation be disjoint with the supports of all the Givens rotations chosen previously during the same step. This can be done by:

- 1) Sorting all the nonzero entries of \mathbf{L}_k only once at each step k and put their ordered indices in a list.
- 2) Choosing $\frac{n}{2}$ Givens rotations to put in the k th factor \mathbf{S}_k by going down the list, making sure each chosen Givens rotation has its support disjoint from all the previously chosen ones in the current factor.

²For clarity's sake, we suppose here that n is even. If n is odd, then \mathbf{S} is composed of $\lfloor \frac{n}{2} \rfloor$ Givens rotation and keeps one coordinate untouched.

Input: matrix \mathbf{L} , number J of Givens rotations.

- 1: $\mathbf{L}_1 \leftarrow \mathbf{L}$
- 2: $j \leftarrow 0, k \leftarrow 1$
- 3: **while** $j < J$ **do**
- 4: Choose $\mathbf{S}_k \in \mathcal{R}_P$ (such that $\|\mathbf{S}_k^T \mathbf{L}_k \mathbf{S}_k\|_{\text{offdiag}}^2 < \|\mathbf{L}_k\|_{\text{offdiag}}^2$)
- 5: $\mathbf{L}_{k+1} \leftarrow \mathbf{S}_k^T \mathbf{L}_k \mathbf{S}_k$
- 6: $j \leftarrow j + \frac{n}{2}, k \leftarrow k + 1$
- 7: **end while**
- 8: $\hat{\mathbf{A}} \leftarrow \text{diag}(\mathbf{L}_k), K \leftarrow k$

Output: sparse orthogonal factors $\mathbf{S}_1, \dots, \mathbf{S}_K$; diagonal factor $\hat{\mathbf{A}}$.

Fig. 4. **Parallel truncated Jacobi algorithm:** Approximate diagonalization algorithm with prescribed complexity and parallel Givens rotations.

A similar approach where parallel elementary rotations are chosen is evoked in [25], with supplementary constraints due to multiresolution. The approach proposed here relaxes these constraints. This approach is also different from the parallel MMF (pMMF) factorization method [26], where the main goal is to accelerate the factorization algorithm itself, by clustering the columns/rows of \mathbf{L}_k every few iterations to reduce the cost of finding the support of the Givens rotations.

E. Computational cost of the approaches

The proposed approaches yield approximate Fourier transforms with potentially better complexity than the exact Fourier transform obtained by exact diagonalization. The cost of performing the associated (approximate) diagonalization is also different, we detail it below for each approach.

Exact diagonalization [9]: exact diagonalization of a symmetric matrix costs $\mathcal{O}(n^3)$ (for example using the classical Jacobi algorithm). In theory, faster diagonalization is possible [30], [31], [32], at the price of higher constants and/or stability issues. Once this diagonalization is computed, and since the obtained eigenvector matrix is in general dense, applying it costs $\mathcal{O}(n^2)$. It is known that the Jacobi algorithm converges at least linearly to a diagonal matrix [29].

Truncated Jacobi: the cost of the truncated Jacobi algorithm of Figure 2 is dominated by the quest for the greatest entry of \mathbf{L}_j (line 1 of the algorithm of Figure 3), that costs $\mathcal{O}(\|\mathbf{L}_j\|_0)$ (which is at worst $\mathcal{O}(n^2)$, but can be much smaller if \mathbf{L}_j is sparse). Since matrices \mathbf{L}_j and \mathbf{L}_{j+1} have identical entries except for the lines and columns p and q , one can reuse computations carried out at the previous step for the current step, using a technique inspired by [22] (a full algorithm implementing this technique is given in appendix A). Using this technique, the worst case complexity of the subproblem resolution remains $\mathcal{O}(\|\mathbf{L}_j\|_0)$, but drops down to $\mathcal{O}(n)$ for most iterations (as soon as the selected coordinates p and q at the current iteration are both different from the two selected at the previous iteration). This results in a complete truncated Jacobi costing $\mathcal{O}(n^2 + nJ)$ arithmetic operations in average.

Once the approximate diagonalization is done, applying the resulting $\hat{\mathbf{U}}$ costs $\mathcal{O}(J)$ arithmetic operations, since applying a Givens rotation costs only 4 multiplications and 2 additions.

Parallel truncated Jacobi: its cost is dominated by the sorting done at each step k . Since there are at most $\frac{n(n-1)}{2}$

| | Obtaining \mathbf{U} or $\hat{\mathbf{U}}$ | Applying \mathbf{U} or $\hat{\mathbf{U}}$ |
|------------------------------------|--|---|
| Exact diagonalization [9] | $\mathcal{O}(n^3)$ | $\mathcal{O}(n^2)$ |
| Truncated Jacobi | $\mathcal{O}(n^2 + nJ)$ | $\mathcal{O}(J)$ |
| — with $J = \mathcal{O}(n \log n)$ | $\mathcal{O}(n^2 \log n)$ | $\mathcal{O}(n \log n)$ |
| Parallel truncated Jacobi | $\mathcal{O}(nJ \log n)$ | $\mathcal{O}(J)$ |
| — with $J = \mathcal{O}(n \log n)$ | $\mathcal{O}(n^2 \log^2 n)$ | $\mathcal{O}(n \log n)$ |

TABLE I
COMPLEXITY OF EXACT vs APPROXIMATE DIAGONALIZATIONS.

entries to sort at each step, the cost of the sorting algorithm can be assumed to be $\mathcal{O}(n^2 \log n)$ (the average complexity of most sorting algorithms with r elements being $\mathcal{O}(r \log r)$). The sorting has to be done $K = \mathcal{O}(\frac{J}{n})$ times, which brings the overall complexity of the greedy diagonalization strategy of Figure 4 to $\mathcal{O}(nJ \log n)$.

As with the plain truncated Jacobi case, once the approximate diagonalization is done, applying the resulting $\hat{\mathbf{U}}$ costs $\mathcal{O}(J)$ arithmetic operations. Yet, since $\frac{n}{2}$ Givens rotations have disjoint supports in each sparse factor, computations can be parallelized leading to much faster approximate FGFTs in practice as we will see in Section V.

Summary: A comparison between the computational complexities of the classical method to obtain the graph Fourier transform (an exact diagonalization of the Laplacian) and the approximate diagonalization methods proposed here is given in Table I. For the approximate diagonalizations we upper bound $\|\mathbf{L}_j\|_0$ and $\|\mathbf{L}_k\|_0$ by $\mathcal{O}(n^2)$, since even if in the beginning the Laplacian is sparse, applying Givens rotations rapidly introduces nonzero entries. This comparison indicates that in order for the approximate diagonalization to be cheaper than the exact diagonalization, one should take $J = \mathcal{O}(n^\alpha)$ with $\alpha < 2$, or even $J = \mathcal{O}(n \log n)$ (which leads to a cost of application of the same order as the classical FFT).

Interestingly, for general symmetric matrices, when diagonalization is done with the Jacobi algorithm, it has been empirically observed that $\mathcal{O}(n^2 \log n)$ Givens rotations are typically required to achieve a prescribed tolerance [29], [33]. One of the main contributions of this paper is to show empirically in the next section that, for certain graph Laplacians, the truncated algorithms (of Figures 2 and 4) provide a good approximate diagonalization with only $J = \mathcal{O}(n \log n)$ Givens rotations, thus enabling an efficient approximate FGFT on the considered graph.

IV. EVALUATION OF THE APPROXIMATE FOURIER TRANSFORMS

In this section, we compare the approximate graph Fourier transforms obtained with the methods introduced in this paper to the factorization methods of [7]. We first introduce the various graphs we use for the evaluation. We then present the different approximate FGFT methods we will compare, as well as the different performance measures. We finish by discussing the obtained results.

A. General evaluation

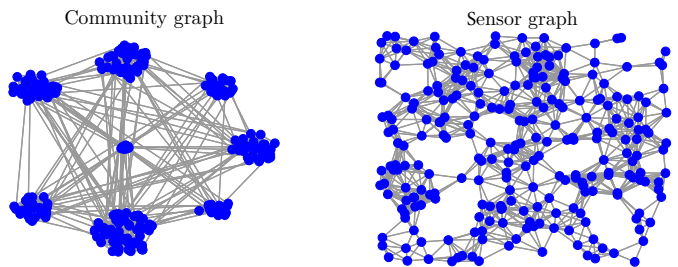


Fig. 5. Examples of a community graph and a sensor graph, of size $n = 256$.

