

# SPARSE REPRESENTATION ON GRAPHS BY TIGHT WAVELET FRAMES AND APPLICATIONS

BIN DONG

**ABSTRACT.** In this paper, we introduce a new (constructive) characterization of tight wavelet frames on non-flat domains in both continuum setting, i.e. on manifolds, and discrete setting, i.e. on graphs; discuss how fast tight wavelet frame transforms can be computed and how they can be effectively used to process graph data. We start with defining the quasi-affine systems on a given manifold  $\mathcal{M}$ . The quasi-affine system is formed by generalized dilations and shifts of a finite collection of wavelet functions  $\Psi := \{\psi_j : 1 \leq j \leq r\} \subset L_2(\mathbb{R})$ . We further require that  $\psi_j$  is generated by some refinable function  $\phi$  with mask  $a_j$ . We present the condition needed for the masks  $\{a_j : 0 \leq j \leq r\}$ , as well as regularity conditions needed for  $\phi$  and  $\psi_j$ , so that the associated quasi-affine system generated by  $\Psi$  is a tight frame for  $L_2(\mathcal{M})$ . The condition needed for the masks is a simple set of algebraic equations which are not only easy to verify for a given set of masks  $\{a_j\}$ , but also make the construction of  $\{a_j\}$  entirely painless. Then, we discuss how the transition from the continuum (manifolds) to the discrete setting (graphs) can be naturally done. In order for the proposed discrete tight wavelet frame transforms to be useful in applications, we show how the transforms can be computed efficiently and accurately by proposing the fast tight wavelet frame transforms for graph data (WFTG). Finally, we consider two specific applications of the proposed WFTG: graph data denoising and semi-supervised clustering. Utilizing the sparse representation provided by the WFTG, we propose  $\ell_1$ -norm based optimization models on graphs for denoising and semi-supervised clustering. On one hand, our numerical results show significant advantage of the WFTG over the spectral graph wavelet transform (SGWT) by [1] for both applications. On the other hand, numerical experiments on two real data sets show that the proposed semi-supervised clustering model using the WFTG is overall competitive with the state-of-the-art methods developed in the literature of high-dimensional data classification, and is superior to some of these methods.

## 1. INTRODUCTION

In recent years, we are experiencing rapid advances in information and computer technology, which contribute greatly to the exponential growth of data. To properly handle, process and analyze such huge and often unstructured data sets, sophisticated mathematical tools and efficient computing methods need to be developed. Such huge data sets, commonly referred to as “big data”, are generally modelled as huge graphs living in very high dimensional spaces. Graphs are commonly understood as a certain discretization or a random sample from some smooth Riemannian manifold [2–6]. To understand and analyze graphs and data on graphs (shall be called graph data), the graph Laplacian is widely used to reveal the geometric properties of the graph and plays an important role in many applications such as graph clustering.

In signal and image processing, many methods are transform based. Sparsity of the signal/image to be recovered under a certain transform is the key to the success of many existing algorithms. One of the successful examples is the wavelet frame transform, especially the tight wavelet frame transform [7–19]. The power of tight wavelet frames lies in their ability to sparsely approximate

---

*Key words and phrases.* Tight wavelet frames, sparse approximation on graphs, spectral graph theory, big data, graph clustering.

Bin Dong is with the Beijing International Center for Mathematical Research(BICMR), Peking University, Beijing, China, 100871.

*Email:* dongbin@math.pku.edu.cn

*Phone:* (8610)62744091.

This work was supported in part by NSF DMS-1418772, and by the Thousand Talents Plan of China.

piecewise smooth functions and the existence of fast decomposition and reconstruction algorithms. Recently, geometric properties of tight wavelet frames were discovered by connecting them to differential operators under variational and PDE frameworks [17–19].

The success of wavelet frames for data defined on flat domains motivates much research on generalizing wavelets and wavelet frames to curved, irregular and unstructured domains. In this paper, we introduce a (constructive) characterization of tight wavelet frames on non-flat domains in both continuum (on manifolds) and discrete (on graphs) setting, discuss how fast tight wavelet frame transforms can be computed and how they can be effectively used to process and analyze graph data. The basic idea is to understand eigenfunctions of Laplace-Beltrami operator (graph Laplacian in discrete setting) as Fourier basis on manifolds (graphs in discrete setting) and the associated eigenvalues as frequency components. This idea was used earlier by [1] in the discrete setting. In this paper, we further observe that Quasi-affine systems generated by dilations and shifts of wavelet functions can be defined on manifolds. When the elements in the quasi-affine system are generated from a refinable function, the transition from continuum (manifolds) to discrete (graphs) setting can be done very naturally. More importantly, such consideration makes the construction of various types of tight wavelet frames on manifolds/graphs totally painless, and it ensures the existence of fast decomposition and reconstruction algorithms which is crucial to many applications.

Given a compact and connected Riemannian manifold  $(\mathcal{M}, g)$ , denote  $L_2(\mathcal{M})$  the space of square integrable functions on  $\mathcal{M}$ . We start with defining the quasi-affine system on  $\mathcal{M}$ . The quasi-affine system is formed by generalized dilations and shifts of a finite collection of wavelet functions  $\Psi := \{\psi_j : 1 \leq j \leq r\} \subset L_2(\mathbb{R})$ . We further restrict our consideration of  $\Psi$  to those that are generated by a set of masks  $\{a_j : 0 \leq j \leq r\} \subset \ell_2(\mathbb{Z})$ . Then, we present the condition needed for the masks  $\{a_j : 0 \leq j \leq r\}$  (i.e. equation (2.15)) so that the associated quasi-affine system generated by  $\Psi$  is a tight frame for  $L_2(\mathcal{M})$  (Theorem 2.1). The condition on the masks is a simple set of algebraic equations which are not only easy to verify for a given set of masks  $\{a_j\}$ , but also make the construction of  $\{a_j\}$  painless. In particular, we show that under suitable conditions, the quasi-affine system on  $\mathcal{M}$  generated by any set of the framelets constructed from the unitary extension principle [20] on  $\mathbb{R}$  is a tight frame for  $L_2(\mathcal{M})$  (Corollary 2.1). In addition, many masks constructed in [18] satisfy the conditions (2.15) in Theorem 2.1 as well, although they may not satisfy the unitary extension principle. Therefore, Theorem 2.1 is not only a rather generic characterization of tight wavelet frames for  $L_2(\mathcal{M})$ , it also provides a simple way of verifying and constructing various types of tight wavelet frame systems on  $L_2(\mathcal{M})$ .

Thanks to the aforementioned special consideration on  $\Psi$ , i.e. associating  $\Psi$  with a set of masks  $\{a_j\}$ , we discuss how the transition from the continuum (manifolds) to the discrete setting (graphs) can be naturally done. We show that inner products with wavelet frame functions on manifolds can be approximated in the discrete setting by “filtering” with the associated masks on graphs. This leads to multi-level discrete tight wavelet frame transforms (decomposition and reconstruction) on graphs, which can be interpreted as an iterative “convolution” of graph data with properly dilated wavelet frame masks (or filters). In order for the proposed discrete tight wavelet frame transforms to be useful in applications, we show how the transforms can be computed efficiently and accurately. Since most masks we use are trigonometric polynomials, they can be accurately approximated by low-degree Chebyshev polynomials, which enables us to compute the decomposition and reconstruction transforms very efficiently and accurately without computing the eigenvalues and eigenvectors of the graph Laplacian. As a result, the proposed fast tight wavelet frame transform on graphs (WFTG) have a similar computational cost as the traditional fast wavelet frame transforms for images with negligible reconstruction error if the underlying graph Laplacian is a relatively sparse matrix. More importantly, numerical simulations show that the WFTG maps graph data to a set of *sparse* coefficients, which indicates that the proposed tight wavelet frames indeed provide sparse representation on graphs.

Finally, we consider two specific applications of the proposed WFTG: graph data denoising and semi-supervised clustering. Utilizing the sparse representation provided by WFTG, we propose  $\ell_1$ -norm based optimization models on graphs for denoising and semi-supervised clustering problems. These models are motivated from models used in image restoration and image segmentation. We compare the proposed WFTG and the spectral graph wavelet transform (SGWT) of [1] using the proposed models for denoising and semi-supervised clustering. Our numerical results show significant advantage of the WFTG over the SGWT for both applications on synthetic graph data. Furthermore, we compare the proposed semi-supervised clustering model using the WFTG with some of the state-of-the-art graph clustering models using two real data sets. Numerical studies show that our approach is overall competitive with these methods and is superior to some of the methods. We also note that, now that we have WFTG as an effective sparsifying transform for graph data, modeling on graphs can be easily generalized from modeling in image processing and analysis, whenever the same application is still relevant on graphs.

**1.1. Brief Literature Review.** Wavelets and their generalizations are well studied in the past thirty years [21–24]. Their success in various applications are mostly due to the sparse representation they provide for piecewise smooth functions. In the past few decades, there has been much endeavour in the community to generalize (bio)orthogonal wavelets and wavelet frames to functions or data defined on non-flat domains, such as spheres, surfaces, graphs, etc.

There has been a relatively rich literature of wavelets and wavelet frames on 2-dimensional surfaces. Wavelets on sphere were first introduced in [25] using the lifting scheme [26], and later in [27, 28] via a group-theoretical approach. Biorthogonal wavelets with high symmetry for surface multiresolution processing have been constructed in [29–34]. Loop’s scheme-based biorthogonal wavelets have been considered in [35] with the biorthogonal dual wavelets constructed in [36]. Redundant representations on surfaces were introduced by [37], where 6-fold symmetric bi-frames with 4 framelets (frame generators) for triangulated surfaces were introduced. More recently in [38], tight wavelet frames on triangulated and quad surfaces were constructed and applications in surface denoising were considered.

In recent years, there has been much interest in constructing wavelet-like representation of graph data. Successful examples include the wavelets on unweighted graphs by [39], the multiscale scheme on graphs based on lifting by [40], the Haar wavelet transform for rooted binary trees [41] and its generalization treelets [42], the diffusion wavelets [43] and diffusion polynomial frames [44], the wavelets on compact differentiable manifolds [45], the spectral graph wavelet transform by [1, 46, 47], Haar transform for coherent matrices [48], and orthogonal polynomial systems for weighted trees [49].

**1.2. Related Work and Contributions.** Coifman and Maggioni [43] introduced diffusion wavelets on smooth manifolds as well as graphs. Their construction is based on repeated applications of a diffusion operator. The major difference between their work and ours is that the diffusion wavelets are orthonormal. Although an orthonormal system is desirable for some applications such as compression, redundant systems such as tight frames are more robust to errors than (bio)orthogonal systems. In many applications of transformation based data processing, most processing is done in the transform domain such as the widely used thresholding. Errors are inevitably introduced no matter how careful the thresholding operator is designed, especially at the presence of noise. However, if a redundant system is used, there is a good chance that these errors are canceled out after transforming back to the physical domain.

Redundant systems were considered by Maggioni and Mhaskar [44] where they developed a theory of diffusion polynomial frames that is related to our framework. However, they did not provide any algorithm for efficient computation of the decomposition and reconstruction transforms which are crucial to many applications. Geller and Mayeli [45] studied a construction for wavelets on compact differentiable manifolds. In particular, their scaling is defined using a pseudodifferential

operator  $t\Delta e^{t\Delta}$ , where  $\Delta$  is the Laplace-Beltrami operator on the given manifold and  $t$  is the scaling parameter. Wavelets are obtained by applying the pseudodifferential operator to a delta impulse. In our framework, the scaling is defined by  $\phi(t\Delta)$  where  $\phi \in L_2(\mathbb{R})$  is a refinable function. The wavelets are obtained by applying  $\psi_j(t\Delta)$ , with  $\psi_j \in L_2(\mathbb{R})$  and  $1 \leq j \leq r$ , to delta impulses at each points of the given manifold. Moreover, we do not need to assume the manifolds are smooth for our approach.

The work by Hammond, Vandergheynst and Gribonval [1] is most related to the present work. In [1], scaling is defined by  $\phi(t\Delta)$  with  $\phi \in L_2(\mathbb{R})$  and wavelets are obtained by applying  $\psi(t\Delta)$ , with  $\psi \in L_2(\mathbb{R})$ , to delta impulses at each vertex of the given graph. Fast decomposition and reconstruction algorithms based on Chebyshev polynomial approximation were proposed as well, which were referred to as the spectral graph wavelet transform (SGWT). The present work is different from [1] in the following ways:

- (1) We consider characterization and construction of tight wavelet frames on both manifolds and graphs, while only graphs were considered by [1].
- (2) We consider wavelet frame functions  $\psi_j$  that are generated by a refinable function  $\phi$  via the associated mask  $a_j$ , which is an entirely new approach. Such consideration grants a natural transition from continuum (manifolds) to discrete (graphs) settings. More importantly, due to the association of  $\psi_j$  with mask  $a_j$ , the proposed fast transform, i.e. WFTG, is more efficient than the SGWT in the discrete setting (graphs), although the essential tool used by both approaches is the Chebyshev polynomial approximation. This is because for WFTG, the function needs to be approximated is  $\hat{a}_j$  (instead of  $\hat{\psi}_j$  itself as in [1]), which is a trigonometric polynomial. Therefore, the WFTG can be computed using low-degree Chebyshev polynomial approximations, instead of high-degree Chebyshev polynomial approximations needed for the SGWT. Efficient and accurate computation of the forward and inverse transforms are rather important in many applications.
- (3) Theorem 2.1 (continuum) and Theorem 3.1 (discrete) provide a constructive characterization of tight wavelet frames on manifolds and graphs. Systematic construction of tight wavelet frames is totally painless using our approach.
- (4) The tight wavelet frame systems constructed by our approach are potentially more effective in applications than those constructed by [1], as indicated by the comparisons of the WFTG and SGWT given in Section 4 on denoising and semi-supervised clustering problems.

**Contributions.** The contribution of this paper is fourfold.

- (1) *Painless construction of tight wavelet frames on manifolds and graphs.* The construction of tight wavelet frames is essentially reduced to the construction of a set of masks  $\{a_j : 0 \leq j \leq r\}$  such that a set of algebraic equations of  $\hat{a}_j$  is satisfied, i.e. equation (2.15). In particular, any masks constructed from the unitary extension principle, such as those in [20, 50–58] and many masks constructed in [18] satisfy the condition (2.15).
- (2) *Fast transforms available, i.e. the fast tight wavelet frame transform on graphs (WFTG).* The key ingredient of the proposed WFTG is the Chebyshev polynomial approximation of  $\hat{a}_j$ . For a given finitely supported mask  $a_j$ , its Fourier series  $\hat{a}_j$  is a trigonometric polynomial and hence is analytic. Therefore, it can be accurately approximated by low-degree Chebyshev polynomials (see e.g. [59]).
- (3) *The WFTG is more effective than the SGWT in solving denoising and semi-supervised clustering problems.* Numerical examples in Section 4 shows that using the proposed denoising and semi-supervised clustering model, better denoising and clustering results are obtained when the WFTG is used.
- (4) *An effective semi-supervised clustering model using the WFTG is proposed and tested on two real data sets.* High-dimensional data clustering is one of the most important problems in machine learning and big data analysis. Numerical experiments in Section 4.2 showed

that the proposed clustering model using the WFTG is overall competitive with the state-of-the-art methods and is superior to some of these methods.

