Aisenstadt Chair Course
CRM September 2009

Part I
Sparse Representations in Signal and Image Processing

Stéphane Mallat

Centre de Mathématiques Appliquées
Ecole Polytechnique

Sparse Approximation Processing

- **Key idea:** approximate signals f as a sparse decomposition in a dictionary $\mathcal{D} = \{\phi_p\}_{p\in\Gamma}$ of waveforms

$$f = \sum_{p\in\Lambda} a[p]\,\phi_p + \epsilon_\Lambda$$

- The signal is characterized by fewer coefficients $a[p]$:
  - Compression capabilites
  - Fast algorithms and memory saving
  - Estimation of fewer coefficients for:
    - noise removal
    - inverse problems
    - pattern recognition ????

A Sparse Tour

- I. Linear versus Non-Linear Representations in Bases
- II. Sparsity in Redundant Dictionaries
- III. Super-resolution for Inverse Problems
- IV. Compressive Sensing
- V. Dictionary Learning & Source Separation

- End: *Grouping to Perceive in an Incompressible World*

- Contributors: many...

- Softwares: *http://www.wavelet-tour.com*

Sparse Linear Versus Non-Linear

- Linear representations are powerful but... limited:
  - Approximations and sampling theorems
  - Principal Component Analysis

- Non-linear approximation in bases:
  - Wavelets and adaptive sampling

- Signal and image compression

- Linear and non-linear noise removal

Linear Representation in a Basis

- Decomposition in an orthonormal basis $\mathcal{B} = \{g_m\}_{m\in\mathbf{N}}$

$$f = \sum_{m=0}^{+\infty} \langle f, g_m \rangle \, g_m$$

- Approximation of f over the first N vectors: projection on the space $\mathbf{U}_N = \text{Vect}\{g_m\}_{0\leq m<N}$

$$f_N = P_{\mathbf{U}_N} f = \sum_{m=0}^{N-1} \langle f, g_m \rangle \, g_m$$

- Error:

$$f - f_N = \sum_{m=N}^{+\infty} \langle f, g_m \rangle \, g_m \quad \text{so} \quad \|f - f_N\|^2 = \sum_{m=N}^{+\infty} |\langle f, g_m \rangle|^2$$

- Depends on the decay of $|\langle f, g_m \rangle|$ as m increases.

Uniform Sampling

- *f(t)* is discretized with a filtering and uniform sampling:

$$f * \phi_s(nT) = \int f(u)\,\phi_s(nT-u)\,du = \langle f(u), \phi_s(nT-u)\rangle$$

- It gives the decomposition coefficients of *f(t)* in a Riesz basis $\{\phi_n(t) = \phi_s(nT-t)\}_{0\leq n<N}$ of a space $\mathbf{U}_N$

$$P_{\mathbf{U}_N} f = f_N = \sum_n \langle f, \phi_n\rangle\,\tilde{\phi}_n$$

- If $\mathbf{U}_N = \mathrm{Vect}\{g_m\}_{0\leq m<N}$ where $\mathcal{B} = \{g_m\}_{m\in\mathbf{N}}$ is an orthonormal basis of the whole signal space then

$$\|f - f_N\|^2 = \sum_{m=N}^{+\infty} |\langle f, g_m\rangle|^2$$

- Sampling theorems...

Approximation in a Fourier Basis

- Fourier basis $\{e^{i2\pi mt}\}_{m\in\mathbf{Z}}$ of $\mathbf{L^2}[0,1]$

$$f(t) = \sum_{m=-\infty}^{+\infty} \hat{f}(2\pi m)\, e^{i2\pi mt}$$

with $\hat{f}(2\pi m) = \int_0^1 f(u)\, e^{-i2\pi mu}\, du$

- Low frequency Fourier approximation:

$$f_N(t) = \sum_{m=-N/2}^{N/2} \hat{f}(2\pi m)\, e^{i2\pi mt}$$

Fourier Approximation Error

- The approximation error is

$$\|f - f_N\|^2 = \sum_{|m|>N/2} |\hat{f}(2\pi m)|^2$$

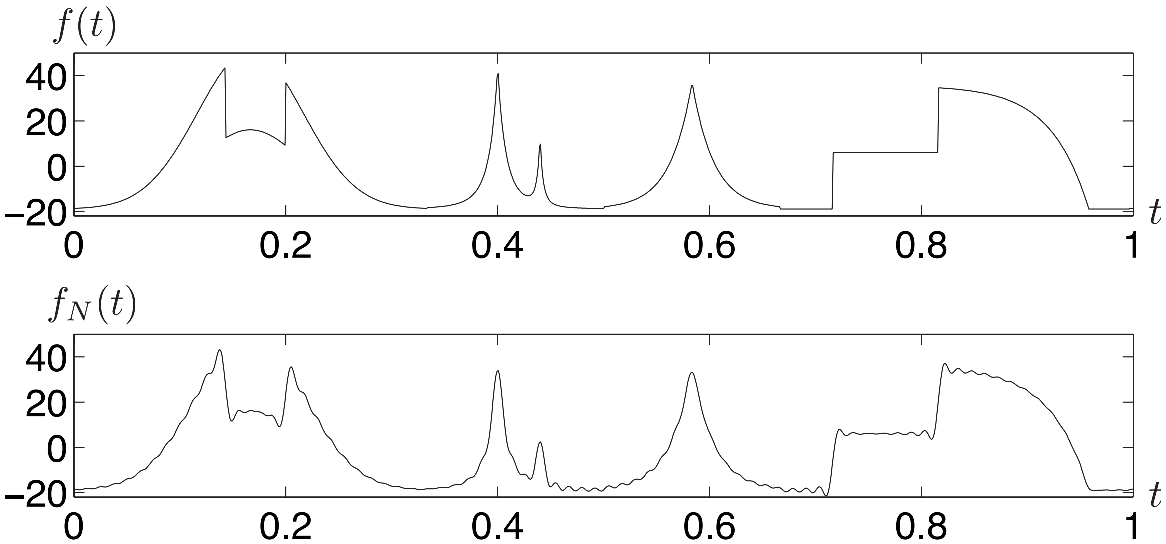
- It depends on the high-frequency decay of $|\hat{f}(2\pi m)|$ which depends on the uniform regularity of f.