Considered graphs. For this experiment, we consider several families of graphs among the most common ones. All graphs used here are generated with the help of the “Graph Signal Processing” toolbox (GSPBOX) [34].

- **Erdős-Rényi:** a totally random graph where every pair of nodes is connected with a unit weight with probability $p = 0.1$.
- **Community:** a graph of size n made of $\sqrt{n}/2$ communities of random sizes. Every community is itself a graph whose nodes correspond to points whose location is drawn uniformly at random on the unit disk, that are connected with a weight inversely proportional to their distance, if this distance is inferior to a certain threshold. Moreover, random weights between any pair of nodes appear with a probability of $1/n$.
- **Sensor:** a graph of size n where the nodes correspond to points whose location is drawn uniformly at random on the unit square, that are connected with a weight inversely proportional to their distance, if this distance is inferior to a certain threshold, zero otherwise.
- **Ring:** the classical ring graph, where each node is connected to the two adjacent nodes with edges of equal weights.

For all these families of graphs, we take graphs of various sizes n , with $n \in \{128, 256, 512, 1024\}$ nodes. As our purpose here is to qualitatively compare the approximate diagonalization method with the factorization methods of [7] (which do not scale well to large graphs, as they involve a full diagonalization), we had to restrict to these graph sizes that may seem small. We expect the conclusions of this section to hold for much bigger graphs. Examples of graphs used in these experiments are shown on Figure 5.

Computation of the Fourier transforms. We consider for each configuration four approximate graph Fourier transforms of complexity $\mathcal{O}(n \log n)$, thus mimicking the complexity of the classical FFT. The corresponding relative complexity gain RCG is of order $\mathcal{O}(n/\log n)$, and is the same for the all the approximate transforms.

- $\hat{\mathbf{U}}_{\text{PALM}}$: computed with the factorization method of [7], where the true Fourier matrix is hierarchically factorized into K factors (sparse but not orthogonal; sparsity levels described in detail in [7]) using the Proximal Alternating Linearized Minimization (PALM) algorithm [35].
- $\hat{\mathbf{U}}_{\text{PALM}_d}$: computed with the version of the factorization method of [7] whose global optimization steps target a diagonalization cost function (aiming at an efficient ap-

proximation of the Fourier transform that approximately diagonalizes the Laplacian matrix).

- $\hat{\mathbf{U}}_{\text{Givens}}$: computed with the truncated Jacobi algorithm.
- $\hat{\mathbf{U}}_{\text{Givens}/\parallel}$: with the parallel truncated Jacobi algorithm.

Performance measures. How can we measure the accuracy of an approximate graph Fourier transform $\hat{\mathbf{U}}$? For this experiment we choose two performance measures. The first one is the relative error on the calculation of the Fourier basis:

$$\text{err}_c(\hat{\mathbf{U}}) \triangleq \frac{\|\mathbf{U} - \hat{\mathbf{U}}\|_F}{\|\mathbf{U}\|_F}.$$

The second one is the relative error on the diagonalization of the Laplacian:

$$\text{err}_d(\hat{\mathbf{U}}) \triangleq \frac{\|\hat{\mathbf{U}}^T \mathbf{L} \hat{\mathbf{U}}\|_{\text{offdiag}}}{\|\mathbf{L}\|_F}.$$

Results. Average results (over 10 realizations) are shown in Table II. In addition to the performance measures, we also show the factorization time T (in seconds, not taking into account the $\mathcal{O}(n^3)$ time needed to obtain \mathbf{U} for the first two methods). Several comments are in order:

- First, it is very clear that the factorization method of [7] gives approximate FGFTs $\hat{\mathbf{U}}_{\text{PALM}}$ that are better in terms of relative calculation error err_c , compared to all the other methods tested here. Conversely, the diagonalization methods presented here gives approximate FGFTs $\hat{\mathbf{U}}_{\text{Givens}}$ and $\hat{\mathbf{U}}_{\text{Givens}/\parallel}$ that are better in terms of relative diagonalization error err_d . This is true for almost all the tested configurations (except for the Erdős-Rényi graph in dimension 512 and 1024 where the approximate FGFT $\hat{\mathbf{U}}_{\text{PALM}_d}$ gives better diagonalization results). This seems quite logical and is concordant with the cost functions respectively considered by the methods. Indeed, the cost function used for the factorization methods is very close to err_c and the cost function used for the diagonalization methods is very close to err_d . Moreover, performance of $\hat{\mathbf{U}}_{\text{Givens}}$ and $\hat{\mathbf{U}}_{\text{Givens}/\parallel}$ are very close, showing that taking parallel Givens rotations does not decrease performance, while providing approximate FGFTs that can be applied (and obtained) much faster.
- Second, regarding the different graph families considered in the experiment, we see that all methods show in general better results for the graphs “Sensor” and “Community”, and poorer results for the graphs “Erdős-Rényi” and “Ring”. It is quite expected that it is difficult to obtain a good approximate FGFT for the graph “Erdős-Rényi”, since it is totally random and unstructured. However, it is a bit more surprising for the graph “Ring”, that is highly structured, and for which we know a fast Fourier transform exists (which is nothing more than the classical butterfly FFT [4]). This indicates that the optimization techniques used to obtain the approximate FGFTs (both the factorizations and the greedy diagonalizations) are in a way suboptimal, in the sense that they do not attain a global optimum of their cost functions. We discuss links between the graph structure and the diagonalization performance in section IV-B. Moreover, the error measure

err_c is greater than one in certain configurations for the approximate FGFTs $\hat{\mathbf{U}}_{\text{PALM}_d}$, $\hat{\mathbf{U}}_{\text{Givens}}$ and $\hat{\mathbf{U}}_{\text{Givens}/\parallel}$. This can be explained by the fact that the cost functions for these methods do not enforce the true Fourier matrix \mathbf{U} and its approximation to be close to each other in Frobenius norm.

- Third, we notice that for $\hat{\mathbf{U}}_{\text{Givens}}$ and $\hat{\mathbf{U}}_{\text{Givens}/\parallel}$, err_d does not increase with the graph size n , although the relative complexity gain increases (we consider here approximate FGFTs whose complexity is $\mathcal{O}(n \log n)$). In other words, the computational benefit grows with n when using the approximate FGFTs, without increasing the relative error: this behavior is typically what is expected from an approximate fast transform. The same behavior was observed [7] for the factorization method, with approximate FGFTs of higher complexity $\mathcal{O}(n^{1.27})$.
- Last, regarding the factorization time T , $\hat{\mathbf{U}}_{\text{Givens}}$ and $\hat{\mathbf{U}}_{\text{Givens}/\parallel}$ are obtained much faster (often by two orders of magnitude) than $\hat{\mathbf{U}}_{\text{PALM}}$ and $\hat{\mathbf{U}}_{\text{PALM}_d}$. The truncated Jacobi algorithms presented in this paper are thus faster than the methods proposed in [7]. Moreover the parallel truncated Jacobi algorithm of Figure 4 (that gives $\hat{\mathbf{U}}_{\text{Givens}/\parallel}$) is faster than the plain truncated Jacobi algorithm of Figure 2 (that gives $\hat{\mathbf{U}}_{\text{Givens}}$), although its theoretical complexity is higher (see Table I). This may be due to the respective implementations of the two methods, or to the constants before the complexity orders being different. For example, it is three to five times faster depending on the graph type for $n = 1024$.

In summary, the approximate FGFTs $\hat{\mathbf{U}}_{\text{PALM}}$ and $\hat{\mathbf{U}}_{\text{PALM}_d}$ computed by the factorization method are more adapted to situations where the knowledge of the actual Fourier coefficients of a signal \mathbf{x} is important ($\mathbf{U}^T \mathbf{x}$ and $\hat{\mathbf{U}}^T \mathbf{x}$ are close). Conversely, the approximate FGFTs $\hat{\mathbf{U}}_{\text{Givens}/\parallel}$ or $\hat{\mathbf{U}}_{\text{Givens}}$ computed by the diagonalization methods of Figures 2 or 4 are more adapted to situations where the “physical” interpretation of the Fourier transform of a signal \mathbf{x} is important ($\hat{\mathbf{U}}$ is an orthogonal matrix whose columns are approximately eigenvectors of the Laplacian \mathbf{L}). However, the algorithm of Figure 4 shows similar performance but is faster than the algorithm of Figure 2, and it gives approximate FGFTs that can be applied faster: it is thus to be preferred in applications. Moreover, all methods seem to perform better on structured graphs (“Sensor” and “Community”) that are anyway more likely to be encountered in real-world applications.

