Multiscale Wavelets on Trees, Graphs and High Dimensional Data
ICML 2010, Haifa

Matan Gavish (Weizmann/Stanford)
Boaz Nadler (Weizmann)
Ronald Coifman (Yale)
Boaz Nadler  Ronald Coifman
Motto

“... the relationships between smoothness and frequency forming the core ideas of Euclidean harmonic analysis are remarkably resilient, persisting in very general geometries.”
- Szlam, Maggioni, Coifman (2008)
Given a dataset $X = \{x_1, \ldots, x_N\}$ with similarity matrix $W_{i,j}$
(or $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$)

"Nonparametric" inference of $f : X \rightarrow \mathbb{R}$

- Denoise: observe $g = f + \varepsilon$, recover $f$
- SSL / classification: extend $f$ from $\tilde{X} \subset X$ to $X$
Given a dataset $X = \{x_1, \ldots, x_N\}$ with similarity matrix $W_{i,j}$
(or $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$)

"Nonparametric" inference of $f : X \rightarrow \mathbb{R}$

- Denoise: observe $g = f + \varepsilon$, recover $f$
- SSL / classification: extend $f$ from $\tilde{X} \subset X$ to $X$
Problem setup: Processing functions on a dataset

Given a dataset $X = \{x_1, \ldots, x_N\}$ with similarity matrix $W_{i,j}$
(or $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$)

"Nonparametric" inference of $f : X \to \mathbb{R}$

- Denoise: observe $g = f + \varepsilon$, recover $f$
- SSL / classification: extend $f$ from $\tilde{X} \subset X$ to $X$
Given a dataset $X = \{x_1, \ldots, x_N\}$ with similarity matrix $W_{i,j}$ (or $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$)

"Nonparametric" inference of $f : X \rightarrow \mathbb{R}$

- Denoise: observe $g = f + \varepsilon$, recover $f$
- SSL / classification: extend $f$ from $\tilde{X} \subset X$ to $X$
Given a dataset $X = \{x_1, \ldots, x_N\}$ with similarity matrix $W_{i,j}$ (or $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$)

"Nonparametric" inference of $f : X \rightarrow \mathbb{R}$

- Denoise: observe $g = f + \varepsilon$, recover $f$
- SSL / classification: extend $f$ from $\tilde{X} \subset X$ to $X$
Problem setup: Processing functions on a dataset

Given a dataset \( X = \{x_1, \ldots, x_N\} \) with similarity matrix \( W_{i,j} \)
(or \( X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d \))

"Nonparametric" inference of \( f : X \to \mathbb{R} \)
- Denoise: observe \( g = f + \varepsilon \), recover \( f \)
- SSL / classification: extend \( f \) from \( \tilde{X} \subset X \) to \( X \)
Problem setup: Processing functions on a dataset

Given a dataset $X = \{x_1, \ldots, x_N\}$ with similarity matrix $W_{i,j}$ (or $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$)

"Nonparametric" inference of $f: X \rightarrow \mathbb{R}$

- Denoise: observe $g = f + \varepsilon$, recover $f$
- SSL / classification: extend $f$ from $\tilde{X} \subset X$ to $X$
Problem setup: Processing functions on a dataset

Given a dataset \( X = \{x_1, \ldots, x_N\} \) with similarity matrix \( W_{i,j} \) (or \( X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d \))

"Nonparametric" inference of \( f : X \to \mathbb{R} \)

- Denoise: observe \( g = f + \varepsilon \), recover \( f \)
- SSL / classification: extend \( f \) from \( \tilde{X} \subset X \) to \( X \)
Problem setup: Data adaptive orthobasis

Can use local geometry $W$, but why reinvent the wheel?

Enter Euclid

- Harmonic analysis wisdom in low dim Euclidean space: use orthobasis $\{\psi_i\}$ for space of functions $f : X \rightarrow \mathbb{R}$
- Popular bases: Fourier, wavelet
- Process $f$ in coefficient domain e.g. estimate, threshold

Exit Euclid

We want to build $\{\psi_i\}$ according to graph $W$
Can use local geometry $W$, but why reinvent the wheel?