## 2. TIGHT WAVELET FRAMES ON MANIFOLD $\{\mathcal{M}, g\}$

We start this section with a brief review of some results in eigenvalue problems of Laplace-Beltrami operator on Riemannian manifolds. For a comprehensive review of the subject, one can refer to the book [60]. Given a Riemannian manifold  $(\mathcal{M}, g)$ , we will define a quasi-affine system generated by finitely many functions, and discuss the conditions needed for a quasi-affine system to form a tight frame for  $L_2(\mathcal{M})$ . We end this section by showing some examples.

**2.1. Spectrum and Eigenfunctions of Laplace-Beltrami Operator.** Let  $\{\mathcal{M}, g\}$  be a compact, connected Riemannian manifold with smooth boundary  $S$ . Let  $m \geq 2$  be the dimension of  $\mathcal{M}$ . Let  $\Delta$  be the Laplace-Beltrami operator on  $\mathcal{M}$  with respect to the metric  $g$ . Let  $\{\lambda_p : p = 0, 1, \dots\}$  and  $\{u_p : p = 0, 1, \dots\}$  be the eigenvalues and eigenfunctions of the following eigenvalue problem

$$(2.1) \quad \Delta u + \lambda u = 0,$$

with Dirichlet boundary condition  $u|_S = 0$ . As convention,  $0 < \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \dots$ . The set of eigenfunctions form an orthonormal basis of  $L_2(\mathcal{M})$ , i.e.

$$\langle u_p, u_{p'} \rangle_{L_2(\mathcal{M})} = \int_{\mathcal{M}} u_p(x) u_{p'}^*(x) dx = \delta_{p,p'},$$

and  $\{u_p\}$  is complete in  $L_2(\mathcal{M})$ . Here,  $f^*$  denotes the complex conjugate of  $f$ . Given  $f \in L_2(\mathcal{M})$ , we define

$$\hat{f}[p] = \langle f, u_p \rangle_{L_2(\mathcal{M})},$$

with  $\hat{f} \in \ell_2(\mathbb{Z}^+)$  and  $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$ . Then, we have the following identity

$$\langle f, g \rangle_{L_2(\mathcal{M})} = \langle \hat{f}, \hat{g} \rangle_{\ell_2(\mathbb{Z}^+)} \quad \text{for } f, g \in L_2(\mathcal{M}),$$

where  $\langle \hat{f}, \hat{g} \rangle_{\ell_2(\mathbb{Z}^+)} := \sum_{p=0}^{\infty} \hat{f}[p] \hat{g}^*[p]$ . In particular, we have the following Parseval's identity

$$\|f\|_{L_2(\mathcal{M})}^2 = \|\hat{f}\|_{\ell_2(\mathbb{Z}^+)}^2.$$

For convenience, we shall drop the subscript in  $\langle \cdot, \cdot \rangle_{L_2(\mathcal{M})}$  and  $\langle \cdot, \cdot \rangle_{\ell_2(\mathbb{Z}^+)}$  whenever there is no confusion.

Finally, we recall the following two results from eigenvalue problems of Riemannian geometry. First, Weyl's asymptotic formula [60, 61] gives us the growth rate of  $\{\lambda_p\}$ :

$$(2.2) \quad \lambda_p \asymp p^{\frac{2}{m}},$$

with  $m$  the dimension of the manifold. Here,  $a_n \asymp b_n$  means  $0 < c = \liminf_n |a_n/b_n| \leq \limsup_n |a_n/b_n| = C < \infty$ . The second result is the uniform bound of the eigenfunctions [62]:

$$(2.3) \quad \|u_p\|_{L_{\infty}(\mathcal{M})} \leq C \lambda_p^{\frac{m-1}{4}}.$$

**2.2. Tight Wavelet Frames for  $L_2(\mathcal{M})$ .** We start with defining a system on  $\mathcal{M}$  that mimics the traditional quasi-affine system [20, 24] generated by finitely many functions defined on  $\mathbb{R}$ . With an abuse of terminology, we shall refer to such systems on  $\mathcal{M}$  as quasi-affine systems as well. The quasi-affine system on manifold  $\mathcal{M}$  is generated by dyadic dilations and continuum translations of finitely many functions defined on the real line  $\mathbb{R}$ . Given  $f \in L_2(\mathbb{R})$ , define dilation and translation of  $f$  as

$$(2.4) \quad f_{n,y}^{\mathcal{M}}(x) := \sum_{p=0}^{\infty} \hat{f}(2^{-n}\lambda_p) u_p^*(y) u_p(x), \quad \text{with } n \in \mathbb{Z}, x \in \mathcal{M}, y \in \mathcal{M}.$$



Here,  $\widehat{f}$  is the Fourier transform on the real line. The index  $n$  of  $f_{n,y}^{\mathcal{M}}$  denotes dilation and  $y$  translation. The dyadic dilation  $2^{-n}$  used here is nonessential and one may change it to any  $\gamma^{-n}$  with  $\gamma > 1$ . As one can see that  $f_{n,y}^{\mathcal{M}}$  is defined in the spectral domain mimicking the dilation and translation of functions on Euclidean domains via Fourier transform, which was first used by [1] in discrete setting, i.e. on graphs. Note that the dilation and translation is only formally defined in (2.4), since in general the infinite summation on the right hand side may diverge. To make sure  $f_{n,y}^{\mathcal{M}} \in L_2(\mathcal{M})$ , proper decay condition on  $\widehat{f}$  is needed. We will return to this with more rigourousness after we define the quasi-affine systems.

Define the quasi-affine system  $X(\Psi) \subset L_2(\mathcal{M})$  generated by  $\Psi := \{\psi_j : 1 \leq j \leq r\} \subset L_2(\mathbb{R})$  as

$$(2.5) \quad X(\Psi) := \{\psi_{j,n,y}^{\mathcal{M}} \in L_2(\mathcal{M}) : 1 \leq j \leq r, n \in \mathbb{Z}, y \in \mathcal{M}\},$$

where  $\psi_{j,n,y}^{\mathcal{M}} \in L_2(\mathcal{M})$  is the dilation and translation of  $\psi_j$  as defined in (2.4):

$$(2.6) \quad \psi_{j,n,y}^{\mathcal{M}}(x) := \sum_{p=0}^{\infty} \widehat{\psi}_j(2^{-n}\lambda_p) u_p^*(y) u_p(x), \quad \text{with } n \in \mathbb{Z}, x \in \mathcal{M}, y \in \mathcal{M}.$$

To ensure  $\psi_{j,n,y}^{\mathcal{M}} \in L_2(\mathcal{M})$ , we need to impose a regularity condition on  $\psi_j \in L_2(\mathbb{R})$  or, in other words, a decay condition on  $\widehat{\psi}_j$  as follows:

$$(2.7) \quad |\widehat{\psi}_j(\xi)| \leq C (1 + |\xi|)^{-s} \quad \text{with } s > \frac{2m-1}{4},$$

for all  $\xi \in \mathbb{R}$ . Then, the two results in geometry (2.2) and (2.3) guarantee that  $\psi_{j,n,y}^{\mathcal{M}} \in L_2(\mathcal{M})$ . Note that (2.6) can be written equivalently in Fourier domain as

$$(2.8) \quad \widehat{\psi_{j,n,y}^{\mathcal{M}}}[p] := \widehat{\psi}_j(2^{-n}\lambda_p) u_p^*(y), \quad \text{with } n \in \mathbb{Z}, p \in \mathbb{Z}^+, y \in \mathcal{M}.$$

The main objective of this section is to discuss conditions on  $\psi_j \in L_2(\mathbb{R})$ ,  $1 \leq j \leq r$ , such that  $X(\Psi)$  defined by (2.5) is a tight frame for  $L_2(\mathcal{M})$ . In case  $X(\Psi)$  is a tight frame for  $L_2(\mathcal{M})$ , we shall call  $X(\Psi)$  a tight wavelet frame for  $L_2(\mathcal{M})$  and the elements in  $\Psi$  framelets.

**2.2.1. Characterization of Tight Wavelet Frames for  $L_2(\mathcal{M})$ .** We consider the set of wavelet functions  $\Psi$  that are finite linear combinations of the shifts of a certain refinable (scaling) function  $\phi \in L_2(\mathbb{R})$ , rather than a general set of wavelet functions. This makes our approach different from the existing characterization and construction of wavelet (frame) systems on manifolds/graphs such as [1, 43, 45–47]. Such consideration makes the construction of tight wavelet frames for  $L_2(\mathcal{M})$  totally painless. In fact, as will be shown by Corollary 2.1 that all tight wavelet frames constructed from the unitary extension principle of [20] can also generate tight wavelet frame systems for  $L_2(\mathcal{M})$ . Furthermore, such consideration ensures the existence of fast and accurate decomposition and reconstruction algorithms that can be easily implemented as well (see Section 3). These properties are rather crucial to many applications.

Let  $\phi \in L_2(\mathbb{R})$  be a compactly supported refinable function with finitely supported refinement mask  $a \in \ell_0(\mathbb{Z})$  satisfying the refinement equation

$$\widehat{\phi}(2\xi) = \widehat{a}(\xi) \widehat{\phi}(\xi).$$

Here,  $\widehat{\phi}$  is the Fourier transform of  $\phi \in L_2(\mathbb{R})$  and  $\widehat{a}$  is the Fourier series of  $a$ , which is a trigonometric polynomial since  $a$  is finitely supported. Same as the wavelet functions  $\psi_j$ , we assume the following decay property on  $\widehat{\phi}$  so that the dilation and translation of  $\phi$  is an element in  $L_2(\mathcal{M})$ :

$$(2.9) \quad |\widehat{\phi}(\xi)| \leq C (1 + |\xi|)^{-s} \quad \text{with } s > \frac{2m-1}{4}.$$

Given such refinable function  $\phi \in L_2(\mathbb{R})$  with mask  $a \in \ell_0(\mathbb{Z})$ , we define a set of compactly supported functions  $\Psi := \{\psi_1, \dots, \psi_r\} \subset L_2(\mathbb{R})$  by their associated masks  $a_j \in \ell_0(\mathbb{Z})$ :

$$\widehat{\psi}_j(2\xi) := \widehat{a}_j(\xi)\widehat{\phi}(\xi), \quad 1 \leq j \leq r.$$

Letting  $\psi_0 := \phi$  and  $a_0 := a$ , we can include the refinement equation for  $\phi$  in the above equations, i.e.

$$(2.10) \quad \widehat{\psi}_j(2\xi) = \widehat{a}_j(\xi)\widehat{\phi}(\xi) \quad \text{for } 0 \leq j \leq r.$$

Given  $\psi_{j,n,y}^{\mathcal{M}} \in L_2(\mathcal{M})$  defined by (2.6), the inner product between  $f \in L_2(\mathcal{M})$  and  $\psi_{j,n,y}^{\mathcal{M}}$  satisfies

$$(2.11) \quad \langle f, \psi_{j,n,y}^{\mathcal{M}} \rangle = \langle \widehat{f}, \widehat{\psi_{j,n,y}^{\mathcal{M}}} \rangle = \sum_{p=0}^{\infty} \widehat{f}[p] \widehat{\psi}_j^*(2^{-n}\lambda_p) u_p(y),$$

where the second equality follows from (2.8). Note that the decay conditions (2.7) and (2.9) implies that  $\langle f, \psi_{j,n,y}^{\mathcal{M}} \rangle \in L_2(\mathcal{M})$  for  $0 \leq j \leq r$ . Then, (2.11) can be written equivalently in Fourier domain as

$$(2.12) \quad \langle \widehat{f}, \widehat{\psi_{j,n,y}^{\mathcal{M}}} \rangle[p] = \widehat{f}[p] \widehat{\psi}_j^*(2^{-n}\lambda_p).$$

Now, we can define the following operator  $\mathcal{P}_{n,j}$  on  $L_2(\mathcal{M})$  as

$$\mathcal{P}_{n,j}f := \int_{\mathcal{M}} \langle f, \psi_{j,n,y}^{\mathcal{M}} \rangle \psi_{j,n,y}^{\mathcal{M}} dy, \quad 0 \leq j \leq r.$$

In particular,

$$\mathcal{P}_n f := \mathcal{P}_{n,0} = \int_{\mathcal{M}} \langle f, \phi_{n,y}^{\mathcal{M}} \rangle \phi_{n,y}^{\mathcal{M}} dy.$$

Now, we show that the operator  $\mathcal{P}_{n,j}$  maps  $L_2(\mathcal{M})$  into itself. Note that (2.6) implies that the Fourier coefficients of  $\psi_{j,n,y}^{\mathcal{M}*}(x)$  (the complex conjugate of  $\psi_{j,n,y}^{\mathcal{M}}(x)$ ) with respect to  $y$  satisfy

$$\widehat{\psi_{j,n,y}^{\mathcal{M}*}}(x)[p] = \widehat{\psi}_j^*(2^{-n}\lambda_p) u_p^*(x).$$

Then, by the above identity and (2.12), we have

$$\begin{aligned} (\mathcal{P}_{n,j}f)(x) &= \langle \langle f, \psi_{j,n,y}^{\mathcal{M}} \rangle, \psi_{j,n,y}^{\mathcal{M}*} \rangle = \langle \langle \widehat{f}, \widehat{\psi_{j,n,y}^{\mathcal{M}}} \rangle, \widehat{\psi_{j,n,y}^{\mathcal{M}*}} \rangle \\ &= \sum_{p=0}^{\infty} \widehat{f}[p] \left| \widehat{\psi}_j(2^{-n}\lambda_p) \right|^2 u_p(x). \end{aligned}$$

By the decay conditions (2.7) and (2.9) on  $\widehat{\psi}_j$  for  $0 \leq j \leq r$ , it is obvious that  $\mathcal{P}_{n,j}f \in L_2(\mathcal{M})$  for every  $f \in L_2(\mathcal{M})$ . Furthermore, we have

$$(2.13) \quad \widehat{\mathcal{P}_{n,j}f}[p] = \widehat{f}[p] \left| \widehat{\psi}_j(2^{-n}\lambda_p) \right|^2, \quad \text{for } 0 \leq j \leq r.$$

By definition,  $\{\psi_j : 0 \leq j \leq r\}$  is uniquely determined by the set of masks  $\{a_j : 0 \leq j \leq r\}$ . Thus, the system  $X(\Psi)$  is uniquely determined by masks  $\{a_j : 0 \leq j \leq r\}$ . The following theorem tells us for what conditions on  $\{a_j : 0 \leq j \leq r\}$ , the corresponding system  $X(\Psi)$  defined by (2.5) is a tight frame for  $L_2(\mathcal{M})$ . In fact, the decay conditions (2.7) and (2.9) can be characterized by the associated masks as well [21]. However, for simplicity and clarity, we will not convert (2.7) and (2.9) to conditions on the masks.