Note. Hereafter, and given these first results, we concentrate on the performance analysis of the approximate FGFT $\hat{\mathbf{U}}_{\text{Givens}/\parallel}$ computed with the algorithm of Figure 4.

B. Link between structure and diagonalization performance

As mentioned in the previous subsection, it seems difficult to establish a direct link between the graph structure and the results obtained by the different methods in terms of diagonalization and factorization error. Focusing on the approximate FGFT $\hat{\mathbf{U}}_{\text{Givens}/\parallel}$ and the performance measure err_d , we perform here approximate diagonalizations on graphs

| | | Erdős-Rényi | | | | Community | | | | Sensor | | | | Ring | | | |
|-------------------------------|-------------------------|-------------------------|---------------------------|---------------------------|----------------------------|-------------------------|---------------------------|---------------------------|----------------------------|-------------------------|---------------------------|---------------------------|----------------------------|-------------------------|---------------------------|---------------------------|----------------------------|
| | | \hat{U}_{PALM} | \hat{U}_{PALM_d} | \hat{U}_{Givens} | $\hat{U}_{\text{Givens}/}$ | \hat{U}_{PALM} | \hat{U}_{PALM_d} | \hat{U}_{Givens} | $\hat{U}_{\text{Givens}/}$ | \hat{U}_{PALM} | \hat{U}_{PALM_d} | \hat{U}_{Givens} | $\hat{U}_{\text{Givens}/}$ | \hat{U}_{PALM} | \hat{U}_{PALM_d} | \hat{U}_{Givens} | $\hat{U}_{\text{Givens}/}$ |
| $n = 128, \text{RCG} = 3.9$ | $\text{err}_c(\hat{U})$ | 0.44 | 0.53 | 1.23 | 1.25 | 0.06 | 0.11 | 0.90 | 1.01 | 0.17 | 0.22 | 0.85 | 0.97 | 0.53 | 0.65 | 1.18 | 1.17 |
| | $\text{err}_d(\hat{U})$ | 0.45 | 0.38 | 0.11 | 0.12 | 0.08 | 0.07 | 0.03 | 0.05 | 0.21 | 0.19 | 0.05 | 0.08 | 0.39 | 0.45 | 0.09 | 0.09 |
| | T | 6.63 | 5.22 | 0.40 | 0.14 | 7.24 | 5.44 | 0.44 | 0.13 | 7.00 | 5.33 | 0.44 | 0.11 | 7.03 | 5.44 | 0.39 | 0.10 |
| $n = 256, \text{RCG} = 7.1$ | $\text{err}_c(\hat{U})$ | 0.61 | 0.86 | 1.31 | 1.31 | 0.19 | 0.33 | 1.08 | 1.14 | 0.27 | 0.36 | 1.07 | 1.15 | 0.69 | 0.79 | 1.24 | 1.24 |
| | $\text{err}_d(\hat{U})$ | 0.45 | 0.34 | 0.10 | 0.11 | 0.21 | 0.18 | 0.05 | 0.07 | 0.30 | 0.24 | 0.06 | 0.08 | 0.40 | 0.50 | 0.08 | 0.08 |
| | T | 56.4 | 47.9 | 1.60 | 0.60 | 60.8 | 47.1 | 1.52 | 0.52 | 59.0 | 45.2 | 1.67 | 0.41 | 55.8 | 44.7 | 1.59 | 0.27 |
| $n = 512, \text{RCG} = 13.1$ | $\text{err}_c(\hat{U})$ | 0.73 | 1.00 | 1.35 | 1.35 | 0.30 | 0.85 | 1.20 | 1.25 | 0.34 | 0.50 | 1.20 | 1.25 | 0.81 | 0.94 | 1.30 | 1.29 |
| | $\text{err}_d(\hat{U})$ | 0.38 | 0.05 | 0.08 | 0.08 | 0.27 | 0.17 | 0.05 | 0.07 | 0.33 | 0.28 | 0.06 | 0.08 | 0.34 | 0.61 | 0.07 | 0.08 |
| | T | 498 | 461 | 14.9 | 5.18 | 511 | 459 | 14.6 | 4.37 | 515 | 457 | 15.0 | 3.06 | 494 | 453 | 14.4 | 1.33 |
| $n = 1024, \text{RCG} = 24.4$ | $\text{err}_c(\hat{U})$ | 0.82 | 1.00 | 1.37 | 1.37 | 0.42 | 1.00 | 1.28 | 1.31 | 0.39 | 0.64 | 1.29 | 1.32 | 0.87 | 1.07 | 1.33 | 1.32 |
| | $\text{err}_d(\hat{U})$ | 0.29 | 0.02 | 0.06 | 0.07 | 0.31 | 0.07 | 0.05 | 0.06 | 0.35 | 0.31 | 0.06 | 0.08 | 0.28 | 0.77 | 0.07 | 0.07 |
| | T | 6e+3 | 5e+3 | 122 | 52.2 | 6e+3 | 5e+3 | 119 | 42.0 | 6e+3 | 5e+3 | 120 | 24.9 | 6e+3 | 5e+3 | 117 | 7.77 |

TABLE II

RESULTS OF THE EVALUATION OF THE APPROXIMATE FAST FOURIER TRANSFORMS. FOR EACH CONFIGURATION, THE RESULTS ARE GIVEN FOR THE FOUR METHODS EVOKED IN THE TEXT, AND THE BEST RESULT IS WRITTEN IN BOLD.

| | $\epsilon = \frac{\epsilon_c}{100}$ | $\epsilon = \frac{\epsilon_c}{25}$ | $\epsilon = \frac{\epsilon_c}{10}$ | $\epsilon = \frac{\epsilon_c}{5}$ | $\epsilon = \frac{\epsilon_c}{2}$ |
|----------------|-------------------------------------|------------------------------------|------------------------------------|-----------------------------------|-----------------------------------|
| $\bar{d} = 4$ | 0.036 | 0.061 | 0.090 | 0.113 | 0.140 |
| $\bar{d} = 8$ | 0.043 | 0.075 | 0.111 | 0.139 | 0.164 |
| $\bar{d} = 16$ | 0.041 | 0.074 | 0.105 | 0.128 | 0.148 |
| $\bar{d} = 32$ | 0.037 | 0.065 | 0.089 | 0.106 | 0.120 |

TABLE III

RELATIVE DIAGONALIZATION ERROR err_d ON A STOCHASTIC BLOCK MODEL (SBM) GRAPH, IN FUNCTION OF THE AVERAGE DEGREE \bar{d} AND THE STRUCTURAL PARAMETER ϵ .

whose level of structure is parameterized in order to gain insight on its influence on the diagonalization performance. To this end, we take Stochastic Block Model (SBM) graphs (see for instance [36, Section 5.1.]) with $n = 1000$ nodes made of twenty communities of equal sizes. We fix the relative complexity gain for this experiment to $\text{RCG} = 10$. We vary the average degree \bar{d} of the graph and the ratio ϵ between in-community and inter-community connection probabilities. Results in average over ten realizations are given on Table III, as a function of \bar{d} and ϵ (ϵ_c is the threshold above which community detection is impossible [37]). First, the average degree does not seem to influence the diagonalization performance. On the other hand, the parameter ϵ seems to influence the result, since the closer it is to ϵ_c , the worst are the diagonalization results. In other words, more structured SBM graphs (with a low ϵ) lead to better diagonalization performance. In contrast with the previous subsection where it was difficult to interpret the result for different graph families, we see here that in a specific graph family such as SBM graphs, global structure is linked to the diagonalization performance.

V. APPLICATION TO SENSOR NETWORKS

In this section, we first discuss the relevance of our method for graphs representing sensor networks, before analyzing its performance on both synthetic and real-world sensor networks.

A. Why sensor networks ?

To obtain the approximate FGFT $\hat{U}_{\text{Givens}/}$, the parallel truncated Jacobi algorithm of Figure 4 costs $\mathcal{O}(nJ \log n)$ arithmetic operations. For instance, it costs $\mathcal{O}(n^2(\log n)^2)$ for $J = \mathcal{O}(n \log n)$ Givens rotations. This is much cheaper than an exact diagonalization, which costs $\mathcal{O}(n^3)$, but substantially more costly than applying the resulting approximate fast transform $\mathcal{O}(n \log n)$.

