GEOMETRIC LEARNING VIA WAVELETS ON MANIFOLDS AND GRAPHS

Yiping Lu.
luyiping9712@pku.edu.cn
School of Mathematic Science, Peking University.
No.5 Yiheyuan Road Haidian District,Beijing,P.R.China 10087.

Abstract. With the development of machine learning, conquering more and more computer vision task seems to be possible via learning.High-dimensional clustering arises frequently from many areas in natural sciences, technical disciplines and social medias. However the lack of theory to express the power of deep learning leads to the huge demand of data while training which is sometimes hard to approach and also leads to huge computation cost. In order to utilize the structure of the datas, we should construct our learning framework beyond the Euclid domain. In this paper, we discussed several wavelet construct on manifolds and their relations and difference. Based on wavelet on manifolds and graphs, we construct a semi-supervised learning algorithm via wavelet structure on manifold to conquer manifold semi-supervised learning tasks.

Keywords. Wavelet, Computer Vision, Machine Learning, Semi-supervisd Learning, Manifold Learning

1. Introduction

As living in the era of big data, data driven learning method have become an important part of our society. With the development of computer and the accumulation of data, many hard tasks can be conquered via learning form huge datasets.As one of the learning method, deep learning has shown its unreasonably powerful expressiveness in a various of computer vision tasks, e.g. object detection, image in-painting and image captioning. However there is no theory that can express the power of the neural networks. As a result, the demand of data is always very huge while training a neural network because they didn't utilize the structure of datasets. While the data is hard to collect, e.g. the medical image labels, or the computational ability is not strong enough, our learning task will be difficult to be compeleted. In order to utilize the structure of the dataset, we should build the learning algorithms on manifold, not on the euclidean domain but a special locally low dimension subspace in the high dimension euclidean domain.

Further more, many data in the real world is not euclidean, e.g. brain, tooth, social network. To conquer this type of data, deep learning is not a good method. In order to solve the problem, we should develop a method to extract feature on

Date: 2017.4.

manifold and complex geometries. Once the method is developed, building deep structure on manifold and graphs becomes possible, that is to say, deep learning may extend its power beyond Euclidean domain.

In image processing and analysis, redundant systems such as wavelet frames have been implemented with excellent results in both classical and more challenging problems. Their applications in classical image restoration problems include image inpainting, super-resolution, deblurring, demosaicing and enhancement. Wavelet frames are also applied to more challenging image restoration problems such as blind deblurring, blind inpainting, and denoising with unknown noise type.Dong et al.(2010) shows that the wavelet frame is similar to differential operators which is also know as a efficient method to extract a picture's feature. In order to extend the method's power to manifold and graphs, in this paper, we constructed a wavelet structure on manifold. Based on the feature extracted by wavelet frame and the sparsity of the wavelet coefficient, we construct a semi-supervised learning algorithm and show high accuracy and strong power.

1.1. **Related Works.**

1.1.1. *Manifold learning.*

The basic assumption of manifold learning methods is that all data in a dataset hold some property is lying on a low dimension manifold in high dimension euclidean space. Kernel PCA, ISOMAP, LLE and etc. are famous method to reduce the 'fat' of data, embedding the manifold to a low dimension euclidean space and take the geometric structure of dataset into consider.

Bertozzi(2010) et al and Osher(2015) et al developed different variational methods via optimizing Ginzburg-Landau functional and total variational functional. The method proposed in this paper gives another variational methods with another regularization term given by wavelet transform. Cai(2010) et al claims that the wavelet regularization can be consider as another way to discretize the GTV regularization. So our method can be consider as another regularization way of total variational functional.

1.1.2. *Wavelets Frame.*

Wavelet frame is based on the theory of tight frame. A countable set $X \subset L_2(R^d)$, with $d \in Z^+$ is called a tight frame of $L_2(R^d)$ if

$$f = \sum_{g \in X} \langle f, g \rangle \, g, \forall f \in L_2(R^d)$$

and wavelet tight frame is a tight frame constructed by a quasi-affine system $X^J(\Psi)$ generated by Ψ is defined by the collection of the dilations and the shifts of Ψ as

$$X^J(\Psi) = \{\psi_{l,n,k} : 1 \le l \le q; n \in Z, k \in Z^d\}$$

where $\psi_{l,n,k}$ is defined by

$$\psi_{l,n,k} = 2^{\frac{nd}{2}} \psi_l(2^n \cdot -k)$$

Moreover we need ψ_l to constructed by a refinable function $\phi = 2^d \sum a_0[k]\phi(2 \cdot -k)$ and $\psi_l = 2^d \sum a_l[k]\phi(2 \cdot -k)$. The unitary extension principle (UEP) provides a general theory of theconstruction of MRA-based tight wavelet frames. Roughly speaking the Fourier series of masks satisy

$$\sum_{l=0}^{q} |\hat{a}_l(\xi)|^2 = 1, \sum_{l=0}^{q} \hat{a}_l(\xi)\overline{\hat{a}_l(\xi + \mu)} = 0$$

1.2. **Contribution.** We constructed a tight wavelet frame on manifold and graphs once the UEP holds, based on the wavelet frame semi-supervised learning can be conquered. Different from Blum A.(1998) 's co-training method, our method doesn't need to add data into the training test. At the same time different from method like Kernel PCA, ISOMAP and LLE, our method doesn't need to calculate the SVD of the Laplace operator and doesn't embedding the data into another space. To utilize the sparsity and local-relating property of Laplace operator, our method is quite efficient.

2. Preliminar

2.1. **Datas On Manifold.**

2.1.1. *Maniflod.*
Manifold is a locally low dimension structure in the high dimension space. To be simple, we only consider the manifold in the euclidean space, not only based on Nash Embedding theorem which shows that every Riemann manifold can be embedded into a high dimension euclidean space but also because real world data is always embedded in a Euclidean space.

Manifolds are generalizations of Euclidean space, it can be seen as locally euclidean topology space. That is to say for every point in the manifold there exists a neighborhood which is homeomorphic to an open set in a Euclidean space.

