Lecture 7

# Introduction to sparsity

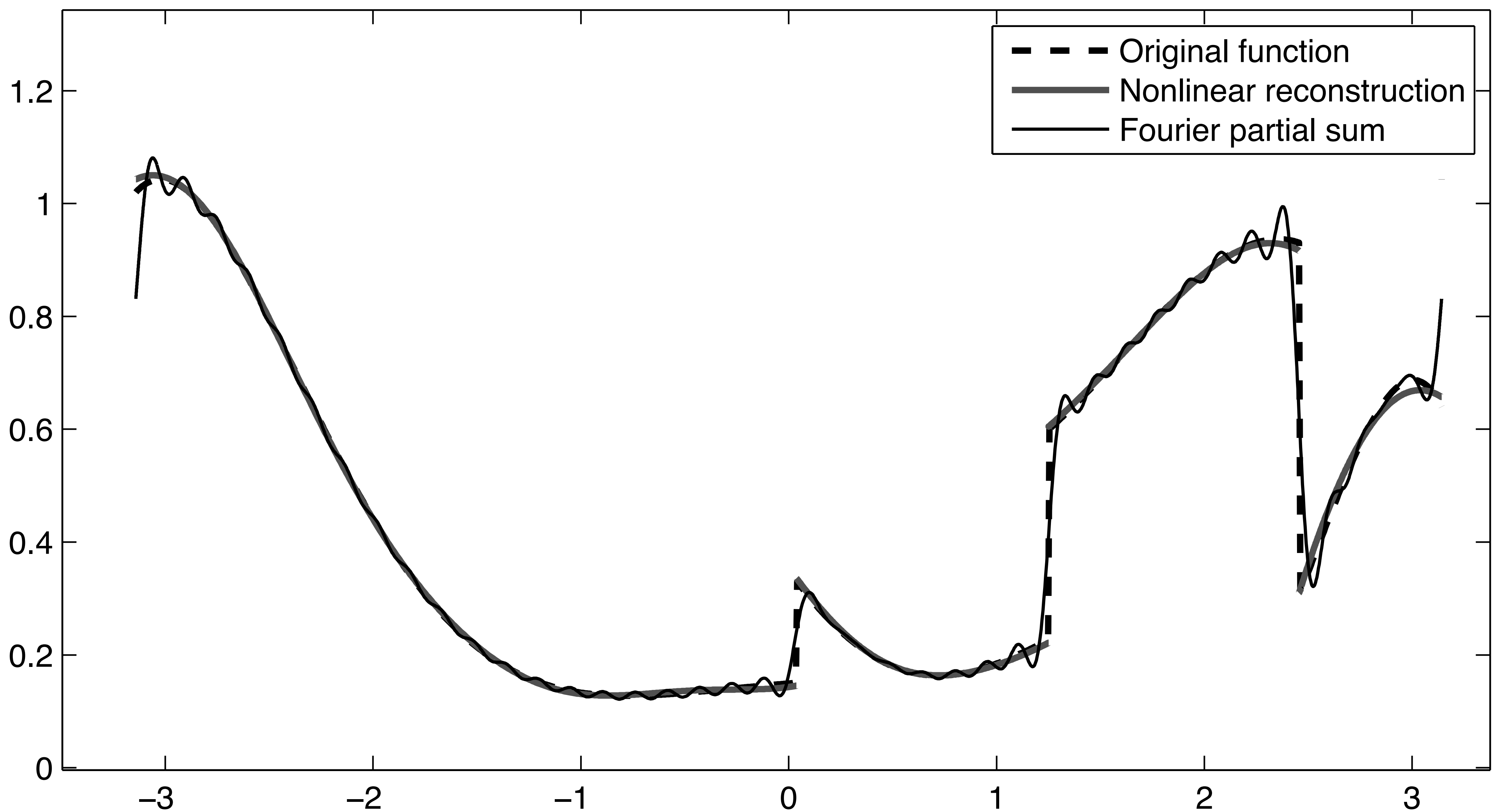
Topics in Inverse Problems

Fall 2021

# Non-smooth problems

- Until now: when solving $Af = g$, $f$ was assumed to be "smooth" (recall source condition $||A^*f|| \leq E$)
- $L_2$-type regularisation is not sufficient in general:
  - Error decays slowly (bad "computationally")
  - Gibbs-type oscillations near the edges (bad "visually")

# Model problem



# Smoothness/decay (Fourier)

- $f \sim \sum_{m=0}^{\infty} \left\langle f, \varphi_m \right\rangle \varphi_m, \quad f_M \sim \sum_{m=0}^{M} \left\langle f, \varphi_m \right\rangle \varphi_m$

- If $f \in C^{\alpha}(\mathbb{T})$ we get $\|f - f_M\| \sim M^{-\alpha}$

- If $f \in PC^{\alpha}(\mathbb{T})$ we get only $\|f - f_M\| \sim M^{-1}$

# Solution #1: (*-)lets

- Look for $f \sim \sum_j c_j \varphi_j$ where $\{c_j\}$ decay fast?
  - May not be possible in general
- Solution: “multi-resolution” representation
  - A few coefficients in “coarse” scale
  - A few coefficients near the “discontinuities”
- Wavelets, curvelets, ridgelets,…
  - “Nice” discretizations and fast computation (usually)

# Wavelets: crash-course

**Wavelet:** $$\int_{-\infty}^{+\infty} \psi(t)dt = \hat{\psi}(0) = 0 \qquad \|\psi\| = 1$$

**Wavelet dictionary:** $$\mathscr{D} = \left\{ \psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left( \underbrace{\frac{t-u}{s}}_{\text{scale}} \right) \right\}_{u\in\mathbb{R}, s\in\mathbb{R}^+}$$

**Wavelet transform:** $$Wf(u,s) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^* \left( \frac{t-u}{s} \right) dt = f \star \bar{\psi}_s(u)$$

$$\bar{\psi}_s(t) = \frac{1}{\sqrt{s}} \psi^* \left( \frac{-t}{s} \right)$$

**Scaling function:** $$|\hat{\phi}(\omega)|^2 = \int_1^{+\infty} |\hat{\psi}(s\omega)|^2 \frac{ds}{s} = \int_{\omega}^{+\infty} \frac{|\hat{\psi}(\xi)|^2}{\xi} d\xi$$

**As $s \to 0$, the magnitude of $Wf(u,s)$ measures regularity of $f$ at $u$**

# Reconstruction formula

**Low-frequency approximation:** $Lf(u,s) = \left\langle f(t), \frac{1}{\sqrt{s}}\phi\left(\frac{t-u}{s}\right)\right\rangle = f \star \bar{\phi}_s(u)$

**Theorem: let** $\psi \in L^2(\mathbb{R})$ **s.t.** $C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < +\infty$ **(admissible). Then any** $f \in L^2(\mathbb{R})$ **satisfies**

$$f(t) = \frac{1}{C_\psi}\int_0^{+\infty}\int_{-\infty}^{+\infty} Wf(u,s)\frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right) du \frac{ds}{s^2},$$

$$f(t) = \frac{1}{C_\psi}\int_0^{s_0} Wf(\cdot\,,s) \star \psi_s(t)\frac{ds}{s^2} + \frac{1}{C_\psi s_0} Lf\left(\cdot\,, s_0\right) \star \phi_{s_0}(t)$$

