Cargese, Corsica, July 2004

On Fourier and Wavelets: Representation, Approximation and Compression

Martin Vetterli
EPFL & UC Berkeley

1. Introduction through History
2. Fourier and Wavelet Representations
3. Wavelets and Approximation Theory
4. Wavelets and Compression
5. Going to Two Dimensions: Non-Separable Constructions
6. Beyond Shift Invariant Subspaces: Finite Rate of Innovation
7. Conclusions and Outlook

Acknowledgements

Collaborations:

- T.Blu, EPFL
- M.Do, UIUC
- P.L.Dragotti, Imperial College
- P.Marziliano, NIT Singapore
- I.Maravic, EPFL
- R.Shukla, EPFL
- C.Weidmann, TRC Vienna

Discussions and Interactions:

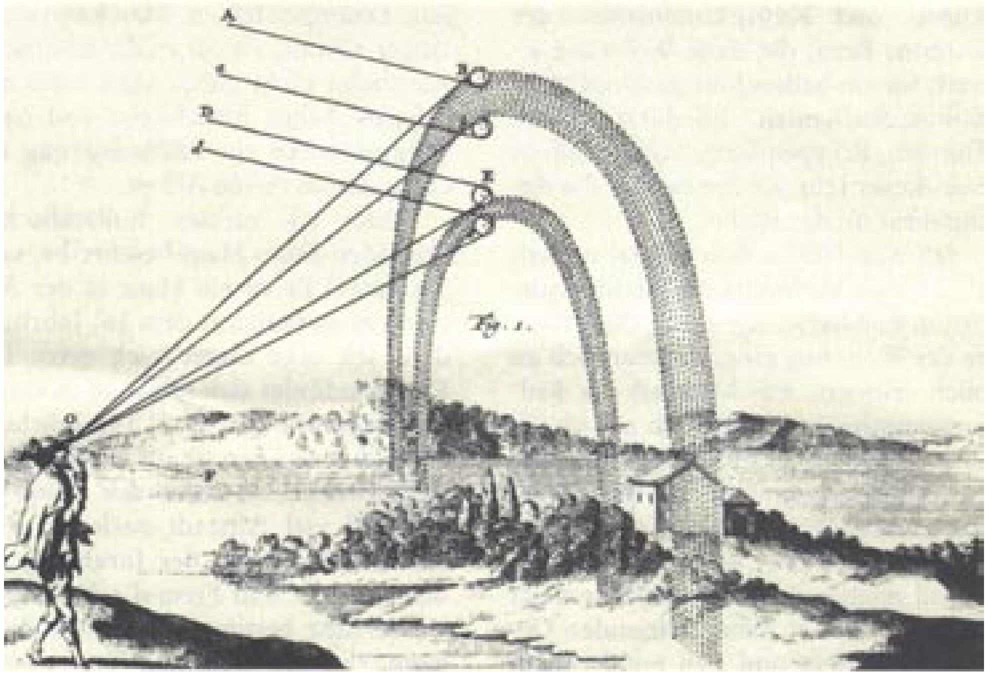
- A.Cohen, Paris VI
- I. Daubechies, Princeton
- R.DeVore, Carolina
- D. Donoho, Stanford
- M.Gastpar, Berkeley
- V.Goyal, MIT
- J. Kovacevic, CMU
- S. Mallat, Polytech. & NYU
- M.Unser, EPFL

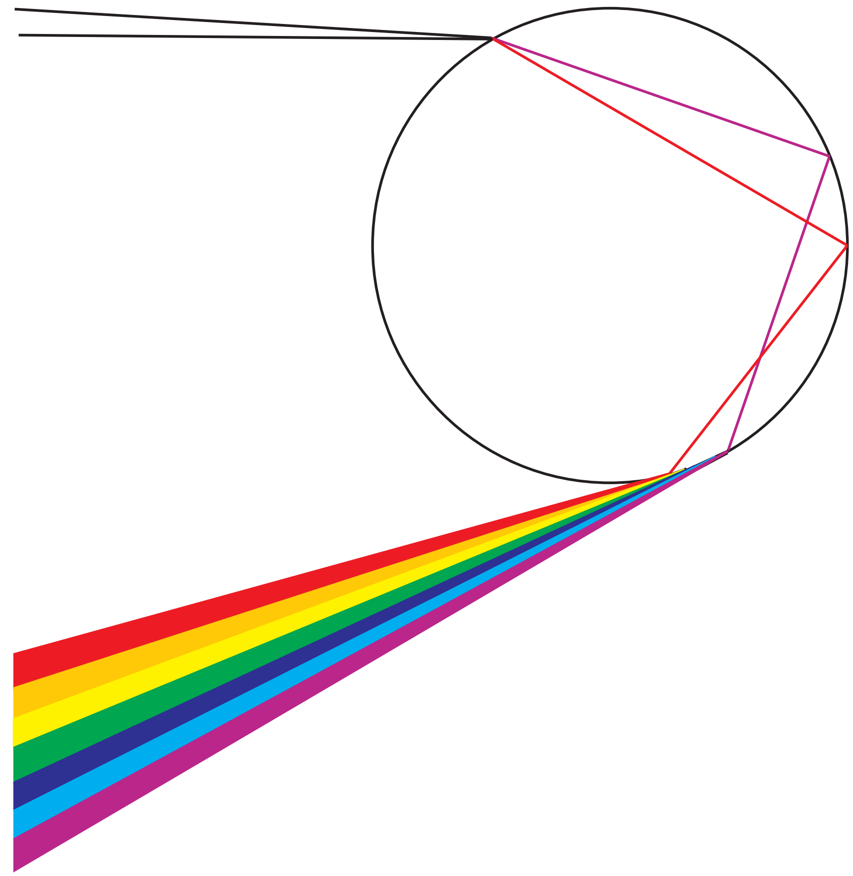
Sponsors: NSF Switzerland

Outline

1. **Introduction through History**
   - From Rainbows to Spectras
   - Signal Representations
   - Approximations
   - Compression
2. **Fourier and Wavelet Representations**
3. **Wavelets and Approximation Theory**
4. **Wavelets and Compression**
5. **Going to Two Dimensions: Non-Separable Constructions**
6. **Beyond Shift Invariant Subspaces**
7. **Conclusions and Outlook**

From Rainbows to Spectras



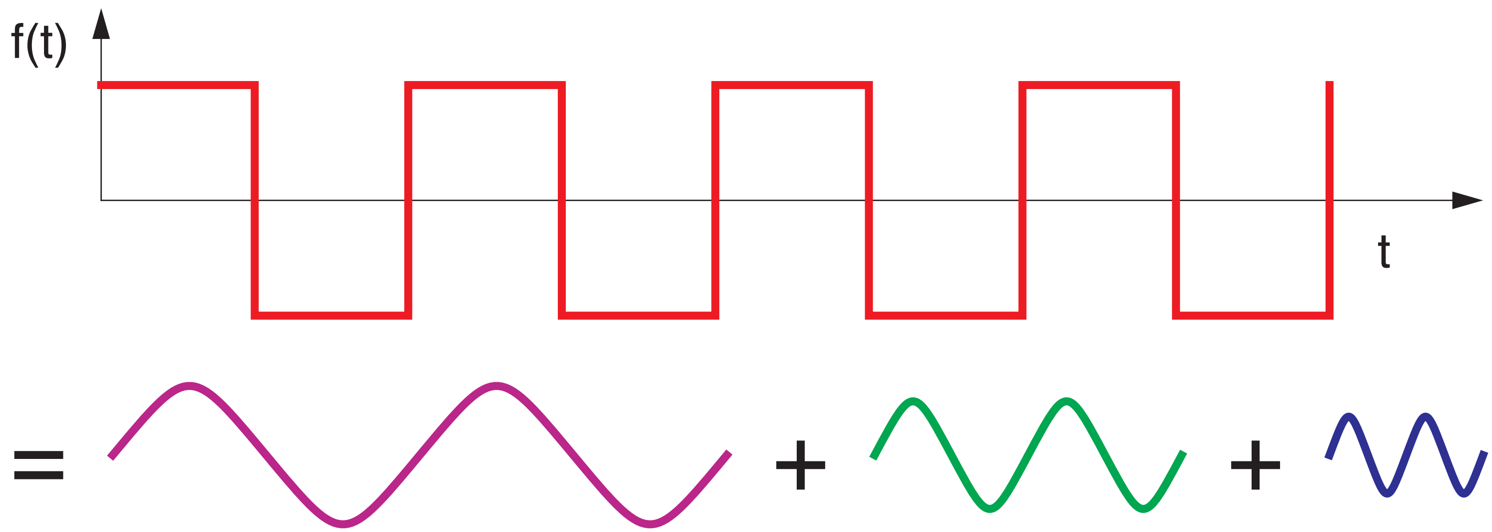


Von Freiberg, 1304: Primary and secondary rainbow

Newton and Goethe

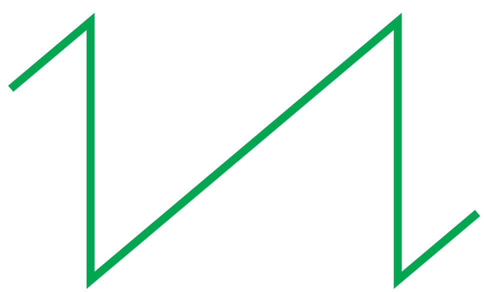
Signal Representations

1807: Fourier upsets the French Academy....



Fourier Series: Harmonic series, frequency changes, f_0, $2f_0$, $3f_0$, ...

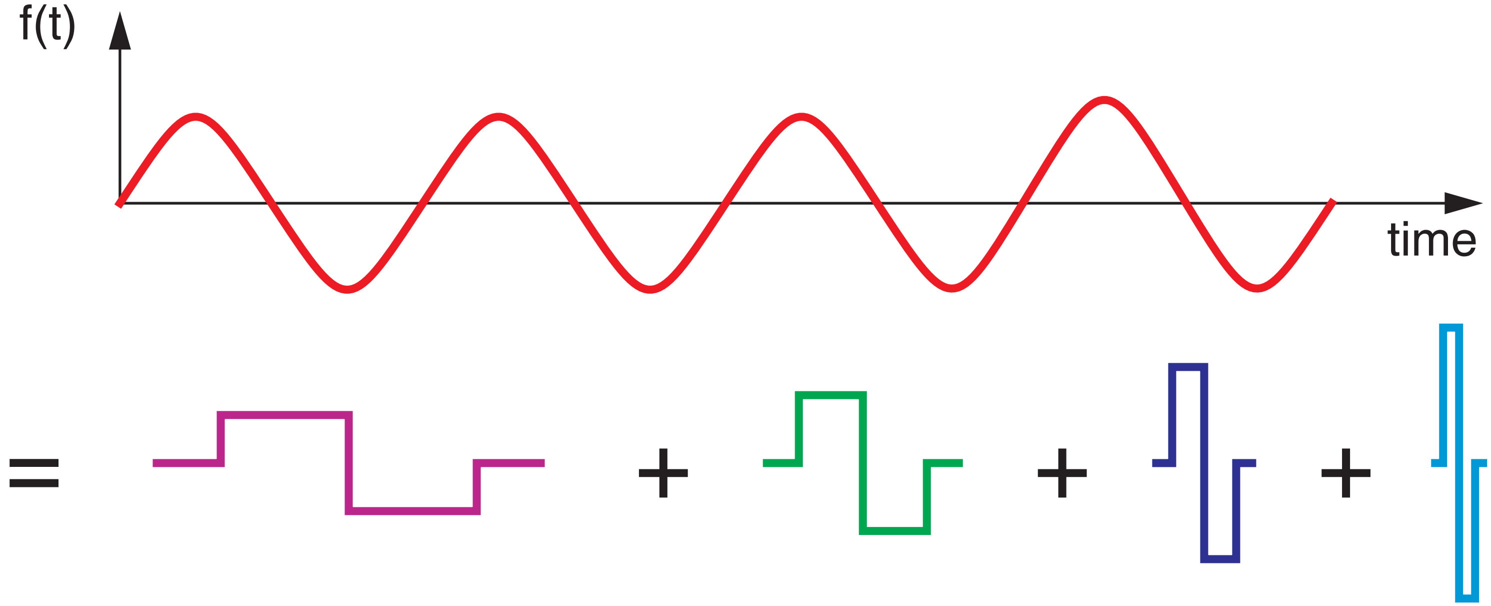
But... 1898: Gibbs' paper



1899: Gibbs' correction

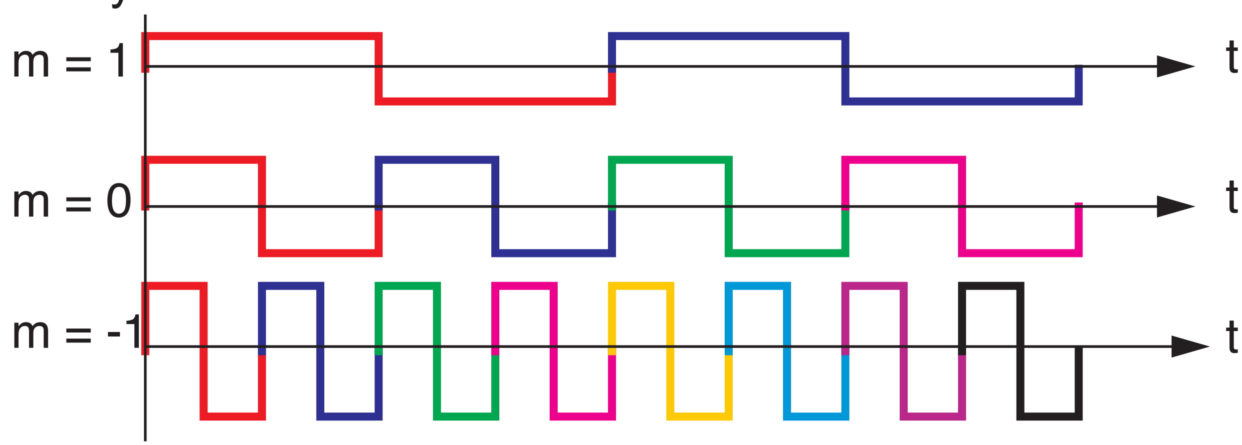
Orthogonality, convergence, complexity

1910: Alfred Haar discovers the Haar wavelet "dual" to the Fourier construction



Haar series:

- Scale changes S_0, $2S_0$, $4S_0$, $8S_0$ ...
- orthogonality



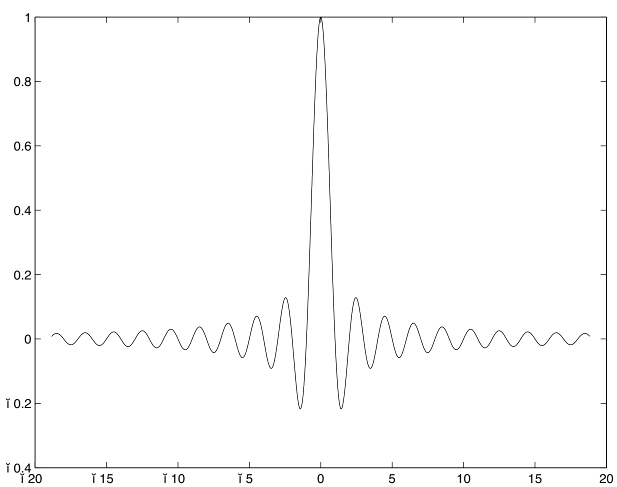
Theorem 1 (Shannon-48, Whittaker-35, Nyquist-28, Gabor-46)

If a function f(t) contains no frequencies higher than W cps, it is completely determined by giving its ordinates at a series of points spaced 1/(2W) seconds apart.

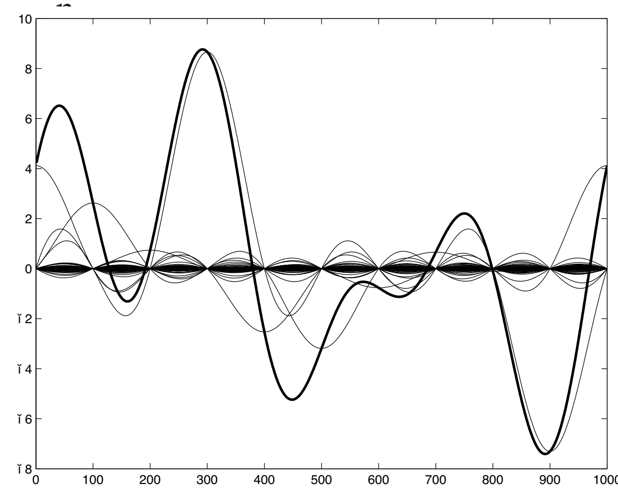
[if approx. T long, W wide, 2TW numbers specify the function]

It is a representation theorem:

- $\{\mathrm{sinc}(t-n)\}_{n \text{ in } Z}$, is an orthogonal basis for $BL[-\pi,\pi]$
- f(t) in $BL[-\pi,\pi]$ can be written as $f(t)) = \sum_n f(n) \cdot \mathrm{sinc}(t-n)$



... slow...!



Note:

- Shannon-BW, BL sufficient, not necessary.
- many variations, non-uniform etc
- Kotelnikov-33!

Representations, Bases and Frames

Ingredients:

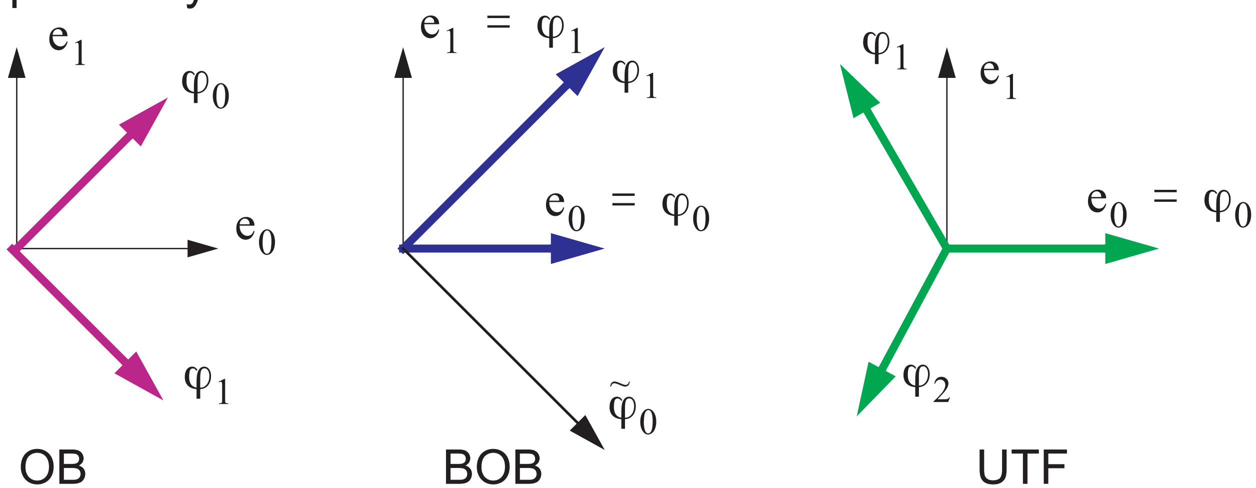
- as set of vectors, or ''atoms'', $\{\varphi_n\}$

- an inner product, e.g. $\langle \varphi_n, f \rangle = \int (\varphi_n \cdot f)$
- a series expansion

$$f(t) = \sum_n \langle \varphi_n, f \rangle \cdot \varphi_n(t)$$

Many possibilities:

- orthonormal bases (e.g. Fourier series, wavelet series)
- biorthogonal bases
- overcomplete systems or frames



Note: no transforms, uncountable

Approximations, aproximation...