Tangent Space. Tangent space can be considered as a affine subspace tangent to the manifold. Tangent vector is a vector in the tangent space, which can represent a tangent line to the manifold.

Here we define tangent vector as the curves on the manifold, where the tangent space of the curve also is the tangent vector of manifold. Following this idea, we define the tangent space and cotangent space as following:

First we define a bilinear function between the function on the manifold and the curves on the manifold. For $f : \Gamma \to R^n$ and $\gamma : I = [0, 1] \to \Gamma$ we define $\langle f, \gamma \rangle = \frac{d}{dt}(f \circ \gamma)|_{t=0}$. (The discuss below is about the tangent space at $\gamma(0)$)

Based on the bilinear function we can define two equivalence relation, first $f_1, f_2 : \Gamma \to R^n$ if we have $\langle f_1, \gamma \rangle = \langle f_2, \gamma \rangle \, \forall \gamma : I = [0, 1] \to \Gamma$ then we say $f_1 \sim f_2$. At the same time, we say $\gamma_1 \sim \gamma_2$ if $\langle f, \gamma_1 \rangle = \langle f, \gamma_2 \rangle \, \forall f : \Gamma \to R^n$.

Now we define the tangent space as the quotient space of all curves on the manifold satisfies the equivalence condition between the curves defined above. Cotangent space is defined as the dual space of the tangent space, which can be also defined as the quotient space of all functions on the manifold satisfies the equivalence condition between the functions defined above. We mark the tangent space as $T_p\Gamma$(here p denotes $\gamma(0)$), and mark the cotangent space as $T_p\Gamma^*$

Riemann Metric. Riemann metric is a metric definite on the tangent space$T_p\Gamma$, that is to say Riemann metric can be defined as a positive defined bilinear function:$g : T_p\Gamma \times T_p\Gamma \to R$

Calculus on manifold. We can define the calculus on the the Riemann manifold. Here we particularly consider two type of functionals: a scalar field $f : \Gamma \to R$ is a smooth real function on the manifold and a tangent vector field $F : \Gamma \to T\Gamma$ is a mapping a tangent vector to each point on manifold. We can show that the tow types of functional can build a Hilbert space respectively.

$$\langle f, g \rangle_{L^2(\Gamma)} = \int_\Gamma f(x)g(x)dx$$
$$\langle F, G \rangle_{L^2(T\Gamma)} = \int_\Gamma \langle F(x), G(x) \rangle_{T_x(\Gamma)} \, dx$$

Here dx can be consider as the d-dimensional volume induced by Riemann metric.

Next we define the differential of f as an operator $df : T\Gamma \to R$ acting on tangent vector fields. At each point x, the differential can be defined as a linear form $df(x) = \langle \nabla f(x), \cdot \rangle_{T_x\Gamma}$. At the same time the divergence is the adjoint operator of the differential operator that is to say $\langle F, \nabla f \rangle_{L^2(T\Gamma)} = \langle -divF, f \rangle_{L^2(\Gamma)}$. Finally, the Laplacian(or Laplace-Beltrami operator in differential geometric jargon) $\Delta : L^2(\Gamma) \to L^2(\Gamma)$ is an operator $\Delta f = -div(\nabla f)$

2.1.2. *Point cloud, weighted graphs and spectral graph theory.*

In order to calculate the functions on the manifold by our computers, we should discretize our manifold to point cloud. As an exampla, In computer graphics and vision applications, two-dimensional manifolds are commonly used to model 3D shapes. The picture bellow is the point cloud discrete from the Stanford Bunny Rabbit.

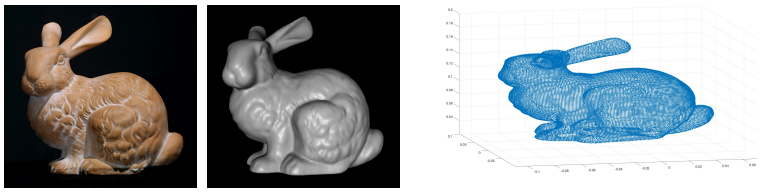


Fig. Stanford Bunny and the sampled point cloud data.

There are several common ways of discretizing such manifolds. First, the manifold is assumed to be sampled at n points. Their embedding corrdinate $x_1, \cdot, x_n$ are referred to as point cloud. Second we construct a graph to represent the geometric of the metric. The edges of the graph represent the local connectivity of the manifold, telling whether two points belong to a neighborhood or not, e.g. with Gaussian edge weights

$$\omega_{i,j} = e^{-||x_i - x_j||^2/2\sigma^2}$$

This simplest discretization of the manifold, however, does not capture correctly the geometry of the underlying continuous manifold (for example, the graph Laplacian would typically not converge to the continuous Laplacian operator of the manifold with the increase of the sampling density. There are also other ways to discretize the manifold by triangular mesh and the weight calculated by the formula

$$\omega_{i,j} = \frac{-l_{i,j}^2 + l - j, k^2 + l_{j,k}^2}{8a_{i,j,k}} + \frac{-l_{i,j}^2 + l - j, h^2 + l_{j,h}^2}{8a_{i,j,h}}$$

$$a_i = \frac{1}{3} \sum_{jk:(i,j,k)\in F} a_{i,j,k}$$

here $l_{i,j} = ||x_i - x_j||_2$ and the weight denotes the cotangent weights $\omega_{i,j} = \frac{1}{2}(\cot \alpha_{i,j} + \cot \beta_{i,j})$ which is commonly used in computer graphics.

In this paper we use the gauss kernel to discretize the manifold to the weighted graph.

