# GRAPH FOURIER TRANSFORM FOR DIRECTED GRAPHS BASED ON LOVÁSZ EXTENSION OF MIN-CUT

*Stefania Sardellitti*[1], *Sergio Barbarossa*[1], *Paolo Di Lorenzo*[2]

[1] Sapienza University of Rome, DIET, Via Eudossiana 18, 00184 Rome, Italy
[2] Dept. of Engineering, University of Perugia, 06125 Perugia, Italy
E-mail: {stefania.sardellitti, sergio.barbarossa}@uniroma1.it, paolo.dilorenzo@unipg.it

## ABSTRACT

A key tool to analyze signals defined over a graph is the so called Graph Fourier Transform (GFT). Alternative definitions of GFT have been proposed, based on the eigen-decomposition of either the graph Laplacian or adjacency matrix. In this paper, we introduce an alternative approach, valid for the general case of directed graphs, that builds the graph Fourier basis as the set of orthonormal vectors that minimize a well-defined continuous extension of the graph cut size, known as Lovász extension. To cope with the non-convexity of the problem, we exploit a recently developed method devised for handling orthogonality constraints, with provable convergence properties.

***Index Terms***— Graph signal processing, graph Fourier transform, total variation, clustering.

## 1. INTRODUCTION

Signal processing on graphs has attracted a lot of interest in the last years because of its many potential applications, from social and economic networks, to smart grids, gene regulatory networks, and so on. Graph signal processing (GSP) represents a promising tool for the representation, processing and analysis of signals defined over the vertices of a (weighted) graph [1–5]. A central role in GSP is the spectral analysis of graph signals, which is based on the so called Graph Fourier Transform (GFT). Alternative definitions of GFT have been proposed: a) one approach, valid for *undirected* graphs, based on the projection onto the eigenvectors of the graph Laplacian, which represent the optimal basis minimizing the $l_2$ norm graph total variation, see, e.g. [6], [3], [7]; b) an alternative approach, proposed in [1], [2] for the more general case of *directed* graphs, built on the Jordan decomposition of the adjacency matrix. This second method is rooted on the association of the graph adjacency matrix with the signal shift operator, which lies at the basis of all shift-invariant linear filtering methods for graph signals [8], [9]. In this paper, we propose an alternative approach valid for *directed* graphs, which, differently from [1], [2], builds a Fourier basis that is unitary and guarantees zero total variation for constant graph signals. We assume as optimization criterion the minimization of the cut size, as a general way to capture clustering properties of graph signals over both undirected and directed graphs. Since the min-cut problem is a combinatorial problem, we exploit the submodularity property of the cut size to derive a convex relaxation of this set function, known as its Lovász extension [10], [11], whose minimization preserves the optimality of the original combinatorial problem. Interestingly, in the undirected graph case, the Lovász extension of the cut size reduces to the widely used $\ell_1$ norm total variation of a graph signal, whereas for directed graph it provides an alternative measure of graph signal variation, which captures the edge directivity. We build the GFT basis as the set of orthonormal vectors that minimize the Lovász extension of the cut size. Since the resulting problem is non-convex, we exploit a recently developed efficient method to handle non-convex orthogonality constraints, namely the proximal alternating minimized augmented Lagrangian (PAMAL) method [12], which is an iterative method with provable convergence property guaranteeing that any limit point is a Karush-Kuhn Tucker point of the original non-convex problem [12].

This work has been supported by TROPIC Project, Nr. ICT-318784. The work of Paolo Di Lorenzo was funded by the "Fondazione Cassa di Risparmio di Perugia".

## 2. MIN-CUT SIZE AND GRAPH FOURIER BASIS

We consider a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ consisting of a set of $N$ vertices (or nodes) $\mathcal{V} = \{1, \ldots, N\}$ along with a set of edges $\mathcal{E} = \{a_{ij}\}_{i,j\in\mathcal{V}}$, such that $a_{ij} > 0$ if there is a direct link from node $j$ to node $i$ or $a_{ij} = 0$, otherwise. We denote with $|\mathcal{V}|$ the cardinality of $\mathcal{V}$, i.e. the number of elements of $\mathcal{V}$. A signal $\boldsymbol{s}$ on a graph $\mathcal{G}$ is defined as a mapping from the vertex set to a real vector of size $N = |\mathcal{V}|$, i.e. $\boldsymbol{s} : \mathcal{V} \to \mathbb{R}$. Let $\mathbf{A}$ denote the $N \times N$ adjacency matrix with entries the edge weights $a_{ij}$ for $i, j = 1, \ldots, N$. The (unnormalized) graph Laplacian is defined as $\mathbf{L} := \mathbf{D} - \mathbf{A}$, where the in-degree matrix $\mathbf{D}$ is a diagonal matrix whose $i$th diagonal entry is $d_i = \sum_j a_{ij}$. One of the basic operations over graphs is clustering, i.e. the partition of the graph onto disjoint subgraphs, such that the vertices within each subgraph (cluster) are highly interconnected, whereas there are only a few links between different