The linear approximation method

Given an orthonormal basis $\{g_n\}$ for a space S and a signal

$$f = \sum_n \langle f, g_n \rangle \cdot g_n ,$$

the best linear approximation is given by the projection onto a fixed sub-space of size M (independent of f!)

$$\hat{f}_M = \sum_{n \in J_M} \langle f, g_n \rangle \cdot g_n$$

The error (MSE) is thus

$$\varepsilon_M = \|f - \hat{f}\|^2 = \sum_{n \notin J_M} |\langle f, g_n \rangle|^2$$

Ex: Truncated Fourier series
project onto first M vectors corresponding to largest expected inner products, typically LP

The Karhunen-Loeve Transform: The Linear View

Best Linear Approximation in an MSE sense:

Vector processes., i.i.d.:

$$X = [X_0, X_1, \ldots, X_{N-1}]^T \qquad E[X_i] = 0 \qquad E[X \cdot X^T] = R_X$$

Consider linear approximation in a basis

$$\hat{X}_M = \sum_{n=0}^{M-1} \langle X, g_n \rangle \cdot g_n \qquad M < N$$

Then:

$$E[\varepsilon_M] = \sum_{n=M}^{N-1} \langle R_X g_n, g_n \rangle$$

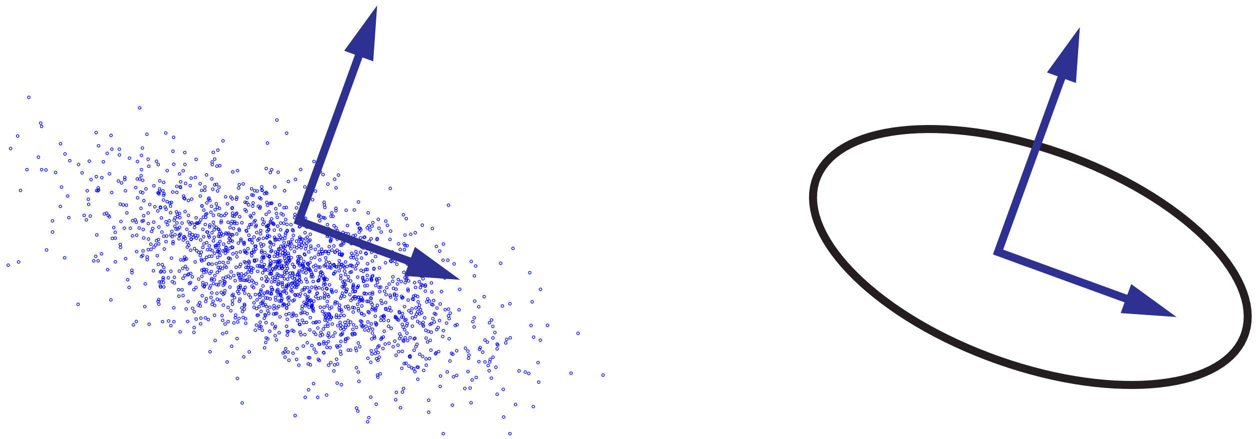
Karhunen-Loeve transform (KLT):

For $0<M<N$, the expected squared error is minimized for the basis $\{g_n\}$ where g_m are the eigenvectors of R_X ordered in order of decreasing eigenvalues.

Proof: eigenvector argument inductively.

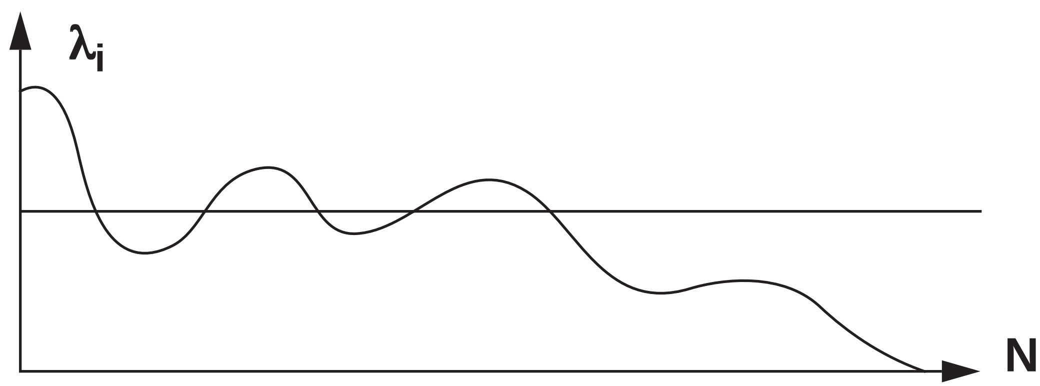
Note: Karhunen-47, Loeve-48, Hotelling-33, PCA, KramerM-56, TC

Geometric intuition: Principal axes of distribution:



Shapes: ellipsoids

To first approximation, keep all coefficients above a threshold:



This can be used in many settings, classification, denoising, and compression (inverse waterfilling thm)

Compression: How many bits for Mona Lisa?

<=> {0,1}

A few numbers...

D.Gabor, September 1959 (Editorial IRE)

"... the 20 bits per second which, the psychologists assure us, the human eye is capable of taking in, ..."

Index all pictures ever taken in the history of mankind

- $100 \text{ years} \cdot 10^{10} \sim 44$ bits

Huffman code Mona Lisa index

- a few bits (Lena Y/N?, Mona Lisa...), what about $R(D)$....

Search the Web!

- http://www.google.com, 5-50 billion images online, or 33-36 bits

JPEG

- 186K... There is plenty of room at the bottom!
- JPEG2000 takes a few less, thanks to wavelets...

Note: $2^{(256x256x8)}$ possible images (D.Field)

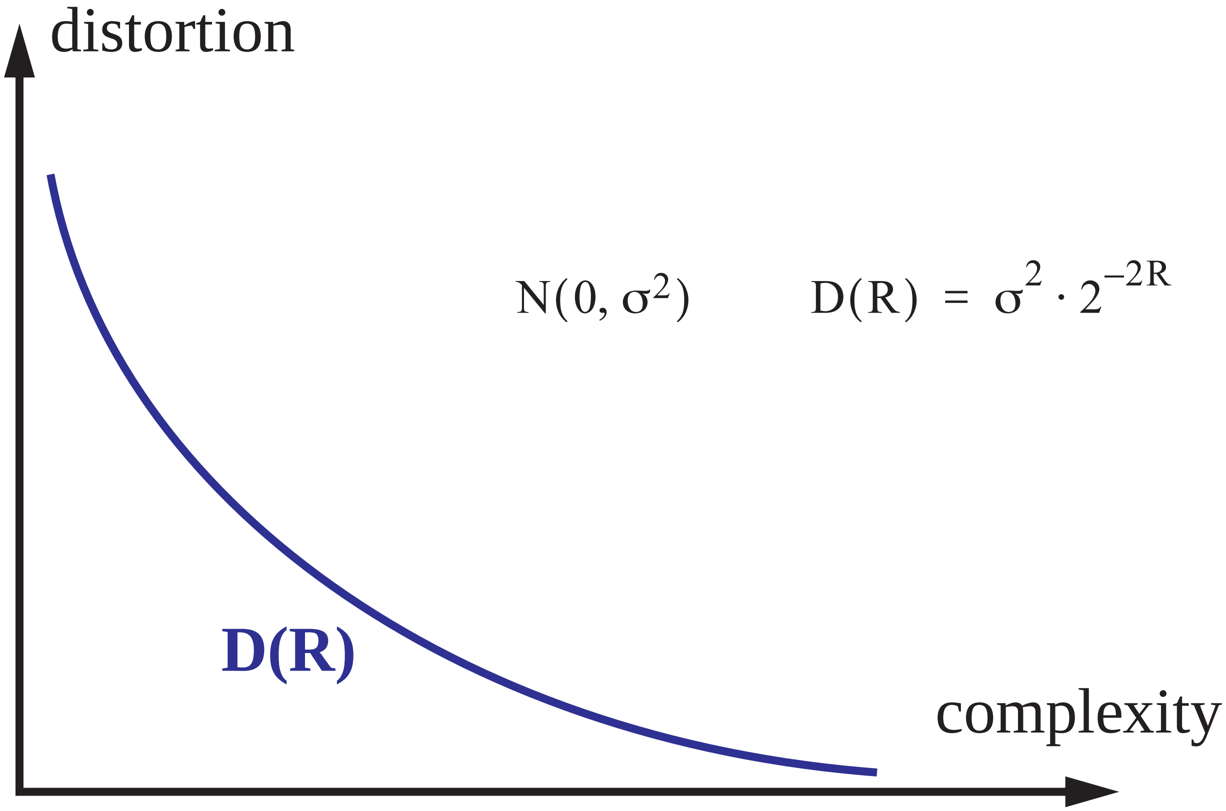
Homework in Cover-Thomas, Kolmogorov, MDL, Occam, DNA, etc

(from a contemporary: 0 bits, I don't care for this modern stuff)

Source Coding: some background

Exchanging description complexity for distortion:

- rate-distortion theory [Shannon:58, Berger:71]
- known in few cases...like i.i.d. Gaussians (but tight: no better way!)



$N(0, \sigma^2)$ $D(R) = \sigma^2 \cdot 2^{-2R}$ or -6dB/bit

- typically: difficult, simple models, high complexity (e.g. VQ)
- high rate results, low rate often unknown

Limitations of the Standard Models

“Splendeurs et misères de la fonction débit-distortion” (after Balsac)

Precise results

- beautiful (maybe too much for its own good)
- upper and lower bounds
- constructive

Problems

- complexity: exponential in code length
- code construction: finding good codes is hard

Paradox:

- Best codes used in practice are suboptimal (Effros)
- transform codes dominate the scene of ‘’real’’ compression

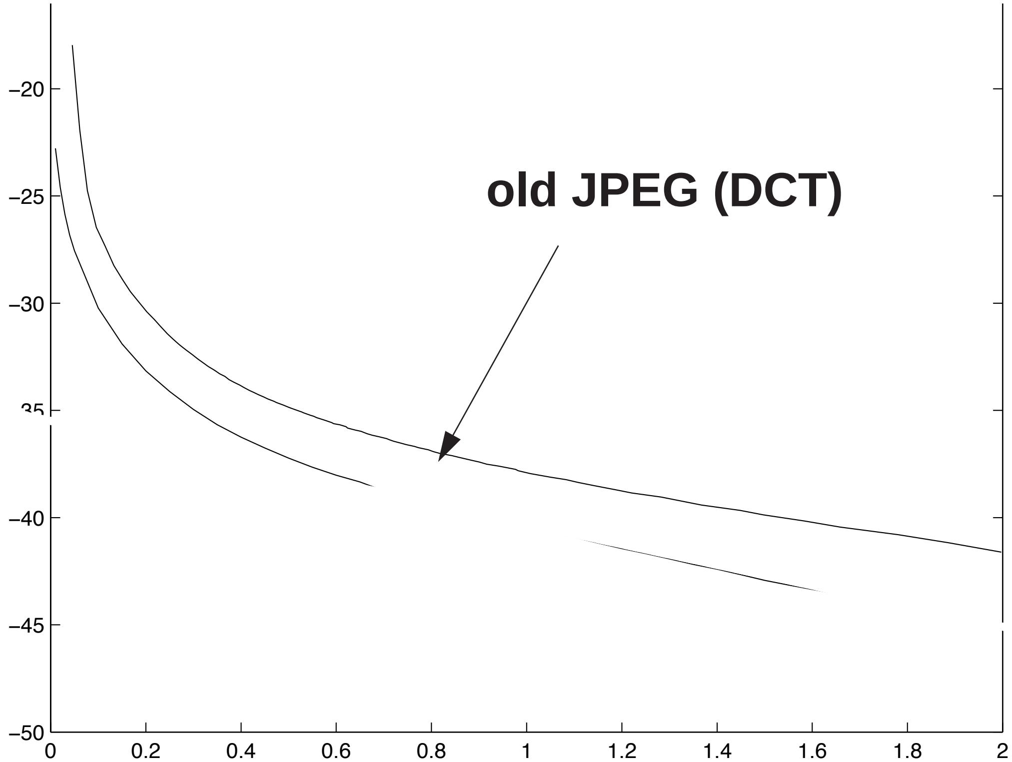
Audio/Image/Video: distortion measures?

So: unlike in lossless compression, lossy compression uses IT in a limited way;)

New image coding standard ... JPEG 2000

Old versus new JPEG: D(R) on log scale

-6 dB/bit



new JPEG
(wavelets)

Main points:

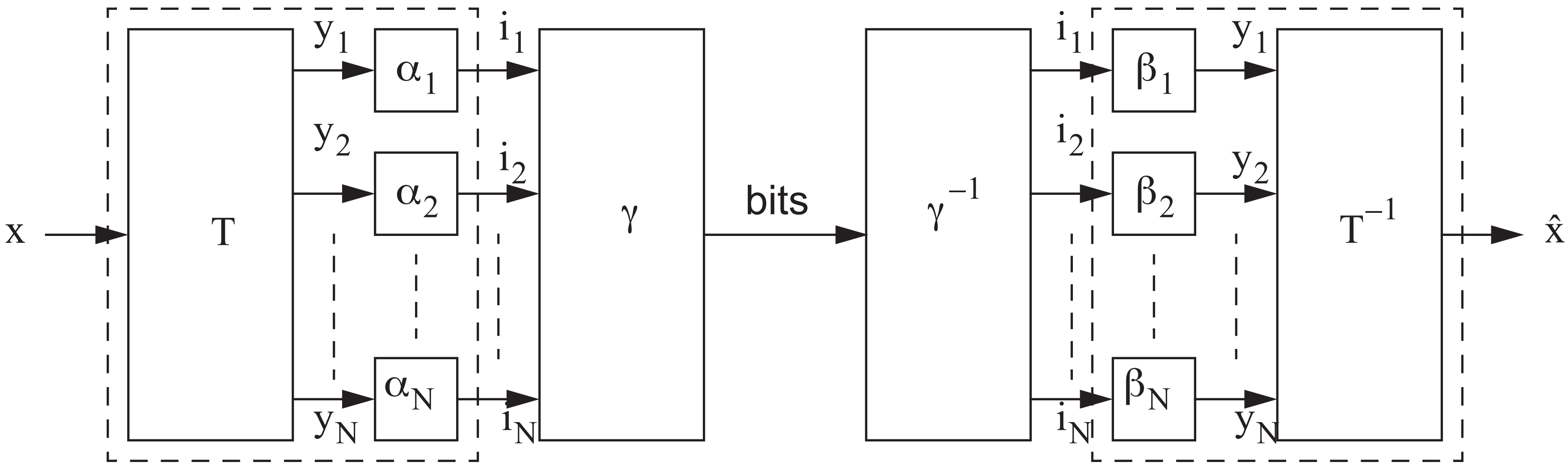
- improvement by a few dB’s
- lot more functionalities (e.g. progressive download on internet)
- at high rate ~ -6db per bit: KLT behavior
- low rate behavior: much steeper: NL approximation effect?
- is this the limit?

The Swiss Army Knife Formula of Transform Coding [Goyal00]

Model: iid vector process of size N, μ=0, R_x, MSE, high rate

- vector quantizer, entropy code

- transform and scalar quantizers, entropy code



$$D(R) = \frac{1}{12N} \cdot \mathrm{tr}(T^{-1}(T^{-1})^T) \cdot 2^{\left(\frac{2}{N}\Sigma h(y_i)\right)} \cdot 2^{-2R}$$

Trace min: ortho; diff. entropy min: independence

- Gaussian case: coincide! but in general not...

Representation, Approximation and Compression: Why does it matter anyway?

Parsimonious or sparse representation of visual information is key in

- storage and transmission
- indexing, searching, classification, watermarking
- denoising, enhancing, resolution change

But: it is also a fundamental question in

- information theory
- signal/image processing
- approximation theory
- vision research

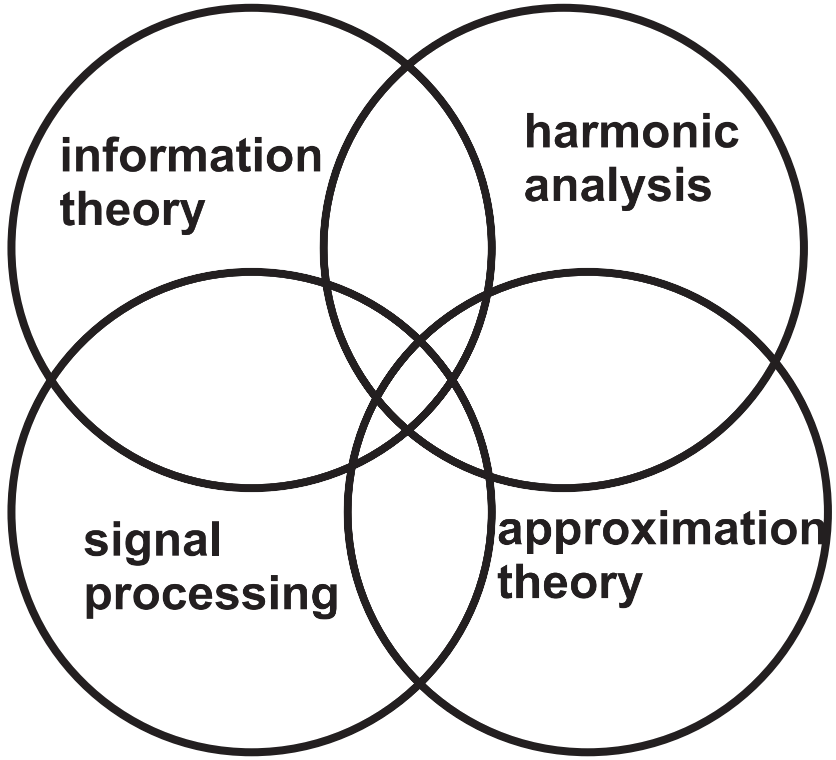
Successes of wavelets in image processing:

- compression (JPEG2000)
- denoising
- enhancement
- classification

Thesis: Wavelet models play an important role

Antithesis: Wavelets are just another fad!

Interaction of topics



- AT: deterministic setting, large classes of fcts
- HA: function classes, existence, embeddings
- IT: boundings, converses, stochastic setting
- SP: bases, algorithms, complexity

The interaction is the fun!

Outline

1. **Introduction through History**
2. **Fourier and Wavelet Representations**
   - Fourier and Local Fourier Transforms
   - Wavelet Transforms
   - Piecewise Smooth Signal Representations
3. **Wavelets and Approximation Theory**
4. **Wavelets and Compression**
5. **Going to Two Dimensions: Non-Separable Constructions**
6. **Beyond Shift Invariant Subspaces: Finite Rate of Innovation**
7. **Conclusions and Outlook**

2. Fourier and Wavelet Representations: Spaces

Norms: $\|x\|_p = \left(\sum_n |x[n]|^p\right)^{1/p}$ $\qquad \|f\|_p = \left(\int_{-\infty}^{\infty} |f(t)|^p dt\right)^{1/p}$

Hilbert spaces: $l_2(Z) = \{x:(\|x\|_2 < \infty)\}$ $\qquad L_2(R) = \{f:(\|f\|_2 < \infty)\}$

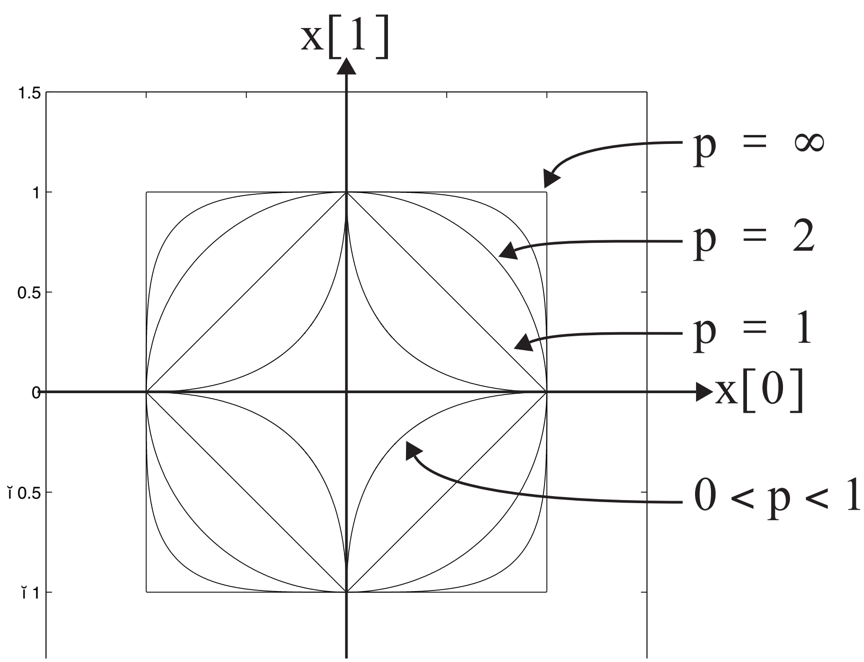
Inner product: $\langle x, y\rangle = \sum_n x^*[n]y[n]$ $\qquad \langle f, g\rangle = \int f^*(t)g(t)dt$

Orthogonality: $x \perp y \Leftrightarrow \langle x, y\rangle = 0$

Banach spaces:

$x, f \;\; \text{s.t.} \;\; \|x\|_p, \|f\|_p < \infty \qquad p \text{ general}$

p-norm = 1



A Tale of Two Representations: Fourier versus Wavelets

Orthonormal Series Expansion

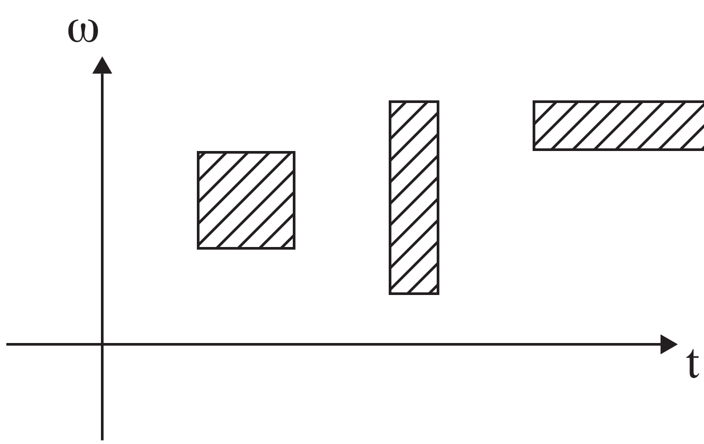
$$f = \sum_{n \in Z} \alpha_n \varphi_n \qquad \alpha_n = \langle \varphi_n, f \rangle \qquad <\varphi_n, \varphi_m> = \delta_{n-m} \qquad \|f\|_2 = \|\alpha\|_2$$

Time-Frequency Analysis and Uncertainty Principle

$$f(t) \leftrightarrow F(\omega) \qquad \Delta^2 t = \int t^2 |f(t)| dt \qquad \Delta^2 \omega = \int \omega^2 |F(\omega)| d\omega$$

Then

$$\Delta^2 t \cdot \Delta^2 \omega \geq \frac{\pi}{2}$$



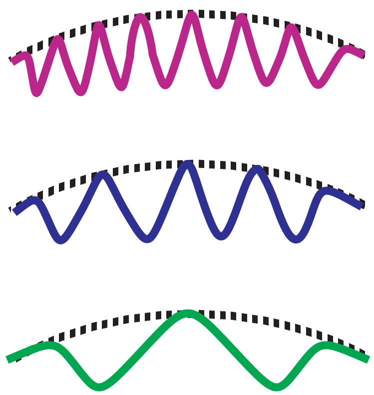
not arbitrarily sharp
in time and frequency

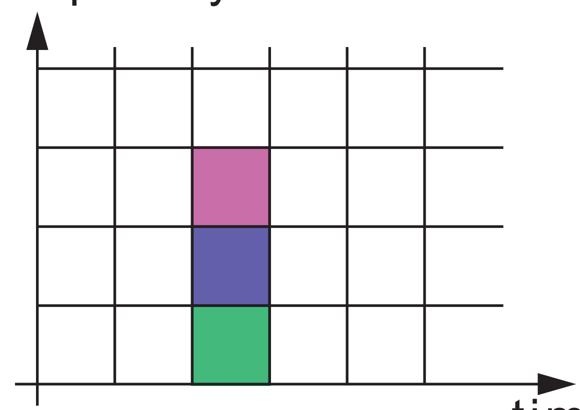
Local Fourier Basis?