2.2. **Classical Wavelet Transform.**

First we give a fast review on classical wavelet transform, in order to construct a sparse approximation by the dilation and translation of a base function which is called "mother wavlet" ψ.The method of constructing the wavelet transform in this section proceeds by producing the wavelets directly in the signal domain, through scaling and translation. The dilation and translation of a function of the mother wavelet is defined below:

$$\psi_{s,a}(x) = \frac{1}{s}\psi(\frac{x-a}{s})$$

Then the analysis operator to calculate the wavelet coefficient at scale s and location a is give by the inner product of f and $\psi_{s,a}$, written by formula:$T^{wav}_{s,a} = \langle f, \psi_{s,a}\rangle = \int_{-\infty}^{\infty} \frac{1}{s}\psi^*(\frac{x-a}{s})f(x)dx$ In order to reconstruct the signal f, the mother wavelet should satisfy the admissibility condition:

$$C_\psi = \int_0^\infty \frac{\left|\hat{\psi}(\omega)\right|^2}{\omega} \leq \infty$$

This condition implies for differentiable $\hat{\psi}$, $\hat{\psi}(0) = \int \psi(x)dx = 0$ which means the mother wavelet should be zero mean.

If the mother wavelet satisfies the admissibility condition, the inverse of the wavelet transform is shown in the formula below, which shows that the primer signal can be reconstructed by the wavelet coefficients:

$$f(x) = \frac{1}{C_\psi}\int_0^\infty \int_{-\infty}^\infty T^{wav}_{s,a} f\psi_{s,a}(x)\frac{dads}{s}$$

2.3. **Supervised and Semi-superivsed Learning:An Interpolation Perspective.**

Supervised and Semi-supervised learning can be consider as an interpolation problem on Euclidean and none-Euclidean domain. Data are only partially sampled due to logistic, economic, or computational constraints: limited number of sensors in seismic data. Moreover, sometimes one may also intentionally sample partial information of the scientific data as a straightforward data compression technique. Learning can be seen as an interpolation algorithm.

3. Tight Wavelet Frames On Riemann Manifold And Point Cloud

3.1. **Tight Wavelet Frames On Riemann Manifold.**

In order to define the tight wavelet on Riemann manifold we should first to define the Fourier analysis or Fourier transform on the Riemann manifold. As an extension of function $e^{i\omega x}$ is the eigenfucntions of one dimension Laplacian operator $\frac{d^2}{dx^2}$, the Fourier base on Riemann manifold can be defined as the eigenvalue function os Laplace-Beltrami Operator Δ on Riemman manifold $\{\Gamma, g\}$ with smooth boundary S. Let $\{\lambda_p : p = 0, 1, \cdots\}$ and $\{u_p : p = 0, 1, \cdots\}$ be the eigenvalues and eigenfunctions of the following eigenvalue problem:$\Delta u + \lambda u = 0$ with Dirichlet boundary condition $u|_S = 0$

For the Laplace-Beltrami Operator is a self-conjugate compact operator, u_p can form an orthonormal basis of $L_2(\Gamma)$, i.e.

$$\langle u_p, u_{p'} \rangle = \int_\Gamma u_p(x) u_{p'}^*(x) dx = \delta_{p,p'}$$

Then we can define the Fourier transform on the Riemann manifold as follow:

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

Like the Fourier transform in the Euclidean space, Fourier transform on manifold is also unitary.

Dilation and Translation. In order to define wavelet frame on manifold we should define dilation and translation of a function on manifold. Now we define a quasi-affine system generated by finitely many functions defined on R. Given $f \in L_2(R)$, now we define its delation and translation on manifold on $\{\Gamma, g\}$ as follow:

$$f_{n,y}^\Gamma(x) = \sum_{p=0}^{\infty} \hat{f}(2^{-n}\lambda_p) u_p^*(y) u_p(x) (\forall n \in Z, x, y \in \Gamma)$$

Rewrite it in the Fourier domain:$\hat{f_{n,y}^\Gamma}[p] = \sum_{p=0}^{\infty} \hat{f}(2^{-n}\lambda_p) u_p^*(y)$

Tight Wavelet Frame On Manifold. Now we define a quasi-affine system $X(\Psi)$ generate by a family of functions: $\Psi = \{\psi_j : 1 \le j \le r\} \subset L_2(R)$ as

$$X(\Psi) = \{\psi_{j,n,y}^\Gamma : 1 \le j \le r, n \in Z, x, y \in \Gamma\}$$

where $\psi_{j,n,y}^\Gamma$ is the dilation and translation of functionψ_j,i.e.

$$\psi_{j,n,y}^\Gamma(x) = \sum_{p=0}^{\infty} \hat{\psi}_j(2^{-n}\lambda_p) u_p^*(y) u_p(x)$$

We also request the system Ψ satisfies the following condition:

$$\hat{\psi}_j(2\xi) = \hat{a}_j(\xi)\hat{\phi}(\xi) (0 \le j \le r)$$

here $a_j \in L_2(Z)$ is the mask related with the system Ψ.

In order to reconstruct the function by its wavelet transform, we should consider the following operator

$$P_{n,j} f = \int_\Gamma \left\langle f, \psi_{j,n,y}^\Gamma \right\rangle \psi_{j,n,y}^\Gamma dy$$

Here we consider $\psi^{\Gamma}_{j,n,\cdot}$ as function from $\Gamma \times \Gamma$ to R, thus $\left\langle f, \psi^{\Gamma}_{j,n,\cdot} \right\rangle$ is a fucntion from Γ to R whose variable is y, i.e. $\left\langle f, \psi^{\Gamma}_{j,n,\cdot} \right\rangle (y) = \left\langle f, \psi^{\Gamma}_{j,n,y} \right\rangle$. Clearly operator $P_{n,j}$ can be written as $P_{n,j}f(x) = \int_{\Gamma} \left\langle f, \psi^{\Gamma}_{j,n,y} \right\rangle \psi^{\Gamma}_{j,n,y}(x)dy$

Now we want to find the relation between f and $P_{n,j}f$:

$$\begin{aligned}(P_{n,j}f)(x) &= \left\langle \left\langle f, \psi^{\Gamma}_{j,n,\cdot} \right\rangle, (\psi^{\Gamma}_{j,n,\cdot})^* \right\rangle = \left\langle \widehat{\left\langle f, \psi^{\Gamma}_{j,n,\cdot} \right\rangle}, \widehat{(\psi^{\Gamma}_{j,n,\cdot})^*} \right\rangle \\ &= \sum_{p=0}^{\infty} \hat{f}[p] \hat{\psi_j}^*(2^{-n}\lambda_p)\psi_j(2^{-n}\lambda_p)u_p(x)\end{aligned}$$

