Image Denoising using Non-local Wavelet Basis

L. Jacques[1], D. K. Hammond[2], P. Vandergheynst[3]

[1] ICTEAM/ELEN [2] Neuroinformatics Center, Univ. Oregon, USA

[3] LTS2/EPFL, Lausanne, Switzerland

UCL Université catholique de Louvain epl TELE

Outline

* Introduction
* Noising Model
* What's Sparsity?
* Denoising: Principle
* Towards Non-locality
  * Non-local Graphs
  * Non-local Spectral Decomposition
* Graph Wavelets
* Non-local Wavelets and Denoising

Introduction

* Torrents of scientific data:

  images, signals, volumes, ADN, graphs, ...

* Sensors under "hard constraints":
  * few detected photons (gamma, rayon X, ...)
  * low transmission, limited longevity, ...
* Consequences:
  * noisy or missing observations, saturation, blur, ...
  * compressed and digitized data, ...
* Problem of interest here: Image Denoising

  (one data restoration method)

"Noising" an image:

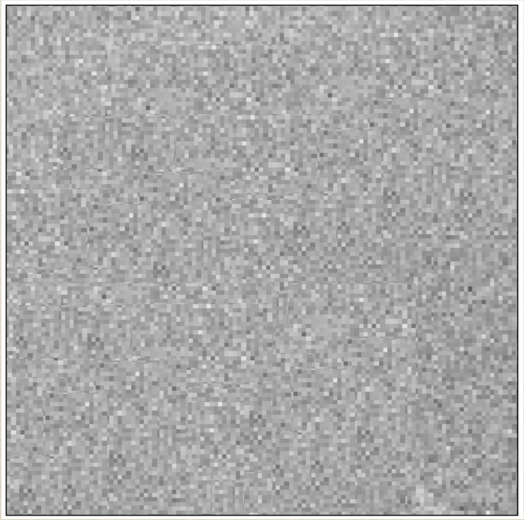
- Simple Model:
  - Noisy image f of N pixels:

(or other distribution)

$$f = f_{\mathrm{pure}} + \varepsilon, \quad \varepsilon(\mathbf{m}) \sim \mathcal{N}(0, \sigma^2)$$

$=$

$+$



f $\qquad$ f_{pure} $\qquad$ ε

structured $\qquad$ not structured

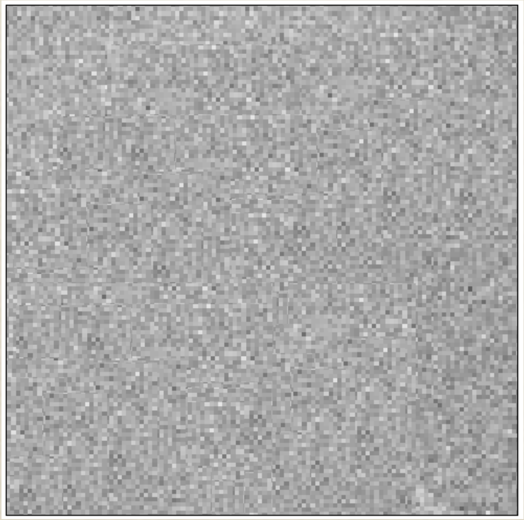
"Noising" an image:

* Simple Model:
  * Noisy image f of N pixels:

$$f = f_{\mathrm{pure}} + \varepsilon, \quad \varepsilon(\mathbf{m}) \sim \mathcal{N}(0, \sigma^2)$$

$=$

$+$



f $\qquad$ f_{pure} $\qquad$ ε

$\Longrightarrow$ Structure $=$ "Sparsity"

What's sparsity?

- Hypothesis: an image (or any signal) can be decomposed in a "sparsity basis" Ψ with few non-zero elements α :

$$f(\mathbf{n}) = \sum_{k=1}^{d} \alpha_k \, \psi_k(\mathbf{n}), \quad \mathbf{n} = (n,m), \;\; \Psi = \{\psi_j : 1 \leqslant j \leqslant d\}$$

- Ψ can be an ONB (e.g. Fourier, Wavelets) or a dictionary

$115{\times}147$ $\approx$ 16k pixels | 10 sparse | 25 sparse | 50 sparse | 250 sparse | 1000 sparse

f

$\boxed{\text{16k pixels} \Rightarrow \text{1k values}}$ $\boxed{f \text{ is sparsely approximable}}$

DeNoising: Principle

- **Signal is sparse (or compressible) in $\boldsymbol{\Psi}$, noise is not!**

- For additive Gaussian Noise model, $(\mathbf{m} = (m, n))$

$$f = f_{\rm pure} + \varepsilon, \quad \varepsilon(\mathbf{m}) \sim \mathcal{N}(0, \sigma^2)$$

- For a ONB: $\Psi = \{\psi_j : 1 \leqslant j \leqslant N\}$

bounded amplitude: $|\langle \varepsilon, \psi_j \rangle| \propto \sigma\sqrt{N}$

$$\alpha_j = \langle f, \psi_j \rangle = \langle f_{\rm pure}, \psi_j \rangle + \langle \varepsilon, \psi_j \rangle$$

with $\langle u, v \rangle = \sum_{\mathbf{m}} u(\mathbf{m}) v(\mathbf{m})$

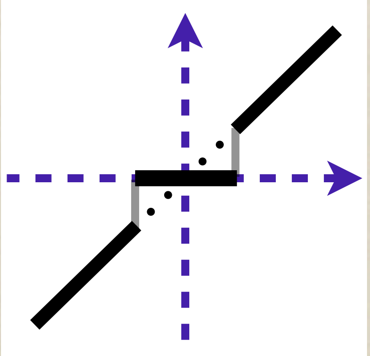
- Shrinkage, or *thresholding:*

$$\boxed{\alpha_j \;\to\; \mathcal{T}(\alpha_j, T(\sigma))}$$

and reconstruct!

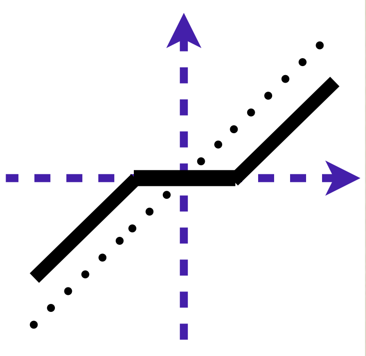
[Donoho, Johnstone, 98]

$\mathcal{T}_h(u, T)$



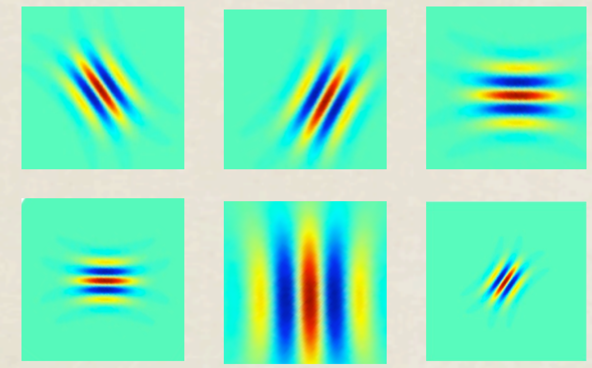
hard

$\mathcal{T}_s(u, T)$



soft

Denoising Example:

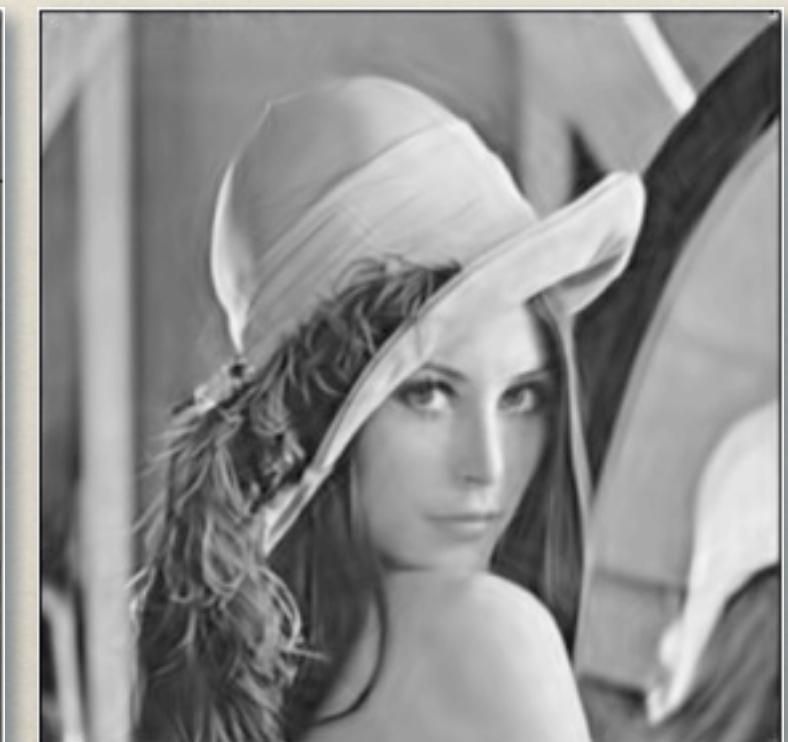


* Sparsity basis = directional wavelets
  [Antoine, Vandergheynst, Jacques]

(or curvelets, [Candes, Donoho, Starck, Demanet] or bandlets, [Mallat, Peyré] or "∗lets") [ask L. Duval, WITS]

Original

Noisy image, SNR: 22.13 dB



Reconstructed, SNR: 32.02 dB

* But most sparsity bases are *local* (in space)

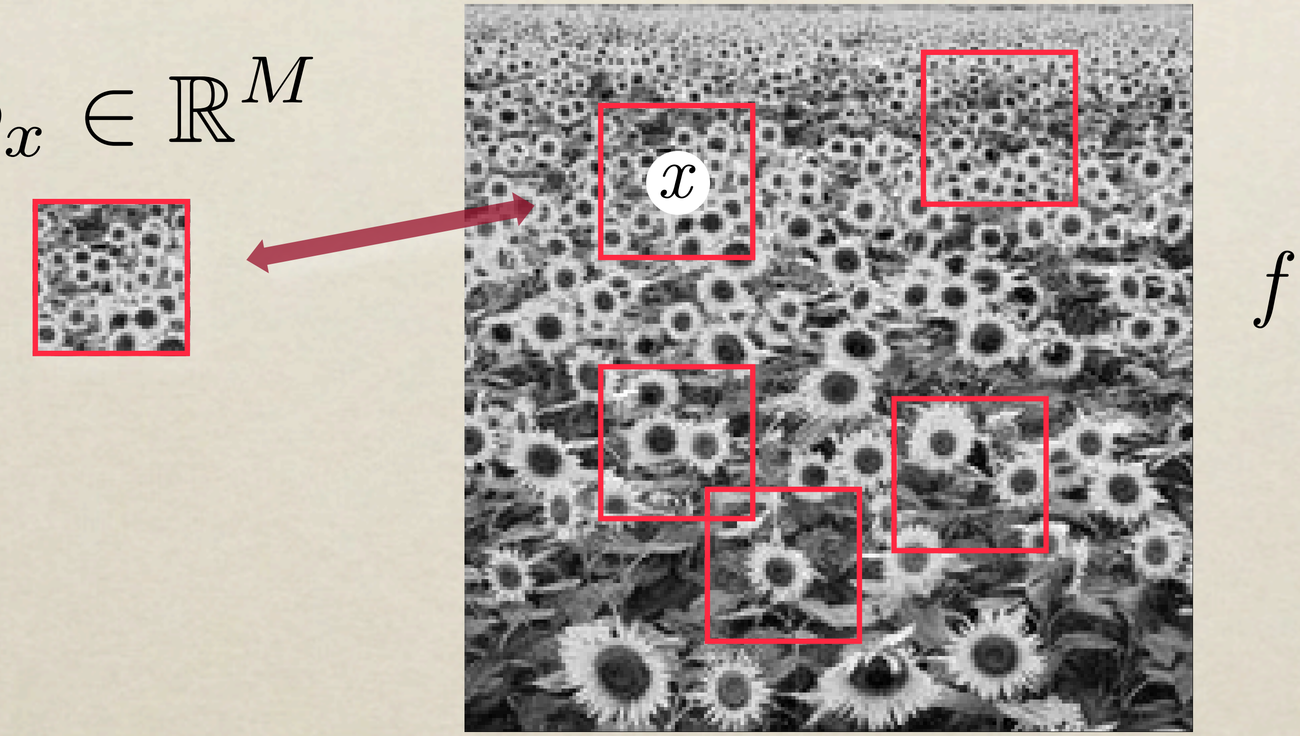
Towards non-locality ...

Towards non-locality ...

Non-local Means [Buades, Coll, Morel, 05]

- Given an image f of N pixels, define *patches* of M pixels located on x :

$$p_x \in \mathbb{R}^M$$

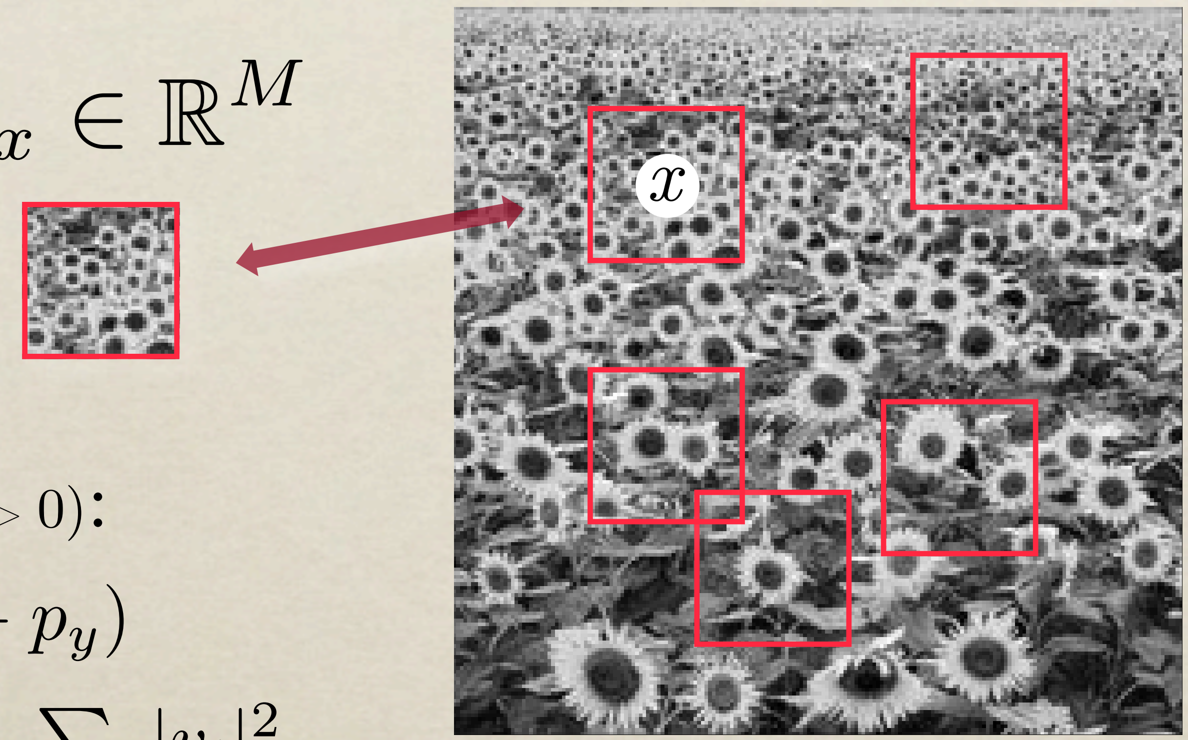


Towards non-locality ...