The Gabor or Short-time Fourier Transform

$$\varphi_{m,n}(t) = w(t-nT)e^{-jm\omega_0(t-nT)}$$

Time-frequency atoms localized at $(nT, m\omega_0)$





When T, ω_0 “small enough”

$$f(t) \approx c \cdot F_{m,n}\varphi_{m,n}(t) \text{ where } F_{m,n} = \langle\varphi_{m,n}, f\rangle$$

Example: Spectrogram

The Bad News...

Balian-Low Theorem

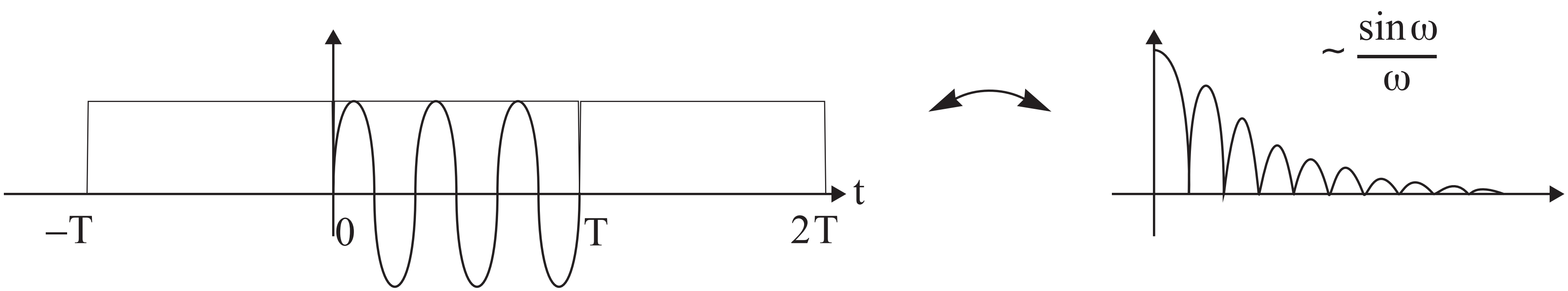
$\varphi_{m,n}$ is a short-time Fourier frame with critical sampling ($T\omega_0 = 2\pi$)

then either

$$\Delta^2 t = \infty \text{ or } \Delta^2 \omega = \infty$$

or: there is no good local orthogonal Fourier basis!

Example of a basis: block based Fourier series



Note: consequence of BL Thm on OFDM, RIAA

The Good News!

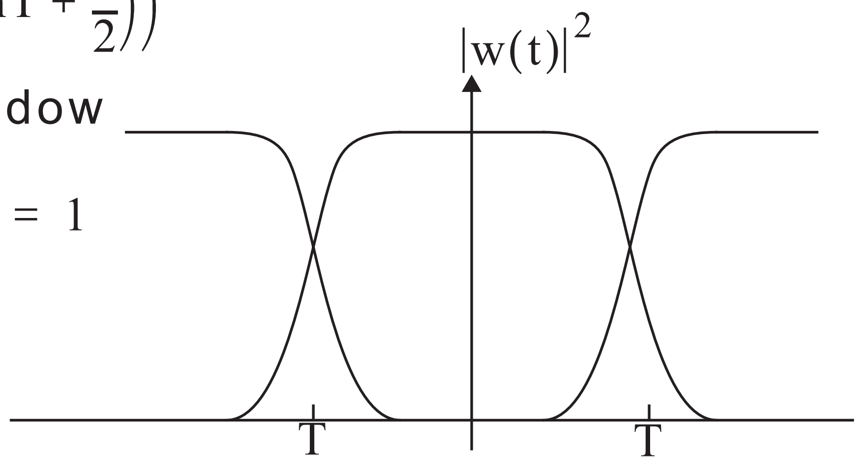
There exist good local cosine bases.

Replace complex modulation ($e^{jm\omega_0 t}$) by appropriate cosine modulation

$$\varphi_{m,n}(t) = w(t-nT)\cos\left(\frac{\pi}{2}\left(m+\frac{1}{2}\right)\left(t-nT+\frac{T}{2}\right)\right)$$

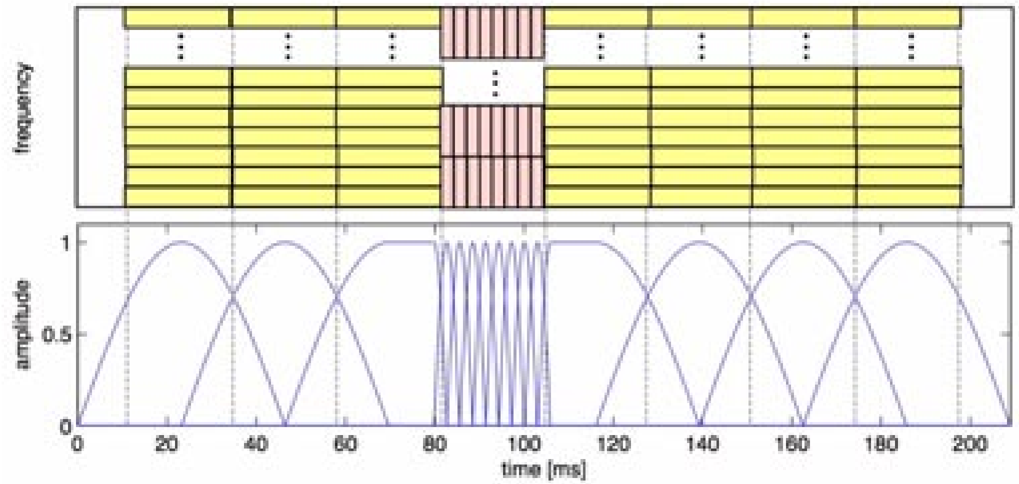
where $w(t)$ is a power complementary window

$$\sum_n |w(t-nT)| = 1$$



Result: MP3!

Many generalisations...



Another Good News!

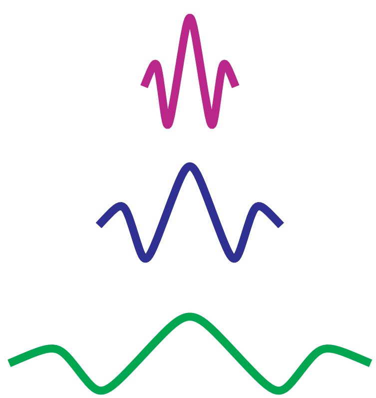
Replace (shift, modulation)

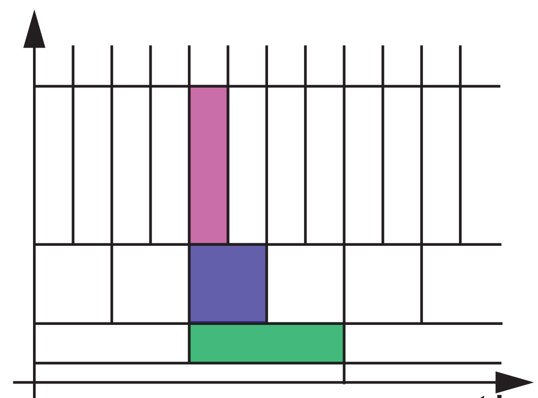
by (shift, scale)

or

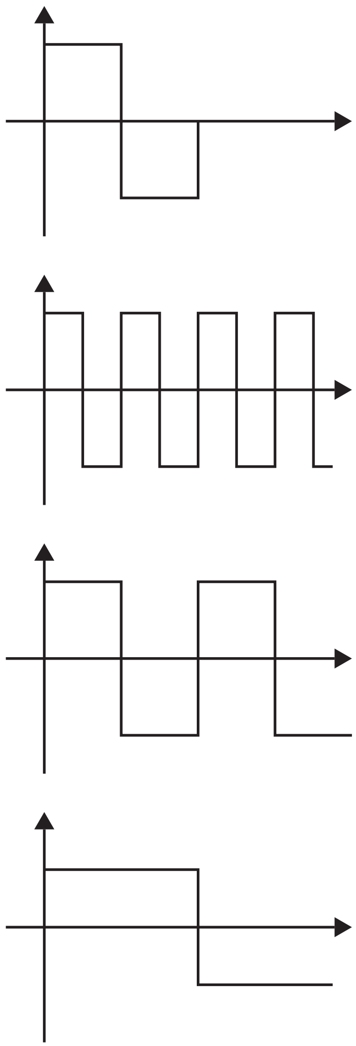
$$\Psi_{m,n}(t) = 2^{-m/2}\Psi\left(\frac{t-2^m n}{2^m}\right) \qquad n, m \in Z$$

then there exist "good" localized orthonormal bases, or wavelet bases

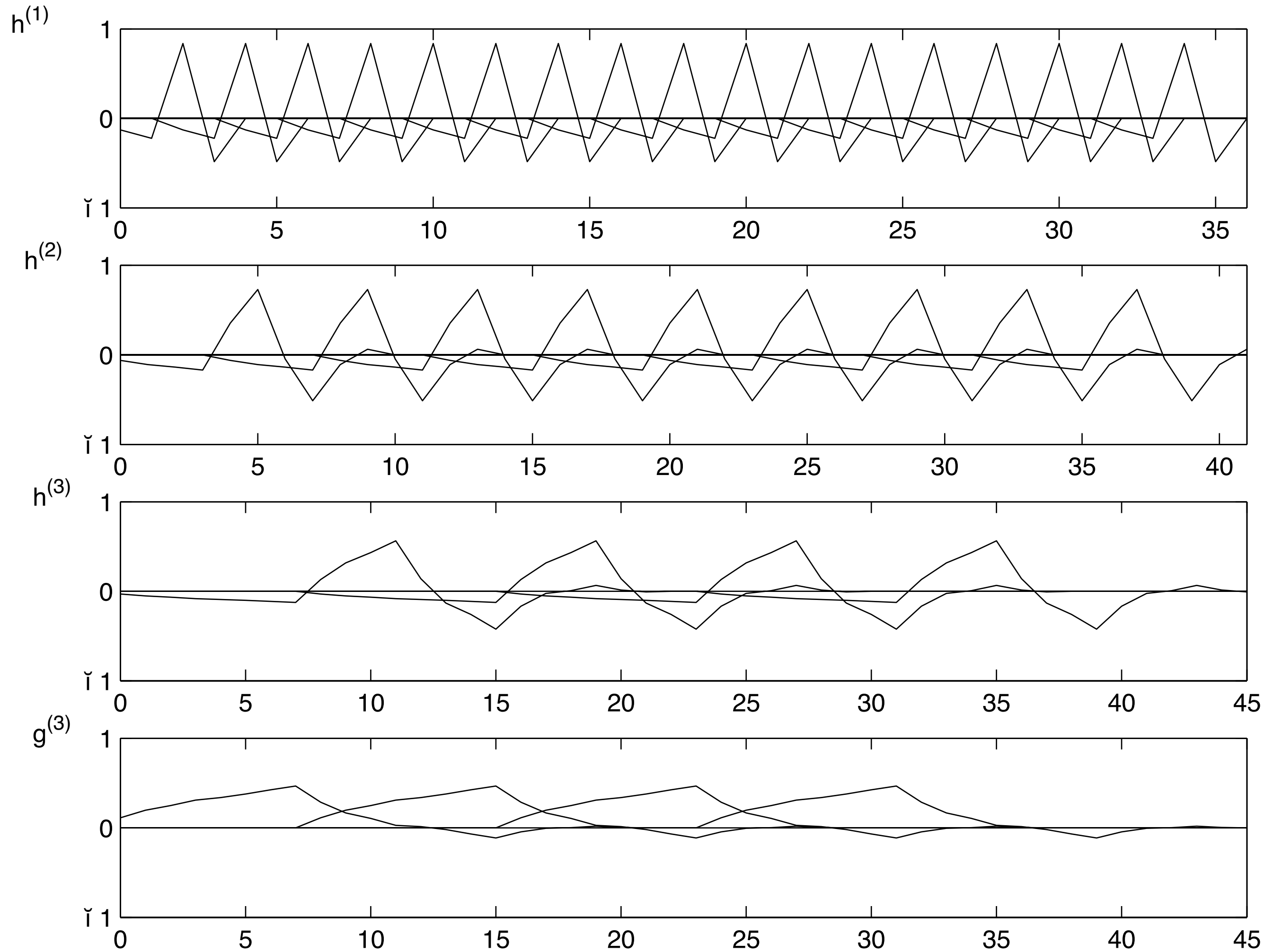




Examples of bases



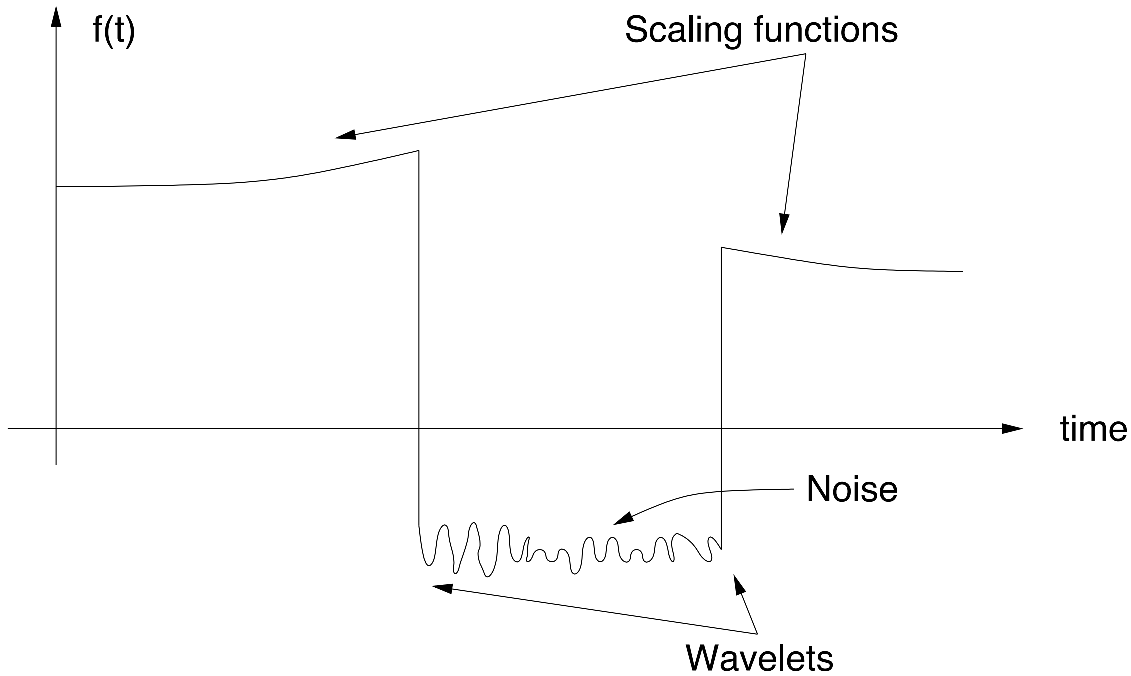
Haar



Daubechies, D_2

Wavelets and representation of piecewise smooth functions

Goal: efficient representation of signals like:



where:

- Wavelet act as singularity detectors
- Scaling functions catch smooth parts
- “Noise” is circularly symmetric

Note: Fourier gets all Gibbs-ed up!

Key characteristics of wavelets and scaling functions

Wavelets derived from filter banks, ortho-LP with N zeroes at π, [Daubechies-88],

$$G(z) = (1 + z^{-1})^N \cdot R(z)$$

Scaling function: $\phi(\omega) = \prod_{i=1}^{\infty} G\left(e^{j(\omega/(2^i))}\right)$

Orthonormal wavelet family: $\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t - n)$

Scaling function and approximations

- Strang-Fix theorem: if $\phi(\omega)$ has N zeros at multiples of 2π (but the origin), then $\{\varphi(t-n)\}_{n \in Z}$ spans polynomials up to degree N-1

$$\sum_n c_n \cdot \varphi(t-n) = t^k \qquad k = 0, 1, \ldots N-1$$

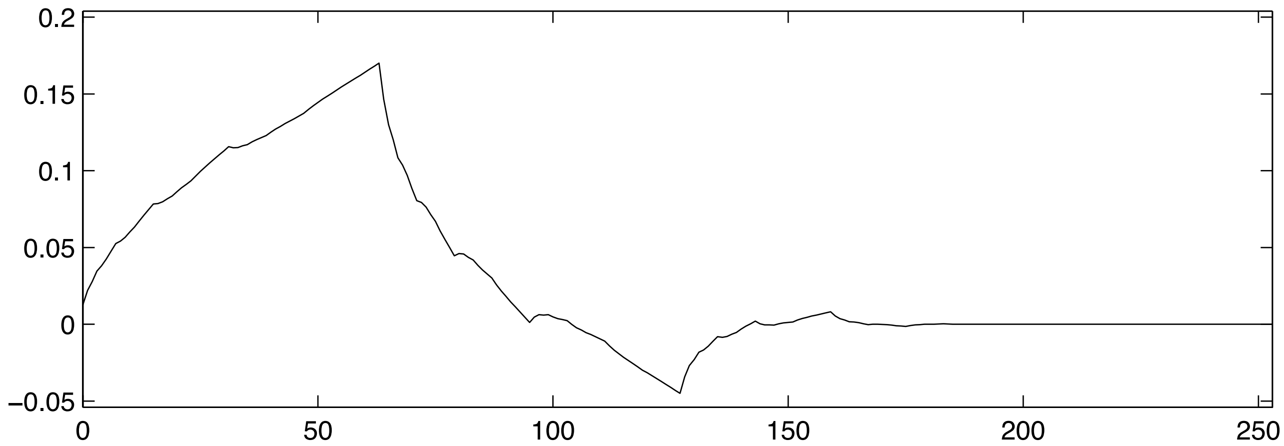
- Two scale equation:

$$\varphi(t) = \frac{1}{\sqrt{2}} \cdot \sum_n g_n \cdot \varphi(2t - n)$$

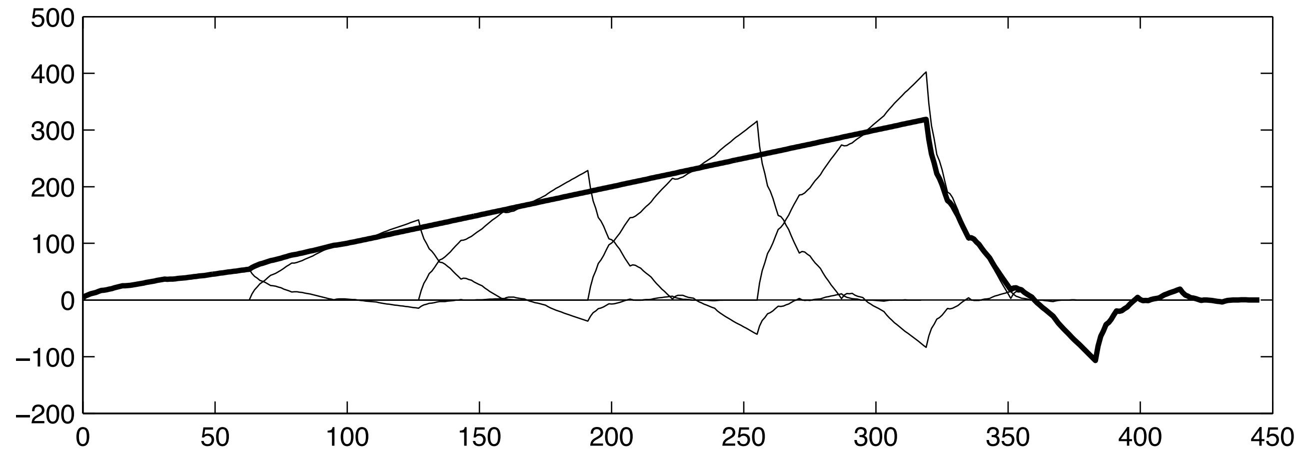
- smoothness: follows from N, α = 0,203 N