The last equation satisfies because $\widehat{\left\langle f, \psi^{\Gamma}_{j,n,\cdot} \right\rangle}[p] = \hat{f}[p]\hat{\psi_j}^*(2^{-n}\lambda_p)$ which is the rewrite of $\left\langle f, \psi^{\Gamma}_{j,n,\cdot} \right\rangle = \left\langle \hat{f}, \widehat{\psi^{\Gamma}_{j,n,\cdot}} \right\rangle = \sum_{p=0}^{\infty} \hat{f}[p]\widehat{\psi^*}(2^{-n}\lambda_p)u_p(y)$ in the Fourier domain.

Rewrite the formula in the Fourier Domain as

$$\widehat{P_{n,j}f}[p] = \hat{f}[p] \left| \hat{\psi}_j(2^{-n}\lambda_p) \right|^2$$

Then the tight frame condition for wavelet on graphs can be easily written as

$$\sum_{j=0}^{r} |\hat{a}_j(\xi)|^2 = 1$$

If the condition above satisfies, it is easily to show that the system $X(\Psi)$ is a tight frame for $L_2(\Gamma)$,*i.e.*

$$f = \sum_{j=1}^{r} \sum_{n \in Z} \int_{\Gamma} \left\langle f, \psi^{\Gamma}_{j,n,y} \right\rangle \psi^{\Gamma}_{j,n,y} dy$$

In the simulation below, we choose the linear tight wavelet frame system, *i.e.*

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

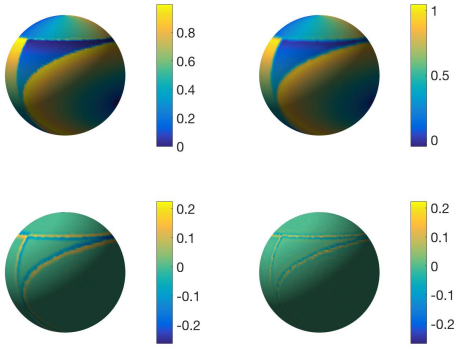


Fig. Wavelet Coefficient On Manifold(Slope), Reconstruction error = 3.6912e-06

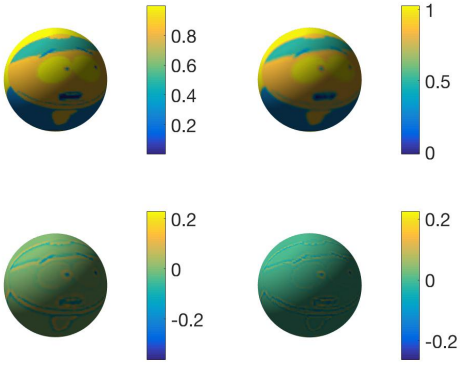


Fig. Wavelet Coefficient On Manifold(EricCartman), Reconstruction error = 3.8833e-06

4. Tight Wavelet Frames On Graphs

In this section we discuss a graph (G, V, E) with weights on the vertical and edges. Here we denote a_j as the weight of vertical v_j and $w_{i,j}$ as the weight of the edge between v_i, v_j. Here we give a assumption that $w_{ij} = -w_{ij}$ which also means if $(i, j) \in E$ then $(j, i) \in E$. For a give graph G we can define two function space with inner product defined as following:

$$\langle f, g \rangle_{L_2(V)} = \sum_{i \in V} a_i f_i g_i$$
$$\langle F, G \rangle_{L_2(E)} = \sum_{(i,j) \in E} \omega_{ij} F_{i,j} G_{i,j}$$

4.1. Laplace Operator And Fourier Analysis On Graphs.

First we define the gradient operator of function $f : V \to R$ as the $\nabla f_{i,j} = f_i - f_j$,

automatically satisfying $(\nabla f)_{i,j} = -(\nabla f)_{j,i}$. The graph divergence is defined as the adjoint operator of the gradient operator, that is to say $div : L^2(E) \to L^2(V)$ doing the converse

$$(divF)_i = \frac{1}{a_i} \sum_{j:(i,j)\in E} \omega_{i,j} F_{i,j}$$

It is to easy to verify that

$$\langle F, \nabla f \rangle_{L^2(E)} = \langle \nabla^* F, f \rangle_{L^2(V)} = \langle -divF, f \rangle_{L^2(V)}$$

The graph Laplacian is an operator $\Delta : L^2(V) \to L^2(V)$ defined as $\nabla = -div\nabla$. That is to say

$$(\Delta f)_i = \frac{1}{a_i} \sum_{(i,j)\in E} \omega_{i,j}(f_i - f_j)$$

Note that formula captures the intuitive geometric interpretation of the Laplacian as the difference between the local average of a function around a point and the value of the function at the point itself.If we denote $W = (\omega_{i,j}), D = diag(\sum_{j\neq i} \omega_{i,j})$ and $A = diag(a_1, \cdots, a_n)$, then $\Delta f = A^{-1}(D - W)f$. Different choice of the matrix A may be choose by $A = I$ or $A = D$.

The Laplacian operator is a self-adjoint positive-semidefinite operator, admitting on a compact domain an eigendecomposition with a discrete set of orthonormal eigenfunctions and non-negative real eigenvalues.

$$\Delta\phi_i = \lambda_i \phi_i, i \in N$$

It is well known that the eigenfunctions are the smoothest functions in the sense of the Dirichlet energy and can be interpreted as a generalization of the standard Fourier basis to the graph(non-Euclidean Domain).It is important to emphasize that the Laplacian eigenbasis is intrinsic due to the intrinsic construction of the Laplacian itself. A smooth square-integrable function f on the domain can be decomposed into Fourier series as

$$f(x) = \sum_{i\geq 0} \langle f, \phi_i \rangle_{L^2(M)} \phi_i(x)$$

where the projection on the basis functions producing a discrete set of Fourier coefficients generalizes the analysis (forward transform) stage in classical signal processing, and summing up the basis functions with these coefficients is the synthesis (inverse transform) stage.

