# 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)

# Problem setup: Processing functions on a dataset

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

**"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

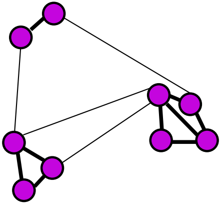
Given a dataset $X = \{x_1, \ldots, x_N\}$ with similarity matrix $W_{i,j}$
(or $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$)

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

- 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 $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$)

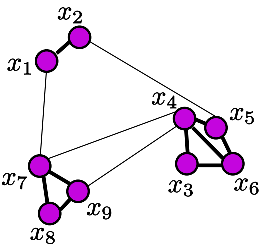


**"Nonparametric"** inference of $f : X \to \mathbb{R}$

- 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 $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$)

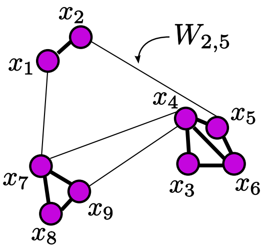



**"Nonparametric"** inference of $f : X \to \mathbb{R}$

- 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 $X = \{x_1, \dots, x_N\}$ with similarity matrix $W_{i,j}$
(or $X = \{x_1, \dots, x_N\} \subset \mathbb{R}^d$)

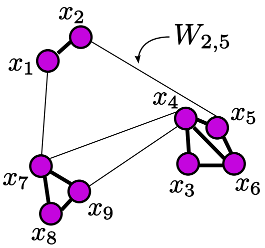



**"Nonparametric"** inference of $f : X \to \mathbb{R}$

- 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 $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$)

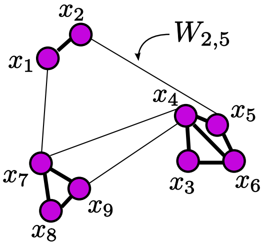



**"Nonparametric"** inference of $f : X \to \mathbb{R}$

- 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 $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$)

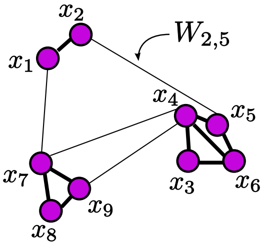



**"Nonparametric"** inference of $f : X \to \mathbb{R}$

- 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 $f : X \to \mathbb{R}$

- 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 $f : X \to \mathbb{R}$
- Popular bases: Fourier, wavelet
- Process $f$ in coefficient domain e.g. estimate, thereshold

**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 $f : X \to \mathbb{R}$
- Popular bases: Fourier, wavelet
- Process $f$ in coefficient domain e.g. estimate, thereshold

**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 $f : X \to \mathbb{R}$
- Popular bases: Fourier, wavelet
- Process $f$ in coefficient domain e.g. estimate, thereshold

## 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 $f : X \to \mathbb{R}$
- Popular bases: Fourier, wavelet
- Process $f$ in coefficient domain e.g. estimate, thereshold

## 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 $f : X \to \mathbb{R}$
- Popular bases: Fourier, wavelet
- Process $f$ in coefficient domain e.g. estimate, thereshold

## 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, thereshold

**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 $f : X \to \mathbb{R}$
- Popular bases: Fourier, wavelet
- Process $f$ in coefficient domain e.g. estimate, thereshold

## Exit Euclid

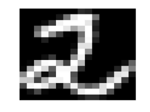
We want to build $\{\psi_i\}$ according to graph $W$

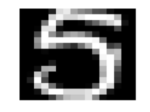
# Toy example

**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 $W_{i,j} = \exp\left(-\|x_i - x_j\|^2\right)$
    - $f : X \to \{1, -1\}$ is the class label.





# Toy example

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

## Example: USPS benchmark

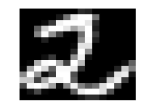
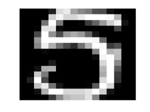
- $X$ is USPS (ML benchmark) as 1500 vectors in $\mathbb{R}^{16\times16} = \mathbb{R}^{256}$
  - Affinity $W_{i,j} = \exp\left(-\|x_i - x_j\|^2\right)$
  - $f : X \to \{1, -1\}$ is the class label.

# Toy example

**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 $W_{i,j} = \exp\left(-\|x_i - x_j\|^2\right)$
  - $f : X \to \{1, -1\}$ is the class label.

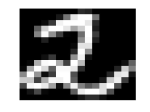



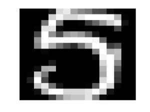
# Toy example

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

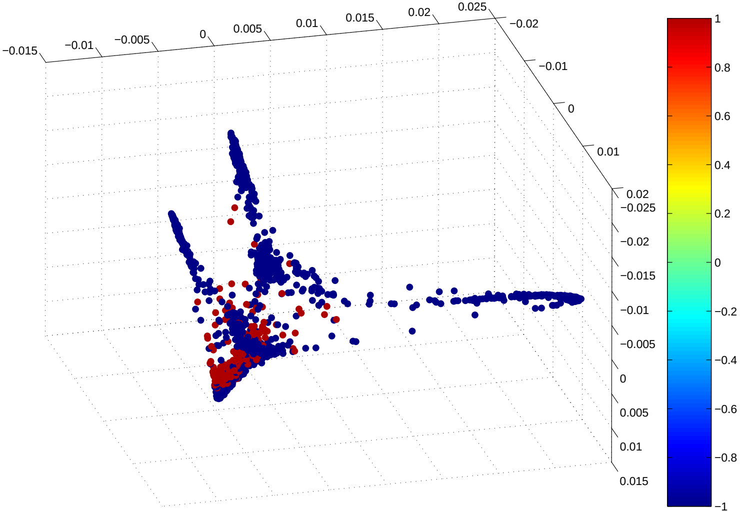
## Example: USPS benchmark

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





# Toy example: visualization by kernel PCA

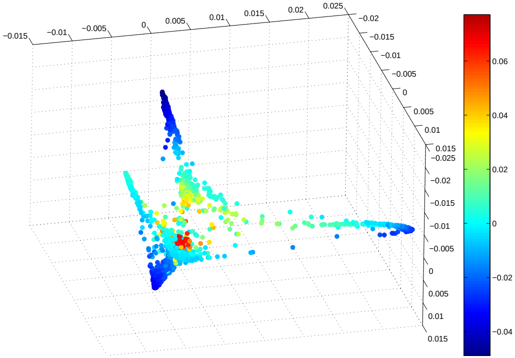


# Fourier Basis for $\{f : X \to \mathbb{R}\}$

**Belkin and Niyogi, Using manifold structure for partially labelled classification, 2003**

## Generalizing Fourier: The Graph Laplacian eigenbasis

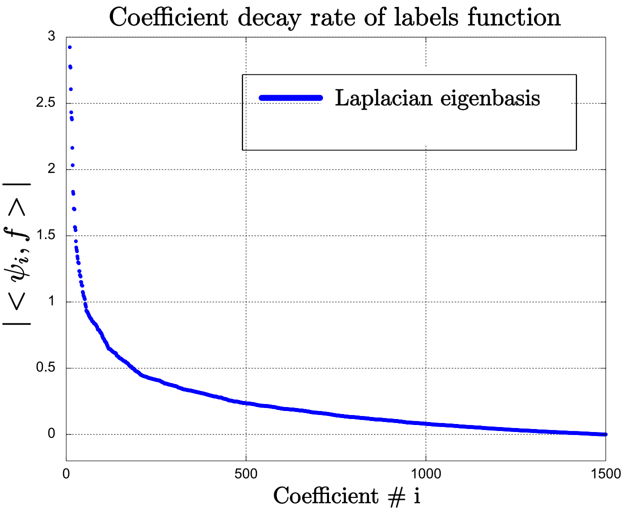
Take $(W - D)\psi_i = \lambda_i \psi_i$ where $D_{i,i} = \sum_j W_{i,j}$



