Graph Wedgelets: Adaptive Data Compression on Graphs Based on Binary Wedge Partitioning Trees and Geometric Wavelets

Wolfgang Erb

***Abstract*—We introduce graph wedgelets - a tool for data compression on graphs based on the representation of signals by piecewise constant functions on adaptively generated binary graph partitionings. The adaptivity of the partitionings, a key ingredient to obtain sparse representations of a graph signal, is realized in terms of recursive wedge splits adapted to the signal. For this, we transfer adaptive partitioning and compression techniques known for 2D images to general graph structures and develop discrete variants of continuous wedgelets and binary space partitionings. We prove that continuous results on best m-term approximation with geometric wavelets can be transferred to the discrete graph setting and show that our wedgelet representation of graph signals can be encoded and implemented in a simple way. Finally, we illustrate that this graph-based method can be applied for the compression of images as well.**

***Index Terms*—Binary graph partitioning, big data compression, geometric wavelets, greedy algorithms on graphs, non-linear approximation, graph wedgelets.**

I. Introduction

IN LINE with the extraordinarily fast growth of stored and transmitted digital information, and the increase in complexity and interdependency of this Big Data, there is a strong need of novel compression techniques that are able to efficiently compress large data sets on unstructured or semi-structured domains. In many cases, these data sets and their interrelations can be organized in terms of networks or graphs as underlying domains. Adaptive algorithms able to compress data based on its intrinsic content as well as on the topological structure of the surrounding graph environment are therefore of main importance.

Efficient storage of data in image and signal processing depends on how sparsely the data can be represented in terms of suitable chosen dictionaries. The most common representation techniques for graph signals have corresponding counterparts in image processing and comprise, for instance, analogs of the Fourier transform, the wavelet transform, or more general space-frequency decompositions. A general overview about some of these techniques in graph signal processing can be found in [29], [39].

Manuscript received 21 October 2021; revised 20 May 2022 and 24 November 2022; accepted 18 December 2022. Date of publication 31 January 2023; date of current version 8 February 2023. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. David I Shuman.

The author is with the Dipartimento di Matematica "Tullio Levi-Civita", Università degli Studi di Padova, 35122 Padova, Italy (e-mail: wolfgang.erb@lissajous.it).

Digital Object Identifier 10.1109/TSIPN.2023.3240899

Focusing on wavelet constructions, there are several approaches that give raise to a wavelet-type multiresolution analysis on graphs. The most prominent works in this direction include diffusion wavelets [2], [9], [42], wavelets and vertex-frequency localized frames based on the spectral decomposition of the graph Laplacian [20], [38], [40], [41], graph wavelet filterbanks [30], lifting based wavelets [24], [37], as well as the construction of wavelets based on partitioning trees [8], [17], [28]. Interesting for us is mainly the latter approach. This is for two reasons: partitions of the graph vertices can be generated very adaptively and efficiently by a multitude of available graph clustering techniques, as, for instance, J-center clustering [18] or spectral clustering [43]; Haar-type wavelets based on hierarchical partitioning trees are easy and cost-efficient to implement, mainly due to the underlying tree structure and the inherent orthogonality of the involved basis functions. Particular construction of Haar wavelets and dictionaries based on hierarchical spectral clustering, k-medoids clustering or spanning trees are, for instance, described in [21], [22], [23], [36]. More general wavelet-type orthogonal systems on weighted partitioning trees have been studied in [5]. Wavelets based on partitioning trees have several applications in machine learning as well, in particular for scattering networks [4], and for semi-supervised learning [17]. Further, in [34] improved Haar wavelets for classes of smooth functions have been computed via a deep learning approach.

The partitioning trees in the works above are solely guided by the topology of the graph and do not take geometric properties of graph signals into account. As shown in [17], the particular structure of the partitions has however a strong impact on how well a signal can be approximated sparsely in terms of the Haar wavelets. For an efficient compression of graph signals it is therefore essential that the partitioning trees are adapted to the signal to be compressed.

Goal of this work is therefore to go one step beyond the established non-adaptive constructions of partitioning trees and to develop and analyze new partitioning strategies for graph wavelets that allow for a signal-driven adaptivity in the refinement of the partitions. This can be regarded as an attempt to introduce a new generation of geometric signal-adapted wavelets intrinsically defined on graphs.

For the compression of images, several approaches for the generation of adaptive partitions are known. Using a function on a continuous 2D domain to describe the image, these approaches

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

usually involve an adaptive segmentation of the image in which the image is approximated by piecewise constant or polynomial functions on the extracted segments. The main idea of this type of compression scheme, and, at the same type also the inherent challenge, is to find a cost-efficient and meaningful splitting procedure such that the resulting segmentation contains only a few relevant elements. On these relevant segments the image is then approximated by simple elementary functions, mainly constant functions or low-order polynomials. If such a meaningful segmentation is found, the resulting compression schemes are highly competitive for low-bit compression [32]. Important examples of such adaptive segmentation schemes are adaptive triangulations [7], [11], [12], [13], quadtree approximations [27], [35], tetralets [26], wedgelets [15], [16], [44] or binary space partitioning trees [31], [32]. The latter two, wedgelets and binary space partitioning trees, will be the most relevant for this work, as their main concepts can be transferred easily to partitions on graphs. In particular, for binary space partitionings, there exists a well-developed theory on the m-term approximation with geometric wavelets [10], [25] that can be translated directly to the graph setting. In this work, the respective discrete partitionings will be called binary graph partitionings (BGPs). Ideas from continuous wedgelet decompositions and binary space partitionings will further lead to the development of the new discrete graph wedgelets.

A. Main Contributions

1) We provide a theoretical framework for the sparse approximation of graph signals with geometric wavelets defined upon adaptive binary graph partitioning (BGP) trees. This will be done in terms of non-linear m-term approximation of functions in discrete Besov-type smoothness classes on graphs. This is an adaption to the discrete graph setting of corresponding continuous results developed in [10], [25].
2) We will give a simple and highly efficient novel construction of BGP trees in terms of recursive wedgelet splits on graphs. We will refer to them as binary wedge partitioning (BWP) trees. The BWP trees can be implemented and stored cost-efficiently by an ordered set of graph nodes.
3) In several experiments, we will study the properties of BWP trees and analyze how well signals on graphs or images can be approximated using adaptive BWPs.

B. Basic Terminology on Graphs

In this work, we consider simple graphs $G = (V, E, \mathbf{A}, \mathrm{d})$ with the following structural components:

1) A set $V = \{\mathrm{v}_1, \ldots, \mathrm{v}_n\}$ consisting of n graph vertices.
2) A set $E \subseteq V \times V$ containing all edges $e_{i,i'} = (\mathrm{v}_i, \mathrm{v}_{i'})$, $i \neq i'$, of the graph G. We will assume that G is undirected.
3) A symmetric adjacency matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ with

$$\begin{aligned} \mathbf{A}_{i,i'} &> 0 \quad \text{if } i \neq i' \text{ and } \mathrm{v}_i, \mathrm{v}_{i'} \text{ are connected,} \\ \mathbf{A}_{i,i'} &= 0 \quad \text{else.} \end{aligned} \tag{1}$$

The positive elements $\mathbf{A}_{i,i'}, i \neq i'$, of the adjacency matrix $\mathbf{A}$ contain the connection weights of the edges $e_{i,i'} \in E$.

4) The graph geodesic distance d on the vertex set V, i.e., the length of the shortest path connecting two graph nodes. The distance d satisfies a triangle inequality and, as G is undirected, defines a metric on V. We assume that G is a connected graph and, thus, that the distance d between two arbitrary nodes is finite.

In this work, we are interested in decompositions of graph signals, i.e. of the functions $x : V \to \mathbb{R}$ on the vertex set V of the graph G. By $\mathcal{L}(V)$, we denote the corresponding n-dimensional vector space of graph signals. As the vertices in V are ordered, we can represent every signal x also as a vector $x = (x(\mathrm{v}_1), \ldots, x(\mathrm{v}_n))^\intercal \in \mathbb{R}^n$. We can endow the space $\mathcal{L}(V)$ with the inner product

$$y^\intercal x := \sum_{i=1}^{n} x(\mathrm{v}_i) y(\mathrm{v}_i). \tag{2}$$

The Hilbert space with the norm $\|x\|^2_{\mathcal{L}^2(V)} = x^\intercal x$, will be denoted as $\mathcal{L}^2(V)$. The system $\{\delta_{\mathrm{v}_1}, \ldots, \delta_{\mathrm{v}_n}\}$ of unit vectors forms a canonical orthonormal basis of $\mathcal{L}^2(V)$, where $\delta_{\mathrm{v}_{i'}}$ are defined as $\delta_{\mathrm{v}_{i'}}(\mathrm{v}_i) = \delta_{i,i'}$ for $i, i' \in \{1, \ldots, n\}$. In addition, we consider the $\mathcal{L}^r(V)$ spaces equipped with the quasi-norms

$$\|x\|_{\mathcal{L}^r(V)} = \left(\sum_{i=1}^{n} |x(\mathrm{v}_i)|^r \right)^{\frac{1}{r}}, \quad r > 0.$$

It is well-known that for $r \geq 1$ the latter quantity satisfies a triangle inequality and, thus, defines a norm.

II. Binary Graph Partitionings (BGPs)

The theory of geometric graph wavelets is based on a signal-driven recursive binary partitioning of the vertex set V. In particular, the graph partitioning will be adapted to the graph topology as well as on the approximated signal. We start with a general theory on binary partitioning trees on graphs.

Definition II.1: A *binary graph partitioning (BGP) tree* $\mathcal{T}$ of the graph G is a binary tree consisting of subsets of the vertex set V that can be ordered recursively in partitions $\mathcal{P}^{(m)}$, $m \in \mathbb{N}$, of V by the following rules:

1) The vertex set V is the root of the BGP tree $\mathcal{T}$ and provides the first trivial partition $\mathcal{P}^{(1)} = \{V\}$.
2) If $\mathcal{P}^{(m)} = \{W_1^{(m)}, \ldots, W_m^{(m)}\}$ is a partition of V consisting of m elements in the BGP tree $\mathcal{T}$, then the next partition $\mathcal{P}^{(m+1)}$ of V in $\mathcal{T}$ is obtained by applying a dyadic split to one of the subsets in $\mathcal{P}^{(m)}$.