Favorable use cases for this method are thus applications where this overhead cost can pay off. This corresponds to cases where the same graph Fourier transform is used a great number of times. Said otherwise, this corresponds to cases where the graph is relatively constant over time, while we process several changing signals over this graph.

Some graphs are changing relatively fast by nature. For example, graphs corresponding to social network friendships, internet links, or movie ratings are of this kind. These types of applications are thus not well suited for the method, since recomputing an approximate diagonalization each time the graph changes would be too costly.

On the other hand, graphs corresponding to sensors in physical infrastructures, such as road networks or power grids are in general quite constant over time. This kind of graph is thus particularly well suited to the method proposed in this paper, since an approximate diagonalization stays valid for a long time (structural changes in a road network or a power grid do not occur very often). In this context, it can be useful to have knowledge (even if only approximate) of the spectrum of graph signals, for example for monitoring purposes or malfunction detection, see e.g. [38, section V.A].

B. Experiments on random sensor graphs

We perform approximate diagonalizations of Laplacians corresponding to random sensor graphs generated with the GSPBOX [34]. For this experiment, we take graphs of various sizes $n \in \{64, 128, 256, 512, 1024, 2048, 4096, 8192\}$, and choose $J = 2n \log n$ Givens rotations, yielding an

| | $n = 64$ | $n = 128$ | $n = 256$ | $n = 512$ | $n = 1024$ | $n = 2048$ | $n = 4096$ | $n = 8192$ |
|--|----------|-----------|-----------|-----------|------------|------------|------------|------------|
| RCG | 1.33 | 2.29 | 4.00 | 7.11 | 12.80 | 23.27 | 42.67 | 78.77 |
| Time gain | 0.04 | 0.05 | 0.11 | 0.26 | 1.56 | 3.88 | 7.57 | 27.16 |
| $\text{err}_d(\hat{\mathbf{U}}_{\text{Givens}})$ | 0.057 | 0.062 | 0.055 | 0.051 | 0.048 | 0.051 | 0.049 | 0.048 |

TABLE IV

RELATIVE COMPLEXITY GAIN, ACTUAL TIME GAIN AND RELATIVE ERROR RESULTS ON RANDOM SENSOR GRAPHS OF VARIOUS SIZES.

approximate Fourier transform of the same complexity as the famous classical butterfly FFT [4]. Once an approximate FGFT $\hat{\mathbf{U}}_{\text{Givens}}$ is obtained this way, its actual time gain is obtained by measuring the mean time taken in MATLAB to perform matrix/vector products $\hat{\mathbf{U}}_{\text{Givens}}\mathbf{x}$ with random vectors $\mathbf{x} \in \mathbb{R}^n$, compared to the time taken to perform the matrix/vector products $\mathbf{U}\mathbf{x}$ (the mean being taken over 100 such products). Results for this experiment are given in Table IV, in terms of Relative Complexity Gain, actual time gain and relative diagonalization error $\text{err}_d(\hat{\mathbf{U}}_{\text{Givens}})$. Several comments are in order:

- First of all, the larger the graph, the higher the actual computational gains. Indeed, the actual time gain is very poor when n is small (0.04 for $n = 64$, which means the approximate FGFT is more than twenty times *slower* than the true Fourier transform), become greater than one for $n = 1024$, and attains 27.16 when $n = 8192$ (the approximate FGFT is then more than 27 times *faster* than the true Fourier transform). Moreover, the ratio between RCG and actual time gain decreases as n grows (it is equal to 46 for $n = 128$ and goes down to 3 for $n = 8192$). This means that the RCG is an overly optimistic estimation of the actual time gain for small n , but becomes of the right order of magnitude for large n .
- Second, the relative diagonalization error err_d is approximately constant (around 0.05) and does not seem to depend on the graph size. This indicates that for random sensor graphs, approximate FGFTs of $\mathcal{O}(n \log n)$ complexity make perfect sense. The compromise between error and complexity could indeed be controlled by a multiplicative constant C , taking $J = C \cdot n \log n$ Givens rotations for the approximate diagonalization.

C. Experiments on real sensor graphs

Let us now consider an application of the approximate diagonalization method to graphs representing real sensor networks. The main idea of this experiment is to compare, for several real-world graphs corresponding to physical infrastructures: a) the approximate FGFT $\hat{\mathbf{U}}_{\text{Givens}}$ obtained for three different number of Givens rotations $J \in \{n \log n, 2n \log n, 6n \log n\}$; b) to the true Fourier transform. To this end, we consider the three following graphs:

- **Minnesota road graph:** a classical graph made of $n = 2642$ nodes, generated with the GSPBOX [34].
- **US power grid graph:** a graph representing the power grid of the western United States, made of $n = 4941$ nodes. This graph was previously studied in [39], and we obtained it from [40].

- **DC road graph:** a graph representing the road network of the District of Columbia, made of $n = 9559$ nodes. This graph was previously used for a shortest path challenge [41], and we obtained it from [42]. The storage of its full Fourier matrix requires 700MB.

Performance metric. In order to evaluate the accuracy of an approximate Fourier matrix $\hat{\mathbf{U}} = \mathbf{S}_1 \dots \mathbf{S}_K$ with respect to the true Fourier matrix \mathbf{U} , we use a measure quite similar to $\text{err}_d(\hat{\mathbf{U}})$, but that is a bit more informative. Its expression is the following:

$$E_\alpha(\hat{\mathbf{U}}) \triangleq \frac{\|(\hat{\mathbf{U}}^T \mathbf{U})_{|i-j| \leq \alpha}\|_F^2}{\|\hat{\mathbf{U}}^T \mathbf{U}\|_F^2} = \frac{1}{n} \left\| (\hat{\mathbf{U}}^T \mathbf{U})_{|i-j| \leq \alpha} \right\|_F^2,$$

where $\mathbf{B} \triangleq \mathbf{A}_{|i-j| \leq \alpha}$ is a matrix of the same size as \mathbf{A} , with $b_{ij} = a_{ij}$ if $|j-i| \leq \alpha$, and $b_{ij} = 0$ otherwise (\mathbf{B} is equal to \mathbf{A} in a band of width $2\alpha+1$ centered around the diagonal and null elsewhere). We argue that the quantity $E_\alpha(\hat{\mathbf{U}})$ is a good measure for the relevance of the approximate Fourier transform $\hat{\mathbf{U}}$. Indeed, it corresponds to the fraction of the total energy of the matrix $\hat{\mathbf{U}}^T \mathbf{U}$ that is contained within a band of width $2\alpha+1$ around the diagonal. Said otherwise, if we imagine a signal $\mathbf{x} \in \mathbb{R}^n$ whose true spectrum is a Dirac located at the m th frequency ($\mathbf{x} = \mathbf{U}\delta_m$), $E_\alpha(\hat{\mathbf{U}})$ correspond to the fraction of the energy of the approximated spectrum $\hat{\delta}_m \triangleq \hat{\mathbf{U}}^T \mathbf{x}$ that is within the neighborhood of width α of m (in expectation over the choice of the frequency m). The measure $E_\alpha(\hat{\mathbf{U}})$ is defined for $\alpha \geq 0$, it is monotonically increasing with respect to the parameter α , we have $0 \leq E_\alpha(\hat{\mathbf{U}}) \leq 1$, and the higher it is, the better is the approximate Fourier matrix $\hat{\mathbf{U}}$. Moreover, we have $E_\alpha(\mathbf{U}) = 1$, $\forall \alpha \geq 0$. In summary the quantity $E_\alpha(\hat{\mathbf{U}})$ does not measure how close the approximation $\hat{\mathbf{U}}$ is of \mathbf{U} in terms of classical relative error, but how close to it is in terms of physical interpretation of the approximate spectrum. Note that the $E_\alpha(\hat{\mathbf{U}})$ measure gives more information than the previously used $\text{err}_d(\hat{\mathbf{U}})$, since it describes quantitatively the energy dispersion of the approximate FGFT. However, the downside of this measure is that since it depends of the parameter α , it requires a plot to be described, whereas a single number was sufficient for $\text{err}_d(\hat{\mathbf{U}})$.

