REPRESENTATIONS OF PIECEWISE SMOOTH SIGNALS ON GRAPHS

Siheng Chen[1], *Rohan Varma*[1], *Aarti Singh*[2], *Jelena Kovačević*[1,3]

[1]Dept. of Electrical and Computer Engineering, [2]Dept. of Machine Learning,
[3]Dept. of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA

ABSTRACT

We study representations of piecewise-smooth signals on graphs. We first define classes for smooth, piecewise-constant, and piecewise-smooth graph signals, followed by a series of multiresolution local sets to analyze those signals by implementing a multiresolution analysis on graphs. Based on these local sets, we propose local-set-based piecewise-constant and piecewise-smooth dictionaries as graph signal representations that, in spirit, resemble the classical Haar wavelet basis and are naturally localized in both graph vertex and graph Fourier domains. Moreover, they promote sparsity when representing piecewise-smooth graph signals. In the experiments, we show that local-set-based dictionaries outperform graph Fourier domain based representations when approximating both simulated and real-world graph signals.

1. INTRODUCTION

Signal representation is one of the most fundamental tasks in our discipline; it is related to approximation, compression, denoising, inpainting, detection, and localization [1]. Representations are particularly crucial for signals with complex, irregular underlying structure that are being generated at an unprecedented rate from various sources, including social, biological, and physical infrastructure [2], among others; we call such signals *graph signals*. While prior work on graph signal representations focused mainly on smooth graph signals [3], such as bandlimited [4, 5], approximately bandlimited [6], and signals with small variation [7], it did not address graph signal localization. As localization properties of graph signals are of interest in a number of applications (e.g. in community detection, the community labels are piecewise-constant on a social network), representations that consider both smoothness and localization are of interest.

We thus study *piecewise-smooth graph signals* and consider both smoothness and localization properties. We first define classes for smooth, piecewise-constant, and piecewise-smooth graph signals, followed by a series of multiresolution local sets to analyze those signals by implementing a multiresolution analysis on graphs. Based on these local sets, we propose local-set-based piecewise-constant and piecewise-smooth dictionaries as graph signal representations that, in spirit, resemble the classical Haar wavelet basis and are naturally localized in both graph vertex and graph Fourier domains. Moreover, they promote sparsity when representing piecewise-smooth graph signals. The main advantages of the proposed local-set-based dictionaries are that they are simple, general, easy to visualize, and effective. In the experiments, we show that local-set-based dictionaries outperform windowed graph Fourier when approximating both simulated and real-world graph signals.

The authors gratefully acknowledge support from the NSF through awards 1130616 and 1421919. Due to the lack of space, the proof of theorems and some details are omitted, and will be shown in a longer version.

2. GRAPH SIGNAL MODELS

Let $G = (\mathcal{V}, \mathcal{E})$ be a graph, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of edges that represent the underlying relations between pairs of nodes. Let $\mathrm{A} \in \mathbb{R}^{N \times N}$ be the adjacency matrix, with $\mathrm{A}_{j,k}$ the edge weight. Let D be the degree matrix, with $(\mathrm{D})_{i,i} = \sum_j \mathrm{A}_{i,j}$, $\mathrm{L} = \mathrm{D} - \mathrm{A}$ the graph Laplacian matrix, and let $\mathrm{P} = \mathrm{D}^{-1}\mathrm{A}$ be the transition matrix. We call $\mathbf{x} = \left[x_0, x_1, \ldots, x_{N-1}\right]^T \in \mathbb{R}^N$ a *graph signal*, with x_i the signal coefficient at the ith node.

Smooth Graph Signals. We start with three smoothness criteria for graph signals; while they have been implicitly mentioned previously, none has been rigorously defined.

Definition 1. A graph signal $\mathbf{x}$ is *pairwise Lipschitz smooth* with parameter L when it satisfies

$$|x_i - x_j| \ \leq \ L\, d(v_i, v_j), \text{ for all } i, j = 0, 1, \ldots, N-1,$$

with $d(v_i, v_j)$ the distance between the ith and the jth nodes.