clusters. Finding a good partition can be formulated as the minimization of the cut size [13]. Consider a subset of vertices $\mathcal{S} \subset \mathcal{V}$ and its complement in $\mathcal{V}$, $\bar{\mathcal{S}}$. The edge boundary of $\mathcal{S}$ is defined as the set of edges with one end in $\mathcal{S}$ and the other end in its complement $\bar{\mathcal{S}}$. The cut set size between $\mathcal{S}$ and $\bar{\mathcal{S}}$ is defined as the sum of the weights over the boundary, i.e. [13]

$$\text{cut}(\mathcal{S}, \bar{\mathcal{S}}) := \sum_{i \in \mathcal{S}, j \in \bar{\mathcal{S}}} a_{ji}. \tag{1}$$

Computing the optimal partition that minimizes the cut size is an NP-hard problem. To overcome this difficulty, we use the property that the cut size is a submodular function [11] and that the Lovász extension [10] of a submodular function is a convex function [11]. More specifically, given the set $\mathcal{V}$ and its power set $2^{\mathcal{V}}$, i.e. the set of all its subsets, let us consider a real-valued set function $F : 2^{\mathcal{V}} \to \mathbb{R}$. The cut size in (1) is an example of set function, with $F(\mathcal{S}) := \text{cut}(\mathcal{S}, \bar{\mathcal{S}})$. The Lovász extension of the graph function $F$ [10], [11] allows the extension of a set-function defined on the vertices of a graph to the real space $\mathbb{R}^N$ according to the following definition.

**Definition 1** *Let $F : 2^{\mathcal{V}} \to \mathbb{R}$ be a set function with $F(\emptyset) = 0$. Let $\boldsymbol{x} \in \mathbb{R}^N$ be ordered w.l.o.g. in increasing sense, so that $x_1 \leq x_2 \leq \ldots \leq x_N$, and define $C_i = \{j \in \mathcal{V} : x_j > x_i\}$, where $C_0 = \mathcal{V}$. Then the Lovász extension $f : \mathbb{R}^N \to \mathbb{R}$ of $F$ is given by:*

$$f(\mathbf{x}) = \sum_{i=1}^{N} x_i (F(C_{i-1}) - F(C_i)) = \sum_{i=1}^{N-1} F(C_i)(x_{i+1} - x_i) + x_1 F(\mathcal{V}).$$

A particular interesting class of set functions are the *submodular* set functions, whose definition is recalled here below.

**Definition 2** *A set function $F : 2^{\mathcal{V}} \to \mathcal{R}$ is submodular if and only if, $\forall \mathcal{A}, \mathcal{B} \subseteq \mathcal{V}$, it satisfies the following inequality*

$$F(\mathcal{A}) + F(\mathcal{B}) \geq F(\mathcal{A} \cup \mathcal{B}) + F(\mathcal{A} \cap \mathcal{B}).$$

A fundamental property of a set function $F$ is that its Lovász extension is convex if and only if $F$ is submodular [ cf. [11], p. 23]. In particular, the cut function is known for being submodular (see, e.g., [11], [14]) and its Lovász extension, in the general case of directed graphs, is

$$f(\mathbf{x}) = \sum_{i,j=1}^{N} a_{ji} [x_i - x_j]_+ := \text{DV}(\mathbf{x}) \tag{2}$$

with $[y]_+ := \max\{y, 0\}$. For undirected graphs, (2) boils down to the widely used $\ell_1$ norm total variation of a graph signal [15], [16]:

$$f(\mathbf{x}) = \sum_{i,j=1, i>j}^{N} a_{ji} |x_i - x_j| := \text{TV}_1(\mathbf{x}). \tag{3}$$

