Wavelets on Graphs via Deep Learning

Raif M. Rustamov and Leonidas Guibas

Presented by Myriam Bontonou

July 1st, 2020

Motivation

- Context: Existing graph wavelet constructions only guided by the structure of the underlying graph, and not the class of signal.

- Idea: Learning to adapt existing Haar wavelets to a given class of signals.

Motivation

- Context: Existing graph wavelet constructions only guided by the structure of the underlying graph, and not the class of signal.

- Idea: Learning to adapt existing Haar wavelets to a given class of signals.

Outline

- Setting details
  - → Hierarchical partitioning.
  - → Haar wavelets.
- Adapting wavelets to a class of signals
  - → Lifting scheme.
  - → Enforcing sparsity on detail coefficients.
- Experiments

Outline

- **Setting details**
  - → Hierarchical partitioning.
  - → Haar wavelets.
- Adapting wavelets to a class of signals
  - → Lifting scheme.
  - → Enforcing sparsity on detail coefficients.
- Experiments

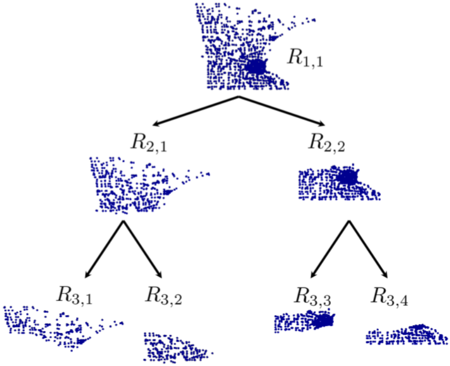
Setting details - Notations

- Connected weighted graph $\mathcal{G}$ with N vertices.
- Signal $f \in \mathbb{R}^N$.
- Edge weight matrix $W \in \mathbb{R}^{N \times N}$ (no self-loop).
- Diagonal matrix of vertex weights $S \in \mathbb{R}^{N \times N}$.

- Integral of f $\int_{\mathcal{G}} f = \sum_i S_{ii} f(i)$.
- Volume of a subset R of vertices $\text{Vol}(R) = \sum_{i \in R} S_{ii}$.

Setting details - Hierarchical partitioning

Hierarchical partitioning of graph $\mathcal{G}$ required!



Setting details - Discrete multi-scale transform

- Set of all signals on graph $\mathcal{G}$: $L^2(\mathcal{G})$,
- Nested sequence of approximation spaces: $V_1 \subset V_2 \cdots \subset V_{l_{\mathsf{max}}} = L^2(\mathcal{G})$,
- Wavelet/detail spaces W_l: such that $V_{l+1} = V_l \oplus W_l$,
- Basis for V_l: scaling functions $\{\phi_{l,k}\}$,
- Basis for W_l: wavelet functions $\{\psi_{l,k}\}$.

Wavelet decomposition of a signal $f \in L^2(\mathcal{G})$ defined as:

$$f = \sum_k a_{l_0,k} \phi_{l_0,k} + \sum_{l=l_0}^{l_{\mathsf{max}}-1} \sum_k d_{l,k} \psi_{l,k} \ .$$

Setting details - Haar wavelets

Given a region of the graph $R_{l,k}$, the Haar approximation of $\bar{a}_{l,k}$ and $\bar{d}_{l,k}$ is used.

- $\bar{a}_{l,k}$ is the average of the graph signal f on $R_{l,k}$:

$$\bar{a}_{l,k} = \mathsf{Vol}(R_{l,k})^{-1} \int_{R_{l,k}} f \ .$$

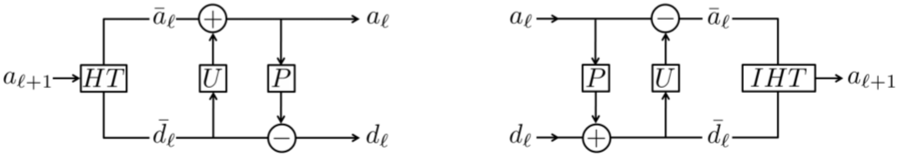
- $\bar{d}_{l,k}$ is associated to a region $R_{l+1,k}$. It is the difference between averages at region $R_{l+1,k}$ and its parent region $R_{l,\mathsf{par}(k)}$:

$$\bar{d}_{l,k} = \bar{a}_{l+1,k} - \bar{a}_{l,\mathsf{par}(k)} \ .$$

Outline

- Setting details
  - → Hierarchical partitioning.
  - → Haar wavelets.
- **Adapting wavelets to a class of signals**
  - → Lifting scheme.
  - → Enforcing sparsity on detail coefficients.
- Experiments

