On the Sparsity of Wavelet Coefficients for Signals on Graphs

Benjamin Ricaud, David I Shuman, and Pierre Vandergheynst

Ecole Polytechnique Fédérale de Lausanne (EPFL), Signal Processing Laboratory (LTS2), Lausanne, Switzerland

{benjamin.ricaud, david.shuman, pierre.vandergheynst}@epfl.ch

ABSTRACT

A number of new localized, multiscale transforms have recently been introduced to analyze data residing on weighted graphs. In signal processing tasks such as regularization and compression, much of the power of classical wavelets on the real line is derived from their theoretically and empirically proven ability to sparsely represent piecewise-smooth signals, which appear to be locally polynomial at sufficiently small scales. As of yet in the graph setting, there is little mathematical theory relating the sparsity of localized, multiscale transform coefficients to the structures of graph signals and their underlying graphs. In this paper, we begin to explore notions of global and local regularity of graph signals, and analyze the decay of spectral graph wavelet coefficients for regular graph signals.

Keywords: Signal processing on graphs, spectral graph wavelets, sparsity, regularity

1. INTRODUCTION

In addition to their use in signal analysis and compression, multiscale transforms such as wavelets are useful tools in ill-posed inverse problems such as denoising, deconvolution, classification, and regression. For signals on the real line, the wavelet transform can capture local behavior of the signal, detect and characterize signals' singularities (e.g., discontinuities, non-differentiable points), and sparsely represent globally smooth or piecewise-smooth (locally regular) signals. Therefore, in ill-posed inverse problems, the sparsity of the wavelet coefficients is often imposed as a regularization term.

Over the past decade, a number of new localized, multiscale transforms have been introduced to analyze data residing on weighted graphs (see Refs. 1-3 for just a few examples, and Ref. 4 for a recent overview). These wavelet constructions generalize ideas from classical wavelets on Euclidean spaces in slightly different manners, but empirically they all seem to sparsely represent smooth and piecewise-smooth signals on graphs reasonably well. However, a major theoretical question still goes largely unanswered: for what classes of graphs signals are the wavelet coefficients sparse? The answer to this question could provide further guidance as to which wavelet transforms are best suited to which signal processing tasks and for what types of signals, as well as insights into how to set various parameters of the wavelet transforms. As always in the area of signal processing on graphs, much of the challenge lies in adapting classical notions such as regularity and smoothness to the discrete, irregular graph domain in a meaningful manner that incorporates the structure of the underlying weighted graph.

In related work, Zhu and Rabbat analyze the decay of graph Fourier coefficients and the M-term linear Fourier approximation error for globally smooth graph signals, which they refer to as signals with bounded variation.[5] Gavish, Nadler, and Coifman define a notion of Hölder regularity for functions residing on trees, and use it to relate the smoothness of functions on trees to the decay of the coefficients of the wavelet-like transforms they propose.[6,7] Führ and Wild analyze wavelet coefficient decay of discrete-time signals on $\mathbb{Z}$ by introducing discrete-time Besov spaces.[8] Maggioni and Mhaskar define Besov spaces for functions residing on metric measure spaces.[9]

In this paper, we begin to explore notions of global and local regularity of graph signals, and analyze the decay of spectral graph wavelet coefficients for regular graph signals.

2. NOTIONS OF GLOBAL REGULARITY FOR SIGNALS ON GRAPHS

A graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, w\}$ consists of a set of vertices $\mathcal{V}$ connected by a set of edges $\mathcal{E}$, and a function $w : \mathcal{E} \rightarrow \mathbb{R}^+$ that assigns a weight, $w(m,n)$, to an edge, $e = (m,n)$ connecting vertices m and n. We let $N = |\mathcal{V}|$ be the number of vertices in the graph. We consider functions f of the form $f : \mathcal{V} \rightarrow \mathbb{R}$ that associate a real value to each vertex of the graph. We can also view such an f as a vector in $\mathbb{R}^N$. This vector contains the data we would like to analyze using a wavelet transform, and we therefore often refer to f as a *graph signal*.