Enter Euclid
- Harmonic analysis wisdom in low dim Euclidean space: use orthobasis $\{\psi_i\}$ for space of functions $f : X \rightarrow \mathbb{R}$
- Popular bases: Fourier, wavelet
- Process $f$ in coefficient domain e.g. estimate, threshold

Exit Euclid
We want to build $\{\psi_i\}$ according to graph $W$
Problem setup: Data adaptive orthobasis

Can use local geometry $W$, but why reinvent the wheel?

Enter Euclid

- Harmonic analysis wisdom in low dim Euclidean space: use orthobasis $\{\psi_i\}$ for space of functions $f : X \rightarrow \mathbb{R}$
- Popular bases: Fourier, wavelet
- Process $f$ in coefficient domain e.g. estimate, threshold

Exit Euclid

We want to build $\{\psi_i\}$ according to graph $W$
Problem setup: Data adaptive orthobasis

Can use local geometry $W$, but why reinvent the wheel?

Enter Euclid

- Harmonic analysis wisdom in low dim Euclidean space: use orthobasis $\{\psi_i\}$ for space of functions $f : X \rightarrow \mathbb{R}$
- Popular bases: Fourier, wavelet
- Process $f$ in coefficient domain e.g. estimate, threshold

Exit Euclid

We want to build $\{\psi_i\}$ according to graph $W$
Problem setup: Data adaptive orthobasis

Can use local geometry $W$, but why reinvent the wheel?

Enter Euclid

- Harmonic analysis wisdom in low dim Euclidean space: use orthobasis $\{\psi_i\}$ for space of functions $f : X \rightarrow \mathbb{R}$
- Popular bases: Fourier, wavelet
- Process $f$ in coefficient domain e.g. estimate, threshold

Exit Euclid

We want to build $\{\psi_i\}$ according to graph $W$
Problem setup: Data adaptive orthobasis

Can use local geometry $W$, but why reinvent the wheel?

Enter Euclid

- Harmonic analysis wisdom in low dim Euclidean space: use orthobasis $\{\psi_i\}$ for space of functions $f : X \rightarrow \mathbb{R}$
- Popular bases: Fourier, wavelet
- Process $f$ in coefficient domain e.g. estimate, threshold

Exit Euclid

We want to build $\{\psi_i\}$ according to graph $W$
Can use local geometry $W$, but why reinvent the wheel?

**Enter Euclid**
- Harmonic analysis wisdom in low dim Euclidean space: use orthobasis $\{\psi_i\}$ for space of functions $f : X \to \mathbb{R}$
- Popular bases: Fourier, wavelet
- Process $f$ in coefficient domain e.g. estimate, threshold

**Exit Euclid**
We want to build $\{\psi_i\}$ according to graph $W$
Example: USPS benchmark

- $X$ is USPS (ML benchmark) as 1500 vectors in $\mathbb{R}^{16 \times 16} = \mathbb{R}^{256}$
  - Affinity $W_{i,j} = \exp \left( - \|x_i - x_j\|^2 \right)$
  - $f : X \to \{1, -1\}$ is the class label.
Toy example

Chapelle, Scholkopf and Zien, Semi-supervised learning, 2006

Example: USPS benchmark

- $X$ is USPS (ML benchmark) as 1500 vectors in $\mathbb{R}^{16 \times 16} = \mathbb{R}^{256}$
  - Affinity $W_{i,j} = \exp \left( -\|x_i - x_j\|^2 \right)$
  - $f : X \rightarrow \{1, -1\}$ is the class label.

![Images of handwritten numbers 2 and 5]
Toy example

Chapelle, Scholkopf and Zien, Semi-supervised learning, 2006

Example: USPS benchmark

- $X$ is USPS (ML benchmark) as 1500 vectors in $\mathbb{R}^{16 \times 16} = \mathbb{R}^{256}$
- Affinity $W_{i,j} = \exp \left( - \|x_i - x_j\|^2 \right)$
- $f : X \rightarrow \{1, -1\}$ is the class label.
Chapelle, Scholkopf and Zien, Semi-supervised learning, 2006

Example: USPS benchmark

- $X$ is USPS (ML benchmark) as 1500 vectors in $\mathbb{R}^{16\times16} = \mathbb{R}^{256}$
- Affinity $W_{i,j} = \exp\left(-\|x_i - x_j\|^2\right)$
- $f : X \to \{1, -1\}$ is the class label.
Toy example: visualization by kernel PCA
Belkin and Niyogi, Using manifold structure for partially labelled classification, 2003