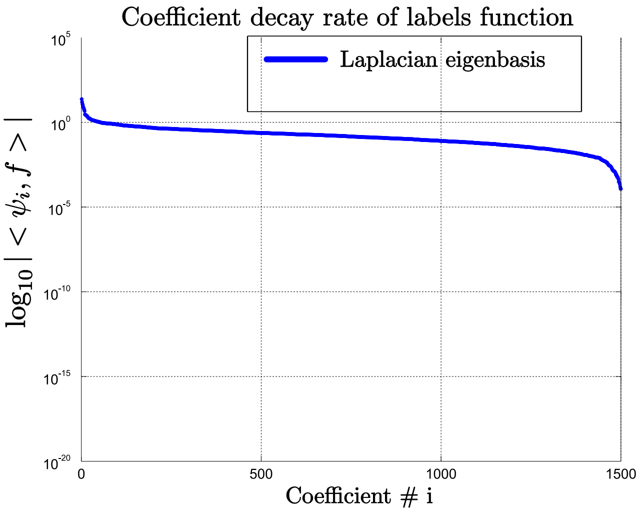
# Cons of Laplacian Eigenbasis

| **Laplacian ("Graph Fourier") basis** |
|---|
| Oscillatory, nonlocalized |
| Uninterpretable |
| Scalability challenging |
| No theoretical bound on $|\langle f, \psi_i \rangle|$ |
| Empirically slow coefficient decay |
| No fast transform |

# Toy example: Graph Laplacian Eigenbasis



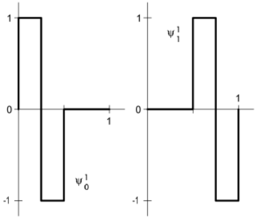
# Toy example: Graph Laplacian Eigenbasis



# Cons of Laplacian Eigenbasis

| **Laplacian eigenbasis** | **“Dream” Basis** |
|---|---|
| Oscillatory, nonlocalized | Localized |
| Uninterpretable | Interpretable |
| Scalability Challenging | Computation scalable |
| No theory bound on $\lvert\langle f, \psi_i\rangle\rvert$ | Provably fast coef. decay |
| Empirically slow decay | Empricially fast decay |
| No fast transform | Fast transform |

# On Euclidean space, Wavelet basis solves this




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

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

- Fast transform

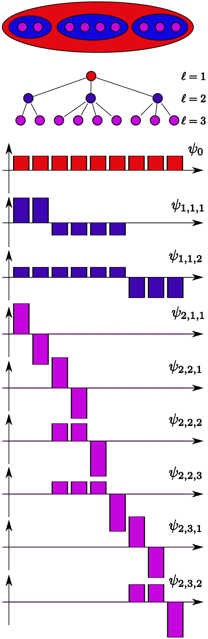
# Wavelet basis for $\{f : X \to \mathbb{R}\}$?

**Prior Art**

- Diffusion wavelets (Coifman, Maggioni)
- Anisotropic Haar bases (Donoho)
- Treelets (Nadler, Lee, Wasserman)

# Main Message

Any *Balanced Partition Tree* whose metric preserves smoothness in $W$ yields an extremely simple Wavelet "Dream" Basis

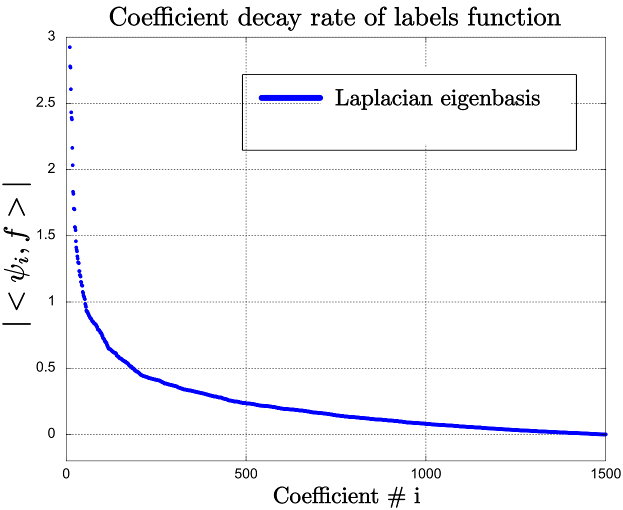


$\ell = 1$
$\ell = 2$
$\ell = 3$
$\psi_0$
$\psi_{1,1,1}$
$\psi_{1,1,2}$
$\psi_{2,1,1}$
$\psi_{2,2,1}$
$\psi_{2,2,2}$
$\psi_{2,2,3}$
$\psi_{2,3,1}$
$\psi_{2,3,2}$

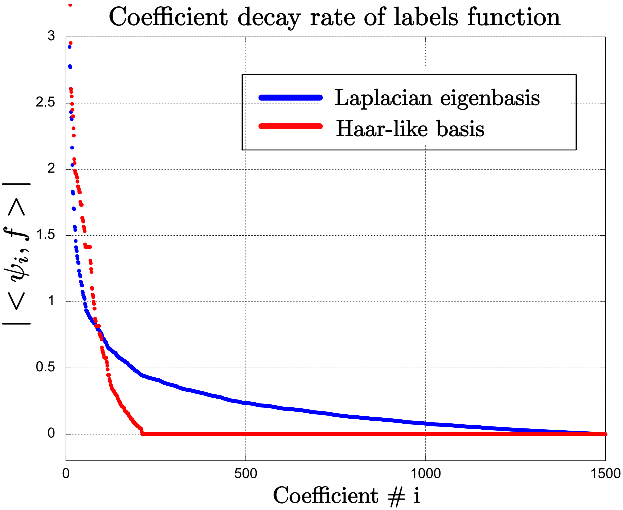
# Cons of Laplacian Eigenbasis

| **Laplacian eigenbasis** | **Haar-like Basis** |
|---|---|
| Oscillatory, nonlocalized | Localized |
| Uninterpretable | Easily interpretable |
| Scalability Challenging | Computation scalable |
| No theory bound on $\lvert\langle f, \psi_i\rangle\rvert$ | $f$ smooth $\Leftrightarrow \lvert\langle f, \psi_{\ell,k}\rangle\rvert \leq c^{-\ell}$ |
| Empirically slow decay | Empricially fast decay |
| No fast transform | Fast transform |