Just as in classical Euclidean settings, notions of regularity and smoothness can be built up from *discrete calculus*[10] notions of derivatives and gradients. The *edge derivative* of a signal $\mathbf{f}$ with respect to edge $e = (m,n)$ at vertex m is defined as

$$\left.\frac{\partial \mathbf{f}}{\partial e}\right|_m := \sqrt{w(m,n)}\,[f(n) - f(m)],$$

and the *graph gradient* of $\mathbf{f}$ at vertex m is the vector

$$\nabla_m \mathbf{f} := \left[\left\{\left.\frac{\partial \mathbf{f}}{\partial e}\right|_m\right\}_{e\in\mathcal{E} \text{ s.t. } e=(m,n) \text{ for some } n\in\mathcal{V}}\right].$$

Summing over the squared norms of the graph gradients at each vertex (the local variations) yields

$$\frac{1}{2}\sum_{m\in V}\|\nabla_m \mathbf{f}\|_2^2 = \frac{1}{2}\sum_{m\in V}\left[\sum_{n\in\mathcal{N}_m} w(m,n)\,[f(n)-f(m)]^2\right] = \sum_{(m,n)\in\mathcal{E}} w(m,n)\,[f(n)-f(m)]^2 = \mathbf{f}^{\mathrm{T}}\mathcal{L}\mathbf{f}, \qquad (1)$$

where $\mathcal{N}_m$ is the set of vertex m's neighbors in the graph, and $\mathcal{L}$ is the graph Laplacian operator, whose action is defined by $(\mathcal{L}f)(n) := \sum_{m=1}^{N} w(m,n)\,[f(n) - f(m)]$. The term $\mathbf{f}^{\mathrm{T}}\mathcal{L}\mathbf{f}$ in (1) is known as the *graph Laplacian quadratic form*,[11] and it gives rise to a semi-norm

$$\|\mathbf{f}\|_{\mathcal{L}} := \|\mathcal{L}^{\frac{1}{2}}\mathbf{f}\|_2 = \sqrt{\mathbf{f}^{\mathrm{T}}\mathcal{L}\mathbf{f}}.$$

From the second-to-last term in (1), it is clear that $\|\mathbf{f}\|_{\mathcal{L}}$ is small if and only if the graph signal $\mathbf{f}$ has similar values at neighboring vertices connected by an edge with a large weight, and the semi-norm $\|\cdot\|_{\mathcal{L}}$ therefore measures the global smoothness of signals. Accordingly, it is often used as a regularization term in Tikhonov regularization problems, in order to enforce prior information that the target signal is globally smooth (see, e.g., Refs. 12-14).

This first notion of global smoothness can be generalized in a few different ways. One common method is to generalize it to the *discrete p-Dirichlet form* (see, e.g., Ref. 14), which is defined as

$$S_p(\mathbf{f}) := \frac{1}{p}\sum_{m\in\mathcal{V}}\|\nabla_m \mathbf{f}\|_2^p = \frac{1}{p}\sum_{m\in V}\left[\sum_{n\in\mathcal{N}_m} w(m,n)\,[f(n)-f(m)]^2\right]^{\frac{p}{2}},$$

where we can see that $S_2(\mathbf{f}) = \mathbf{f}^{\mathrm{T}}\mathcal{L}\mathbf{f}$. In this paper, we generalize (1) in a slightly different manner. Denoting the real eigenvalues and associated eigenvectors of the graph Laplacian $\mathcal{L}$ by $\{\lambda_\ell, u_\ell\}_{\ell=0,1,\ldots,N-1}$,* we define a *discrete Sobolev semi-norm* as

$$\|f\|_{\mathcal{H}^p} := \|\mathcal{L}^p f\|_2 = \|\widehat{\mathcal{L}^p f}\|_2 = \sqrt{\sum_{\ell}|\lambda_\ell|^{2p}|\hat{f}(\ell)|^2}, \qquad (2)$$

where $\hat{f}(\ell) := \langle f, u_\ell\rangle$ is the graph Fourier transform of f. A few brief remarks about $\|\cdot\|_{\mathcal{H}^p}$ are in order. First, $\|\cdot\|_{\mathcal{H}^{\frac{1}{2}}} = \|\cdot\|_{\mathcal{L}}$. Second, like $\|\cdot\|_{\mathcal{L}}$, $\|\cdot\|_{\mathcal{H}^p}$ is only a semi-norm because $\|f\|_{\mathcal{H}^p} = 0$ for any graph signal f whose values are the same on every vertex. Third, in classical Euclidean settings it is common to define Sobolev spaces

*In order to omit complex conjugates for notational simplicity, we assume that the graph signal f and the eigenvectors $\{u_\ell\}$ are all real.