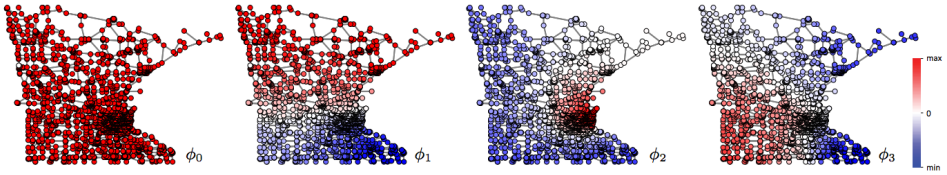


Fig. The Fourier Basis On Graph

A centerpiece of classical Euclidean signal processing is the property of the Fourier transform diagonalizing the convolution operator, colloquially referred to as the Convolution Theorem, allowing to express the convolution $f \circledast_G g$ of two functions in the spectral domain as the element-wise product of their Fourier transforms

$$\widehat{f \circledast g} = \widehat{f} * \widehat{g}$$

4.2. **Wavelet On Graphs.**

Consider function $\psi : V \to R$, like the translation and delation on the manifold, we can define the translation and delation on the graph as:

$$\psi_{t,n}(m) = T_g^t(\delta_n)(m) = \sum_{l=0}^{N=1} g(t\lambda_l)\chi_l^*(n)\chi_l(m)$$

here χ_i and λ_l are the eigenvalue eigenfunction of the Laplace operator.

Based on the definition of the translation and delation, it is easy to define the discre tight wavelet frame decomposition $Wf_G := \{W_{j,l}f_G : (j,l) \in B\}$, here $B = \{(0,L)\} \cup \{1,2,\cdots,r\} \times \{1,2,\cdots,r\}$

$$\widehat{W_{j,1}f_G}[k] := \hat{a}_j^*(2^{-N}\lambda_k)\widehat{f_G}[k]$$
$$\widehat{W_{j,l}f_G}[k] := \hat{a}_j^*(2^{-N+l-1}\lambda_k)a_0^*(2^{-N+l-3}\lambda_k)\cdots a_0^*(2^{-N}\lambda_k)\widehat{f_G}[k]$$

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

At the same time we should define the reconstruction transform $W^T\alpha$ which is defined by the following iterative procedure int the frequency domain

$$\hat{\alpha}_{0,l-1}[k] = \sum_{j=0}^{r} \hat{a}_j(2^{-N+l-1}\lambda_k)\hat{\alpha}_{j,l}[k], l = L, L-1, \cdots, 1$$

where $\alpha_{0,0} := W^T\alpha$ is the reconstructed graph data from α. Here W^T satisfies $\langle Wf_G, \alpha\rangle = \langle f_G, W^T\alpha\rangle$ and $W^TWf_G = f_G$

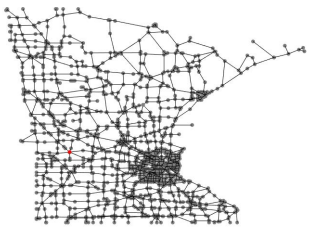
4.3. **Fast Tight Wavelet Frame Transform on Graphs.**
The Fourier analysis on graphs require the full set of eigenvectors and eigenvalues of the graph laplacian, which is computationally expensive to obtain for large graphs. To avoid to calculate the eigenvectors, a trick is to use polynomial approximation of the masks, so that the eigenvalue decomposition of the graph Laplacian is not needed. In this paper we use the Chebyshev polynomials to give a uniformly approximation. Recall that the Chebyshev polynomials on $[0, \pi]$ are defined as

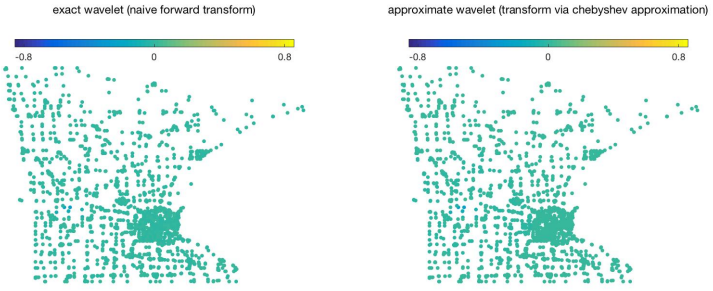
$$T_0 = 1, T_1(\xi) = \frac{\xi - \pi/2}{\pi/2}, T_k(\xi) = \frac{4}{\pi}(\xi - \pi/2)T_{k-1} - T_{k-1}(\xi)$$

Then we can use $T_j^n(\xi) = \frac{1}{2}c_{j,0} + \sum_{k=1}^{n-1} c_{j,k}T_k$ as the approximation of $\hat{a}_j$, here $c_{j,k} = \frac{2}{\pi}\int_0^\pi \cos(k\theta)\hat{a}_j(\frac{\pi}{2}(\cos(\theta)+1)d\theta)$. Then the fast waveelt frame transformation on graphs can be denoted as

$$\widehat{W_{j,1}f_G}[k] := \hat{T}_j^{n*}(2^{-N}\lambda_k)\widehat{f_G}[k]$$
$$\widehat{W_{j,l}f_G}[k] := \hat{T}_j^{n*}(2^{-N+l-1}\lambda_k)T_0^{n*}(2^{-N+l-3}\lambda_k)\cdots T_0^{n*}(2^{-N}\lambda_k)\widehat{f_G}[k]$$

At the same time, the numerical simulations is shown below.