Generalizing Fourier: The Graph Laplacian eigenbasis

Take \((W - D)\psi_i = \lambda_i \psi_i\) where \(D_{i,i} = \sum_j W_{i,j}\)
## Cons of Laplacian Eigenbasis

<table>
<thead>
<tr>
<th>Laplacian (“Graph Fourier”) basis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oscillatory, nonlocalized</td>
</tr>
<tr>
<td>Uninterpretable</td>
</tr>
<tr>
<td>Scalability challenging</td>
</tr>
<tr>
<td>No theoretical bound on $</td>
</tr>
<tr>
<td>Empirically slow coefficient decay</td>
</tr>
<tr>
<td>No fast transform</td>
</tr>
</tbody>
</table>
Toy example: Graph Laplacian Eigenbasis
Toy example: Graph Laplacian Eigenbasis

Coefficient decay rate of labels function

\[ \log_{10} | \langle \psi_i, f \rangle | \]

Coefficient # i
### Cons of Laplacian Eigenbasis

<table>
<thead>
<tr>
<th>Laplacian eigenbasis</th>
<th>“Dream” Basis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oscillatory, nonlocalized</td>
<td>Localized</td>
</tr>
<tr>
<td>Uninterpretable</td>
<td>Interpretable</td>
</tr>
<tr>
<td>Scalability Challenging</td>
<td>Computation scalable</td>
</tr>
<tr>
<td>No theory bound on $</td>
<td>\langle f, \psi_i \rangle</td>
</tr>
<tr>
<td>Empirically slow decay</td>
<td>Empirically fast decay</td>
</tr>
<tr>
<td>No fast transform</td>
<td>Fast transform</td>
</tr>
</tbody>
</table>
On Euclidean space, Wavelet basis solves this

- Localized
- Interpretable - scale/shift of same function
- **Fundamental wavelet property on** $\mathbb{R}$ **- coeffs decay:**
  If $\psi$ is a regular wavelet and $0 < \alpha < 1$, then

  $$|f(x) - f(y)| \leq C |x - y|^\alpha \iff |\langle f, \psi_{\ell,k} \rangle| \leq \tilde{C} \cdot 2^{-\ell(\alpha + \frac{1}{2})}$$

- Fast transform
Wavelet basis for \( \{f : X \rightarrow \mathbb{R}\} \)?

Prior Art

- Diffusion wavelets (Coifman, Maggioni)
- Anisotropic Haar bases (Donoho)
- Treelets (Nadler, Lee, Wasserman)
Any Balanced Partition Tree whose metric preserves smoothness in $W$ yields an extremely simple Wavelet “Dream” Basis.
## Cons of Laplacian Eigenbasis

<table>
<thead>
<tr>
<th>Laplacian eigenbasis</th>
<th>Haar-like Basis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oscillatory, nonlocalized</td>
<td>Localized</td>
</tr>
<tr>
<td>Uninterpretable</td>
<td>Easily interpretable</td>
</tr>
<tr>
<td>Scalability Challenging</td>
<td>Computation scalable</td>
</tr>
<tr>
<td>No theory bound on $</td>
<td>\langle f, \psi_i \rangle</td>
</tr>
<tr>
<td>Empirically slow decay</td>
<td>Empirically fast decay</td>
</tr>
<tr>
<td>No fast transform</td>
<td>Fast transform</td>
</tr>
</tbody>
</table>

$f$ smooth $\Leftrightarrow |\langle f, \psi_{\ell,k} \rangle| \leq c^{-\ell}$
Toy example: Haar-like coeffs decay

Coefficient decay rate of labels function

$L_{\psi_i, f}$

Coefficient # $i$
Toy example: Haar-like coeffs decay
Toy example: Haar-like coeffs decay
Toy example: Haar-like coeffs decay