**Theorem 2.1.** *Given the set of compactly supported functions  $\{\psi_j : 0 \leq j \leq r\} \subset L_2(\mathbb{R})$  and the associated trigonometric polynomials (or masks)  $\{\widehat{a}_j(\xi) : 0 \leq j \leq r\}$  satisfying (2.10), assume that the decay conditions (2.7) and (2.9) are satisfied,*

$$(2.14) \quad |\widehat{a}_0(\xi) - 1| \leq C|\xi|$$

for  $\xi$  near the origin, and

$$(2.15) \quad \sum_{j=0}^r |\widehat{a}_j(\xi)|^2 = 1.$$

Then, the system  $X(\Psi)$  is a tight frame for  $L_2(\mathcal{M})$ , i.e.

$$f = \sum_{j=1}^r \sum_{n \in \mathbb{Z}} \int_{\mathcal{M}} \langle f, \psi_{j,n,y}^{\mathcal{M}} \rangle \psi_{j,n,y}^{\mathcal{M}} dy \quad \text{for every } f \in L_2(\mathcal{M}).$$

*Proof.* By (2.13), we have

$$\begin{aligned} \widehat{\mathcal{P}_n f}[p] &= \widehat{\mathcal{P}_{n,0} f}[p] = \widehat{f}[p] \left| \widehat{\phi}(2^{-n} \lambda_p) \right|^2 \\ (\text{by (2.15)}) \quad &= \sum_{j=0}^r \widehat{f}[p] \left| \widehat{a}_j(2^{-n} \lambda_p) \widehat{\phi}(2^{-n} \lambda_p) \right|^2 \\ &= \sum_{j=0}^r \widehat{f}[p] \left| \widehat{\psi}_j(2^{-n+1} \lambda_p) \right|^2 = \sum_{j=0}^r \widehat{\mathcal{P}_{n-1,j} f}[p]. \end{aligned}$$

Thus,

$$\mathcal{P}_n f = \mathcal{P}_{n-1} f + \sum_{j=1}^r \mathcal{P}_{n-1,j} f$$

and hence

$$\mathcal{P}_{n_1} f = \mathcal{P}_{n_2} f + \sum_{j=1}^r \sum_{n=n_2}^{n_1} \mathcal{P}_{n,j} f.$$

We first show that

$$\mathcal{P}_{n_1} f \rightarrow f \quad \text{in } L_2(\mathcal{M}) \text{ as } n_1 \rightarrow \infty.$$

By the assumption (2.14), we have  $\widehat{\phi}(\xi) \rightarrow 1$  as  $\xi \rightarrow 0$  (see e.g. [63, 64]). Then,

$$\|\mathcal{P}_{n_1} f - f\|_{L_2(\mathcal{M})}^2 = \sum_{p=0}^{\infty} \left| \widehat{f}[p] |\widehat{\phi}(2^{-n_1} \lambda_p)|^2 - \widehat{f}[p] \right|^2 = \sum_{p=0}^{\infty} |\widehat{f}[p]|^2 \left| |\widehat{\phi}(2^{-n_1} \lambda_p)|^2 - 1 \right|^2.$$

Since  $\widehat{\phi}(2^{-n_1} \lambda_p) \rightarrow 1$  for each  $p \geq 0$  as  $n_1 \rightarrow \infty$ , and  $\widehat{\phi}$  is bounded, we have  $\|\mathcal{P}_{n_1} f - f\|_{L_2(\mathcal{M})}^2 \rightarrow 0$  as  $n_1 \rightarrow \infty$  (by the dominated convergence theorem).

Finally, we show that

$$\mathcal{P}_{n_2} f \rightarrow 0 \quad \text{in } L_2(\mathcal{M}) \text{ as } n_2 \rightarrow -\infty.$$

Indeed, since  $\widehat{\phi}(\xi) \rightarrow 0$  as  $\xi \rightarrow \infty$ , we have  $\widehat{\mathcal{P}_{n_2} f}[p] = \widehat{f}[p] |\widehat{\phi}(2^{-n_2} \lambda_p)|^2 \rightarrow 0$  as  $n_2 \rightarrow -\infty$ , for  $p \geq 0$ . Thus, the boundedness of  $\widehat{\phi}$  leads to  $\mathcal{P}_{n_2} f \rightarrow 0$  in  $L_2(\mathcal{M})$  as  $n_2 \rightarrow -\infty$  (by the dominated convergence theorem). This concludes the proof of the theorem.  $\square$

### Remark 2.1.

- (1) From Theorem 2.1, we can see that by considering  $\Psi = \{\psi_j\}_j$  that is obtained from the identity (2.10), we can simply start with a properly chosen refinable function  $\phi$  and find a set of masks  $\{a_j\}_j$  such that (2.15) is satisfied. The decay conditions can also be verified using the masks [21]. This makes the construction of tight wavelet frames for  $L_2(\mathcal{M})$  very simple. Such construction of  $\Psi$  is known as the multiresolution analysis (MRA) [65–68] based construction in the literature of wavelets [69, 69, 70] and wavelet frames [20, 56, 58, 71] on flat domains.



- (2) We can see from the above proof that, if Neumann boundary condition is chosen for the eigenvalue problem (2.1) or the underlying manifold  $\mathcal{M}$  has no boundary at all, we will have  $\widehat{\mathcal{P}_{n_2}f}[0] \rightarrow \widehat{f}[0]$  and  $\widehat{\mathcal{P}_{n_2}f}[p] \rightarrow 0$  for  $p \geq 1$  as  $n_2 \rightarrow -\infty$ , since  $\lambda_0 = 0$ . Therefore, the system  $X(\Psi)$  obtained from Theorem 2.1 is a tight wavelet frame for the subspace  $\{f \in L_2(\mathcal{M}) : \langle f, u_0 \rangle = 0\}$  instead of  $L_2(\mathcal{M})$  itself.

By Theorem 2.1, it is immediate that all the compactly supported tight wavelet frames for  $L_2(\mathbb{R})$  constructed from the unitary extension principle (UEP) [20] can generate a tight wavelet frame for  $L_2(\mathcal{M})$  provided that the decay conditions (2.7) and (2.9) are satisfied. Recall that a set of masks  $\{a_j \in \ell_0(\mathbb{Z}) : j = 0, 1, \dots, r\}$  is said to satisfy the UEP [20] if

$$(2.16) \quad \sum_{j=0}^r |\widehat{a}_j(\xi)|^2 = 1 \quad \text{and} \quad \sum_{j=0}^r \widehat{a}_j(\xi) \widehat{a}_j^*(\xi + \pi) = 0.$$

Interested reader should consult [20, 56] for details.

**Corollary 2.1.** *Let  $\{\psi_j : 0 \leq j \leq r\}$  and the associated trigonometric polynomials (or masks)  $\{\widehat{a}_j : 0 \leq j \leq r\}$  be constructed from the UEP. Then  $X(\Psi)$  defined by (2.5) is a tight frame for  $L_2(\mathcal{M})$  provided that the decay conditions (2.7) and (2.9) are satisfied and  $|\widehat{a}_0(\xi) - 1| \leq C|\xi|$  for  $\xi$  near the origin.*

**Remark 2.2.** *By Theorem 2.1, the second condition of the UEP is not needed for  $X(\Psi)$  defined by (2.5) to form a tight frame for  $L_2(\mathcal{M})$ . This is because the system  $X(\Psi)$  is entirely translation invariant. The second condition of the UEP is to make sure aliasing can be canceled, which is not needed for a translation invariant wavelet system (see e.g. [72, Theorem 6.4]). Therefore, other than the masks constructed from the UEP, there are many more examples of  $\{a_j\}_j$  whose corresponding function  $\psi_j$  generate a tight wavelet frame for  $L_2(\mathcal{M})$ . For example, many masks constructed by [18] for discretization of nonlinear PDEs only satisfy (2.15). If their corresponding refinable and wavelet functions satisfy the decay conditions, they generate tight frame systems for  $L_2(\mathcal{M})$  by Theorem 2.1.*

**2.2.2. Examples of Tight Wavelet Frames on  $\mathcal{M}$ .** Theorem 2.1 tell us that to construct tight wavelet frames for  $L_2(\mathcal{M})$ , we can simply start with finding a finite set of filters such that (2.15) is satisfied. In particular, Corollary 2.1 tells us that, under suitable conditions, tight wavelet frames constructed from the UEP can generate tight wavelet frames for  $L_2(\mathcal{M})$ . This leads us to a huge collection of tight wavelet frames, because in the literature of wavelet frames, numerous sets of masks with various different properties are constructed based on the UEP [20, 50–58]. Here, we shall focus on the sets of masks constructed from B-splines in [20].

Consider a B-spline  $\phi$  of order  $r$  with  $r \geq 1$ :

$$(2.17) \quad \widehat{\phi}(\xi) = e^{-ij\frac{\xi}{2}} \left( \frac{\sin(\xi/2)}{\xi/2} \right)^r,$$

with  $\tau = 0$  when  $r$  is even and  $\tau = 1$  when  $r$  is odd. The corresponding refinement mask  $\widehat{a}_0$  is given as

$$\widehat{a}_0(\xi) = e^{-i\tau\frac{\xi}{2}} \cos^r(\xi/2).$$

We define  $r$  wavelet masks as

$$(2.18) \quad \widehat{a}_j(\xi) = -i^j e^{-i\tau\frac{\xi}{2}} \sqrt{\binom{r}{j}} \sin^j(\xi/2) \cos^{r-j}(\xi/2), \quad 1 \leq j \leq r.$$

It is easy to check that (2.15) is satisfied:

$$\sum_{j=0}^r |\widehat{a}_j(\xi)|^2 = (\cos^2(\xi/2) + \sin^2(\xi/2))^r = 1.$$

Therefore, by Theorem 2.1, the system  $X(\Psi)$  generated by the  $r$  framelets defined by

$$(2.19) \quad \widehat{\psi}_j = -i^j e^{-i\tau \frac{\xi}{2}} \sqrt{\binom{r}{j}} \frac{\cos^{r-j}(\xi/4) \sin^{r+j}(\xi/4)}{(\xi/4)^r}, \quad 1 \leq j \leq r,$$

forms a tight frame for  $L_2(\mathcal{M})$  for all  $r \geq \frac{2m-1}{4}$  with  $m$  the dimension of the manifold  $\mathcal{M}$ .

**Remark 2.3.** When the dimension of  $\mathcal{M}$  is 2, i.e.  $m = 2$ , the framelets  $\Psi$  given by (2.19) for all  $r \geq 1$  can generate a tight wavelet frame for  $L_2(\mathcal{M})$ . However, when  $m \geq 3$ , according to the assumption (2.7), only the wavelet frame systems generated by higher order B-splines can form tight frames. Of course, this assumption may be weakened so that the systems generated by low order B-splines also form a tight frame for  $L_2(\mathcal{M})$ . In the next section, we shall show that in the discrete setting, the set of masks (2.18) for any  $r \geq 1$  always generates a discrete tight frame on graphs. More generally, any set of masks  $\{a_j : 0 \leq j \leq r\}$  satisfying (2.15) can generate a discrete tight wavelet frame on graphs (see Theorem 3.1).

### 3. DISCRETE TIGHT WAVELET FRAME TRANSFORMS

**3.1. Motivations.** Let  $h_{j,n}(y) := \langle f, \psi_{j,n,y}^{\mathcal{M}} \rangle$ . Observe that

$$(3.1) \quad \begin{aligned} \widehat{h}_{j,n-1}[p] &= \widehat{f}[p] \widehat{\psi}_j^*(2^{-n+1}\lambda_p) \\ &= \widehat{f}[p] \widehat{a}_j^*(2^{-n}\lambda_p) \widehat{\phi}^*(2^{-n}\lambda_p) \\ &= \widehat{a}_j^*(2^{-n}\lambda_p) \widehat{h}_{0,n}[p]. \end{aligned}$$

This shows that  $h_{j,n-1}$ , i.e. the continuous wavelet frame coefficients at level  $n-1$  and band  $j$ , can be obtained from  $h_{0,n}$ , i.e. the low frequency coefficients at level  $n$ , by “convolving”  $h_{0,n}$  with the masks  $a_j$ . As further discussed below, we can understand discrete function  $f_G$  on a graph (which is a certain discretization of  $\mathcal{M}$ ) as a sampling of the underlying function  $f$  by sampling  $h_{0,n}(y)$  at certain scale  $n > 0$ . Then, the discrete wavelet frame transforms can be simply defined as “convolutions” of  $f_G$  with filters  $\{a_j\}$ .

**3.2. Discrete Tight Wavelet Frame Transforms on Graphs.** We denote a graph as  $G := \{E, V, w\}$ , where  $V := \{v_k \in \mathcal{M} : k = 1, \dots, K\}$  is a discretization of a given manifold  $\mathcal{M}$ ,  $E \subset V \times V$  is an edge set, and  $w : E \mapsto \mathbb{R}^+$  denotes a weight function. In this paper, we choose the following commonly used weight function

$$w(v_k, v_{k'}) := e^{-\|v_k - v_{k'}\|_2^2 / \sigma}, \quad \sigma > 0.$$

Let  $A := (a_{k,k'})$  be the adjacency matrix

$$a_{k,k'} := \begin{cases} w(v_k, v_{k'}) & \text{if } v_k \text{ and } v_{k'} \text{ are connected by an edge in } E \\ 0 & \text{otherwise,} \end{cases}$$

and  $D := \text{diag}\{d[1], d[2], \dots, d[K]\}$  where  $d[k]$  is the degree of node  $v_k$  defined by  $d[k] := \sum_{k'} a_{k,k'}$ . Let  $\mathcal{L}_0$  be the (unnormalized) graph Laplacian, which takes the following form

$$\mathcal{L}_0 := D - A.$$

In the literature, normalized graph Laplacians were also considered

$$\mathcal{L}_1 := I - D^{-1}A \quad \text{or} \quad \mathcal{L}_2 := I - D^{-1/2}AD^{1/2}.$$

Here, we shall use  $\mathcal{L}$  to denote one of the above three graph Laplacians. In our numerical computations, we will use the unnormalized graph Laplacian. The consistency of the graph Laplacian to the Laplace-Beltrami operator was studied in [3–5].

Denote  $\{(\lambda_k, u_k)\}_{k=0}^{K-1}$  the set of pairs of eigenvalues and eigenfunctions of  $\mathcal{L}$ . Assuming the graph is connected, then we have  $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{K-1}$ . The eigenfunctions form an orthonormal basis for all functions on the graph:

$$\langle u_k, u_{k'} \rangle := \sum_{n=1}^K u_k[n] u_{k'}[n] = \delta_{k,k'}.$$

Let  $f_G : V \mapsto \mathbb{R}$  be a function on the graph  $G$ . Then its Fourier transform is given by

$$\widehat{f_G}[k] := \sum_{n=1}^K f_G[n] u_k[n].$$

**Transition from continuum to discrete.** Suppose  $G$  is a certain discretization of  $\mathcal{M}$  with  $V \subset \mathcal{M}$ . Given a function  $f_G$  on graph  $G$ , we assume that  $f_G$  is sampled from the underlying function  $f : \mathcal{M} \mapsto \mathbb{R}$  by

$$f_G[k] := \langle f, \phi_{N, v_k} \rangle, \quad v_k \in V,$$

with the dilation scale  $N$  being the smallest integer such that

$$\lambda_{\max} := \lambda_{K-1} \leq 2^N \pi.$$

Note that the scale  $N$  is selected such that  $2^{-N} \lambda_k \in [0, \pi]$  for  $0 \leq k \leq K-1$ .