Non-local Means [Buades, Coll, Morel, 05]

* Given an image f of N pixels, define *patches* of M pixels located on x :

$$p_x \in \mathbb{R}^M$$



f

* Define similarity (for $r > 0$):

$$w(x,y) = G_r(p_x - p_y)$$

$$G_r(v) := \exp - \tfrac{1}{2\,r^2} \textstyle\sum_i |v_i|^2$$

* NL Means:

$$[M_r f](x) \;=\; \frac{1}{\sum_y w(x,y)} \sum_y w(x,y) f(y)$$

Towards non-locality ...

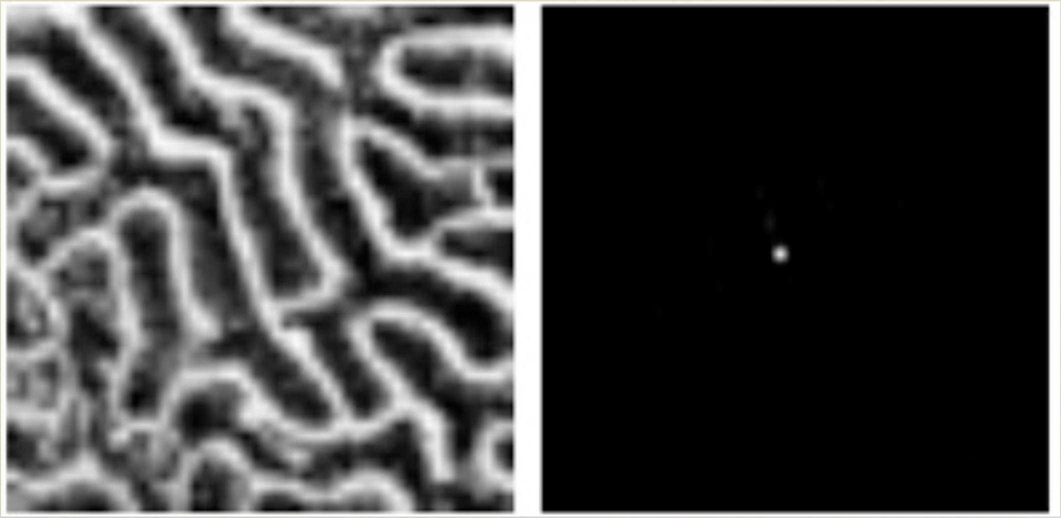
* Example: Non-local Diffusion [Source: G. Peyré, 08]

Original image:
computing similarity
$w(x,y)$

Towards non-locality ...

* Example: Non-local Diffusion [Source: G. Peyré, 08]

Original image:
computing similarity
$w(x,y)$

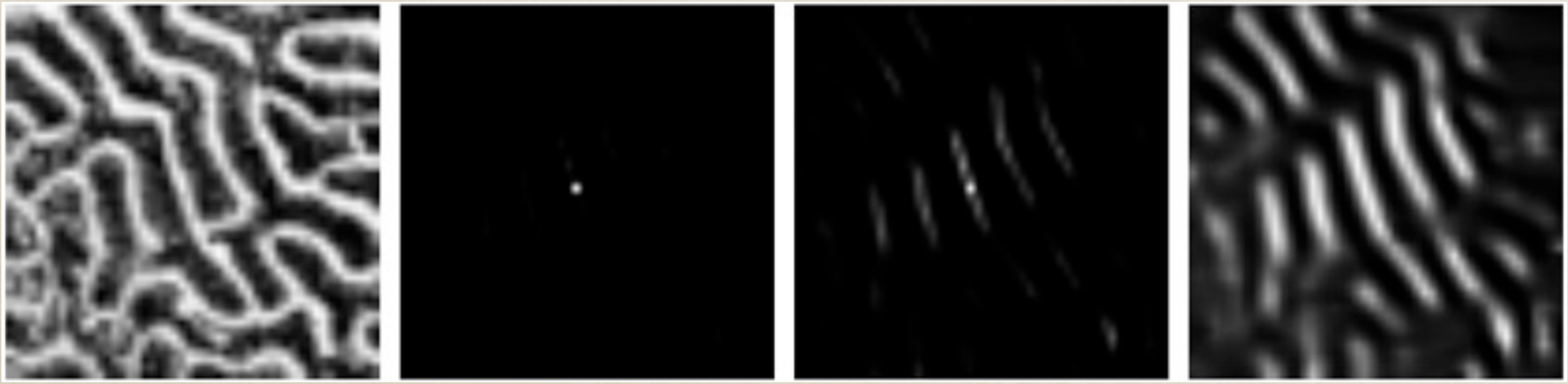


$g = \delta(x - c)$

Towards non-locality ...

- Example: Non-local Diffusion [Source: G. Peyré, 08]

Original image:
computing similarity
$w(x,y)$

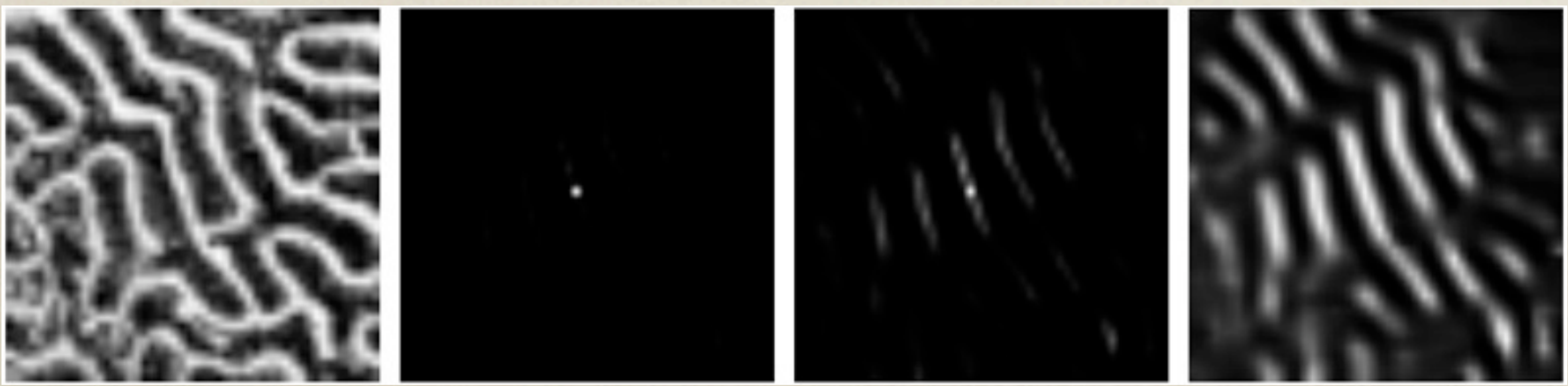


$g = \delta(x - c)$ $\quad M_4\, g$ $\quad M_8\, g$

Towards non-locality ...