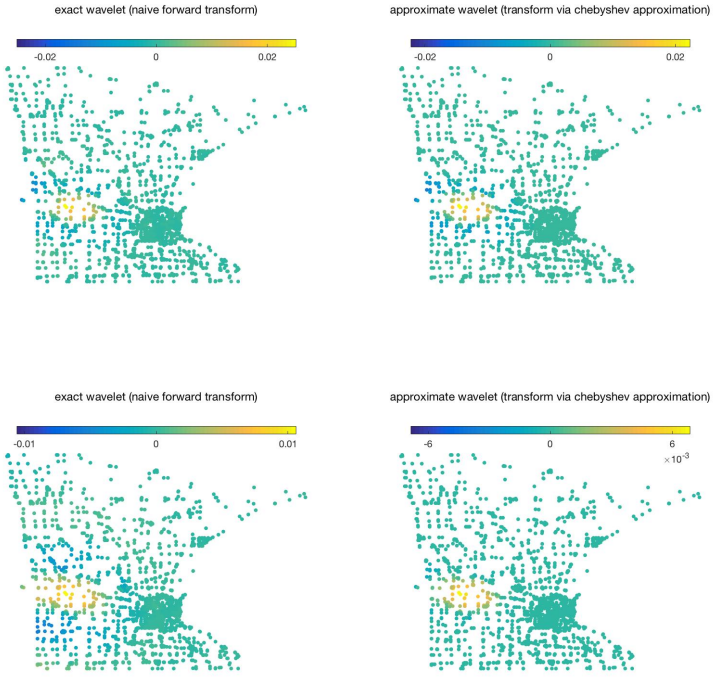


Fig. Translation and Dilation On Graph At Scale 0,2,5.

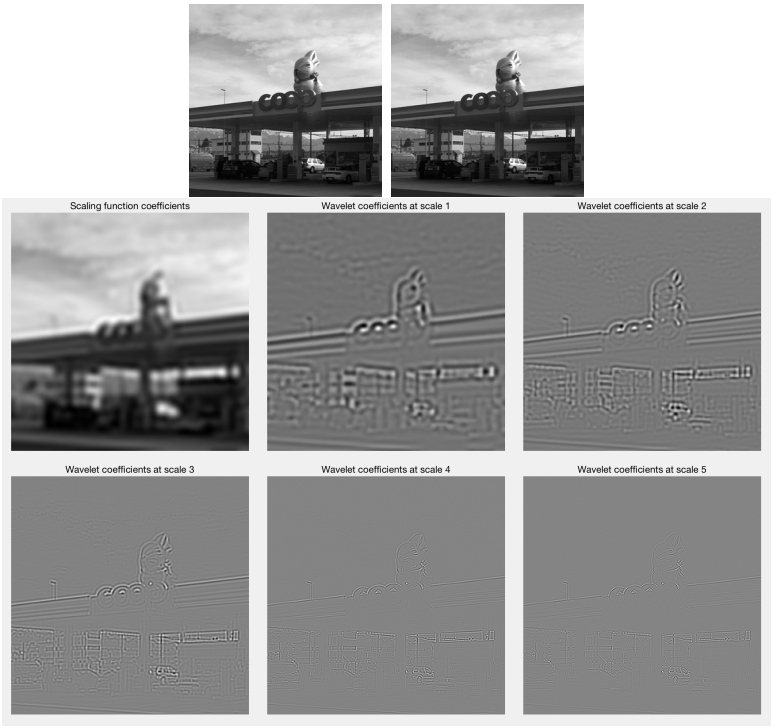


Fig. (a)Original Image (b)Reconstructed Image (c)Wavelet Cofficient Considering the picture as an manifold, Error=3.1123e-6

5. Numerical Test

5.1. **Semi-supervised Learning:Geometric Approach.**

In this section, we tested the a model to some real data sets. In particular, we shall observe how the learning algorithms perform on MNIST dataset of handwritten digits. MNIST dataset contains 70, 000 28×28 gray scale digit images. We view digits 0 - 9 as ten classes. Each image can be seen as a point in 784-dimensional Euclidean space. Throughout this section, we use Haar tight wavelet frame system, *i.e.*

$$\hat{a}_0(\xi) = \cos(\xi/2), \hat{a}_1(\xi) = \sin(\xi/2)$$

In our test, we label 100, 70 and 50 images respectively. The labeled images are selected at random in 70,000 images. For each case, we do 5 independent tests and the results are shown below.

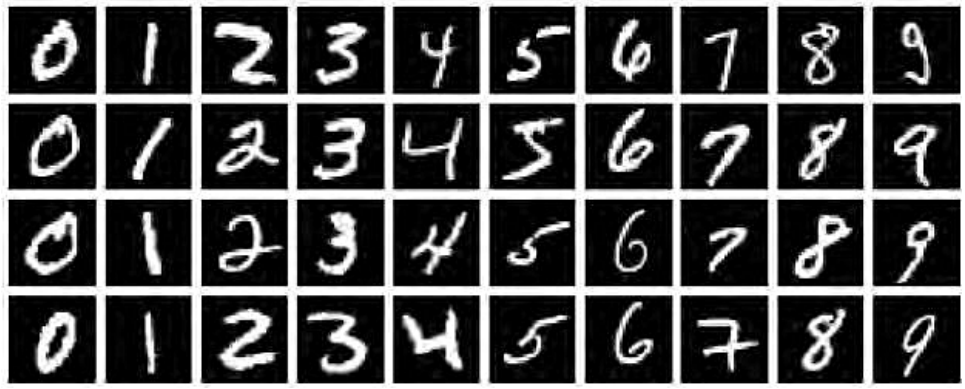


Fig. The MNIST dataset

Based on the sparse approximation given by the wavelet transform, our learning model can be seen as minimizing the following functional.

$$\min_{u\in[0,1]} ||\lambda \cdot Wu||_{1,G} + \frac{1}{2} ||u|_\Gamma - f|_\Gamma||^2_{2,G}$$

The first term impose a regularization of the level sets of u as the total variation regularization, while the second term impose the data term. To optimize the functional we apply the derivation of the ADMM algorithm as

$$
\begin{aligned}
u^{k+1} &= arg \min_{u\in[0,1]} \frac{1}{2} ||u|_\Gamma - f|_\Gamma||^2_{2,G} + \frac{\mu}{2} \left|\left|Wu - d^k + b^k\right|\right|^2_2 \\
d^{k+1} &= S_{\lambda G/\mu}(Wu^{k+1} + b^k) \\
b^{k+1} &= b^k + Wu^{k+1} - d^{k+1}
\end{aligned}
$$

here we need to solve a subproblem $u^{k+1} = arg\min_{u\in[0,1]} \frac{1}{2} ||u|_\Gamma - f|_\Gamma||^2_{2,G} + \frac{\mu}{2} \left|\left|Wu - d^k + b^k\right|\right|^2_2$ for u^{k+1} and the approximation is done by the following algorithm.