**and**

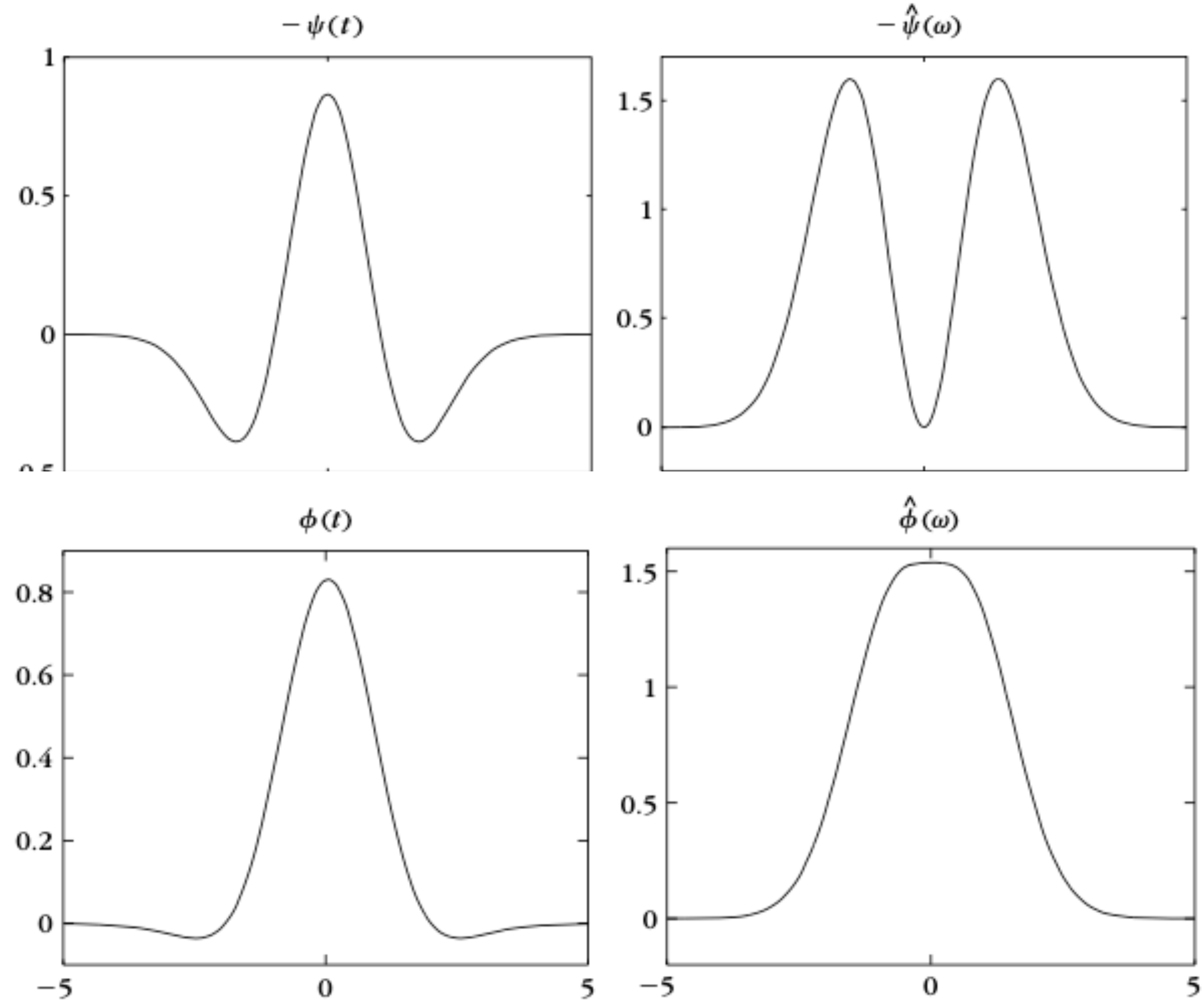
$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{C_\psi}\int_0^{+\infty}\int_{-\infty}^{+\infty} |Wf(u,s)|^2 du \frac{ds}{s^2}$$

# Mexican hat

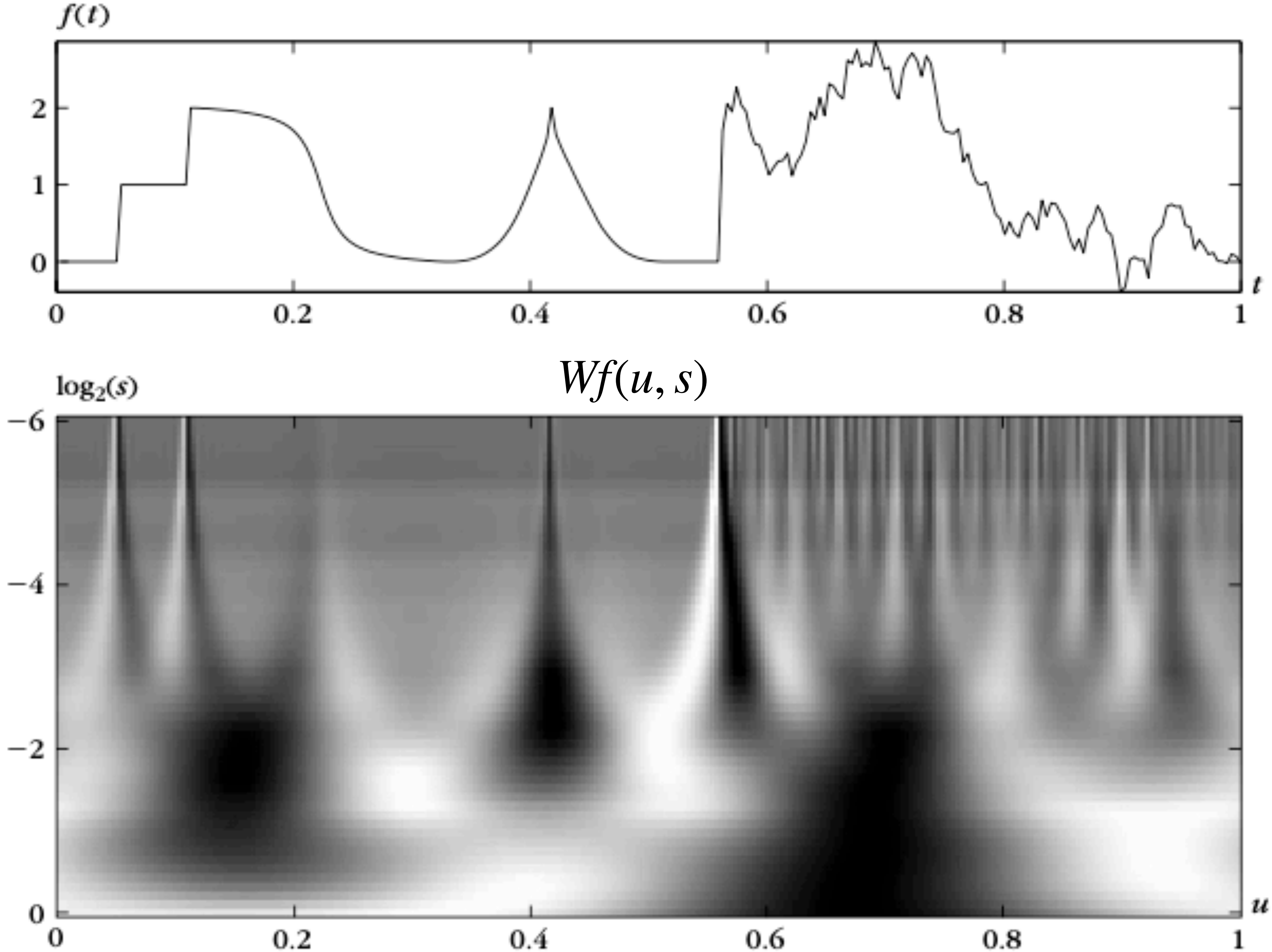
$$\psi(t) = \frac{2}{\pi^{1/4}\sqrt{3\sigma}}\left(\frac{t^2}{\sigma^2} - 1\right)\exp\left(\frac{-t^2}{2\sigma^2}\right)$$