as classes of signals with a finite Sobolev norm. In the graph setting, $\|f\|_{\mathcal{H}^p}$ is always finite, and, moreover, $\frac{\|f\|_{\mathcal{H}^p}}{\|f\|_2} \leq \lambda_{\max}^p$ for all $f \in \mathbb{R}^N$. Fourth, the fact that we can represent $\|f\|_{\mathcal{H}^p}$ straightforwardly in the graph spectral domain, as in (2), is important for our ability to relate this smoothness measure with the sparsity of the spectral graph wavelets, which are defined via the graph spectral domain. Fifth, this definition is a discrete analog to the quantity used in the continuous setting to define the space $\mathbb{W}^p(\mathbb{R})$ of p-times differentiable Sobolev functions (see, e.g., Ref. 15, pp. 438-9), which are those functions satisfying

$$\int_{-\infty}^{\infty} |\omega|^{2p} |\hat{f}(\omega)|^2 d\omega < \infty.$$

Finally, we should mention that previous works have also presented slightly different definitions of discrete Sobolev norms and semi-norms. These include Ref. 16, which defines a proper norm $\sqrt{\|f\|_{\mathcal{L}}^2 + \|f\|_2^2}$, and Ref. 17, which for $p \geq 1$ defines a weighted Sobolev semi-norm as $\|f\| := \left(\sum_{(m,n)\in\mathcal{E}} w(m,n)|f(m)-f(n)|^p\right)^{\frac{1}{p}}$.

3. SPECTRAL GRAPH WAVELETS

A spectral graph wavelet[3] at scale s and centered at vertex n is of the form

$$\psi_{s,n}(i) := (T_n D_s g)(i) = \sum_{\ell=0}^{N=1} \hat{g}(s\lambda_\ell) u_\ell(n) u_\ell(i), \; \forall i \in \mathcal{V}, \tag{3}$$

where D_s is a generalized dilation operator satisfying $\widehat{\mathcal{D}_s g}(\lambda) = \hat{g}(s\lambda)$, and T_n is a generalized translation operator defined as

$$(T_n g)(i) := \sum_{\ell=0}^{N-1} \hat{g}(\lambda_\ell) u_\ell(n) u_\ell(i).$$

In this paper, we assume that the wavelet kernel $\hat{g} : \mathbb{R}^+ \to \mathbb{R}$ is a continuous function defined on the entire non-negative real line, and it satisfies $\hat{g}(0) = 0$, $\lim_{\lambda\to\infty} \hat{g}(\lambda) = 0$, and the admissibility condition in Lemma 5.1 of Ref. 3. The spectral graph wavelet transform coefficient of a graph signal f at scale s and location n is then given by

$$\Psi f(s,n) = \langle f, \psi_{s,n} \rangle = \sum_\ell \hat{g}(s\lambda_\ell) u_\ell(n) \hat{f}(\lambda_\ell).$$

4. WAVELET COEFFICIENT DECAY OF GLOBALLY REGULAR GRAPH SIGNALS

In this section, we examine different relations between the global regularity of a graph signal and the decay of its spectral graph wavelet coefficients. We start with a simple motivating example. Consider spectral graph wavelets of the form (3), with a mother wavelet $\hat{g}(\lambda) := \mathbb{1}_{\{1\leq\lambda<2\}}$, and a discrete set of wavelet scales $\{s_j\}_{j=J_{\min},J_{\min}+1,\ldots,J_{\max}}$, where $s_j := 2^{-j}$ and $J_{\min}, J_{\max} \in \mathbb{Z}$. Then we have

$$\begin{aligned}
\sum_{j=J_{\min}}^{J_{\max}} s_j^{-2} \sum_{i=1}^{N} |\langle \psi_{s_j,i}, f\rangle|^2 &= \sum_{j=J_{\min}}^{J_{\max}} s_j^{-2} \sum_{\ell=0}^{N-1} |\hat{g}(s_j\lambda_\ell)|^2 |\hat{f}(\lambda_\ell)|^2 \\
&= \sum_{\ell=0}^{N-1} \left(\sum_{j=J_{\min}}^{J_{\max}} \left[\frac{\hat{g}(s_j\lambda_\ell)}{s_j}\right]^2 \right) |\hat{f}(\lambda_\ell)|^2 \\
&\leq \sum_{\ell=0}^{N-1} \lambda_\ell^2 |\hat{f}(\lambda_\ell)|^2 = \|\mathcal{L}f\|_2^2 = \|f\|_{\mathcal{H}^1}^2.
\end{aligned} \tag{4}$$

If the right-hand side of (4) is small, then the left-hand side must also be small, and, in particular, for small values of $s_j = 2^{-j}$ (large indices j), $|\langle \psi_{s_j,i}, f\rangle|$ must be small. That is, for this choice of wavelet kernels, the wavelet coefficients of globally smooth functions decay rapidly. The next proposition generalizes this example to show that this decay property is not restricted to kernels with that specific form.