If W' is an element of $\mathcal{P}^{(m+1)}$ obtained from a dyadic split of a set $W \in \mathcal{P}^{(m)}$, then $W' \subset W$ corresponds to a *child* of W in the tree $\mathcal{T}$. We call two elements $W', W'' \in \mathcal{T}$ *siblings* if both are children of the same $W \in \mathcal{T}$. Note that, as $\mathcal{T}$ is binary, a set $W \in \mathcal{T}$ can only have two children or no children at all. In the latter case we call W a *leave* of the tree $\mathcal{T}$.

We call a BGP tree $\mathcal{T}$ *balanced* if there exists $\frac{1}{2} \leq \rho < 1$ such that for every child W' of an element $W \in \mathcal{T}$ we have

$$(1-\rho)|W| \leq |W'| \leq \rho |W|.$$

We call a BGP tree $\mathcal{T}$ *complete* if it has n leaves, each containing a single vertex of the graph G. A complete and balanced BGP tree will be referred to as *BGP(ρ) tree*.

To see whether a graph signal f can be approximated sparsely by piecewise constant functions on the elements of a BGP tree, we will analyze the $\mathcal{L}^2$-error $\|f - \mathcal{S}_m(f)\|_{\mathcal{L}^2(V)}$, where $\mathcal{S}_m(f)$ denotes the best m-term approximation

$$\mathcal{S}_m(f) = \sum_{i=1}^{m} \psi_{W_i}(f)(\mathrm{v}) \tag{3}$$

of the function f in terms of m wavelets $\psi_{W_i}(f)$, $i \in \{1, \dots, m\}$. These Haar-type wavelets are determined by the elements W_i of a BGP tree $\mathcal{T}$, and sorted descendingly in terms of the $\mathcal{L}^2$-norm:

$$\|\psi_{W_1}(f)\|_{\mathcal{L}^2(V)} \geq \|\psi_{W_2}(f)\|_{\mathcal{L}^2(V)} \geq \|\psi_{W_3}(f)\|_{\mathcal{L}^2(V)} \geq \cdots.$$

The wavelets with respect to a BGP tree $\mathcal{T}$ are defined in the following way: let $W,' W \in \mathcal{T}$ such that W' is a child of W. Then, the wavelet component $\psi_{W'}(f)$ is given as the signal

$$\psi_{W'}(f)(\mathrm{v}) = \left(\frac{\langle f, \chi_{W'} \rangle}{|W'|} - \frac{\langle f, \chi_W \rangle}{|W|} \right) \chi_{W'}(\mathrm{v}), \tag{4}$$

where $\chi_{W'}$ denotes the indicator function of the set W'. In this way, we obtain for every child W' in $\mathcal{T}$ a wavelet component $\psi_{W'}(f)$ of f. For the root $V \in \mathcal{T}$, we additionally set

$$\psi_V(f)(\mathrm{v}) = \frac{\langle f, \chi_V \rangle}{|V|}.$$

Now, picking the m components with the largest $\mathcal{L}^2$-norm, we obtain exactly the non-linear m-term approximation $\mathcal{S}_m(f)$ of f given in (3). If the BGP tree $\mathcal{T} = \mathcal{T}(f)$ depends on f the respective wavelets are called *geometric wavelets*.

To study the convergence of the m-term approximation the following energy functional is of main relevance (see [14] for a general overview). It is the discrete counterpart of a corresponding functional given in [10] for binary space partitionings in hypercubes. In wavelet theory, it is usually used in the characterization of Besov spaces and measures in some sense the sparseness of the wavelet representation of a signal f. In our case, this sparseness is strongly related to the partitions given within the BGP tree $\mathcal{T}$.

Definition II.2: For $0 < r < \infty$, we define the r-energy of the wavelet components of f with respect to a BGP tree $\mathcal{T}$ as

$$\mathcal{N}_r(f, \mathcal{T}) = \left(\sum_{W \in \mathcal{T}} \|\psi_W(f)\|^r_{\mathcal{L}^2(V)} \right)^{\frac{1}{r}}.$$

We can say the following about the decomposition of f in terms of BGP wavelets. The proof is given in the Appendix.

Theorem II.3: Let G be a graph with n nodes, and $\mathcal{T}$ a BGP(ρ) tree on G, i.e., $\mathcal{T}$ is complete and balanced. Then:

i) The tree $\mathcal{T}$ contains $2n - 1$ elements.
ii) For every signal $f \in \mathcal{L}(V)$ we have

$$f = \sum_{j=1}^{2n-1} \psi_{W_j}(f),$$

i.e., f can be decomposed in terms of $2n - 1$ wavelets.
iii) For $0 < r < 2$, we have

$$\|f\|_{\mathcal{L}^2(V)} \leq C \mathcal{N}_r(f, \mathcal{T})$$

with a constant $C > 0$ depending only on ρ.

III. m-Term Approximation Error for Geometric Wavelets on Near Best BGP Trees

Similar as the r-energy functional $\mathcal{N}_r(f, \mathcal{T})$ also the following Besov-type smoothness term quantifies how well a function f can be approximated with piecewise constant functions on the elements of a BGP tree.

Definition III.1: For $\alpha > 0$, $\frac{1}{2} \leq \rho < 1$, and $0 < r < \infty$, we define the geometric Besov-type smoothness measure $|\cdot|_{\mathcal{GB}^\alpha_r}$ of a function $f \in \mathcal{L}(V)$ as

$$|f|_{\mathcal{GB}^\alpha_r} = \inf_{\mathcal{T} \in \mathrm{BGP}(\rho)} \left(\sum_{W \in \mathcal{T}} |W|^{-\alpha r} \sup_{\mathrm{w} \in W} \sum_{\mathrm{v} \in W} |f(\mathrm{v}) - f(\mathrm{w})|^r \right)^{\frac{1}{r}}.$$

In [10] (and similarly in [25]), the corresponding spaces of functions have been referred to as geometric B-spaces. In contrast to the r-energy introduced in Definition II.2, the smoothness measure $|f|_{\mathcal{GB}^\alpha_r}$ is not linked to one particular BGP tree but allows to quantify the sparseness of f with respect to a largy family of BGP(ρ) trees. This can be taken into account also for the r-energy $\mathcal{N}_r(f, \mathcal{T})$ by calculating the infimum over all possible BGP(ρ) trees.

In practice, it might not be possible to determine the infimum over all trees, only an approximate solution might be feasible. Therefore, we say that an f-adapted BGP(ρ) tree $\mathcal{T}_r(f)$ is a near best BGP(ρ) tree if there exists a constant $C > 0$ such that

$$\mathcal{N}_r(f, \mathcal{T}_r(f)) \leq C \inf_{\mathcal{T} \in \mathrm{BGP}(\rho)} \mathcal{N}_r(f, \mathcal{T}).$$

This is the setting we have in mind when we design greedy algorithms in the next section to create adaptive partitionings for the compression of graph signals. For near best BGP(ρ) trees, we have the following relation.

Theorem III.2: Let $\alpha > 0$, $\frac{1}{2} \leq \rho < 1$ and $1/r = \alpha + 1/2$. Further, let $\mathcal{T}_r(f)$ be a near best BGP(ρ) tree. Then, we have the equivalences

$$C_1 \mathcal{N}_r(f, \mathcal{T}_r(f)) \leq |f|_{\mathcal{GB}^\alpha_r} \leq C_2 \mathcal{N}_r(f, \mathcal{T}_r(f))$$

with constants C_1 and C_2 that depend only on α and ρ.

We can finally conclude that if f is smooth with respect to the Besov measure given in Definition III.1, it suffices to find a near best BGP(ρ) tree to obtain the following m-term approximation rates.

Theorem III.3 (Jackson estimate): Let $\alpha > 0$ and $r > 0$ be related by $1/r = \alpha + 1/2$. Then, for a graph signal $f \in \mathcal{L}(V)$ and geometric wavelets with respect to a near best BGP(ρ) tree $\mathcal{T}_r(f)$, we obtain the m-term approximation error

$$\|f - \mathcal{S}_m(f)\|_{\mathcal{L}^2(V)} \leq C m^{-\alpha} |f|_{\mathcal{GB}^\alpha_r}$$

with a constant $C > 0$ that depends only on r and ρ.

Both, Theorem III.2 and III.3 are discrete versions of respective continuous results given for geometric wavelets on binary

space partitionings [10] and piecewise polynomial approximation on nested triangulations [25]. The proofs are provided in the appendix of this article. We note that all the results of this work hold true also more generally for discrete metric spaces. Within this class, graphs equipped with the graph geodesic distance as a metric are however the most relevant examples for us.

IV. BWP Trees and Graph Wedgelets

After the general theory on BGPs, we are interested in finding explicit constructions of adaptive partitioning trees on graphs. For an efficient coding of the BGP trees we require simple splitting strategies on graphs that can be adapted to the geometry of a graph signal. For this, we propose as building elements the following elementary wedge splits.

Definition IV.1: We call a dyadic partition $\{V,' V''\}$ of the vertex set V a *wedge split* of V if there exist two distinct nodes v' and v'' of V such that V' and V'' have the form

$$V' = \{\mathrm{v} \in V \mid \mathrm{d}(\mathrm{v},\mathrm{v}') \leq \mathrm{d}(\mathrm{v},\mathrm{v}'')\}, \quad \text{and}$$
$$V'' = \{\mathrm{v} \in V \mid \mathrm{d}(\mathrm{v},\mathrm{v}') > \mathrm{d}(\mathrm{v},\mathrm{v}'')\}.$$

A key advantage of the just defined wedge splits is that they can be encoded very compactly in terms of the two nodes v' in V' and v'' in V''. They have the following basic properties:

Property IV.2: If $\{V,' V''\}$ is a wedge split of V, then
1) V' and V'' are uniquely determined by v' and v''.
2) $V' \cap V'' = \emptyset$ and $V' \cup V'' = V$.
3) If the vertex set V is connected, then also V' and V'' are connected subsets of the graph G.

While the first two properties follow immediately from the Definition IV.1, the third property is a consequence of the fact that d is the shortest path distance on G.