Alternative definitions of GFT have been proposed in the literature. In case of *undirected* graphs, the GFT of a vector $\boldsymbol{s}$ was defined as [3]

$$\hat{\boldsymbol{s}} = \mathbf{U}^T \boldsymbol{s} \tag{4}$$

where the columns of matrix $\mathbf{U}$ are the eigenvectors of the Laplacian matrix $\mathbf{L}$, i.e. $\mathbf{L} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$. This definition is basically rooted on the clustering properties of these eigenvectors [17]. In fact, by definition of eigenvector, the Fourier basis used in (4) can be thought as the solution of the following sequence of optimization problems for $k = 2, \ldots, N$

$$\begin{aligned} \boldsymbol{u}_k = &\arg\min_{\boldsymbol{u}_k \in \mathbb{R}^N} \boldsymbol{u}_k^T \mathbf{L} \boldsymbol{u}_k := \arg\min_{\boldsymbol{u}_k \in \mathbb{R}^N} \text{TV}_2(\boldsymbol{u}_k) \\ &\text{s.t.} \qquad \boldsymbol{u}_k^T \boldsymbol{u}_\ell = \delta_{k\ell}, \;\; \ell = 1, \ldots, k \end{aligned} \tag{5}$$

where $\delta_{k\ell}$ is the Kronecker symbol and we used the property that the quadratic form built on the Laplacian is the $\ell_2$-norm total variation, i.e. $\text{TV}_2(\mathbf{x}) := \sum_{i,j=1, i>j}^{N} a_{ji}(x_i - x_j)^2$. Since the first eigenvector of $\mathbf{L}$ is the constant vector, we assume in (5) $\boldsymbol{u}_1 = b\,\mathbf{1}$, with $b = 1/\sqrt{N}$. In all applications where graph signals exhibit a *cluster behavior*, meaning that the signal is relatively smooth within each cluster, whereas it can vary arbitrarily from cluster to cluster, the GFT defined as in (4) helps emphasizing the presence of clusters [17].

Alternatively, for *directed* graphs, the GFT was defined in [2] as

$$\hat{\boldsymbol{s}} = \mathbf{V}^{-1} \boldsymbol{s} \tag{6}$$

where $\mathbf{V}$ comes from the Jordan decomposition of the asymmetric adjacency matrix $\mathbf{A}$, i.e. $\mathbf{A} = \mathbf{V}\mathbf{J}\mathbf{V}^{-1}$. To estimate variations of the graph Fourier basis, the total variation of a vector was defined in [2] as

$$\text{TV}_A(\boldsymbol{s}) = \|\boldsymbol{s} - \mathbf{A}_{\text{norm}}\, \boldsymbol{s}\|_1 \tag{7}$$

where $\mathbf{A}_{\text{norm}} := \mathbf{A}/|\lambda_{\max}(\mathbf{A})|$. Although giving rise to the elegant theory of algebraic signal processing over graphs [1], [2], [8], the definition of GFT as in (6) raises some important issues: i) in general, the transformation matrix $\mathbf{V}$ is not unitary; ii) the definition of total variation in (7) may give rise to a nonzero total variation for a constant graph signal; iii) the computation of the Jordan decomposition incurs into serious and intractable numerical instabilities even for moderate graph sizes [18].

To cope with these issues, we define the Fourier basis as the set of $N$ orthonormal basis vectors $\mathbf{x}_i, i = 1, \ldots, N$, that minimizes the total variation defined in (2), since it represents the continuous convex Lovász extension of the graph cut size. The first vector is set equal to the constant vector, i.e. $\mathbf{x}_1 = b\,\mathbf{1}$, as this (unit-norm) vector yields a total variation equal to zero. Then, the matrix $\mathbf{X} := (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ containing all the basis vectors, is found as the solution of the following optimization problem

$$\begin{aligned} \min_{\mathbf{X} \in \mathbb{R}^{N \times N}} \quad & \text{DV}(\mathbf{X}) := \sum_{k=1}^{N} \text{DV}(\mathbf{x}_k) \qquad (\mathcal{P}) \\ \text{s.t.} \quad & \mathbf{X}^T \mathbf{X} = \mathbf{I}, \;\; \mathbf{x}_1 = b\mathbf{1} \end{aligned}$$

where $\text{DV}(\mathbf{x}_k) = \sum_{i,j=1}^{N} a_{ji} \left[x_k(i) - x_k(j)\right]_+$. The nonlinear constraints are used to get an orthonormal basis and to prevent the all-zeros solution. Unfortunately, they make $\mathcal{P}$ a a non-convex problem.