**(high-pass filter)**

**(low-pass filter)**



# Example decomposition



# Linear vs nonlinear approximation with wavelets

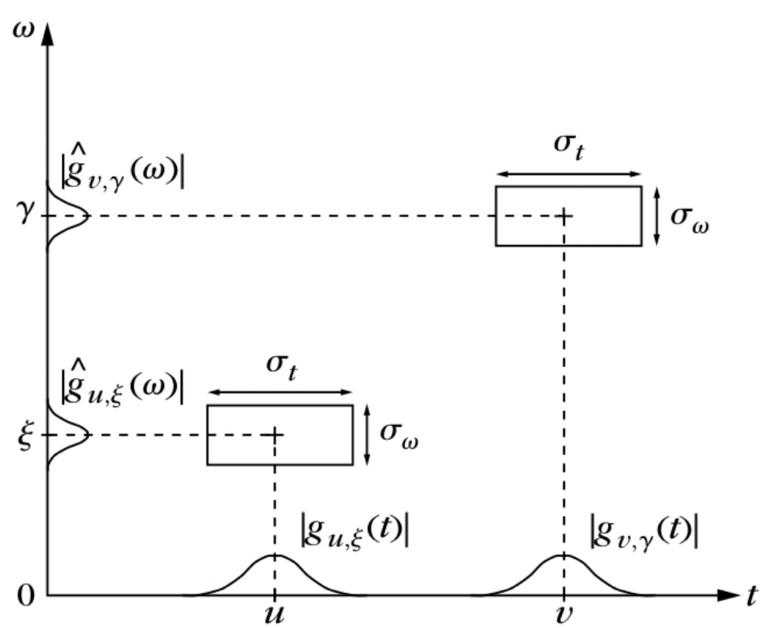
- Linear: use the first $N$ coefficients
- Nonlinear: use the first $N$ **largest** coefficients (best $N$-term approximation)
- Q: characterize functions whose best $N$-term approximation error ($\sigma_N$) decays like $N^{-\alpha}$
- A: **Besov** spaces
- Fact: for piecewise $C^\alpha$ functions $f$, $\sigma_N(f) \asymp N^{-\alpha}$ when using wavelets

# Time-frequency analysis

$$\sigma_t^2 = \frac{1}{\|f\|^2}\int_{-\infty}^{+\infty}(t-u)^2\,|f(t)|^2\,dt \qquad\qquad \sigma_\omega^2 = \frac{1}{2\pi\|f\|^2}\int_{-\infty}^{+\infty}(\omega-\xi)^2\,|\hat{f}(\omega)|^2\,d\omega$$

**Heisenberg Uncertainty Principle:** $\sigma_t^2\sigma_\omega^2 \geqslant \frac{1}{4}$

**Short-Time FT (Gabor)**



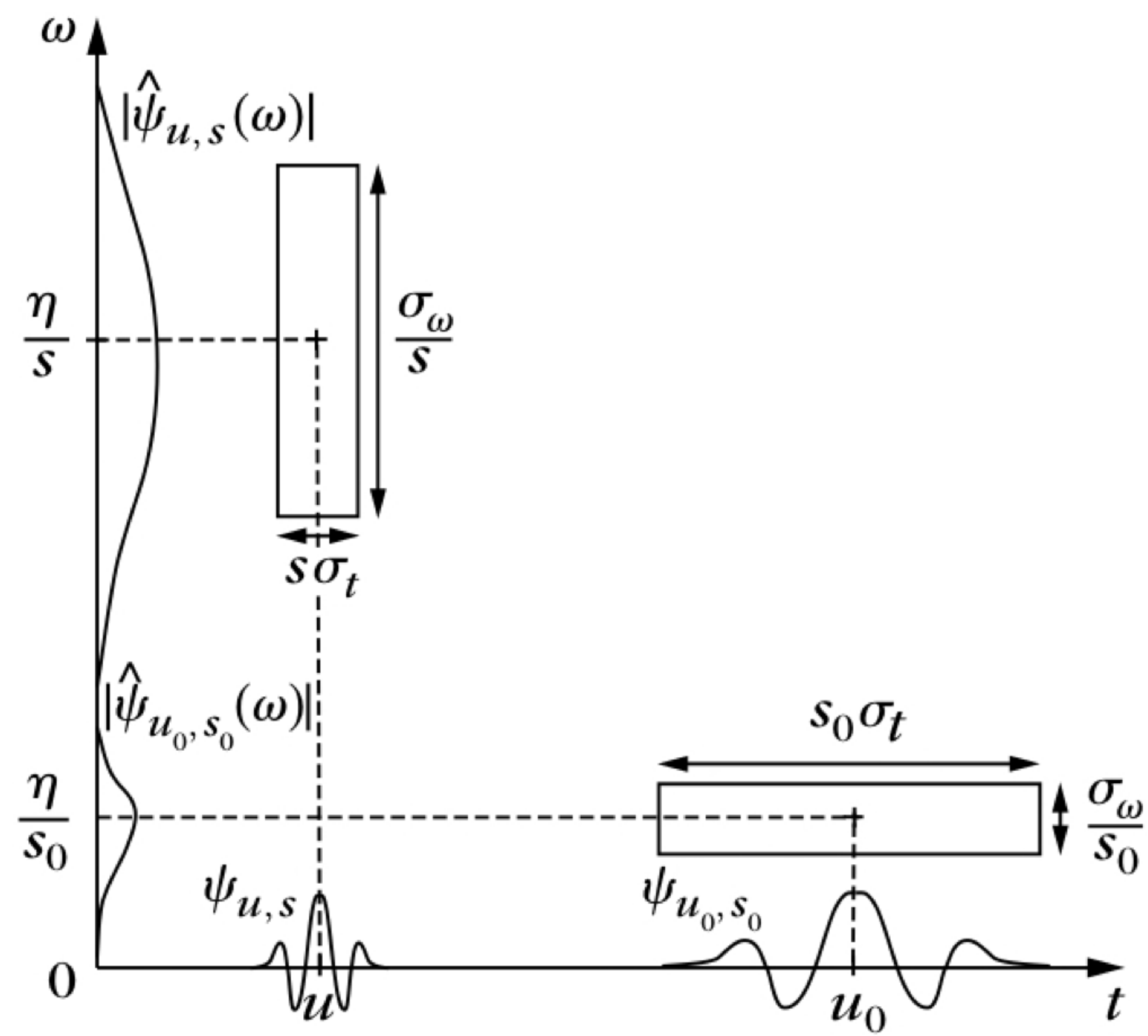

$$g_{u,\xi}(t) = e^{i\xi t} g(t-u). \tag{4.10}$$

It is normalized $\|g\| = 1$ so that $\|g_{u,\xi}\| = 1$ for any $(u,\xi) \in \mathbb{R}^2$. The resulting windowed Fourier transform of $f \in \mathbf{L^2}(\mathbb{R})$ is

$$Sf(u,\xi) = \langle f, g_{u,\xi}\rangle = \int_{-\infty}^{+\infty} f(t)\,g(t-u)\,e^{-i\xi t}\,dt. \tag{4.11}$$