Figure 6 shows $E_\alpha(\hat{\mathbf{U}})$ versus α/n , for the three considered graphs. Several comments are in order:

- First, it is clear that the higher the Relative Complexity Gain RCG, the lower the relevance measure $E_\alpha(\hat{\mathbf{U}})$, for any graph and any value of α . This is quite expected and shows that there exists a compromise between computational efficiency and relevance of the approximate FGFTs.
- Second, the results seem a bit better for the power grid graph than for the two other graphs, that show very sim-

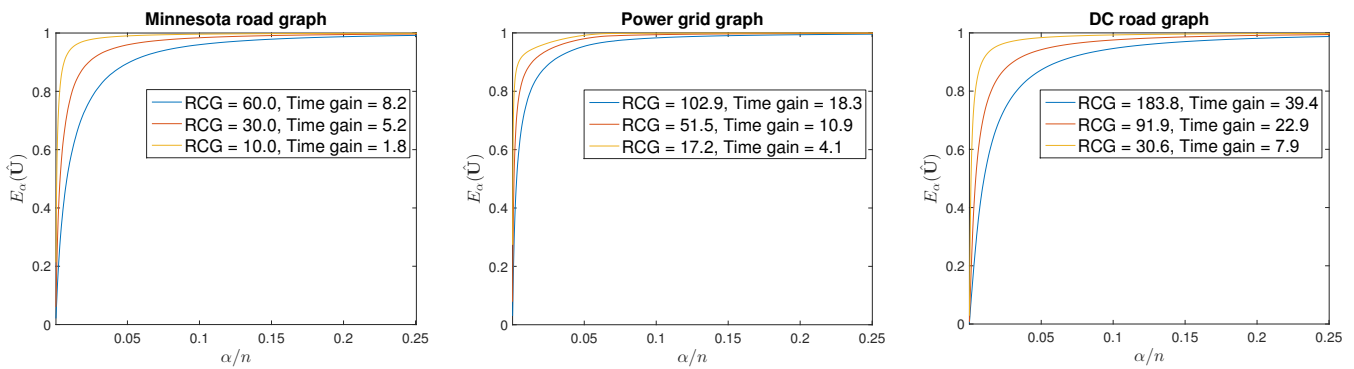


Fig. 6. Relevance of several approximate FGFTs computed for various graphs corresponding to sensor networks. Three approximate FGFTs (corresponding to different colors) are considered for each of the three graphs. The results are shown for different values of the parameter $\alpha \in [0, \frac{\pi}{4}]$.

ilar performance and both correspond to road networks. For example, if we consider the approximate FGFTs made by the product of $J = n \log n$ Givens rotations (the blue curves), we can see that we have $E_\alpha(\hat{\mathbf{U}}) \approx 0.8$ for $\alpha/n = 0.05$ for the two road networks, whereas we have $E_\alpha(\hat{\mathbf{U}}) \approx 0.9$ for the same value of α/n for the power grid network. This is an interesting observation that may mean that the inherent structure of power grid networks is more suited to approximate FGFTs than the structure of road networks. This hypothesis would of course require a more comprehensive study to be statistically validated.

- Third, we can see that the larger the graph, the faster the approximate FGFTs for the same error. Indeed, for the DC road graph (of size $n = 9559$), the fastest approximate FGFT (the blue curve) exhibits an RCG of 183.8 and an actual time gain of 39.4. Its performance in terms of the relevance measure $E_\alpha(\hat{\mathbf{U}})$ is almost the same as an approximate FGFT on the Minnesota road graph (of size $n = 2642$) that exhibits an RCG of 10 and an actual time gain of only 1.8. This means that approximate FGFTs of complexity $\mathcal{O}(n \log n)$ are relevant on these examples.

The storage of the FGFT with an RCG of 30 only requires 60MB, compared to the 700MB of the original Fourier matrix.

Finally, let us illustrate the expected tradeoff between accuracy and computational efficiency of the $\hat{\mathbf{U}}_{\text{Givens}}$ approximate FGFT. To this end, we consider the DC road graph and generate a random structured spectrum $\mathbf{y} \in \mathbb{R}^{9559}$, transform it to the node domain to get $\mathbf{x} = \mathbf{U}\mathbf{y}$, and compute its approximate spectrum $\hat{\mathbf{y}} = \hat{\mathbf{U}}_{\text{Givens}}^T \mathbf{x}$. Results are shown in Figure 7 for different RCG values. As expected, the larger the RCG value, the less accurate is the approximate spectrum; yet the approximation gives good results with an actual time gain of 7.9, and fair results with an actual time gain of 22.9.

VI. APPLICATION EXAMPLE: FILTERING

In this last section, as an example of the versatility of approximate FGFTs, we focus on one of its many possible

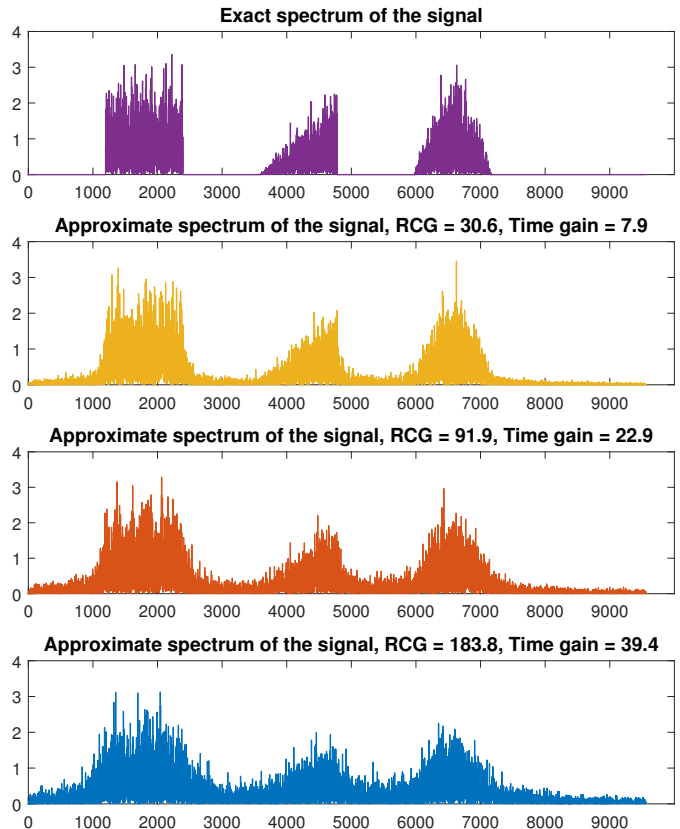


Fig. 7. Example of spectrum computations on the DC road graph.

applications: graph signal filtering. Given a graph signal \mathbf{x} , its filtered version \mathbf{y} is computed as

$$\mathbf{y} = \underbrace{\mathbf{U}\mathbf{H}\mathbf{U}^T}_{\mathbf{G}} \mathbf{x},$$

where \mathbf{H} is a diagonal matrix whose diagonal entries are the frequency response of the filter and we denote $\mathbf{G} \triangleq \mathbf{U}\mathbf{H}\mathbf{U}^T$ the filtering matrix. We see that filtering requires to apply the Fourier transform and its inverse (here, its transpose as \mathbf{U} is orthogonal). An efficient approximation of the Fourier matrix would thus allow to have a complexity gain on the filtering that is of the order of the relative complexity gain RCG.