Based on our earlier observation (3.1), we define the *discrete  $L$ -level tight wavelet frame decomposition* as

$$\mathbf{W} f_G := \{W_{j,l} f_G : (j, l) \in \mathbb{B}\}$$

with

$$(3.2) \quad \mathbb{B} := \{(1, 1), (2, 1), \dots, (r, 1), (1, 2), \dots, (r, L)\} \cup \{(0, L)\}$$

and

$$(3.3) \quad \widehat{W_{j,l} f_G}[k] := \begin{cases} \widehat{a_j^*}(2^{-N} \lambda_k) \widehat{f_G}[k] & l = 1, \\ \widehat{a_j^*}(2^{-N+l-1} \lambda_k) \widehat{a_0^*}(2^{-N+l-2} \lambda_k) \dots \widehat{a_0^*}(2^{-N} \lambda_k) \widehat{f_G}[k] & 2 \leq l \leq L. \end{cases}$$

The index  $j$  denotes the band of the transform with  $j = 0$  the low frequency component and  $1 \leq j \leq r$  the high frequency components. The index  $l$  denotes the level of the transform.

Given a graph function  $f_G$ , let  $\alpha := \mathbf{W} f_G := \{\alpha_{j,l} : (j, l) \in \mathbb{B}\}$ , with  $\alpha_{j,l} := W_{j,l} f_G$ , be its tight wavelet frame coefficients. We denote the *discrete tight wavelet frame reconstruction* as  $\mathbf{W}^\top \alpha$ , which is defined by the following iterative procedure in frequency domain

$$(3.4) \quad \widehat{\alpha_{0,L-1}}[k] = \sum_{j=0}^r \widehat{a_j}(2^{-N+L-1} \lambda_k) \widehat{\alpha_{j,L}}[k] \quad \text{for } l = L, L-1, \dots, 1,$$

where  $\alpha_{0,0} := \mathbf{W}^\top \alpha$  is the reconstructed graph data from  $\alpha$ . Note that  $\mathbf{W}$  is obviously a linear transformation, and it is easy to verify that the linear transformation  $\mathbf{W}^\top$  defined by (3.4) is indeed the adjoint of  $\mathbf{W}$  satisfying  $\langle \mathbf{W} f_G, \alpha \rangle = \langle f_G, \mathbf{W}^\top \alpha \rangle$  for all  $f_G$  and  $\alpha$ .

Perfect reconstruction from  $\alpha$  to  $f_G$  through  $\mathbf{W}^\top$  can be verified as long as the masks  $\{a_j : 0 \leq j \leq r\}$  satisfy the condition (2.15) in Theorem 2.1. We have the following theorem stating that  $\{a_j : 0 \leq j \leq r\}$  generates a discrete tight wavelet frame on graphs.

**Theorem 3.1.** *Given a set of masks  $\{a_j : 0 \leq j \leq r\} \subset \ell_0(\mathbb{Z})$ , suppose condition (2.15) is satisfied. Then, the discrete tight wavelet frame transforms  $\mathbf{W}$  and  $\mathbf{W}^\top$  defined on  $G = \{E, V, w\}$  by (3.3) and (3.4) satisfy*

$$\mathbf{W}^\top \mathbf{W} f_G = f_G, \quad \text{for all } f_G : V \mapsto \mathbb{R}.$$

*Proof.* For simplicity, we prove the theorem for the case  $L = 2$ . The proof for general  $L$  is analogous. For  $L = 2$ , we have  $\boldsymbol{\alpha} := \{\alpha_{j,l} : (j,l) \in \mathbb{B}\}$  with  $\alpha_{j,l} := W_{j,l}f_G$  and  $\mathbb{B} = \{(j,l) : 1 \leq j \leq r, l = 1, 2\} \cup \{(0,2)\}$ . By (3.3), we have

$$\widehat{\alpha}_{j,2}[k] = \widehat{a}_j^*(2^{-N+1}\lambda_k)\widehat{a}_0^*(2^{-N}\lambda_k)\widehat{f}_G[k].$$

Then, letting  $l = 2$  in (3.4), we have

$$\begin{aligned}\widehat{\alpha}_{0,1}[k] &= \sum_{j=0}^r \widehat{a}_j(2^{-N+1}\lambda_k)\widehat{\alpha}_{j,2}[k] \\ &= \sum_{j=0}^r |\widehat{a}_j(2^{-N+1}\lambda_k)|^2 \widehat{a}_0^*(2^{-N}\lambda_k)\widehat{f}_G[k] \\ &= \widehat{a}_0^*(2^{-N}\lambda_k)\widehat{f}_G[k],\end{aligned}$$

where the last identity follows from (2.15). Letting  $l = 1$  in (3.4) and similarly, we have

$$\begin{aligned}\widehat{\alpha}_{0,0}[k] &= \sum_{j=0}^r \widehat{a}_j(2^{-N}\lambda_k)\widehat{\alpha}_{j,1}[k] \\ &= \sum_{j=0}^r |\widehat{a}_j(2^{-N}\lambda_k)|^2 \widehat{f}_G[k] \\ &= \widehat{f}_G[k].\end{aligned}$$

This shows that  $\mathbf{W}^\top \mathbf{W} f_G = f_G$ . □

We present some sets of masks satisfying (2.15), and hence generate discrete tight wavelet frames on graphs according to Theorem 3.1. Note that these masks are modified from those given by (2.18) by removing the complex factors.

**Example 3.1.** We present three sets of masks that satisfy (2.15) and hence generate discrete tight wavelet frames on graphs. We shall refer to them as the “Haar”, “linear” and “quadratic” tight wavelet frame systems, and the masks as the “Haar”, “linear” and “quadratic” masks. These masks slice the spectrum in a different but similar way as shown in Figure 1.

(1) **Haar.**

$$\widehat{a}_0(\xi) = \cos(\xi/2) \quad \text{and} \quad \widehat{a}_1(\xi) = \sin(\xi/2).$$

(2) **Linear.**

$$\widehat{a}_0(\xi) = \cos^2(\xi/2), \quad \widehat{a}_1(\xi) = \frac{1}{\sqrt{2}} \sin(\xi) \quad \text{and} \quad \widehat{a}_2(\xi) = \sin^2(\xi/2).$$

(3) **Quadratic.**

$$\begin{aligned}\widehat{a}_0(\xi) &= \cos^3(\xi/2), \quad \widehat{a}_1(\xi) = \sqrt{3} \sin(\xi/2) \cos^2(\xi/2) \\ \widehat{a}_2(\xi) &= \sqrt{3} \sin^2(\xi/2) \cos(\xi/2) \quad \text{and} \quad \widehat{a}_3(\xi) = \sin^3(\xi/2).\end{aligned}$$

**3.3. Fast Tight Wavelet Frame Transform on Graphs (WFTG).** The discrete tight wavelet frame transforms given by (3.3) and (3.4) require the full set of eigenvectors and eigenvalues of the graph Laplacian, which is computationally expensive to obtain for large graphs. A solution to such computation challenge is to use polynomial approximation of the masks, such as the Chebyshev polynomials, so that eigenvalue decomposition of the graph Laplacian is not needed. This approach was first proposed by [1]. However, the masks  $\widehat{a}_j(\xi)$  that we use, as well as those constructed in the literature from the UEP, are trigonometric polynomials (such as (2.18)). Therefore,  $\widehat{a}_j(\xi)$  can be accurately approximated by *low-degree* Chebyshev polynomials (see e.g. [59]) which significantly

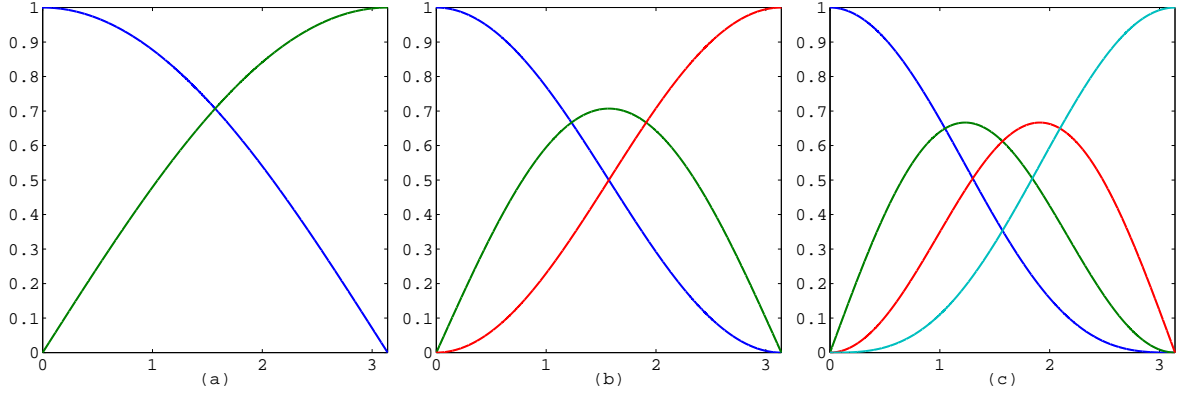


FIGURE 1. Plots of  $\hat{a}_j(\xi)$  for  $\xi \in [0, \pi]$ . Plot (a) presents the “Haar” masks with  $\hat{a}_0$  in blue and  $\hat{a}_1$  in green. Plot (b) presents the “linear” masks with  $\hat{a}_0$  in blue,  $\hat{a}_1$  in green and  $\hat{a}_2$  in red. Plot (c) presents the “quadratic” masks with  $\hat{a}_0$  in blue,  $\hat{a}_1$  in green,  $\hat{a}_2$  in red and  $\hat{a}_3$  in light blue.

reduces the computation cost of the decomposition and reconstruction algorithms. In this section, we describe the details of the fast tight wavelet frame transform on graphs (WFTG) based on polynomial approximation. Comparisons of our approach with that of [1] will be presented later in Section 3.4 and Section 4.

We start with approximating a given function  $g(\xi)$  with  $\xi \in [0, \pi]$  using Chebyshev polynomials  $\{T_k(\xi) : k = 0, 1, \dots\}$ . Recall that the Chebyshev polynomials on  $[0, \pi]$  are defined iteratively by

$$(3.5) \quad T_0 = 1, \quad T_1(\xi) = \frac{\xi - \pi/2}{\pi/2}, \quad T_k(\xi) = \frac{4}{\pi}(\xi - \pi/2)T_{k-1}(\xi) - T_{k-2}(\xi), \quad \text{for } k = 2, 3, \dots$$

Then, given  $g(\xi)$ , we have the following expansion

$$g = \frac{1}{2}c_0 + \sum_{k=1}^{\infty} c_k T_k, \quad \text{where } c_k = \frac{2}{\pi} \int_0^{\pi} \cos(k\theta) g\left(\frac{\pi}{2}(\cos(\theta) + 1)\right) d\theta.$$

When  $g(\xi)$  is smooth, we can use partial sums to accurately approximate  $g(\xi)$ :

$$g(\xi) \approx \mathcal{T}^n(\xi) = \frac{1}{2}c_0 + \sum_{k=1}^{n-1} c_k T_k.$$

We denote the Chebyshev polynomial approximation of  $\hat{a}_j(\xi)$  as

$$\hat{a}_j(\xi) \approx \mathcal{T}_j^n(\xi) = \frac{1}{2}c_{j,0} + \sum_{k=1}^{n-1} c_{j,k} T_k, \quad c_{j,k} = \frac{2}{\pi} \int_0^{\pi} \cos(k\theta) \hat{a}_j\left(\frac{\pi}{2}(\cos(\theta) + 1)\right) d\theta.$$

We denote the Chebyshev approximation of  $\hat{a}_j^*$  as  $\mathcal{T}_j^{n*}$ . Since the masks  $\{\hat{a}_j\}$  we will be using are real-valued (those in Example 3.1), we have  $\mathcal{T}_j^n = \mathcal{T}_j^{n*}$ .

Let us see how to speed up the transformation (3.3) using the above approximation. The Laplacian  $\mathcal{L}$  admits the eigenvalue decomposition  $\mathcal{L} = U\Lambda U^\top$  where  $\Lambda = \text{diag}\{\lambda_0, \lambda_1, \dots, \lambda_{K-1}\}$  and columns of  $U$  are the eigenvectors. Then we can rewrite (3.3) in the following matrix form in physical domain:

$$(3.6) \quad W_{j,l} f_G = \begin{cases} U \hat{a}_j^*(2^{-N}\Lambda) U^\top f_G & l = 1, \\ U \hat{a}_j^*(2^{-N+l-1}\Lambda) \hat{a}_0^*(2^{-N+l-2}\Lambda) \dots \hat{a}_0^*(2^{-N}\Lambda) U^\top f_G & l \geq 2. \end{cases}$$

Here,

$$\hat{a}_j^*(\gamma\Lambda) := \text{diag}\{\hat{a}_j^*(\gamma\lambda_0), \hat{a}_j^*(\gamma\lambda_1), \dots, \hat{a}_j^*(\gamma\lambda_{K-1})\}.$$

If we substitute the approximation  $\hat{a}_j(\xi) \approx \mathcal{T}_j^n(\xi)$  in (3.6) and use the fact that  $\mathcal{T}_j^n$  are polynomials, we obtain the WFTG:

**Fast Tight Wavelet Frame Transform on Graphs (WFTG).**  $Wf_G = \{W_{j,l}f_G : (j,l) \in \mathbb{B}\}$  where

$$(3.7) \quad W_{j,l}f_G := \begin{cases} \mathcal{T}_j^{n*}(2^{-N}\mathcal{L})f_G & l = 1, \\ \mathcal{T}_j^{n*}(2^{-N+l-1}\mathcal{L})\mathcal{T}_0^{n*}(2^{-N+l-2}\mathcal{L}) \dots \mathcal{T}_0^{n*}(2^{-N}\mathcal{L})f_G & l \geq 2. \end{cases}$$

Reconstruction transform  $\mathbf{W}^\top$  can be defined similarly and we have  $\mathbf{W}^\top \mathbf{W} \approx \mathbf{I}$ . Note that, for computational efficiency, the operations  $\mathcal{T}_j^*(s\mathcal{L})f_G$  for decomposition and  $\mathcal{T}_j(s\mathcal{L})f_G$  for reconstruction are computed via the iterative definition of the Chebyshev polynomial (3.5). Therefore, in WFTG, only matrix-vector multiplications are involved.

**Remark 3.1.** As we mentioned earlier, since the masks  $\hat{a}_j(\xi)$  are trigonometric polynomials, they can be accurately approximated by low-degree Chebyshev polynomials (see Table 1 for example). In our simulations, we choose  $n = 8$  which is sufficient for the applications we considered. Note that, if a higher order B-spline tight wavelet frame system is used, Chebyshev polynomials with a higher degree may be needed to achieve a given approximation accuracy. However, just as in image processing, only the lower order systems are mostly used because they provide a better balance between quality and computation efficiency.