Using wedge splits we can define elementary wedgelets $\omega^{+}_{(\mathrm{v},'\mathrm{v}'')}$ and $\omega^{-}_{(\mathrm{v},'\mathrm{v}'')}$ as the following indicator functions:

$$\omega^{+}_{(\mathrm{v},'\mathrm{v}'')}(\mathrm{v}) = \chi_{V'}(\mathrm{v}) = \begin{cases} 1, & \text{if } \mathrm{d}(\mathrm{v},\mathrm{v}') \leq \mathrm{d}(\mathrm{v},\mathrm{v}''), \\ 0, & \text{otherwise,} \end{cases}$$

$$\omega^{-}_{(\mathrm{v},'\mathrm{v}'')}(\mathrm{v}) = \chi_{V''}(\mathrm{v}) = \begin{cases} 1, & \text{if } \mathrm{d}(\mathrm{v},\mathrm{v}') > \mathrm{d}(\mathrm{v},\mathrm{v}''), \\ 0, & \text{otherwise.} \end{cases}$$

Using wedge splits, we generate now the following BGP trees:

Definition IV.3: A binary wedge partitioning (BWP) tree $\mathcal{T}_Q$ of G with respect to the ordered set $Q = \{\mathrm{q}_1, \ldots, \mathrm{q}_M\} \subset V$ is a BGP tree constructed recursively as follows:
1) The root of $\mathcal{T}_Q$ is the set V. It forms the trivial partition $\mathcal{P}^{(1)} = \{V^{(1)}_{\mathrm{q}_1}\} = \{V\}$ and is associated to $\mathrm{q}_1 \in Q$.
2) For a partition $\mathcal{P}^{(m)} = \{V^{(m)}_{\mathrm{q}_1}, \ldots, V^{(m)}_{\mathrm{q}_m}\}$ of V in $\mathcal{T}_Q$ associated to $\mathrm{q}_i \in V^{(m)}_{\mathrm{q}_i}$, $i \in \{1, \ldots, m\}$, $m < M$, consider the node $\mathrm{q}_{m+1} \in V^{(m)}_{\mathrm{q}_j}$ for a $j \in \{1, \ldots, m\}$. We split $V^{(m)}_{\mathrm{q}_j}$ by a wedge split based on q_j and q_{m+1} into two disjoint sets $V^{(m)+}_{(\mathrm{q}_j,\mathrm{q}_{m+1})}$ (containing q_j) and $V^{(m)-}_{(\mathrm{q}_j,\mathrm{q}_{m+1})}$ (containing q_{m+1}) and obtain the new partition
$$\mathcal{P}^{(m+1)} = \{V^{(m+1)}_{\mathrm{q}_1}, \ldots, V^{(m+1)}_{\mathrm{q}_{m+1}}\}$$
with $V^{(m+1)}_{\mathrm{q}_i} = V^{(m)}_{\mathrm{q}_i}$ if $i \neq \{j, m+1\}$, $V^{(m+1)}_{\mathrm{q}_j} = V^{(m)+}_{(\mathrm{q}_j,\mathrm{q}_{m+1})}$ and $V^{(m+1)}_{\mathrm{q}_{m+1}} = V^{(m)-}_{(\mathrm{q}_j,\mathrm{q}_{m+1})}$.

A BWP tree $\mathcal{T}_Q$ as given in Definition IV.3 is uniquely determined by the ordered node set Q. This allows to store $\mathcal{T}_Q$ compactly in terms of the M nodes of Q. Further, adaptive refinements of a BWP tree can be coded directly in terms of the centers Q, i.e., in every refinement step an adaptive selection of a node $\mathrm{q}_j \in Q$ and the generation of new node q_{m+1} is necessary. We collect some properties of BWP trees that follow directly from Definition IV.3, Theorem II.3 and the general Definition II.1 for BGPs.

Proposition IV.4: Let $\mathcal{T}_Q$ be a BWP tree determined by the ordered node set $Q = \{\mathrm{q}_1, \ldots, \mathrm{q}_M\}$.
1) A BWP tree $\mathcal{T}_Q$ contains $2\,M - 1$ elements: 1 root and $2\,M - 2$ children.
2) The M leaves of the binary tree $\mathcal{T}_Q$ are given by the elements of the M-th. partition $\mathcal{P}^{(M)} = \{V^{(M)}_{\mathrm{q}_1}, \ldots, V^{(M)}_{\mathrm{q}_M}\}$.
3) A BWP tree $\mathcal{T}_Q$ is complete if and only if $|Q| = |V|$.
4) A BWP tree $\mathcal{T}_Q$ is balanced with $\frac{1}{2} \leq \rho \leq \frac{n-1}{n}$.
5) The characteristic function of the subset $V^{(m)}_{\mathrm{q}_i}$ can be written as a product of m elementary wedgelets $\omega^{\pm}_{(\mathrm{q}_i,\mathrm{q}_j)}$, with $\mathrm{q}_i, \mathrm{q}_j \in \{\mathrm{q}_1, \ldots, \mathrm{q}_m\}$, $i < j$.

Definition IV.5: The characteristic functions

$$\omega^{(m)}_{\mathrm{q}_i}(\mathrm{v}) = \chi_{V^{(m)}_{\mathrm{q}_i}}(\mathrm{v}), \quad 1 \leq i \leq m,\ 1 \leq m \leq M,$$

of the sets $V^{(m)}_{\mathrm{q}_i}$ will be referred to as *wedgelets* with respect to the BWP tree $\mathcal{T}_Q$. The wedgelets $\left\{\omega^{(m)}_{\mathrm{q}_i} \,:\, 1 \leq i \leq m\right\}$ form an orthogonal basis for the piecewise constant functions on the partition $\mathcal{P}^{(m)}$ (using the inner product (2) in $\mathcal{L}^2(V)$).

Remark IV.6: It is possible to define larger families of BWP trees than in Definition IV.3 (giving larger sets of wedgelets in Definition IV.5, respectively) by allowing wedge splits inside the subsets that are not linked to the centers q_i. In this case, the tree can however not be represented with a simple node set Q. It leads also to a larger computational cost in the selection of a proper wedge split when calculating the adaptive tree.

V. Adaptive Greedy Generation of BWP Trees

To generate BWP trees $\mathcal{T}_Q$ that are adapted to a given graph signal f, every refinement step requires two principal pieces of information. At each partition level m one of the sets $V^{(m)}_{\mathrm{q}_j}$, $j \in \{1, \ldots, m\}$, has to be selected, and a new node $\mathrm{q}_{m+1} \in V^{(m)}_{\mathrm{q}_j}$ is required for the consequent elementary wedge split of $V^{(m)}_{\mathrm{q}_j}$. In general, both choices can be made in an f-adapted or in a non-adaptive way. As adaptive refinement procedures, we consider the following three greedy methods:

Max-distance (MD) greedy wedge splitting: at stage m, the domain $V^{(m)}_{\mathrm{q}_j}$ is chosen f-adaptively by selecting j such that

$$j = \operatorname*{argmax}_{i \in \{1,\ldots,m\}} \|f - \bar{f}_{V^{(m)}_{\mathrm{q}_i}}\|_{\mathcal{L}^2\left(V^{(m)}_{\mathrm{q}_i}\right)}, \tag{5}$$

where

$$\bar{f}_{V^{(m)}_{\mathrm{q}_i}} = \frac{\langle f, \omega^{(m)}_{\mathrm{q}_i}\rangle}{|V^{(m)}_{\mathrm{q}_i}|} = \frac{1}{|V^{(m)}_{\mathrm{q}_i}|} \sum_{\mathrm{v} \in V^{(m)}_{\mathrm{q}_i}} f(\mathrm{v})$$

denotes the mean value of f over the set $V^{(m)}_{\mathrm{q}_i}$. This first selection rule ensures that the chosen set $V^{(m)}_{\mathrm{q}_j}$ is the one with the maximal

$\mathcal{L}^2$-error in (5). As soon as j, or equivalently, q_j are determined, a non-adaptive way to choose the subsequent node set q_{m+1} is by the selection rule

$$\mathrm{q}_{m+1} = \arg\max_{\mathrm{v}\in V_{\mathrm{q}_j}^{(m)}} \mathrm{d}(\mathrm{q}_j, \mathrm{v}),$$

i.e., q_{m+1} is the vertex in $V_{\mathrm{q}_j}^{(m)}$ furthest away from q_j. This choice and the corresponding split can be interpreted as a two center clustering of $V_{\mathrm{q}_j}^{(m)}$ in which the first node q_j is fixed (see a previous work [3] for more details on greedy J-center clustering). One heuristic reason for this selection is that the resulting binary partitions in the BWP tree might be more balanced with a smaller constant ρ compared to the theoretical upper bound $1 - 1/n$ in Proposition IV.4.

Fully-adaptive (FA) greedy wedge splitting: in the FA-greedy procedure the subset to be split is selected according to (5), but also the node q_{m+1} determining the wedge split is chosen according to an adaptive rule. If $\{V_{(\mathrm{q}_j,\mathrm{q})}^{(m)+}, V_{(\mathrm{q}_j,\mathrm{q})}^{(m)-}\}$ denotes the partition of $V_{\mathrm{q}_j}^{(m)}$ for the wedge split determined by q_j and a second node q, we choose q_{m+1} such that

$$\|f - \bar{f}_{V_{(\mathrm{q}_j,\mathrm{q})}^{(m)+}}\|^2_{\mathcal{L}^2\left(V_{(\mathrm{q}_j,\mathrm{q})}^{(m)+}\right)} + \|f - \bar{f}_{V_{(\mathrm{q}_j,\mathrm{q})}^{(m)-}}\|^2_{\mathcal{L}^2\left(V_{(\mathrm{q}_j,\mathrm{q})}^{(m)-}\right)} \tag{6}$$

is minimized over all $\mathrm{q} \in V_{\mathrm{q}_j}^{(m)}$. Compared to the semi-adaptive MD-greedy procedure, the FA-greedy method is computationally more expensive. On the other hand, as the wedge splits are more adapted to the particular form of the underlying function f, we expect a better approximation behavior for the FA-greedy scheme. This expectation will be confirmed in the numerical experiments performed in the last section.

Randomized (R) greedy wedge splitting: If the size of the subsets $V_{\mathrm{q}_j}^{(m)}$ is large it might be too time-consuming to find the global minimum of the quantity (6) in the FA-greedy scheme. A quasi-optimal alternative to the fully-adaptive procedure is a randomized splitting strategy, in which the minimization of (6) is performed on a subset of $1 \le R \le |V_{\mathrm{q}_j}^{(m)}|$ randomly picked nodes of $V_{\mathrm{q}_j}^{(m)}$. In this strategy, the parameter R acts as a control parameter giving a result close or identical to FA-greedy if R is chosen large enough.

The just described adaptive selection rules to generate BWP trees and the respective wedgelet encoding and decoding variants are summarized in Algorithms 1 and 2.