- Nyquist sampling theorem: $\phi_n(t) = \dfrac{\sin(\pi t/T - n)}{\pi t/T - n}$

- If f is s times differentiable in the sense of Sobolev then

$$\|f - f_N\|^2 = o(N^{-2s})$$

Example of Fourier Approximation



Wavelet Tour of Signal Processing, 3rd ed. Top: Original signal f. Middle: Signal f_N approximated from $N = 128$ lo
rier coefficients, with $\|f - f_N\|/\|f\| = 8.63\,10^{-2}$.

Principal Component Analysis

- Find a best approximation basis from signal examples.
- Signals are realization of a random vector $F[p] \in \mathbf{R}^P$
- Linear approximation in a basis $\{g_m\}_{0 \le m < P}$

$$F_N = \sum_{m=0}^{N-1} \langle F, g_m \rangle \, g_m$$

- Find the basis which minimizes the expected error:

$$E\{\|F - F_N\|^2\} = \sum_{m=N}^{P} E\{|\langle F, g_m \rangle|^2\}$$

Karhunen-Loeve Basis

- The covariance matrix $R_F[n,m] = E\{F[n]\,F[m]\}$ is diagonal in an orthonormal basis (Karhunen-Loeve).

- **Theorem:** The approximation error

$$E\{\|F - F_N\|^2\} = \sum_{m=N}^{P} E\{|\langle F, g_m\rangle|^2\}$$

is minized by projecting F on the N vectors of the Karhunen-Loeve basis with largest eigenvalues (variance).

PCA Properties

- The Karhunen-Loeve basis is easy to compute

- But it does not always provide a good approximation.
- Example: random shift signals

$$F[p] = f[(n - X) \bmod P]$$

are stationary

$$R_F[n,m] = R_F[n-m] = \frac{1}{P} f \star \tilde{f}[n-m]$$

the Karhunen-Loeve basis is thus a Fourier basis, which is not always effective...

Non-Linear Approximation

- *Adaptive sampling*: put samples where they are needed.
- How ?
- Sparse non-linear approximation in a basis $\mathcal{B} = \{g_m\}_{m \in \mathbf{N}}$

$$f_M = \sum_{m \in \Lambda} \langle f, g_m \rangle \, g_m \quad \text{with} \quad |\Lambda| = M \ .$$

- Since

$$\|f - f_M\|^2 = \sum_{m \notin \Lambda} |\langle f, g_m \rangle|^2$$

- The minimum error is obtained by thresholding:

$$\Lambda = \{m \ : \ |\langle f, g_m \rangle| > T(M)\} \ .$$

Non-Linear Approximation Error

- $\{\langle f, g_{m_k}\rangle\}_k$ sorted with decreasing amplitude

$$|\langle f, g_{m_{k+1}}\rangle| \leq |\langle f, g_{m_k}\rangle|.$$

- Sparse non-linear approximation:

$$f_M = \sum_{k=1}^{M} \langle f, g_{m_k}\rangle\, g_{m_k}$$

and

$$\|f - f_M\|^2 = \sum_{k=M+1}^{N} |\langle f, g_{m_k}\rangle|^2 \ .$$

If $|\langle f, g_{m_k}\rangle| = O(k^{-\alpha})$ then $\|f - f_M\|^2 = O(M^{1-2\alpha})$.

Wavelet Bases

- Wavelet orthonormal basis of $\mathbf{L^2}[0,1]$

$$\left\{ \psi_{j,n}(t) = \frac{1}{\sqrt{2^j}}\, \psi\left(\frac{t - 2^j n}{2^j}\right) \right\}_{j<0,\, 2^j n \in [0,1]}$$