We can choose the geodesic distance, the diffusion distance [8], or some other distance metric for $d(\cdot,\cdot)$. Similarly to the traditional Lipschitz criterion [9], the pairwise Lipschitz smoothness criterion emphasizes pairwise smoothness, which zooms onto the difference between each pair of adjacent nodes.

Definition 2. A graph signal $\mathbf{x}$ is *total Lipschitz smooth* with parameter L when it satisfies

$$\sum_{(i,j)\in\mathcal{E}} \mathrm{A}_{i,j}(x_i - x_j)^2 \ \leq \ L.$$

The total Lipschitz smoothness criterion generalizes the pairwise Lipschitz smoothness criterion while still emphasizing pairwise smoothness, but in a less restricted manner; it is also known as the Laplacian smoothness criterion [10].

Definition 3. A graph signal $\mathbf{x}$ is *neighboring smooth* with parameter L when it satisfies

$$\sum_i \left(x_i - \frac{1}{\sum_{j:(i,j)\in\mathcal{E}} \mathrm{A}_{i,j}} \sum_{j:(i,j)\in\mathcal{E}} \mathrm{A}_{i,j}\, x_j \right)^2 \leq L.$$

The neighboring smoothness criterion emphasizes neighboring smoothness, which compares each node to the average of its immediate neighbors.

The three criteria quantify smoothness in different ways: the pairwise and the total Lipschitz ones focus on the variation of two signal coefficients connected by an edge with the pairwise Lipschitz one more restricted, while the neighboring smoothness criterion focuses on comparing a node to the average of its neighbors. Following the three criteria quantifying smoothness, we now construct three signal classes satisfying each of the three criteria. We first construct polynomial graph signals to satisfy the Lipschitz smoothness criterion.

Definition 4. A graph signal $\mathbf{x}$ is polynomial with degree K when

$$\mathbf{x} = \mathrm{D}_K\, \mathbf{a} = \begin{bmatrix} \mathbf{1} & \mathrm{D}^{(1)} & \mathrm{D}^{(2)} & \ldots & \mathrm{D}^{(K)} \end{bmatrix} \mathbf{a} \in \mathbb{R}^N,$$

where $\mathbf{a} \in \mathbb{R}^{KN+1}$ and D_K is a polynomial dictionary with $\mathrm{D}^{(k)}_{i,j} = d^k(v_i, v_j)$. Denote this class by $\mathrm{PL}(K)$.

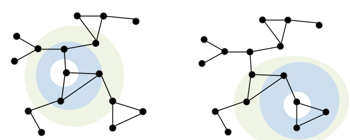


Fig. 1: Different origins lead to different coordinate systems; white, blue, and green denote the origin, nodes with geodesic distance 1 from the origin, and nodes with geodesic distance 2 from the origin, respectively.

In classical signal processing, polynomial time signals can be expressed as $x_n = \sum_{k=0}^{K} a_k n^k$, $n = 0, \ldots, N-1$; we can rewrite this as in the above definition as $\mathbf{x} = \mathrm{D}_K\, \mathbf{a}$, with $(D_K)_{n,k} = n^k$. The columns of D_K are denoted as $\mathrm{D}^{(k)}$, $k = 0, \ldots, K$, and called *atoms*; the elements of each atom $\mathrm{D}^{(k)}$ are n^k. Since polynomial time signals are shift-invariant, we can set any time point as the origin; such signals are thus characterized by $K+1$ degrees of freedom a_k, $k = 0, \ldots, K$. This is not true for graph signals, however; they are not shift-invariant and any node can serve as the origin (see Figure 1). In the above definition, $\mathrm{D}^{(k)}$ are now matrices with the number of atoms equal to the number of nodes N (with each atom corresponding to the node serving as the origin). The dictionary D_K thus contains $KN+1$ atoms. We can show that signals in $\mathrm{PL}(1)$ are pairwise Lipschitz smooth.

We now construct bandlimited signals to satisfy the total Lipschitz and neighboring smoothness criteria.

Definition 5. A graph signal $\mathbf{x}$ is bandlimited with respect to a graph Fourier basis V with bandwidth K when

$$\mathbf{x} = \mathrm{V}_{(K)}\, \mathbf{a},$$

where $\mathbf{a} \in \mathbb{R}^K$ and $\mathrm{V}_{(K)}$ is a submatrix containing the first K columns of V. Denote this class by $\mathrm{BL}_{\mathrm{V}}(K)$ [6].