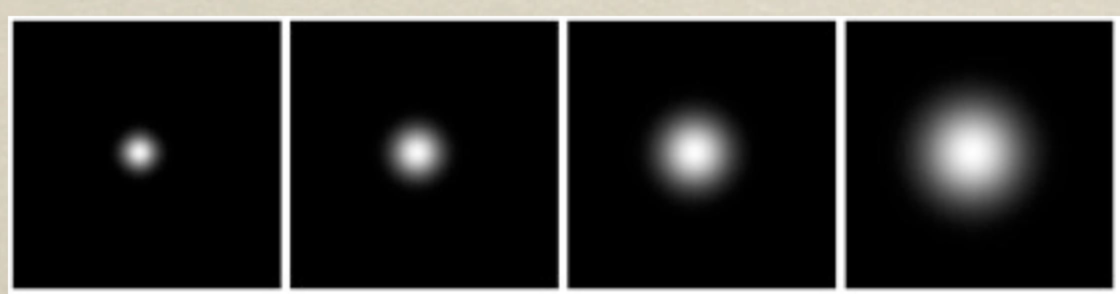
* Example: Non-local Diffusion [Source: G. Peyré, 08]

Original image:
computing similarity
$w(x,y)$



$g = \delta(x - c)$ $\quad M_4\, g \quad$ $M_8\, g$

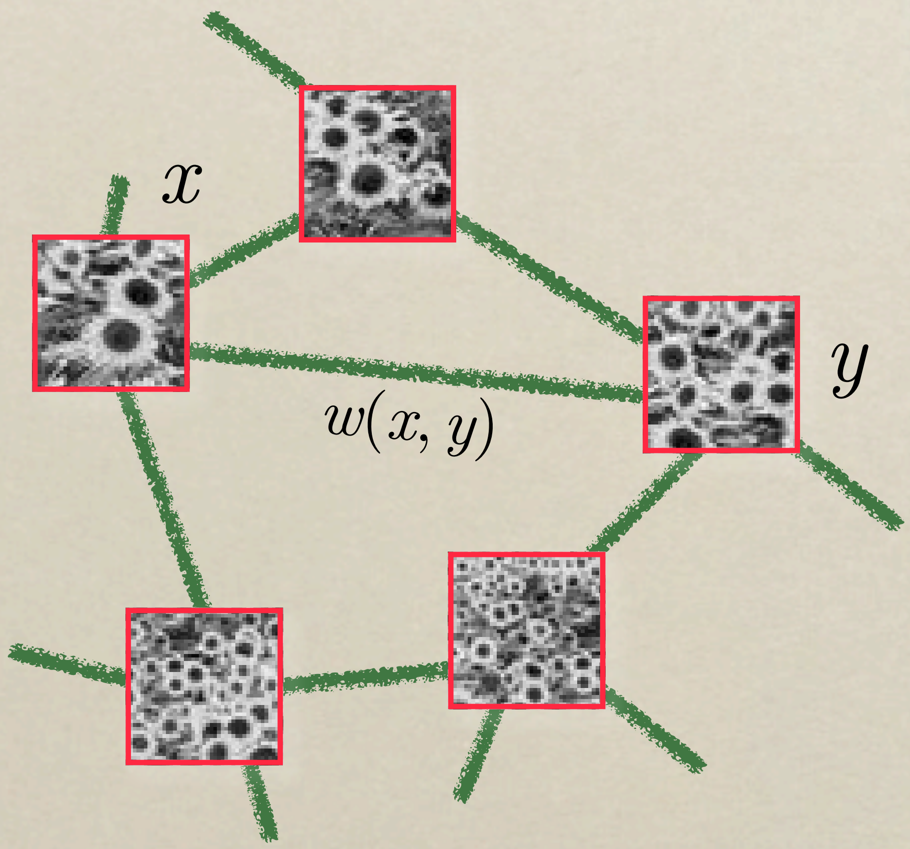
Same for Local diffusion:



Non-local Graph

* Behind NL Means, a graph $G = (V, E, w)$ with:

$$V = \{x \in \text{im. dom.}\},\ E \subset V \times V,\ w : E \to \mathbb{R}$$



Non-local Graph

* Behind NL Means, a graph $G = (V,\, E,\, w)$ with:

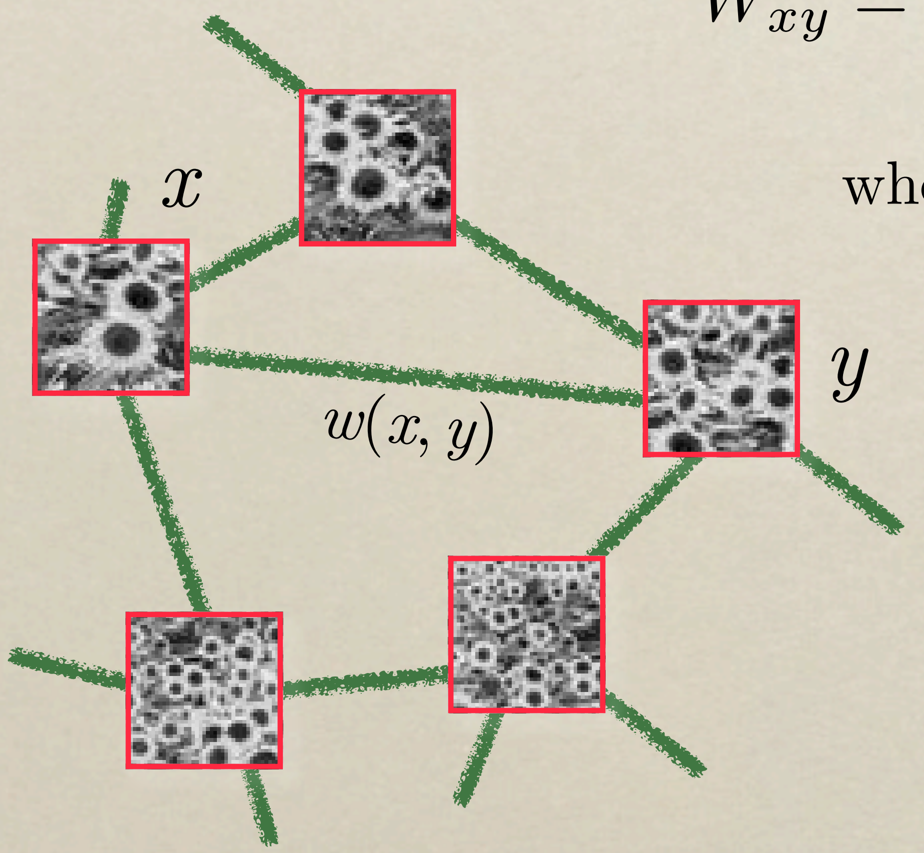
$$V = \{x \in \text{im. dom.}\},\ E \subset V \times V,\ w : E \to \mathbb{R}$$

* Similarity is linked to the *connectivity* matrix W:

$$W_{xy} = \begin{cases} w(x,y) & \text{if } (x,y) \in E \\ 0 & \text{else.} \end{cases}$$

where: E is defined generally by thresholding w, and

$$M_r f = D^{-1} W f,\ D_{xy} = \delta_{xy} \sum_y W_{xy}$$



Non-local Spectral Decomposition

* Connectivity induces a Laplacian: (with $D_{ij} = \delta_{ij} \sum_k W_{ik}$)

$$L = D - W \quad \text{or} \quad L^{\text{norm}} = \text{Id} - D^{-1/2} W D^{-1/2}$$

* Eigenvectors of L, i.e. the set

$$\Gamma = \{\gamma_\ell \in \mathbb{R}^N : L\,\gamma_\ell = \lambda_\ell \gamma_\ell,\ 0 \leq \ell \leq N-1\},$$

with $\lambda_0 = 0 \leqslant \lambda_1 \leqslant \cdots \leqslant \lambda_{N-1}$, form an ONB for functions on V, i.e. $\mathbb{R}^N$

Example: Euclidean Graph $\leftrightarrow$ 2-D Fourier

* By analogy: The Fourier transform on a graph:

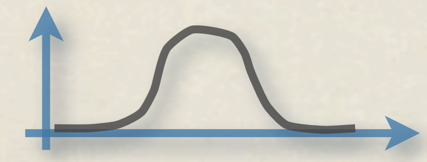
$$\hat{f}_\ell = \langle \gamma_\ell, f \rangle = \sum_i \gamma_\ell^*(i)\, f(i) \in \mathbb{C}, \quad \text{(forward)}$$

$$f(i) = \sum_\ell \hat{f}_\ell\, \gamma_\ell(i), \quad \text{(inverse)}$$

Graph Wavelets

- Use the Graph Wavelets [Hammond, Gribonval, Vandergheynst, 09]