Proposition 1. *Let $p \geq 1$, and assume that $C_p := \int_0^\infty |\hat{g}(s)|^2/s^{2p} ds < \infty$. Then*

$$\int_0^\infty s^{-2p} \sum_n |\langle f, \psi_{s,n}\rangle|^2 ds = C_p \|f\|_{\mathcal{H}^{(2p-1)/2}}.$$

Proof. Once again using the fact that $\sum_n |\langle f, \psi_{s,n}\rangle|^2 = \sum_\ell |\hat{g}(s\lambda_\ell)|^2 |\hat{f}(\lambda_\ell)|^2$, we have

$$\begin{aligned}\int_0^\infty s^{-2p} \sum_n |\langle f, \psi_{s,n}\rangle|^2 ds &= \sum_\ell |\hat{f}(\lambda_\ell)|^2 \int_0^\infty s^{-2p} |\hat{g}(s\lambda_\ell)|^2 ds \\ &= \sum_\ell |\hat{f}(\lambda_\ell)|^2 \lambda_\ell^{2p-1} \int_0^\infty t^{-2p} |\hat{g}(t)|^2 dt = C_p \|f\|_{\mathcal{H}^{(2p-1)/2}},\end{aligned}$$

where the first equality in the second line follows from the change of variable $t = s\lambda_\ell$. ☐

Finally, the next proposition shows that when the wavelet kernel is a polynomial with p "vanishing spectral moments," then the wavelet coefficients decay rapidly at a rate dominated by the term $s^p \|f\|^2_{\mathcal{H}^p}$.

Proposition 2. *Assume that $\hat{g}(\lambda) = \sum_{k=p}^q a_k \lambda^k$ for some $p \geq 1$ (implying $\hat{g}(0) = 0$). Then*

$$|\Psi f(s,n)| = |\langle f, \psi_{s,n}\rangle| \leq \sum_{k=p}^q |a_k| s^k \|f\|_{\mathcal{H}^k}.$$

Proof.

$$\begin{aligned}|\Psi f(s,n)| = |\langle f, \psi_{s,n}\rangle| = \left|\sum_\ell \hat{g}(s\lambda_\ell) u_\ell(n) \hat{f}(\lambda_\ell)\right| &= \left|\sum_\ell \sum_{k=p}^q a_k (s\lambda_\ell)^k u_\ell(n) \hat{f}(\lambda_\ell)\right| \\ &\leq \sum_{k=p}^q |a_k| \sum_\ell \left|(s\lambda_\ell)^k u_\ell(n) \hat{f}(\lambda_\ell)\right| \\ &\leq \sum_{k=p}^q |a_k| \sqrt{\sum_\ell |u_\ell(n)|^2} \sqrt{\sum_{\ell'} (s\lambda_{\ell'})^{2k} |\hat{f}(\lambda_{\ell'})|^2} \\ &= \sum_{k=p}^\alpha |a_k| s^k \|f\|_{\mathcal{H}^k},\end{aligned}$$

where the last inequality follows from the Cauchy-Schwarz inequality, and the final equality follows from the orthonormality of the graph Laplacian eigenvectors. ☐

5. ONGOING WORK: LOCAL REGULARITY AND WAVELET COEFFICIENT DECAY OF LOCALLY REGULAR GRAPH SIGNALS

As mentioned in Section 1, much of the recent success of efficient information extraction from high-dimensional data on Euclidean spaces is rooted in the development of transforms that sparsely represent signals with certain properties. For example, in classical wavelet analysis of signals on the real line, piecewise-smooth signals have sparse wavelet coefficients because they appear to be locally polynomial at sufficiently small scales. Such a local notion of smoothness is often captured by Besov or Hölder norms, the latter of which is based on Lipschitz regularity. The Lipschitz regularity of a function around a given point is intimately related via the Taylor series to successive derivatives of the function at the point of interest. A highly regular function may be differentiated a large number of times at that point, whereas an irregular function cannot be differentiated many times at that point.