Lowpass filters and scaling functions reproduce polynomials

- Iterate of Daubechies L=4 lowpass filter reproduces linear ramp



scaling function



linear ramp

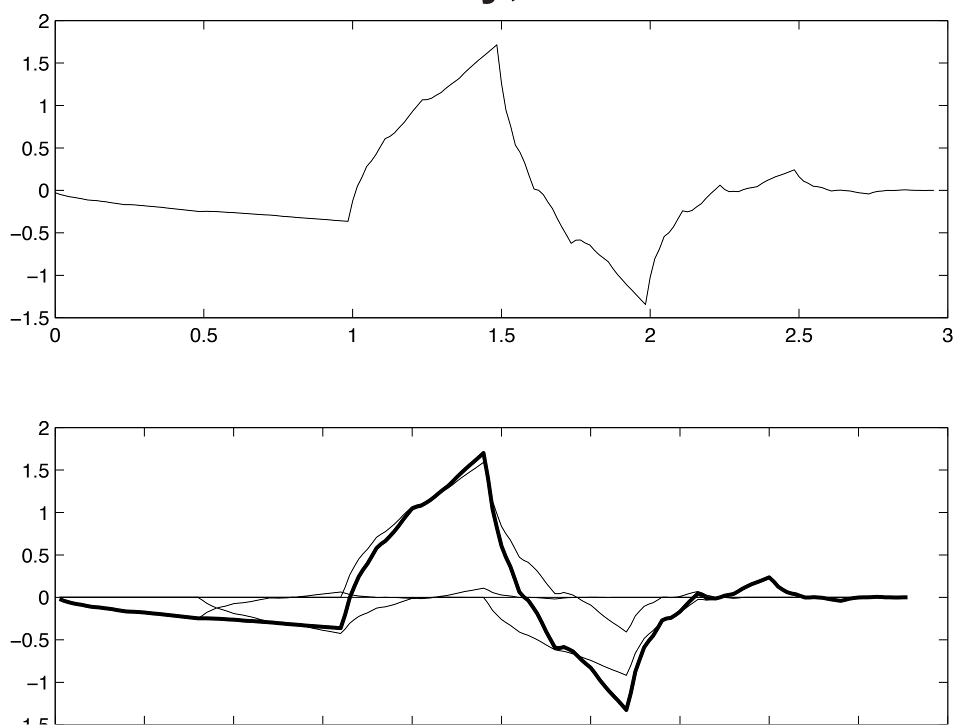
Scaling functions catch ‘’trends’’ in signals

Wavelet approximations

- wavelet ψ has N zero moments, kills polynomials up to deg. N-1
- wavelet of length L= 2N-1, or 2N-1 coeffs influenced by singularity at each scale, wavelet are singularity detectors,
- wavelet coefficients of smooth functions decays fast, e.g. f in c^p,m << 0

$$\langle \psi_{m,n}, f \rangle = c 2^{m\left(p - \frac{1}{2}\right)}$$

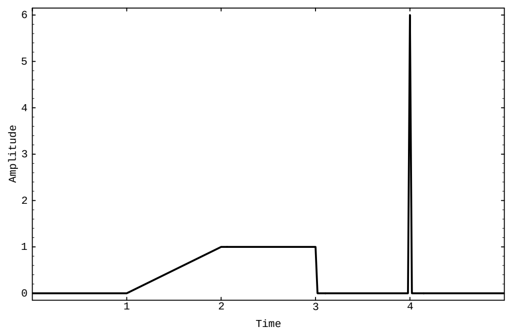
Note: all this is in 1 dimension only, 2D is another story...

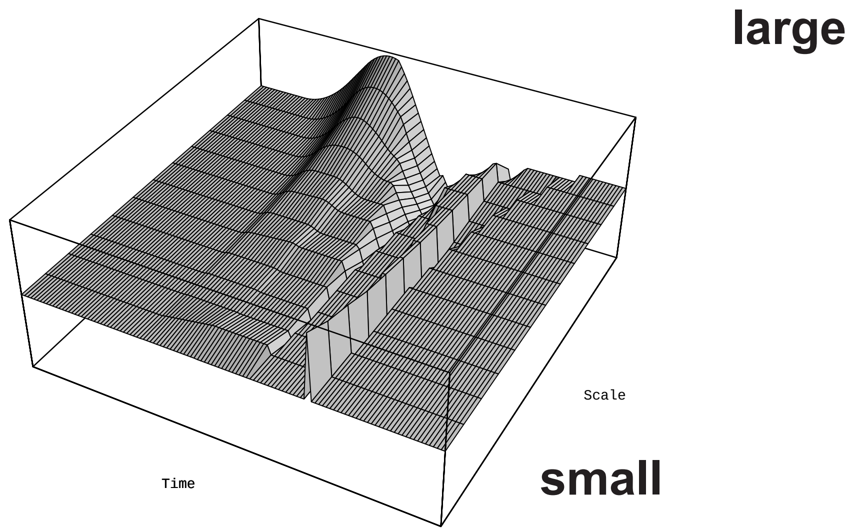


How about singularities?

If we have a singularity of order n at the origin (0: Dirac, 1: Heaviside,...), the CWT transform behaves as

$$X(a, 0) = c_n \cdot a^{\left(n - \frac{1}{2}\right)}$$



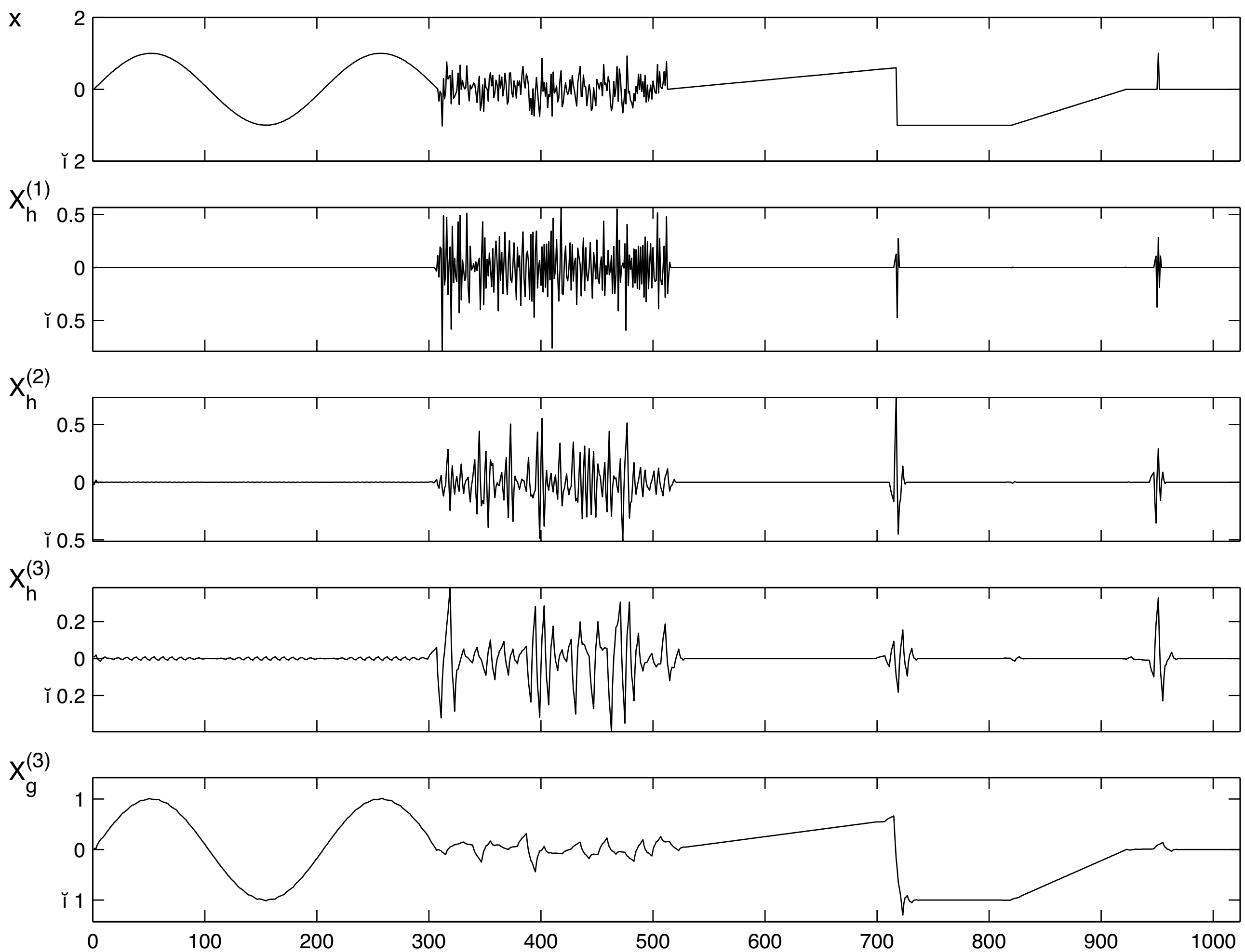


In the orthogonal wavelet series: same behavior, but only $L=2N-1$ coefficients influenced at each scale!

- e.g. Dirac/Heaviside: behavior as $2^{-m/2}$ and $2^{m/2}$, m<<0

Wavelets catch and characterize singularities!

Thus: a piecewise smooth signal expands as:



- lowpass catches trends, polynomials
- a singularity influences only L wavelets at each scale
- wavelet coefficients decay fast

Outline

1. **Introduction through History**

2. **Fourier and Wavelet Representations**

3. **Wavelets and Approximation Theory**
    - Sobolev and Besov spaces
    - Non-linear approximation
    - Fourier versus wavelet, LA versus NLA

4. **Wavelets and Compression**

5. **Going to Two Dimensions: Non-Separable Constructions**

6. **Beyond Shift Invariant Subspaces: Finite Rate of Innovation**

7. **Conclusions and Outlook**

More Spaces

C^p spaces: p-times diff. with bounded derivatives
-> Taylor expansions

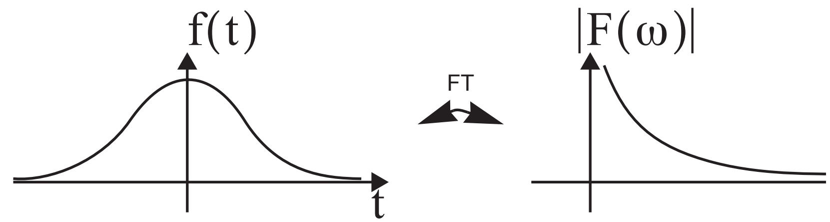
Holder/Lipschitz α: locally α smooth (non-integer)

Sobolev Spaces $W^s(R)$

$$f \in l^2(R) \qquad \int_{-\infty}^{\infty} |\omega|^{2s} |F(\omega)|^2 d\omega < \infty$$

If $s > n + \frac{1}{2}$ then f is n-times continuously differentiable

Equivalently $F(\omega)$ decays at $\dfrac{1}{(1+|\omega|)^{s+1/2+\varepsilon}}$



Besov Spaces $B_p(I)$ with respect to a basis (typically wavelets)

$$f \in l^2(I)$$

$$\|f\|_{B,p} = \left(\sum_m \sum_n |\langle \Psi_{m,n}, f \rangle|^p \right)^{1/p} < \infty$$

or wavelet expansion has finite l_p norm

From linear to non-linear approximation theory

The non-linear approximation method

Given an orthonormal basis $\{g_n\}$ for a space S and a signal

$$f = \sum_n \langle f, g_n \rangle \cdot g_n ,$$

the best nonlinear approximation is given by the projection onto an adapted subspace of size M (dependent on f!)

$$\tilde{f}_M = \sum_{n \in I_M} \langle f, g_n \rangle \cdot g_n$$

$$I_M: \qquad \left|\langle f, g_n \rangle\right|_{n \in I_M} \geq \left|\langle f, g_m \rangle\right|_{m \notin I_M} \qquad \text{set of } M \text{ largest } \langle\ ,\ \rangle$$

The error (MSE) is thus

$$\tilde{\varepsilon}_M = \left\|f - \tilde{f}\right\|^2 = \sum_{n \notin I_M} \left|\langle f, g_n \rangle\right|^2$$

and $\tilde{\varepsilon}_M \leq \varepsilon_M$.

Difference: take the first M coeffs (linear) or
take the largest M coeffs (non-linear)

Nonlinear approximation

- This is a simple but nonlinear scheme
- Clearly, if $A_M(.)$ is the NL approximation scheme:

$$A_M(x) + A_M(y) \neq A_M(x + y)$$

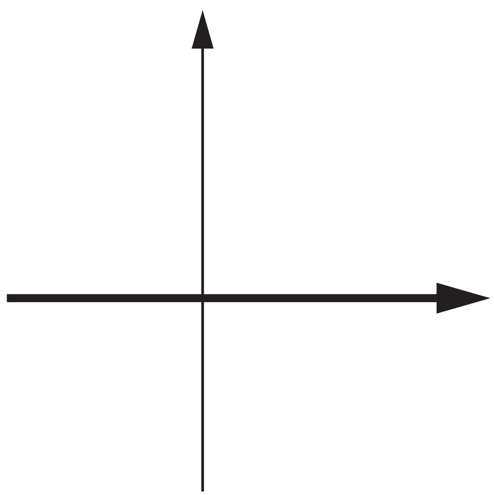
This could be called "adaptive subspace fitting"

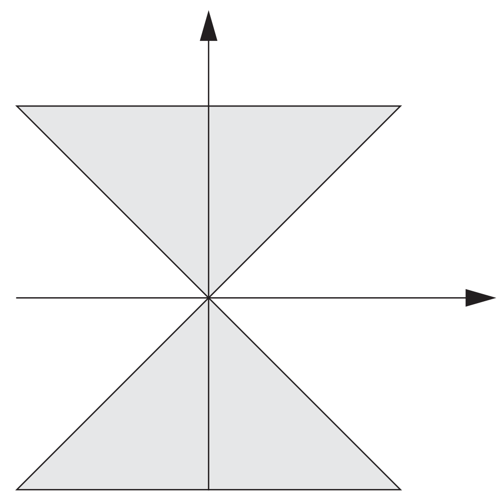
From a compression point of view, you "pay" for the adaptivity

- in general, this will cost

$$\log\left(\begin{bmatrix} N \\ k \end{bmatrix}\right) \text{ bits}$$

which cannot be spent on coefficient representation anymore



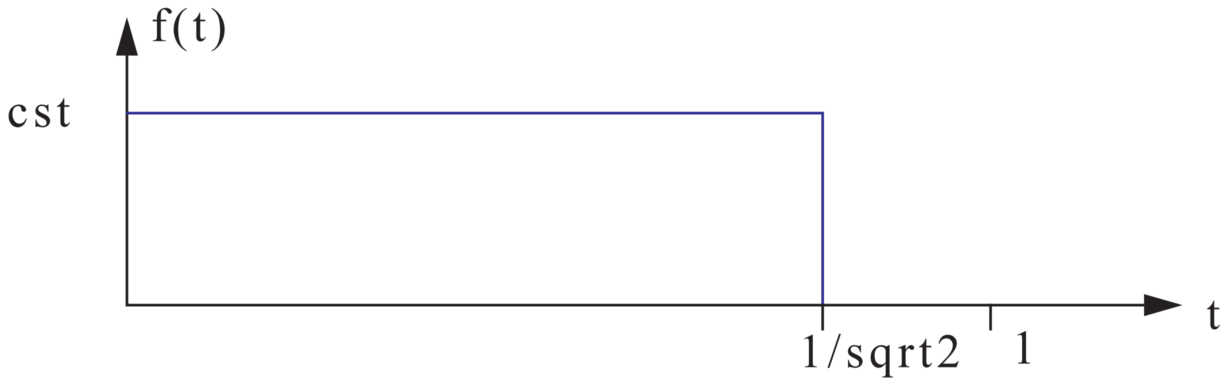


LA: pick a subspace a priori NLA pick after seeing the data

Non-Linear Approximation Example

Nonlinear approximation power depends on basis

Example:



Two different bases for $[0,1]$:

- Fourier series $\{e^{j2\pi kt}\}_{k \in \mathfrak{I}}$
- Wavelet series: Haar wavelets

Linear approximation in Fourier or wavelet bases

$$\varepsilon_M \sim 1/M$$

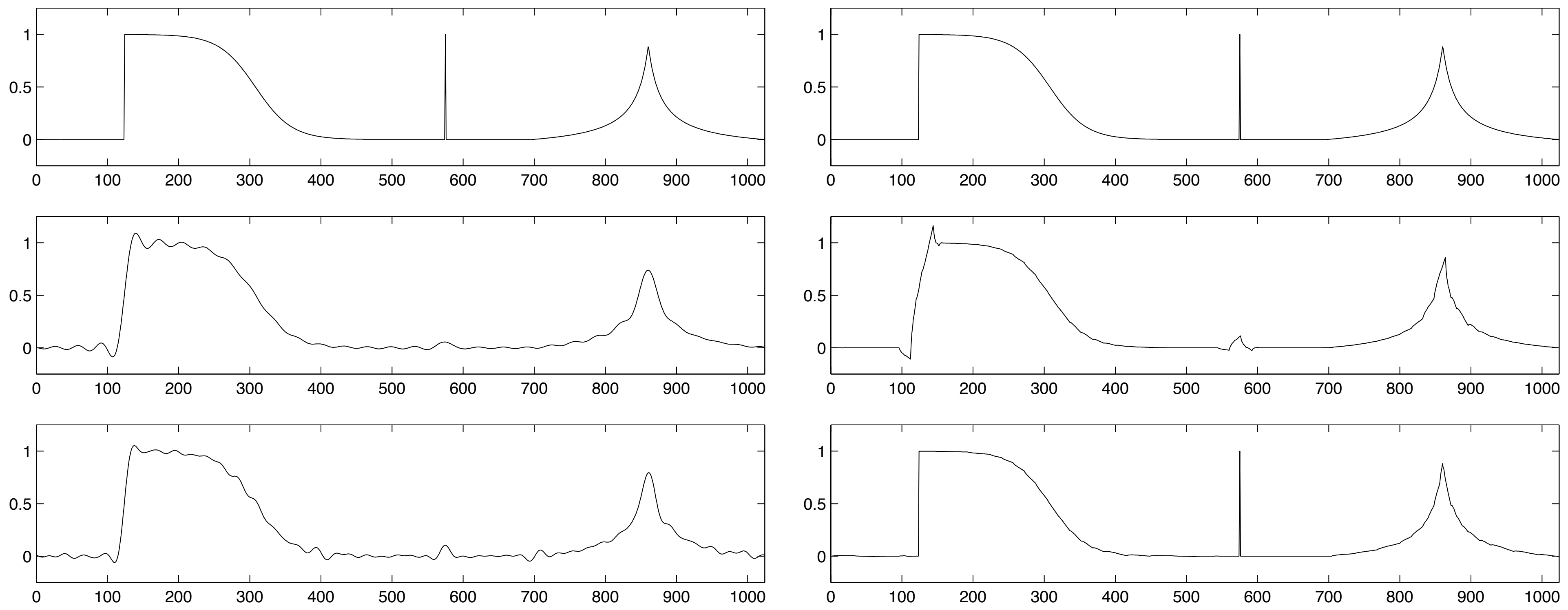
Nonlinear approximation in a Fourier basis

$$\tilde{\varepsilon}_M \sim 1/M$$

Nonlinear approximation in a wavelet basis

$$\tilde{\varepsilon}_M \sim 1/2^M$$

Fourier versus Wavelet bases, LA versus NLA

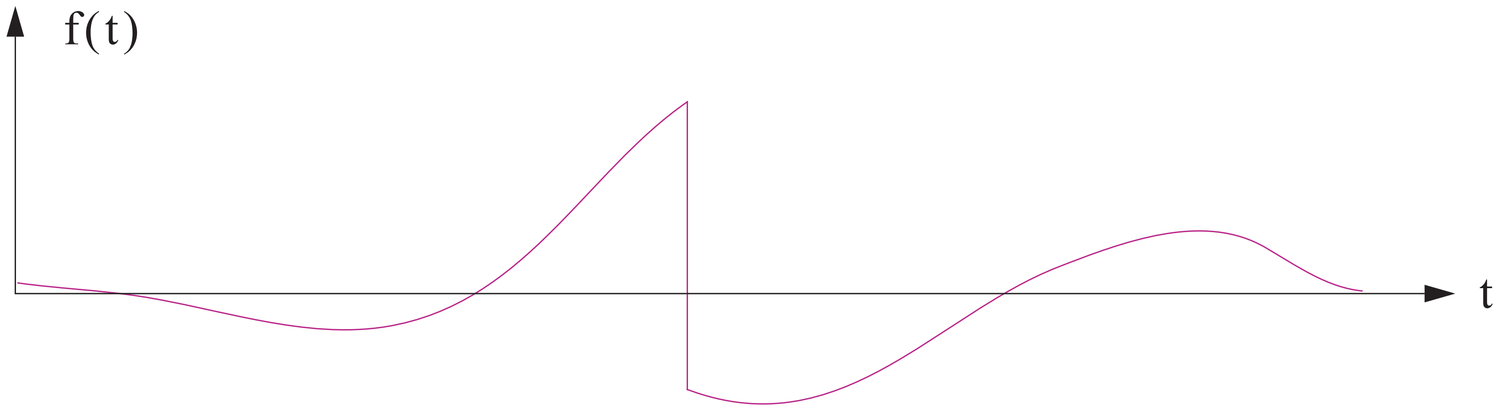


N= 1024, M=64

Fourier (left): LA versus NLA does not matter

Wavelets (right): NLA does orders of magnitude better!