* Given a kernel $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, $\approx$ band pass



* Extension of g to the L operator:

$$\begin{aligned} T_g^t &:= g(tL) && (\Leftrightarrow 1/t \text{ acts as a scale}) \\ \widehat{g(tL)f} &:= g(t\lambda_\ell)\hat{f}_\ell && (\Leftrightarrow \text{convolution by } g \text{ in freq.}) \\ [g(tL)\, f](i) &= \sum_\ell g(t\lambda_\ell)\hat{f}_\ell\, \gamma_\ell(i) && (\Leftrightarrow \text{filtering of } f\,) \end{aligned}$$

* Wavelet? a wavelet localized on the graph node j is

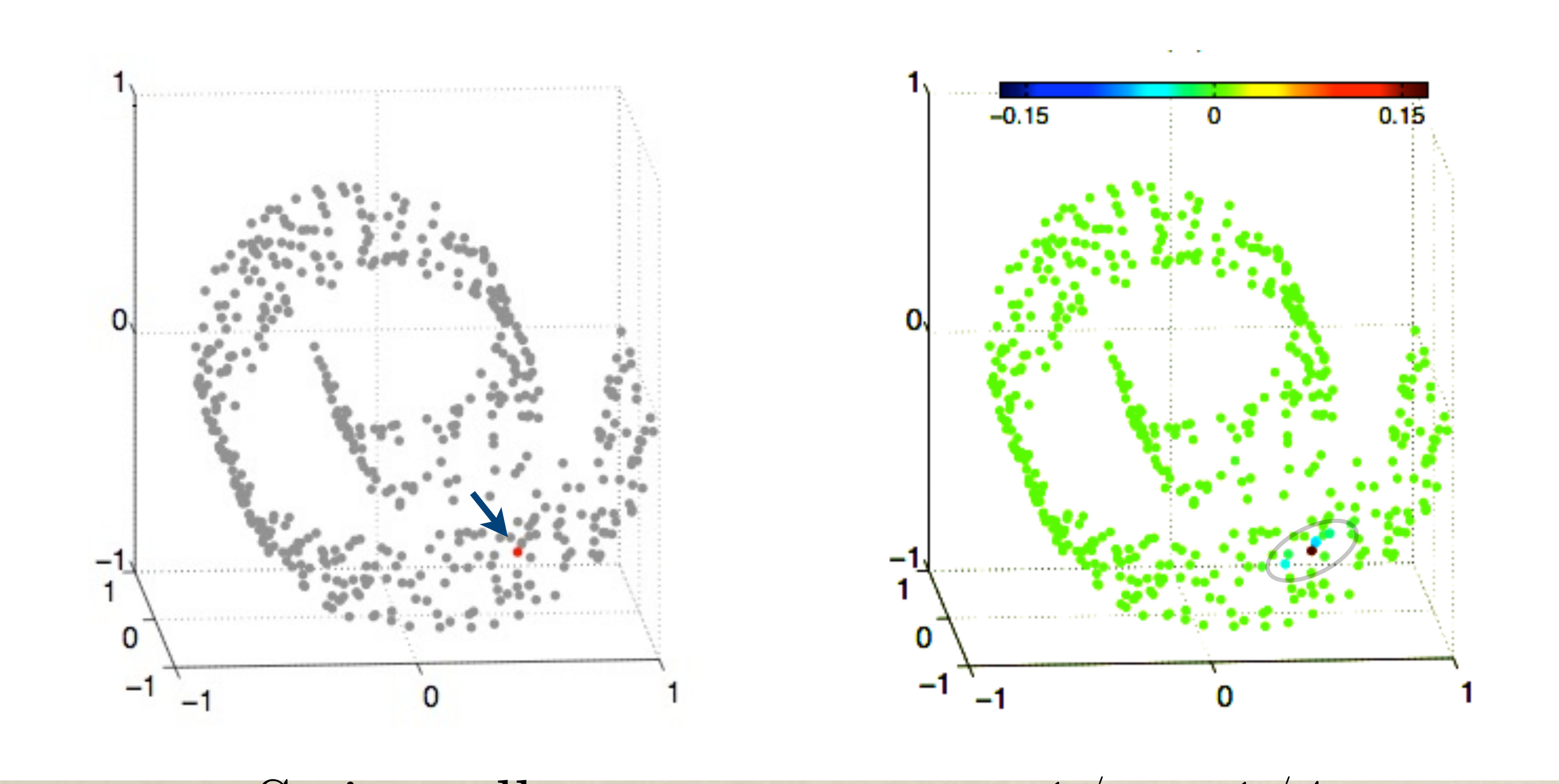
$$\psi_{t,j}(i) = T_g^t\, \delta_j(i)$$

* and the Wavelet Transform:

$$W_f(t,j) = \langle \psi_{t,j}, f \rangle = T_g^t f$$

Graph Wavelets

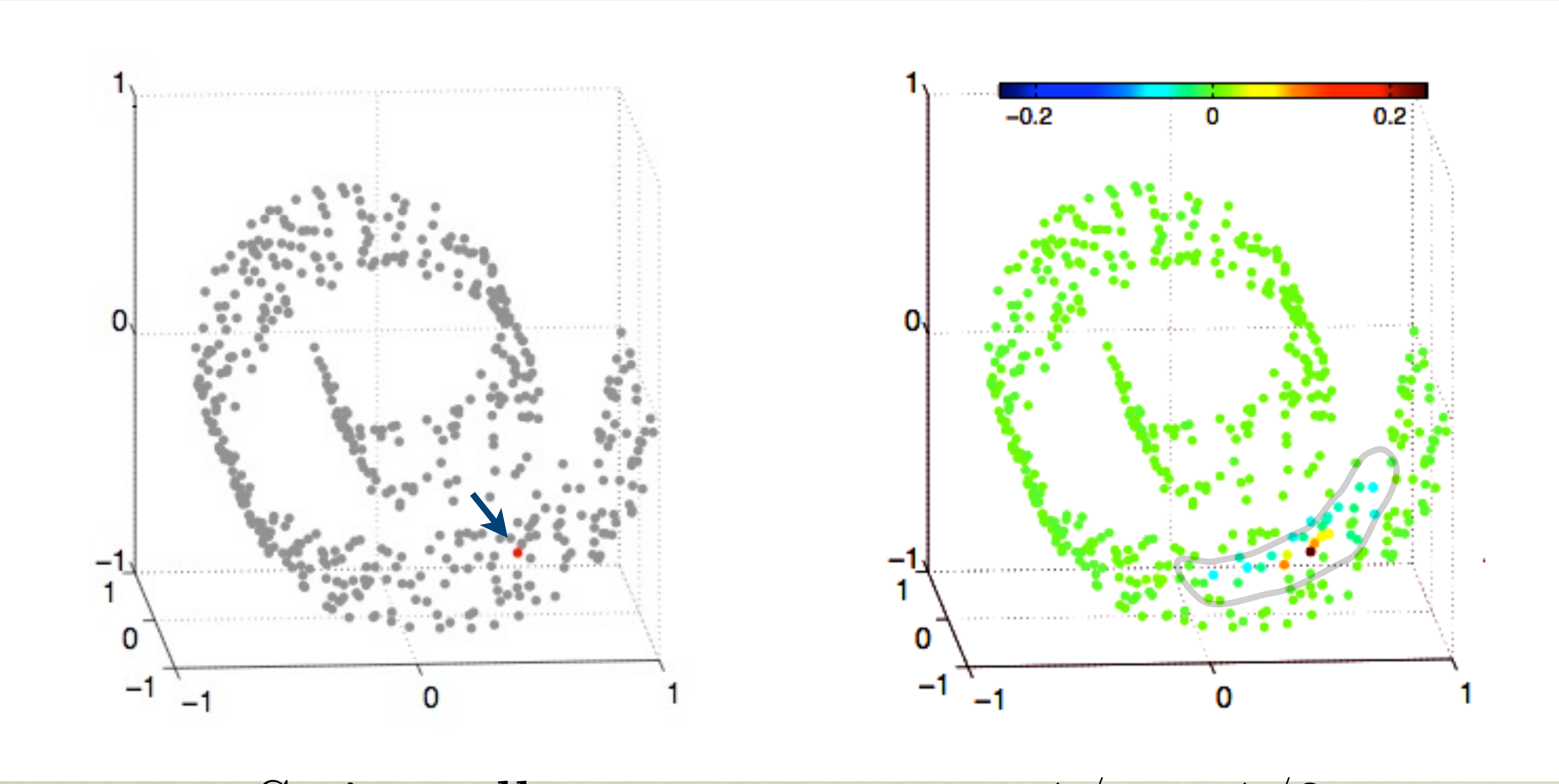
* Wavelet localization: (also proved theoretically)