# Wavelet transform

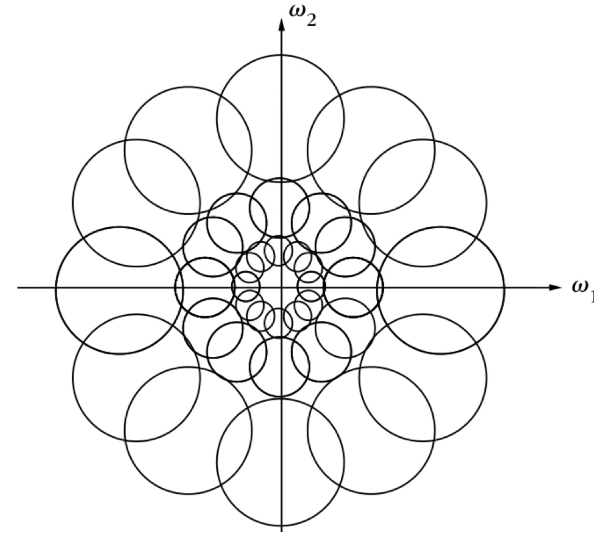
$$\phi_{u,\gamma}(t) = \psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right)$$

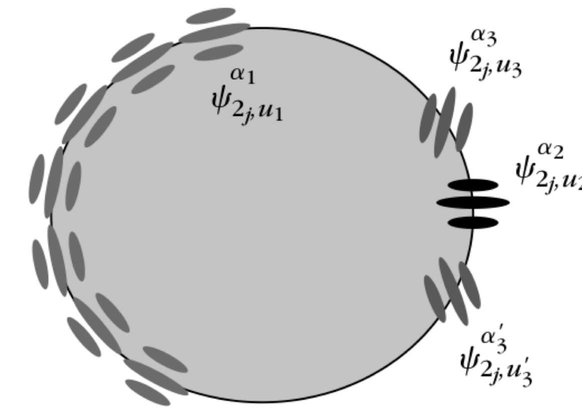



$$\sigma_t^2\sigma_\omega^2 \geqslant \frac{1}{4}$$

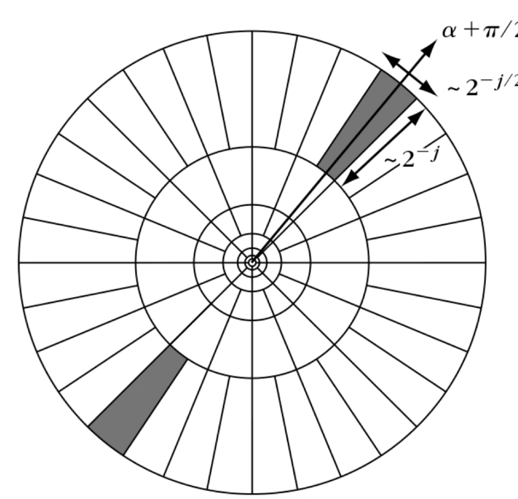
# Higher dimensions

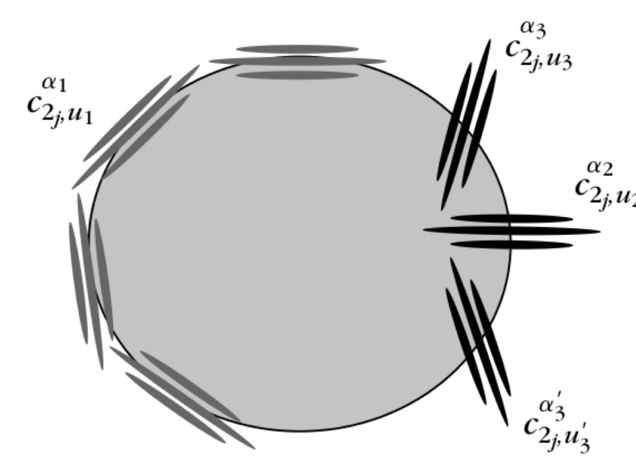
**Directional wavelets**



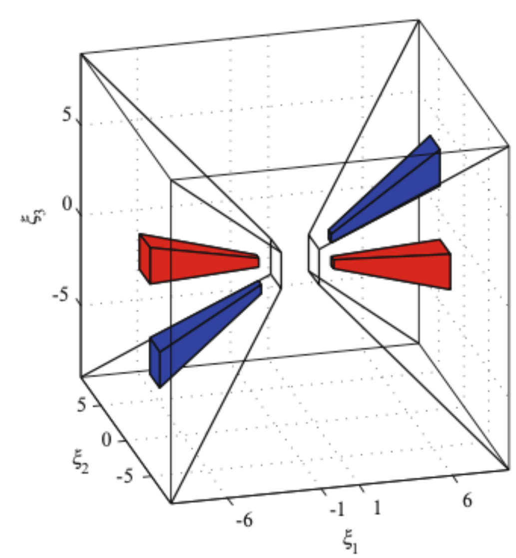




**Curvelets**




**Shearlets**



# Detecting singularities

# Smoothness/decay: Fourier

- $f$ is $\alpha$-Lipschitz at $t$ if $\exists K > 0$ and $p_t$ of degree $m = \lfloor \alpha \rfloor$ s.t. $\forall s \in \mathbb{R}$, $\quad \left| f(s) - p_t(s) \right| \leqslant K |s - t|^{\alpha}$

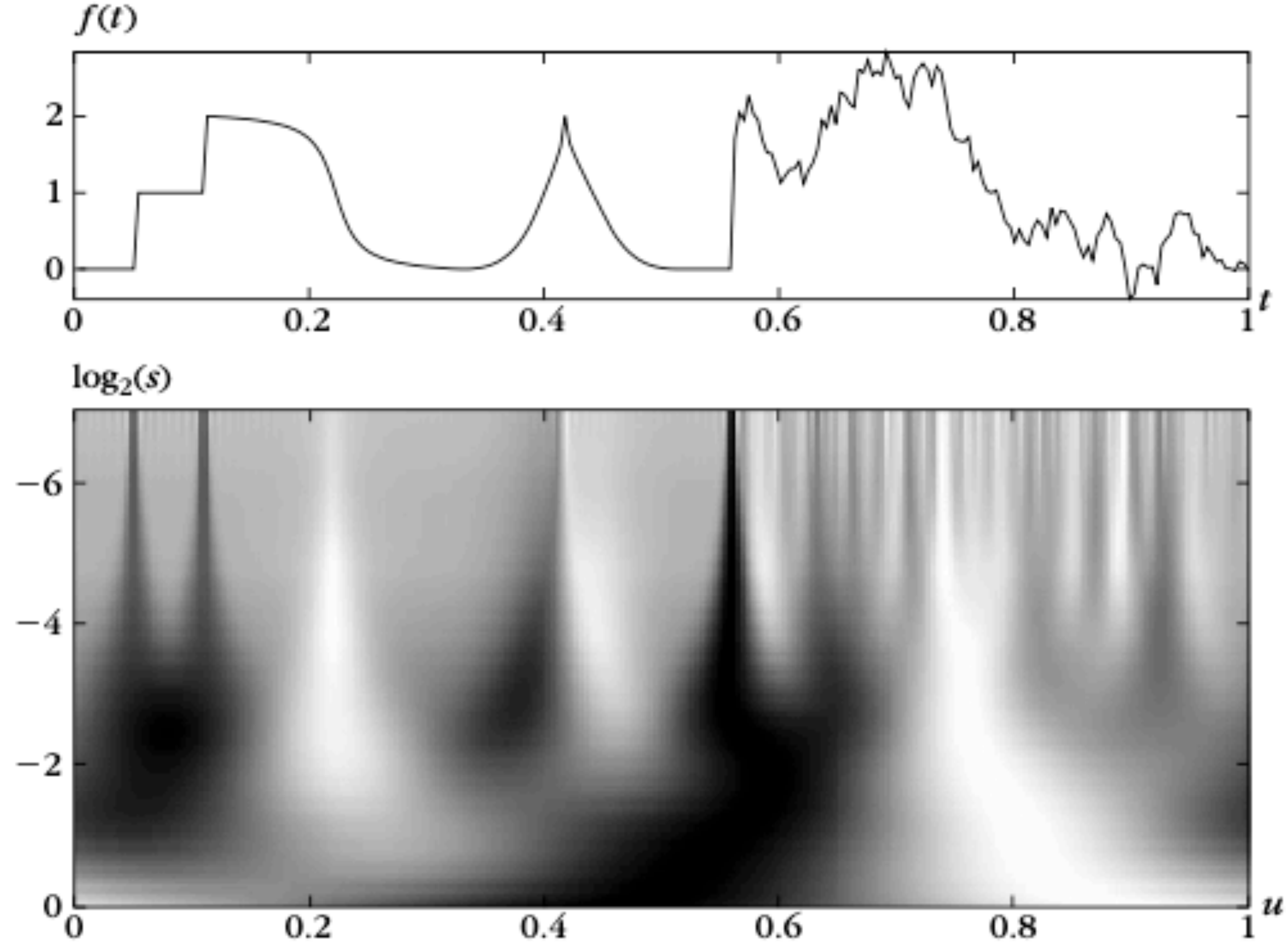
- $f$ is $\alpha$-Lip. on $\mathbb{R}$ and bounded if
$$\int_{-\infty}^{+\infty} |\hat{f}(\omega)| \left(1 + |\omega|^{\alpha}\right) d\omega < +\infty$$