Upper bounds for the computational cost: We assume that the distance between two nodes can be calculated in $\mathcal{O}(1)$ operations. Further, in a worst-case scenario, the calculation of a mean value over $V_{\mathrm{q}_j}^{(m)}$ requires $\mathcal{O}(n)$ operations. The computational expenses of Algorithm 1 can thus be bounded by $\mathcal{O}(Mn^2)$, $\mathcal{O}(MRn)$ and $\mathcal{O}(Mn)$ operations for FA-greedy, R-greedy and MD-greedy, respectively. In Fig. 8(e), the computational times required by the three BWP variants are plotted for an image decomposition. The measured times indicate that the dependence of the cost on M is rather sublinear than linear.

Acceleration possibilities: For very large n, FA-greedy and R-greedy (for large R) might be too expensive. In this case, a possibility to increase the calculational speed is to split the graph a priori into J subgraphs. Then, the adaptive BWP methods can be applied (also in a parallelized form) separately to each subgraph. For this procedure, clustering algorithms as J-center clustering [18] are available. In [3], this clustering method has been used for partition of unity methods on graphs.

Algorithm 1. Wedgelet Encoding of a Graph Signal.

Input: Graph signal f, initial node $\mathrm{q}_1 \in V$, first partition $\mathcal{P}^{(1)} = \{V\} = \{V_{\mathrm{q}_1}^{(1)}\}$ and final partition size M.

for $m = 2$ *to* M **do**

1) **Greedy selection of subset:** calculate j according to the rule (5) as

$$j = \arg\max_{i\in\{1,\dots,m-1\}} \left\|f - \bar{f}_{V_{\mathrm{q}_i}^{(m-1)}}\right\|_{\mathcal{L}^2(V_{\mathrm{q}_i}^{(m-1)})};$$

2) Conduct one of the following alternatives:
Max-distance (MD) greedy procedure: select new node $\mathrm{q}_m = \arg\max_{\mathrm{v}\in V_{\mathrm{q}_j}^{(m)}} \mathrm{d}(\mathrm{q}_j, \mathrm{v})$ farthest away from q_j and add it to the node set Q;
Fully-adaptive (FA) greedy procedure: determine new node q_m such that the squared $\mathcal{L}^2$-error term (6) is minimized and add it to the node set Q;
Randomized (R) greedy procedure: determine q_m such that (6) is minimized over a subset of R randomly chosen points and add it to Q;

3) According to Definition IV.3, generate the **new partition** $\mathcal{P}^{(m)}$ from the partition $\mathcal{P}^{(m-1)}$ by a wedge split of the subset $V_{\mathrm{q}_j}^{(m-1)}$ into the children sets $V_{(\mathrm{q}_j,\mathrm{q}_m)}^{(m-1)+}$ and $V_{(\mathrm{q}_j,\mathrm{q}_m)}^{(m-1)-}$;

4) Compute **mean values** $\bar{f}_{V_{\mathrm{q}_i}^{(m)}}$, $i \in \{1,\dots,m\}$, for the new partition $\mathcal{P}^{(m)}$ by an update from $\mathcal{P}^{(m-1)}$.

Output: $Q = \{\mathrm{q}_1,\dots,\mathrm{q}_M\}$, $\{\bar{f}_{V_{\mathrm{q}_1}^{(M)}},\dots,\bar{f}_{V_{\mathrm{q}_M}^{(M)}}\}$.

Algorithm 2: Wedgelet Decoding of a Graph Signal.

Input: $Q = \{\mathrm{q}_1,\dots,\mathrm{q}_M\}$, $\{\bar{f}_{V_{\mathrm{q}_1}^{(M)}},\dots,\bar{f}_{V_{\mathrm{q}_M}^{(M)}}\}$.

Calculate the partition $\mathcal{P}^{(M)} = \{V_{\mathrm{q}_1}^{(M)},\dots,V_{\mathrm{q}_M}^{(M)}\}$ of V by elementary wedge splits along the BWP tree $\mathcal{T}_Q$ according to the recursive procedure in Definition IV.3.

Output: The wedgelet approximation

$$\mathcal{W}_M f(\mathrm{v}) = \sum_{i=1}^{M} \bar{f}_{V_{\mathrm{q}_i}^{(M)}}\, \omega_{\mathrm{q}_i}^{(M)}(\mathrm{v})$$

of f. For $M = n$, $\mathcal{W}_n f = f$ is reconstructed.

A. Geometric Wavelets Based on Wedge Splits

Instead of storing the mean values $\{\bar{f}_{V_{\mathrm{q}_1}^{(M)}},\dots,\bar{f}_{V_{\mathrm{q}_M}^{(M)}}\}$ of the wedgelet approximation $\mathcal{W}_M f$, we can alternatively encode $\mathcal{W}_M f$ using the geometric wavelets introduced in (4). This alternative representation is particularly suited if a further compression of the signal is desired, for instance by using an m-term approximation of the signal with $m < M$.

For a BWP tree $\mathcal{T}_Q$ and $2 \le m \le M$, we define the index $j = j(m) < m$ such that $V_{\mathrm{q}_j}^{(m)}$ is the sibling of $V_{\mathrm{q}_m}^{(m)}$ in the BWP tree $\mathcal{T}_Q$. Then, the geometric wavelets with respect to the

Algorithm 3: Wedgelet Encoding With BWP Wavelets.

Input: Graph signal f, initial node $\mathrm{q}_1 \in V$, first partition $\mathcal{P}^{(1)} = \{V\} = \{V_{\mathrm{q}_1}^{(1)}\}$ and final partition size M.

for $m = 2$ *to* M **do**

1) 2) & 3) as in Algorithm 1;

4) Compute **geometric wavelet coefficients** $c^{+}_{(\mathrm{q}_j,\mathrm{q}_m)}$ and $c^{-}_{(\mathrm{q}_j,\mathrm{q}_m)}$ (one of the two is sufficient).

Output: $Q = \{\mathrm{q}_1, \ldots, \mathrm{q}_M\}$, $\{c_{\mathrm{q}_1}, c^{\pm}_{(\mathrm{q}_j,\mathrm{q}_2)}, \ldots, c^{\pm}_{(\mathrm{q}_j,\mathrm{q}_M)}\}$.

Algorithm 4: Wedgelet Decoding With BWP Wavelets

Input: $Q = \{\mathrm{q}_1, \ldots, \mathrm{q}_M\}$, $\{c_{\mathrm{q}_1}, c^{\pm}_{(\mathrm{q}_j,\mathrm{q}_2)}, \ldots, c^{\pm}_{(\mathrm{q}_j,\mathrm{q}_M)}\}$.

$\mathcal{W}_1 f = c_{\mathrm{q}_1}$.

for $m = 2$ *to* M **do**

Calculate the partition $\mathcal{P}^{(m)} = \{V_{\mathrm{q}_1}^{(m)}, \ldots, V_{\mathrm{q}_m}^{(m)}\}$ of V from the partition $\mathcal{P}^{(m-1)}$ by an elementary wedge split of the set $V_{\mathrm{q}_j}^{(m-1)}$. **Update** the wedgelet approximation

$$\mathcal{W}_m f = \mathcal{W}_{m-1} f + c^{+}_{(\mathrm{q}_j,\mathrm{q}_m)} \omega_{\mathrm{q}_j}^{(m)} + c^{-}_{(\mathrm{q}_j,\mathrm{q}_m)} \omega_{\mathrm{q}_m}^{(m)}.$$

Output: The wedgelet approximation

$$\mathcal{W}_M f = c_{\mathrm{q}_1} + \sum_{m=2}^{M} \left(c^{+}_{(\mathrm{q}_j,\mathrm{q}_m)} \omega_{\mathrm{q}_j}^{(m)} + c^{-}_{(\mathrm{q}_j,\mathrm{q}_m)} \omega_{\mathrm{q}_m}^{(m)} \right)$$

of f. For $M = n$, $\mathcal{W}_n f = f$ is reconstructed.

BWP tree $\mathcal{T}_Q$ are defined as the signals

$$\psi^{+}_{(\mathrm{q}_j,\mathrm{q}_m)}(f) = \left(\bar{f}_{V_{\mathrm{q}_j}^{(m)}} - \bar{f}_{V_{\mathrm{q}_j}^{(m-1)}} \right) \omega_{\mathrm{q}_j}^{(m)} =: c^{+}_{(\mathrm{q}_j,\mathrm{q}_m)}(f)\, \omega_{\mathrm{q}_j}^{(m)},$$

$$\psi^{-}_{(\mathrm{q}_j,\mathrm{q}_m)}(f) = \left(\bar{f}_{V_{\mathrm{q}_m}^{(m)}} - \bar{f}_{V_{\mathrm{q}_j}^{(m-1)}} \right) \omega_{\mathrm{q}_m}^{(m)} =: c^{-}_{(\mathrm{q}_j,\mathrm{q}_m)}(f)\, \omega_{\mathrm{q}_m}^{(m)}.$$

Further, in the particular case $m = 1$, we set

$$\psi_{\mathrm{q}_1}(f) = \bar{f}_V\, \omega_{\mathrm{q}_1}^{(1)} =: c_{\mathrm{q}_1}(f)\, \omega_{\mathrm{q}_1}^{(1)}.$$

In this way, we obtain $2M - 1$ geometric wavelets for $\mathcal{T}_Q$. Beside the altered notation, this definition corresponds to the general definition of geometric wavelets for BGP trees given in (4). For $\psi^{+}_{(\mathrm{q}_j,\mathrm{q}_m)}(f)$ and $\psi^{-}_{(\mathrm{q}_j,\mathrm{q}_m)}(f)$, we have the relation

$$\begin{aligned} &\sum_{\mathrm{v} \in V} \left(\psi^{+}_{(\mathrm{q}_j,\mathrm{q}_m)}(f)(\mathrm{v}) + \psi^{-}_{(\mathrm{q}_j,\mathrm{q}_m)}(f)(\mathrm{v}) \right) \\ &\quad = \bar{f}_{V_{\mathrm{q}_m}^{(m)}} |V_{\mathrm{q}_m}^{(m)}| + \bar{f}_{V_{\mathrm{q}_j}^{(m)}} |V_{\mathrm{q}_j}^{(m)}| - \bar{f}_{V_{\mathrm{q}_j}^{(m-1)}} |V_{\mathrm{q}_j}^{(m-1)}| \\ &\quad = \sum_{\mathrm{v} \in V_{\mathrm{q}_m}^{(m)}} f(\mathrm{v}) + \sum_{\mathrm{v} \in V_{\mathrm{q}_j}^{(m)}} f(\mathrm{v}) - \sum_{\mathrm{v} \in V_{\mathrm{q}_j}^{(m-1)}} f(\mathrm{v}) = 0. \end{aligned}$$

This implies that $\langle \psi^{+}_{(\mathrm{q}_j,\mathrm{q}_m)}(f), 1 \rangle = -\langle \psi^{-}_{(\mathrm{q}_j,\mathrm{q}_m)}(f), 1 \rangle$ and that $c^{+}_{(\mathrm{q}_j,\mathrm{q}_m)}(f)$ and $c^{-}_{(\mathrm{q}_j,\mathrm{q}_m)}(f)$ are related by

$$c^{+}_{(\mathrm{q}_j,\mathrm{q}_m)}(f) = -\tfrac{|V_{\mathrm{q}_m}^{(m)}|}{|V_{\mathrm{q}_j}^{(m)}|} c^{-}_{(\mathrm{q}_j,\mathrm{q}_m)}(f).$$

In particular, when computing the coefficients of the geometric wavelets, it suffices to store one of the two, either $c^{+}_{(\mathrm{q}_j,\mathrm{q}_m)}(f)$ or $c^{-}_{(\mathrm{q}_j,\mathrm{q}_m)}(f)$. The encoding and decoding of a graph signal f in terms of wedge-based geometric wavelets is compactly described in Algorithms 3 and 4.