Swiss roll $\qquad\qquad$ $1/t = 1/4$

Graph Wavelets

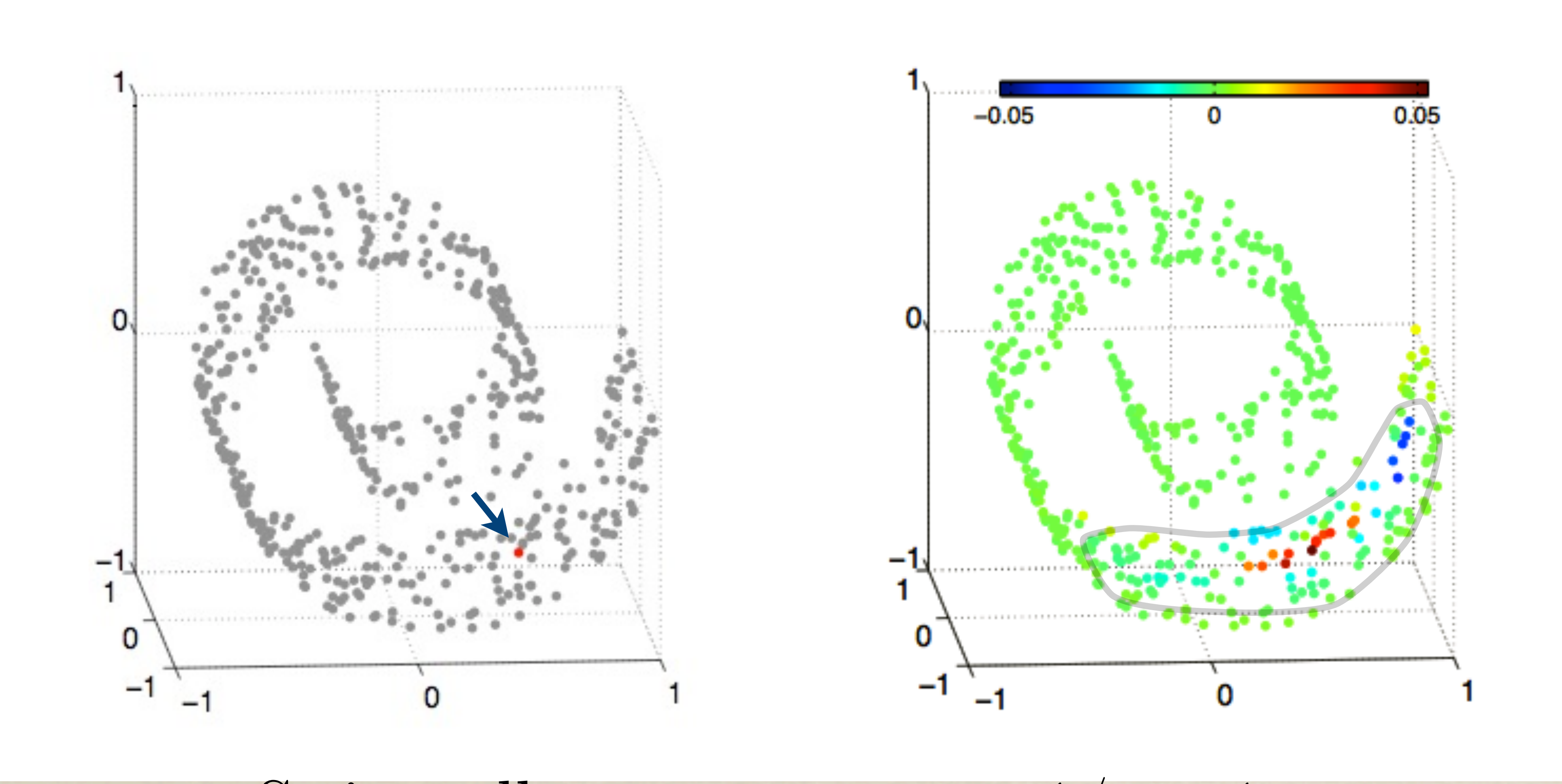
* ✳ Wavelet localization: (also proved theoretically)



Swiss roll $\qquad\qquad$ $1/t = 1/2$

Graph Wavelets

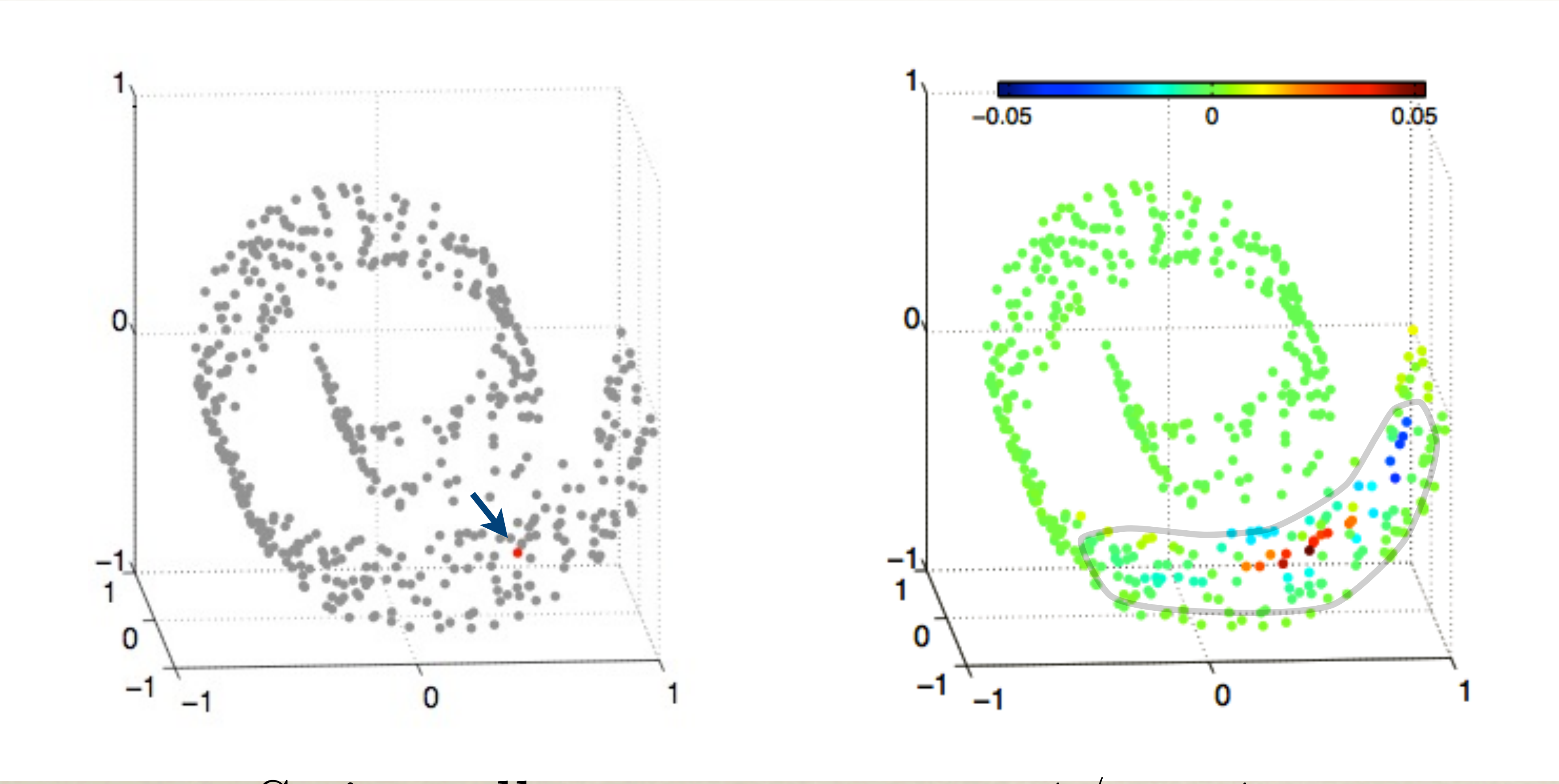
* Wavelet localization: (also proved theoretically)