**3.4. Numerical Simulations of WFTG.** Given a graph  $G = \{E, V, w\}$  with  $|V| = K$ , we shall use the unnormalized graph Laplacian  $\mathcal{L} = D - A$  as an example. We compute only the largest eigenvalue  $\lambda_{K-1}$ , and choose the initial dilation scale as  $N = \log_2(\frac{\lambda_{K-1}}{\pi})$ .

In our simulations, we choose the “linear” tight wavelet frame system given in Example 3.1, i.e.

$$\hat{a}_0(\xi) = \cos^2(\xi/2), \quad \hat{a}_1(\xi) = \frac{1}{\sqrt{2}}\sin(\xi) \quad \text{and} \quad \hat{a}_2(\xi) = \sin^2(\xi/2).$$

We present approximation errors of  $\mathcal{T}_j^n$  to  $\hat{a}_j$  for several values of  $n$  in the following Table 1, where one can easily see that the  $\hat{a}_j$  can be accurately approximated by low-degree Chebyshev polynomials. In our simulations, we fix  $n = 8$ , i.e. we use Chebyshev polynomials of degree 7 to approximate the masks given in Example 3.1.

TABLE 1. Approximation error  $\|\mathcal{T}_j^n - \hat{a}_j\|_\infty$  with  $\hat{a}_j$  the ‘linear’ masks given in Example 3.1.

Errors	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$
$n = 4$	$2.273 \times 10^{-3}$	$2.022 \times 10^{-2}$	$2.273 \times 10^{-3}$
$n = 5$	$2.273 \times 10^{-3}$	$4.267 \times 10^{-4}$	$2.273 \times 10^{-3}$
$n = 6$	$3.417 \times 10^{-5}$	$4.267 \times 10^{-4}$	$3.417 \times 10^{-5}$
$n = 7$	$3.417 \times 10^{-5}$	$4.775 \times 10^{-6}$	$3.417 \times 10^{-5}$
$n = 8$	$3.762 \times 10^{-7}$	$4.775 \times 10^{-6}$	$3.762 \times 10^{-7}$

To illustrate the effects of the WFTG using the “linear” masks in Example 3.1, we consider a graph  $G = \{E, V, w\}$ , with  $V$  being sampled from a unit sphere. The number of vertices in  $V$  is 16,728. The adjacency matrix  $A$  is generated using the weight function  $w(v_i, v_j) = e^{-\|v_i - v_j\|_2^2/\sigma}$  for  $v_i, v_j \in V$ . In our experiments, we fix  $\sigma = 10$ . Then we threshold  $A$  to limit the number of nearest neighbors of each vertex to 10. The functions  $f_G : V \mapsto \mathbb{R}$  are generated by mapping two images, “Slope” and “Eric”, onto the graph  $G$  (see Figure 2). We perform 4 levels of WFTG and use  $n = 8$  for the Chebyshev polynomial approximation of the masks. The tight wavelet frame coefficients  $W_{j,l}f_G$  for  $0 \leq j \leq 2$  and  $1 \leq l \leq 4$  are shown in Figure 3 and 4. It is worth noticing that the high frequency coefficients  $W_{j,l}f_G$  for  $j = 1, 2$  and  $1 \leq l \leq 4$  are very sparse, which is consistent with



what we normally observe from wavelet frame coefficients of images. In other words, the proposed tight wavelet frames provide a *sparse* representation for graph functions. Sparsity is crucial to many applications. In the next section, we shall consider denoising and semi-supervised clustering utilizing the sparse representation provided by the proposed WFTG.

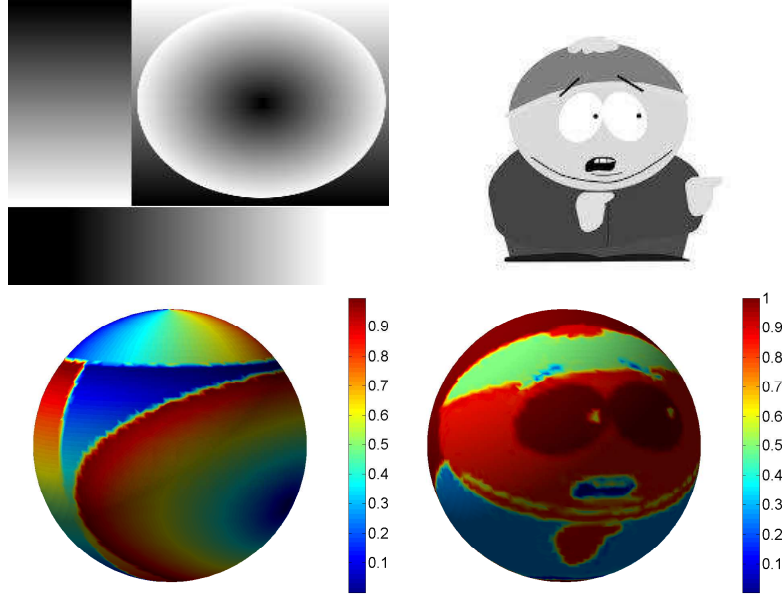


FIGURE 2. This figure shows two images (first row), “Slope” and “Eric”, that are mapped to the graph of the unit sphere to form the graph data  $f_G$  (second row).

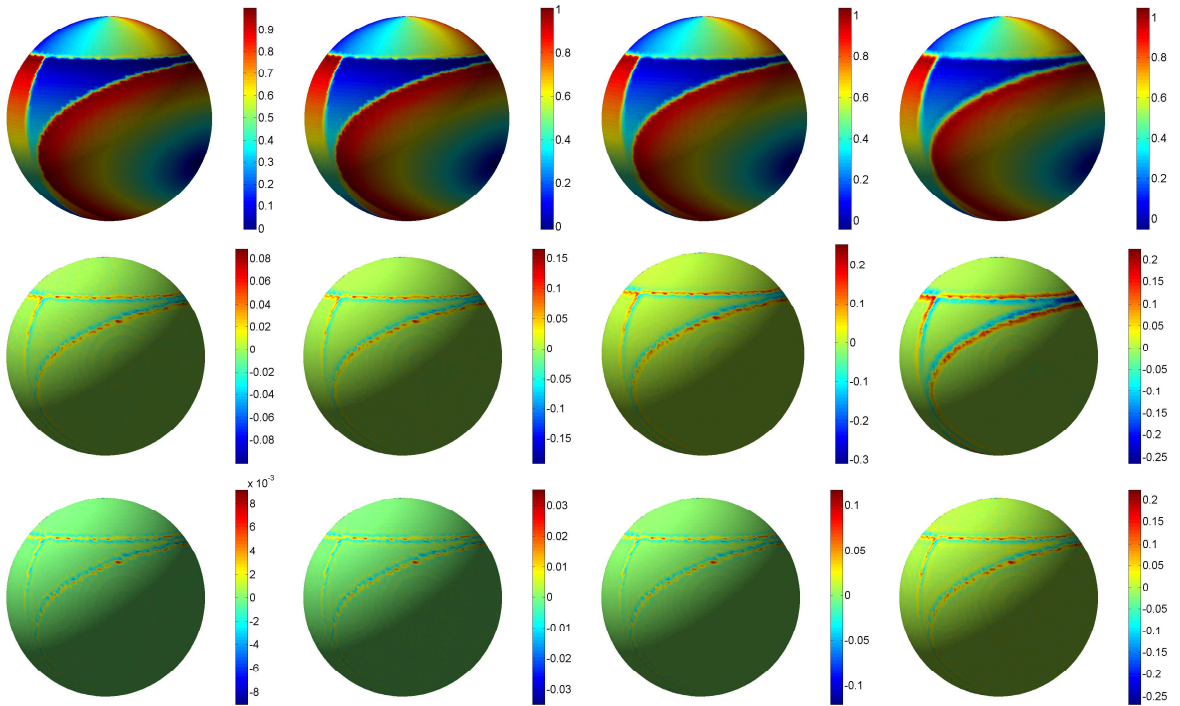


FIGURE 3. This figure shows the images of the tight wavelet frame coefficients  $W_{j,l}f_G$  for  $0 \leq j \leq 2$  (row 1-3) and  $1 \leq l \leq 4$  (column 1-4).

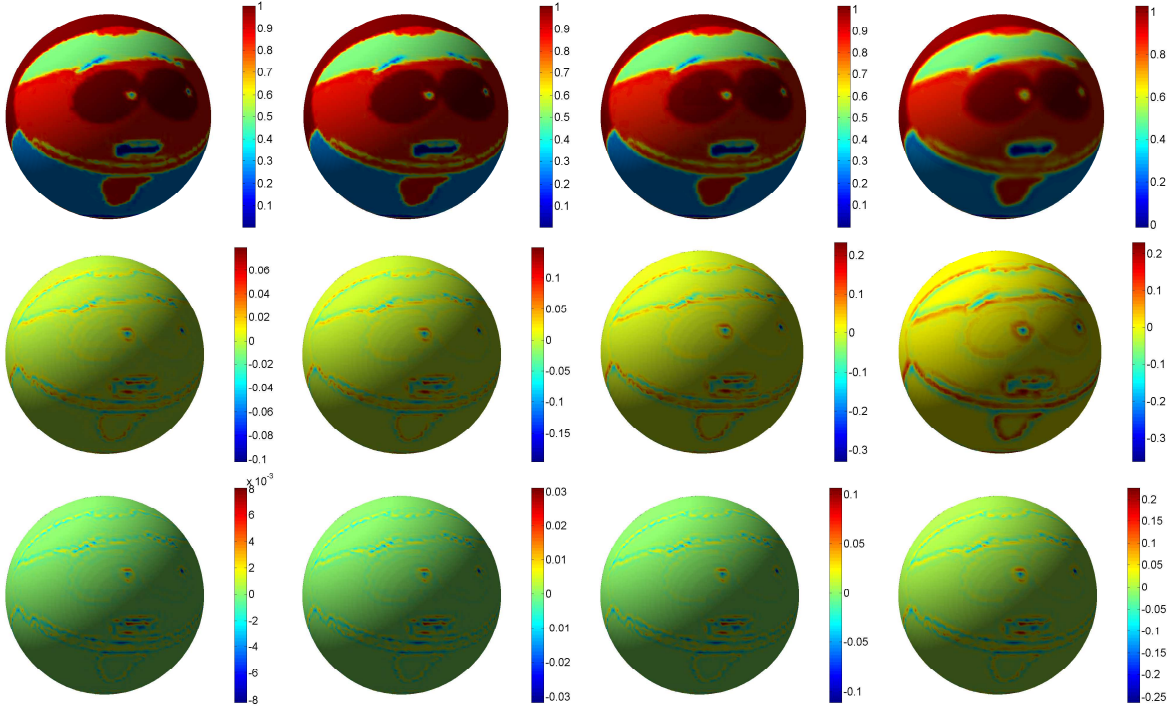


FIGURE 4. This figure shows the images of the tight wavelet frame coefficients  $W_{j,l}f_G$  for  $0 \leq j \leq 2$  (row 1-3) and  $1 \leq l \leq 4$  (column 1-4).

Finally, we compare the computation efficiency of the WFTG with that of the spectral graph wavelet transform (SGWT) of [1], since SGWT is most related to the proposed WFTG. All simulations are conducted on a laptop with Intel(R) Core(TM) i7-4600U CPU (2.1GHz & 2.7GHz) and 8GB RAM running Windows 7 (64-bit). The masks used for WFTG are those given in Example 3.1. The codes for SGWT is downloaded from “<http://wiki.epfl.ch/sgwt>”. The number of scales of SGWT is chosen to be 4 and order of polynomial approximation is chosen to be 25 (by default). The specific graph wavelet used for SGWT is generated by a cubic spline (see [1] for detail).

The computation time and reconstruction errors of a 4-level decomposition and reconstruction of the WFTG and SGWT are summarized in Table 2. For a similar reconstruction error, the WFTG using “Haar” masks is faster than the SGWT; the WFTG using “linear” masks is comparable in speed to SGWT; and the WFTG with “quadratic” masks is slower. However, the WFTG is a more redundant transformation than the SGWT in general. For example, after 4 levels of decomposition, the size of the data in transform domain is 5 times of the size of the original data for “Haar”, 9 times for “linear” and 13 times for “quadratic”, whereas the size of the data in the transform domain of SGWT is 5 times the size of the original data. Therefore, taking the amount of redundancy into account, WFTG is in fact faster than the SGWT in general. In Section 4, we will further show that the WFTG is more effective than the SGWT in graph data denoising and semi-supervised clustering, which indicates that the additional redundancy by the WFTG is beneficial in applications.

In practice, we often need to properly balance between computation efficiency and redundancy when choosing a system to use for a specific application. Higher redundancy may lead to better results while sacrificing efficiency. Taking the denoising problem in Section 4.1 as an example, we can see from Table 3 that using “quadratic” does not significantly improve denoising quality while we generally face a 50% increase in total computation time. Therefore, taking both efficiency and redundancy into account, one may stick to “Haar” and “Linear” or other set of masks with comparable redundancy in applications.

The computation efficiency of the WFTG is mainly due to the efficient approximation of the masks by Chebyshev polynomial, because the masks constructed by our method are analytic which leads to geometric decay of the coefficients of the Chebyshev polynomial approximation (see e.g. [59]). In other words, the masks  $\hat{a}_j(\xi)$  can be accurately approximated by *low-degree* Chebyshev polynomials. In addition, since tight frames are used by WFTG, the reconstruction step is straightforward, and we do not need to solve a least-squares problem to find the inverse as what is used by the SGWT.

TABLE 2. Comparison of computation efficiency of WFTG and SGWT [1]. The reconstruction time is in seconds and errors are measured by the  $\ell_\infty$ -norm.

Image Name	WFTG (Haar)		WFTG (Linear)		WFTG (Quadratic)		SGWT	
	Time (sec.)	Error	Time (sec.)	Error	Time (sec.)	Error	Time (sec.)	Error
Slope	0.246	$3.80 \times 10^{-6}$	0.450	$7.16 \times 10^{-6}$	0.773	$3.80 \times 10^{-6}$	0.519	$5.02 \times 10^{-6}$
Eric	0.262	$3.81 \times 10^{-6}$	0.462	$7.63 \times 10^{-6}$	0.782	$3.81 \times 10^{-6}$	0.509	$1.21 \times 10^{-6}$

#### 4. APPLICATIONS: DENOISING AND SEMI-SUPERVISED CLUSTERING

In the previous section, we discussed how we can define sparse representations on graphs by tight wavelet frames, and how to efficiently compute the decomposition and reconstruction transforms using polynomial approximation of the masks. There are many potential applications of the proposed representation. For example, one may consider a general linear inverse problem on graphs whenever the solution we seek is sparse under the proposed tight wavelet frame transform. In this section, we shall focus on the applications of the WFTG to graph data denoising and semi-supervised clustering.