## 3. DIRECTED VARIATION MINIMIZATION

To tackle the non-convex orthogonality constraint in $\mathcal{P}$, in this section we proposed the PAMAL algorithm, developed in [12], which solves the orthogonality constrained problem by iteratively updating the primal variables and the multipliers estimates. If we introduce an auxiliary variable $\mathbf{P} = \mathbf{X}$ to split the orthogonality constraint, problem $\mathcal{P}$ becomes

$$\begin{aligned} \min_{\mathbf{X},\mathbf{P}\in\mathbb{R}^{N\times N}} \quad & \mathrm{DV}(\mathbf{X}) \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{P}, \mathbf{x}_1 = b\mathbf{1}, \;\; \mathbf{P}^T\mathbf{P} = \mathbf{I} \end{aligned} \tag{8}$$

which is equivalent to the following one

$$\begin{aligned} \min_{\mathbf{X},\mathbf{P}\in\mathbb{R}^{N\times N}} \quad & f(\mathbf{X},\mathbf{P}) \triangleq \mathrm{DV}(\mathbf{X}) + \delta_{\mathcal{S}_1}(\mathbf{x}_1) + \delta_{\mathcal{S}_t}(\mathbf{P}) \qquad (\mathcal{P}_e) \\ \text{s.t.} \quad & \mathbf{H}(\mathbf{X},\mathbf{P}) \triangleq \mathbf{P} - \mathbf{X} = \mathbf{0}, \end{aligned}$$

where $\delta_{\mathcal{S}}(\mathbf{X})$ denotes the indicator function of set $\mathcal{S}$, i.e.

$$\delta_{\mathcal{S}}(\mathbf{X}) = \begin{cases} 0, & \text{if } \mathbf{X}\in\mathcal{S} \\ +\infty, & \text{otherwise} \end{cases} \tag{9}$$

and the set $\mathcal{S}_\mathbf{1}$ in $\mathcal{P}_e$ is defined as $\mathcal{S}_\mathbf{1} \triangleq \{\mathbf{x}\in\mathbb{R}^N : \|\mathbf{x}\|_F^2 = 1, \mathbf{x}\in \mathrm{span}\{\mathbf{1}\}\}$, whereas $\mathcal{S}_t \triangleq \{\mathbf{P}\in\mathbb{R}^{N\times N} \;:\; \mathbf{P}^T\mathbf{P} = \mathbf{I}\}$ denotes the Stiefel manifold. The basic idea proposed in [12] to solve $\mathcal{P}_e$ is to combine the augmented Lagrangian method [19], [20] and the alternating proximal minimization algorithm, known as the PAM method [21], which deals with non-smooth, non-convex optimization problems. Following the augmented Lagrangian method, we add a penalty term to the objective function so as to assign a high cost to infeasible points. The augmented Lagrangian function associated to problem $\mathcal{P}_e$ is then

$$\mathcal{L}(\mathbf{X},\mathbf{P},\mathbf{\Lambda}) = f(\mathbf{X},\mathbf{P}) + \langle\mathbf{\Lambda},\mathbf{H}(\mathbf{X},\mathbf{P})\rangle + \frac{\rho}{2}\|\mathbf{H}(\mathbf{X},\mathbf{P})\|_F^2$$

where $\rho$ is a positive penalty coefficient, $\mathbf{\Lambda}\in\mathbb{R}^{N\times N}$ represents the multipliers matrix, while the inner matrix product is defined as $\langle\mathbf{A},\mathbf{B}\rangle \triangleq \mathrm{tr}(\mathbf{A}^T\mathbf{B})$. The proposed method reduces problem $\mathcal{P}_e$ to a sequence of problems that, as described in Algorithm 1, alternately update, at each iteration $k$, the following three steps:

1. Compute the critical point $(\mathbf{X}^k,\mathbf{P}^k)$ of the function $\mathcal{L}(\mathbf{X},\mathbf{P},\mathbf{\Lambda}^k;\rho^k)$ by solving
$$(\mathbf{X}^k,\mathbf{P}^k) \triangleq \min_{\mathbf{X},\mathbf{P}\in\mathbb{R}^{N\times N}} \quad \mathcal{L}(\mathbf{X},\mathbf{P},\mathbf{\Lambda}^k;\rho^k); \tag{10}$$
2. Update the multiplier estimates $\mathbf{\Lambda}^k$;
3. Update the penalty parameter $\rho^k$.