When V is the eigenvector matrix of the graph Laplacian matrix, we denote it as V_{L} and can show that signals in $\mathrm{BL}_{\mathrm{V}_{\mathrm{L}}}(K)$ are total Lipschitz smooth; when V is the eigenvector matrix of the transition matrix, we denote it as V_{P} and can show that signals in $\mathrm{BL}_{\mathrm{V}_{\mathrm{P}}}(K)$ are neighboring smooth (we omit details due to limited space). Of these three classes, $\mathrm{PL}(K)$ is novel; $\mathrm{BL}_{\mathrm{V}}(K)$ has been introduced in [6].

Piecewise-constant Graph Signals. Piecewise-constant graph signals have been used in many applications related to graphs without having been explicitly defined; for example, in community detection, community labels form a piecewise-constant graph signal for a social network; in semi-supervised learning, classification labels form a piecewise-constant graph signal for a graph constructed from the dataset. While smooth graph signals emphasize slow transitions, piecewise-constant graph signals emphasize fast transitions (corresponding to boundaries) and localization (corresponding to signals being constant in a local neighborhood).

We construct piecewise-constant graph signals by using local sets, which have been used previously in graph cuts and graph signal reconstruction [11, 12].

Definition 6. Let $\{S_c\}_{c=1}^{C}$ be the partition of the node set $\mathcal{V}$. We call $\{S_c\}_{c=1}^{C}$ *local sets* when they satisfy that the subgraph corresponding to each local set is connected, that is, when G_{S_c} is connected for all c.

Definition 7. A graph signal $\mathbf{x}$ is piecewise-constant with C pieces when

$$\mathbf{x} = \sum_{c=1}^{C} a_c \mathbf{1}_{S_c},$$

where $\{S_c\}_{c=1}^{C}$ forms a series of local sets and $(\mathbf{1}_S)_i = 1$, when $v_i \in S$; 0, otherwise. Denote this class by $\mathrm{PC}(C)$.

Let Δ be the *graph difference operator* (the oriented incidence matrix of G), whose rows correspond to edges [13, 14]. For example, if e_i is a directed edge that connects the jth node to the kth node ($j < k$), the elements of the ith row of Δ are $\Delta_{i,\ell} = -1$ when $\ell = j$; $\Delta_{i,\ell} = 1$ when $\ell = k$; and $\Delta_{i,\ell} = 0$, otherwise. $\Delta\mathbf{x}$ then measures the difference between each pair of adjacent signal coefficients. When the value of the graph signal on each local set is different, $\|\Delta\mathbf{x}\|_0$ counts the total number of edges connecting nodes between local sets.

Piecewise-smooth Graph Signals. To be able to deal with as wide a class of real-world graphs signals as possible, we combine smooth and piecewise-constant graph signals into piecewise-smooth graph signals.

Definition 8. A graph signal $\mathbf{x}$ is piecewise-polynomial with C pieces and degree K when

$$\mathbf{x} = \sum_{c=1}^{C} \mathbf{x}^{(c)} \mathbf{1}_{S_c},$$

where $\mathbf{x}^{(c)}$ is a kth order polynomial signal on the subgraph G_{S_c} with $x_i^{(c)} = a_c + \sum_{j\in S_c} \sum_{k=1}^{K} a_{k,j,c} d^k(v_i, v_j)$. Denote this class by $\mathrm{PPL}(C, K)$.

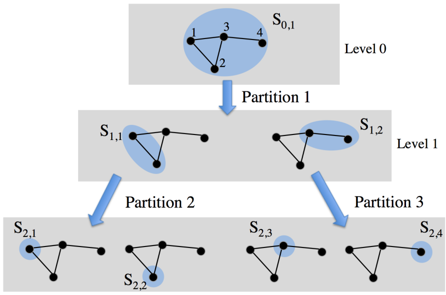


Fig. 2: Local set decomposition tree. In each partition, we decompose a node set into two disjoint connected sets.