Given a graph data  $f$ , its Euclidean  $\ell_p$ -norm ( $p = 1$  or  $2$ ) is defined by

$$\|f\|_p := \left( \sum_{k=1}^K |f[k]|^p \right)^{\frac{1}{p}},$$

and its graph- $\ell_p$ -norm is defined by

$$\|f\|_{p,G} := \left( \sum_{k=1}^K |f[k]|^p d[k] \right)^{\frac{1}{p}},$$

where  $d[k]$  is the degree of the node  $v_k \in V$ . Let  $D := \text{diag}\{d[1], d[2], \dots, d[K]\}$ . Obviously, we have  $\|D^{1/p}f\|_p = \|f\|_{p,G}$  with  $D^s := \text{diag}\{d^s[1], d^s[2], \dots, d^s[K]\}$  for  $s \in \mathbb{R}$ . Note that we have dropped the subscript “ $G$ ” of  $f_G$  that we used earlier to differentiate functions on graphs from functions on manifolds, since we shall stay in the discrete setting throughout this section.

What is crucial to the success of the proposed WFTG in solving denoising and clustering problems is the sparse approximation provided by the underlying tight frame system. Because of this, the models we propose for both of the problems take the following generic form

$$(4.1) \quad \min_u \nu \|\mathbf{W}u\|_{1,G} + F(u, f),$$

where  $F(\cdot, \cdot)$  is some fidelity function that takes different forms for different applications, and  $f$  is the observed graph data. The use of the  $\ell_1$ -norm of  $\mathbf{W}u$  is because the desirable solution we seek can be sparsely approximated by the underlying wavelet frame system.

Note that model (4.1) resembles some variational models used in the literature for image and graph data processing and analysis, such as the total variation (TV) based models [73–78]. However, the modeling philosophy of (4.1), as well as most wavelet frame based models for image processing, is based on sparsity of the coefficients in transform domain, whereas for variational models such as TV models, data to be recovered are characterized by carefully chosen function spaces. Nonetheless, comparing most wavelet (frame) based models and variational models, it seems that the major

difference between them is the choice of the underlying transformation that maps the data to a certain domain where it can be sparsely approximated.

On flat domains, fundamental connections between wavelet frame based approach (taking the form of (4.1) with a specific fidelity function  $F(\cdot, \cdot)$ ) and variational methods were established in [17, 19]. In particular, connections to the total variation model [73] was established in [17], and to the Mumford-Shah model [79] was established in [19]. Furthermore, in [18], the authors established a generic connection between iterative wavelet frame shrinkage and general nonlinear evolution PDEs. The series of three papers [17–19] showed that wavelet frame transforms are discretizations of differential operators in both variational and PDE frameworks, and such discretization is superior to some of the traditional finite difference schemes for image restoration.

One interesting question to ask is whether (4.1) is a discrete form of a certain variational model, or in other words, whether the WFTG discretizes differential operators on manifolds? If so, what type of convergence can we establish between the discrete model (4.1) and its corresponding variational model? We do not attempt to answer these questions in this paper since a rigorous justification may be rather technical. But we observe from Figure 3 and 4 that the WFTG does behave like differential operators.

Finally, we note that the quality of model (4.1) for a given application depends on the quality of the transformation  $\mathbf{W}$ . Taking denoising as an example, redundant systems are often better than non-redundancy systems. Furthermore, two different redundant systems may produce very different results even when the same model as (4.1) is used. In this section, we will compare the WFTG and the SGWT for graph denoising and graph clustering problems. Our results will show that WFTG is significantly more effective than the SGWT at least for denoising and clustering problems. Furthermore, we will compare our proposed graph clustering model based on the WFTG with some state-of-the-art clustering models using two real data sets.

**4.1. Denoising.** Denoising on graphs may not be a problem as important as image denoising. However, a good representation together with the modeling based on such representation should be robust to noise in order for them to be useful in practice. Therefore, denoising is a necessary test problem. On the other hand, if the graph data collected in practice is corrupted by a certain type of noise that needs to be removed, denoising is a helpful step to take before any further data analysis. Given a graph  $G = \{E, V, w\}$  with  $|V| = K$ , and a graph function  $\bar{u} : V \mapsto \mathbb{R}$ , the observed data or the data we collect is  $f = \bar{u} + \eta$ , where  $\eta$  is some noise, which shall be assumed to be Gaussian white noise.

Now, we propose to use the following analysis based model for denoising, which was originally proposed in image processing [16, 80, 81]:

$$(4.2) \quad \min_u \|\boldsymbol{\nu} \cdot \mathbf{W}u\|_{1,G} + \frac{1}{2}\|u - f\|_{2,G}^2,$$

where  $\mathbf{W}$  is the tight wavelet frame transform (3.3) and

$$\|\boldsymbol{\nu} \cdot \mathbf{W}u\|_{1,G} := \sum_{(j,l) \in \mathbb{B}} \nu_{j,l} \|W_{j,l}u\|_{1,G}$$

with  $\nu_{j,l} \geq 0$  being the tuning parameters and  $\mathbb{B}$  defined in (3.2). In our simulations, we choose

$$(4.3) \quad \nu_{j,l} = \begin{cases} 4^{-l+1}\nu & \text{for } j \neq 0, 1 \leq l \leq L, \\ 0 & \text{for } j = 0, l = L, \end{cases}$$

where  $\lambda > 0$  is a scalar tuning parameter. Note that model (4.2) can be solved efficiently using the split Bregman algorithm [16, 82], which is also equivalent to the alternating direction method of multipliers (ADMM) [83–85]. Applying the derivation of split Bregman algorithm to our model

(4.2), we have

$$(4.4) \quad \begin{cases} u^{k+1} = \arg \min_u \frac{1}{2} \|u - f\|_{2,G}^2 + \frac{\mu}{2} \|\mathbf{W}u - d^k + b^k\|_2^2, \\ d^{k+1} = \arg \min_d \|\boldsymbol{\nu} \cdot d\|_{1,G} + \frac{\mu}{2} \|d - (\mathbf{W}u^{k+1} + b^k)\|_2^2, \\ b^{k+1} = b^k + \mathbf{W}u^{k+1} - d^{k+1}. \end{cases}$$

Both of the sub-optimization problems in (4.4) have closed-form solutions. Replacing them by their closed-form solutions, we obtain the following denoising algorithm solving (4.2):

$$(4.5) \quad \begin{cases} u^{k+1} = (D + \mu I)^{-1} (Df + \mu \mathbf{W}^\top (d^k - b^k)), \\ d^{k+1} = \mathcal{S}_{\nu_G/\mu}(\mathbf{W}u^{k+1} + b^k), \\ b^{k+1} = b^k + \mathbf{W}u^{k+1} - d^{k+1}, \end{cases}$$

where  $(D + \mu I)^{-1} = \text{diag}\{\frac{1}{d[1]+\mu}, \dots, \frac{1}{d[K]+\mu}\}$ ,  $\nu_G := \{\nu_{j,l} D : (j,l) \in \mathbb{B}\}$  and the soft-thresholding operator  $\mathcal{S}$  is defined pointwise as  $\mathcal{S}_x(y) := \frac{y}{|y|} \max\{|y| - x, 0\}$ . In our numerical implementation, we choose zero initialization of the algorithm, i.e.  $u^0 = 0$  and  $b^0 = d^0 = 0$ .

**Remark 4.1.** We specifically choose the Euclidean  $\ell_2$ -norm in the second term of the first two equations of (4.4), which corresponds to the choice of the Euclidean linear and augmented terms in the derivation of the split Bregman algorithm using the augmented Lagrangian method (see [83–85] or [24, Chapter 4] for details). Another choice, perhaps a more straightforward choice, is to use  $\|\cdot\|_{2,G}^2$ . However, this will lead to an inversion of the non-diagonal matrix  $D + \mathbf{W}^\top \mathbf{D} \mathbf{W}$  in solving for  $u^{k+1}$ , where  $\mathbf{D}$  is a  $(rL + 1)K \times (rL + 1)K$  diagonal matrix given by  $\mathbf{D} = \text{diag}\{D, D, \dots, D\}$ . Therefore, for computation efficiency, we shall use the algorithm (4.5) to solve (4.2).

The graph we consider is the one formed by sampling a unit sphere as described in Section 3.4. We select four examples of graph data as the noise-free data  $\bar{u}$ , where “Slope” and “Eric” were given in Figure 2, and the other two graph data, “Barbara” and “Lena”, are generated by mapping these two widely used images on the discrete sphere. Gaussian white noise is added to  $\bar{u}$  with a standard deviation 0.05. The noise-free and noisy data are visualized in Figure 5, first and second row respectively. The only reason that spherical data is used instead of general graph data is for visualization purpose, so that one can visually judge the denoising quality by examining the images.

In our simulations, when the WFTG is chosen for  $\mathbf{W}$ , we fix the level of decomposition to 1, since using higher decomposition levels only slightly improves the denoising quality while the computation cost is noticeably increased. We choose “Haar”, “linear” and “quadratic” masks given by Example 3.1 for the WFTG and compare the denoising results. For each choice of masks and each graph data, the parameter  $\nu$  of (4.3) is adjusted manually for optimal denoising quality.

For comparison purpose, we also choose the SGWT as the transform  $\mathbf{W}$  in model (4.2). The implementation of the algorithm (4.4) is exactly the same as using WFTG for  $\mathbf{W}$ . For SGWT, we choose the number of scales to be 5 and order of the Chebyshev polynomial approximation as 25. Note that one may choose a different number of scales. However, the denoising quality varies only slightly when larger number of scales are chosen. The specific graph wavelet systems of SGWT used for the comparisons are generated by a cubic spline and the Mexican hat wavelet (see [1] for further details). For optimal denoising quality, we select the same parameter  $\nu$  across all different scales (instead of scaling it like in (4.3)), where the specific value of  $\nu$  is manually chosen for different graph data.

To see the denoising effect using the WFTG and SGWT with different frame systems, we summarize the denoising errors in Table 3. The denoising error is defined by

$$\text{Denoising Error} = \frac{\|u - \bar{u}\|_2}{\|\bar{u}\|_2},$$

where  $u$  is the denoised data and  $\bar{u}$  is the original data. From Table 3, it is obvious that the denoising results using the WFTG are significantly better than those of the SGWT. For WFTG,



we can see that “linear” produces better results than “Haar”, while “quadratic” is comparable with “linear”. However, WFTG using “Quadratic” masks is slower than “Haar” and “Linear”. Therefore, the “linear” tight wavelet frame system seems to be a good balance between computational cost and reconstruction quality for denoising problems. However, when the graph data can be well approximated by a piecewise constant function, such as “Slope”, the “Haar” system seems to be the best choice. For visual examination, denoising results using the “linear” tight wavelet frame system for the WFTG, cubic spline and Mexican hat wavelet for SGWT are shown in the 3rd-5th row of Figure 5. It is obvious from Figure 5 that the visual quality of the denoising results of the WFTG is significantly better than that of the SGWT.

Finally, we note that if one has a linear inverse problem to solve on graph, i.e.  $f = A\bar{u} + \eta$ , then we can change the second term of (4.2) to  $\frac{1}{2}\|Au - f\|_{2,G}^2$  (see e.g. [16, 17]). Also, if the additive noise  $\eta$  is non-Gaussian, such as Poisson, one may modify the norm of the second term of (4.2) to properly incorporate the correct noise statistics (see e.g. [86–88]).

TABLE 3. Comparisons: denoising errors of WFTG and SGWT.

Errors	WFTG “Haar”	WFTG “Linear”	WFTG “Quadratic”	SGWT “Cubic”	SGWT “Mexican Hat”
“Slope”	0.04279	0.04224	0.04307	0.06278	0.05989
“Eric”	0.02995	0.02872	0.02886	0.03926	0.03787
“Barbara”	0.07587	0.07419	0.07376	0.08857	0.08743
“Lena”	0.07972	0.07864	0.07829	0.08692	0.08830

**4.2. Semi-Supervised Clustering.** We consider semi-supervised clustering, where the labeling of a small set of the data is provided in advance. We introduce an optimization model based on the proposed WFTG, along with a fast numerical algorithm. The proposed model and algorithm are motivated by the earlier work on wavelet frame based image segmentation [89] and surface reconstruction [90], as well as the variational frameworks for image segmentation [91, 92] and graph clustering [76, 78]. We shall focus on the binary clustering problem (i.e. two classes). Generalization to multiple-classes clustering can be done following a similar idea as in multiphase image segmentation [93–95] which will be considered in our future studies.

Suppose the given graph  $G = \{E, V, w\}$  contains two clusters. Let  $|V| = K$  and  $\Gamma \subset \{1, 2, \dots, K\}$  be the set of labels where, for  $i \in \Gamma$ , we know which cluster  $v_i$  belongs to. Let  $\Gamma := \Gamma_0 \cup \Gamma_1$  where  $\Gamma_0$  is the index set for cluster 1 and  $\Gamma_1$  for cluster 2. Define  $f : \Gamma \mapsto \mathbb{R}$  as

$$f[i] := \begin{cases} 0 & \text{for } i \in \Gamma_0, \\ 1 & \text{for } i \in \Gamma_1. \end{cases}$$

Our objective is to recover a function  $u : V \mapsto [0, 1]$  with  $u|_{\Gamma} \approx f$ , such that the two sets  $\{i \mid u[i] \geq \beta\}$  and  $\{i \mid u[i] < \beta\}$  provide a good clustering of the given graph for some  $\beta \in (0, 1)$ . To find such  $u$ , we propose the following model

$$(4.6) \quad \min_{u \in [0,1]} \|\boldsymbol{\nu} \cdot \mathbf{W}u\|_{1,G} + \frac{1}{2}\|u|_{\Gamma} - f\|_{2,G}^2,$$

where the first term impose a regularization of the level sets of  $u$ , while the second is the fidelity term making sure that  $u|_{\Gamma} \approx f$ . The parameter  $\boldsymbol{\nu}$  is chosen as in (4.3).