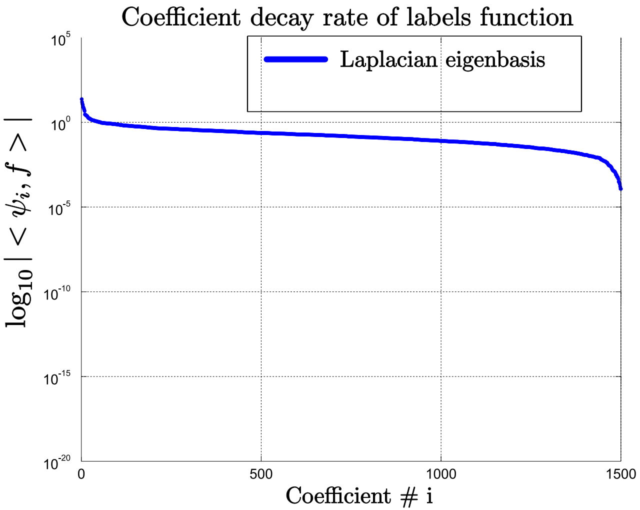
# Toy example: Haar-like coeffs decay



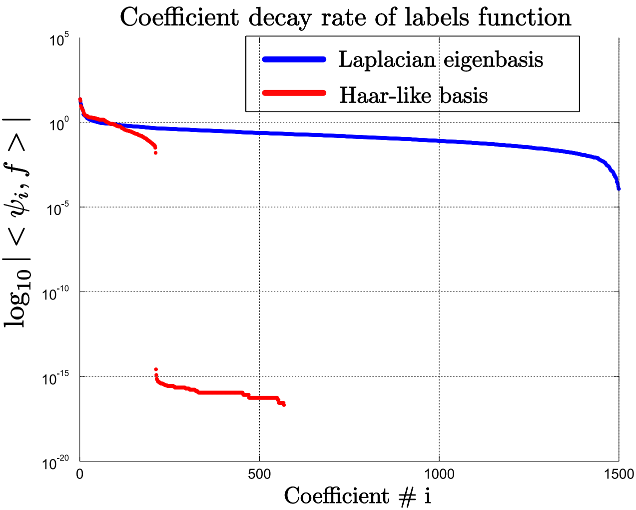
# Toy example: Haar-like coeffs decay



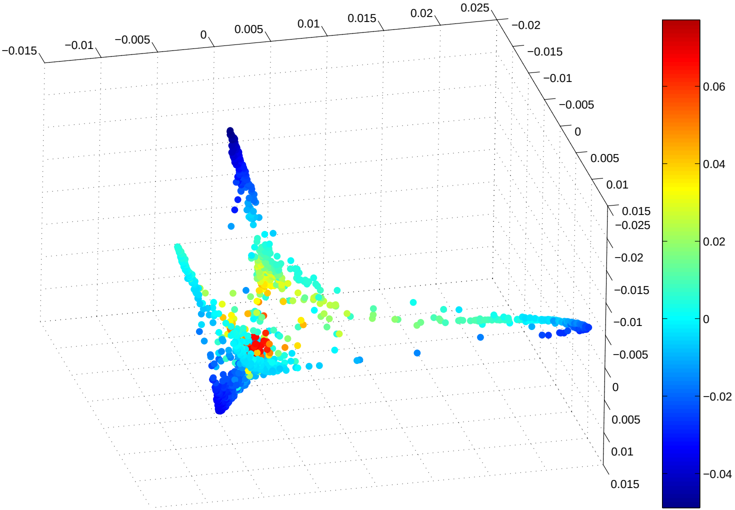
# Toy example: Haar-like coeffs decay



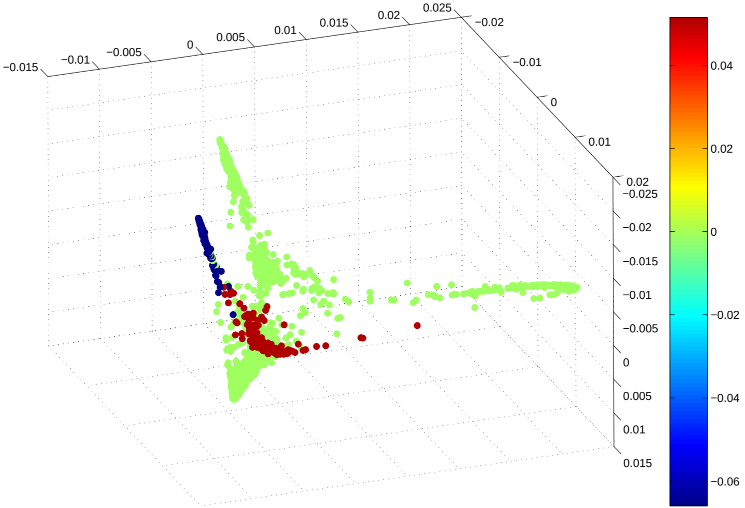
# Toy example: Haar-like coeffs decay



# Eigenfunctions are oscillatory



# Toy example: Haar-like basis function

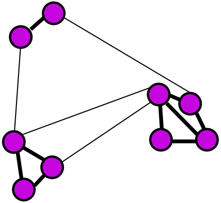


# Main Message

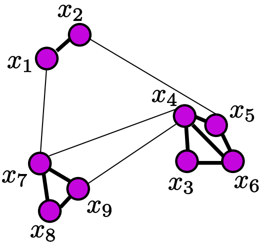
**Any** *Balanced* **Partition Tree**, whose metric preserves smoothness in $W$,**yields an extremely simple Basis**

# A Partition Tree on the nodes

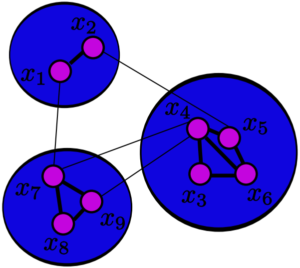
# A Partition Tree on the nodes



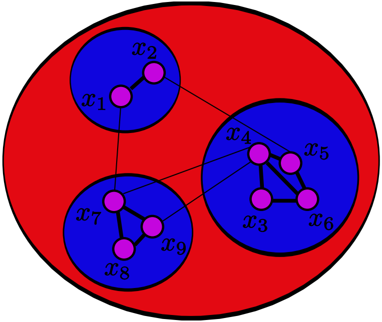
# A Partition Tree on the nodes



# A Partition Tree on the nodes

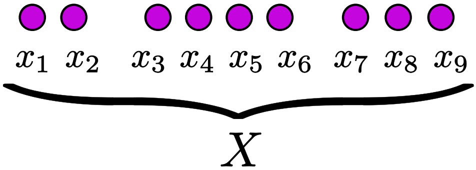


# A Partition Tree on the nodes



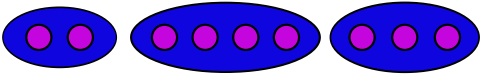
# Partition Tree (Dendrogram)

# Partition Tree (Dendrogram)

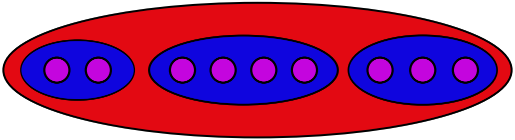


# Partition Tree (Dendrogram)

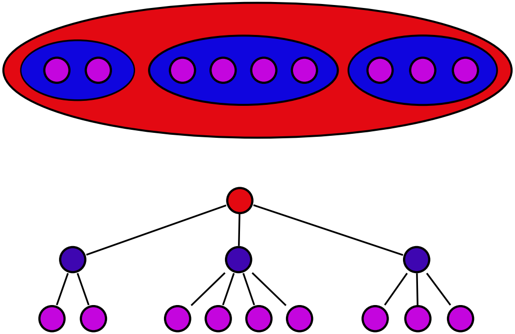
# Partition Tree (Dendrogram)



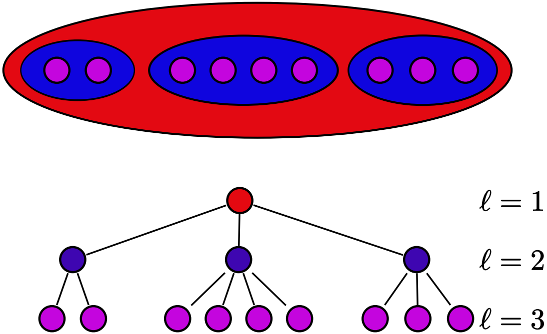
# Partition Tree (Dendrogram)



# Partition Tree (Dendrogram)



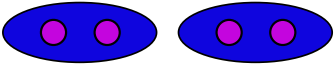
# Partition Tree (Dendrogram)