- Local information is "lost" in $\hat{f}$ (or is it?)

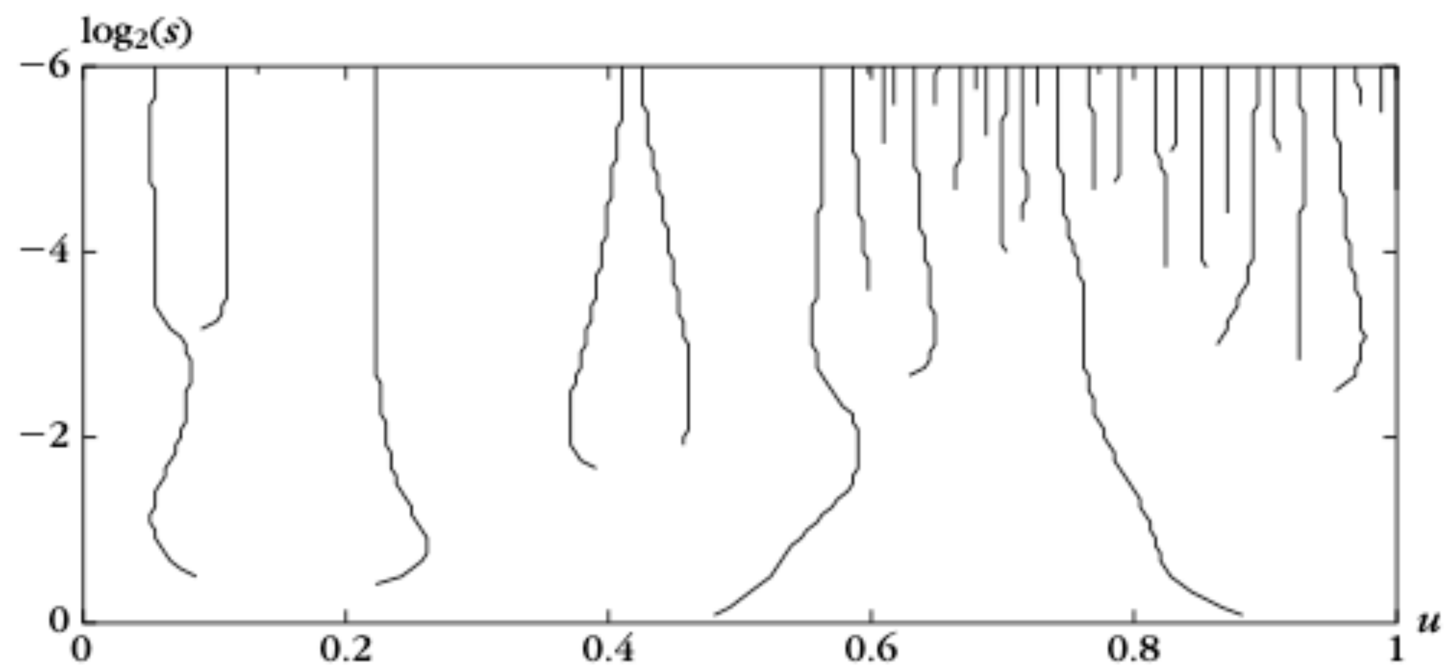
# “Wavelet zoom”

- Can characterize local Lipschitz exponents via decay of $Wf(t, s)$ as $s \to 0$ (each local maxima must satisfy $|Wf(u, s)| \leqslant As^{\alpha+1/2}$)

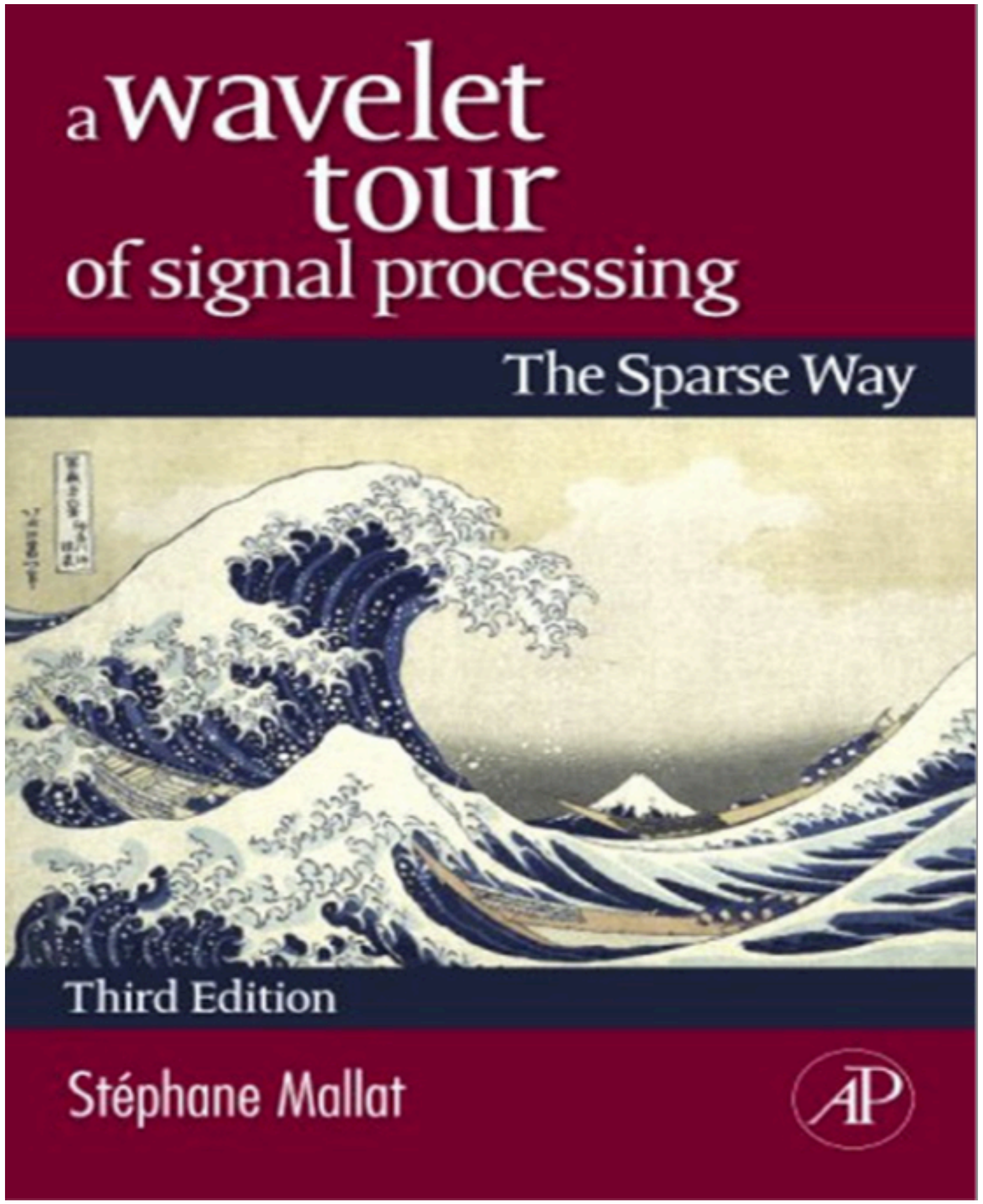
- Conversely, can detect singularities by “tracking” the local maxima of $|Wf|$ and making sure they are large enough