To solve (4.6), we apply the derivation of the split Bregman algorithm again, and obtain the following algorithm

$$(4.7) \quad \begin{cases} u^{k+1} = \arg \min_{u \in [0,1]} \frac{1}{2}\|u|_{\Gamma} - f\|_{2,G}^2 + \frac{\mu}{2}\|\mathbf{W}u - d^k + b^k\|_2^2, \\ d^{k+1} = \mathcal{S}_{\boldsymbol{\nu}_G/\mu}(\mathbf{W}u^{k+1} + b^k), \\ b^{k+1} = b^k + \mathbf{W}u^{k+1} - d^{k+1}, \end{cases}$$



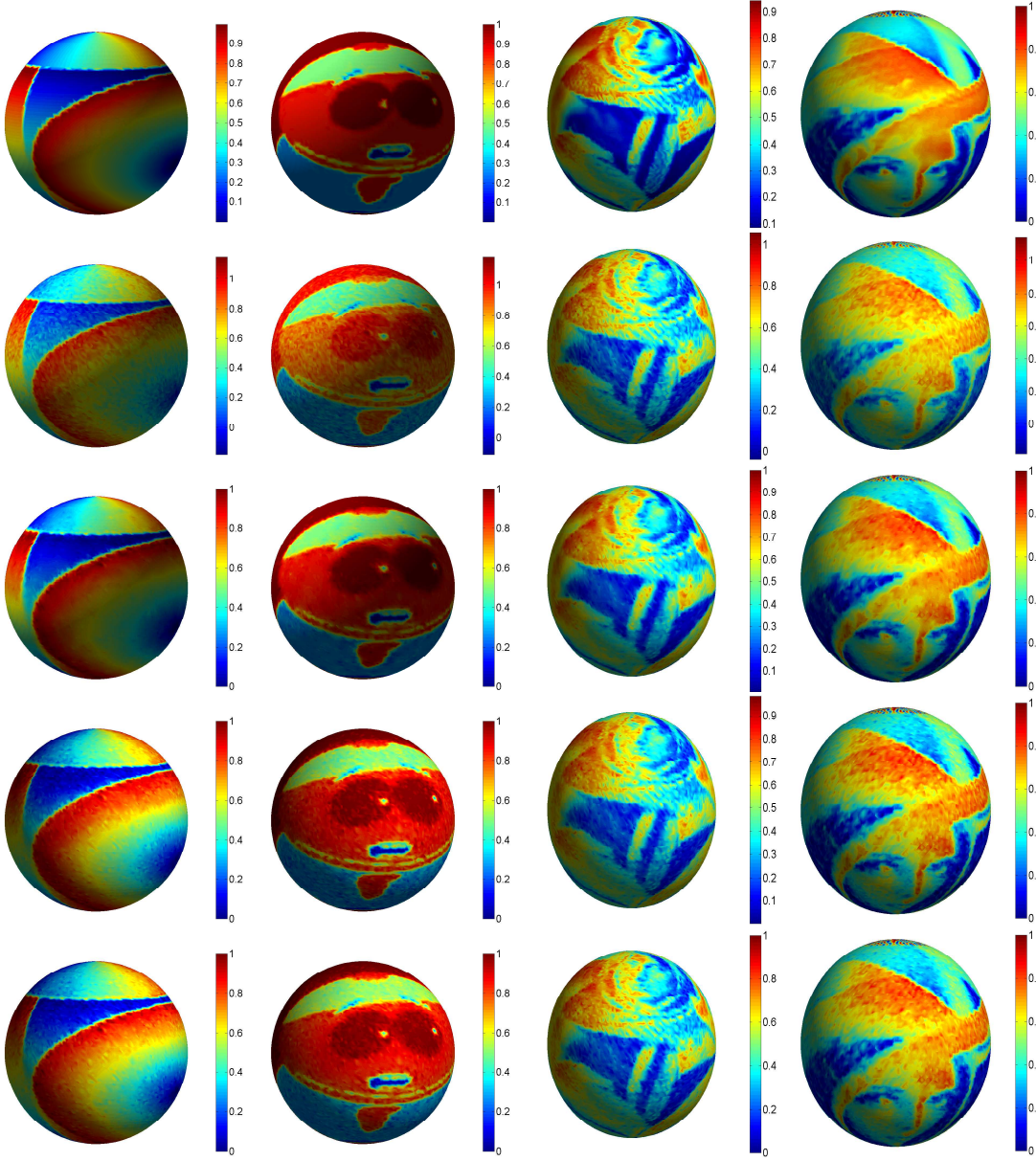


FIGURE 5. This figure shows the four noise-free graph data (first row), the corresponding noisy data (second row) and the denoising results (3rd-5th row) of WFTG (“linear”), SGWT (cubic spline) and SGWT (Mexican hat).

where the subproblem for  $u^{k+1}$  is approximated by the following steps

$$\begin{cases} \left(u^{k+\frac{1}{2}}\right)_{|\Gamma^c} = (\mathbf{W}^\top(d^k - b^k))_{|\Gamma^c}, \\ \left(u^{k+\frac{1}{2}}\right)_{|\Gamma} = (D_\Gamma + \mu I)^{-1} \left(f + \mu (\mathbf{W}^\top(d^k - b^k))_{|\Gamma}\right), \\ u^{k+1} = \max \left\{ \min \left\{ u^{k+\frac{1}{2}}, 1 \right\}, 0 \right\}, \end{cases}$$

where  $D_\Gamma = \text{diag}\{d[k] : k \in \Gamma\}$ .

4.2.1. *Synthetic Data.* Now, we test algorithm (4.7) using a synthetic data set known as the “two moons” which was first used in [96]. This data set is constructed from two half unit circles in  $\mathbb{R}^2$

with centers at  $(0, 0)$  and  $(1, 0.5)$ . For each circle, 1000 points are uniformly chosen and lifted to  $\mathbb{R}^{100}$  by adding i.i.d. Gaussian white noise with variance = 0.02 to each coordinate. The graph is formed from the point set using the exact same way as we discussed earlier in Section 3.4.

Since the solution  $u$  of (4.6) is ideally binary, it is more suitable to use the “Haar” tight wavelet frame system given in Example 3.1. The level of decomposition is chosen to be 1. Parameters are properly tuned for optimal clustering results. Algorithm 4.7 is initialized by

$$(4.8) \quad u^0[k] := \begin{cases} 0, & u_1[k] > 0, \\ 1, & u_1[k] \leq 0, \end{cases}$$

where  $u_1$  is the second eigenvector of  $\mathcal{L}$ , and  $d^0 = b^0 = \mathbf{W}u^0$ .

Same as denoising, we compare the clustering results of the model (4.6) (implemented by algorithm (4.7)) using the proposed WFTG (“Haar”) and the SGWT. For SGWT, Meyer wavelet is used because numerically it produces the best clustering results among all available choices given by the codes online. Figure 6 shows the ground truth of the two clusters and the clustering result by our algorithm using WFTG (“Haar”) and SGWT (“Meyer”) assuming 10% of the labels of the vertices (chosen randomly) are known. More clustering results assuming different percentages of known labels are summarized in Table 4. These results suggest that, for the proposed clustering model (4.6), the WFTG is more suitable than the SGWT. In the next subsection, we shall test model (4.6) with WFTG on two realistic data sets and compare the results with some state-of-the-art clustering methods.

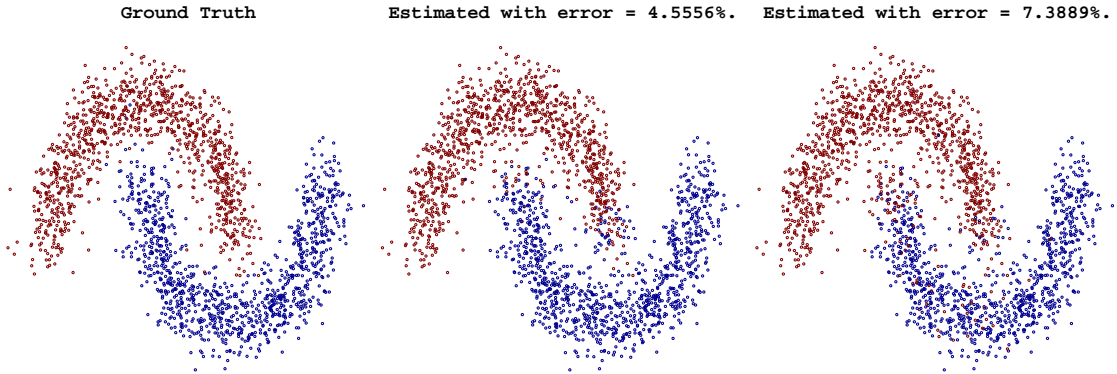


FIGURE 6. This figure shows the ground truth of the two clusters (left) and the estimated clusters using WFTG (middle) and SGWT (right) with error rate 4.5556% and 7.3889% respectively.

TABLE 4. Clustering errors (%) of the data set “two moons” using the WFTG and SGWT.

Known Labels (%)	15	10	5	3
WFTG	4.1765	4.5556	5.9474	6.3402
SGWT	7.4118	7.3889	7.6842	8.7629

**4.2.2. Real Data Sets.** We test the proposed model (4.6) and algorithm (4.7) using the proposed WFTG on two real data sets that are often used as test sets by many recent clustering methods. We compare our method with four recently proposed semi-supervised clustering algorithms, which are: the **max-flow** method and primal augmented Lagrangian method (**PAL**) by [78], the **binary MBO** scheme by [77], and the method based on Ginzburg-Landau (**GL**) functional by [76]. All these methods are based on solving or approximately solving a total variation (TV) based model on graphs, which take the same form as (4.6) with the first term, i.e. the regularization term, of

(4.6) replaced by the graph BV (bounded variation) semi-norm of  $\mathbf{u}$ . Comparisons with some other methods proposed earlier in the literature will also be mentioned.

The first real data set we use is the MNIST data set [97], which is available at <http://yann.lecun.com/exdb/mnist/>. It contains 70000  $28 \times 28$  images of handwritten digits from 0-9. Since our clustering method is binary (2-classes clustering), we choose the subset with digits 4 and 9 to classify since these digits are harder to distinguish. This created a data set of 13782 image vectors, which is either 4 or 9. In our numerical simulations, we randomly draw 500 image vectors from the data set (about 3.62%) and use them to create the known label set  $\Gamma$  and graph data  $f$  of (4.6). We repeat this process 100 times and report the average classification errors (in %) and computation time.

The second data set is the banknote authentication data set from the UCI machine learning repository [98]. It is a data set of 1372 features extracted from images ( $400 \times 400$  pixels) of genuine and forged banknotes using wavelet transforms. The goal is to separate the banknotes into being either genuine or forged. In our numerical simulations, we randomly draw 50 data vectors from the entire data set (about 3.64%) and use them to create the known label set  $\Gamma$  and graph data  $f$  of (4.6). We repeat this process 100 times and report the average classification errors (in %) and computation time.

For both MNIST and banknote data sets, we use “Haar” masks of Example 3.1 for the WFTG. Level of decomposition is chosen to be 1. The parameters  $(\nu, \mu)$  ( $\nu$  is given by (4.3)) and the total number of iterations are chosen to be (2, 0.01) and 200 for MNIST data set, and (0.02, 0.02) and 100 for banknote data set. In all the simulations, the algorithm (4.7) is initialized using the second eigenvector of the graph Laplacian as given by (4.8).

Summary of the clustering results using both data sets is given in Table 5. The results of the methods max-flow, PAL, binary MBO and GL are also included in Table 5, which are taken from [78] where the exact same data sets and experimental settings were used. Our method is only slightly worse than the four latest clustering method for MNIST data set, while it is comparable to the max-flow and PAL methods and much better than binary MBO and GL for banknote data set.

There are many other clustering methods proposed in the literature that were tested on the MNIST data set as well. For example, an unsupervised clustering method was recently proposed by [99], where the classification of the digits 4 and 9 were tested. A purity score (measures the fraction of the vertices that have been assigned to the correct class) of 97.7 was reported. Also, methods based on k-nearest neighbors [97, 100], neural or convolutional nets [97, 100, 101], and SVM [102] report clustering errors of 2.83%-5.00%, 0.35%-4.70%, and 0.68%-1.40%, respectively. We note that all of these approaches are supervised learning methods using 60,000 out of 70,000 digits as a training set, i.e. assuming about 85.7% of the labels are already known. However, our method only takes 3.6% of the entire data set as training set, and the performance is still competitive with the above approaches. Furthermore, no preprocessing or initial dimension reduction on the data set are required for our method, unlike most of the aforementioned algorithms.

TABLE 5. Clustering errors (%) of MNIST and banknote data set. Computation time (in seconds) for our method is included in parenthesis.

Errors (%)	Our Method	Max-Flow	PAL	Binary MBO	GL
MNIST	2.76 (8.5 sec.)	1.52	1.56	1.64	1.75
Banknote	1.64 (2.9 sec.)	1.17	1.71	6.52	3.90

#### ACKNOWLEDGEMENT

The author would like to thank the anonymous reviewers for their constructive suggestions and comments that helped tremendously in improving the presentation, clarity and rigorousness of this paper.

## REFERENCES

- [1] 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.
- [2] R. R. Coifman and S. Lafon, “Diffusion maps,” *Applied and computational harmonic analysis*, vol. 21, no. 1, pp. 5–30, 2006.
- [3] M. Belkin and P. Niyogi, “Towards a theoretical foundation for laplacian-based manifold methods,” in *Learning theory*, pp. 486–500, Springer, 2005.
- [4] M. Hein, J.-Y. Audibert, and U. Von Luxburg, “From graphs to manifolds—weak and strong pointwise consistency of graph laplacians,” in *Learning theory*, pp. 470–485, Springer, 2005.
- [5] E. Giné, V. Koltchinskii, *et al.*, “Empirical graph laplacian approximation of laplace–beltrami operators: Large sample results,” in *High dimensional probability*, pp. 238–259, Institute of Mathematical Statistics, 2006.
- [6] A. Singer, “From graph to manifold laplacian: The convergence rate,” *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 128–134, 2006.
- [7] A. Chai and Z. Shen, “Deconvolution: A wavelet frame approach,” *Numerische Mathematik*, vol. 106, no. 4, pp. 529–587, 2007.
- [8] R. Chan, T. Chan, L. Shen, and Z. Shen, “Wavelet algorithms for high-resolution image reconstruction,” *SIAM Journal on Scientific Computing*, vol. 24, no. 4, pp. 1408–1432, 2003.
- [9] R. Chan, S. Riemenschneider, L. Shen, and Z. Shen, “Tight frame: an efficient way for high-resolution image reconstruction,” *Applied and Computational Harmonic Analysis*, vol. 17, no. 1, pp. 91–115, 2004.
- [10] J. Cai, R. Chan, L. Shen, and Z. Shen, “Restoration of chopped and noddied images by framelets,” *SIAM J. Sci. Comput.*, vol. 30, no. 3, pp. 1205–1227, 2008.
- [11] J. Cai, R. Chan, L. Shen, and Z. Shen, “Convergence analysis of tight framelet approach for missing data recovery,” *Advances in Computational Mathematics*, vol. 31, no. 1, pp. 87–113, 2009.
- [12] J. Cai, R. Chan, and Z. Shen, “A framelet-based image inpainting algorithm,” *Applied and Computational Harmonic Analysis*, vol. 24, no. 2, pp. 131–149, 2008.
- [13] J. Cai, R. Chan, and Z. Shen, “Simultaneous cartoon and texture inpainting,” *Inverse Problems and Imaging (IPI)*, vol. 4, no. 3, pp. 379–395, 2010.
- [14] J. Cai and Z. Shen, “Framelet based deconvolution,” *J. Comp. Math.*, vol. 28, no. 3, pp. 289–308, 2010.
- [15] R. Chan, Z. Shen, and T. Xia, “A framelet algorithm for enhancing video stills,” *Applied and Computational Harmonic Analysis*, vol. 23, no. 2, pp. 153–170, 2007.
- [16] J. Cai, S. Osher, and Z. Shen, “Split Bregman methods and frame based image restoration,” *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, vol. 8, no. 2, pp. 337–369, 2009.
- [17] J. Cai, B. Dong, S. Osher, and Z. Shen, “Image restorations: total variation, wavelet frames and beyond,” *Journal of American Mathematical Society*, vol. 25(4), pp. 1033–1089, 2012.
- [18] B. Dong, Q. Jiang, and Z. Shen, “Image restoration: Wavelet frame shrinkage, nonlinear evolution pdes, and beyond,” *UCLA CAM Report*, vol. 13-78, 2013.
- [19] J. Cai, B. Dong, and Z. Shen, “Image restorations: a wavelet frame based model for piecewise smooth functions and beyond,” *Applied and Computational Harmonic Analysis*, 2015. <http://dx.doi.org/10.1016/j.acha.2015.06.009>.
- [20] A. Ron and Z. Shen, “Affine systems in  $L_2(\mathbb{R}^d)$ : The analysis of the analysis operator,” *Journal of Functional Analysis*, vol. 148, no. 2, pp. 408–447, 1997.
- [21] I. Daubechies, *Ten lectures on wavelets*, vol. CBMS-NSF Lecture Notes, SIAM, nr. 61. Society for Industrial and Applied Mathematics, 1992.
- [22] C. K. Chui, *An introduction to wavelets*, vol. 1. Academic press, 1992.