The stationary solution $(\mathbf{X}^k,\mathbf{P}^k)$ of problem (10) is found through an approximate algorithm, i.e. an algorithm tolerating a prescribed error value $\epsilon^k$, so that

$$\| \partial\mathcal{L}(\mathbf{X}^k,\mathbf{P}^k,\mathbf{\Lambda}^k;\rho^k) \|_\infty \leq \epsilon^k \tag{11}$$

or, equivalently, by finding a subgradient point $\mathbf{\Theta}^k \in \partial\mathcal{L}$ satisfying (11) with $\mathbf{P}^k \in \mathcal{S}_t$. To find such a point, we adopt a coordinate-descent method with proximal regularization based on the PAM method proposed in [22]. More specifically, at the $k$-th outer iteration of the algorithm, we compute $(\mathbf{X}^k,\mathbf{P}^k)$ by iteratively solving, at each inner iteration $n$, the following proximal regularization of a two blocks Gauss-Seidel method:

$$\begin{aligned} \mathbf{X}^{k,n} = \quad & \underset{\mathbf{X}\in\mathbb{R}^{N\times N},\mathbf{x}_1=b\mathbf{1}}{\arg\min} \quad \mathcal{L}(\mathbf{X},\mathbf{P}^{k,n-1},\mathbf{\Lambda}^k;\rho^k) + \\ & \qquad \frac{c_1^{k,n-1}}{2}\| \mathbf{X} - \mathbf{X}^{k,n-1} \|_F^2 \qquad (\tilde{\mathcal{P}}_{k,n}) \\ \mathbf{P}^{k,n} = \quad & \underset{\mathbf{P}\in\mathbb{R}^{N\times N}}{\arg\min} \quad \mathcal{L}(\mathbf{X}^{k,n-1},\mathbf{P},\mathbf{\Lambda}^k;\rho^k) + \\ & \qquad \frac{c_2^{k,n-1}}{2}\| \mathbf{P} - \mathbf{P}^{k,n-1} \|_F^2 \qquad (\tilde{\mathcal{Q}}_{k,n}) \end{aligned}$$

where the proximal parameters $c_i^{k,n}$ can be arbitrarily chosen as long as they satisfy

$$0 < \underline{c} \leq c_i^{k,n} \leq \bar{c} < \infty, \quad k,n\in\mathbb{N}, i = 1,2, \quad \underline{c} > 0, \bar{c} > 0.$$

The first convex problem $\tilde{\mathcal{P}}_{k,n}$ can be solved through any convex optimization numerical tool, whereas the second problem in $\tilde{\mathcal{Q}}_{k,n}$ admits a closed-form solution as stated in the following proposition (for the proof see Appendix A in [23]).

**Proposition 1** *Define the matrix* $\mathbf{F} \triangleq (c_2^{k,n-1}\mathbf{P}^{k,n-1} + \rho^k\mathbf{X}^{k,n} - \mathbf{\Lambda}^k)(\rho^k + c_2^{k,n-1})^{-1}$ *with SVD decomposition* $\mathbf{F} = \mathbf{Q}\mathbf{\Sigma}\mathbf{T}^T$ *where* $\mathbf{Q},\mathbf{T}\in\mathbb{R}^{N\times N}$ *are unitary matrices, while* $\mathbf{\Sigma}$ *is a diagonal matrix with entries the singular values of* $\mathbf{F}$. *The optimal solution of the non-convex problem* $\tilde{\mathcal{Q}}_{k,n}$ *is given by* $\mathbf{P}^{k,n} = \mathbf{Q}\mathbf{T}^T$.

Algorithm 2 describes the method to solve problems $\tilde{\mathcal{P}}_{k,n}$, $\tilde{\mathcal{Q}}_{k,n}$ in step 1 of Algorithm 1. The inner iterations terminate when there exists a subgradient point $\mathbf{\Theta}^{k,n}\in\partial\mathcal{L}(\mathbf{X}^{k,n},\mathbf{P}^{k,n},\mathbf{\Lambda}^k;\rho^k)$ satisfying $\| \mathbf{\Theta}^{k,n} \|_\infty \leq \epsilon^k$, $\mathbf{P}^{k,n}\in\mathcal{S}_t$, where $\mathbf{\Theta}^{k,n} \triangleq (\mathbf{\Theta}_1^{k,n},\mathbf{\Theta}_2^{k,n})$ with the subgradients given by

$$\begin{aligned} \mathbf{\Theta}_1^{k,n} &= c_1^{k,n-1}(\mathbf{X}^{k,n-1} - \mathbf{X}^{k,n}) + \rho^k(\mathbf{P}^{k,n-1} - \mathbf{P}^{k,n}) \\ \mathbf{\Theta}_2^{k,n} &= c_2^{k,n-1}(\mathbf{P}^{k,n-1} - \mathbf{P}^{k,n}). \end{aligned}$$