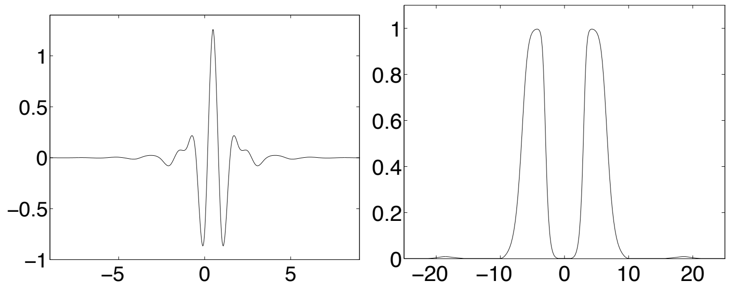
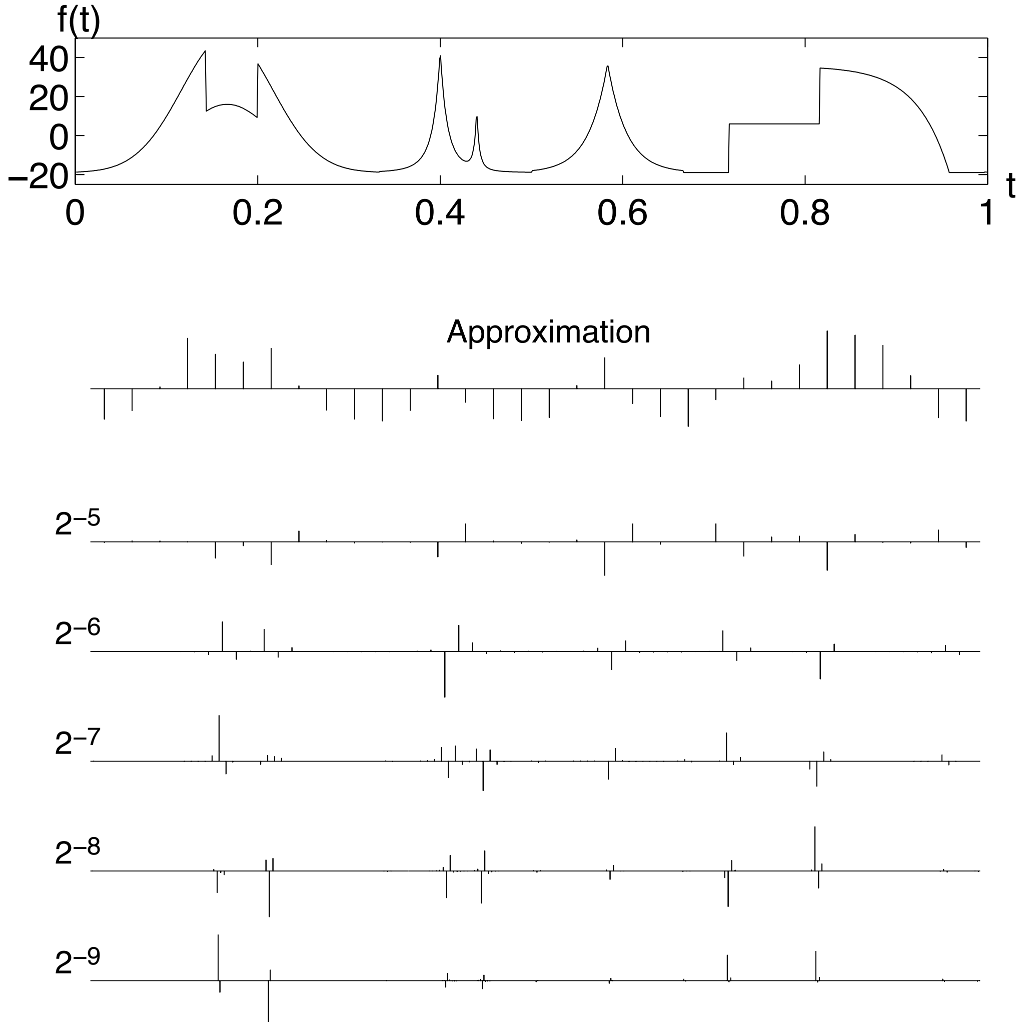


Fig. 7.5. A Wavelet Tour of Signal Processing, 3rd ed. Battle-Lemarié cubic spline wavelet ψ and its Fourier transform modulus.

- Fast algorithm in *O(N)* to compute N wavelet coefficients $\langle f, \psi_{j,n} \rangle$
- $|\langle f, \psi_{j,n} \rangle|$ is large where f is irregular.