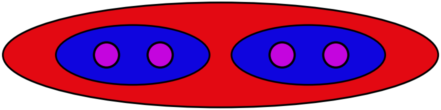
# The Haar Basis on $[0, 1]$

# The Haar Basis on $[0, 1]$

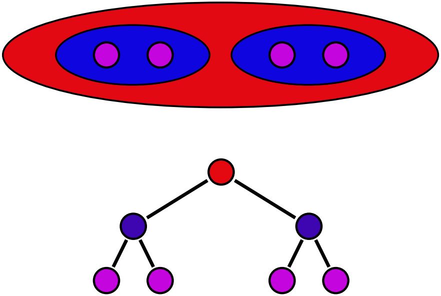
# The Haar Basis on $[0, 1]$



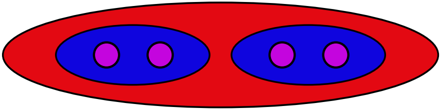
# The Haar Basis on $[0, 1]$

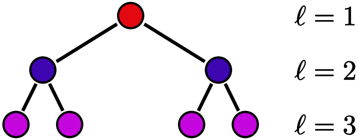


# The Haar Basis on $[0, 1]$

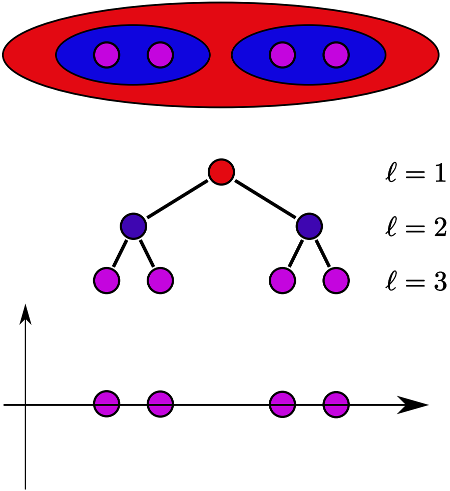


# The Haar Basis on $[0, 1]$

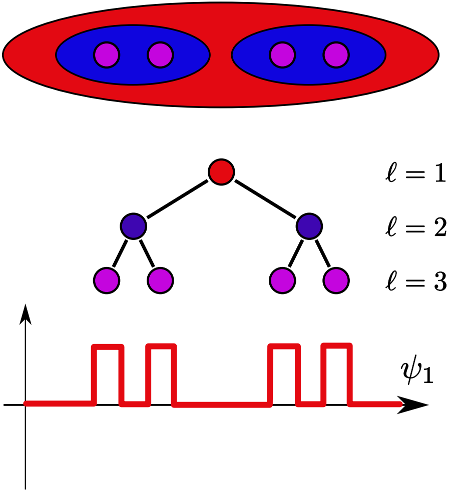




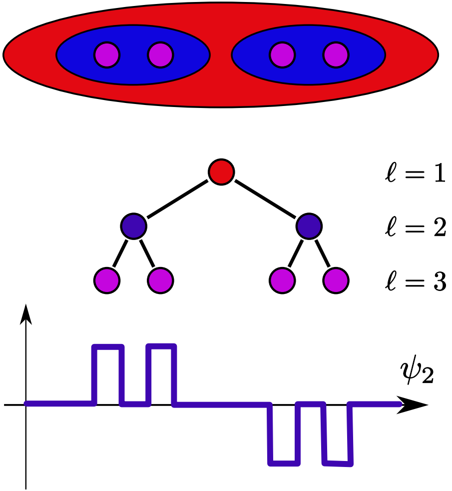
# The Haar Basis on $[0, 1]$



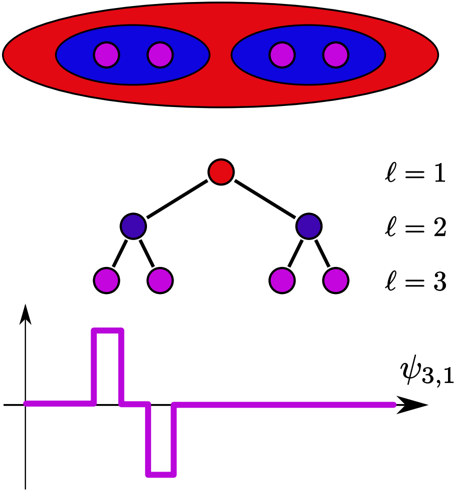
# The Haar Basis on $[0, 1]$



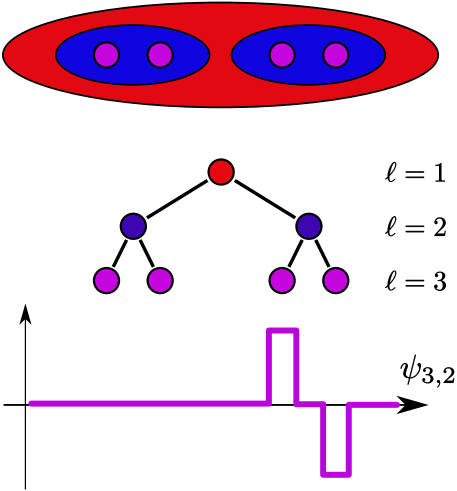
# The Haar Basis on [0, 1]



# The Haar Basis on [0, 1]

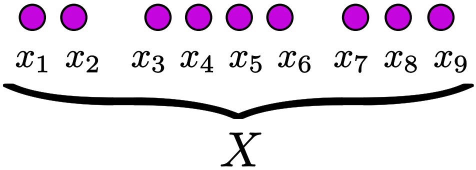


# The Haar Basis on $[0, 1]$



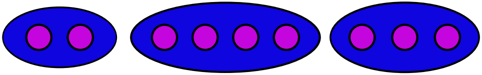
# Partition Tree (Dendrogram)

# Partition Tree (Dendrogram)

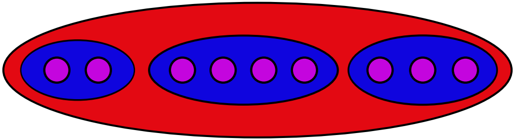


# Partition Tree (Dendrogram)

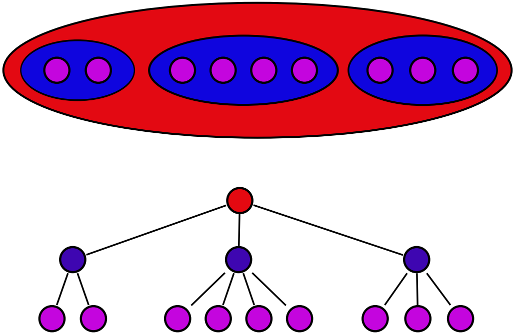
# Partition Tree (Dendrogram)



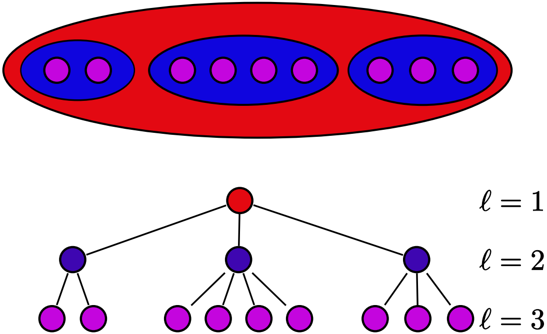
# Partition Tree (Dendrogram)



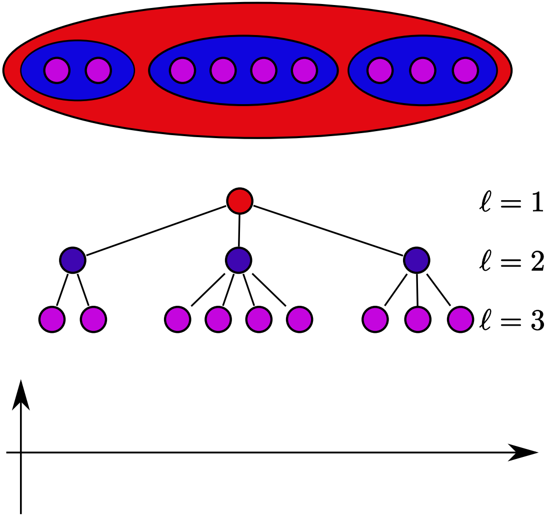
# Partition Tree (Dendrogram)