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

In our test 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 and use them to create the known label set and graph data. Our method approach 98.2% accuracy with only 3.67% known label.

5.2. **Two Moons Data.** The two moon data set is constructed from two half unit circles in R^2 with centers at $(0, 0)$ and $(1, 0.5)$. For each circle, 1000points are uniformly chosen and lifted to 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 and the learning algorithm using the exact same way. In our test we also use the Haar tight wavelet frame system given before. The error rate with different number of known labels is given below.

Known Labels	15	10	5	3
Error Rate	4.177	4.556	5.947	6.340

Table. Two Moon Dataset Semi-supervised Learning Test Via Wavelet

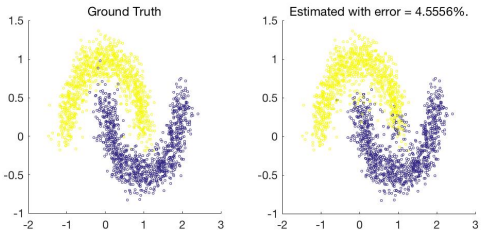


Table. Two Moon Dataset Semi-supervised Learning Test

5.3. **BankNote Dataset.**

BankNote Dataset were extracted from images that were taken from genuine and forged banknote-like specimens. For digitization, an industrial camera usually used for print inspection was used. The final images have 400× 400 pixels. Due to the object lens and distance to the investigated object gray-scale pictures with a resolution of about 660 dpi were gained. We also apply wavelet transform tool to extract features from images. The algorithm we use here is the same to the learning proposed before. On this dataset we can beat several state-of-art methods

Methods	Our Method	PAL	Binary MBO	GL
Error Rate	**1.64**	1.71	6.52	3.90

Table. Compare to state-of-art methods.

However the numerical test on MNIST is far from the stat-of-art methods. In order to explain the difference between the two datasets, we use PCA to project the dataset to a two dimension subspace that can preserve the geometry properties mostly. We can see that the BankNote Dataset can be divided into two parts by geometry properties while the MNIST Dataset cannot. I consider it as the reason of the better performance on BankNote Dataset.

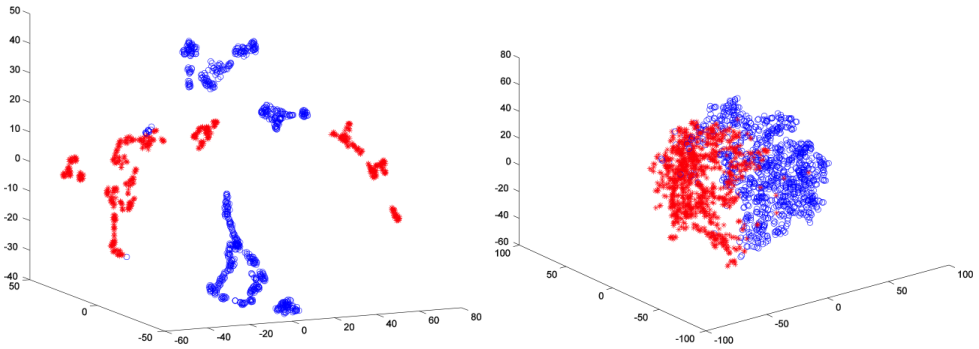


Fig. Different Geometry Properties Of Datasets (a) BankNote (b) MNIST

6. Further Work

Cai et al(2010) discovers the relationship between the variational method and wavelet frame based approach. Osher et al.(2015) constructed a variational method to do the semi-supervised learning tasks and image inpainting. However the wavelet on manifolds may not holds the same the same property. In Shi and Osher's Model, the regularization of the energy functional can be seen as the l_2 norm of ∇f, here orders the function f lies in the Soblev Space on manifold. As a conclusion, the Soblev Space is highly related to the Fouier coefficient of the function which can formula as: Suppose $\left|\hat{f}[p]\right| = O(\lambda_p^{-\alpha})$ with $\alpha > \frac{2s+m}{4}$ then $f \in H^S(M)$. Back to our model, we define the sample operator on manifold as $T^n_{\psi_j} := \psi_j(2^n\Delta)$, for $\psi^M_{j,n,y} = T^n_{\psi_j}\delta_y = \sum_{p\geq 0}\hat{\psi}_j(2^{-n}\lambda_p)u_p^*(y)u_p(\cdot)$, then

$$\begin{aligned} T^n_{\psi_j} f &= \sum_p \hat{f}[p]\hat{\psi}_j^*(2^{-n}\lambda_p)u_p^*(y)u_p(\cdot) \\ &= \sum \hat{f}[p]\hat{a}_j^*(2^{-n-1}\lambda_p)\hat{\phi}(2^{-n-1}\lambda_p)u_p(\cdot) \\ &= (a_j[2^{n+1}\Delta] \circledast_M T^{n+1}_{\phi})f \end{aligned}$$

and it converge to $D_j f = \sum_p \lambda_p^{s_j}\hat{f}[p]u_p(\cdot)$ where $D : H^s \to H^{s-2s_j}$ is an even order differential operator. Lack of first order differential operator's regularization, the expressive of model becomes week and lack of robustness at the same time. We should consider to modify our model to take the odd order differential operator into consider which may have huge improvement to the expressive. The asymptotic analysis is also need to be proved.