Nonlinear approximation theory and wavelets



Approximation results for piecewise smooth fcts

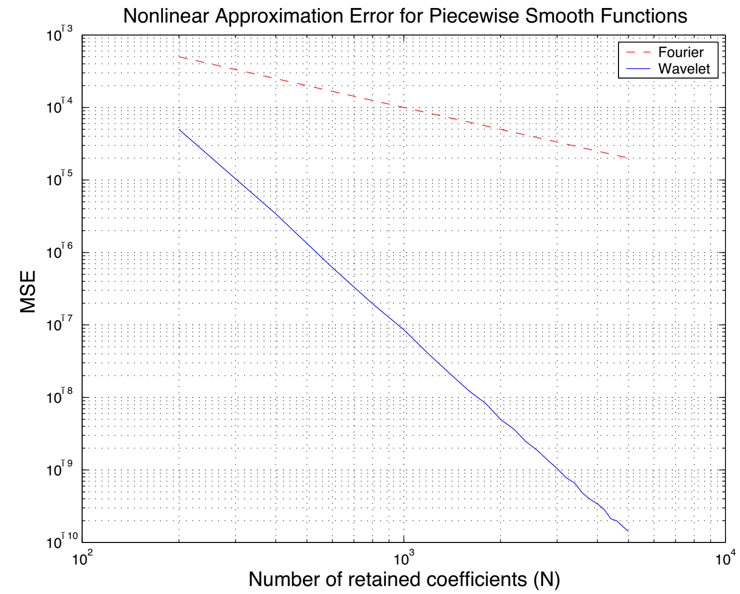
- between discontinuities, behavior by Sobolev or Besov regularity
- k derivatives $\Rightarrow$ coeffs $\sim 2^{m(k-1/2)}$ when $m \ll 0$
- Besov spaces can be defined with wavelet bases. If

$$\|f\|_{G,p} = \left(\sum |\langle f, g_n \rangle|^p\right)^{1/p} < \infty \qquad 0 < p < 2$$

then [DeVoreJL92]:

$$\varepsilon_M = o(M^{1-2/p})$$

Approximation in Sobolev and Besov Spaces



Linear Approximation, $W^s[0,N]$

- Sobolev-s: uniformly smooth
- Fourier: $\varepsilon_M = M^{-2s-\delta} \qquad \delta > 0$
- Wavelets: $\varepsilon_M = M^{-2s-\delta} \qquad \delta > 0$

Non-Linear Approximation

- Besov-s: smooth between a finite # of discontinuities
- Fourier: does not work, $\varepsilon_M = M^{-1}$
- Wavelets: approximation power given by the smoothness!
- Key: effect of discontinuities limited, because wavelets are concentrated around discontinuties
- f(t) in $W^s(0,N)$ between finite # of discontinuities, then f(t) in $B_p(0,N)$ (wavelet of compact support)
- Then:

$$\varepsilon_M = M^{\left(1-\frac{2}{p}\right)} \qquad \frac{1}{p} < s$$

- result can be refined to get $\varepsilon_M = M^{-2s-\delta} \qquad \delta > 0$

Outline

1. **Introduction through History**
2. **Fourier and Wavelet Representations**
3. **Wavelets and Approximation Theory**
4. **Wavelets and Compression**
   - A small but instructive example
   - piecewise polynomials and D(R)
   - piecewise smooth and D(R)
   - improved wavelet schemes
5. **Going to Two Dimensions: Non-Separable Constructions**
6. **Beyond Shift Invariant Subspaces: Finite Rate of Innovation**
7. **Conclusions and Outlook**

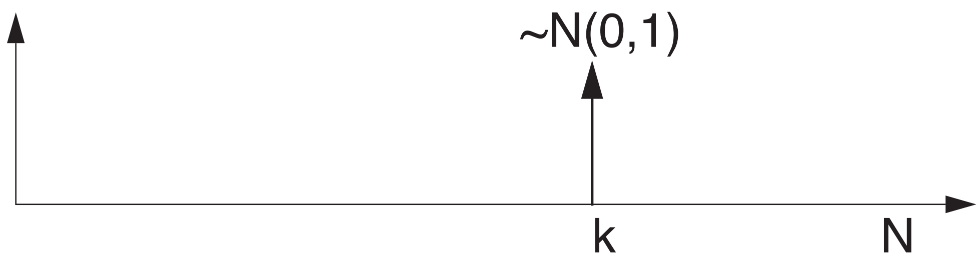
4. Wavelets and Compression

Compression is just one bit trickier than approximation...

A small but instructive example:

Assume

- $x[n] = \alpha\, \delta[n-k]$, signal is of length N, k is U[0,N-1] and α is N(0,1).
- This is a Gaussian RV at location k



- Note: $R_x = I$!

Linear approximation:

$$\varepsilon_M = \frac{1}{M}$$

Non-linear approximation, M > 0:

$$\tilde{\varepsilon}_M = 0$$

Given budget R for block of size N:

1. Linear approximation and KLT: equal distribution of R/N bits

$$D(R) = c \cdot \sigma^2 \cdot 2^{-2(R/N)}$$

This is the optimal linear approximation and compression!

2. Rate-distortion analysis [Weidmann:99]

High rate case:

- Obvious scheme: pointer + quantizer

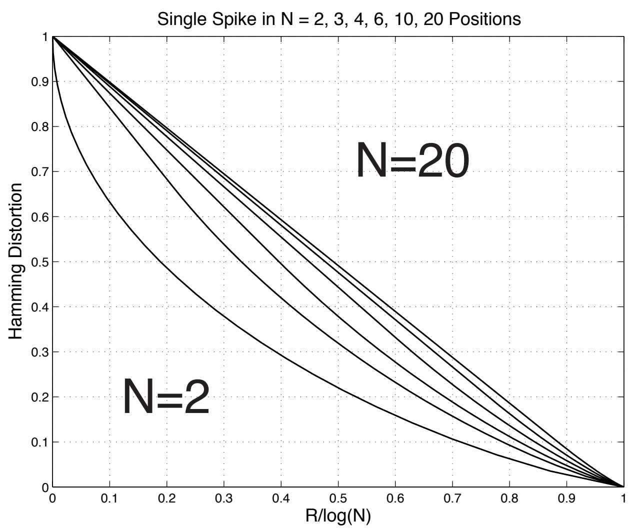
$$D(R) = c \cdot \sigma^2 \cdot 2^{-2(R - \log N)}$$

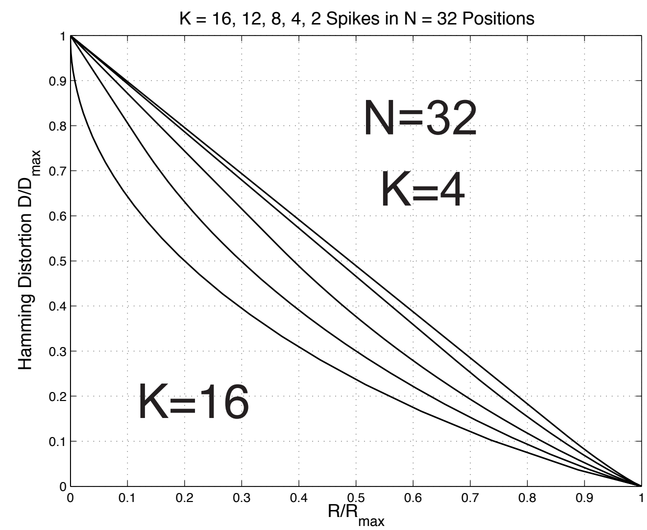
- This is the R(D) behavior for R >> Log N
- Much better than linear approximation

Low rate case:

- Hamming case solved, inc. multiple spikes:
  - there is a linear decay at low rates
- L_2 case: upper bounds that beat linear approx.

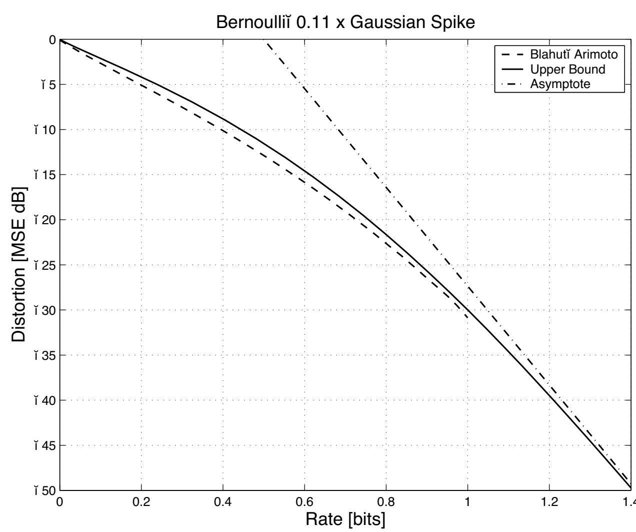
Example 1: Binary, Hamming, 1 and k spikes



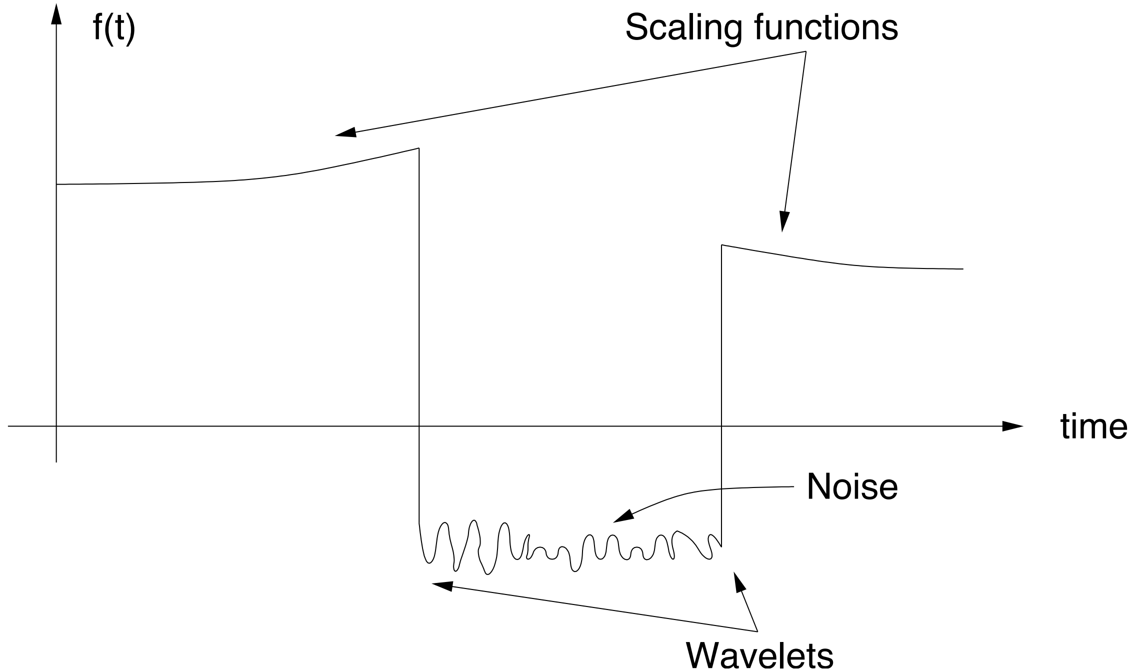


Example 2: Bernoulli-Gaussian

p=0.11



Piecewise smooth functions: pieces are Lipschitz-α



The following D(R) behavior is reachable [CohenDGO:02]:

$$D(R) = c_1 \cdot R^{-2\alpha} + c_3 \cdot \sqrt{R} \cdot 2^{-c_4 \cdot \sqrt{R}}$$

There are 2 modes:

- $R^{-2\alpha}$ corresponding to the Lipschitz-α pieces
- $\sqrt{R} \cdot 2^{-c \cdot \sqrt{R}}$ corresponding to the discontinuities

Lipschitz-α pieces: Linear Approximation

The wavelet transform at scale j decays as (j << 0)

$$w_j \approx 2^{j(\alpha + 1/2)}$$

Keep coefficients up to scale J, or choose a stepsize Δ for a quantizer

$$\Delta \approx 2^{J(\alpha + 1/2)}$$

Therefore, M ~ 2^J coefficients

Squared error:

$$\sum_{j=-\infty}^{-J} 2^{-j} \cdot 2^{2j(\alpha + 1/2)} \sim 2^{-2J\alpha} \sim M^{-2\alpha}$$

Rate:

- number of coefficients $c \cdot M$

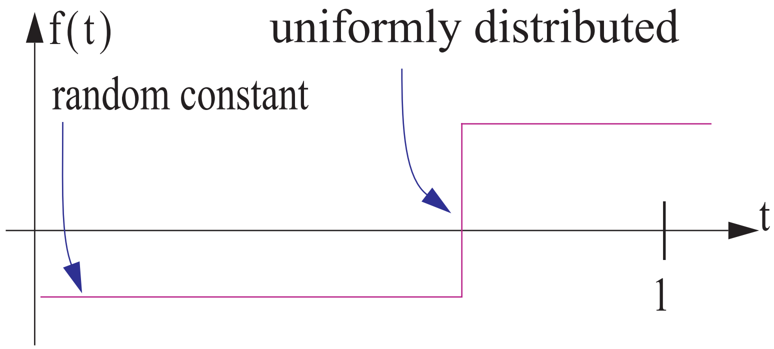
Thus

$$D(R) \sim c \cdot R^{-2\alpha}$$

Just as good as Fourier (~$R^{-2\alpha}$)**, but local!**

Rate-distortion bounds for piecewise polynomial functions

D(R) behavior of nonlinear approximation with wavelets

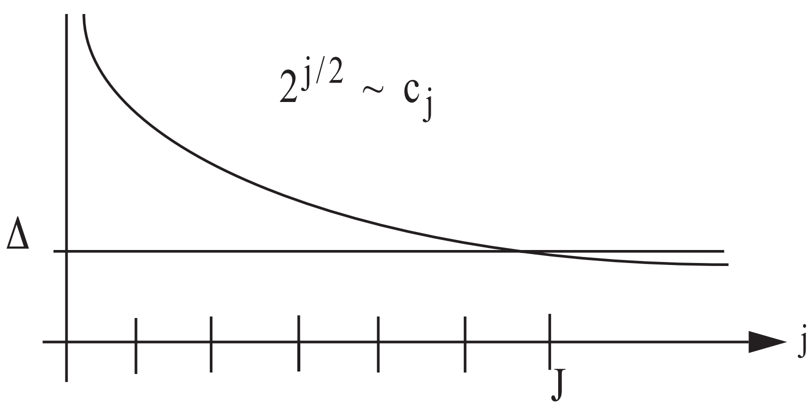


Consider the simplest case: Haar!

Recall that

$$\tilde{\varepsilon}_M \cong 2^{-M} \quad c_j \cong 2^{j/2}$$

and consider describing the significantcoefficients



Choose a stepsize Δ for a quantizer. Therefore

- number of scales J before coefs set to zero $\sim \log(1/\Delta)$
- number of bits per coefficient $\sim \log(1/\Delta)$, thus $R \sim J^2$

Distortion: number of scales times $\cdot \Delta^2 \sim J \cdot 2^{-J}$

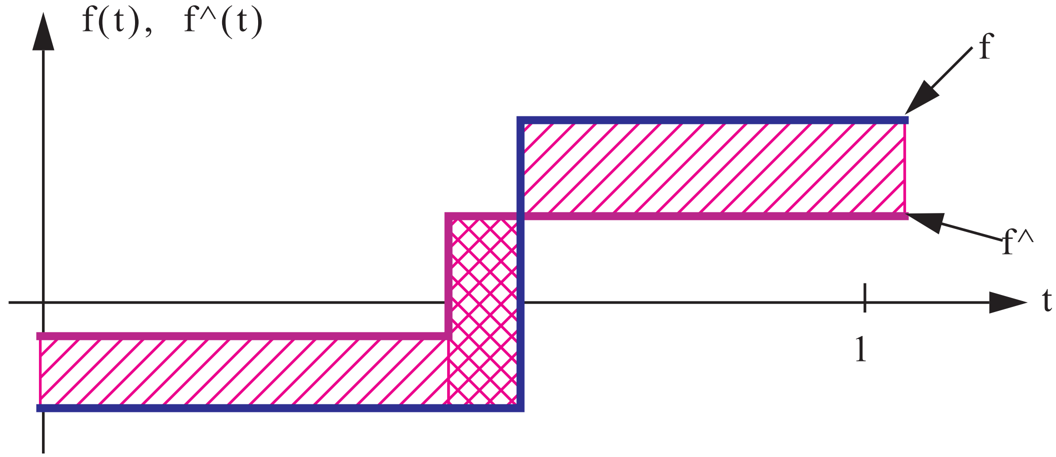
Thus

$$D_w(R) = C_3 \cdot \sqrt{R} \cdot 2^{-c_2 \cdot \sqrt{R}}$$

Rate-distortion behavior using an oracle

An oracle decides to optimally code a piecewise polynomial by allocating bits ''where needed'':

Consider the simplest case



Two approximation errors

- Δ_t: quantization of step location
- Δ_a: quantization of amplitude

Rate allocation: R_t versus R_a

Result:

$$D_p(R) = C_1 \cdot 2^{-R/2}$$

Piecewise polynomial, with max degree N

A. Nonlinear approximation with wavelets having $N+1$ zero moments

$$D_w(R) = C'_w \cdot (1 + \alpha\sqrt{C_w R}) \cdot 2^{-\sqrt{C_w R}}$$

B. Oracle-based method

$$D_p(R) = C'_p \cdot 2^{-(C_p \cdot R)}$$

Thus

- wavelets are a generic but suboptimal scheme
- oracle method asymptotically superior but dependent on the model

Conclusion on compression of piecewise smooth functions:

D(R) behavior has two modes:

$$D(R) = c_1 \cdot R^{-2\alpha} + c_3 \cdot \sqrt{R} \cdot 2^{-c_4 \cdot \sqrt{R}}$$

- 1/polynomial decay: cannot be (substantially) improved
- exponential mode: can be improved, important at low rates

Can we improve wavelet compression? Footprints!