Coefficient decay rate of labels function

- **Laplacian eigenbasis**
- **Haar-like basis**

Log scale plot showing the decay of coefficients with index $i$.
Eigenfunctions are oscillatory
Toy example: Haar-like basis function
Any Balanced Partition Tree, whose metric preserves smoothness in $W$, yields an extremely simple Basis.
A Partition Tree on the nodes
A Partition Tree on the nodes
A Partition Tree on the nodes
A Partition Tree on the nodes
A Partition Tree on the nodes
Partition Tree (Dendrogram)
Partition Tree (Dendrogram)
Partition Tree (Dendrogram)
Partition Tree (Dendrogram)
Partition Tree (Dendrogram)
Partition Tree (Dendrogram)
Partition Tree (Dendrogram)
The Haar Basis on $[0, 1]$
The Haar Basis on $[0, 1]$
The Haar Basis on \([0, 1]\)
The Haar Basis on $[0, 1]$
The Haar Basis on $[0, 1]$
The Haar Basis on $[0, 1]$
The Haar Basis on $[0, 1]$
The Haar Basis on $[0, 1]$
The Haar Basis on $[0, 1]$
The Haar Basis on $[0, 1]$
The Haar Basis on $[0, 1]$
Partition Tree (Dendrogram)
Partition Tree (Dendrogram)

\[ x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \quad x_7 \quad x_8 \quad x_9 \]

\[ X \]
Partition Tree (Dendrogram)
Partition Tree (Dendrogram)
Partition Tree (Dendrogram)
Partition Tree (Dendrogram)
Partition Tree $\Rightarrow$ Haar-like basis

\[
\ell = 1 \\
\ell = 2 \\
\ell = 3
\]
Partition Tree $\Rightarrow$ Haar-like basis

\[ \ell = 1 \]
\[ \ell = 2 \]
\[ \ell = 3 \]
Partition Tree $\Rightarrow$ Haar-like basis

\[ \ell = 1 \]
\[ \ell = 2 \]
\[ \ell = 3 \]

$\psi_1$
Partition Tree $\Rightarrow$ Haar-like basis

$\ell = 1$

$\ell = 2$

$\ell = 3$

$\psi_{2,1}$
Partition Tree $\Rightarrow$ Haar-like basis

\[ \psi_{2,2} \]

\[ \ell = 1 \]
\[ \ell = 2 \]
\[ \ell = 3 \]
Partition Tree $\Rightarrow$ Haar-like basis
Partition Tree $\Rightarrow$ Haar-like basis

\[
\ell = 1 \\
\ell = 2 \\
\ell = 3 \\
\psi_{3,2}
\]
Partition Tree $\Rightarrow$ Haar-like basis

$\ell = 1$

$\ell = 2$

$\ell = 3$

$\psi_{3,3}$
Partition Tree $\Rightarrow$ Haar-like basis
Partition Tree $\Rightarrow$ Haar-like basis

\[ \ell = 1 \]
\[ \ell = 2 \]
\[ \ell = 3 \]

\[ \psi_{3,5} \]
Partition Tree $\Rightarrow$ Haar-like basis

\[ \ell = 1 \]
\[ \ell = 2 \]
\[ \ell = 3 \]

\[ \psi_{3,6} \]
Any Balanced Partition Tree, whose metric preserves smoothness in $W$, yields an extremely simple Wavelet Basis.
How to define smoothness

- Partition tree $T$ induces natural tree (ultra-) metric $d$
- Measure smoothness of $f : X \to \mathbb{R}$ w.r.t $d$

Theorem

Let $f : X \to \mathbb{R}$. Then

$$|f(x) - f(y)| \leq C \cdot d(x, y)^{\alpha} \iff |\langle f, \psi_{\ell,k} \rangle| \leq \tilde{C} \cdot |\text{supp}(\psi_{\ell,k})|^{(\alpha + \frac{1}{2})}$$

for any Haar-like basis $\{\psi_{\ell,k}\}$ based on the tree $T$.

- If the tree is balanced $\Rightarrow |\text{offspring folder}| \leq q \cdot |\text{parent folder}|$
- Then

$$|f(x) - f(y)| \leq C \cdot d(x, y)^{\alpha} \iff |\langle f, \psi_{\ell,k} \rangle| \leq \tilde{C} \cdot q^\ell(\alpha + \frac{1}{2})$$

(for classical Haar, $q = \frac{1}{2}$).
How to define smoothness

- Partition tree $T$ induces natural tree (ultra-) metric $d$
- Measure smoothness of $f : X \rightarrow \mathbb{R}$ w.r.t $d$

**Theorem**

Let $f : X \rightarrow \mathbb{R}$. Then

$$|f(x) - f(y)| \leq C \cdot d(x, y)^\alpha \iff |\langle f, \psi_{\ell,k} \rangle| \leq \tilde{C} \cdot |\text{supp}(\psi_{\ell,k})|^{(\alpha + \frac{1}{2})}$$

for any Haar-like basis $\{\psi_{\ell,k}\}$ based on the tree $T$.