Empirically, the magnitudes of spectral graph wavelet coefficients of locally smooth graph signals decay as the scale s_j decreases to zero. A simple example of this with a piecewise-constant graph signal is shown in Figure 1. In this section, we begin to investigate potential definitions of local regularity of graph signals and make some preliminary connections between local regularity and spectral graph wavelet coefficient decay.

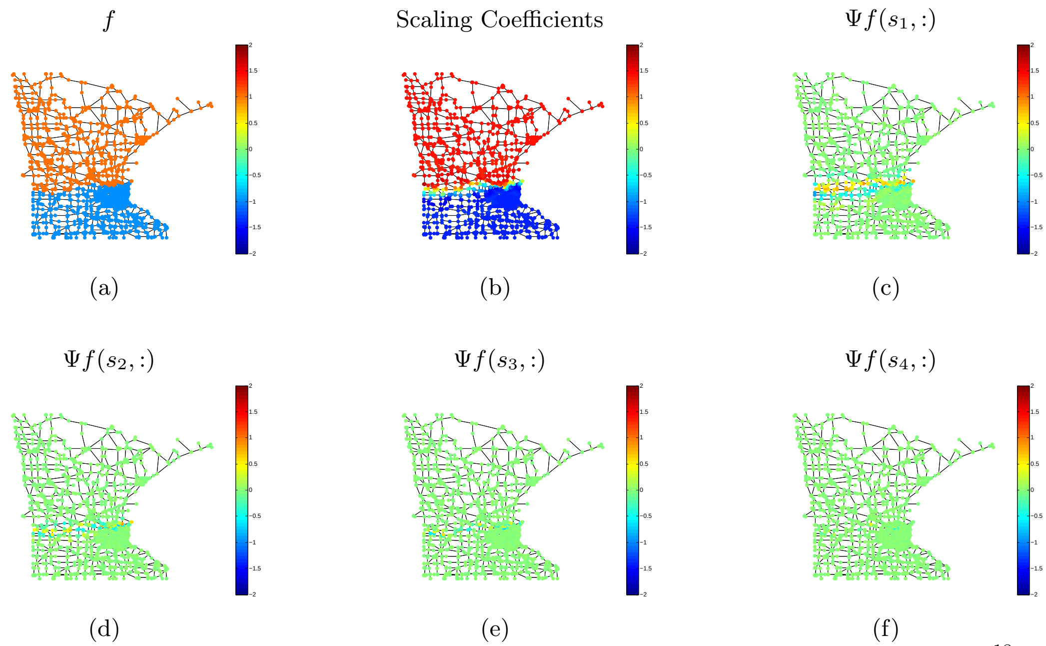

Figure 1. Spectral graph wavelet coefficients of a piecewise-smooth signal on the Minnesota road graph.[18] (a) The signal f is equal to 1 on the top half of the graph and -1 on the bottom half of the graph. (b) The scaling coefficients are not sparse. (c)-(f) The spectral graph wavelet coefficients cluster around the discontinuity and have magnitudes that decay as the wavelet scale s_j decreases. Note that $s_1 > s_2 > s_3 > s_4$.

5.1 Notions of Local Regularity

One notion of local regularity is the *local variation* at vertex m (see, e.g., Ref. 14)

$$\|\nabla_m \mathbf{f}\|_2 := \left[\sum_{e \in \mathcal{E} \text{ s.t. } e=(m,n) \text{ for some } n \in \mathcal{V}} \left(\frac{\partial \mathbf{f}}{\partial e}\Big|_m \right)^2 \right]^{\frac{1}{2}} = \left[\sum_{n \in \mathcal{N}_m} w(m,n) \left[f(n) - f(m) \right]^2 \right]^{\frac{1}{2}},$$

which is small when the function $\mathbf{f}$ has similar values at m and all neighboring vertices of m, and therefore provides a measure of local smoothness of $\mathbf{f}$ around vertex m. However, this notion only captures information about the signal within a one-hop radius around the point of interest.

A second notion of local regularity is to generalize the Hölder regularity definition of Refs. 6 and 7 to general graphs as follows.