$\mathrm{PPL}(1, K)$ is the polynomial class with degree K, $\mathrm{PL}(K)$ from Definition 4, and $\mathrm{PPL}(C, 0)$ is the piecewise-constant class with C pieces, $\mathrm{PC}(C)$ from Definition 7. The degrees of freedom for a local set S_c at the polynomial degree k is the number of origins, that is, $\left\| \begin{bmatrix} a_{k,1,c} & a_{k,2,c} & \ldots & a_{k,|S_c|,c} \end{bmatrix} \right\|_0$.

Definition 9. A graph signal $\mathbf{x}$ is piecewise-bandlimited with C pieces and bandwidth K when

$$\mathbf{x} = \sum_{c=1}^{C} \mathbf{x}^{(c)} \mathbf{1}_{S_c},$$

where $\mathbf{x}^{(c)}$ is a bandlimited signal on the subgraph G_{S_c} with $x_i^{(c)} = \sum_{k=0}^{K} a_{k,c} \mathrm{V}_{i,k}^{(c)}$, and $\mathrm{V}^{(c)}$ is a graph Fourier basis of G_{S_c}. Denote this class by $\mathrm{PBL}(C, K)$.

We use zero padding to ensure $\mathrm{V}^{(c)} \in \mathbb{R}^{N \times N}$ for each G_{S_c}. Still, $\mathrm{V}^{(c)}$ can be the eigenvector matrix of either the graph Laplacian matrix or the transition matrix.

3. MULTIRESOLUTION REPRESENTATIONS

We now discuss representations for piecewise-smooth graph signals based on multiresolution local sets.

Multiresolution Local Sets. Our aim is to construct a series of local sets in a multiresolution fashion. We initialize $S_{0,1} = \mathcal{V}$ to correspond to the 0th level subspace V_0, that is, $V_0 = \{c_0 \mathbf{1}_{S_{0,1}}, c_0 \in \mathbb{R}\}$. We then partition $S_{0,1}$ into two disjoint local sets $S_{1,1}$ and $S_{1,2}$, corresponding to the first level subspace V_1, where $V_1 = \{c_1 \mathbf{1}_{S_{1,1}} + c_2 \mathbf{1}_{S_{1,2}}, c_1, c_2 \in \mathbb{R}\}$. We then recursively partition each larger local set into two smaller local sets. For the ith level subspace, we have $V_i = \sum_{j=1}^{2^i} c_j \mathbf{1}_{S_{i,j}}$ and then, we partition $S_{i,j}$ into $S_{i+1,2j-1}, S_{i+1,2j}$ for all $j = 1, 2, \ldots, 2^i$. We call $S_{i,j}$ the parent set of $S_{i+1,2j-1}, S_{i+1,2j}$ and $S_{i+1,2j-1}, S_{i+1,2j}$ are the children sets of $S_{i,j}$. When $|S_{i,j}| \leq 1$, $S_{i+1,2j-1} = S_{i,j}$ and $S_{i+1,2j} = \emptyset$. At the finest resolution, each local set corresponds to an individual node or an empty set. In other words, we build a binary decomposition tree that partitions a graph structure into multiple local sets. The ith level of the decomposition tree corresponds to the ith level subspace. The depth of the decomposition T depends on how local sets are partitioned; T ranges from N to $\lceil \log N \rceil$, where N corresponds to partitioning one node at a time and $\lceil \log N \rceil$ corresponds to an even partition at each level. We show an example in Figure 2.

This graph partitioning is the key step in constructing the local sets. The proposed construction does not restrict us to any particular graph partitioning algorithm; depending on the application, the partitioning step can be implemented by many existing graph partition algorithms, the only requirement is to satisfy Definition 6. Some candidate algorithms are the graph cuts [11] and the balance cut in the spanning tree [15].

The proposed construction of local sets mimics the classical multiresolution analysis to some extent. The initial subspace V_0 is at the coarsest resolution. Through partitioning, local sets zoom into increasingly finer resolutions in the graph vertex domain. The subspace V_T at the finest resolution zooms into each individual node and covers the entire $\mathbb{R}^N$. Classical scale invariance requires that when $f(t) \in V_0$, then $f(2^m t) \in V_m$, which is ill-posed in the graph domain because graphs are finite and discrete; the classical translation invariance requires that when $f(t) \in V_0$, then $f(t-n) \in V_0$, which is again ill-posed, this time because graphs are irregular. The essence of scaling and translation invariance, however, is to use the same function and its scales and translates to span different subspaces, which is what the proposed construction promotes.