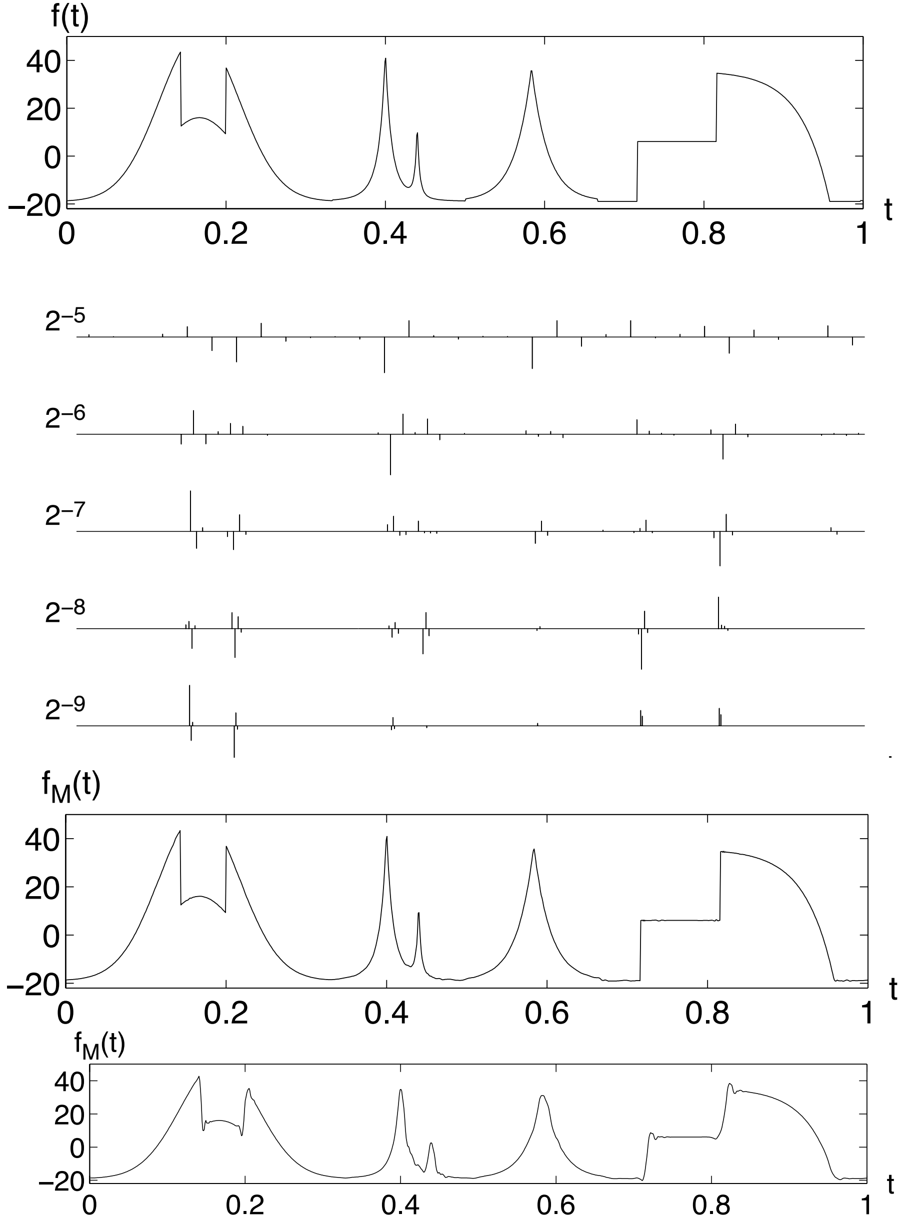
Wavelet Coefficients

•



Wavelet coefficients $\langle f, \psi_{j,n} \rangle$

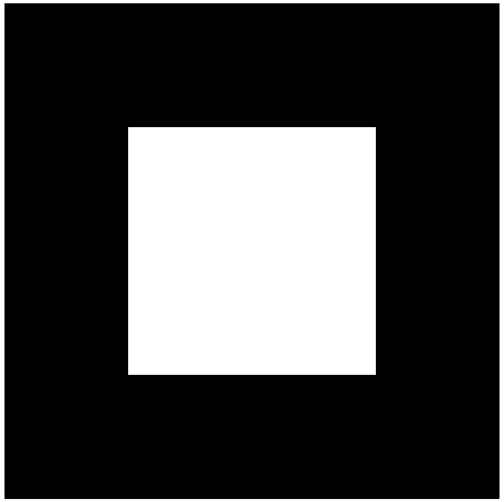
Non-Linear Wavelet Approximation

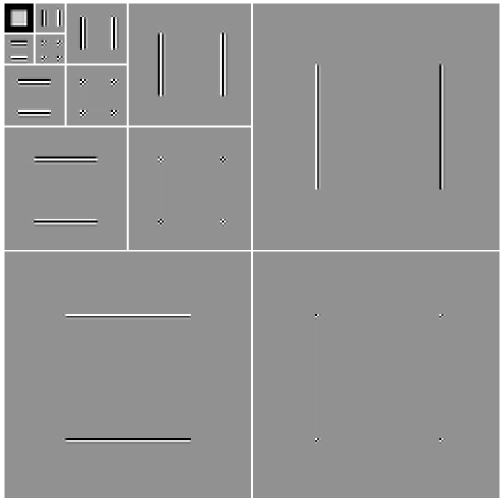


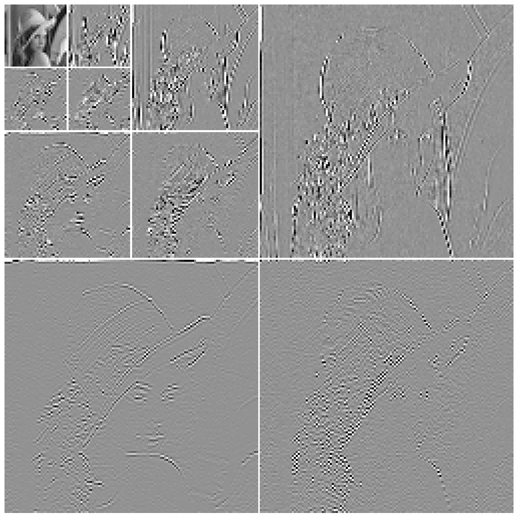
$$Non-linear:$$
$$\|f - f_M\|^2 = 5.1\,10^{-3}$$
$$Linear:$$
$$\|f - f_M\|^2 = 8.5\,10^{-2}$$

Wavelet Bases of Images

- Wavelet basis of $\mathbf{L^2}[0,1]^2$: $\left\{ \frac{1}{2^j} \psi^k \left(\frac{x - 2^j n}{2^j} \right) \right\}_{\substack{1 \leq k \leq 3, j < 0 \\ 2^j n \in [0,1]^2}}$