For updating the multipliers matrix in step 2 of Algorithm 1, we adopt the classical first-order approximation by imposing that the estimates of multipliers must be bounded. For the setting of the remaining parameters of the proposed algorithm we will assume that: i) the sequence of positive tolerance parameters $\{\epsilon^k\}_{k\in\mathbb{N}}$ is chosen such that $\lim_{k\to\infty}\epsilon^k = 0$; ii) the penalty parameter $\rho^k$ is updated according to the infeasibility degree by following the rule described in step 3 of Algorithm 1 [12], [20]. For the convergence proof of Algorithm 2 to a stationary solution of problem $\mathcal{P}_e$, see [23].

We present now some numerical results to assess the effectiveness of the proposed strategy to find a GFT basis. The

**Algorithm 1** :PAMAL method

Given the parameters $\{\epsilon^k\}_{k\in\mathbb{N}}$, $0 < \epsilon^k < 1$, $\tau \in [0,1)$, $\gamma > 1$, $k = 1$, $\rho^k > 0$, $\mathbf{\Lambda}^k \in \mathbb{R}^{N\times N}$, $\mathbf{\Lambda}_{min} \leq \mathbf{\Lambda}^k \leq \mathbf{\Lambda}_{max}$ with $-\infty < [\mathbf{\Lambda}_{min}]_{i,j} \leq [\mathbf{\Lambda}_{max}]_{i,j} < \infty, \forall i,j$

**Repeat**

**Step.1** Compute $(\mathbf{X}^k, \mathbf{P}^k)$ as in Algorithm 2 such that there exists $\mathbf{\Theta}^k \in \partial\mathcal{L}(\mathbf{X}^k, \mathbf{P}^k, \mathbf{\Lambda}^k; \rho^k)$ with $\| \mathbf{\Theta}^k \|_\infty \leq \epsilon^k$, $(\mathbf{P}^k)^T\mathbf{P}^k = \mathbf{I}$

**Step.2** Update the multiplier estimates

$$\mathbf{\Lambda}^{k+1} = [\mathbf{\Lambda}^k + \rho^k(\mathbf{P}^k - \mathbf{X}^k)]_{\mathcal{T}}$$

where $[\cdot]_{\mathcal{T}}$ is the projection on $\mathcal{T} \triangleq \{\mathbf{\Lambda} : \mathbf{\Lambda}_{min} \leq \mathbf{\Lambda} \leq \mathbf{\Lambda}_{max}\}$

**Step.3** Set $\mathbf{R}^k = \mathbf{P}^k - \mathbf{X}^k$, and update the penalty parameter

$$\rho^{k+1} = \begin{cases} \rho^k & \text{if} \quad \| \mathbf{R}^k \|_\infty \leq \tau \| \mathbf{R}^{k-1} \|_\infty \\ \gamma\rho^k & \text{otherwise} \end{cases}$$

$k = k + 1$

**until convergence**

**Algorithm 2** :PAM method for solving step 1 in Algorithm 1

Let $(\mathbf{X}^{1,0}, \mathbf{P}^{1,0})$ any finite initialization. For $k \geq 2$, set $(\mathbf{X}^{k,0}, \mathbf{P}^{k,0}) = (\mathbf{X}^{k-1}, \mathbf{P}^{k-1})$, $n = 0$.