Swiss roll $1/t = 1$

Graph Wavelets

* Wavelet localization: (also proved theoretically)



Swiss roll $\qquad\qquad$ $1/t = 1$

* Fast computation: $g(u) \approx \mathrm{Pol(u)} \;\Leftrightarrow\;$ Apply $(Lf)^n$ (with L very sparse)

e.g. Taylor, Chebyshev

Graph Wavelets

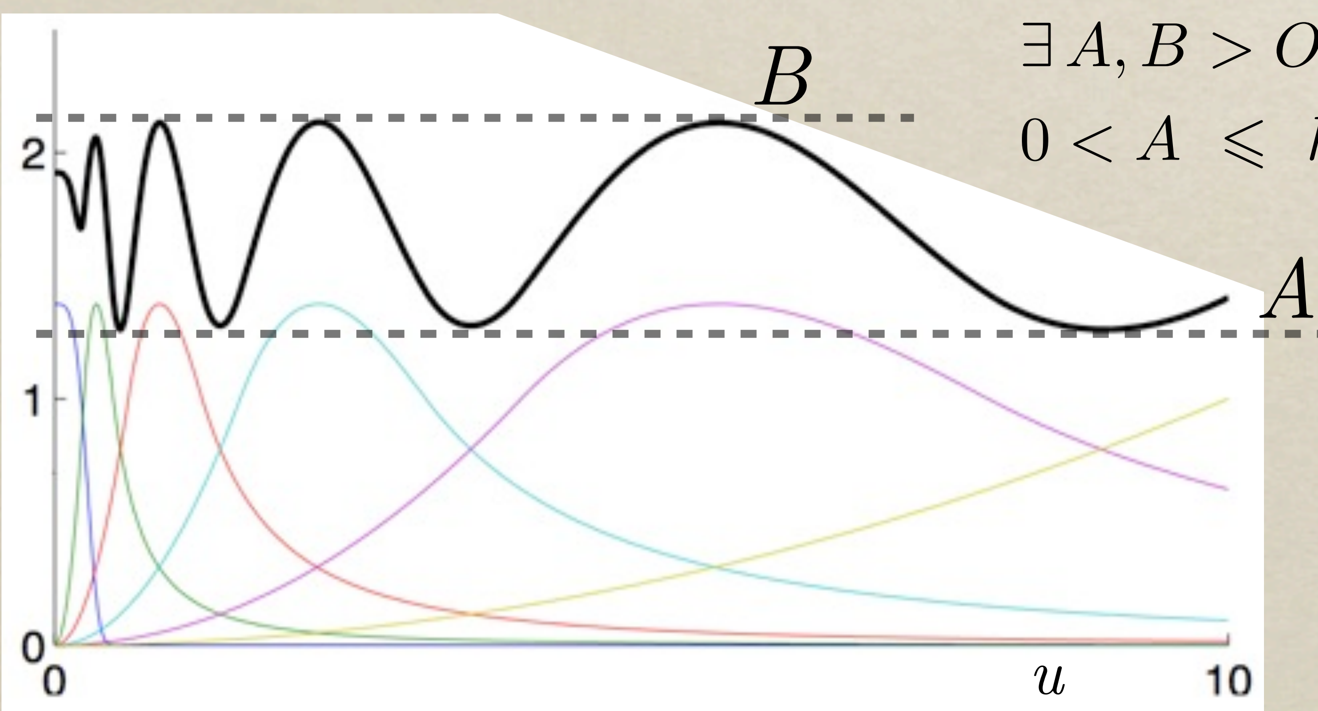
* Reconstruction possible (similar to common CWT):

$$
\begin{aligned}
f(j) &= \langle \gamma_0, f \rangle \gamma_0 + \tfrac{1}{C_g} \textstyle\sum_i \int_0^\infty W_f(t,i)\, \psi_{t,i}(j)\, \tfrac{\mathrm{d}t}{t}, \\
C_g &= \int_0^\infty g^2(u)/u \,\mathrm{d}u
\end{aligned}
$$

* For finite scales, i.e. $\{t_s : 1 \leqslant s \leqslant J,\ t_s < t_{s+1}\}$, reconstruction possible if frame (in scale), i.e.

$$
\exists\, A, B > O,\ \exists h : \mathbb{R}_+ \to \mathbb{R}_+ \text{ (i.e. scaling function)}
$$

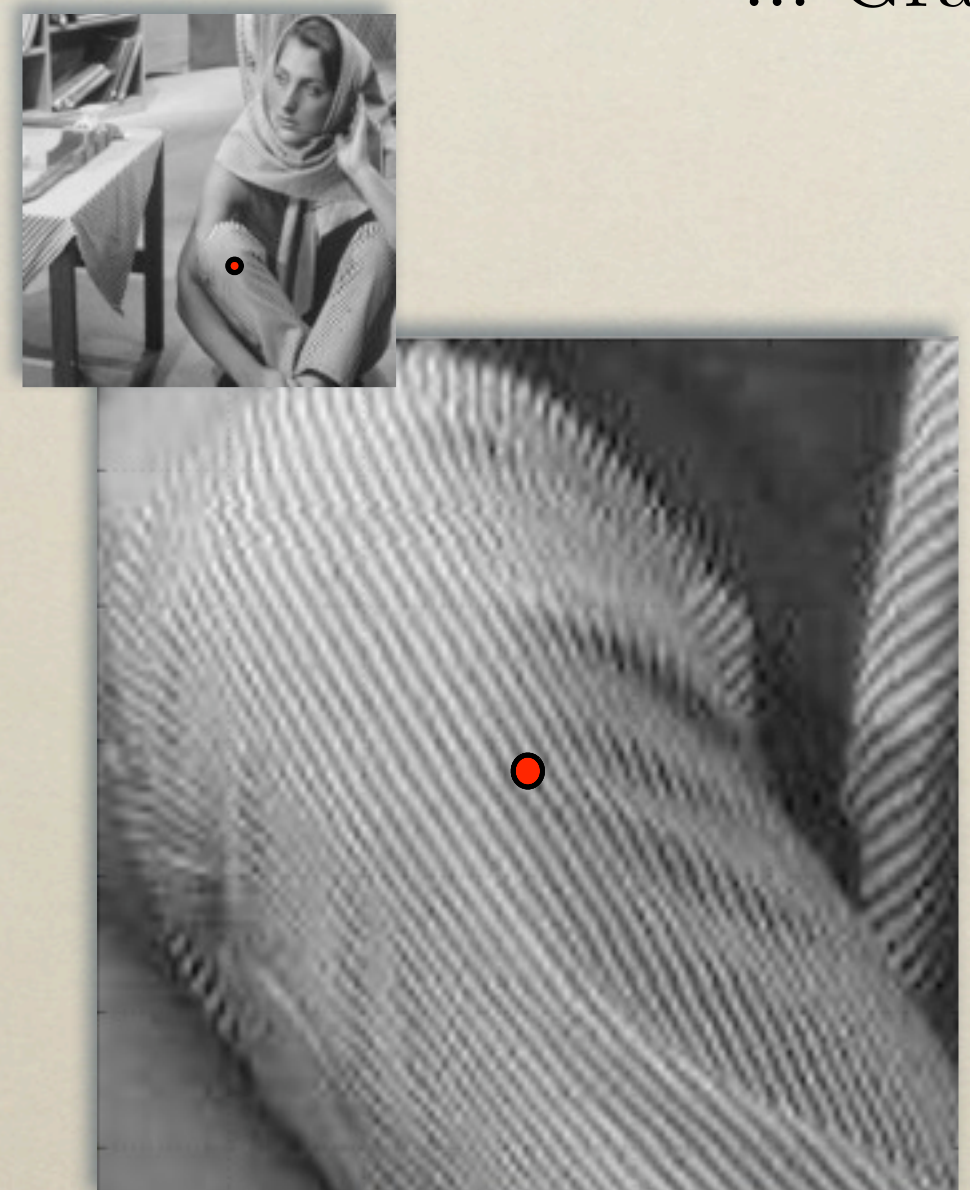
$$
0 < A \leqslant h^2(u) + \textstyle\sum_s g(t_s u)^2 \leqslant B < \infty
$$