Wavelet coefficients

$k = 1, 2, 3$

$j = -1, -2, -3, -4$

$2^j n \in [0, 1]^2$

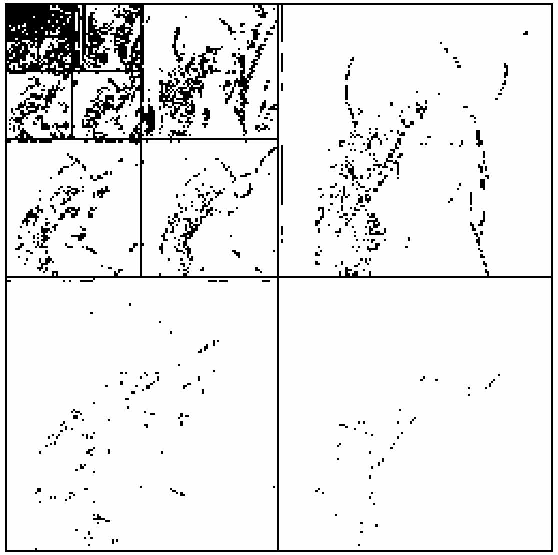
Wavelet Image Approximations

Original Image

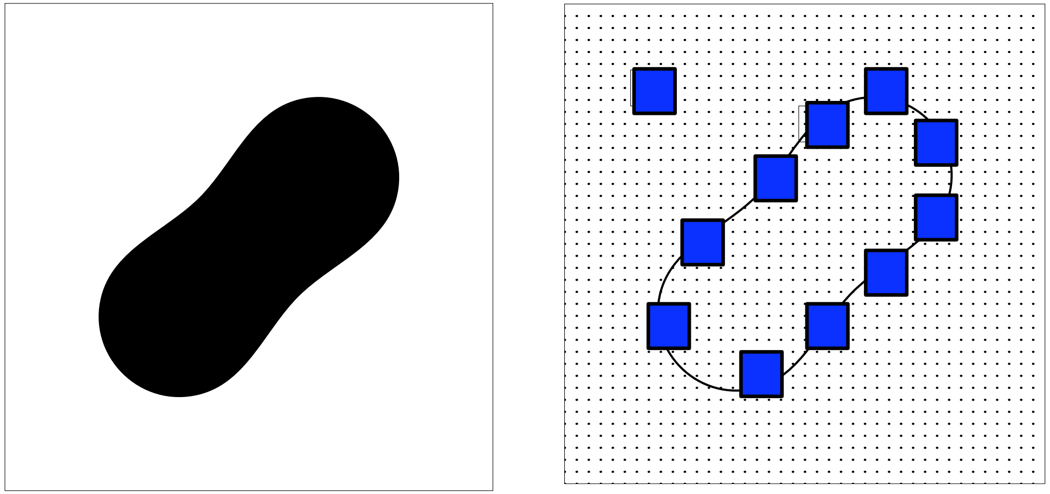
Non-linear Approximation

Linear Approximation

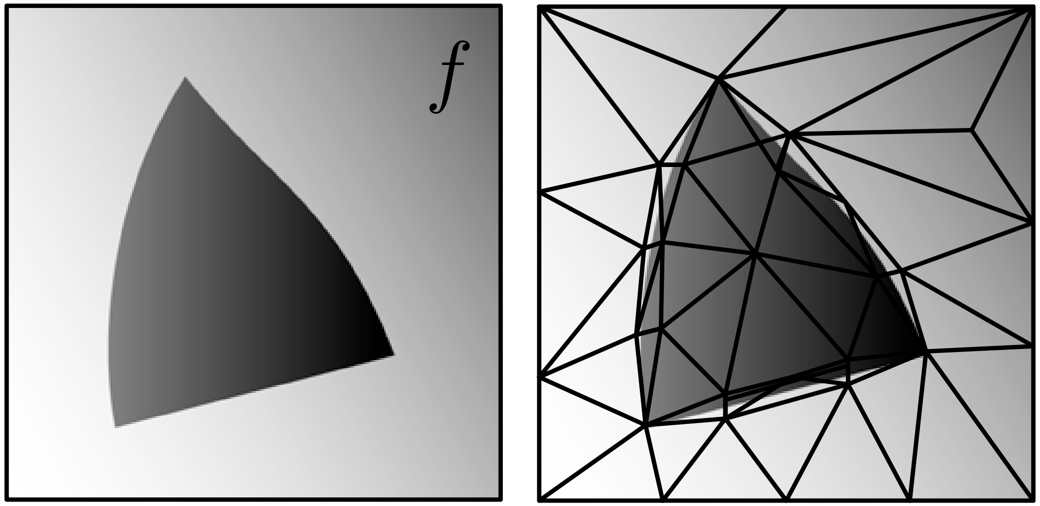
M = N/16 largest wavelet coeffs.



Good but Not Optimal



The number of large wavelet coefficient is proportional to the length of the contour.



Need less adapted triangles if the contour geometry is regular.