Adaptation - Lifting scheme



Lifting scheme: one step of forward (left) and backward (right) transform. Here a_l and d_l denote the vectors of all approximation and detail coefficients of the lifted transform at level l. U and P are linear update and predict operators. HT and IHT are the Haar transform and its inverse.

Adaptation - Haar wavelets

Given a region of the graph $R_{l,k}$, the Haar approximation is now part of the lifting scheme.

- The average of $R_{l,k}$ is computed as the average of its children regions:

$$\bar{a}_{l,k} = \mathsf{Vol}(R_{l,k})^{-1} \sum_{j,\mathsf{par}(j)=k} a_{l+1,j}\mathsf{Vol}(R_{l+1,j}) \ .$$

- $\bar{d}_{l,k}$ is also a function of a_{l+1}:

$$\bar{d}_{l,k} = a_{l+1,k} - \bar{a}_{l,\mathsf{par}(k)} \ .$$

Adaptation - Enforcing sparsity

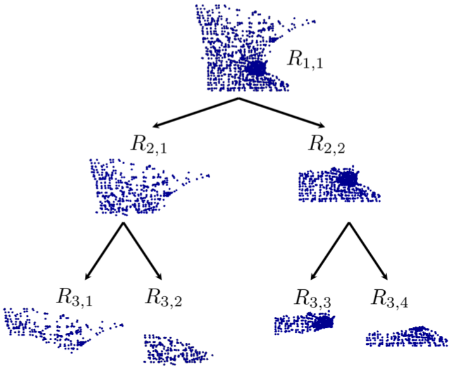
Goal: Design wavelets that yield approximately sparse expansions - i.e., details coefficients should be mostly small.

- Set of training functions $\{f^n\}_{n=1}^{n_{\text{max}}}$
- Given a training function f^n, a level l, and s a sparsity penalty function, the following minimization problem is considered:

$$\{U, P\} = \text{arg}\min_{U,P} \sum_n s(d_l^n) = \text{arg}\min_{U,P} \sum_n s(\bar{d}_l^n - P(\bar{a}_l^n + U\bar{d}_l^n)) \ .$$

- Training from the finest level, where $a_{l+1}^n = f^n$, to coarse level.

To sum up...



1. Choose a hierarchical partitioning algorithm and a maximal level l_{max}.
2. Compute the average of the signal f on the sub-graphs at level l_{max}.
3. For l in range $l_{\text{max}} - 1$ to 1:
   - Using a_{l+1}, compute the Haar coefficients $\bar{a}_l$ and $\bar{d}_l$.
   - Learn the matrices U and P.
   - Retrieve a_l and d_l.

Satisfied requirements?

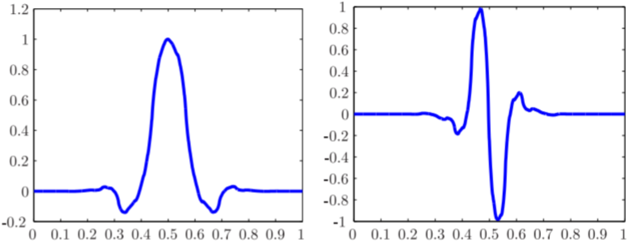
Usual requirements for a proper multi-scale transform

- ✓ critical (or at least controlled) sampling
- ✓ perfect recovery
- ✓ orthogonality
- ✓ applicable to any arbitrary graphs