**(a)**




**(b)**


a wavelet tour of signal processing
The Sparse Way
Third Edition
Stéphane Mallat
AP


Applied and Numerical Harmonic Analysis
$\hat{f}(\gamma)=\int f(x)e^{-2\pi i x\gamma}dx$
Gitta Kutyniok
Demetrio Labate
Editors
Shearlets
Multiscale Analysis for
Multivariate Data
Birkhäuser

# Sparse representations / sparse regularization

Finite-dim

Prior: $x$ can be well-approximated by a few large coeff's (in some basis)

$\|x\|_0$ "$\ell_0$ norm" $= \#\{i : x_i \neq 0\}$

Denoising $\quad y^\delta = x_0 + e \quad , \quad \|e\| \leq \varepsilon$

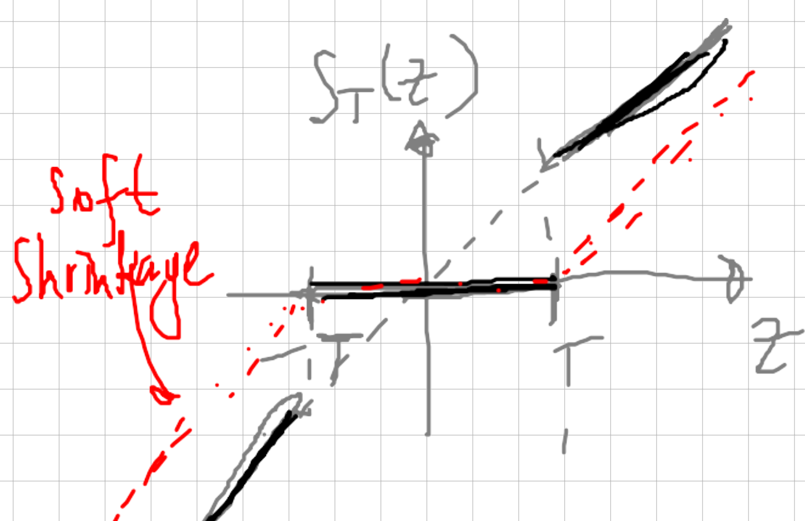
$$\boxed{\underline{\arg} \min_x \|x\|_0 \quad \text{s.t} \quad \|x - y^\delta\| \leq \varepsilon}$$

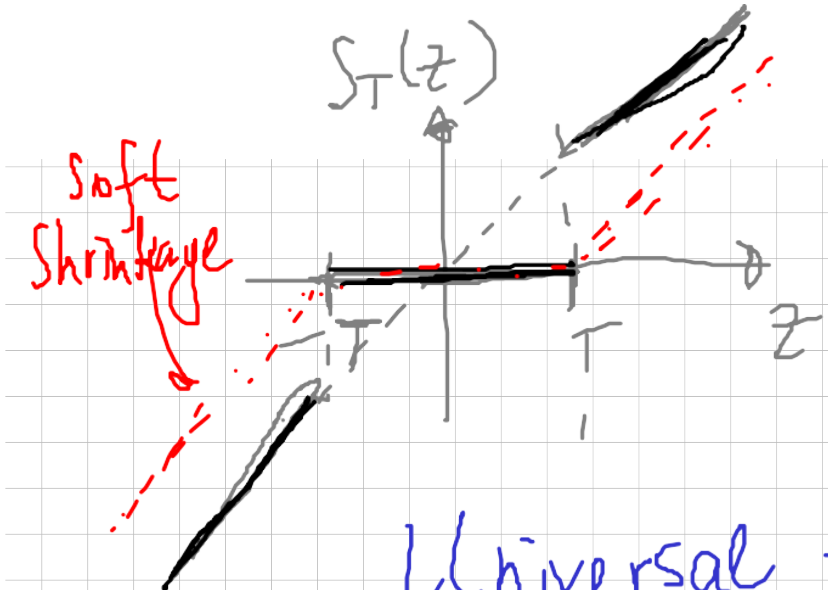
# Hard thresholding

Fix a threshold $T(\varepsilon, x) = T$ $\quad x \in \mathbb{R}^N$

$$x^{\delta,T} = \sum_{i=1} S^0_T(\langle y^\delta, \psi_i \rangle)\psi_i \quad , \text{where}$$

$$S^0_T(z) = \begin{cases} z & |z| \geq T \\ 0 & \text{otherwise} \end{cases}$$






Universal threshold, ideal shrinkage

Donoho & Johnstone, '94

$f$ is given by $N$ samples with additive noise of variance $\sigma$

Then thresholding with $T^* = \sigma\sqrt{2\log N}$ is optimal (up to constants)

$$E\{\|x^{\delta,T^*} - y^\delta\|^2\} \lesssim \log N \cdot \delta^{2-\frac{1}{\alpha+1/2}} \quad \text{for } f \in PC^\alpha$$

**Applications:** Compression, denoising.

What about Inverse Problems?

$$(P_{0,\varepsilon}) \quad \arg\min_{x} \|x\|_0 \quad \text{s.t.} \quad \|Ax - y^\delta\| \le \varepsilon$$

Can be solved by a non-tractable algorithm.

coupling

We don't know where are the nonzeros.

Thm: the problem is NP-hard.

$\|x\|_1 = \sum_{i=1}^{N} |x_i|$

$(P_{1,\varepsilon})$ (1) $\arg\min_x \|x\|_1$ s.t. $\|Ax - y^\delta\|_2 \le \varepsilon$ Basis Pursuit (BP)

(2) $\arg\min_x \|x\|_1 + \lambda \|Ax - y^\delta\|_2$ BP denoising

(3) $\arg\min_x \|Ax - y^\delta\|_2$ s.t. $\|x\|_1 \le \tau$ LASSO

Th: a) $x^*$ solves (2) with some $\lambda > 0$.