VI. Memory Requirements for Wedgelet Encoding

To encode a graph signal in terms of M wedgelets, we need to store the mean values $\{\bar{f}_{V_{\mathrm{q}_1}^{(M)}}, \ldots, \bar{f}_{V_{\mathrm{q}_M}^{(M)}}\}$ on the leaves of the BWP tree as well as the geometric information of the wedge splits provided by the set $Q = \{\mathrm{q}_1, \ldots, \mathrm{q}_M\}$. Using a quantization of K different values, every mean value $\bar{f}_{V_{\mathrm{q}_i}^{(M)}}$ can be stored with $\log_2(K)$ bits. The same holds true if the BWP wavelet coefficients $c^{+}_{(\mathrm{q}_j,\mathrm{q}_m)}$ replace the mean values. As we have n nodes, we can further store any node v_j using its index $1 \leq j \leq n$. This requires at most $\log_2(n)M$ bits to store the node set Q. Consequently, we get the following upper bound for the memory requirements of wedgelet encoding.

Theorem VI.1: Assume that the mean values $\bar{f}_{V_{\mathrm{q}_i}^{(M)}}$ (or the coefficients c_{q_1} and $c^{+}_{(\mathrm{q}_j,\mathrm{q}_i)}$) are given in a quantized form with at most K different values. Then, the wedgelet encodings in Algorithms 1 and 3 require a memory of at most

$$\frac{\lceil \log_2(n) + \log_2(K) \rceil M}{n} \quad \text{bits per node.}$$

In the particular case of an image with $512 \times 512 = 2^{18}$ pixels and an image depth of $K = 2^8 = 256$ colors we get by Theorem VI.1 that a representation with $M = 1000$ wedgelets requires a memory of less than 0.1 bits per pixel.

VII. Examples on Graphs and Images

A. Examples of BWPs on Graphs

We will consider two test graphs, the Minnesota road network G_1 and an Erdős–Rényi graph G_2. The dataset for G_1 has been retrieved from [33] and consists of $n = 2642$ vertices and 3304 edges. The graph G_2 with $n = 100$ vertices and 236 edges was generated according to the $G(n, p)$ Erdős–Rényi random graph model with probability $p = 0.05$. The distance metric on both graphs is the shortest-path distance. As test signals on G_1 we consider the two binary functions

$$f_1 = 2\chi_{V_1} - 1, \quad f_2 = 2\chi_{V_2} - 1,$$

based on the characteristic functions of the node sets

$$V_1 = \{\mathrm{v} \in V \mid x_{\mathrm{v}} < -94\},$$

$$V_2 = \{\mathrm{v} \in V \mid 0.75(x_{\mathrm{v}} + 93.3)^2 + (y_{\mathrm{v}} - 44.95)^2 < 0.35\},$$

where $(x_{\mathrm{v}}, y_{\mathrm{v}})$ denote the Cartesian coordinates of the node $\mathrm{v} \in V$. The test signal f_3 on G_2 describes a clustering of G_2 based on three integer values from 1 to 3. The functions f_1, f_2 and f_3 are illustrated in Fig. 2.

Starting from a random node q_1 we use Algorithm 1 to generate the BWP tree as well as a piecewise constant approximation of the functions. The initial part of the BWP tree for the

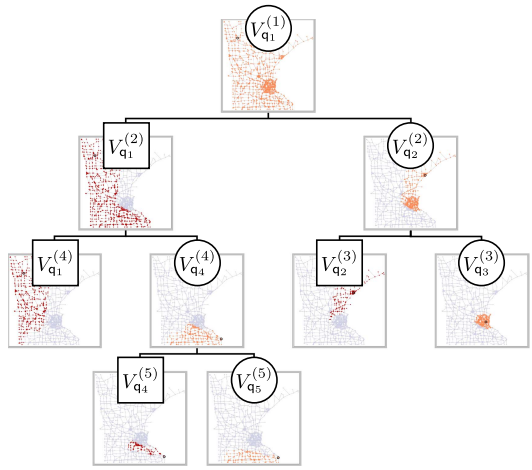

Fig. 1. A BWP tree for the adaptive approximation of the test function f_2 in Fig. 2 (middle) on the Minnesota graph.

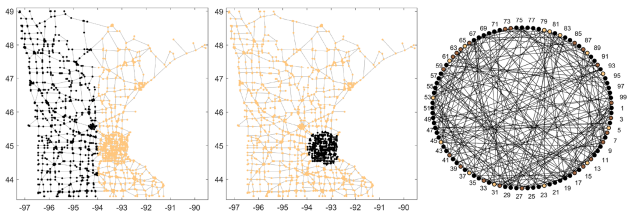

Fig. 2. Test signals f_1 (left), f_2 (middle) and f_3 (right) to generate BWPs.

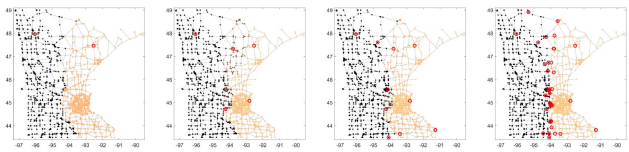

Fig. 3. Approximation of f_1 with 1,4,9 and 39 wedge splits (from left to right). The red rings indicate the center nodes Q. The number of wrongly classified nodes equals 356, 286, 110, and 12, respectively.

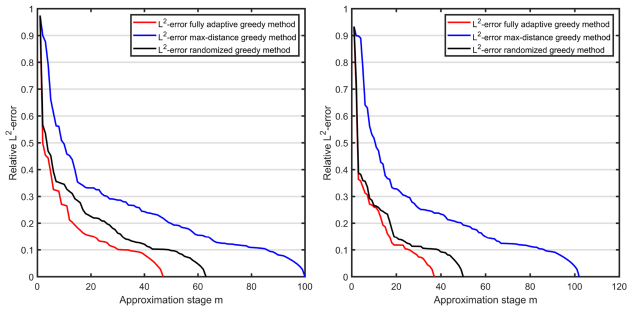

Fig. 4. L^2-approximation errors for the BWP approximation $\mathcal{W}_m f$ of f_1 (left) and f_2 (right) using FA-greedy, MD-greedy and R-greedy with $R = 50$.

approximation of f_2 (using FA-greedy) is shown in Fig. 1. In Fig. 3, approximations $\mathcal{W}_m f_1$ of f_1 are illustrated for different partitioning stages m (the number of wedge splits equals $m-1$, we used FA-greedy to generate the BWP tree).

The three variants MD-greedy, R-greedy and FA-greedy for the test functions f_1 and f_2 are compared in Fig. 4. In these examples, FA-greedy performs best for most partitioning stages m,

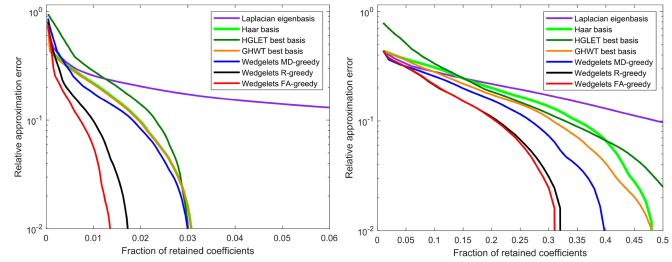

Fig. 5. Comparison between best m-term approximation using BWP wavelets (FA-greedy, R-greedy with $R = 50$ and MD-greedy) and non-adaptive Haar-type wavelet dictionaries for the test functions f_2 (left) and f_3 (right).

followed by R-greedy and MD-greedy. The FA-greedy variant is also the most cost-intensive of the three. Our tests showed that with $R = 100$ randomly chosen nodes, the outcome of R-greedy is already very similar to FA-greedy.

In a second test, we compare the best m-term approximation of the adaptive BWP wavelets with four non-adaptive basis systems from the literature: a non-adaptive Haar basis, a Laplacian eigenbasis, a hierarchical graph Laplacian eigensystem (HGLET, [21]) and a generalized Haar-Walsh basis system (GHWT, [22], [23]). The latter two are Haar-type dictionaries that contain, beside Haar functions, also Laplacian eigenfunctions on subgraphs (HGLETs) and generalized Walsh basis functions (GHWTs). For the comparison with the four non-adaptive basis systems we used the MTSG toolbox developed for the works [21], [22], [23]. We can see that the data-adaptive FA-greedy and R-greedy schemes are able to generate sparser representations of the test functions in which a considerably smaller amount of the wavelet coefficients is required compared to the non-adaptive basis systems. While this higher sparseness is an aimed-at desirable property for data compression, adaptive basis systems require additional memory to store the respective geometric information. Therefore, when using adaptive BWPs for the compression of graph data also the additional storage costs of the BWPs as estimated in Section VI have to be taken into account.