Sparse Signal Compression

- Signal $f[n] \in \mathbf{R}^N$ decomposed in a basis $\mathcal{B} = \{g_m\}_{0 \leq m < N}$

$$f = \sum_{m=0}^{N-1} \langle f, g_m \rangle \, g_m$$

- Coefficients approximated by a uniform quantifier:

$$Q(x) = n\,\Delta \quad \text{if} \quad x \in [(n - 1/2)\Delta, (n + 1/2)\Delta)$$

- Restored signal from quantized coefficients:

$$\tilde{f} = \sum_{m=0}^{N-1} Q(\langle f, g_m \rangle) \, g_m$$

-

Bit Budget

- Need R bits for a binary entropy coding of

$$\{Q(\langle f, g_m \rangle)\}_{0 \leq m < N}$$

includes only M non-zero coefficients $m \in \Lambda$

$$Q(\langle f, g_m \rangle) = 0 \;\; \text{if} \;\; |\langle f, g_m \rangle| \leq \Delta/2 \; .$$

Distortion-Rate

- Compression distortion:

$$D(R) = \|f - \tilde{f}\|^2 = \sum_{m=9}^{N-1} |\langle f, g_m \rangle - Q(\langle f, g_m \rangle)|^2$$

$$= \sum_{|\langle f,g_m \rangle| < \Delta/2} |\langle f, g_m \rangle|^2 + \sum_{|\langle f,g_m \rangle| \geq \Delta/2} |\langle f, g_m \rangle - Q(\langle f, g_m \rangle)|^2$$

$$D(R) \leq \|f - f_M\|^2 + M \frac{\Delta^2}{4}.$$

- Bit budget:

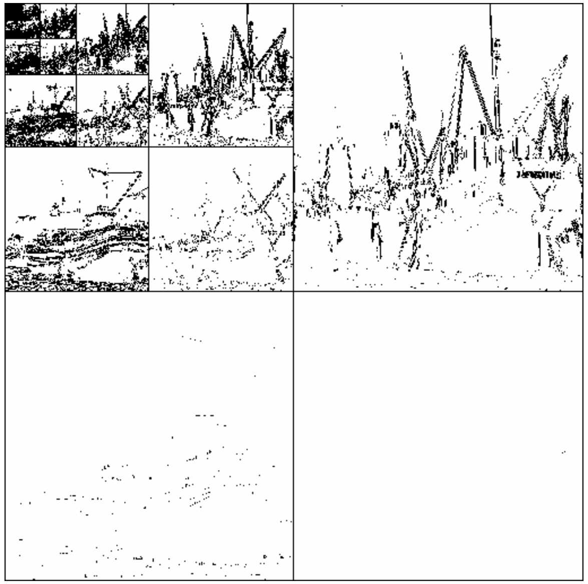
$$R = \log_2 \binom{N}{M} + \mu\, M$$

$$R \sim M \log_2(N/M)$$

- Compression depends on non-linear approximation.

Compression with JPEG-2000

Non-zero wavelet coefficients



0.2 bit/pixel

0.05 bit/pixel

Noise Removal

- Measure a signal plus noise

$$X[n] = f[n] + W[n] \quad \text{for} \quad 0 \leq n < N \ .$$

- Deterministic signal model: $f \in \Theta$

- Estimator: $\tilde{F} = D\,X$
- Risk: $r(D, f) = E\{\|\tilde{F} - f\|^2\}$
- Maximum risk: $r(\Theta, D) = \sup_{f \in \Theta} r(D, f)$

- Minimax risk: $r_{\min}(\Theta) = \inf_{D} r(\Theta, D)$
- How to construct nearly minimax estimators ?

Diagonal Estimator in a Basis

- Decompose $X = f + W$ in a basis $\mathcal{B} = \{g_m\}_{0 \le m < N}$

$$X = \sum_{m=0}^{N-1} \langle X, g_m \rangle \, g_m$$

- Diagonal attenuation of each coefficient

$$\tilde{F} = DX = \sum_{m=0}^{N-1} a_m \, \langle X, g_m \rangle \, g_m \quad \text{with} \quad a_m \le 1 \ .$$

- Risk if W is a Gaussian white noise of variance σ^2

$$r(D, f) = \sum_{m=1}^{N} |\langle f, g_m \rangle|^2 (1 - a_m)^2 + \sum_{m=1}^{N} \sigma^2 \, |a_m|^2.$$

- Linear if a_m does not depend upon X
- How efficient are non-linear diagonal estimators ?

Linear Estimators

$$a_m = 1 \quad \text{for} \quad 0 \le m < M \quad \text{and} \quad a_m = 0 \quad \text{for} \quad M \le m \ .$$

- The risk depends upon the linear approximation error:

$$\begin{aligned} r(D, f) &= \sum_{m=M}^{N-1} |\langle f, g_m \rangle|^2 + M\,\sigma^2 \\ &= \|f - f_M\|^2 + M\,\sigma^2 \end{aligned}$$