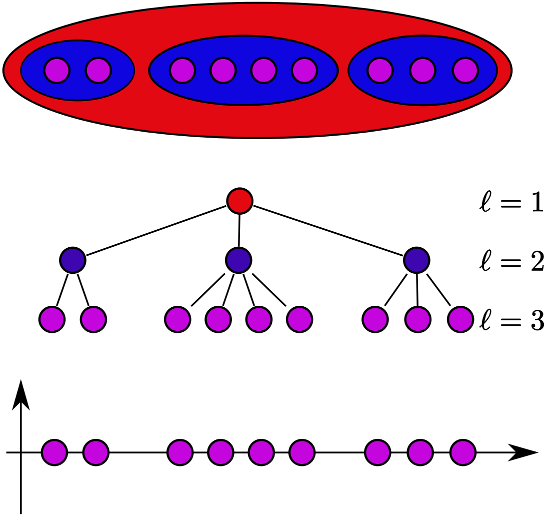
# Partition Tree (Dendrogram)



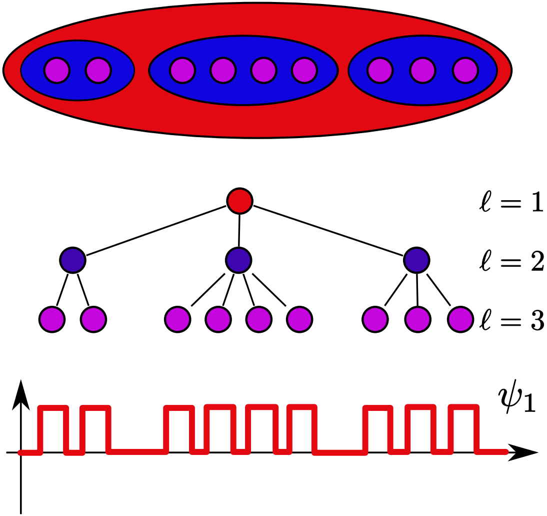
# Partition Tree ⇒ Haar-like basis



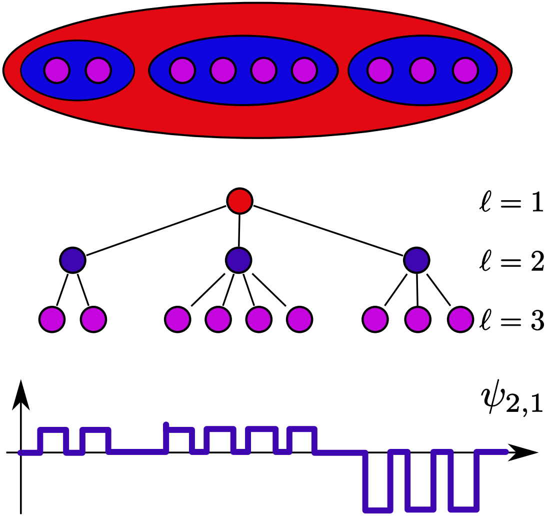
# Partition Tree ⇒ Haar-like basis



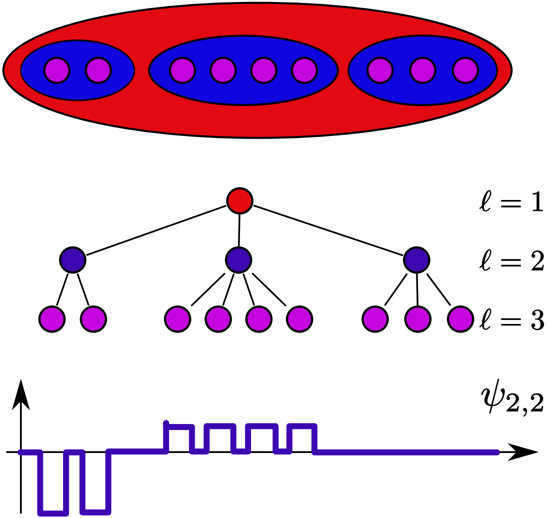
# Partition Tree $\Rightarrow$ Haar-like basis



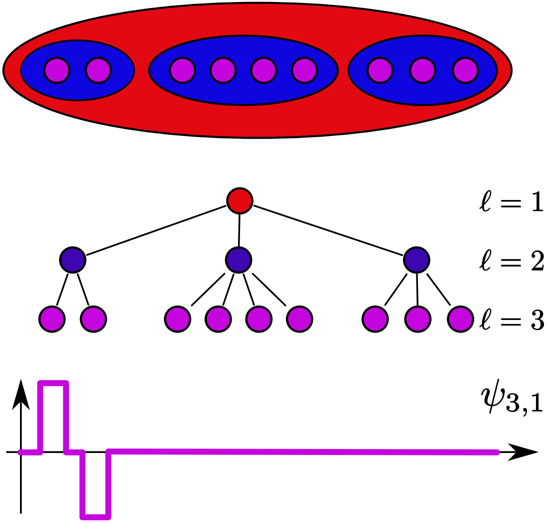
# Partition Tree ⇒ Haar-like basis



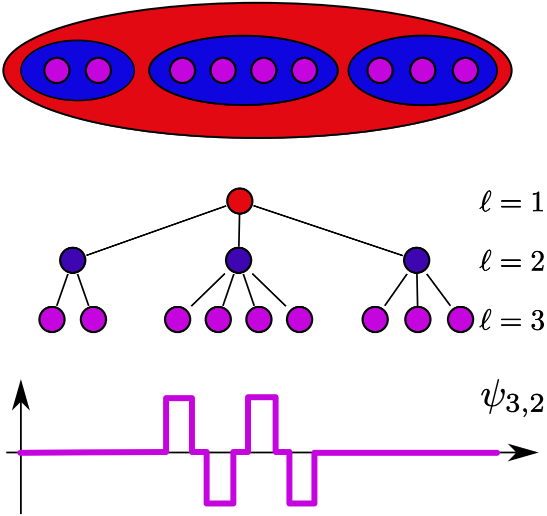
# Partition Tree ⇒ Haar-like basis



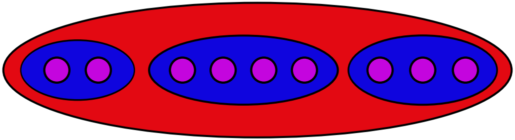
# Partition Tree $\Rightarrow$ Haar-like basis