In a further test, we compare the same basis systems on the Minnesota graph for the approximation of piecewise smooth test functions. For this, we modify the piecewise constant signal f_1 by adding a gradient term, i.e., we consider the signal

$$f_4(\mathrm{v}) = f_1(\mathrm{v}) + \alpha(x_\mathrm{v} - \bar{x}_V), \quad \alpha \in \mathbb{R},$$

where $\bar{x}_V$ denotes the mean value of x_v over V. As the parameter α increases, the function f_4 gets dominated by the gradient term and a dictionary of piecewise constant signals will struggle to approximate f_4 in a sparse way. This is confirmed by our simulation shown in Fig. 6. As α gets larger, the best m-term approximation errors using a pure Haar basis system (adaptive or non-adaptive) deteriorate. The difference between adaptive and non-adaptive basis systems is however still significant. It is also visible that basis systems that are better adapted to piecewise smooth functions as the HGLETs display an improved performance.

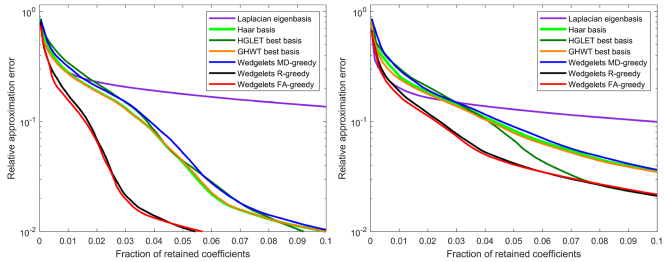

Fig. 6. Comparison of best m-term approximations using Haar-type wavelet dictionaries for the piecewise smooth test function f_4 and gradient parameter $\alpha = 0.1$ (left) and $\alpha = 0.5$ (right).

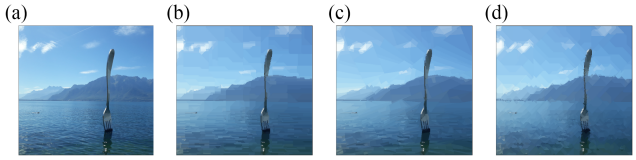

Fig. 7. Role of the metric in BWP (a) original image with 500×451 pixels; (b)(c)(d) R-greedy compression with $M = 1000$ nodes, $R = 500$ as well as (b) the 1-norm, (c) the 2-norm and (d) the infimum norm for the pixel distance.

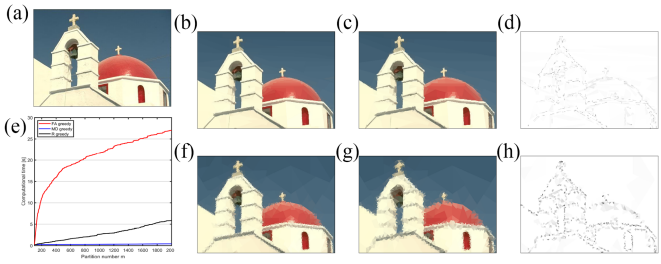

Fig. 8. BWP image compression. (a) original 481×321-image; (b)(c) FA-greedy BWP compression for $M = 2000$, $M = 1000$; (d) wavelet details between (b) and (c); (e) Computational times of the BWP variants; (f)(g) MD-greedy BWP compression for $M = 2000$, $M = 1000$; (h) wavelet details between (f) and (g).

B. BWPs for the Compression of 2D Images

As adaptive partitioning tools for discrete domains, BWPs can also be used for the piecewise approximation and compression of images. An image can be naturally thought of as a finite rectangular grid of pixels and interpreted as a graph. Pixels close to each other are therein linked by a weighted edge. The structure and the weights of the single edges determine the local dependencies in the image and have therefore a strong influence on the outcome of the greedy algorithms.

A simple qualitative comparison of the role of the used metric is given in Fig. 7. The 1-norm and the infimum norm for the distance of the pixels lead to partitions with a rather rectangular or rhomboid wedge structure. On the other hand, wedges generated by the 2-norm seem to be more anisotropic and slightly better adapted to the edges of the image.

We next compare the performance of the FA-greedy and the MD-greedy method for the compression of images. In the example given in Fig. 8, we see that, as expected, the FA-greedy performs considerably better. In particular, the wavelet details in the FA-greedy scheme are smaller and more distinguished than

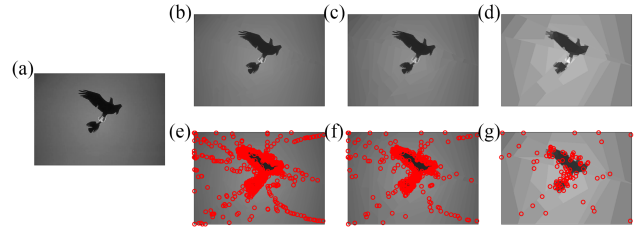

Fig. 9. BWP image encoding. (a) original with 481×321 pixels; (b)(c)(d) R-greedy compression with 1000, 500 and 100 nodes, $R = 500$; (e)(f)(g) respective node distributions for the approximations in (b)(c)(d). The corresponding PSNR values are (b) 40.762 dB, (c) 37.935 dB, and (d) 31.827 dB.

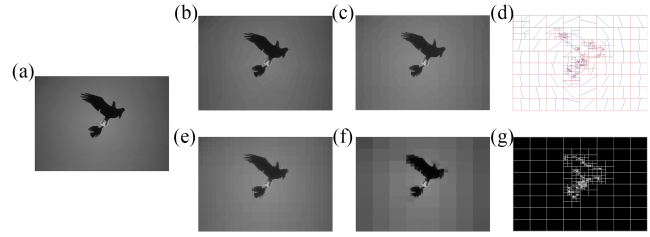

Fig. 10. Comparison of 4 image compression techniques based on piecewise approximation: (a) original 481×321 image; (b) graph wedgelet compression using 500 most relevant BWP wavelet coefficients (PSNR: 38.297 dB) (c)(d) continuous wedgelet compression using 506 wedges (PSNR: 36.828 dB) (e) Haar wavelet compression using 500 most relevant coefficients (PSNR: 34.764 dB) (f)(g) quadtree compression with 505 blocks (PSNR: 31.662 dB).

for MD-greedy when using the same number of wedge splits. Regarding the distribution of the center nodes Q, we see further in Fig. 9 that the adaptive BWP scheme (in this case a R-greedy scheme) selects the new nodes increasingly closer to the edges of the image such that most refinements of the partitions are performed in those regions where the gradients are large.

Finally, in Fig. 10, we compare the approximation quality of our adaptive BWP algorithm with three well-known segmentation based compression schemes from the literature. The first is a classical bivariate Haar wavelet transform in which the image is decomposed in uniform dyadic blocks and 6 hierarchical levels. The most relevant wavelet coefficients of the image are then selected according to the Birgé-Massart strategy [1]. In Fig. 10(e) the respective compression for 500 coefficients is illustrated. We compare this with a compression using the most relevant geometric wavelets in a R-greedy BWP tree ($R = 500$). Generating the BWP tree for $M = 4000$ and selecting the 500 most relevant coefficient pairs $c^{\pm}_{(\mathrm{q}_j,\mathrm{q}_i)}$ provided by Algorithm 3, we obtain the compressed image in Fig. 10(b). Qualitatively, the contours of the image are more pronounced for the adapted BWP wavelets, while in the classical Haar wavelet approach block artifacts are visible. A further indication for the higher image quality of BWP compression is the larger peak signal to noise ratio (PSNR).

The two other compression methods are a continuous wedgelet decomposition (we use the implementation of [16]) and a quadtree decomposition [35] into adaptively generated dyadic blocks. The resulting image approximations shown in Fig. 10(c)(d)(f)(g) (using 506 and 505 segments, respectively) are compared with the graph wedgelet approximation in

Fig. 9(c)(f) (using 500 wedgelets). It is visible that the number of quadtree blocks and continuous wedges is still quite low for a good resolution of the original image. This is also indicated by the lower PSNR of the resulting approximations. The graph wedgelets on the other hand display a higher adaptivity to the original image with a larger PSNR value. Our numerical tests also indicate that for smaller values of M the generation of the wedgelet approximation alone, without thresholding the small geometric wavelet components, provides already good compression results. This is visible in the comparison between Fig. 9(c) and 10(b), where almost no difference between the two approximations is visible.

VIII. Conclusion

We introduced graph wedgelets: a novel type of geometric Haar-type basis functions on graphs that are able to efficiently capture the geometric information of signals using adaptive binary wedge splits that minimize a local $\mathcal{L}^2$-error. These discrete wedgelets are organized in terms of a binary wedge partitioning tree that can be encoded compactly in terms of a finite ordered sequence of graph nodes. We illustrated that these BWP trees provide promising dictionaries for the sparse representation of graph signals and can be applied to the compression of signals and images. From a theoretical point of view we showed that the geometric wavelets based on a near-best wedgelet partitioning tree provide quickly converging m-term approximants if the approximated signal is in a properly defined Besov-type smoothness class.

Appendix

A. Proof of Theorem II.3

i) As $\mathcal{T}$ is complete and binary, it contains n leaves, and at least two of them are siblings (assuming that $n > 1$, for $n = 1$ the tree contains just a single root). Pruning the tree $\mathcal{T}$ by removing this pair of siblings, we obtain a reduced tree with $n-1$ leaves. An induction argument over the number n of leaves therefore tells us that $|\mathcal{T}| = 2n-1$.

ii) The completeness of $\mathcal{T}$ implies that we can decompose any signal f in terms of the n leaves of $\mathcal{T}$ as

$$f = \sum_{i=1}^{n} f(\mathrm{v}_i)\chi_{\{\mathrm{v}_i\}} = \sum_{W\in\mathcal{P}^{(n)}} \frac{\langle f, \chi_W\rangle}{|W|}\chi_W.$$

Now, by the definition of the wavelets $\psi_W(f)$ in (4), we recursively obtain

$$\begin{aligned} f &= \sum_{W\in\mathcal{P}^{(n)}\setminus\mathcal{P}^{(n-1)}} \psi_W(f) + \sum_{W\in\mathcal{P}^{(n-1)}} \frac{\langle f, \chi_W\rangle}{|W|}\chi_W \\ &= \sum_{k=2}^{n} \sum_{W\in\mathcal{P}^{(k)}\setminus\mathcal{P}^{(k-1)}} \psi_W(f) \;+\; \psi_V(f) = \sum_{W\in\mathcal{T}} \psi_W(f). \end{aligned}$$

Thus, using the ordered elements $W_i \in \mathcal{T}$, $i \in \{1,\dots,2n-1\}$ (according to the $\mathcal{L}^2$-norm of the wavelets), we obtain the representation in (ii).

iii) In the proof of Theorem III.3, we will show that for every BGP(ρ) tree $\mathcal{T}$ the inequality

$$\|f - \mathcal{S}_m(f)\|_{\mathcal{L}^2(V)} \le Cm^{-\alpha}\mathcal{N}_r(f,\mathcal{T})$$

holds true with a constant $C > 0$ that depends only on ρ (Theorem III.3 is actually formulated in terms of near-best BGP(ρ) trees, the first part of the proof works however for all BGP(ρ) trees). This implies that

$$\begin{aligned} \|f\|_{\mathcal{L}^2(V)} &\le \|f - \mathcal{S}_1(f)\|_{\mathcal{L}^2(V)} + \|\mathcal{S}_1(f)\|_{\mathcal{L}^2(V)} \\ &\le (C+1)\mathcal{N}_r(f,\mathcal{T}), \end{aligned}$$

and, thus, the statement of the theorem.