Dictionary Construction. We collect local sets by level in ascending order in a dictionary, with atoms corresponding to each local set, that is, $\mathrm{D}_{\mathrm{LSPC}} = \{\mathbf{1}_{S_{i,j}}\}_{i=0,j=1}^{i=T,j=2^i}$. We call it the *local-set-based piecewise-constant dictionary*. After removing empty sets, the dictionary has $2N-1$ atoms, that is, $\mathrm{D}_{\mathrm{LSPC}} \in \mathbb{R}^{N \times (2N-1)}$; each atom is a piecewise-constant graph signal with various sizes and localizing various parts of a graph. While for an arbitrary piecewise-constant signal we do not know the support of its underlying pieces, $\mathrm{D}_{\mathrm{LSPC}}$ still provides sparse representations.

Theorem 1. For all $\mathbf{x} \in \mathbb{R}^N$, we have $\|\mathbf{a}^*\|_0 \leq 2T \|\Delta \mathbf{x}\|_0$, where T is the maximum level of the decomposition and

$$\mathbf{a}^* = \arg\min_{\mathbf{a}} \|\mathbf{a}\|_0, \ \text{subject to}: \ \mathbf{x} = \mathrm{D}_{\mathrm{LSPC}} \, \mathbf{a}.$$

When $\mathbf{x}$ is piecewise-constant, $\|\Delta \mathbf{x}\|_0$ is small; thus, $\mathrm{D}_{\mathrm{LSPC}}$ is particularly good for representing piecewise-constant graph signals. The graph partitioning influences the quality of representation; the even partition of each local set optimizes the worst case scenario. For piecewise-constant graph signals, the sizes of the local sets matter, not the shape.

To represent piecewise-smooth graph signals, we use multiple atoms for each local set. We take the piecewise-polynomial signals as an example. For each local set,

$$\mathrm{D}_{S_{i,j}} = \begin{bmatrix} \mathbf{1} & \mathrm{D}_{S_{i,j}}^{(1)} & \mathrm{D}_{S_{i,j}}^{(2)} & \ldots & \mathrm{D}_{S_{i,j}}^{(K)} \end{bmatrix},$$

where $(\mathrm{D}_{S_{i,j}}^{(k)})_{m,n} = d^k(v_m, v_n)$, when $v_m, v_n \in S_{i,j}$; and 0, otherwise. The number of atoms in $\mathrm{D}_{S_{i,j}}^{(k)}$ is $1 + K|S_{i,j}|$. We collect the sub-dictionaries for all the multiresolution local sets to form the *local-set-based piecewise-smooth dictionary*, that is, $\mathrm{D}_{\mathrm{LSPS}} = \{\mathrm{D}_{S_{i,j}}\}_{i=0,j=1}^{i=T,j=2^i}$. The number of atoms in $\mathrm{D}_{\mathrm{LSPS}}$ is $O(KNT)$, where K is the maximum degree of polynomial, N is the size of the graph and T is the maximum level of the decomposition. When we use even partitioning, the total number of atoms is $O(KN \log N)$.

Similarly, to model piecewise-bandlimited signals, we replace $\mathrm{D}_{S_{i,j}}$ by the graph Fourier basis of each subgraph $G_{S_{i,j}}$. The total number of atoms of the corresponding $\mathrm{D}_{\mathrm{LSPS}}$ is then $O(NT)$. For piecewise-smooth signals, we cannot use the sparse coding to do exact approximation. To minimize the approximation error, both the sizes and the shapes of the local sets matter for piecewise-smooth graph signals.

Theorem 2. For all $\mathbf{x} \in \mathrm{PBL}(C, K)$, we have $\|\mathbf{a}^*\|_0 \leq 2KT \|\Delta \mathbf{x}_{\mathrm{PC}}\|_0$, where T is the maximum level of the decomposition, $\mathbf{x}_{\mathrm{PC}}$ is a piecewise-constant signal that corresponds the same local sets with $\mathbf{x}$ and

$$\mathbf{a}^* = \arg\min_{\mathbf{a}} \|\mathbf{a}\|_0, \ \text{subject to} \ \|\mathbf{x} - \mathrm{D}_{\mathrm{LSPS}}\, \mathbf{a}\|_2^2 \leq \epsilon \|\mathbf{x}\|_2^2,$$

where ϵ is a constant determined by graph partitioning.