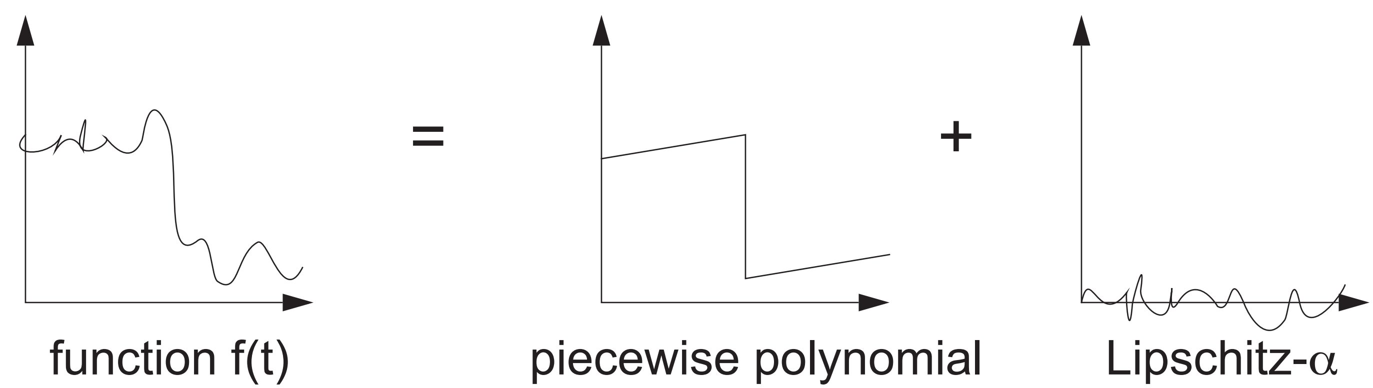
Key: Remove depencies accross scales:

- dynamic programming: Viterbi-like algorithm
- tree based algorithms: pruning and joining
- wavelet footprints: wavelet vector quantization

Theorem [DragottiV:03]:

Consider a piecewise smooth signal f(t), where pieces are Lipschitz-α. There exists a piecewise polynomial p(t) with pieces of maximum degree $\lfloor\alpha\rfloor$ such that the residual $r_\alpha(t) = f(t) - p(t)$ is uniformly Lipschitz-α.

This is a generic split into piecewise polynomial and smooth residual



Footprint Basis and Frames

Suboptimality of wavelets for piecewise polynomials is due to independent coding of dependent wavelet coefficients

$$D_w(R) \sim C \cdot \sqrt{R} \cdot 2^{-\sqrt{R}}$$

Compression with wavelet footprints

Theorem: [DragottiV:03]

Given a bounded piecewise polynomial of deg D with K discontinuities. Then,a footprint based coder achieves

$$D(R) = c_1 \cdot 2^{-(c_2 \cdot R)}$$

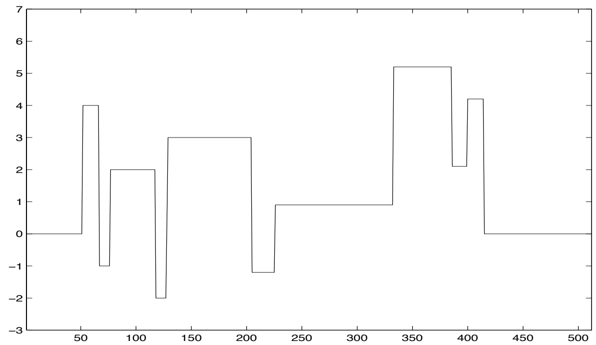
This is a computational effective method to get oracle performance

What is more, the generic split “piecewise smooth” into “uniformly smooth + piecewise polynomial” allows to fix wavelet scenarios, to obtain

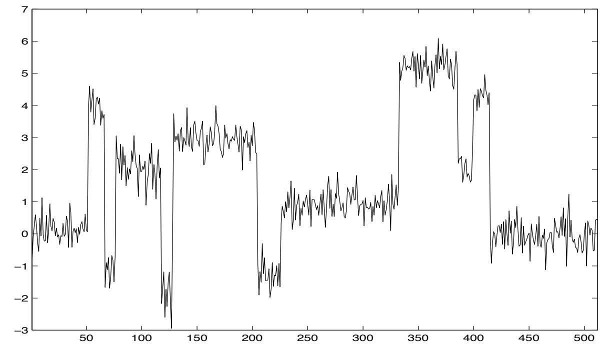
$$D(R) = c_1 \cdot R^{-2\alpha} + c_2 \cdot 2^{-c_3 \cdot R}$$

This can be used for denoising and superresolution

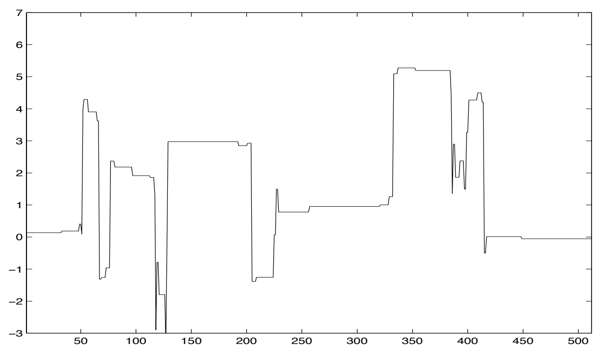
Denoising (use coherence across scale)



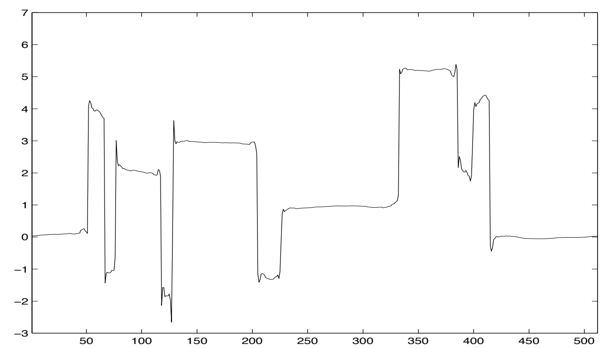
Original signal



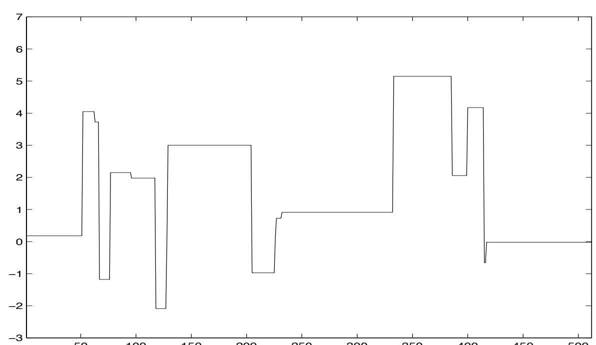
Noisy Signal (SNR=15.62dB)



Hard-Thresholding (SNR=21.3dB)



Cycle-Spinning (SNR=25.4dB)



Denoising with Footprints (SNR=27.2dB)

This is a vector thresholding method adapted to wavelet singularities

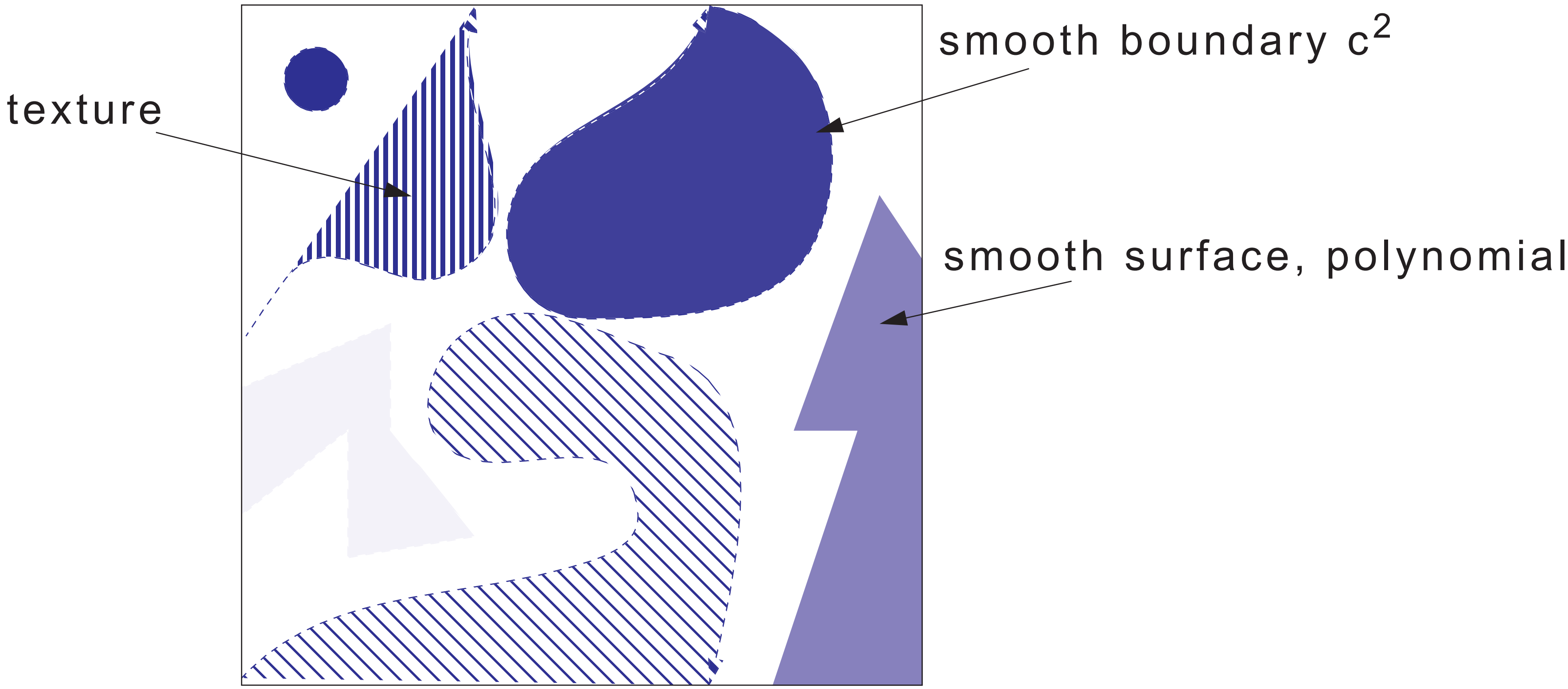
Outline

1. **Introduction through History**
2. **Fourier and Wavelet Representations**
3. **Wavelets and Approximation Theory**
4. **Wavelets and Compression**
5. **Going to Two Dimensions: Non-Separable Constructions**
    - the need for truly two-dimensional constructions
    - tree based methods
    - non-separable bases and frames
6. **Beyond Shift Invariant Subspaces: Finite Rate of Innovation**
7. **Conclusions and Outlook**

5. Going to Two Dimensions: Non-Separable Constructions

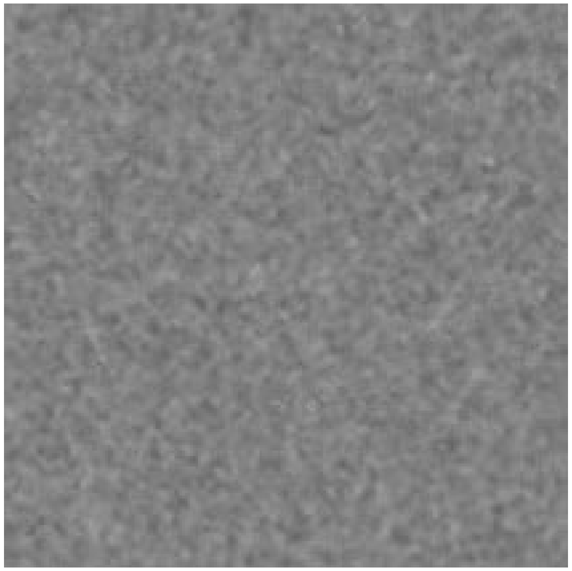
Going to two dimensions requires non-separable bases

Objects in two dimensions we are interested in



- textures: $D(R) = C_0 \cdot 2^{-2R}$ per pixel
- smooth surfaces: $D(R) = C_1 \cdot 2^{-2R}$ per object!

Models of the world:



Gauss-Markov

Piecewise polynomial

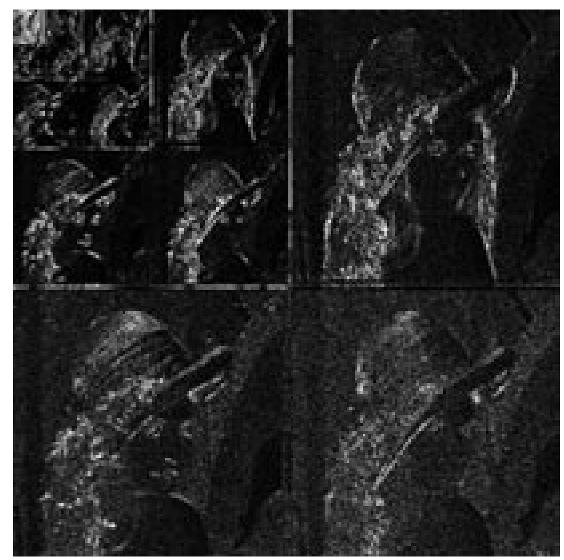
the usual suspect

Many proposed models:

- mathematical difficulties
- one size fits all...
- Lena is not PC, but is she BV?

But: Fourier, DCT, wavelets use a separable approach (line/column...)

=> geometry based image processing



Recent work on geometric image processing

Long history: compression, vision, filter banks

Current affairs:

Signal adapted schemes

- Bandelets [LePennec & Mallat]: wavelet expansions centered at at discontinuity as well as along smooth edges
- Non-linear tilings [Cohen, Mattei]: adaptive segmentation
- Tree structured approaches [Shukla et al, Baraniuk et al]

Bases and Frames

- Wedgelets [Donoho]: Basic element is a wedge
- Ridgelets [Candes, Donoho]: Basic element is a ridge
- Curvelets [Candes, Donoho]
  Scaling law: width ~length^2
  $L(R^2)$ set up
- Multidirectional pyramids and contourlets [Do et al]
  Discrete-space set-up, $l(Z^2)$
  Tight frame with small redundancy
  Computational framework

This is where the action is!

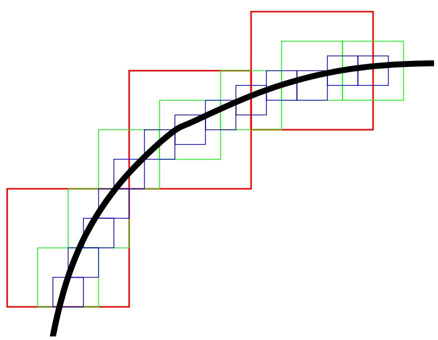
Nonseparable schemes and approximation

Approximation properties:

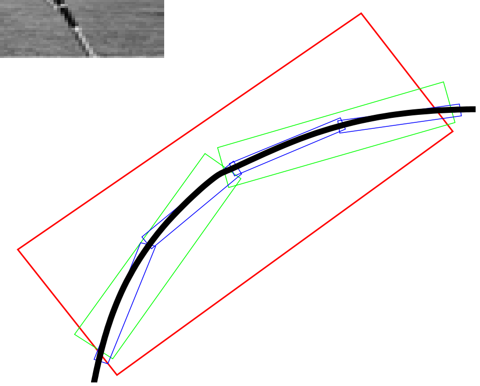
- wavelets good for point singularities
- ridgelets good for ridges
- curvelets good for curves

Consider c^2 boundary between two csts

 # wavelet coeffs $O(2^j)$



 # curvelet coeffs $O(2^{j/2})$

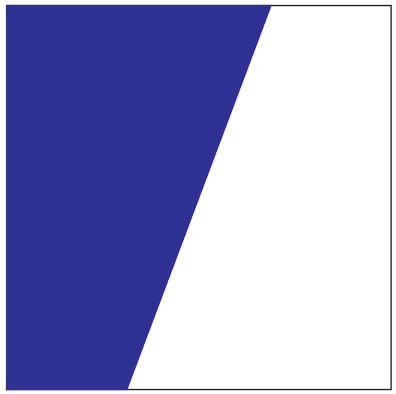


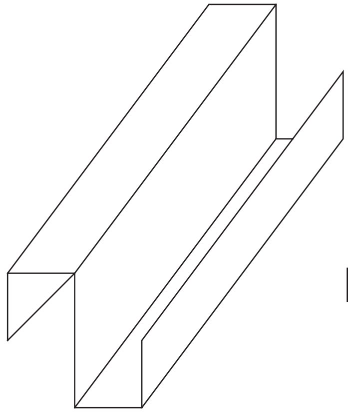
Rate of approximation, M-term NLA in bases, c^2 boundary

- Fourier: $O(M^{-1/2})$
- Wavelets: $O(M^{-1})$
- Curvelets: $O(M^{-2})$

Compression of non-separable objects

Objects we know how to compress....





Approximation

- Wavelets $E_M \sim M^{-1}$
- Ridgelets $E_M \sim 2^{-M}$

Rate/distortion

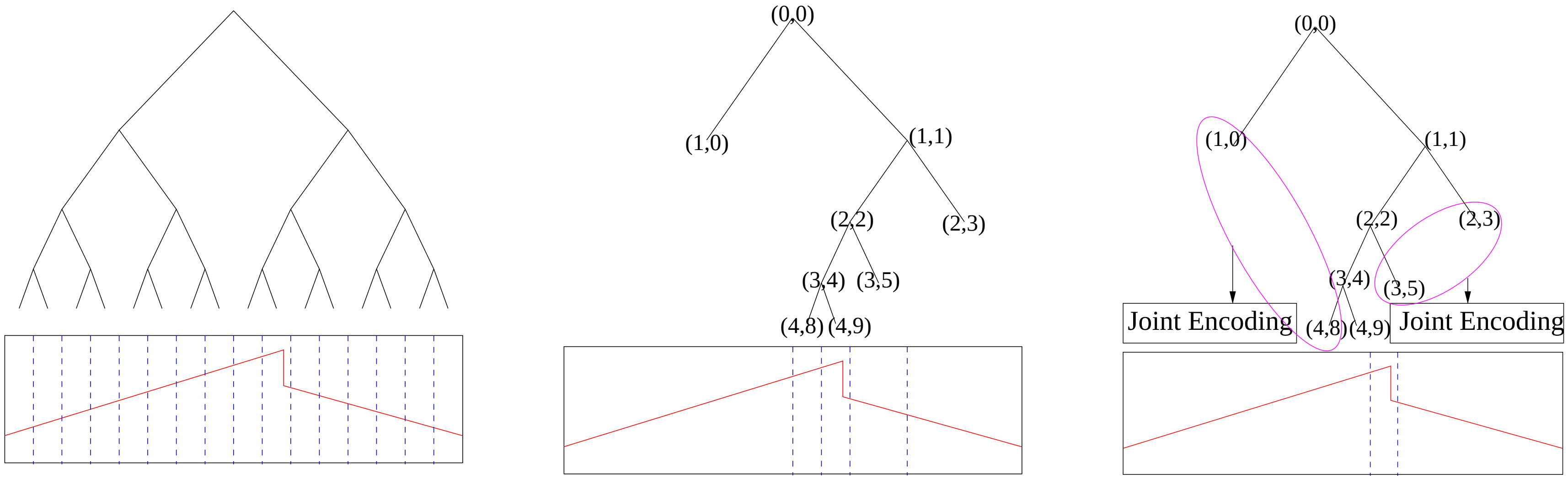
- Oracle $D(R) = C \cdot 2^{-2R}$
- Wavelets....poor
- Ridgelets....suboptimal
- adaptive schemes: close to oracle
- fixed basis: under investigation

Tree Based Geometric Compression [ShuklaDDV:03]

Idea

- tree and quadtree algorithms popular, many pruning algorithms optimality proofs for wedgelets [Donoho:99]
- new pruning and joining algorithm