Definition 5.1. *A graph signal f is (C, α, r)-Hölder regular with respect to the graph $\mathcal{G}$ at vertex $n \in \mathcal{V}$ if*

$$|f(n) - f(m)| \leq C[d_{\mathcal{G}}(m,n)]^{\alpha}, \ \forall m \in \mathcal{N}(n,r),$$

where $\mathcal{N}(n,r)$ is the neighborhood of all vertices within a shortest-path geodesic distance of r hops from vertex n.

In Def. 5.1, $d_{\mathcal{G}}(m,n)$ may be any distance on the graph, including for example the shortest-path geodesic distance (number of hops), a weighted distance metric that reflects the weights of the edges comprising the shortest path, or the resistance distance, which also reflects the number of paths connecting m and n.

A third idea is to use the graph Laplacian operator as a substitute for derivation, and use the quantity $(\mathcal{L}^k f)(n)$ as a measure of local regularity of the graph signal f in a neighborhood of radius k around vertex n. For example, if we let the function f be constant on all vertices in a neighborhood $\mathcal{N}(n,p)$ of vertices around n, then $(\mathcal{L}^k f)(n) = 0$ for $k \leq p$ and $k \neq 0$. For $k > p$, $|(\mathcal{L}^k f)(n)|$ will be larger when there are sharp jumps in f near the boundary of $\mathcal{N}(n,p)$. This quantity is potentially useful, because if we have a polynomial kernel as in Proposition 2, a spectral graph wavelet takes the form

$$\psi_{s,n}(i) = \sum_{\ell} \hat{g}(s\lambda_\ell) u_\ell(n) u_\ell(i) = \sum_{\ell} \sum_{k=p}^{q} a_k \lambda_\ell^k s^k u_\ell(n) u_\ell(i) = \sum_{k=p}^{q} a_k s^k (\mathcal{L}^k)_{n,i},$$

and thus the wavelet transform of a signal f reads:

$$\Psi f(s,n) = \langle f, \psi_{s,n} \rangle = \sum_{i=1}^{N} f(i) \sum_{k=p}^{q} a_k s^k (\mathcal{L}^k)_{n,i} = \sum_{k=p}^{q} a_k s^k \big(\mathcal{L}^k f\big)(n).$$

5.2 Wavelet Coefficient Decay of Locally Regular Graph Signals

We now use Def. 5.1 to generate a loose bound on the magnitude of the spectral graph wavelet coefficient centered at a vertex n and at scale s (s small) of a graph signal that is locally regular around n. The main idea is that as the scale s goes towards zero, the spectral graph wavelet is more strongly localized around the center vertex n. This fact bounds the magnitude of the inner product of the signal and the wavelet over vertices farther away from n. Meanwhile, the local regularity allows us to bound the magnitude of the inner product of the signal and the wavelet over vertices close to n.

Proposition 3. *Assume that f is (C,α,r)-Hölder regular for some $r \geq 1$, and let $\hat{g}(\lambda) = \sum_{k=r}^{q} a_k \lambda^k$ for some coefficients $\{a_k\}_{k=r,r+1,\ldots,q}$. Then there exist constants C_2 and $\bar{s}$ such that for all $s < \bar{s}$, we have*

$$|\Psi f(s,n)| \leq C r^\alpha \sum_{m \in \mathcal{N}(n,r)} |\psi_{s,n}(m)| + C_2 s^{r+1} \sum_{m \notin \mathcal{N}(n,r)} |f(m) - f(n)|.$$

Proof.

$$\begin{aligned}
|\Psi f(s,n)| = |\langle f, \psi_{s,n} \rangle| &= \left| \sum_{m=1}^{N} f(m) \psi_{s,n}(m) \right| \\
&= \left| \sum_{m=1}^{N} [f(m) - f(n)] \psi_{s,n}(m) \right| && (5) \\
&= \left| \sum_{m \in \mathcal{N}(n,r)} [f(m) - f(n)] \psi_{s,n}(m) + \sum_{m \notin \mathcal{N}(n,r)} [f(m) - f(n)] \psi_{s,n}(m) \right| \\
&\leq \sum_{m \in \mathcal{N}(n,r)} |f(m) - f(n)| |\psi_{s,n}(m)| + \sum_{m \notin \mathcal{N}(n,r)} |f(m) - f(n)| |\psi_{s,n}(m)| \\
&\leq \sum_{m \in \mathcal{N}(n,r)} C [d_{\mathcal{G}}(m,n)]^\alpha |\psi_{s,n}(m)| + \sum_{m \notin \mathcal{N}(n,r)} |f(m) - f(n)| |\psi_{s,n}(m)| \\
&\leq C r^\alpha \sum_{m \in \mathcal{N}(n,r)} |\psi_{s,n}(m)| + \sum_{m \notin \mathcal{N}(n,r)} |f(m) - f(n)| |\psi_{s,n}(m)| && (6) \\
&\leq C r^\alpha \sum_{m \in \mathcal{N}(n,r)} |\psi_{s,n}(m)| + C_2 s^{r+1} \sum_{m \notin \mathcal{N}(n,r)} |f(m) - f(n)|, && (7)
\end{aligned}$$