B. Proof of Theorem III.2.

We adapt the proof of [10, Theorem 3.5], in which a similar equivalence has been shown for binary space partitions in a continuous setting. To obtain the result of Theorem III.2 for near best BGP(ρ) trees, it suffices to show that for every BGP(ρ) tree $\mathcal{T}$ the following equivalence holds:

$$C_1\mathcal{N}_r(f,\mathcal{T}) \le \mathcal{M}_r(f,\mathcal{T}) \le C_2\mathcal{N}_r(f,\mathcal{T}), \tag{7}$$

where

$$\mathcal{M}_r(f,\mathcal{T}) = \left(\sum_{W\in\mathcal{T}} |W|^{-\alpha r} \sup_{\mathrm{w}\in W} \sum_{\mathrm{v}\in W} |f(\mathrm{v}) - f(\mathrm{w})|^r\right)^{\frac{1}{r}},$$

and $C_1, C_2 > 0$ are constants that depend only on ρ and r. We start with the first inequality in (7). If W' is a child of W in $\mathcal{T}$, we can estimate the r-norm $\|\psi_{W'}(f)\|_{\mathcal{L}^r(V)}$ as

$$\begin{aligned} \|\psi_{W'}(f)\|_{\mathcal{L}^r(V)} &= \left\|\left(\tfrac{\langle f,\chi_{W'}\rangle}{|W'|} - \tfrac{\langle f,\chi_W\rangle}{|W|}\right)\chi_{W'}(\mathrm{v})\right\|_{\mathcal{L}^r(V)} \\ &\le C\left(\left\|(\bar{f}_{W'} - f)\chi_{W'}\right\|_{\mathcal{L}^r(V)} + \left\|(\bar{f}_W - f)\chi_{W'}\right\|_{\mathcal{L}^r(V)}\right) \\ &= C\left(\left\|\bar{f}_{W'} - f\right\|_{\mathcal{L}^r(W')} + \left\|\bar{f}_W - f\right\|_{\mathcal{L}^r(W')}\right), \end{aligned} \tag{8}$$

where $\bar{f}_{W'}$ and $\bar{f}_W$ denote the means of f over the sets W' and W, respectively. The constant C in this estimate depends only on r. The estimate (8) allows us to proceed as follows and to obtain the first of the two inequalities in (7):

$$\begin{aligned} \mathcal{N}_r(f,\mathcal{T}) &= \left(\sum_{W\in\mathcal{T}} \|\psi_W(f)\|^r_{\mathcal{L}^2(V)}\right)^{\frac{1}{r}} \\ &\le \left(\sum_{W\in\mathcal{T}} |W|^{r/2-1}\|\psi_W(f)\|^r_{\mathcal{L}^r(V)}\right)^{\frac{1}{r}} \\ &\le C'\left(\sum_{W\in\mathcal{T}} |W|^{r/2-1}\|\bar{f}_W - f\|^r_{\mathcal{L}^r(W)}\right)^{\frac{1}{r}} \\ &\le C''\left(\sum_{W\in\mathcal{T}} |W|^{r/2-2}\sum_{\mathrm{v}\in W}\|f(\mathrm{v}) - f\|^r_{\mathcal{L}^r(W)}\right)^{\frac{1}{r}} \\ &\le C''\left(\sum_{W\in\mathcal{T}} |W|^{-\alpha r}\sup_{\mathrm{w}\in W}\sum_{\mathrm{v}\in W}|f(\mathrm{v}) - f(\mathrm{w})|^r\right)^{\frac{1}{r}} \\ &= C''\mathcal{M}_r(f,\mathcal{T}). \end{aligned}$$

Here, for the first inequality we used a general form of Hölder's inequality. For the second inequality, we combined the bound in (8) with the fact that a BGP(ρ) tree is balanced resulting in a constant C' that depends on $0 < \rho < 1$ and on r. Also the subsequent constant C'' depends only on ρ and r. In the last inequality, the relation $1/r = \alpha + 1/2$ was included.

We consider now the second inequality in (7). Based on the expansion $f = \sum_{W\in\mathcal{T}} \psi_W(f)$, we get

$$\begin{aligned}
&\sup_{\mathrm{w}\in W} \sum_{\mathrm{v}\in W} |f(\mathrm{v}) - f(\mathrm{w})|^r \\
&\quad = \sup_{\mathrm{w}\in W} \sum_{\mathrm{v}\in W} |\sum_{W'\in\mathcal{T}, W'\subset W} \psi_{W'}(f)(\mathrm{v}) - \psi_{W'}(f)(\mathrm{w})|^r \\
&\quad \le C \sum_{\mathrm{v}\in W} \sum_{W'\in\mathcal{T}, W'\subset W} |\psi_{W'}(f)(\mathrm{v})|^r \\
&\quad = C \sum_{W'\in\mathcal{T}, W'\subset W} \|\psi_{W'}(f)\|^r_{\mathcal{L}^r(W')},
\end{aligned}$$

with a constant C that depends only on r. This provides for $\mathcal{M}_r(f,\mathcal{T})$ the bound

$$\begin{aligned}
\mathcal{M}_r(f,\mathcal{T})^r &\le C \sum_{W\in\mathcal{T}} |W|^{-\alpha r} \sum_{W'\in\mathcal{T}, W'\subset W} \|\psi_{W'}(f)\|^r_{\mathcal{L}^r(W')} \\
&= C \sum_{W\in\mathcal{T}} \sum_{W'\subset W} \left(\frac{|W'|}{|W|}\right)^{\alpha r} |W'|^{-\alpha r} \|\psi_{W'}(f)\|^r_{\mathcal{L}^r(W')} \\
&= C \sum_{W'\in\mathcal{T}} |W'|^{-\alpha r} \|\psi_{W'}(f)\|^r_{\mathcal{L}^r(W')} \sum_{W\supset W'} \left(\frac{|W'|}{|W|}\right)^{\alpha r}.
\end{aligned}$$

The balancedness of $\mathcal{T}$ implies that $|W'| \le \rho|W|$ for every parent W of W', and thus $|W'| \le \rho^k |W^{(k)}|$ for every parent $W^{(k)}$ in the k-th generation before W'. Thus,

$$\sum_{W\supset W'} \left(\frac{|W'|}{|W|}\right)^{\alpha r} \le \sum_{k=1}^{\infty} \rho^{k\alpha r} \le \frac{\rho^{\alpha r}}{1-\rho^{\alpha r}}, \tag{9}$$

and we can conclude that

$$\begin{aligned}
\mathcal{M}_r(f,\mathcal{T})^r &\le \frac{C\rho^{\alpha r}}{1-\rho^{\alpha r}} \sum_{W'\in\mathcal{T}} |W'|^{-\alpha r} \|\psi_{W'}(f)\|^r_{\mathcal{L}^r(W')} \\
&\le C' \sum_{W'\in\mathcal{T}} \|\psi_{W'}(f)\|^r_{\mathcal{L}^2(W')} = C'\mathcal{N}_r(f,\mathcal{T})^r,
\end{aligned}$$

with a constant C' that depends only on ρ and r. This gives the second inequality in (7).

C. Proof of Theorem III.3

The subsequent proof is an adaption of the techniques developed for the derivation of [25, Theorem 3.4] in which a Jackson estimate for piecewise polynomial approximation on nested triangulations in $\mathbb{R}^2$ is formulated. The same techniques have been used in [10, Theorem 3.6] to derive a respective Jackson estimate for geometric wavelets with respect to binary space partitionings of convex domains.

To subdivide the geometric wavelets of the tree $\mathcal{T}_r(f)$ into dyadic blocks, we define for $\mu \in \mathbb{Z}$ the index sets

$$I_\mu := \left\{ i \;\middle|\; \tfrac{\mathcal{N}_r(f,\mathcal{T}_r(f))}{2^\mu} \le \|\psi_{W_i}(f)\|_{\mathcal{L}^2(V)} < \tfrac{\mathcal{N}_r(f,\mathcal{T}_r(f))}{2^{\mu-1}} \right\}.$$

For the union of these index sets, we get

$$\bigcup_{\nu\le\mu} I_\nu = \left\{ i \mid \|\psi_{W_i}(f)\|_{\mathcal{L}^2(V)} \ge 2^{-\mu}\mathcal{N}_r(f,\mathcal{T}_r(f)) \right\}.$$

Thus, by Definition II.2 of the r-energy $\mathcal{N}_r(f,\mathcal{T}_r(f))$, we have

$$|I_\mu| \le \sum_{\nu\le\mu} |I_\nu| = |\bigcup_{\nu\le\mu} I_\nu| \le 2^{\mu r}.$$

Now, setting $m := \sum_{\nu\le\mu} |I_\nu|$, we obtain the estimate

$$\begin{aligned}
\|f - \mathcal{S}_m(f)\|_{\mathcal{L}^2(V)} &\le \left\| \sum_{\nu>\mu} \sum_{i\in I_\nu} |\psi_{W_i}(f)| \right\|_{\mathcal{L}^2(V)} \\
&\le \sum_{\nu>\mu} \left\| \sum_{i\in I_\nu} |\psi_{W_i}(f)| \right\|_{\mathcal{L}^2(V)}.
\end{aligned}$$