- If the tree is balanced $\Rightarrow |\text{offspring folder}| \leq q \cdot |\text{parent folder}|$
- Then

$$|f(x) - f(y)| \leq C \cdot d(x, y)^\alpha \iff |\langle f, \psi_{\ell,k} \rangle| \leq \tilde{C} \cdot q^{\ell(\alpha + \frac{1}{2})}$$

(for classical Haar, $q = \frac{1}{2}$).
How to define smoothness

- Partition tree $T$ induces natural tree (ultra-) metric $d$
- Measure smoothness of $f: X \rightarrow \mathbb{R}$ w.r.t $d$

Theorem

Let $f: X \rightarrow \mathbb{R}$. Then

$$|f(x) - f(y)| \leq C \cdot d(x, y)^{\alpha} \iff |\langle f, \psi_{\ell,k} \rangle| \leq \tilde{C} \cdot |\text{supp}(\psi_{\ell,k})|^{(\alpha + \frac{1}{2})}$$

for any Haar-like basis $\{\psi_{\ell,k}\}$ based on the tree $T$.

- If the tree is balanced $\Rightarrow |\text{offspring folder}| \leq q \cdot |\text{parent folder}|$
- Then

$$|f(x) - f(y)| \leq C \cdot d(x, y)^{\alpha} \iff |\langle f, \psi_{\ell,k} \rangle| \leq \tilde{C} \cdot q^{\ell(\alpha + \frac{1}{2})}$$

(for classical Haar, $q = \frac{1}{2}$).
$f$ smooth in tree metric $\iff$ coefs decay

How to define smoothness

- Partition tree $T$ induces natural tree (ultra-) metric $d$
- Measure smoothness of $f : X \to \mathbb{R}$ w.r.t $d$

Theorem

Let $f : X \to \mathbb{R}$. Then

$$|f(x) - f(y)| \leq C \cdot d(x, y)^\alpha \iff |\langle f, \psi_{\ell,k} \rangle| \leq \tilde{C} \cdot |\text{supp}(\psi_{\ell,k})|^{(\alpha + \frac{1}{2})}$$

for any Haar-like basis $\{\psi_{\ell,k}\}$ based on the tree $T$.

- If the tree is balanced $\Rightarrow |\text{offspring folder}| \leq q \cdot |\text{parent folder}|$

- Then

$$|f(x) - f(y)| \leq C \cdot d(x, y)^\alpha \iff |\langle f, \psi_{\ell,k} \rangle| \leq \tilde{C} \cdot q^{\ell(\alpha + \frac{1}{2})}$$

(for classical Haar, $q = \frac{1}{2}$).
How to define smoothness

- Partition tree $T$ induces a natural tree (ultra-) metric $d$
- Measure smoothness of $f : X \to \mathbb{R}$ w.r.t $d$

Theorem

Let $f : X \to \mathbb{R}$. Then

$$|f(x) - f(y)| \leq C \cdot d(x, y)^\alpha \iff |\langle f, \psi_{\ell,k} \rangle| \leq \tilde{C} \cdot |\text{supp}(\psi_{\ell,k})|^{(\alpha+\frac{1}{2})}$$

for any Haar-like basis $\{\psi_{\ell,k}\}$ based on the tree $T$.

- If the tree is balanced $\Rightarrow \ |\text{offspring folder}| \leq q \cdot |\text{parent folder}|$
- Then

$$|f(x) - f(y)| \leq C \cdot d(x, y)^\alpha \iff |\langle f, \psi_{\ell,k} \rangle| \leq \tilde{C} \cdot q^{\ell(\alpha+\frac{1}{2})}$$

(for classical Haar, $q = \frac{1}{2}$).
Any partition tree on $X$ induces “wavelet” Haar-like bases ✓

“Balanced” tree $\Rightarrow f$ smooth equals fast coefficient decay ✓

Application to semi-supervised learning

Beyond basics: Comparing trees, Tensor product of Haar-like bases
Application: Semi-supervised learning

Classification/Regression with Haar-like basis

- Task: Given values of smooth $f$ on $\tilde{X} \subset X$, extend $f$ to $X$.
- Step 1: Build a partition tree s.t. $f$ is smooth w.r.t tree metric
- Step 2: Construct a Haar-like basis $\{\psi_{\ell,i}\}$
- Step 3: Estimate $\hat{f} = \sum \langle f, \psi_{\ell,i} \rangle \psi_{\ell,i}$

- Control over coefficient decay $\Rightarrow$ non-parametric risk analysis
- Bound on $\mathbb{E} \left\| f - \hat{f} \right\|^2$ depends only on smoothness $\alpha$ of the target $f$ and $\# \text{labeled points}$
Task: Given values of smooth $f$ on $\tilde{X} \subset X$, extend $f$ to $X$.