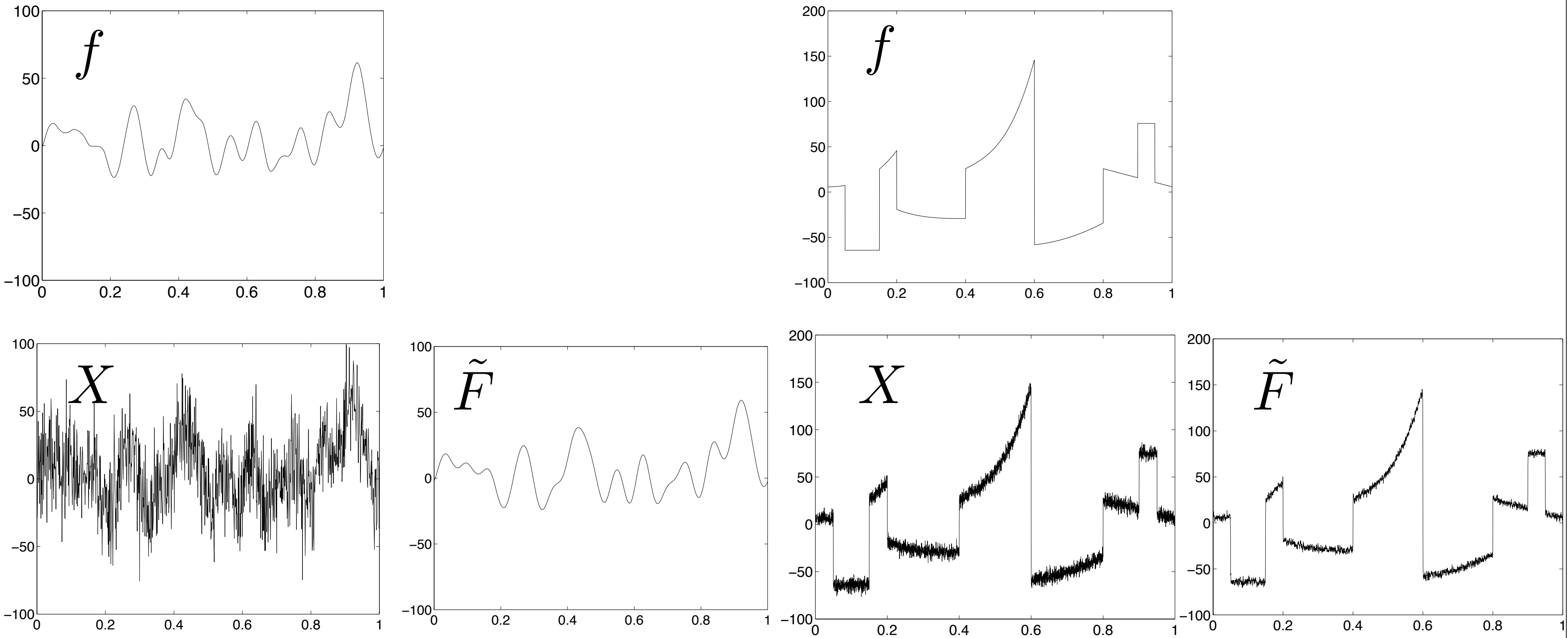
- M is adjusted so that

$$\|f - f_M\|^2 \sim M\sigma^2$$

Linear in a Fourier Basis

- In a discrete Fourier basis: $\{g_m[n] = N^{-1/2}\, e^{i2\pi mn/N}\}_{0 \le m < N}$

$$\tilde{F} = DX = X \star h \ \text{ with } \ \hat{h}[m] = a_m \ .$$



Non-Linear Oracle Estimation

- The risk of a diagonal estimation is:

$$r(D,f) = \sum_{m=1}^{N} |\langle f, g_m\rangle|^2 (1-a_m)^2 + \sum_{m=1}^{N} \sigma^2\, |a_m|^2$$

$$\text{with} \quad a_m \in \{0,1\} \ .$$

- To minimize the risk, an oracle will choose:

$$a_m = 1 \ \text{ if } \ |\langle f, g_m\rangle| \geq \sigma \ \text{ and } \ a_m = 0 \quad \text{otherwise} \quad .$$

The minimum risk depends upon the non-linear approximation error:

$$\begin{aligned} r_o(f) &= \sum_{|\langle f, g_m\rangle| \leq \sigma} |\langle f, g_m\rangle|^2 + M\sigma^2 \\ &= \|f - f_M\|^2 + M\sigma^2. \end{aligned}$$

Thresholding Estimation

A thresholding estimator D defined by

$$a_m(\langle X, g_m\rangle) = \begin{cases} 1 & \text{if } |\langle X, g_m\rangle| \geq T \\ 0 & \text{otherwise} \end{cases}$$