Intuition: full tree **dyadic tree** **pruned & joined tree**



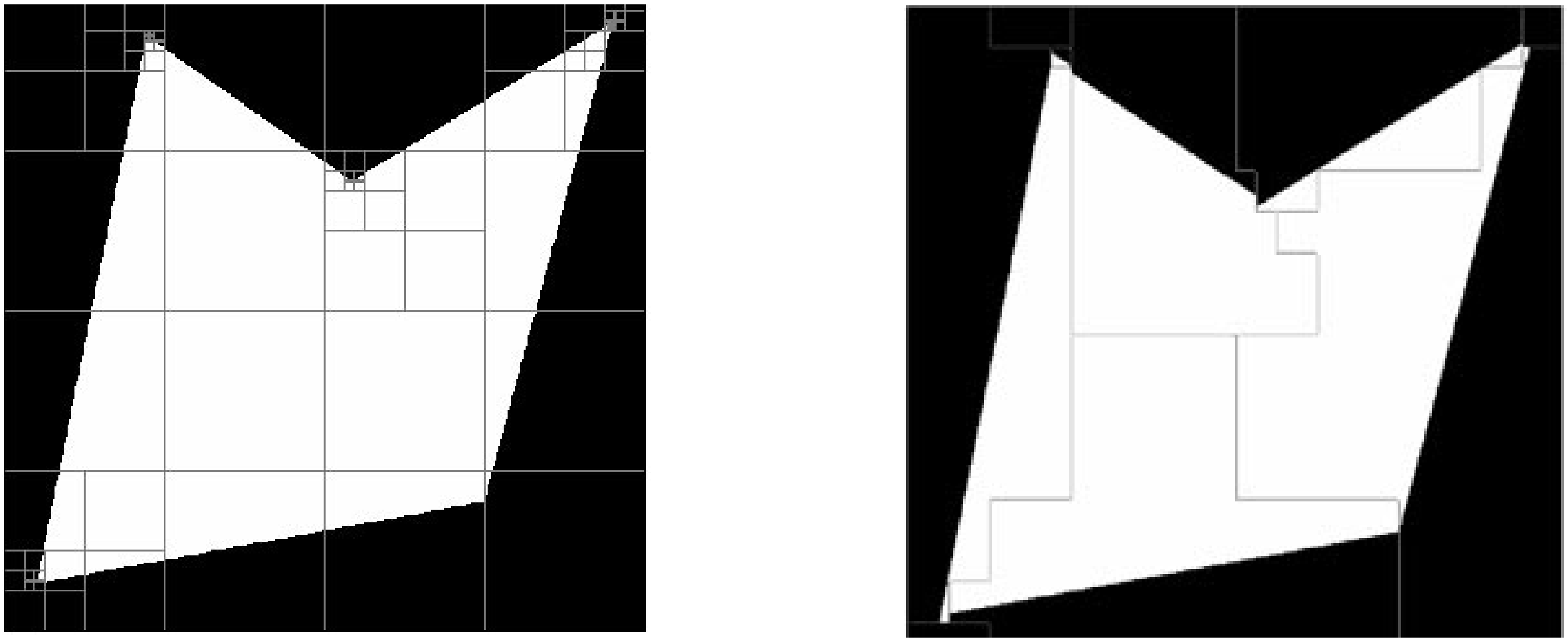
$N_J \sim 2^J$ $D(R) \sim R^{-1}$ $N_J \sim J$ $D(R) \sim \sqrt{R} \cdot 2^{-c_1 \cdot \sqrt{R}}$ $N_J \sim J^0$ $D(R) \sim 2^{-c_2 R}$

Results: Rate-distortion optimal for piecewise polynomials

$D(R) = c_1 \cdot 2^{-(c_2 \cdot R)}$ **that is, like an oracle method (up to constants)**

Extension to Quadtree:

- Example



Results:

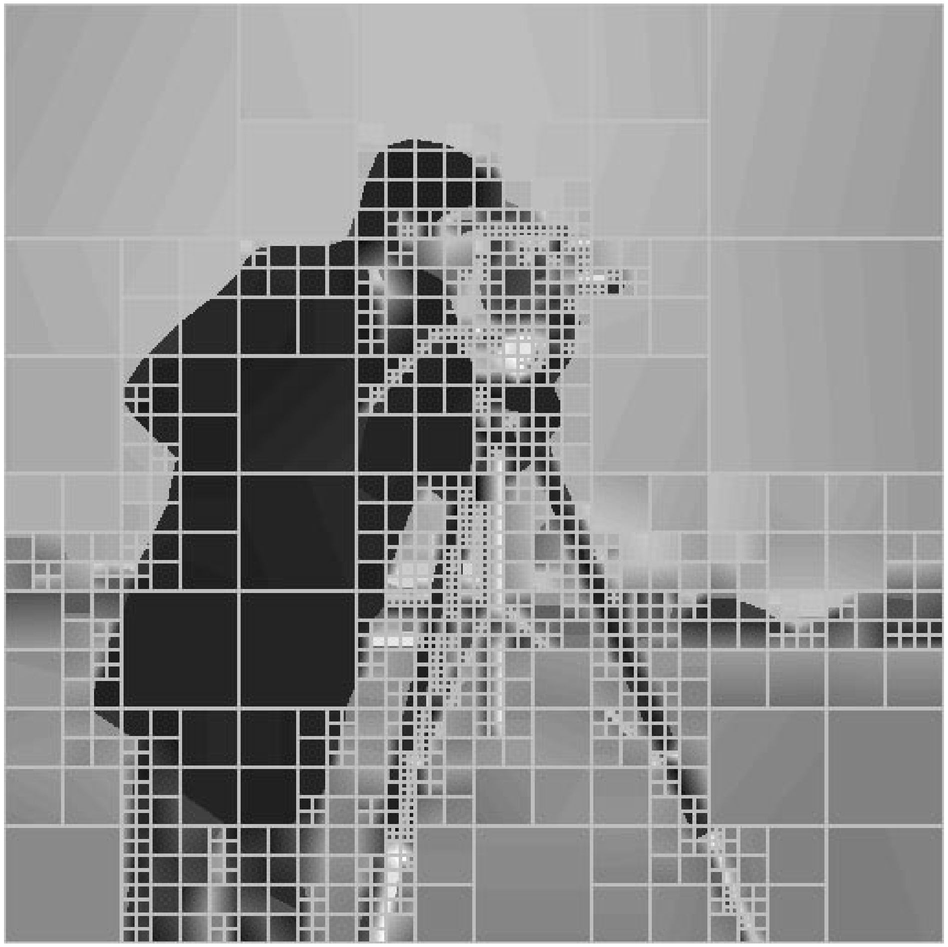
- consider a piecewise polynomial 2D signal, with polynomial boundaries, the following rate-distortion behavior is achieved

$$D(R) = c_3 \cdot 2^{-(c_4 \cdot R)}$$

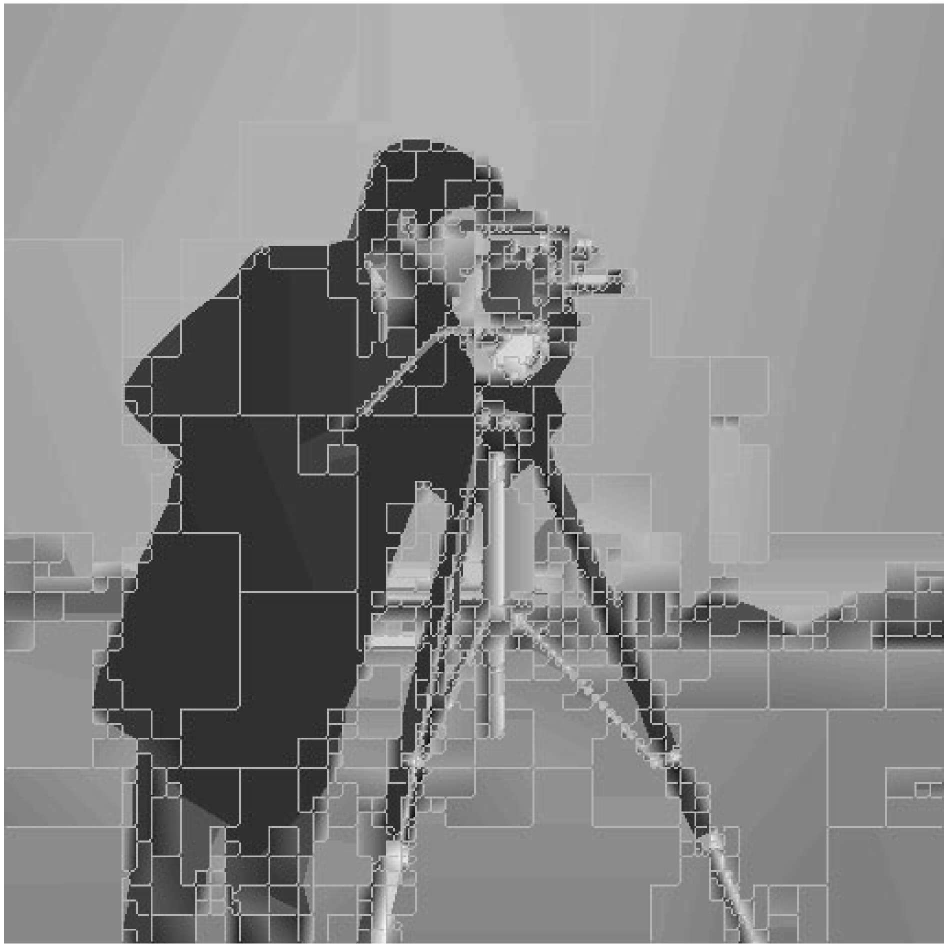
- this is like an oracle method, and >> than prune algorithms which have a $\sqrt{R}$ penalty
- complexity: polynomial

The prune-join quadtree algorithm

- polynomial fit to surface and to boundary on a quadtree
- rate-distortion optimal tree pruning and joining



quadtree with R(D) pruning



R(D) Joining of ''similar'' leaves

Note: careful R(D) optimization!

Geometric Compression versus JPEG2000 at 0.11 bits/pixel, PSNR:

28.95

27.75

30.01

29.22

pruned-joined quadtree

JPEG2000

Behavior of tree algorithms on piecewise smooth fcts

ppf: piecewise polynomial functions

psf: piecewise smooth functions, a-smooth

Signal Class	Oracle Coder	Wavelet Coder	Prune tree Coder	Prune-join tree Coder
1-D PPF	2^{-cR}	$2^{-c_1\sqrt{R}}$	$2^{-c_2\sqrt{R}}$	$2^{-c_3 R}$
2-D PPF	2^{-dR}	$\frac{\log R}{R}$	$2^{-c_4\sqrt{R}}$	$2^{-c_5 R}$
1-D PSF	R^{-2a}	R^{-2a}	$\left(\frac{\log R}{R}\right)^{2a}$	$\left(\frac{\log R}{R}\right)^{2a}$
2-D PSF	R^{-a}	$\frac{\log R}{R}$	$\left(\frac{\log R}{R}\right)^{a}$	$\left(\frac{\log R}{R}\right)^{a}$

at most log penalty with polynomial complexity
(and a bit more work gets rid of logs...)

Interesting scaling laws, good behavior in practice!

Directional bases and contourlets [M.Do]

Goal: find a discrete-space construction that has good approximation properties for smooth functions with smooth boundaries

- directional analysis as in a Radon transform
- multiresolution as in wavelets and pyramids
- computationally easy
- bases or low redundancy frames

Background:

- curvelets [Candes-Donoho] indicate that ''good'' fixed bases do exist for approximation of piecewise smooth 2D functions
- a frequency-direction relationship indicates a scaling law $d \sim j^{1/2}$

Idea:

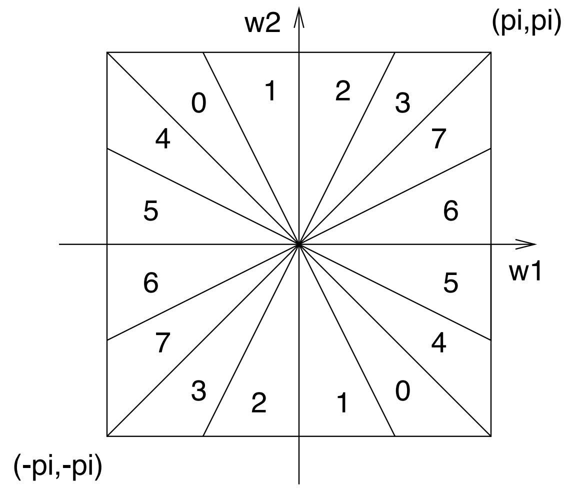
- directional analysis: directions are key
- multiresolution analysis

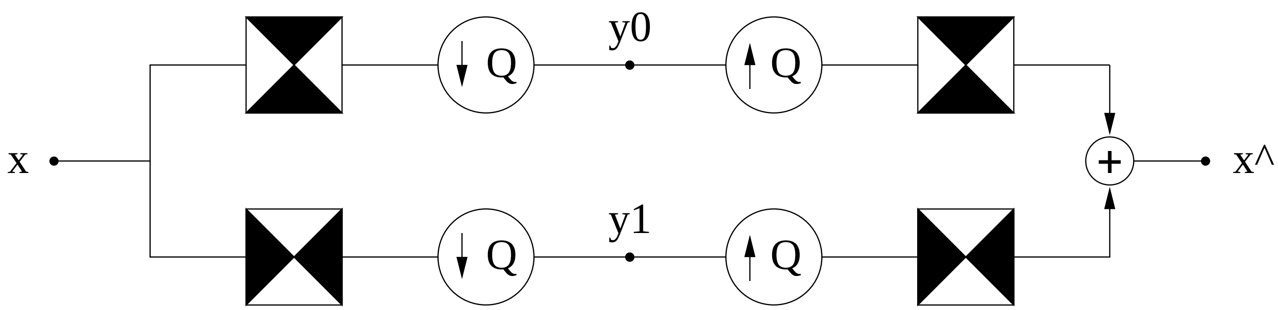
Result:

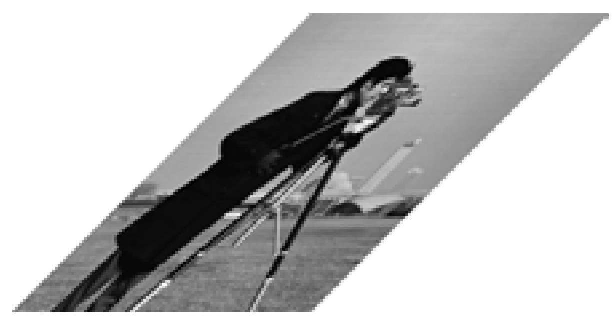
- one-more-let: contourlets!

Directional Filter Banks [BambergerS:92, DoV:02]

- divide 2-D spectrum into slices with iterated tree-structured f-banks

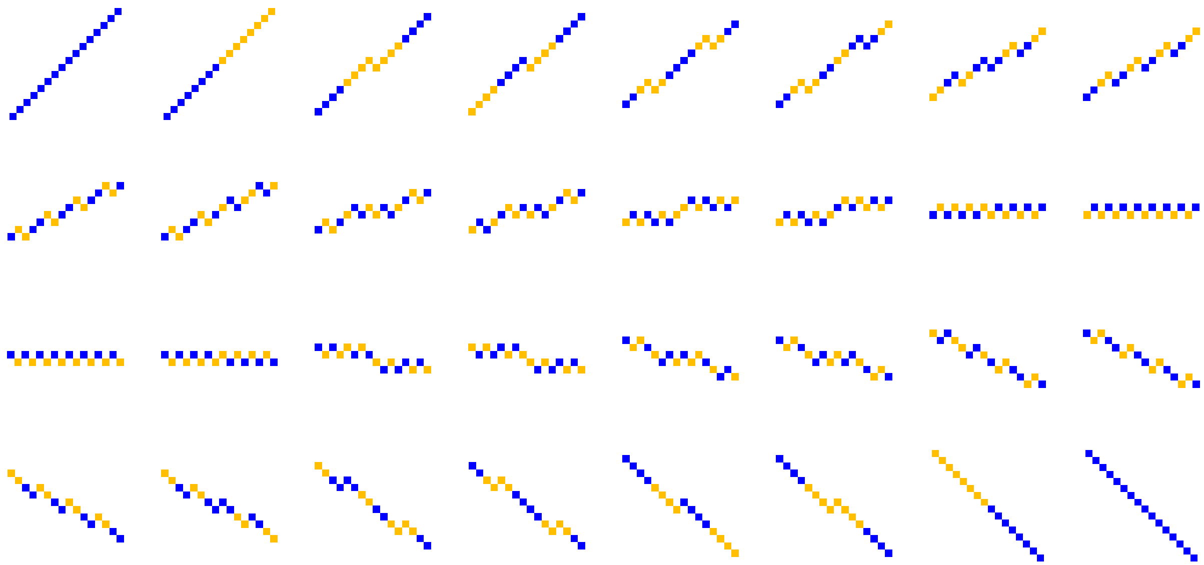






Example of directional basis functions

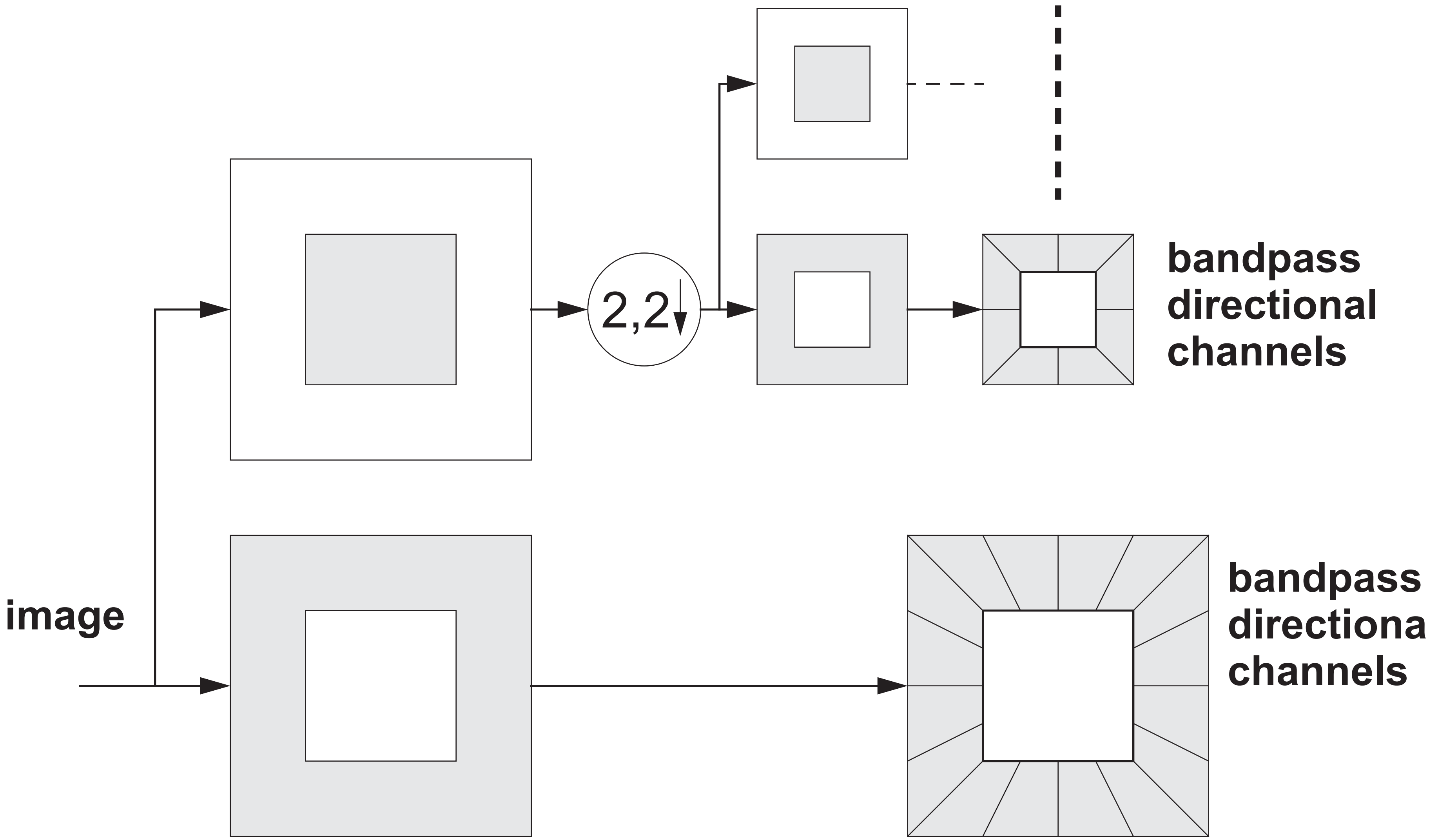
- 64 channels, elementary filters are Haar filters
- orthonormal directional basis
- 64 equivalent filters, the 32 ''mostly horizontal'' ones are shown



This ressembles a ''local Radon transform'', or radonlets!