**Repeat**

**Step.1** Set $n = n + 1$. Compute $\mathbf{X}^{k,n}$ by solving problem $\tilde{\mathcal{P}}_{k,n}$

**Step.2** $\mathbf{P}^{k,n} = \mathbf{Q}\mathbf{T}^T$ where $\mathbf{Q}, \mathbf{T}$ come from the following SVD decomposition

$$\mathbf{Q}\mathbf{\Sigma}\mathbf{T}^T = \frac{c_2^{k,n-1}\mathbf{P}^{k,n-1} + \rho^k\mathbf{X}^{k,n} - \mathbf{\Lambda}^k}{\rho^k + c_2^{k,n-1}}$$

**Step.3** Set $(\mathbf{X}^k, \mathbf{P}^k) = (\mathbf{X}^{k,n}, \mathbf{P}^{k,n})$, $\mathbf{\Theta}^k = \mathbf{\Theta}^{k,n}$

**until** $\| \mathbf{\Theta}^{k,n} \|_\infty \leq \epsilon^k$

parameters of the PAMAL method are set as $\tau = 0.5$, $\gamma = 1.5$, $\rho^1 = 50$, $\epsilon^k = (0.9)^k, \forall k \in \mathbb{N}$, $\mathbf{\Lambda}_{\min} = -1000 \cdot \mathbf{I}$ $\mathbf{\Lambda}_{\max} = 1000 \cdot \mathbf{I}$, $\mathbf{\Lambda}^1 = \mathbf{0}$, $\underline{c} = c_i^{k,n} = \bar{c} = 0.5$, $\forall i,k,n$. As a first example, we consider a directed graph composed of $N = 15$ nodes. In Figure 1, we plot the basis vectors $\{\mathbf{x}_k\}_{k=1}^{15}$, obtained by running Algorithm 1 (the dashed lines correspond to directed edges). The intensity of the vector entries is encoded in the color associated to each vertex and the basis vectors are ordered according to increasing values of the directed variation $\mathrm{DV}(\mathbf{x}_k)$. We can notice how the basis favors the appearance of clusters. Furthermore, the basis vectors assume an exactly constant value within each cluster, in contrast with the Laplacian method where the eigenvectors of $\boldsymbol{L}$ are only slowly varying within the clusters, but not exactly constant. We also compare our method to the two basic definitions of GFT proposed in [3], [2]. As a graph model, we assume scale-free graphs with $N = 20$ nodes. To make the comparison possible, since the Laplacian-based method of [3] assumes undirected graphs, and also to avoid complex vectors for which the directed total variation DV does not represent a valid metric, we focus on undirected scale-free graphs, where the total variation is as in (3). In Fig. 2 we plot the average total variations versus the average minimum degree, for the following three cases: a) $\mathrm{TV}_1(\mathbf{X}^*)$ derived by solving problem $\mathcal{P}$ using the PAMAL method; b) $\mathrm{TV}_1(\mathbf{V})$ where $\mathbf{V}$ are the eigenvectors of the adjacency matrix according to the GFT