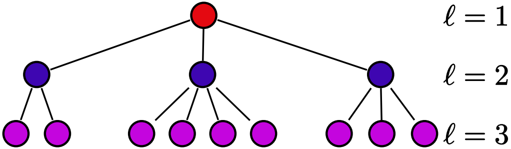


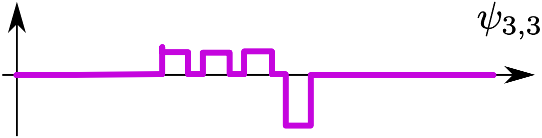
# Partition Tree ⇒ Haar-like basis



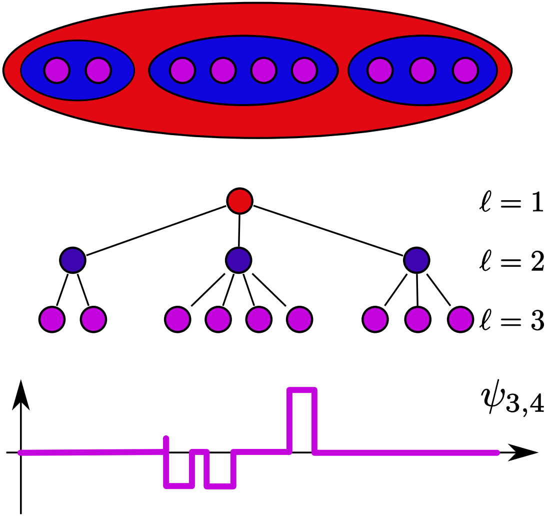
# Partition Tree ⇒ Haar-like basis



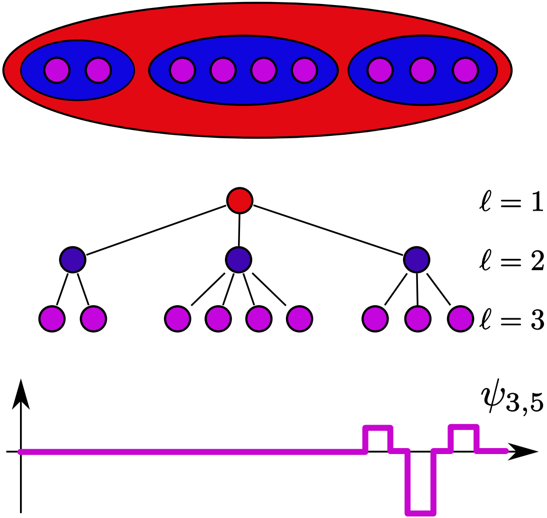


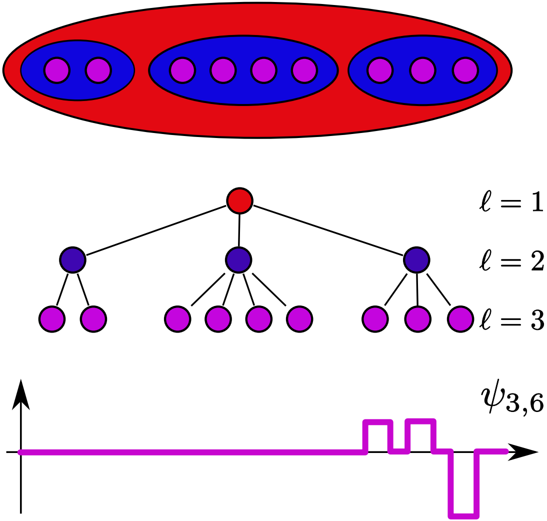
# Partition Tree ⇒ Haar-like basis

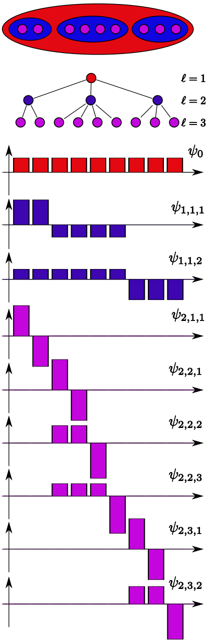


# Partition Tree $\Rightarrow$ Haar-like basis



# Partition Tree $\Rightarrow$ Haar-like basis




$\ell = 1$
$\ell = 2$
$\ell = 3$
$\psi_0$
$\psi_{1,1,1}$
$\psi_{1,1,2}$
$\psi_{2,1,1}$
$\psi_{2,2,1}$
$\psi_{2,2,2}$
$\psi_{2,2,3}$
$\psi_{2,3,1}$
$\psi_{2,3,2}$

# Main Message

Any **Balanced Partition Tree,whose metric preserves smoothness in** $W$, yields an extremely simple WaveletBasis

# $f$ smooth in tree metric $\iff$ coefs decay

## How to define smoothness

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

## Theorem

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

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

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

- If the tree is *balanced* $\Rightarrow$ $|offspring\ folder| \leq q \cdot |parent\ folder|$
- Then $|f(x) - f(y)| \leq C \cdot d(x, y)^{\alpha} \iff |\langle f, \psi_{\ell,k} \rangle| \leq \tilde{C} \cdot q^{\ell\left(\alpha + \frac{1}{2}\right)}$ (for classical Haar, $q = \frac{1}{2}$).

# $f$ smooth in tree metric $\iff$ coefs decay

## How to define smoothness

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

## Theorem

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

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

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

- If the tree is *balanced* $\Rightarrow$ $|$*offspring folder*$| \leq q \cdot |$*parent folder*$|$
- Then $|f(x) - f(y)| \leq C \cdot d(x,y)^{\alpha} \iff |\langle f, \psi_{\ell,k} \rangle| \leq \tilde{C} \cdot q^{\ell\left(\alpha + \frac{1}{2}\right)}$ (for classical Haar, $q = \frac{1}{2}$).

# $f$ smooth in tree metric $\iff$ coefs decay

## How to define smoothness

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

## Theorem

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

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

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

- If the tree is *balanced* $\Rightarrow$ $|offspring\ folder| \leq q \cdot |parent\ folder|$
- 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 natrual tree (ultra-) metric $d$
- Measure smoothness of $f : X \to \mathbb{R}$ w.r.t $d$

**Theorem**

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

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

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

- If the tree is *balanced* $\Rightarrow$ $|offspring\ folder| \leq q \cdot |parent\ folder|$
- 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 natrual tree (ultra-) metric $d$
- Measure smoothness of $f : X \to \mathbb{R}$ w.r.t $d$

## Theorem

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

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

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

- If the tree is *balanced* $\Rightarrow$ $|offspring\ folder| \leq q \cdot |parent\ folder|$
- Then
  $|f(x) - f(y)| \leq C \cdot d(x,y)^{\alpha} \iff |\langle f, \psi_{\ell,k} \rangle| \leq \tilde{C} \cdot q^{\ell\left(\alpha+\frac{1}{2}\right)}$
  (for classical Haar, $q = \frac{1}{2}$).

# Results

## Results

1. Any partition tree on $X$ induces "wavelet" *Haar-like bases* $\surd$
2. "Balanced" tree $\Rightarrow$ $f$ smooth **equals** fast coefficient decay $\surd$
3. Application to semi-supervised learning
4. 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 $\hat{f} = \sum \widehat{\langle f, \psi_{\ell,i} \rangle} \psi_{\ell,i}$

- 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 $\hat{f} = \sum \widehat{\langle f, \psi_{\ell,i} \rangle} \psi_{\ell,i}$

- 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 $\hat{f} = \sum \widehat{\langle f, \psi_{\ell,i} \rangle} \psi_{\ell,i}$

- 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 $\hat{f} = \sum \widehat{\langle f\,,\,\psi_{\ell,i}\rangle}\,\psi_{\ell,i}$

- 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 $\hat{f} = \sum \widehat{\langle f\,,\,\psi_{\ell,i}\rangle}\,\psi_{\ell,i}$

- 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 $\hat{f} = \sum \langle \widehat{f\,,\,\psi_{\ell,i}} \rangle\, \psi_{\ell,i}$