Filtering with an approximate FGFT. We consider the approximate FGFT $\hat{\mathbf{U}}_{\text{Givens}}//$ to perform filtering: the approximate filtering matrix thus reads:

$$\hat{\mathbf{G}}_{\text{FGFT}} \triangleq \hat{\mathbf{U}}_{\text{Givens}}// \mathbf{H} \hat{\mathbf{U}}_{\text{Givens}}//^T.$$

Filtering with polynomial approximations. We compare this proposition to the usual approximate filtering method based on the polynomial approximation of the filter's frequency response [43]. Starting from the frequency response of an ideal filter $h(\lambda)$ (where λ represents the continuous frequency variable), the idea is to approximate it with a polynomial function $\hat{h}(\lambda) = \sum_{i=0}^p \alpha_i \lambda^i$, such that the approximate filtering matrix has the expression

$$\hat{\mathbf{G}}_{\text{poly}} \triangleq \mathbf{U} \hat{h}(\mathbf{\Lambda}) \mathbf{U}^T = \sum_{i=0}^p \alpha_i \mathbf{L}^i$$

where \hat{h} is applied entry-wise to the diagonal entries of $\mathbf{\Lambda}$. This provides a computationally efficient way of applying the filter, since its application cost is $p(\|\mathbf{L}\|_0 + n) + n$ arithmetic operations. With this polynomial approximation method, the Relative Complexity Gain reads $\text{RCG} = \frac{n^2}{p(\|\mathbf{L}\|_0 + n) + n}$. Note that obtaining the coefficients α_i characterizing $\hat{\mathbf{G}}_{\text{poly}}$ for a given filter costs $\mathcal{O}(p)$ arithmetic operations, and one needs to recompute these coefficients as soon as the filter changes. On the other hand, with the greedy diagonalizations, obtaining $\hat{\mathbf{G}}_{\text{FGFT}}$ costs $\mathcal{O}(nJ \log n)$ (as denoted in Table I), but can be done once and for all and reused for as many filters as needed, as long as the same graph is used.

Application to denoising. We perform a first experiment on the Minnesota road network graph presented in Section V. We will compare the performance of the two approximation methods for three RCG values: 35, 17 and 12. This corresponds to a number of Givens rotations $J = 50000, 100000, 150000$ for the FGFT-based approximation, and an order $p = 14, 28, 40$ for the polynomial-based approximation. In this first experiment, a low-frequency signal \mathbf{x} is generated randomly in the graph Fourier domain: coefficients of its spectrum \mathbf{y} are independent and follow a normal distribution $y_i \sim \mathcal{N}(0, \theta_i)$ with $\theta_i = 1$ if $i \leq 300$, $\theta_i = 0$ otherwise, where λ_i is the i th eigenvalue of the Laplacian. We then obtain a signal in the graph node domain $\mathbf{x} = \mathbf{U}\mathbf{y}$, that is corrupted with a white Gaussian noise \mathbf{n} with $n_i \sim \mathcal{N}(0, \sigma)$, to get the noisy signal $\tilde{\mathbf{x}} = \mathbf{x} + \mathbf{n}$. The noisy signal is then low-pass filtered in order to remove part of the noise. The filter used has a frequency response given by $h_i = 1$ if $i \leq 300$, $h_i = 0$ otherwise, where h_i is the i th entry on the diagonal of the filter matrix \mathbf{H} . Filtering is done with the true Fourier matrix \mathbf{U} (with the exact filter $\mathbf{G}\tilde{\mathbf{x}} = \mathbf{U}\mathbf{H}\mathbf{U}^T\tilde{\mathbf{x}}$), to which we compare $\hat{\mathbf{G}}_{\text{poly}}\tilde{\mathbf{x}}$ and $\hat{\mathbf{G}}_{\text{FGFT}}\tilde{\mathbf{x}}$.

First of all, for illustration purposes, an example of graph filtering with $\hat{\mathbf{U}}_{\text{Givens}}//$ on a realisation of noisy signal with $\sigma = 0.4$ is shown in Figure 8. Moreover, for different noise levels σ , average results over 100 realizations are given in

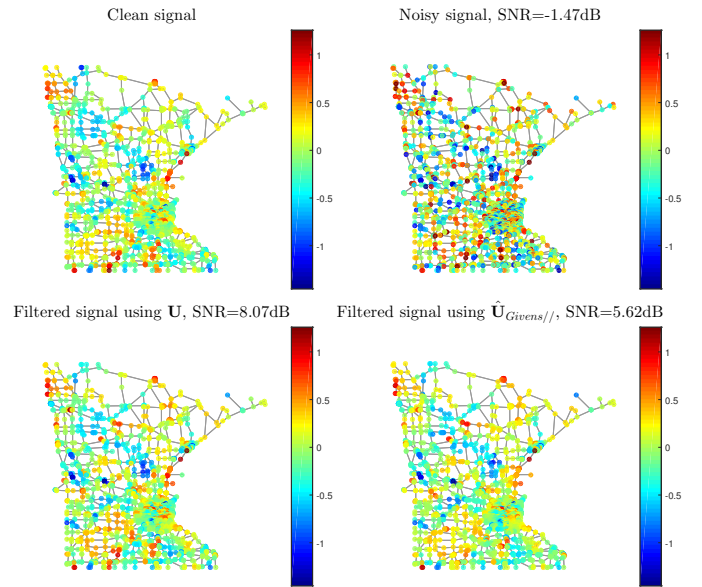


Fig. 8. Filtering example on the Minnesota road graph, the approximate FGFT $\hat{\mathbf{U}}_{\text{Givens}}//$ is approximately 35 times more efficient than the true Fourier matrix \mathbf{U} .

Table V, in terms of Signal to Noise Ratio (SNR) in decibel (dB):

$$\text{SNR} = 10 \log_{10} \left(\frac{\|\hat{\mathbf{x}}\|_2^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2} \right).$$

Both methods show similar performance at constant RCG, and their run time are comparable. However, approximate FGFTs can be considered as more versatile than polynomial approximation, in the sense that they give access to an approximate spectrum, and thus allow to perform in a computationally efficient way any task requiring access to this spectrum, which is not the case of polynomial approximations that can be used for filtering only. Besides, the performance of polynomial approximations depends heavily on the considered type of filter, as is discussed in the next paragraph.

Comparison of the approximate filtering matrices. In order to better understand the relative strengths and weaknesses of the proposed method compared to polynomial approximation, we now analyze the approximate filtering matrix $\hat{\mathbf{G}}_{\text{FGFT}}$ and compare it to $\hat{\mathbf{G}}_{\text{poly}}$ obtained with equal relative complexities. We evaluate the accuracy of these approximate filtering matrices, generically written $\hat{\mathbf{G}}$, by comparing them with the exact filtering matrix \mathbf{G} , with the following measure:

$$\frac{\|\mathbf{G} - \hat{\mathbf{G}}\|_F}{\|\mathbf{G}\|_F}.$$

In this second experiment, we consider a community graph of size $n = 2048$ made of 23 communities, and two different types of filters, representing two extreme situations with respect to polynomial approximation: an ideal low-pass filter and a heat kernel (exponentially decreasing filter). The ideal low-pass filter has a frequency response equal to one for the first r frequencies and zero after that (we consider

| | $\sigma = 0.2$ | $\sigma = 0.25$ | $\sigma = 0.3$ | $\sigma = 0.4$ | $\sigma = 0.5$ | $\sigma = 0.6$ |
|---|---|---|---|---|---|---|
| Noisy signal | 4.50 | 2.61 | 0.98 | -1.45 | -3.47 | -4.99 |
| Filtered with U | 13.97 | 12.08 | 10.40 | 7.97 | 6.00 | 4.48 |
| | $\hat{\mathbf{G}}_{\text{poly}} \hat{\mathbf{G}}_{\text{FGFT}}$ | $\hat{\mathbf{G}}_{\text{poly}} \hat{\mathbf{G}}_{\text{FGFT}}$ | $\hat{\mathbf{G}}_{\text{poly}} \hat{\mathbf{G}}_{\text{FGFT}}$ | $\hat{\mathbf{G}}_{\text{poly}} \hat{\mathbf{G}}_{\text{FGFT}}$ | $\hat{\mathbf{G}}_{\text{poly}} \hat{\mathbf{G}}_{\text{FGFT}}$ | $\hat{\mathbf{G}}_{\text{poly}} \hat{\mathbf{G}}_{\text{FGFT}}$ |
| RCG = 12, $T=2.4\text{ms}$ 2.9ms | 11.84 11.39 | 10.70 10.25 | 9.43 9.05 | 7.44 7.16 | 5.72 5.46 | 4.34 4.11 |
| RCG = 17, $T=1.7\text{ms}$ 2.1ms | 0.0017 10.35 | 8.59 9.43 | 7.82 8.44 | 6.49 6.79 | 5.16 5.18 | 4.01 3.90 |
| RCG = 35, $T=1.0\text{ms}$ 1.1ms | 8.10 7.45 | 7.72 7.00 | 7.10 6.39 | 6.08 5.25 | 4.89 4.21 | 3.87 3.10 |

TABLE V

FILTERING RESULTS, THE SNR IN DB (MEAN OVER 100 REALIZATIONS) IS GIVEN, AND THE BEST RESULT IS WRITTEN IN BOLD. THE AVERAGE FILTERING TIME T (IN MILLISECONDS) FOR BOTH METHOD AT EACH RCG IS ALSO GIVEN.