The wavelets $\psi_{W_i}(f)$ are constant on the sets W_i, thus $\|\psi_{W_i}(f)\|_{\mathcal{L}^2(V)} = \|\psi_{W_i}(f)\|_{\mathcal{L}^\infty(V)} |W_i|^{1/2}$. This together with the definition of the index sets I_ν gives the bound

$$\begin{aligned}
\|f - \mathcal{S}_m(f)\|_{\mathcal{L}^2(V)} &\le \sum_{\nu>\mu} \left\| \sum_{i\in I_\nu} \|\psi_{W_i}(f)\|_{\mathcal{L}^\infty(V)} \chi_{W_i} \right\|_{\mathcal{L}^2(V)} \\
&= \sum_{\nu>\mu} \left\| \sum_{i\in I_\nu} \frac{\|\psi_{W_i}(f)\|_{\mathcal{L}^2(V)}}{|W_i|^{1/2}} \chi_{W_i} \right\|_{\mathcal{L}^2(V)} \\
&\le \sum_{\nu>\mu} 2^{-\nu+1} \mathcal{N}_r(f,\mathcal{T}_r(f)) \left\| \sum_{i\in I_\nu} \frac{\chi_{W_i}}{|W_i|^{1/2}} \right\|_{\mathcal{L}^2(V)}.
\end{aligned}$$

The balancedness of the BGP(ρ) tree implies that the $\mathcal{L}^2$-norm in the last inequality can be bounded by

$$\begin{aligned}
\left\| \sum_{i\in I_\nu} \frac{\chi_{W_i}}{|W_i|^{1/2}} \right\|_{\mathcal{L}^2(V)} &\le \left(\sum_{i\in I_\nu} \sum_{\mathrm{v}\in W_i} \frac{1}{|W_i|} \left(\sum_{W\supseteq W_i} \frac{|W_i|^{1/2}}{|W|^{1/2}} \right)^2 \right)^{\frac{1}{2}} \\
&\le \left(\sum_{i\in I_\nu} \sum_{\mathrm{v}\in W_i} \frac{1}{|W_i|} \frac{1}{1-\rho^{1/2}} \right)^{\frac{1}{2}} \\
&= \left(1-\rho^{1/2}\right)^{-\frac{1}{2}} |I_\nu|^{1/2},
\end{aligned}$$

where, in the second inequality, we bounded the interior sum similarly as in (9). Combining the last two estimates and the previous bound on the complexity $|I_\nu|$, we can conclude that

$$\begin{aligned}
\|f - \mathcal{S}_m(f)\|_{\mathcal{L}^2(V)} &\le C \sum_{\nu>\mu} 2^{-\nu} \mathcal{N}_r(f,\mathcal{T}_r(f)) |I_\nu|^{1/2} \\
&\le C\mathcal{N}_r(f,\mathcal{T}_r(f)) \sum_{\nu>\mu} 2^{-\nu(1-r/2)} \\
&= C\mathcal{N}_r(f,\mathcal{T}_r(f)) 2^{-\mu(1-r/2)} \\
&\le C m^{-1/r+1/2} \mathcal{N}_r(f,\mathcal{T}_r(f)) \\
&= C m^{-\alpha} \mathcal{N}_r(f,\mathcal{T}_r(f)),
\end{aligned}$$

with $C = 2(1-\rho^{1/2})^{-1/2}$. Now, the bound of Theorem III.3 follows from the first inequality in Theorem III.2.

Acknowledgment

The author thanks the INδAM research group GNCS, the Italian Research Network on Approximation (RITA), and the UMI-TAA research group for their support.

References

[1] L. Birgé and P. Massart, "From model selection to adaptive estimation," in *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics*. Berlin, Germany: Springer, pp. 55–88, 1997.

[2] J. C. Bremer, R. Coifman, M. Maggioni, and A. D. Szlam, "Diffusion wavelet packets," *Appl. Comput. Harmon. Anal.*, vol 21, no. 1, pp. 95–112, 2006.

[3] R. Cavoretto, A. De Rossi, and W. Erb, "Partition of unity methods for signal processing on graphs," *J. Fourier Anal. Appl.*, vol. 27, 2021, Art. no. 66.

[4] X. Cheng, X. Chen, and S. Mallat, "Deep Haar scattering networks," *Inf. Inference*, vol. 5, no. 2, pp. 105–33, 2016.

[5] C. K. Chui, F. Filbir, and H. N. Mhaskar, "Representation of functions on Big Data: Graphs and trees," *Appl. Comput. Harmon. Anal.*, vol. 38, no. 3, pp. 489–509, 2015.

[6] F. R. K. Chung, *Spectral Graph Theory*. Providence, RI, USA: AMS, 1997.

[7] A. Cohen, N. Dyn, F. Hecht, and J.-M. Mirebeau, "Adaptive multiresolution analysis based on anistropic triangulations," *Math. Comput.*, vol. 81, no. 278, pp. 789–810, 2012.

[8] R. Coifman and M. Gavish, "Harmonic analysis of digital data bases," in *Wavelets and Multiscale Analysis*. Berlin, Germany: Springer, pp. 161–197, 2011.

[9] R. Coifman and M. Maggioni, "Diffusion wavelets," *Appl. Comput. Harmon. Anal*. vol. 21, pp. 53–94, 2006.

[10] S. Dekel and D. Leviatan, "Adaptive multivariate approximation using binary space partitions and geometric wavelets," *SIAM J. Numer. Anal.*, vol. 43, no. 2, pp. 707–732, 2005.

[11] L. Demaret, N. Dyn, and A. Iske, "Image compression by linear splines over adaptive triangulations," *Signal Process.*, vol. 86, no. 7, pp. 1604–1616, 2006.

[12] L. Demaret and A. Iske, "Adaptive image approximation by linear splines over locally optimal delaunay triangulations," *IEEE Signal Proc. Lett.*, vol. 13, no. 5, pp. 281–284, May 2006.

[13] L. Demaret and A. Iske, "Optimal N-term approximation by linear splines over anisotropic Delaunay triangulations," *Math. Comput.*, vol. 84, no. 293, pp. 1241–1264, 2015.

[14] R. DeVore, "Nonlinear approximation," *Acta Numer.*, vol. 7, pp. 51–150, 1998.

[15] D. L. Donoho, "Wedgelets: Nearly minimax estimation of edges," *Ann. Statist.*, vol. 27, no. 3, pp. 859–897, 1999.

[16] F. Friedrich, L. Demaret, H. Führ, and K. Wicker, "Efficient moment computation over polygonal domains with an application to rapid wedgelet approximation," *SIAM J. Sci. Comput.*, vol. 29, no. 2, pp. 842–863, 2007.

[17] M. Gavish, B. Nadler, and R. Coifman, "Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 367–374.

[18] T. F. Gonzalez, "Clustering to minimize the maximum intercluster distance," *Theor. Comput. Sci.*, vol. 38, pp. 293–306, 1985.

[19] G. Godsil and G. Royle, *Algebraic Graph Theory*. New York, NY, USA: Springer, 2001.

[20] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, 2011.

[21] J. Irion and N. Saito, "Hierarchical graph laplacian eigen transforms," *JSIAM Lett.*, vol. 6, pp. 21–24, 2014.

[22] J. Irion and N. Saito, "The generalized Haar-Walsh transform," in *Proc. IEEE Workshop Stat. Signal Process.*, 2014, pp. 472–475.

[23] J. Irion and N. Saito, "Efficient approximation and denoising of graph signals using the multiscale basis dictionaries," *IEEE Trans. Signal Inf. Process. Netw*., vol. 3, no. 3, pp. 607–616, Sep. 2017.

[24] M. Jansen, G. P. Nason, and B. W. Silverman, "Multiscale methods for data on graphs and irregular multidimensional situations," *J. Roy. Stat. Soc.*, vol. 71, no. 1, pp. 97–125, 2009.

[25] B. Karaivanov and P. Petrushev, "Nonlinear piecewise polynomial approximation beyond Besov spaces," *Appl. Comput. Harmon. Anal.*, vol. 15, no. 3, pp. 177–223, 2003.

[26] J. Krommweh, "Tetrolet transform: A new adaptive Haar wavelet algorithm for sparse image representation," *J. Vis. Commun. Image Representations*, vol. 21, no. 4, pp. 364–374, 2010.

[27] R. Leonardi and M. Kunt, "Adaptive split-and-merge for image analysis and coding," *Proc. SPIE*, vol. 594, pp. 2–9, 1986.

[28] F. Murtagh, "The Haar wavelet transform of a dendrogram," *J. Classification*, vol. 24, pp. 3–32, 2007.

[29] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, May 2018.

[30] S. K. Narang and A. Ortega, "Perfect reconstruction two-channel wavelet filter-banks for graph structured data," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2786–2799, Jun. 2012.

[31] H. Radha, R. Leonardi, B. Naylor, and M. Vetterli, "Image representation using binary space partitioning trees," *Proc. SPIE*, vol. 1360, pp. 639–650, 1990.

[32] H. Radha, M. Vetterli, and R. Leonardi, "Image compression using binary space partitioning trees," *IEEE Trans. Image Process.*, vol. 5, no. 12, pp. 1610–1624, Dec. 1996.

[33] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 4292–4293. [Online]. Available: http://networkrepository.com

[34] R. Rustamov and L. J. Guibas, "Wavelets on graphs via deep learning," in *Advances in Neural Information Processing Systems*, vol. 26. New York, NY, USA: Curran Associates Inc. 2013.

[35] H. Samet, "Data structures for quadtree approximation and compression," *Commun. ACM*, vol. 28, no. 9, pp. 973–993, 1985.

[36] J. Sharpnack and A. Singh, "Near-optimal and computationally efficient detectors for weak and sparse graph-structured patterns," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2013, pp. 443–446.

[37] G. Shen and A. Ortega, "Transform-based distributed data gathering," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3802–3815, Jul. 2010.

[38] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "A windowed graph fourier transform," in *Proc. IEEE Stat. Signal Process. Workshop*, 2012, pp. 133–136.

[39] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[40] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Appl. Comput. Harmon. Anal.*, vol. 40, no. 2, pp. 260–291, 2016.

[41] D. I. Shuman, "Localized spectral graph filter frames: A unifying framework, survey of design considerations, and numerical comparison," *IEEE Signal Process. Mag.*, vol. 37, no. 6, pp. 43–63, Nov. 2020.

[42] A. D. Szlam, M. Maggioni, R. R. Coifman, and J. C. Bremer, "Diffusion-driven multiscale analysis on manifolds and graphs: Top-down and bottom-up constructions," *Proc. SPIE*, vol. 5914, pp. 445–455, 2005.

[43] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, pp. 395–416, 2007.

[44] M. Wakin, J. Romberg, H. Choi, and R. Baraniuk, "Geometric methods for wavelet-based image compression," *Proc. SPIE*, vol. 5207, pp. 507–520, 2003.

Open Access provided by 'Università degli Studi di Padova' within the CRUI CARE Agreement