Relation to Prior Work. Graph signal representation has been considered previously in, for example, multiscale wavelets on trees, which provide a hierarchy tree representation for a dataset [16]. It proposes a wavelet-like orthonormal basis, focuses on high-dimensional data and constructs a decomposition tree bottom up; in our work, we focus on graphs and construct a decomposition tree top down, which is useful for capturing clusters; spanning tree wavelet basis proposes a localized basis on a spanning tree, which uses a special graph partitioning algorithm [15]. Some other representations have been proposed without demonstrating their advantages in representing any specific class of graph signals [17, 18, 19, 20]; in our work, we target piecewise-smooth signals.

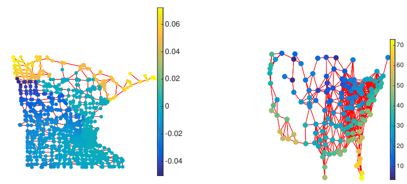
4. APPLICATIONS

Good representations for piecewise-smooth graph signals are potentially useful in many applications, such as visualization, denoising, active sampling [21, 5] and semi-supervised learning [22]. Here we consider approximation, whose goal is to use a few expansion coefficients to approximate a graph signal. We compare the windowed graph Fourier transform [20] with the local-set-based dictionaries. We use the balance cut of the spanning tree to obtain the local sets [15]. To do approximation, we solve the following sparse coding problem by using orthogonal matching pursuit [23],

$$\mathbf{x}' = \arg\min_{\mathbf{a}} \|\mathbf{x} - \mathrm{D}\, \mathbf{a}\|_2^2, \ \text{subject to} \ \|\mathbf{a}\|_0 \leq s, \qquad (1)$$

where D is a dictionary and $\mathbf{a}$ is a sparse code.

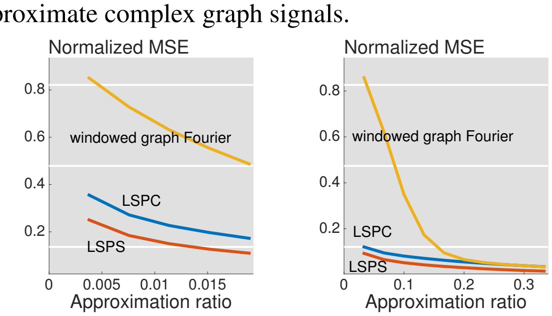
Experiments. We test the representations on two datasets, the Minnesota road graph [24] and the U.S city graph [7]. The Minnesota road graph is a standard dataset including 2642 nodes and 3304 undirected edges [24]. We simulate 100 piecewise-constant graph signals as follows: we random choose three nodes as cluster centers and assign all other nodes to their nearest cluster centers based on the geodesic distance. We assign a random integer to each cluster. We further obtain 100 piecewise-polynomial graph signals by element-wise multiplying a polynomial function, $-d^2(v_0, v) + 12d(v_0, v)$, where v_0 is a reference node that assigns randomly. As an example, see Figure 3(a). The U.S weather station graph is a network representation of 150 weather stations across the U.S. We assign an edge when two weather stations are within 500 miles. The graph includes 150 nodes and 1033 undirected, unweighted edges. Each weather station has 365 days of recordings (one recording per day), for a total of 365 graph signals. As an example, see Figure 3(b).



(a) Minnesota. (b) U.S weather station.

Fig. 3: Graph signal.

The approximation error is measured by the normalized mean square error, $\text{Normalized MSE} = \|\mathbf{x}' - \mathbf{x}\|_2^2 \,/\, \|\mathbf{x}\|_2^2$, where $\mathbf{x}'$ is the approximation signal and $\mathbf{x}$ is the original signal. Figure 4 shows the averaged approximation errors. LSPC denotes local-set-based piecewise-constant dictionary and LSPS denotes local-set-based piecewise-smooth dictionary. For the windowed graph Fourier transform, we use 15 filters; for LSPS, three piecewise-smooth models provide tight performances; here we show the results of the piecewise-polynomial smooth model with degree $K = 2$. We see that the local-set-based dictionaries perform better than the windowed graph Fourier transform; local-set-based piecewise-smooth dictionary is slightly better than local-set-based piecewise-constant dictionary; even though the windowed graph Fourier transform is solid in theory, provides highly redundant representations and is useful for visualization, it does not well approximate complex graph signals.