- 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

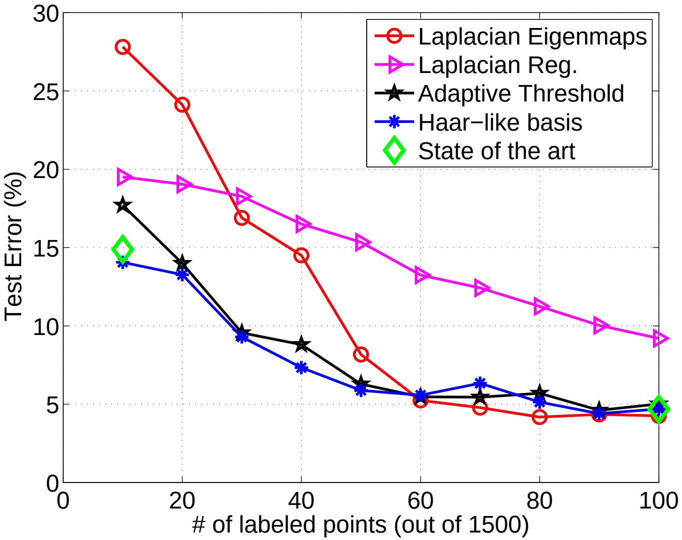
# Application: Semi supervised learning

## Classification/Regression with Haar-like basis

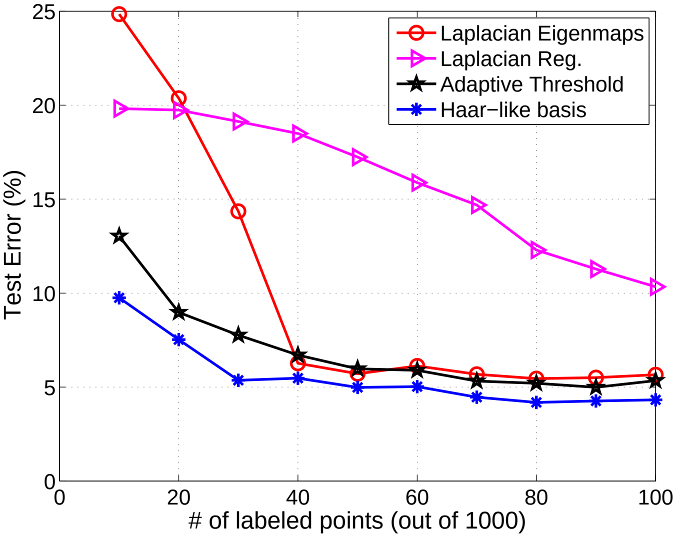
- Task: Given values of smooth $f$ on $\tilde{X} \subset X$, extend $f$ to $X$.
- Step 1: Build a partition tree s.t. $f$ is smooth w.r.t tree metric
- Step 2: Construct a Haar-like basis $\{\psi_{\ell,i}\}$
- Step 3: Estimate $\hat{f} = \sum \widehat{\langle f , \psi_{\ell,i} \rangle} \psi_{\ell,i}$