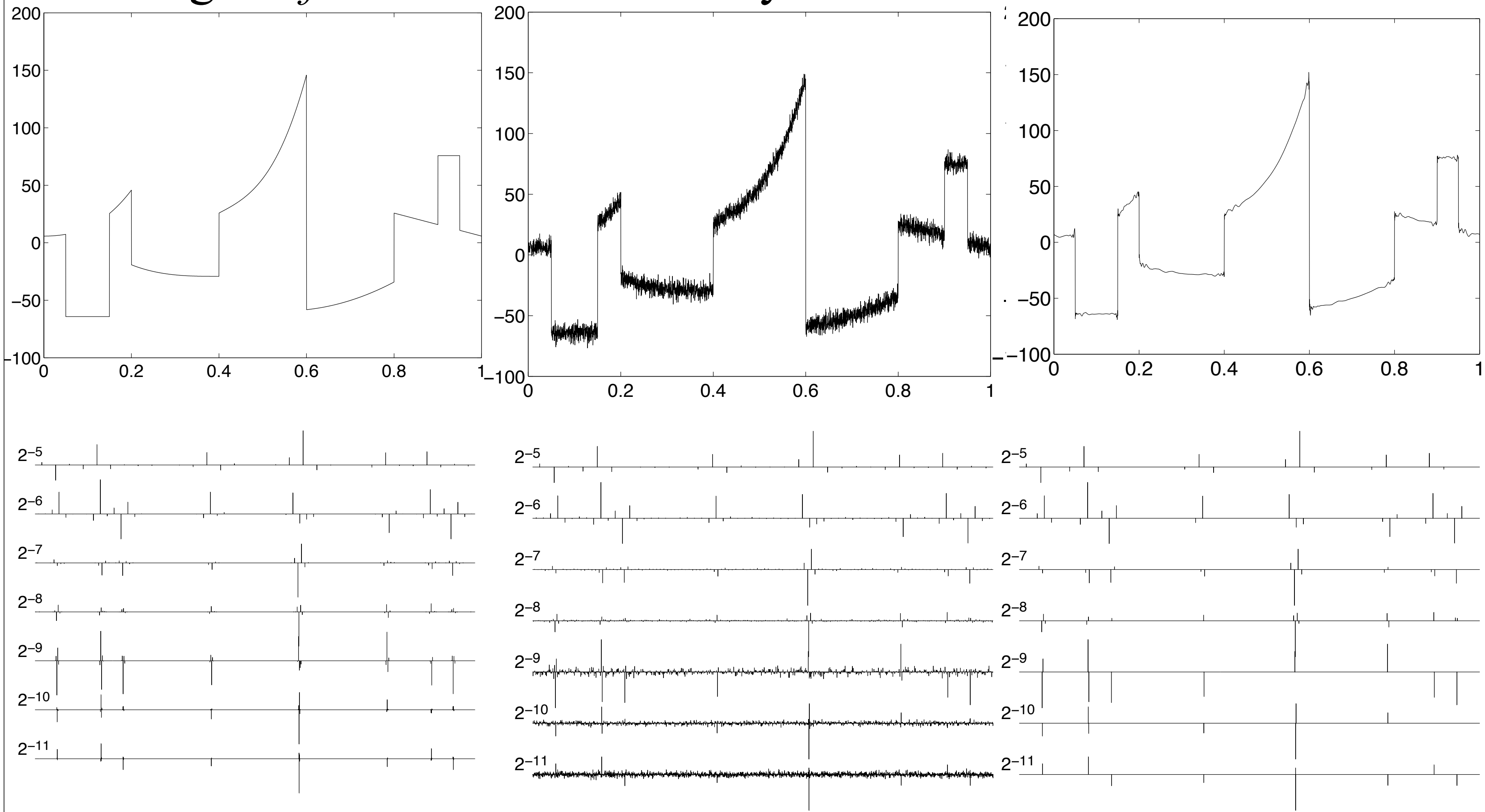
is nearly as good as an oracle estimator.

Theorem: If $T = \sigma\sqrt{2\log_e N}$ then

$$r(D,f) \leq (2\log_e N + 1)\left(\sigma^2 + r_o(f)\right) .$$

Wavelet Thresholding

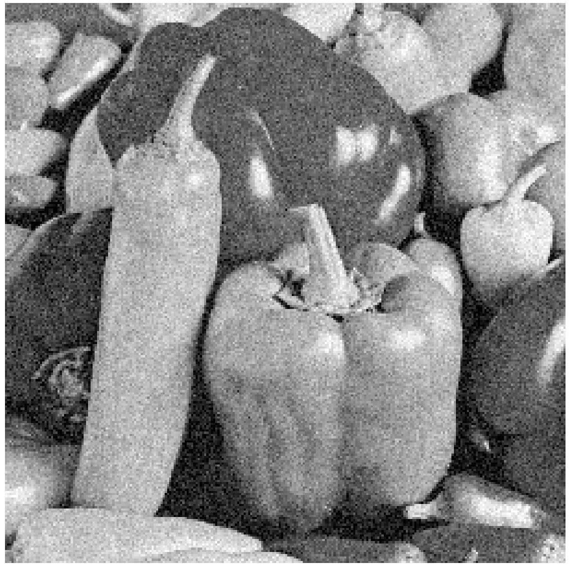
Original f | Noisy X | Transl Invariant Thresholded DX



Wavelet Image Thresholding

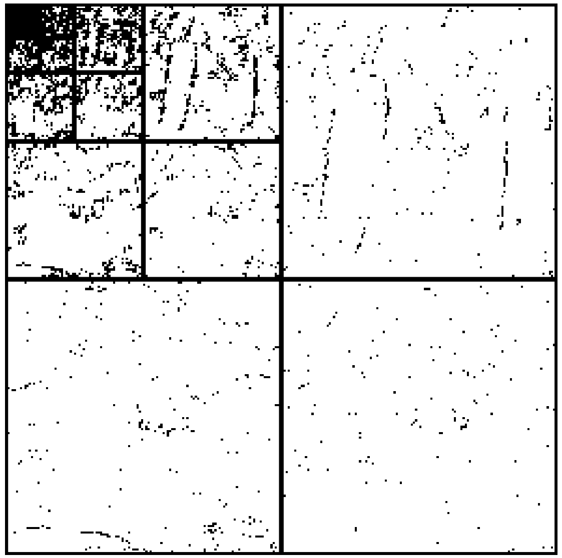
Original image f

Noisy image X



Translat. Invariant Thesh. estim. DX

Wavelet coeff. above T



1st Conclusion

- Sparse representation provide efficient compression and denoising estimators with simple diagonal operators.
- Linear approximation are sparse for "uniformly regular signals". Linear estimators are then nearly optimal.
- Non-linear approximations can adapt to more complex regularity.
- Wavelet are nearly optimal for piecewise regular one-dimensional signals. Good but not optimal for images.