(a) Minnesota. (b) U.S weather station.

Fig. 4: Approximation error. Approximation ratio is the percentage of used coefficients (s in (1)).

5. CONCLUSIONS

We proposed representations to analyze piecewise-smooth graph signals by defining classes and using multiresolution local sets as a tool to analyze such graph signals. We then proposed local-set based dictionaries as graph signal representations. In the experiments, we showed that local-set-based dictionaries outperform graph Fourier domain based representations when approximating both simulated and real-world graph signals.

6. REFERENCES

[1] M. Vetterli, J. Kovačević, and V. K. Goyal, *Foundations of Signal Processing*, Cambridge University Press, 2014, http://www.fourierandwavelets.org/.

[2] M. Newman, *Networks: An Introduction*, Oxford University Press, 2010.

[3] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, pp. 83–98, May 2013.

[4] A. Anis, A. Gadde, and A. Ortega, "Towards a sampling theorem for signals on arbitrary graphs," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2014, pp. 3864–3868.

[5] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, 2015, To appear.

[6] S. Chen, R. Varma, A. Singh, and J. Kovačević, "Signal recovery on graphs: Random versus experimentally designed sampling," in *SampTA*, Washington, DC, May 2015.

[7] S. Chen, A. Sandryhaila, J. M. F. Moura, and J. Kovačević, "Signal recovery on graphs: Variation minimization," *IEEE Trans. Signal Process.*, 2015, To appear.

[8] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 21, pp. 5–30, July 2006.

[9] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, New York, NY, third edition, 2009.

[10] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neur. Comput.*, vol. 13, pp. 1373–1396, 2003.

[11] U. V. Luxburg, "A tutorial on spectral clustering," *Statistics and Computing.*, vol. 17, pp. 395–416, 2007.

[12] X. Wang, P. Liu, and Y. Gu, "Local-set-based graph signal reconstruction," *IEEE Trans. Signal Process.*, vol. 63, no. 9, May 2015.

[13] J. Sharpnack, A. Rinaldo, and A. Singh, "Sparsistency of the edge lasso over graphs," in *AISTATS*, La Palma, Canary Islands, Apr. 2012.

[14] Yu-Xiang Wang, James Sharpnack, Alex Smola, and Ryan J. Tibshirani, "Trend filtering on graphs," in *AISTATS*, San Diego, CA, May 2015.

[15] A. Krishnamurthy J. Sharpnack and A. Singh, "Detecting activations over graphs using spanning tree wavelet bases," in *AISTATS*, Scottsdale, AZ, Apr. 2013.

[16] 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 *Proc. Int. Conf. Mach. Learn.*, Haifa, Israel, June 2010, pp. 367–374.

[17] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Appl. Comput. Harmon. Anal.*, pp. 53–94, July 2006.

[18] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, pp. 129–150, Mar. 2011.

[19] M. Crovella and E. Kolaczyk, "Graph wavelets for spatial traffic analysis," in *Proc. IEEE INFOCOM*, Mar. 2013, vol. 3, p. 18481857.

[20] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Appl. Comput. Harmon. Anal.*, February 2015, To appear.

[21] R. Castro, R. Willett, and R. Nowak, "Faster rates in regression via active learning," in *Proc. Neural Information Process. Syst.*, Vancouver, Dec. 2005.

[22] X. Zhu, "Semi-supervised learning literature survey," Tech. Rep. 1530, Univ. Wisconsin-Madison, 2005.

[23] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signal, Syst. Comput.*, Pacific Grove, CA, Nov. 1993, vol. 1, pp. 40–44.

[24] D. Gleich, "The matlabbgl matlab library," http://www.cs.purdue.edu/homes/dgleich/packages/matlab bgl/index.html.