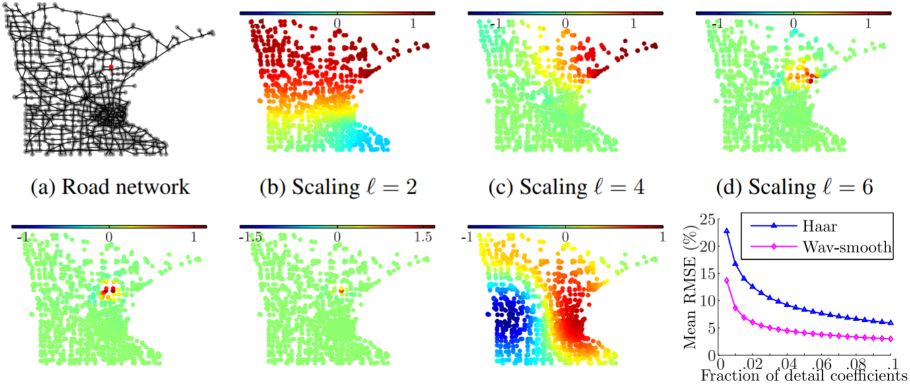
Outline

- Setting details
  - → Hierarchical partitioning.
  - → Haar wavelets.
- Adapting wavelets to a class of signals
  - → Lifting scheme.
  - → Enforcing sparsity on detail coefficients.
- **Experiments**

Experiments



Scaling (left) and wavelet (right) functions on a periodic interval.

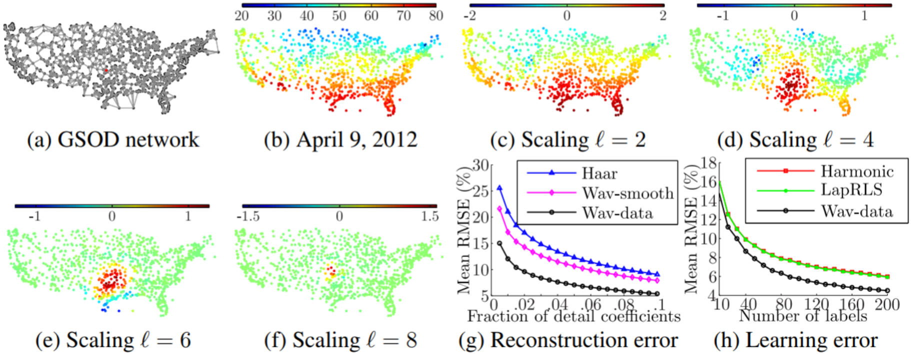


(a) Road network (b) Scaling $\ell = 2$ (c) Scaling $\ell = 4$ (d) Scaling $\ell = 6$

(e) Scaling $\ell = 8$ (f) Scaling $\ell = 10$ (g) Sample function (h) Reconstruction error

Our construction trained with smooth prior on the network (a), yields the scaling functions (b,c,d,e,f). A sample continuous function (g) out of 100 total test functions. Better average reconstruction results (h) for our wavelets (Wav-smooth) indicate a good generalization performance.

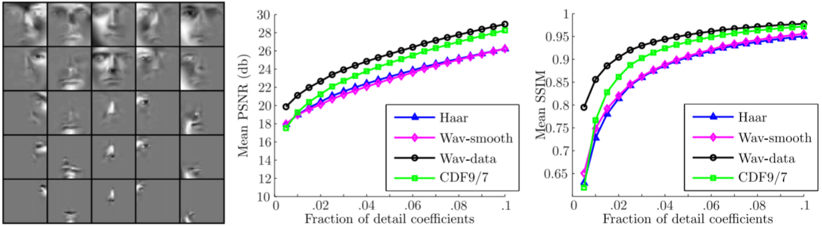
Experiments



(a) GSOD network (b) April 9, 2012 (c) Scaling $\ell = 2$ (d) Scaling $\ell = 4$

(e) Scaling $\ell = 6$ (f) Scaling $\ell = 8$ (g) Reconstruction error (h) Learning error

Our construction on the station network (a) trained with daily temperature data (e.g. (b)), yields the scaling functions (c,d,e,f). Reconstruction results (g) using our wavelets trained on data (Wav-data) and with smooth prior (Wav-smooth). Results on semi-supervised learning (h).



(a) Scaling functions (b) PSNR (c) SSIM

The scaling functions (a) resulting from training on a face images dataset. These wavelets (Wav-data) provide better sparse reconstruction quality than the CDF9/7 wavelets filterbanks (b,c).

Conclusion

- Introduction of a new wavelet transform, which takes into account graph structures, and learns to adapt to classes of signals.
- Graph structures handled with a hierarchical partitioning algorithm and Haar wavelets.
- Classes of signals learned over training functions, through an optimization procedure enforcing a sparse decomposition.