- 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



# MNIST Digits 8 vs. {3,4,5,7}



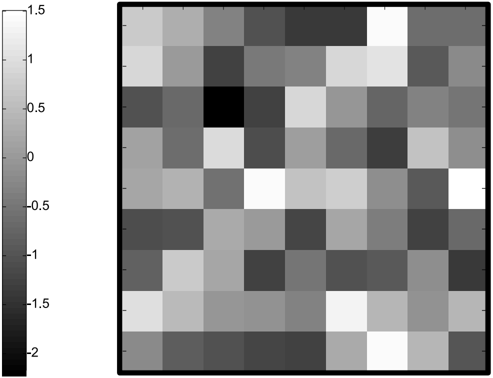
# Results

## Results

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

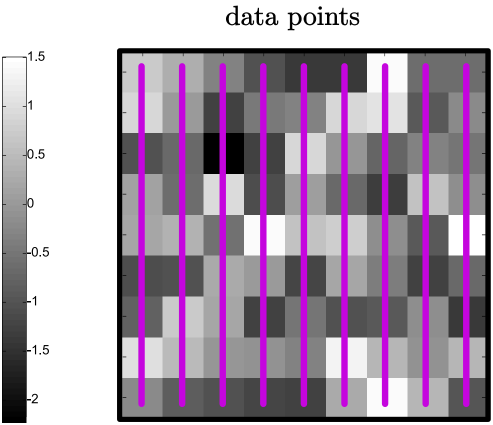
# Tensor product of Haar-like bases

**Coifman and G, Harmonic Analysis of Digital Data Bases (2010)**



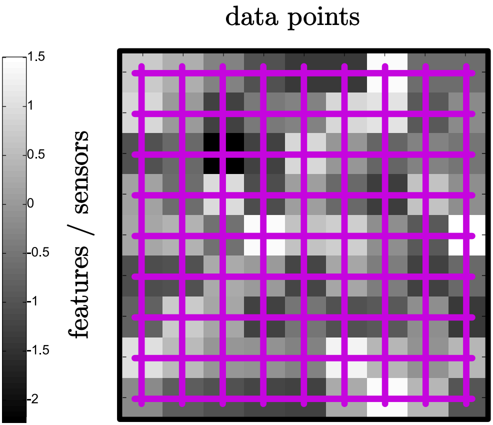
# Tensor product of Haar-like bases

**Coifman and G, Harmonic Analysis of Digital Data Bases (2010)**



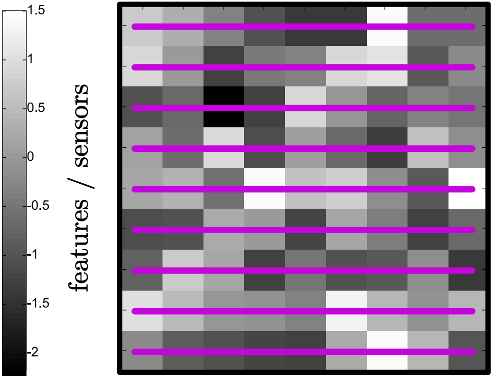
# Tensor product of Haar-like bases

**Coifman and G, Harmonic Analysis of Digital Data Bases (2010)**



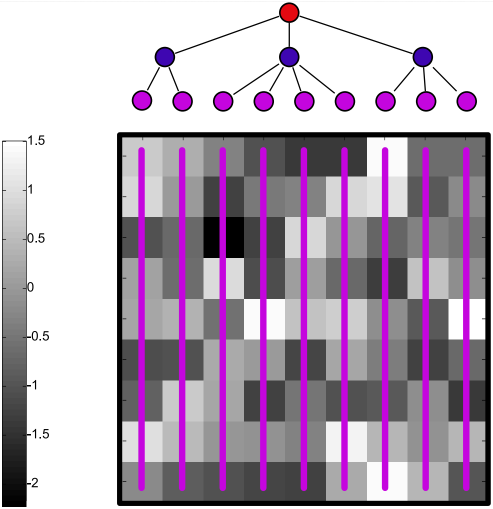
# Tensor product of Haar-like bases

**Coifman and G, Harmonic Analysis of Digital Data Bases (2010)**



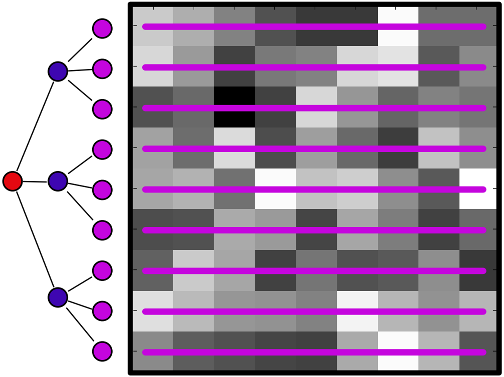
# Tensor product of Haar-like bases

**Coifman and G, Harmonic Analysis of Digital Data Bases (2010)**



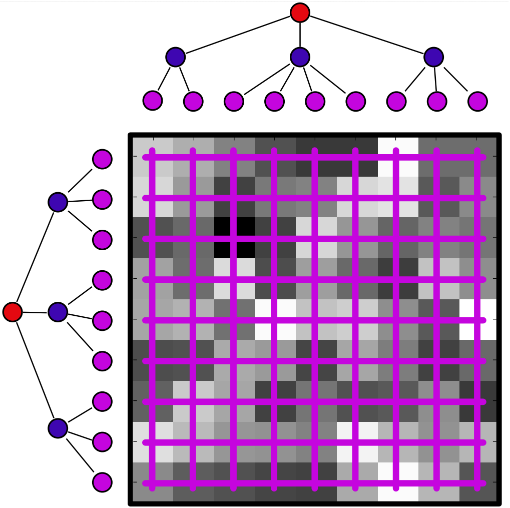
# Tensor product of Haar-like bases

**Coifman and G, Harmonic Analysis of Digital Data Bases (2010)**



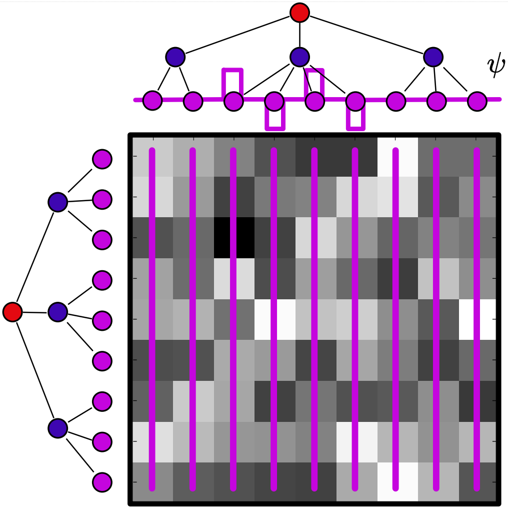
# Tensor product of Haar-like bases

**Coifman and G, Harmonic Analysis of Digital Data Bases (2010)**



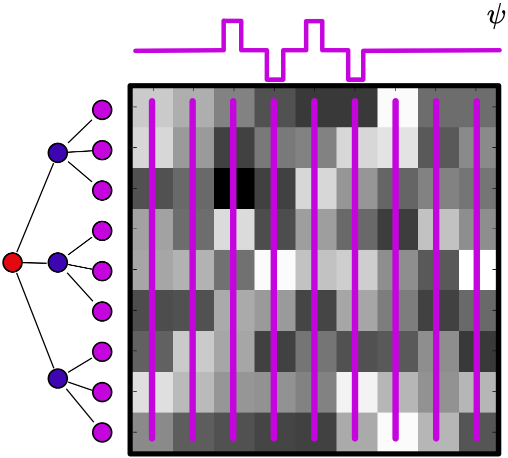
# Tensor product of Haar-like bases

**Coifman and G, Harmonic Analysis of Digital Data Bases (2010)**



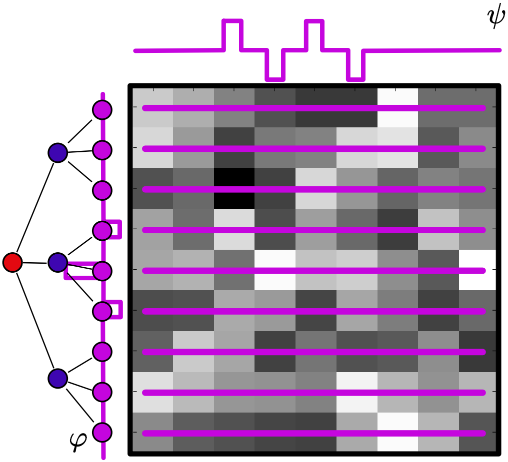
# Tensor product of Haar-like bases

**Coifman and G, Harmonic Analysis of Digital Data Bases (2010)**



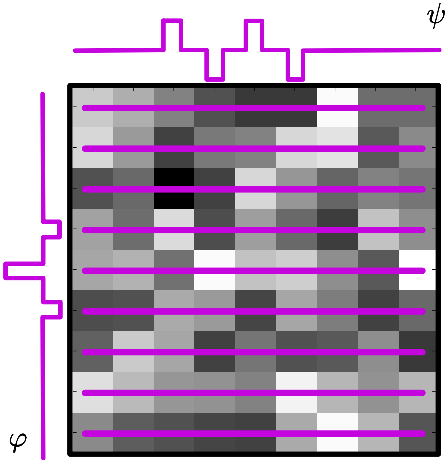
# Tensor product of Haar-like bases

**Coifman and G, Harmonic Analysis of Digital Data Bases (2010)**



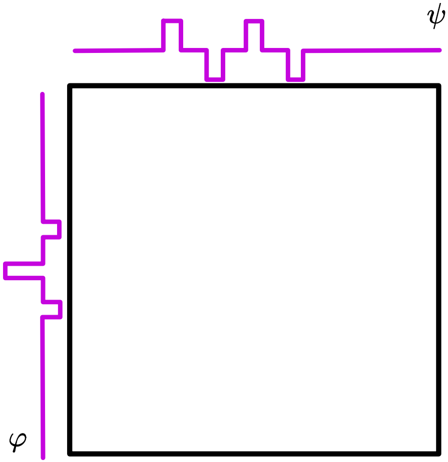
# Tensor product of Haar-like bases

**Coifman and G, Harmonic Analysis of Digital Data Bases (2010)**



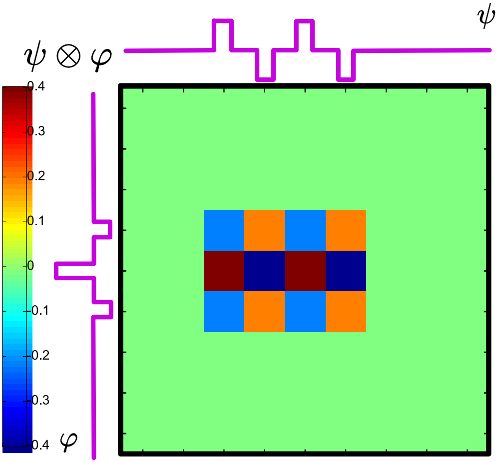
# Tensor product of Haar-like bases

**Coifman and G, Harmonic Analysis of Digital Data Bases (2010)**



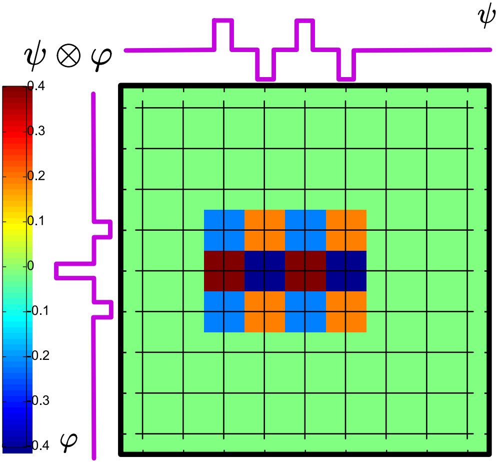
# Tensor product of Haar-like bases

**Coifman and G, Harmonic Analysis of Digital Data Bases (2010)**

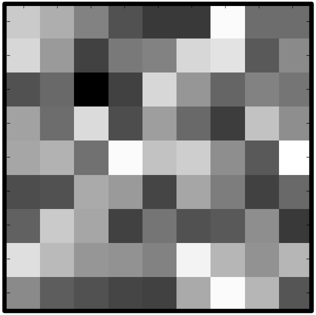


# Tensor product of Haar-like bases

**Coifman and G, Harmonic Analysis of Digital Data Bases (2010)**



# Tensor product of Haar-like bases



$$= \sum_{i,j} a_{ij} \psi_i \otimes \varphi_j$$

# The Big Picture

**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 $X \Longleftrightarrow$ bases for $\{f : X \to \mathbb{R}\}$ ("Spaces of Homogeneous Type")

# The Big Picture

**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 $\{f : X \to \mathbb{R}\}$ ("Spaces of Homogeneous Type")

# The Big Picture

**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 $\{f : X \to \mathbb{R}\}$ ("Spaces of Homogeneous Type")

# The Big Picture

**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 $\{f : X \to \mathbb{R}\}$ ("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
- G, Nadler and Coifman, Inference using Haar-like wavelets, preprint (2010)
- Coifman and G, *Harmonic analysis of digital data bases*, to appear in **Wavelets: Old and New Perspectives**, Springer (2010)
- Coifman and Weiss, *Extensions of Hardy spaces and their use in analysis*, **Bul. Of the AMS**, 83 #4, pp. 569-645 (1977)
- Singh, Nowak and Calderbank, Detecting weak but hierarchically-structured patterns in networks, Proceedings of AISTATS 2010