At the same time, data-driven tight frame construction is also a hot topic in applied and computational harmonic analysis. However the size of Laplace matrix becomes a barrier to data-driven tight frame on manifolds. In order to construct data-driven tight frame, more efficient numerical linear algebra method should be developed. Moreover, the data-driven method may help computer scientists to utilize deep convolutional neural networks on graphs and manifolds, these deep structures may become a powerful weapon in the future.

At last, co-training is an popular way in semi-supervised learning. How to combine the co-training frame work and wavelet on manifold may be an interesting method.

Appendix.
The code of wavelet on graphs and semi-supervised learning will soon be published on the author's Github.

Acknowledgement.
I would like thank Prof.Bin Dong for recently drawing my attention this project,

giving me a lot of help and instruction and providing GPU for our test. I also want to acknowledge Prof. Renjie Feng for his help on teaching me the knowledge of Eigen Geometry. At last I would like to thank Daozhe Lin for useful Discussion.

7. Reference

Bronstein M M, Bruna J, Lecun Y, et al. Geometric deep learning: going beyond Euclidean data[J]. 2016.

Dong B. Sparse representation on graphs by tight wavelet frames and applications [J]. Applied & Computational Harmonic Analysis, 2014.

Dong B, Hao N. Semi-supervised high-dimensional clustering by tight wavelet frames[C]// Spie, Wavelets& Sparsity Xvi. 2015.

Chern S S, Chen W H, Lam K S. Lectures on Differential Geometry[J]. Annales De L Institut Henri Poincare Physique Theorique, 2014, 40(40):329-342.

Hammond D K, Vandergheynst P, Gribonval R. Wavelets on graphs via spectral graph theory[J]. Applied & Computational Harmonic Analysis, 2011, 30(2):129-150.

Shaham U, Cloninger A, Coifman R R. Provable approximation properties for deep neural networks[J]. Applied & Computational Harmonic Analysis, 2015.

Dong, B. and Zhang, Y., An efficient algorithm for ?0 minimization in wavelet frame based image restoration, Journal of Scientific Computing 54 (2-3), 350-368 (2013).

Zhang, Y., Dong, B., and Lu, Z., l0 minimization of wavelet frame based image restoration, Mathematics of Computation 82, 995-1015 (2013).

Hou, L., Ji, H., and Shen, Z., Recovering over-/underexposed regions in photographs., SIAM J. Imaging Sciences 6(4), 2213-2235 (2013).

Cai, J., Osher, S., and Shen, Z., Split Bregman methods and frame based image restoration, Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal 8(2), 337?369 (2009).

Cai, J., Chan, R., and Shen, Z., A framelet-based image inpainting algorithm, Applied and Computational Harmonic Analysis 24(2), 131?149 (2008).

Zhang Q, Benveniste A. Wavelet networks[J]. IEEE Transactions on Neural Networks, 1992, 3(6):889-898.

Blum A. Combining labeled and unlabeled data with co-training[C]// Eleventh Conference on Computational Learning Theory. ACM, 2000:92-100.

Tao Sun, Robert Hannah, Wotao Yin, Asynchronous Coordinate Descent under More Realistic Assumptions, May 2017

Siwei Yu, Stanley Osher, Jianwei Ma, Zuoqiang Shi, Noise attenuation in a low dimensional manifold. October 2016

Stanley Osher, Zuoqiang Shi, and Wei Zhu, Low Dimensional Manifold Model for Image Processing, January 2016.

Bin Dong, Zuowei Shen and Peichu Xie, Image restoration: a general wavelet frame based model and its asymptotic analysis, SIAM Journal on Mathematical Analysis, 49(1), 421-445, 2017.

Jian-Feng Cai, Bin Dong and Zuowei Shen, Image restoration: a wavelet frame based model for piecewise smooth functions and beyond, Applied and Computational Harmonic Analysis, 41(1), 94-138, 2016.

Coifman R R, Lafon S. Diffusion maps[J]. Applied & Computational Harmonic Analysis, 2006, 21(1):5-30.

Lafon S, Lee A B. Diffusion maps and coarse-graining: a unified framework for dimensionality reduction, graph partitioning, and data set parameterization[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2006, 28(9):1393-1403.

Nadler B, Lafon S, phane, et al. Diffusion Maps, spectral clustering and eigenfunctions of Fokker-Planck operators[C]// International Conference on Neural Information Processing Systems. MIT Press, 2005:955-962.

Lafon S, Keller Y, Coifman R R. Data fusion and multicue data matching by diffusion maps[C]// IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006:2006.

Coifman R R, Maggioni M. Diffusion wavelets[J]. Applied & Computational Harmonic Analysis, 2006, 21(1):53-94.

Jian-Feng Cai, Bin Dong, Stanley Osher and Zuowei Shen, Image restoration: total variation; wavelet frames; and beyond, Journal of the American Mathematical Society, 25(4), 1033-1089, 2012.

Weickert J, Steidl G, Mrazek P, et al. Diffusion Filters and Wavelets: What Can They Learn from Each Other?[M]// Handbook of Mathematical Models in Computer Vision. Springer US, 2006:1-16.

Bertozzi, A. L. and Flenner, A., Diffuse interface models on graphs for classification of high dimensional data, Multiscale Modeling & Simulation 10(3), 1090?1118 (2012).

Merkurjev, E., Kostic, T., and Bertozzi, A. L., An mbo scheme on graphs for classification and image processing, SIAM Journal on Imaging Sciences 6(4), 1903?1930 (2013).

Merkurjev, E., Bae, E., Bertozzi, A., and Tai, X. C., Global binary optimization on graphs for classification of high dimensional data, UCLA CAM Report 14-72 (2014).

Wei E, Ozdaglar A. Distributed Alternating Direction Method of Multipliers[C]// Decision and Control. IEEE, 2012:5445-5450.

Boyd S. Alternating Direction Method of Multipliers[J]. 2011.

Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.