Then $x^*$ solves (1) with $\varepsilon = \|Ax^* - y^\delta\|$

easy

b) $x^*$ is a unique sol. of (1) with $\varepsilon \ge 0$

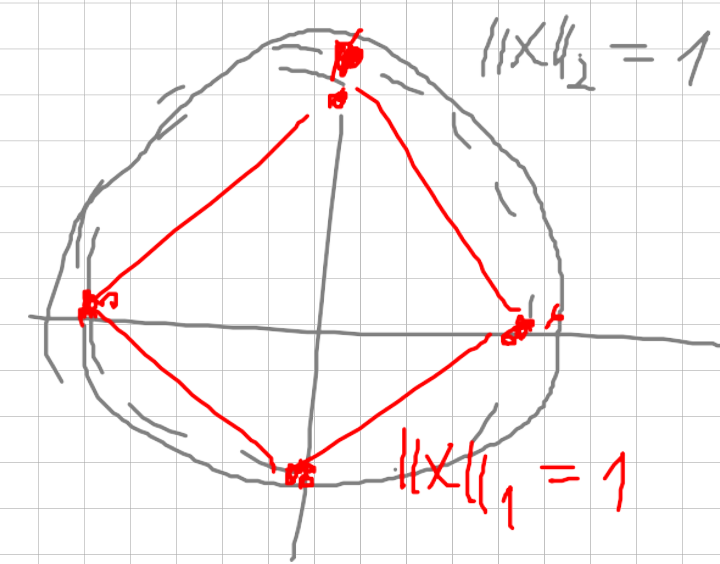
Then $x^*$ is a unique sol. of (3) with $\tau = \|x^*\|_1$

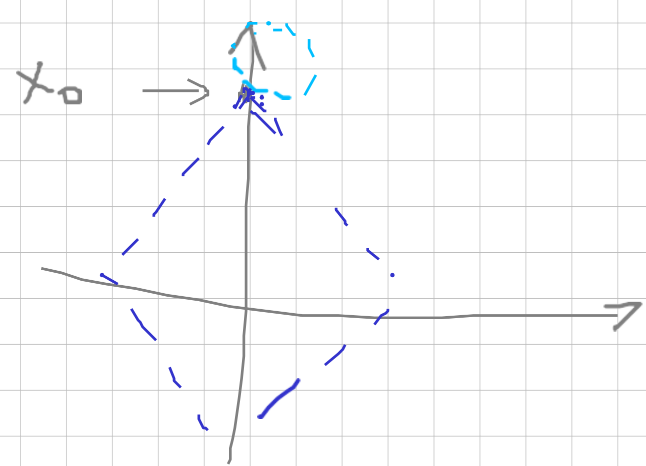
c) $x^*$ solves (3) with $\tau > 0 \Rightarrow \exists \lambda(x^*)$ s.t. $x^*$ solves (2)

with this $\lambda$.

$\boxed{\ell_1 \text{ "promotes sparsity"}}$

$$\arg\min_x \|x\|_1 \quad \text{s.t.} \quad \|Ax - y^\delta\|_2 \le \varepsilon$$




Prior
$y = Ax_0$ where
$\|x_0\|_0$ is small

Remarks; „Applications"

- denoising / compression → $A = I$
- inverse problems → $A$ is fixed
- „overcomplete representations" $A$ may be unknown

  $A$ is a „dictionary" which can be learned from data

One can consider

$R_\lambda$ reg operator

$$R_\lambda y^\delta = \arg\min_x \|Ax - y^\delta\| + \lambda \|x\|_1$$

2004 Daubechies, Defrise, Demol

convergence: if we select $\lambda(\delta)$ s.t. $\begin{cases} \lambda(\delta) \xrightarrow[\delta \to 0]{} 0 \\ \dfrac{\delta^2}{\lambda(\delta)} \xrightarrow[\delta \to 0]{} 0 \end{cases}$

then $\|\delta x\| \to 0$

Also can obtain convergence rates stability etc

ISTA algorithm (Iterative Shrinkage & thresholding)

$$\Phi_\lambda(x) = \|Ax - y^\delta\|_2^2 + \lambda\|x\|_1 \to \min.$$

General idea: decouple the terms which are $\sim \|Ax\|^2$

„Surrogate functional" (assuming $\|A^*A\| < 1$)

$$\Phi_\lambda^{SUR}(x;\xi) = \|Ax - y^\delta\|_2^2 + \lambda\|x\|_1 - \|Ax - A\xi\|^2 + \|x - \xi\|^2 \geq \Phi_\lambda(x)$$

and also $\Phi_\lambda^{SUR}(x;x) = \Phi_\lambda(x)$

We can show that

$$\phi_y^{SUR}(x;\xi) = \|y^\delta\|^2 - 2x^T \overbrace{A^T(y^\delta - A\xi)}^{u} - \|A\xi\|^2 + \lambda\|x\|_1 + \|\xi - x\|^2$$

constant

Can optimize this coordinate-wise!

$$\phi_y^{SUR}(x;\xi) = \sum_{i=1}^{N} \left( x_i z_i + \lambda|x_i| + x_i^2 \right) + \text{STUFF NOT DEPENDING ON } x.$$

How do we minimize this?

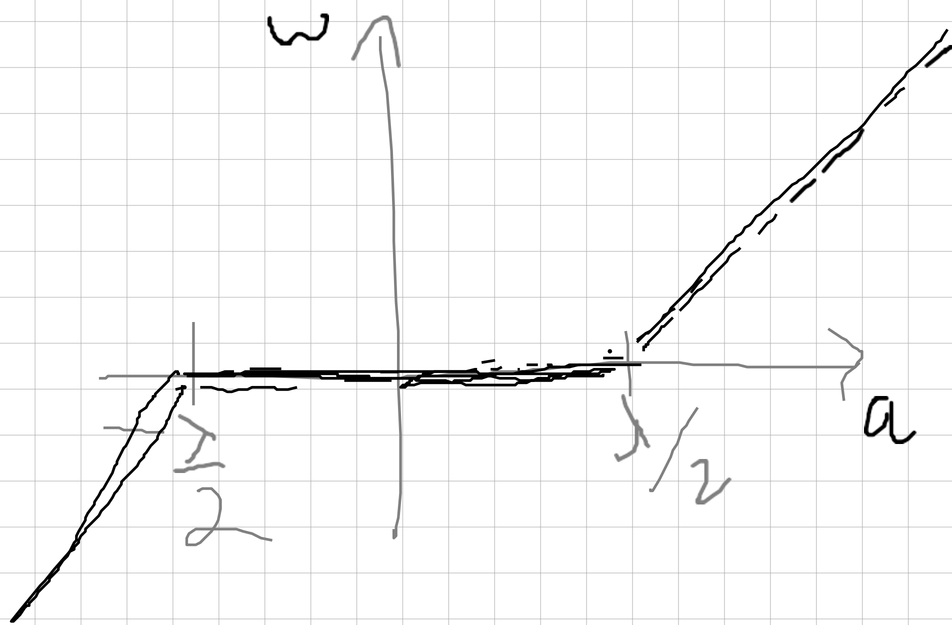
$$\underbrace{\arg\min_w (w-a)^2 + \lambda|w|}_{L} \quad , w \in \mathbb{R}$$