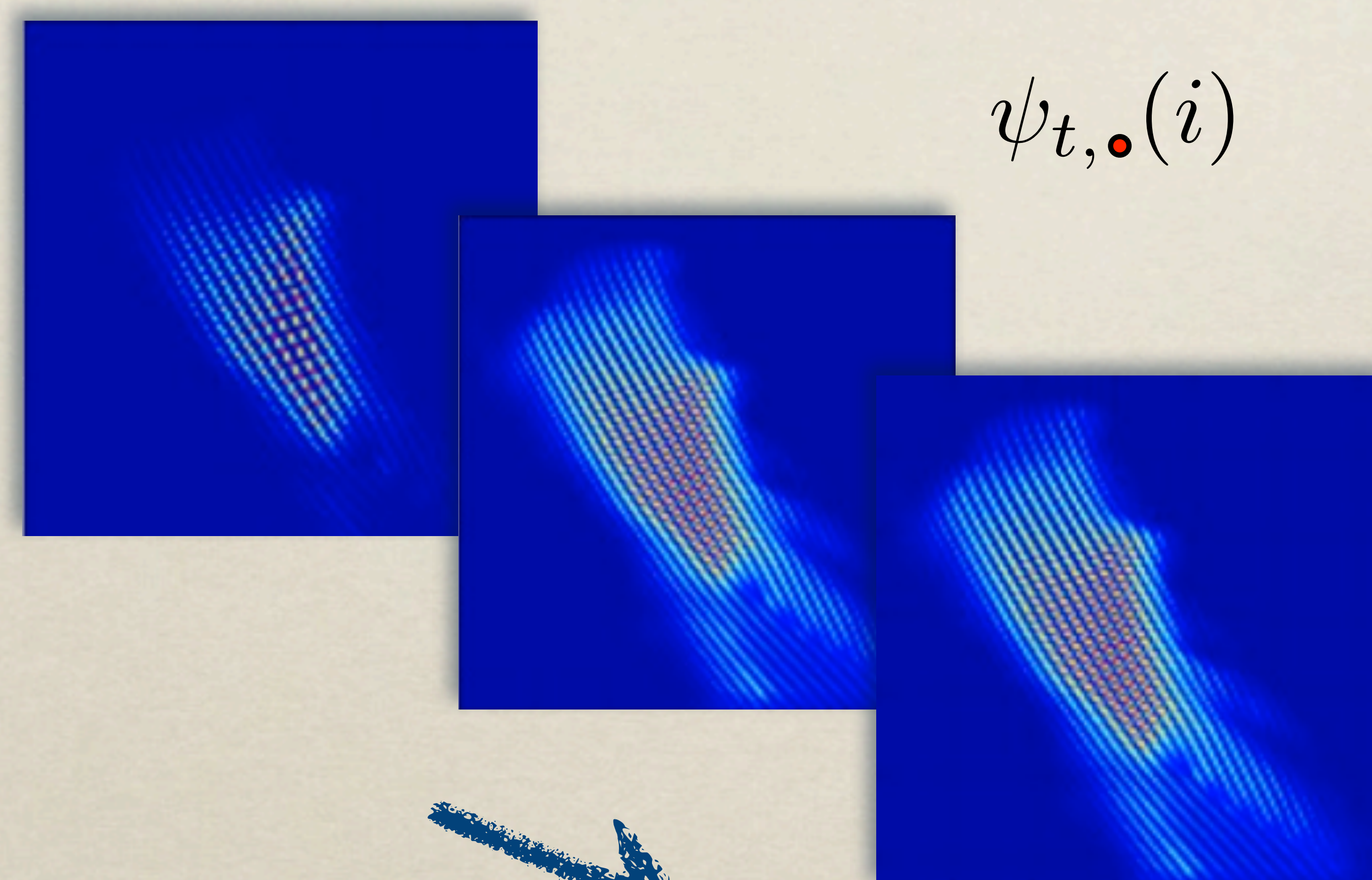


Non-local Wavelet Basis (frame)

* Non-local Wavelets are ...

... Graph Wavelets on Non-Local Graph





$\psi_{t,\bullet}(i)$

$1/t$ increasing

Interest: good *adaptive* sparsity basis

Non-Local Wavelet Denoising

* Of course graph depends on the (pure) image
* But graph not too much affected by noise

- Denoising Method:

1. (pre-filter the image for cleaner graph)
2. Compute $G_f = (V, E, w)$ from $f = f_{\text{pure}} + \varepsilon$
3. Decompose f on the wavelet frame

$$\Psi = \{\phi\} \cup \{\psi_{t_s,j} \in \mathbb{R}^N : t_s = 2^s, 1 \leqslant s \leqslant S, \leqslant j \leqslant N\}$$

5. Threshold the coefficients
6. Reconstruct f^* (by frame inversion, i.e. conjugate gradient)

(remark: more general formulation available: e.g. Iterative Soft Thresholding)

Non-Local Wavelet Denoising

- Boat Image Non-Local Denoising: (preliminary result)

Original

Noisy image, PSNR: 19.62dB

Reconstructed, PSNR: 25.78 dB

- More tests to be done (comparisons, parameters influence, ...)

Conclusion and Perspectives

* Sparsity is key feature of efficient denoising techniques.
* Non-local Wavelets offer a good adaptive sparsity basis.

* Non-local Wavelet Processing is a very promising.
* However:
  * Quantifying the effect on noise on graph structure?
  * Other noise distributions? (e.g. NL means promising for Poisson)
  * Application to other image restorations (e.g. deconvolution, inpainting, texture generation, ...)?

Thank you!

Some References:

- A. Buades, B. Coll, and J.M. Morel. "A Non-Local Algorithm for Image Denoising". Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society Conference on, Vol. 2, 60, 2005.
- D. K. Hammond, P. Vandergheynst, Rémi Gribonval. "Wavelets on graphs via spectral theory", to be submitted.
- D. L. Donoho and I. Johnstone. "Minimax estimation by wavelet shrinkage". Ann. Statist., Vol. 26, pp. 879–921, 1998.
- G. Peyré, "Image Processing with Non-local Spectral Bases", SIAM Multiscale Modeling and Simulation, Vol. 7(2), p.703-730, 2008.
- D. Hammond, K. Raoaroor, L. Jacques, P. Vandergheynst. "Image Denoising with Nonlocal Spectral Graph Wavelets". SIAM Conference on Imaging Science (IS10), April 12-14, 2010 (+ Journal paper in preparation)