- [23] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Access Online via Elsevier, 2008.
- [24] B. Dong and Z. Shen, “MRA-Based Wavelet Frames and Applications,” *IAS Lecture Notes Series, Summer Program on “The Mathematics of Image Processing”, Park City Mathematics Institute*, 2010.
- [25] P. Schröder and W. Sweldens, “Spherical wavelets: Efficiently representing functions on the sphere,” in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 161–172, ACM New York, NY, USA, 1995.
- [26] W. Sweldens, “The lifting scheme: A construction of second generation wavelets,” *SIAM Journal on Mathematical Analysis*, vol. 29, no. 2, p. 511, 1998.
- [27] J.-P. Antoine, P. Vandergheynst, *et al.*, “Wavelets on the n-sphere and related manifolds,” *Journal of Mathematical Physics*, vol. 39, no. 8, p. 3987, 1998.
- [28] J.-P. Antoine and P. Vandergheynst, “Wavelets on the 2-sphere: A group-theoretical approach,” *Applied and Computational Harmonic Analysis*, vol. 7, no. 3, pp. 262–291, 1999.
- [29] M. Bertram, “Biorthogonal loop-subdivision wavelets,” *Computing*, vol. 72, no. 1-2, pp. 29–39, 2004.
- [30] M. Bertram, M. A. Duchaineau, B. Hamann, and K. I. Joy, “Generalized b-spline subdivision-surface wavelets for geometry compression,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 10, no. 3, pp. 326–338, 2004.
- [31] Q. Jiang, “Biorthogonal wavelets with 4-fold axial symmetry for quadrilateral surface multiresolution processing,” *Advances in Computational Mathematics*, vol. 34, no. 2, pp. 127–165, 2011.
- [32] Q. Jiang, “Biorthogonal wavelets with 6-fold axial symmetry for hexagonal data and triangle surface multiresolution processing,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 9, 2011.
- [33] H. Wang, K. Qin, and K. Tang, “Efficient wavelet construction with catmull-clark subdivision,” *The Visual Computer*, vol. 22, no. 9-11, pp. 874–884, 2006.
- [34] H. Wang, K. Qin, and H. Sun, “ $\sqrt{3}$ -subdivision-based biorthogonal wavelets,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 13, no. 5, pp. 914–925, 2007.
- [35] A. Khodakovsky, P. Schröder, and W. Sweldens, “Progressive geometry compression,” in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pp. 271–278, Citeseer, 2000.
- [36] B. Han and Z. Shen, “Wavelets from the loop scheme,” *Journal of Fourier Analysis and Applications*, vol. 11, no. 6, pp. 615–637, 2005.
- [37] Q. Jiang and D. K. Pounds, “Highly symmetric bi-frames for triangle surface multiresolution processing,” *Applied and Computational Harmonic Analysis*, vol. 31, no. 3, pp. 370–391, 2011.
- [38] B. Dong, Q. Jiang, C. Liu, and Z. Shen, “Multiscale representation of surfaces by tight wavelet frames with applications to denoising,” *Applied and Computational Harmonic Analysis*, 2015. <http://dx.doi.org/10.1016/j.acha.2015.03.005>.
- [39] M. Crovella and E. Kolaczyk, “Graph wavelets for spatial traffic analysis,” in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 3, pp. 1848–1857, IEEE, 2003.
- [40] M. Jansen, G. P. Nason, and B. W. Silverman, “Multiscale methods for data on graphs and irregular multidimensional situations,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 71, no. 1, pp. 97–125, 2009.
- [41] F. Murtagh, “The haar wavelet transform of a dendrogram,” *Journal of Classification*, vol. 24, no. 1, pp. 3–32, 2007.
- [42] A. B. Lee, B. Nadler, and L. Wasserman, “Treelets: an adaptive multi-scale basis for sparse unordered data,” *The Annals of Applied Statistics*, pp. 435–471, 2008.

- [43] R. R. Coifman and M. Maggioni, “Diffusion wavelets,” *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 53–94, 2006.
- [44] M. Maggioni and H. Mhaskar, “Diffusion polynomial frames on metric measure spaces,” *Applied and Computational Harmonic Analysis*, vol. 24, no. 3, pp. 329–353, 2008.
- [45] D. Geller and A. Mayeli, “Continuous wavelets on compact manifolds,” *Mathematische Zeitschrift*, vol. 262, no. 4, pp. 895–927, 2009.
- [46] M. Gavish, B. Nadler, and R. R. Coifman, “Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning,” in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 367–374, 2010.
- [47] N. Leonardi and D. Van De Ville, “Tight wavelet frames on multislice graphs,” *Signal Processing, IEEE Transactions on*, vol. 61, no. 13, pp. 3357–3367, 2013.
- [48] M. Gavish and R. R. Coifman, “Sampling, denoising and compression of matrices by coherent matrix organization,” *Applied and Computational Harmonic Analysis*, vol. 33, no. 3, pp. 354–369, 2012.
- [49] C. Chui, F. Filbir, and H. Mhaskar, “Representation of functions on big data: graphs and trees,” *Applied and Computational Harmonic Analysis*, 2014.
- [50] A. Ron and Z. Shen, “Compactly supported tight affine spline frames in  $L_2(\mathbb{R}^d)$ ,” *Mathematics of Computation*, vol. 67, no. 221, pp. 191–207, 1998.
- [51] K. Gröchenig and A. Ron, “Tight compactly supported wavelet frames of arbitrarily high smoothness,” *Proceedings of the American Mathematical Society*, vol. 126, no. 4, pp. 1101–1107, 1998.
- [52] C. Chui and W. He, “Compactly supported tight frames associated with refinable functions,” *Applied and Computational Harmonic Analysis*, vol. 8, no. 3, pp. 293–319, 2000.
- [53] A. Petukhov, “Explicit construction of framelets,” *Applied and Computational Harmonic Analysis*, vol. 11, no. 2, pp. 313–327, 2001.
- [54] I. Selesnick, “Smooth Wavelet Tight Frames with Zero Moments\* 1,” *Applied and Computational Harmonic Analysis*, vol. 10, no. 2, pp. 163–181, 2001.
- [55] A. Petukhov, “Symmetric framelets,” *Constructive Approximation*, vol. 19, no. 2, p. 309, 2003.
- [56] I. Daubechies, B. Han, A. Ron, and Z. Shen, “Framelets: MRA-based constructions of wavelet frames,” *Applied and Computational Harmonic Analysis*, vol. 14, pp. 1–46, Jan 2003.
- [57] B. Dong and Z. Shen, “Pseudo-splines, wavelets and framelets,” *Applied and Computational Harmonic Analysis*, vol. 22, no. 1, pp. 78–104, 2007.
- [58] B. Han and Z. Shen, “Dual wavelet frames and Riesz bases in Sobolev spaces,” *Constructive Approximation*, vol. 29, no. 3, pp. 369–406, 2009.
- [59] J. C. Mason and D. C. Handscomb, *Chebyshev polynomials*. CHAPMAN & HALL/CRC Press, 2010.
- [60] I. Chavel, *Eigenvalues in Riemannian geometry*, vol. 115. Academic press, 1984.
- [61] H. Weyl, “Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung),” *Mathematische Annalen*, vol. 71, no. 4, pp. 441–479, 1912.
- [62] D. Grieser, “Uniform bounds for eigenfunctions of the laplacian on manifolds with boundary\*,” *Communications in Partial Differential Equations*, vol. 27, no. 7-8, pp. 1283–1299, 2002.
- [63] A. Cavaretta, W. Dahmen, and C. Micchelli, *Stationary subdivision*. Amer Mathematical Society, 1991.
- [64] Q. Jiang and Z. Shen, “On existence and weak stability of matrix refinable functions,” *Constructive Approximation*, vol. 15, no. 3, pp. 337–353, 1999.
- [65] S. Mallat, “Multiresolution approximations and wavelet orthonormal bases of  $L_2(\mathbb{R})$ ,” *Transactions of the American Mathematical Society*, vol. 315, no. 1, pp. 69–87, 1989.



- [66] Y. Meyer, *Wavelets and operators*. Translated by DH Salinger. Cambridge Studies in Advanced Mathematics, 1992.
- [67] C. De Boor, R. DeVore, and A. Ron, “On the construction of multivariate (pre) wavelets,” *Constructive approximation*, vol. 9, no. 2, pp. 123–166, 1993.
- [68] R. Jia and Z. Shen, “Multiresolution and wavelets,” *Proc. Edinb. Math. Soc., II. Ser.*, vol. 37, no. 2, pp. 271–300, 1994.
- [69] I. Daubechies, “Orthonormal bases of compactly supported wavelets,” *Commun. Pure Appl. Math.*, vol. 41, no. 7, pp. 909–996, 1988.
- [70] A. Cohen, I. Daubechies, and J. Feauveau, “Biorthogonal Bases of Compactly Supported Wavelets,” *Communications on Pure and Applied Mathematics*, vol. 45, no. 5, pp. 485–560, 1992.
- [71] C. Chui, W. He, and J. Stöckler, “Compactly supported tight and sibling frames with maximum vanishing moments,” *Applied and computational harmonic analysis*, vol. 13, no. 3, pp. 224–262, 2002.
- [72] Z. Fan, A. Heinecke, and Z. Shen, “Duality for frames,” *Journal of Fourier Analysis and Applications*, 2015, DOI: 10.1007/s00041-015-9415-0.
- [73] L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Phys. D*, vol. 60, pp. 259–268, 1992.
- [74] G. Gilboa and S. Osher, “Nonlocal operators with applications to image processing,” *Multiscale Model Sim.*, vol. 7, pp. 1005–1028, Jan 2008.
- [75] G. Gilboa and S. Osher, “Nonlocal linear image regularization and supervised segmentation,” *Multiscale Modeling and Simulation*, vol. 6, no. 2, pp. 595–630, 2008.
- [76] A. L. Bertozzi and A. Flenner, “Diffuse interface models on graphs for classification of high dimensional data,” *Multiscale Modeling & Simulation*, vol. 10, no. 3, pp. 1090–1118, 2012.
- [77] E. Merkurjev, T. Kostic, and A. L. Bertozzi, “An mbo scheme on graphs for classification and image processing,” *SIAM Journal on Imaging Sciences*, vol. 6, no. 4, pp. 1903–1930, 2013.
- [78] E. Merkurjev, E. Bae, A. Bertozzi, and X. C. Tai, “Global binary optimization on graphs for classification of high dimensional data,” *UCLA CAM Report*, vol. 14-72, 2014.
- [79] D. Mumford and J. Shah, “Optimal approximations by piecewise smooth functions and associated variational problems,” *Communications on pure and applied mathematics*, vol. 42, no. 5, pp. 577–685, 1989.
- [80] M. Elad, J. Starck, P. Querre, and D. Donoho, “Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA),” *Applied and Computational Harmonic Analysis*, vol. 19, no. 3, pp. 340–358, 2005.
- [81] J. Starck, M. Elad, and D. Donoho, “Image decomposition via the combination of sparse representations and a variational approach,” *IEEE transactions on image processing*, vol. 14, no. 10, pp. 1570–1582, 2005.
- [82] T. Goldstein and S. Osher, “The split Bregman algorithm for L1 regularized problems,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.
- [83] D. Gabay and B. Mercier, “A dual algorithm for the solution of nonlinear variational problems via finite element approximation,” *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1976.
- [84] D. Bertsekas and J. Tsitsiklis, *Parallel and distributed computation: numerical methods*. Prentice-Hall, Inc., 1989.
- [85] J. Eckstein and D. Bertsekas, “On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators,” *Mathematical Programming*, vol. 55, no. 1, pp. 293–318, 1992.
- [86] T. Le, R. Chartrand, and T. J. Asaki, “A variational approach to reconstructing images corrupted by poisson noise,” *Journal of Mathematical Imaging and Vision*, vol. 27, no. 3, pp. 257–263, 2007.

- [87] R. H. Chan and K. Chen, "Multilevel algorithm for a poisson noise removal model with total-variation regularization," *International Journal of Computer Mathematics*, vol. 84, no. 8, pp. 1183–1198, 2007.
- [88] B. Zhang, J. M. Fadili, and J.-L. Starck, "Wavelets, ridgelets, and curvelets for poisson noise removal," *Image Processing, IEEE Transactions on*, vol. 17, no. 7, pp. 1093–1108, 2008.
- [89] B. Dong, A. Chien, and Z. Shen, "Frame based segmentation for medical images," *Communications in Mathematical Sciences*, vol. 9(2), pp. 551–559, 2010.
- [90] B. Dong and Z. Shen, "Frame based surface reconstruction from unorganized points," *Journal of Computational Physics*, vol. 230, pp. 8247–8255, 2011.
- [91] F. Chan and L. Vese, "Active contours without edges," *IEEE Transactions on image processing*, vol. 10, no. 2, pp. 266–277, 2001.
- [92] F. Chan and L. Vese, "An active contour model without edges," *Scale-Space Theories in Computer Vision*, vol. 1682, pp. 141–151, 1999.
- [93] B. Sandberg and T. F. Chan, "A logic framework for active contours on multi-channel images," *Journal of Visual Communication and Image Representation*, vol. 16, no. 3, pp. 333–358, 2005.
- [94] E. Bae, J. Yuan, and X.-C. Tai, "Global minimization for continuous multiphase partitioning problems using a dual approach," *International journal of computer vision*, vol. 92, no. 1, pp. 112–129, 2011.
- [95] C. Tai, X. Zhang, and Z. Shen, "Wavelet frame based multiphase image segmentation," *SIAM Journal on Imaging Sciences*, vol. 6, no. 4, pp. 2521–2546, 2013.
- [96] T. Bühler and M. Hein, "Spectral clustering based on the graph p-laplacian," in *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 81–88, ACM, 2009.
- [97] Y. LeCun and C. Cortes, "The mnist database of handwritten digits," URL <http://yann.lecun.com/exdb/mnist/>, 1998.
- [98] K. Bache and M. Lichman, "Uci machine learning repository," URL <http://archive.ics.uci.edu/ml>, vol. 901, 2013.
- [99] H. Hu, T. Laurent, M. A. Porter, and A. L. Bertozzi, "A method based on total variation for network modularity optimization using the mbo scheme," *SIAM Journal on Applied Mathematics*, vol. 73, no. 6, pp. 2224–2246, 2013.
- [100] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [101] D. C. Ciresan, U. Meier, J. Masci, L. Maria Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, vol. 22, p. 1237, 2011.
- [102] D. Decoste and B. Schölkopf, "Training invariant support vector machines," *Machine learning*, vol. 46, no. 1-3, pp. 161–190, 2002.