$(w = x_i)$

$$\frac{\partial}{\partial w} L(w) = 0$$

1) $w > 0 \Rightarrow w = a - \frac{\lambda}{2}$

2) $w < 0 \Rightarrow w = a + \frac{\lambda}{2}$




Soft thresholding

$$S \leftarrow S^1_\lambda(a)$$

ISTA:

iterates are $f_0 = 0, f_1, f_2 \longrightarrow \arg\min_x \Phi_\lambda$

$$f_{k+1} = S_\lambda^1 \left( f_k + A^T (y^\delta - A f_k) \right)$$

Projected Landweber Iteration

# Tomography example

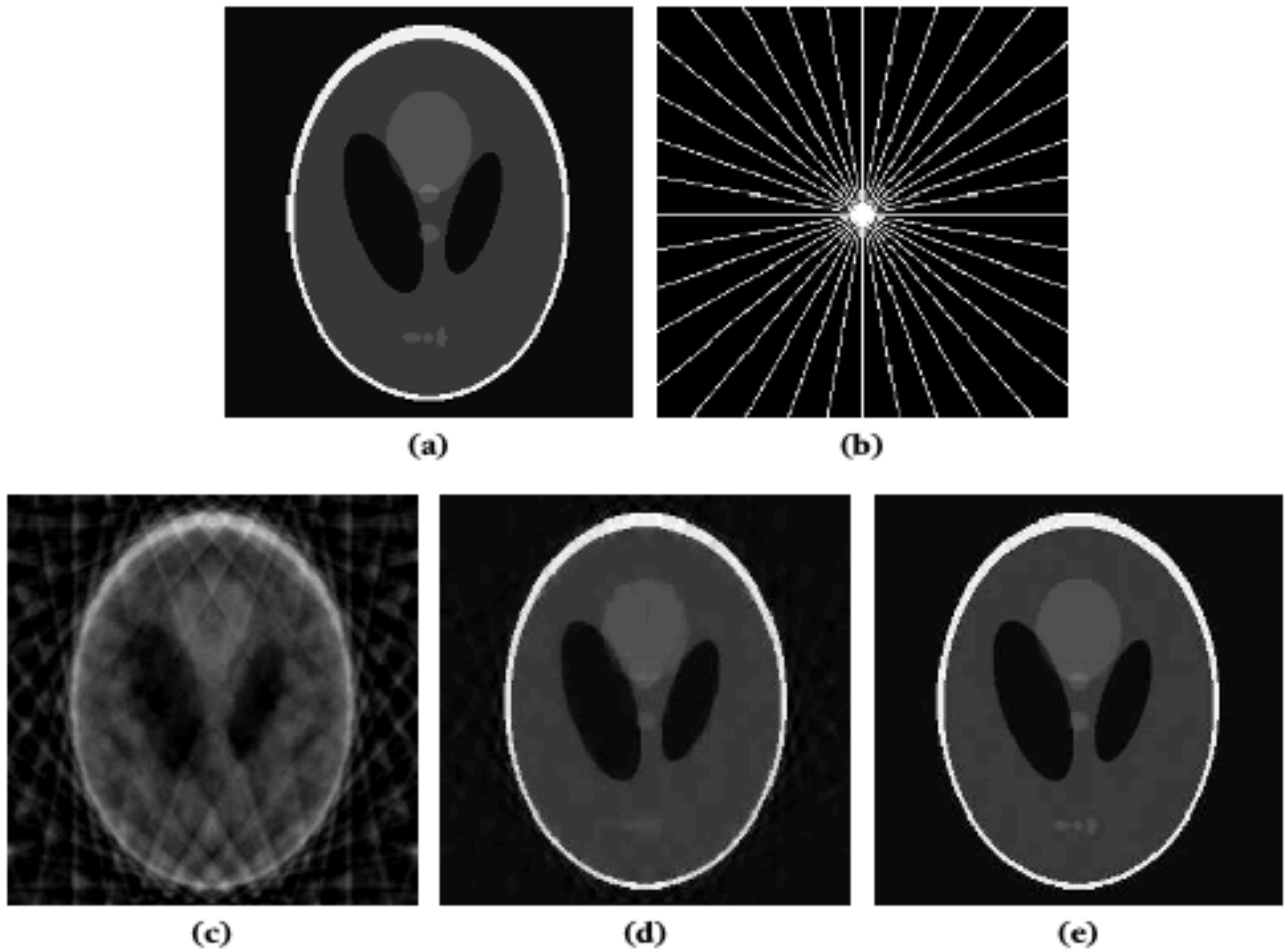
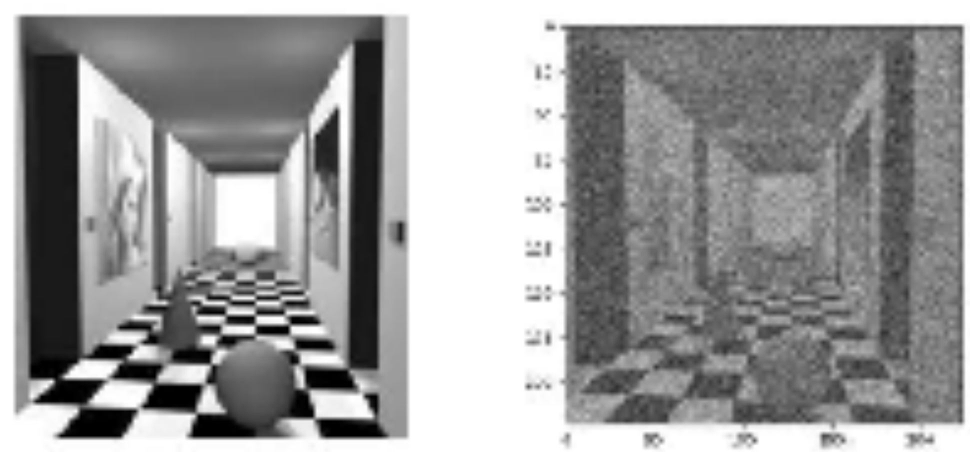


**FIGURE 13.7**

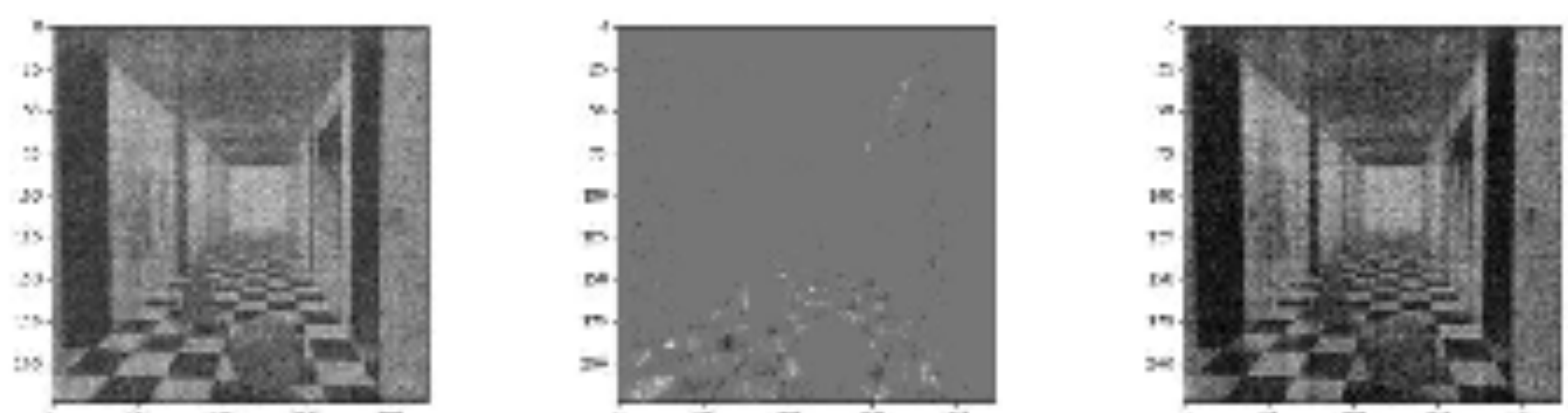
**(a)** Original phantom image. **(b)** Frequency plane showing in white the frequency rays in $\Omega$.
**(c)** Reconstruction with a linear orthogonal projection computed with a backprojection.
**(d)** Lagrangian pursuit estimation in a Haar translation-invariant wavelet dictionary.
**(e)** Inversion with a total variation regularization.

# Mixed penalties

- $\Phi_{\lambda,\mu}(u, v) := \|K(u + v) - g\|^2 + \lambda\|Wv\|^2 + \mu\|u\|^p$



Original Image　　Noisy Image



K=500　　U　　V