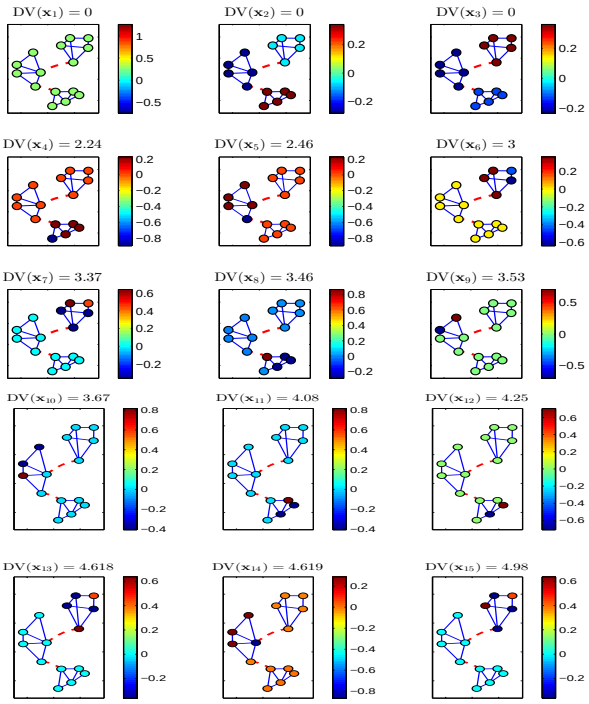


**Fig. 1**: Optimal basis vector for Algorithm 1.

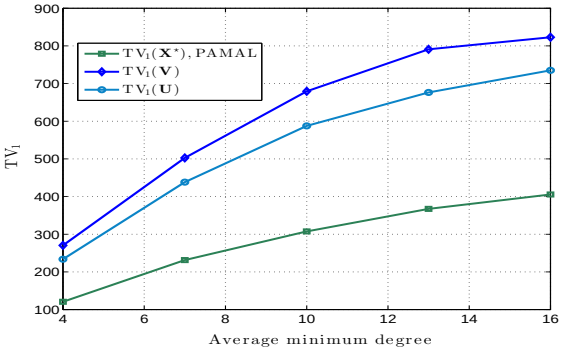


**Fig. 2**: Averaged total variation versus the average minimum degree according to alternative GFT definitions.

defined in (6); c) $\mathrm{TV}_1(\mathbf{U})$ where $\mathbf{U}$ are the eigenvectors of the Laplacian matrix by assuming the GFT as in (4). As we can notice from Fig. 2, our GFT basis leads to a substantial performance improvement in terms of total variation minimization. This is due to the capability of the proposed approach to favor exactly constant values within each cluster. In summary, the proposed method, although maybe loosing the elegance of [1], yields a GFT basis which is unitary by construction, it is numerically robust and it helps to identify clusters, an operation implicit in many applications of GSP.

## 4. REFERENCES


[1] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.

[2] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.

[3] 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," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[4] S. K. Narang, G. Shen, and A. Ortega, "Unidirectional graph-based wavelet transforms for efficient data gathering in sensor networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Dallas, TX, Mar. 2010, pp. 2902–2905.

[5] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4845–4860, Sep. 2016.

[6] I. Pesenson, "Sampling in Paley-Wiener spaces on combinatorial graphs," *Trans. of the American Math. Society*, vol. 360, no. 10, pp. 5603–5627, Oct. 2008.

[7] X. Zhu and M. Rabbat, "Approximating signals supported on graphs," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 3921–3924.

[8] M. Püschel and J. M. F. Moura, "Algebraic signal processing theory: Foundation and 1-D time," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3572–3585, 2008.

[9] M. Püschel and J. M. F. Moura, "Algebraic signal processing theory: 1-D space," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3586–3599, 2008.

[10] L. Lovász, "Submodular functions and convexity," *in A. Bachem et al. (eds.) Math. Program. The State of the Art, Springer Berlin Heidelberg*, pp. 235–257, 1983.

[11] F. Bach, *Learning with Submodular Functions: A Convex Optimization Perspective*, vol. 6, Foundations and Trends in Machine Learning, 2013.

[12] W. Chen, H. Ji, and Y. You, "An augmented Lagrangian method for $l_1$-regularized optimization problems with orthogonality constraints," *SIAM J. Scientific Computing*, vol. 38, no. 4, pp. B570–B592, 2016.

[13] M. Newman, *Networks: An Introduction*, New York, NY, USA: Oxford Univ. Press, 2010.

[14] L. Jost, S. Setzer, and M. Hein, "Nonlinear eigenproblems in data analysis: Balanced graph cuts and the ratioDCA-Prox," *in Extraction of Quantifiable Information from Complex Systems, Springer Intern. Publishing*, vol. 102, pp. 263–279, 2014.

[15] A. Chambolle and J. Darbon, "A parametric maximum flow approach for discrete total variation regularization," *in O. Lezoray, L. Grady Image Processing and Analysis with Graphs: Theory and Practice. CRC Press*, pp. 93–110, 2012.

[16] F. Lozes, A. Elmoataz, and O. Lézoray, "Partial difference operators on weighted graphs for image processing on surfaces and point clouds," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3896–3909, Sep. 2014.

[17] F. R. K. Chung, *Spectral Graph Theory*, American Mathematical Society, 1997.

[18] G. H. Golub and J. H. Wilkinson, "Ill-conditioned eigensystems and computation of the Jordan canonical form," *SIAM Review*, vol. 18, no. 4, pp. 578–619, Oct. 1976.

[19] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt, "On augmented Lagrangian methods with general lower–level constraints," *SIAM J. Optimiz.*, vol. 18, no. 4, pp. 1286–1309, 2007.

[20] D. P. Bertsekas, *Constraint optimization and Lagrange multiplier methods*, Belmont Massachusetts: Athena Scientific, 1999.

[21] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Math. Program.*, pp. 1–36, 2013.

[22] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi–algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods," *Math. Program.*, vol. 137, no. 1–2, pp. 91–129, 2013.

[23] S. Sardellitti, S. Barbarossa, and P. Di Lorenzo, "On the graph Fourier transform for directed graphs," *submitted to IEEE J. Sel. Topics Signal Process.*, 2016, Available: https://arxiv.org/pdf/1601.05972v2.pdf.