| | RCG = 35 | RCG = 17 | RCG = 12 |
|---------------------------|---|---|---|
| | $\hat{\mathbf{G}}_{\text{poly}} \hat{\mathbf{G}}_{\text{FGFT}}$ | $\hat{\mathbf{G}}_{\text{poly}} \hat{\mathbf{G}}_{\text{FGFT}}$ | $\hat{\mathbf{G}}_{\text{poly}} \hat{\mathbf{G}}_{\text{FGFT}}$ |
| Ideal low-pass $r = 7$ | 1.00 1.03 | 1.00 0.94 | 1.39 0.93 |
| Ideal low-pass $r = 23$ | 1.00 0.64 | 0.78 0.34 | 0.42 0.23 |
| Ideal low-pass $r = 1000$ | 0.68 0.25 | 0.48 0.19 | 0.50 0.16 |
| Exponential | 0.05 0.09 | 1e-4 0.06 | 0.00 0.05 |

TABLE VI

COMPARISON (RELATIVE FROBENIUS NORM ERROR) BETWEEN THE FILTERING OPERATORS OBTAINED WITH THE APPROXIMATE FGFT AND WITH POLYNOMIAL APPROXIMATIONS, THE BEST RESULT BEING WRITTEN IN BOLD.

that the r th eigenvalue is perfectly known for the polynomial approximation). We consider three choices for the index r of the cutting frequency: $r = 7$ which corresponds to a region where the density of graph frequencies is high, $r = 23$ which corresponds to a region where there is a gap in the graph frequencies density (it corresponds to the number of communities) and $r = 1000$ which corresponds to a region where the density of graph frequencies is lower. The heat kernel has a frequency response $h(\lambda) = \exp(-\lambda)$

Table VI shows the resulting relative approximation errors for various RCG values. For the ideal low-pass filter with $r = 7$, the two approximate filtering methods perform quite poorly, with a small advantage for the FGFT-based approximation. This can be explained by the fact that this filtering operation is intrinsically difficult to approximate because of the high density of eigenvalues around the cutting frequency. On the other hand, with $r = 23$, results are better for the FGFT-based approximation, but remain quite poor for the polynomial approximations. Finally, for $r = 1000$, results are better for both methods while the FGFT-based approximation shows a clear advantage for all tested values of RCG. Conversely, for the exponential filter, polynomial approximations give much better results than approximate FGFTs (it is essentially perfect). This can be explained by the fact that the performance of polynomial approximations depends heavily on the type of filter: it works well for filters whose frequency response is well approximated by polynomials of low degree (typically the exponential filter), and works not as well for filters whose frequency response is not easily approximated (typically the ideal low-pass filter). Indeed, polynomials of low degree can approximate well only relatively smooth functions [44]. On the other hand, the performance of approximate FGFTs seems

less dependent on the type of filters.

It is thus expected that in general the polynomial filter approximation is the best method for filtering signals on graphs, but in specific situations (such as the ideal low-pass filter), it may be outperformed by an FGFT-based approximation. Interestingly, the ideal low-pass filter is important in certain applications such as spectral clustering [36], where one needs to apply an ideal low-pass filter with r being equal to the desired number of clusters.

VII. CONCLUSION

In this paper, we proposed a method to obtain approximate Fast Fourier Transforms on graphs via an approximate diagonalization of the Laplacian matrix with an approximate eigenvector matrix constrained to be a product of a relatively low number of Givens rotations. The method was first described and analyzed in detail. It was then compared to another method we previously proposed on various popular graphs, and applied to sensor networks, showing promising results. Finally, an application to filtering was proposed, showing that approximate FGFTs can perform as good as polynomial filter approximations for certain filters.

In the future, it would be very interesting to reduce the gap between actual time gain and RCG, to have approximate FGFTs whose benefits are closer to the theoretical gain, whose usability would be enhanced.

It would also be beneficial to the method to imagine clever ways to update the approximate FGFT when the graph changes, without requiring to recompute a complete approximate diagonalization. This could indeed open up the method to new applications (other than sensor networks) where the graphs are varying by nature (social networks, movie ratings, etc.).

We could also imagine applying the method to approximately diagonalize other symmetric matrices besides graph Laplacians. For example, applying it directly to a graph filtering matrix \mathbf{G} (computed for example by polynomial approximation, as defined in Section VI, to avoid any costly diagonalization) could allow for a faster application of the filter, which is of interest if the filter has to be applied a great number of times.

Furthermore, an efficient implementation that would reduce the approximate diagonalization time of the proposed methods would be very interesting. Indeed, efficient numerical linear algebra routines exist that implement the Jacobi eigenvalue

algorithm, so that modifying them directly could potentially greatly reduce this factorization time. Moreover, the use of more advanced techniques similar to those introduced in [26] for the pMMF method may also be used to accelerate the diagonalization. This way, the cost of approximate diagonalization (whose theoretical time complexity is $\mathcal{O}(n^2 \log n)$ for $J = \mathcal{O}(n \log n)$ Givens rotations) may even become competitive with that of the cost of applying the true Fourier transform (whose theoretical time complexity is $\mathcal{O}(n^2)$). This would open new fields of applications for the proposed methods.

Finally, a theoretical analysis of the proposed greedy diagonalization would be of great interest. For example, a result linking a graph structure and the minimum number of Givens rotations required to diagonalize its Laplacian up to a prescribed accuracy would be very useful to get a deeper understanding of the method. More generally, characterizing which symmetric matrices admit good approximate diagonalization by products of few Givens (or more generally few sparse) factors is an interesting question for approximation theory.

APPENDIX A EFFICIENT GREEDY STRATEGY

We give here a computationally efficient version of the algorithm of Figure 2.

Input: Matrix \mathbf{L} , desired number of factors J .

```

1:  $\mathbf{L}_1 \leftarrow \mathbf{L}$ 
2: for  $r = 1$  to  $n$  do
3:   for  $s = r + 1$  to  $n$  do
4:      $c_{r,s} \leftarrow -|l_{r,s}^j|$ 
5:   end for
6:    $d(r) \leftarrow \min \mathbf{C}(r, :)$ 
7:    $e(r) \leftarrow \operatorname{argmin} \mathbf{C}(r, :)$ 
8: end for
9: for  $j = 1$  to  $J$  do
10:   $p \leftarrow \operatorname{argmin} d(r)$ 
11:   $q \leftarrow e(p)$ 
12:   $\theta \leftarrow \frac{1}{2} \arctan\left(\frac{l_{pq}^j - l_{pp}^j}{2l_{pq}^j}\right) + \frac{\pi}{4}$ 
13:   $\mathbf{S}_j \leftarrow \mathbf{G}_{p,q,\theta}$ 
14:   $\mathbf{L}_{j+1} \leftarrow \mathbf{S}_j^T \mathbf{L}_j \mathbf{S}_j$ 
15:  for  $r = p, q$  do
16:    for  $s = r + 1$  to  $n$  do
17:       $c_{r,s} \leftarrow -|l_{r,s}^j|$ 
18:    end for
19:     $d(r) \leftarrow \min \mathbf{C}(r, :)$ 
20:     $e(r) \leftarrow \operatorname{argmin} \mathbf{C}(r, :)$ 
21:  end for
22:  for  $s = p, q$  do
23:    for  $r = 1$  to  $s - 1$  do
24:       $c_{r,s} \leftarrow -|l_{r,s}^j|$ 
25:      if  $c_{r,s} < d(r)$  then
26:         $d(r) \leftarrow c_{r,s}$ 
27:         $e(r) \leftarrow s$ 
28:      else
29:        if  $e(r) = s$  then
30:           $d(r) \leftarrow \min \mathbf{C}(r, :)$ 
31:           $e(r) \leftarrow \operatorname{argmin} \mathbf{C}(r, :)$ 
32:        end if
33:      end if
34:    end for
35:  end for
36: end for
37:  $\hat{\mathbf{A}} \leftarrow \operatorname{diag}(\mathbf{L}_{J+1})$ 

```

Output: Sparse orthogonal factors $\{\mathbf{S}_j\}_{j=1}^J$; diagonal factor $\hat{\mathbf{A}}$.