where (5) follows from the fact that the spectral graph wavelets have zero mean, (6) follows from Def. 5.1, and (7) follows from Eq. (41) in the proof of Theorem 5.5 of Ref. 3. $\square$

Such a rough analysis is just the tip of the iceberg with regards to theoretical connections between classes of graph signals, the underlying graph structure, and sparsity of the wavelet coefficients. Open issues include which distance metrics to use in definitions such as Def. 5.1, which classes of signals are sparsely represented by spectral graph wavelets and other types of wavelets on graphs, converse theorems showing that the sparsity of wavelet coefficients implies some regularity of the signal, and how to use such lines of analysis for guidance on how to optimally design the spectral graph wavelet filters or parameterize other types of wavelets on graphs.

ACKNOWLEDGMENTS

This work was supported by FET-Open grant number 255931 UNLocX.

REFERENCES

[1] Crovella, M. and Kolaczyk, E., "Graph wavelets for spatial traffic analysis," in [*Proc. IEEE INFOCOM*], **3**, 1848–1857 (Mar. 2003).

[2] Coifman, R. R. and Maggioni, M., "Diffusion wavelets," *Appl. Comput. Harmon. Anal.* **21**(1), 53–94 (2006).

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

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

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

[6] Gavish, M., Nadler, B., and Coifman, R. R., "Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning," in [*Proc. Int. Conf. Mach. Learn.*], 367–374 (Jun. 2010).

[7] Coifman, R. R. and Gavish, M., "Harmonic analysis of digital data bases," in [*Wavelets and Multiscale Analysis: Theory and Applications*], Cohen, J. and Zayed, A. I., eds., 161–197, Springer (2011).

[8] Führ, H. and Wild, M., "Characterizing wavelet coefficient decay of discrete-time signals," *Appl. Comput. Harmon. Anal.* **20**, 184–201 (Mar. 2006).

[9] Maggioni, M. and Mhaskar, H. N., "Diffusion polynomial frames on metric measure spaces," *Appl. Comput. Harmon. Anal.* **24**, 329–353 (May 2008).

[10] Grady, L. J. and Polimeni, J. R., [*Discrete Calculus*], Springer (2010).

[11] Spielman, D., "Spectral graph theory," in [*Combinatorial Scientific Computing*], Chapman and Hall / CRC Press (2012).

[12] Zhou, D. and Schölkopf, B., "A regularization framework for learning from graph data," in [*Proc. ICML Workshop Stat. Relat. Learn. and Its Connections to Other Fields*], 132–137 (Jul. 2004).

[13] Belkin, M., Matveeva, I., and Niyogi, P., "Regularization and semi-supervised learning on large graphs," in [*Learn. Theory*], *Lect. Notes Comp. Sci.*, 624–638, Springer-Verlag (2004).

[14] Elmoataz, A., Lezoray, O., and Bougleux, S., "Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing," *IEEE Trans. Image Process.* **17**, 1047–1060 (Jul. 2008).

[15] Mallat, S. G., [*A Wavelet Tour of Signal Processing*], Academic Press (2008).

[16] Mahadevan, S. and Maggioni, M., "Value function approximation with diffusion wavelets and Laplacian eigenfunctions," in [*Adv. Neural Inf. Process. Syst. 18*], Weiss, Y., Schölkopf, B., and Platt, J., eds., 843–850, MIT Press, Cambridge, MA (2006).

[17] Ostrovskii, M. I., "Sobolev spaces on graphs," *Quaestiones Mathematicae* **28**(4), 501–523 (2005).

[18] Gleich, D., "The MatlabBGL Matlab library," http://www.cs.purdue.edu/homes/dgleich/.