- Step 1: Build a partition tree s.t. $f$ is smooth w.r.t tree metric
- Step 2: Construct a Haar-like basis $\{\psi_{\ell,i}\}$
- Step 3: Estimate $\hat{f} = \sum \langle f, \psi_{\ell,i} \rangle \psi_{\ell,i}$

Control over coefficient decay $\Rightarrow$ non-parametric risk analysis

Bound on $\mathbb{E} \left\| f - \hat{f} \right\|^2$ depends only on smoothness $\alpha$ of the target $f$ and # labeled points
Application: Semi supervised learning

Classification/Regression with Haar-like basis

- Task: Given values of smooth $f$ on $\tilde{X} \subset X$, extend $f$ to $X$.
- Step 1: Build a partition tree s.t. $f$ is smooth w.r.t tree metric
- Step 2: Construct a Haar-like basis $\{\psi_{\ell,i}\}$
- Step 3: Estimate $\hat{f} = \sum \langle f, \psi_{\ell,i} \rangle \psi_{\ell,i}$

- Control over coefficient decay \(\Rightarrow\) non-parametric risk analysis
- Bound on $\mathbb{E} \left\| f - \hat{f} \right\|^2$ depends only on smoothness $\alpha$ of the target $f$ and \# labeled points
Application: Semi supervised learning

Classification/Regression with Haar-like basis

- Task: Given values of smooth $f$ on $\tilde{X} \subset X$, extend $f$ to $X$.
- Step 1: Build a partition tree s.t. $f$ is smooth w.r.t tree metric
- Step 2: Construct a Haar-like basis $\{\psi_{\ell,i}\}$
- Step 3: Estimate $\hat{f} = \sum \langle f, \psi_{\ell,i} \rangle \psi_{\ell,i}$

- Control over coefficient decay $\Rightarrow$ non-parametric risk analysis
- Bound on $\mathbb{E} \left\| f - \hat{f} \right\|^2$ depends only on smoothness $\alpha$ of the target $f$ and $\#$ labeled points
Application: Semi supervised learning

Classification/Regression with Haar-like basis

- Task: Given values of smooth $f$ on $\tilde{X} \subset X$, extend $f$ to $X$.
- Step 1: Build a partition tree s.t. $f$ is smooth w.r.t tree metric
- Step 2: Construct a Haar-like basis $\{\psi_{\ell,i}\}$
- Step 3: Estimate $\hat{f} = \sum \langle f, \psi_{\ell,i} \rangle \psi_{\ell,i}$

- Control over coefficient decay $\Rightarrow$ non-parametric risk analysis
- Bound on $\mathbb{E} \left\| f - \hat{f} \right\|^2$ depends only on smoothness $\alpha$ of the target $f$ and $\#$ labeled points
Classification/Regression with Haar-like basis

- Task: Given values of smooth $f$ on $\tilde{X} \subset X$, extend $f$ to $X$.
- Step 1: Build a partition tree s.t. $f$ is smooth w.r.t tree metric
- Step 2: Construct a Haar-like basis $\{\psi_{\ell,i}\}$
- Step 3: Estimate $\hat{f} = \sum \langle f, \psi_{\ell,i} \rangle \psi_{\ell,i}$

- Control over coefficient decay $\Rightarrow$ non-parametric risk analysis
- Bound on $E \left\| f - \hat{f} \right\|^2$ depends only on smoothness $\alpha$ of the target $f$ and $\#$ labeled points
Classification/Regression with Haar-like basis

- Task: Given values of smooth $f$ on $\tilde{X} \subset X$, extend $f$ to $X$.
- Step 1: Build a partition tree s.t. $f$ is smooth w.r.t tree metric
- Step 2: Construct a Haar-like basis $\{\psi_\ell,i\}$
- Step 3: Estimate $\hat{f} = \sum \langle f, \psi_\ell,i \rangle \psi_\ell,i$

- Control over coefficient decay $\Rightarrow$ non-parametric risk analysis
- Bound on $\mathbb{E} \left\| f - \hat{f} \right\|^2$ depends only on smoothness $\alpha$ of the target $f$ and # labeled points
Toy Example benchmark

![Graph showing test error (%) against the number of labeled points (out of 1500)].