Fig. 9. **Truncated Jacobi algorithm (efficient version)** : Algorithm for the approximate greedy diagonalization of the Laplacian matrix (efficient version).

ACKNOWLEDGMENT

The authors would like to thank Igal Rozenberg for discussions and collaboration on the topic of graph signal processing during his internship with the team.

REFERENCES

- [1] J. A. Bondy and U. S. R. Murty, *Graph theory with applications*. Elsevier Science Ltd/North-Holland, 1976, vol. 290.
- [2] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *Signal Processing Magazine, IEEE*, vol. 30, no. 3, pp. 83–98, 2013.
- [3] A. Sandryhaila and J. Moura, "Discrete signal processing on graphs," *Signal Processing, IEEE Transactions on*, vol. 61, no. 7, pp. 1644–656, 2013.
- [4] J. Cooley and J. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Mathematics of Computation*, vol. 19, no. 90, pp. 297–301, 1965.
- [5] J. Morgenstern, "The linear complexity of computation," *J. ACM*, vol. 22, no. 2, pp. 184–194, Apr. 1975. [Online]. Available: <http://doi.acm.org/10.1145/321879.321881>
- [6] L. Le Magoarou and R. Gribonval, "Flexible multilayer sparse approximations of matrices and applications," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 688–700, 2016.
- [7] —, "Are there approximate fast fourier transforms on graphs?" in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 4811–4815.
- [8] W. Givens, "Computation of plane unitary rotations transforming a general matrix to triangular form," *Journal of the Society for Industrial and Applied Mathematics*, vol. 6, no. 1, pp. pp. 26–50, 1958. [Online]. Available: <http://www.jstor.org/stable/2098861>
- [9] C. G. J. Jacobi, "Über ein leichtes verfahren, die in der theorie der säkularstörungen vorkommenden gleichungen numerisch aufzulösen," *J. reine angew. Math.*, vol. 30, pp. 51–94, 1846.
- [10] R. Rubinstein, M. Zibulevsky, and M. Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation," *Signal Processing, IEEE Transactions on*, vol. 58, no. 3, pp. 1553–1564, March 2010.
- [11] O. Chabiron, F. Malgouyres, J.-Y. Tourneret, and N. Dobigeon, "Toward fast transform learning," *International Journal of Computer Vision*, pp. 1–22, 2014. [Online]. Available: <http://dx.doi.org/10.1007/s11263-014-0771-z>
- [12] J. Sulam, B. Ophir, M. Zibulevsky, and M. Elad, "Trainlets: Dictionary learning in high dimensions," *IEEE Transactions on Signal Processing*, 2016.
- [13] O. Chabiron, F. Malgouyres, H. Wendt, and J.-Y. Tourneret, "Optimization of a Fast Transform Structured as a Convolutional Tree," Jan. 2016, working paper or preprint. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01258514>
- [14] S. Lyu and X. Wang, "On algorithms for sparse multi-factor NMF," in *Advances in Neural Information Processing Systems 26*, 2013, pp. 602–610.
- [15] G. Beylkin, R. Coifman, and V. Rokhlin, "Fast wavelet transforms and numerical algorithms i," *Communications on pure and applied mathematics*, vol. 44, no. 2, pp. 141–183, 1991.
- [16] V. Rokhlin, "Rapid solution of integral equations of classical potential theory," *Journal of Computational Physics*, vol. 60, no. 2, pp. 187 – 207, 1985. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0021999185900026>
- [17] W. Hackbusch, "A Sparse Matrix Arithmetic Based on H-matrices. Part I: Introduction to H-matrices," *Computing*, vol. 62, no. 2, pp. 89–108, May 1999. [Online]. Available: <http://dx.doi.org/10.1007/s006070050015>
- [18] E. Candès, L. Demanet, and L. Ying, "Fast computation of fourier integral operators," *SIAM Journal on Scientific Computing*, vol. 29, no. 6, pp. 2464–2493, 2007. [Online]. Available: <http://dx.doi.org/10.1137/060671139>
- [19] C. Rusu, N. Gonzalez-Prelcic, and R. W. J. Heath, "Fast orthonormal sparsifying transforms based on householder reflectors," *Submitted to IEEE Transactions on Signal Processing*, 2016.
- [20] A. S. Householder, "Unitary triangularization of a nonsymmetric matrix," *J. ACM*, vol. 5, no. 4, pp. 339–342, Oct. 1958. [Online]. Available: <http://doi.acm.org/10.1145/320941.320947>

- [21] A. B. Lee, B. Nadler, and L. Wasserman, "Treelets - an adaptive multi-scale basis for sparse unordered data," *The Annals of Applied Statistics*, vol. 2, no. 2, pp. 435–471, July 2008.
- [22] G. Cao, L. Bacheega, and C. Bouman, "The sparse matrix transform for covariance estimation and analysis of high dimensional signals," *Image Processing, IEEE Transactions on*, vol. 20, no. 3, pp. 625–640, 2011.
- [23] H. Chen, S. Zhu, and B. Zeng, "Design of non-separable transforms for directional 2-d sources," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 3697–3700.
- [24] H. Chen and B. Zeng, "Design of low-complexity, non-separable 2-d transforms based on butterfly structures," in *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 2921–2924.
- [25] R. Kondor, N. Teneva, and V. Garg, "Multiresolution matrix factorization," in *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, 2014, pp. 1620–1628.
- [26] N. Teneva, P. K. Mudrakarta, and R. Kondor, "Multiresolution matrix compression," in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, 2016, pp. 1441–1449.
- [27] A. Sandryhaila and J. M. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *Signal Processing Magazine, IEEE*, vol. 31, no. 5, pp. 80–90, Sept 2014.
- [28] G. H. Golub and H. A. Van der Vorst, "Eigenvalue computation in the 20th century," *Journal of Computational and Applied Mathematics*, vol. 123, no. 1, pp. 35–65, 2000.
- [29] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU Press, 2012, vol. 3.
- [30] V. Strassen, "Gaussian elimination is not optimal," *Numerische Mathematik*, vol. 13, no. 4, pp. 354–356, 1969.
- [31] D. Coppersmith and S. Winograd, "Matrix multiplication via arithmetic progressions," *Journal of symbolic computation*, vol. 9, no. 3, pp. 251–280, 1990.
- [32] F. L. Gall, "Powers of tensors and fast matrix multiplication," in *International Symposium on Symbolic and Algebraic Computation, ISSAC '14, Kobe, Japan, July 23-25, 2014*, 2014, pp. 296–303. [Online]. Available: <http://doi.acm.org/10.1145/2608628.2608664>
- [33] R. P. Brent and F. T. Luk, "The solution of singular-value and symmetric eigenvalue problems on multiprocessor arrays," *SIAM Journal on Scientific and Statistical Computing*, vol. 6, no. 1, pp. 69–84, 1985.
- [34] N. Perraudin, J. Paratte, D. Shuman, V. Kalofolias, P. Vandergheynst, and D. K. Hammond, "GSPBOX: A toolbox for signal processing on graphs," *ArXiv e-prints*, Aug. 2014.
- [35] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 459–494, 2014. [Online]. Available: <http://dx.doi.org/10.1007/s10107-013-0701-9>
- [36] N. Tremblay, G. Puy, R. Gribonval, and P. Vandergheynst, "Compressive spectral clustering," in *Machine Learning, Proceedings of the Thirty-third International Conference (ICML 2016), June, 2016*, pp. 20–22.
- [37] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborov, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Phys. Rev. E*, vol. 84, no. 6, p. 066106, Dec. 2011.
- [38] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs: Frequency analysis," *Signal Processing, IEEE Transactions on*, vol. 62, no. 12, pp. 3042–3054, 2014.
- [39] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 1998.
- [40] "Us power grid network dataset." [Online]. Available: <http://konect.uni-koblenz.de/networks/opsahl-powergrid>
- [41] C. Demetrescu, A. Goldberg, and D. Johnson, "9th dimacs challenge on shortest paths," 2007, <http://www.dis.uniroma1.it/challenge9/>.
- [42] "Dc road network dataset." [Online]. Available: <http://www.dis.uniroma1.it/challenge9/data/tiger/>
- [43] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129 – 150, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1063520310000552>
- [44] T. Dupont and R. Scott, "Polynomial approximation of functions in sobolev spaces," *Mathematics of Computation*, vol. 34, no. 150, pp. 441–463, 1980.