- Laplacian Eigenmaps
- Laplacian Reg.
- Adaptive Threshold
- Haar–like basis
- State of the art

Legend:
- Red circle: Laplacian Eigenmaps
- Pink triangle: Laplacian Reg.
- Black star: Adaptive Threshold
- Blue asterisk: Haar–like basis
- Green diamond: State of the art
MNIST Digits 8 vs. \{3,4,5,7\}

![Graph showing test error vs. number of labeled points for different methods.](image)
### Results

1. Any partition tree on $X$ induces "wavelet" Haar-like bases ✓
2. "Balanced" tree $\Rightarrow f$ smooth equals fast coefficient decay ✓
3. Application to semi-supervised learning ✓
4. Beyond basics: Tensor product of Haar-like bases, Coefficient thresholding
Tensor product of Haar-like bases

Tensor product of Haar-like bases


data points
Tensor product of Haar-like bases

Tensor product of Haar-like bases

Tensor product of Haar-like bases

Tensor product of Haar-like bases

Tensor product of Haar-like bases

Tensor product of Haar-like bases

Tensor product of Haar-like bases

Tensor product of Haar-like bases

Tensor product of Haar-like bases

Tensor product of Haar-like bases

Tensor product of Haar-like bases

Tensor product of Haar-like bases

\[ = \sum_{i,j} a_{ij} \psi_i \otimes \varphi_j \]
Coifman and Weiss, *Extensions of Hardy spaces and their use in analysis*, 1979

- **Geometric** tools for Data analysis widely recognized
- **Analysis** tools (e.g. function spaces, wavelet theory) in graph or general geometries valuable and largely unexplored in ML context
- Deep theory, long tradition: geometry of $X \iff$ bases for $\{ f : X \rightarrow \mathbb{R} \}$ (“Spaces of Homogeneous Type”)
Coifman and Weiss, *Extensions of Hardy spaces and their use in analysis*, 1979

- **Geometric** tools for Data analysis widely recognized
- **Analysis** tools (e.g. function spaces, wavelet theory) in graph or general geometries valuable and largely unexplored in ML context
- Deep theory, long tradition: geometry of $X \leftrightarrow$ bases for \{ $f : X \to \mathbb{R}$ \} ("Spaces of Homogeneous Type")
Coifman and Weiss, Extensions of Hardy spaces and their use in analysis, 1979

- **Geometric** tools for Data analysis widely recognized
- **Analysis** tools (e.g. function spaces, wavelet theory) in graph or general geometries valuable and largely unexplored in ML context

  - Deep theory, long tradition: geometry of $X \iff \text{bases for } \{ f : X \to \mathbb{R} \}$ (“Spaces of Homogeneous Type”)
Coifman and Weiss, Extensions of Hardy spaces and their use in analysis, 1979

- **Geometric** tools for Data analysis widely recognized
- **Analysis** tools (e.g. function spaces, wavelet theory) in graph or general geometries valuable and largely unexplored in ML context
- Deep theory, long tradition: geometry of $X \iff$ bases for $\{ f : X \rightarrow \mathbb{R} \}$ (“Spaces of Homogeneous Type”)
Summary

Motto
“... the relationships between smoothness and frequency forming the core ideas of Euclidean harmonic analysis are remarkably resilient, persisting in very general geometries.”
- Szlam, Maggioni, Coifman (2008)

Main message
Any Balanced Partition Tree whose metric preserves smoothness in $W$ yields an extremely simple “Dream” Wavelet Basis

Fascinating open question
Which graphs admit Balanced Partition Trees, whose metric preserves smoothness in $W$?
Supporting Information - proofs & code:
www.stanford.edu/~gavish ;
www.wisdom.weizmann.ac.il/~nadler/

G, Nadler and Coifman, *Multiscale wavelets on trees, graphs and high dimensional data*, Proceedings of ICML 2010


Singh, Nowak and Calderbank, Detecting weak but hierarchically-structured patterns in networks, Proceedings of AISTATS 2010