- changes of sign (for orthonormality)
- approximate lines (discretizations)

Multiresolution directional pyramid



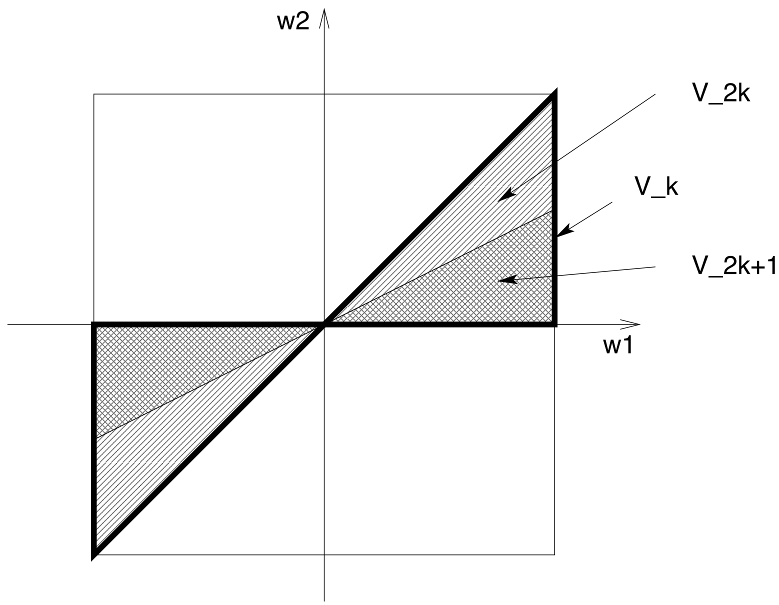
Result:

- “tight” pyramid and orthogonal directional channels => tight frame
- low redundancy < 4/3, computationally efficient

A directional multiresolution analysis

Theorem [Do:01]: For a finite number of directions, this generates a tight frame for $L_2(R^2)$ with frame bound equal 1

Method: Define embedded lowpass directional spaces $V_{j,k}$ and directional bandpass spaces $W_{j,k}$

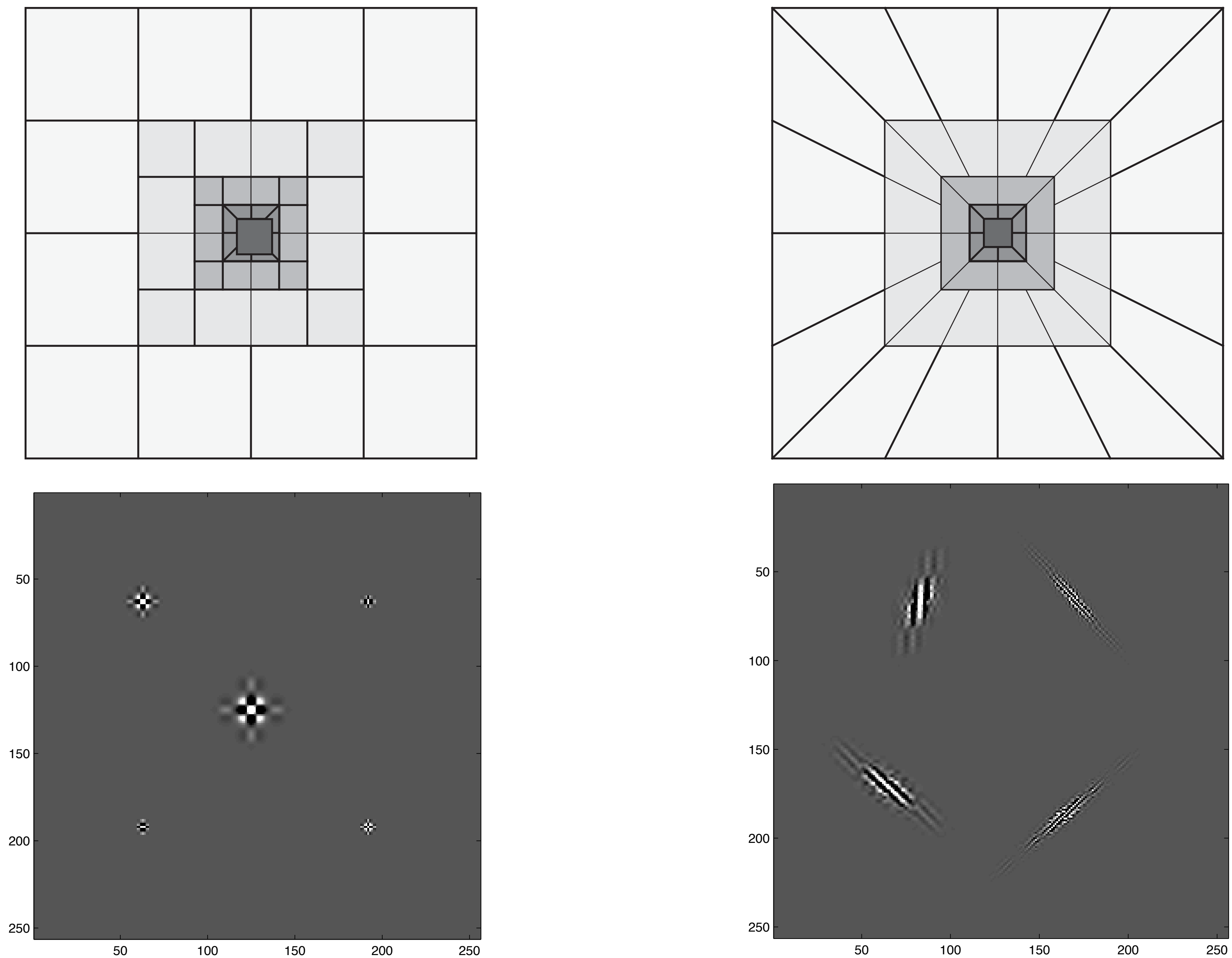


This defines contourlets: how do they compare to wavelets?

Approximation: M-term NLA satisfies $\left|f - f_{contourlet}\right|^2 \sim \frac{1}{M^2}$ [CandesD:00]

possible with sinc filters,
... open problem if compact support contourlets exist....

Basis functions: wavelets versus contourlets

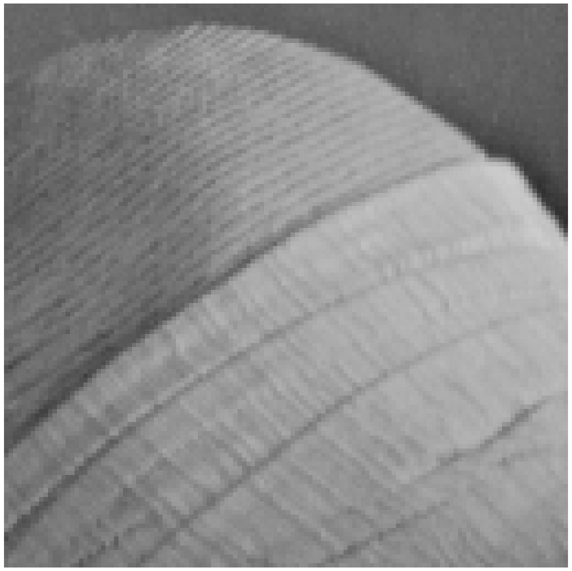


Expansion Example

Pepper image and its expansion

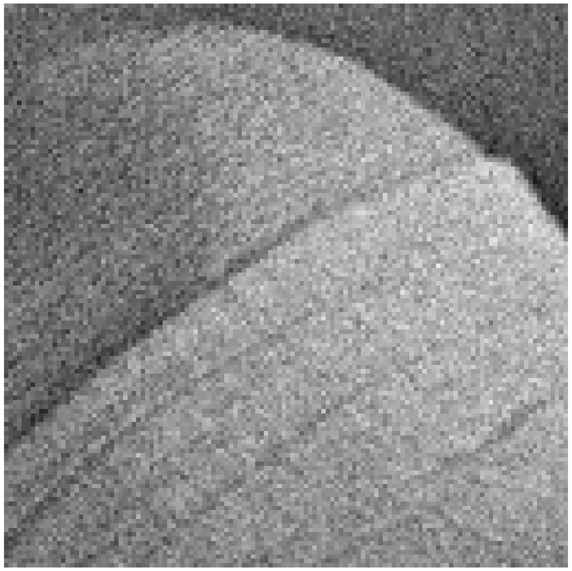
Compression, denoising, inverse problems:
if it is sparse, it is a good start!

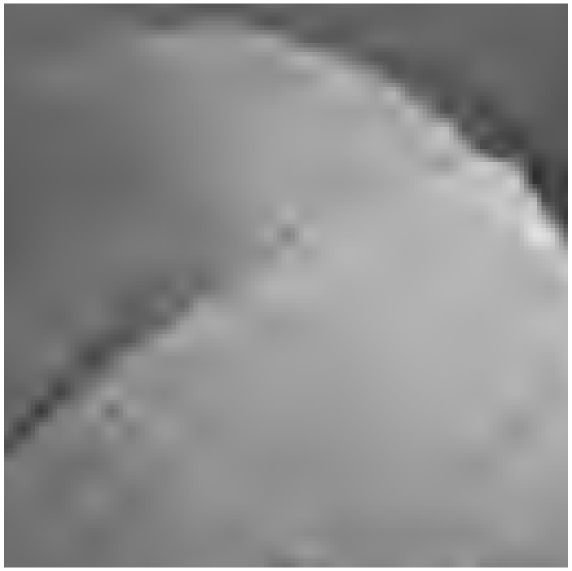
Example: denoising with contourlets



original

noisy





wavelet
13.8 dB

contourlets
15.4 dB

Outline

1. **Introduction through History**
2. **Fourier and Wavelet Representations**
3. **Wavelets and Approximation Theory**
4. **Wavelets and Compression**
5. **Going to Two Dimensions: Non-Separable Constructions**
6. **Beyond Shift Invariant Subspaces: Finite Rate of Innovation**
   - Shift-Invariance and Multiresolution Analysis
   - A Variation on a Theme by Shannon
   - A Representation Theorem
7. **Conclusions and Outlook**

Shift-Invariance and Multiresolution Analysis

Most sampling results require shift-invariant subspaces

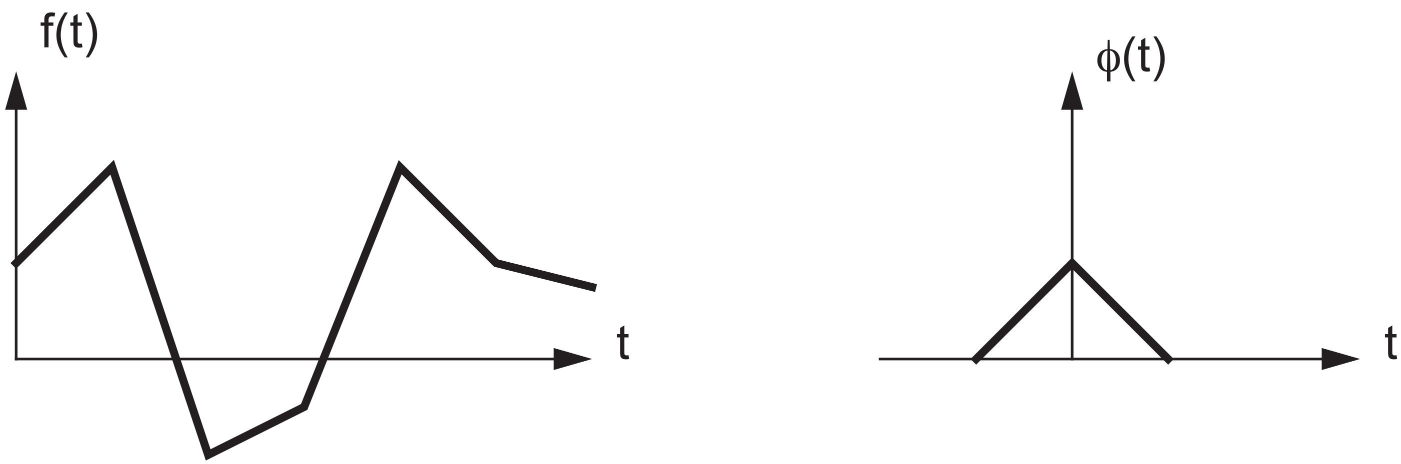
- $f(t) \in V \Leftrightarrow f(t - nT) \in V \qquad n \in Z$

Wavelet constructions rely in addition on scale-invariance

- $f(t) \in V_0 \Leftrightarrow f(2^m t) \in V_{-m} \qquad m \in Z$

Multiresolution analysis (Mallat, Meyer) gives a powerful framework. Yet it requires a subspace structure...

Example: uniform or B-splines



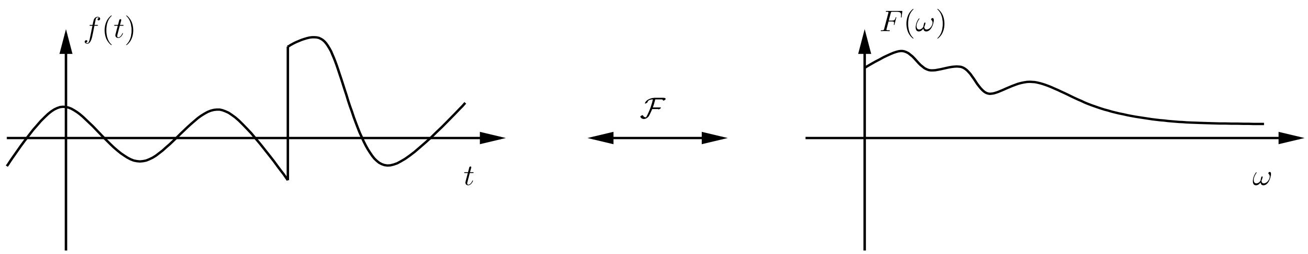
Question: can sampling be generalized beyond subspaces?

Note: Shannon BW sufficient, not necessary!

A Variation on a Theme by Shannon

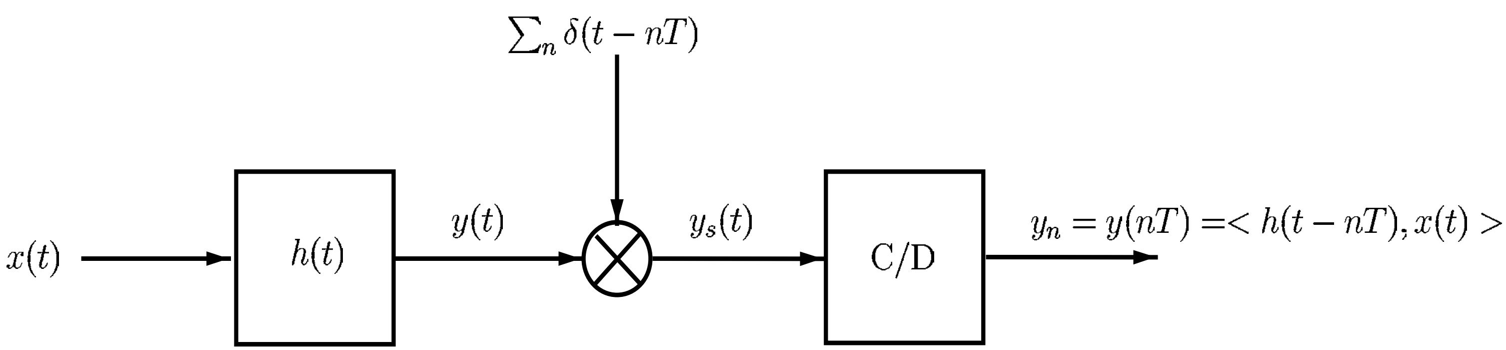
Shannon, BL case: $f(t) = \sum_{n \in Z} f(nT)\mathrm{sinc}(t/T - n)$ or 1/T degrees of freedom per unit of time

But: a single discontinuity, and no more sampling theorem...



Q: Are there other signals with finite number of degrees of freedom per unit of time that allow exact sampling results?
=> Finite rate of innovation

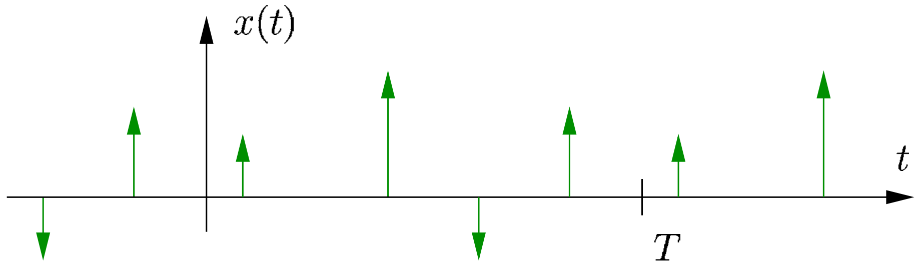
Usual setup:



$x(t)$**: signal,** $h(t)$**:sampling kernel,** $y(t)$**:filtering of** $x(t)$ **and** y_n **:samples**

A Toy Example

K Diracs on the interval: 2K degrees of freedom. Periodic case:



$$x(t) = \sum_{n \in Z} \sum_{k=0}^{K-1} c_k \delta(t - t_k - n\tau) = \sum_{k=0}^{K-1} c_k \frac{1}{\tau} \sum_{m \in Z} e^{\frac{j2\pi m(t - t_k)}{\tau}}$$

Key: The Fourier series is a weighted sum of K exponentials

$$X[m] = \frac{1}{\tau} \sum_{k=0}^{K-1} c_k e^{\frac{-j2\pi m t_k}{\tau}}$$

Result: Taking 2k+1 samples from a lowpass version of BW-(2K+1) allows to perfectly recover x(t)

Method: Yule-Walker system, annihilating filter, Vandermonde system

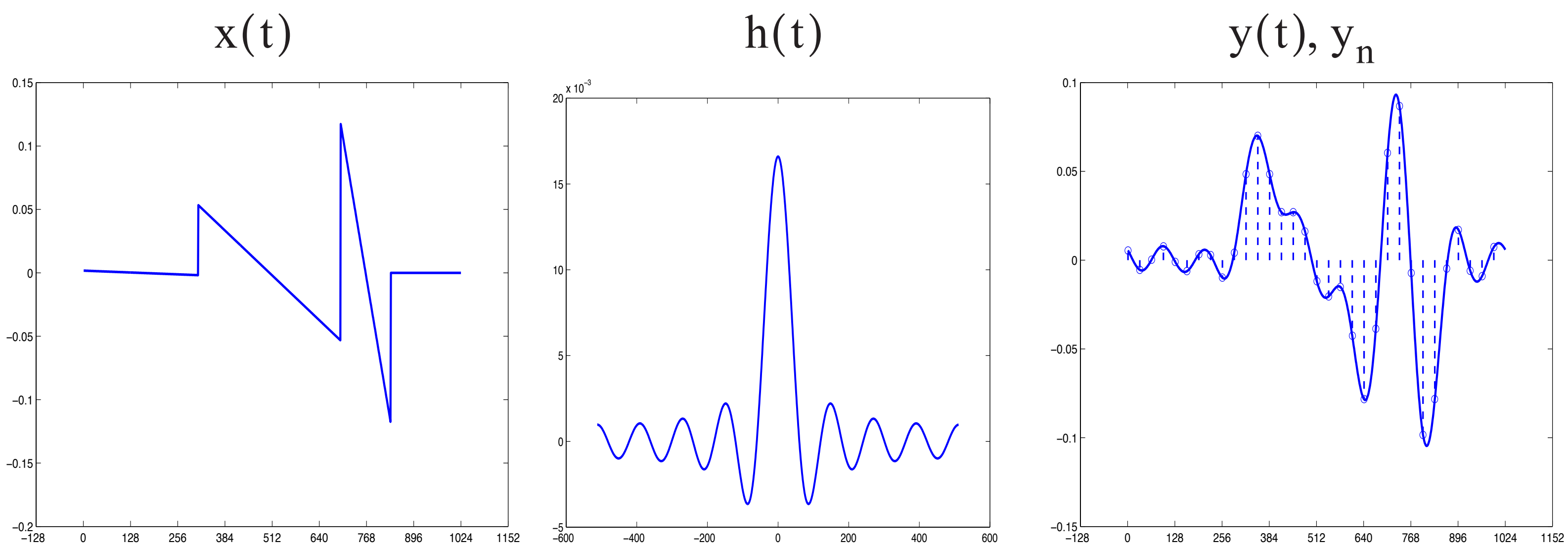
Note: Relation to spectral estimation and ECC (Berlekamp-Massey)

A Representation Theorem [VMB:02]

For the class of periodic FRI signals which includes

- sequences of Diracs
- non-uniform or free knot splines
- piecewise polynomials

there exist sampling schemes with a sampling rate of the order of the rate of innovation with perfect reconstruction at polynomial cost.

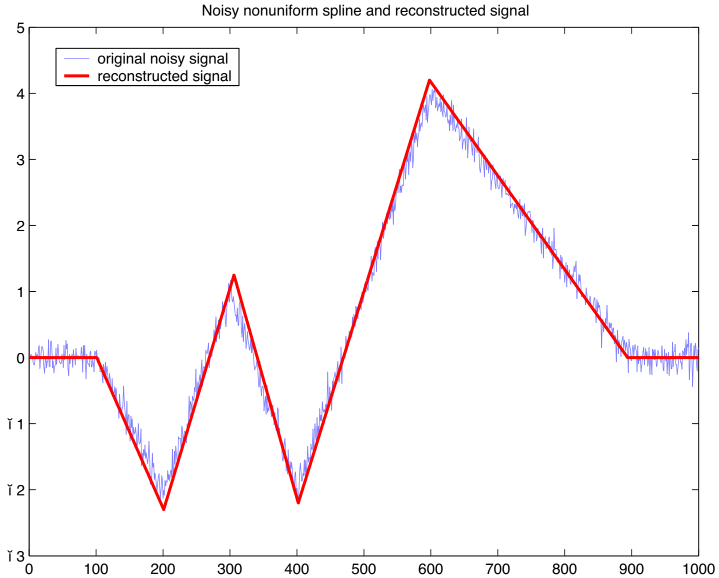


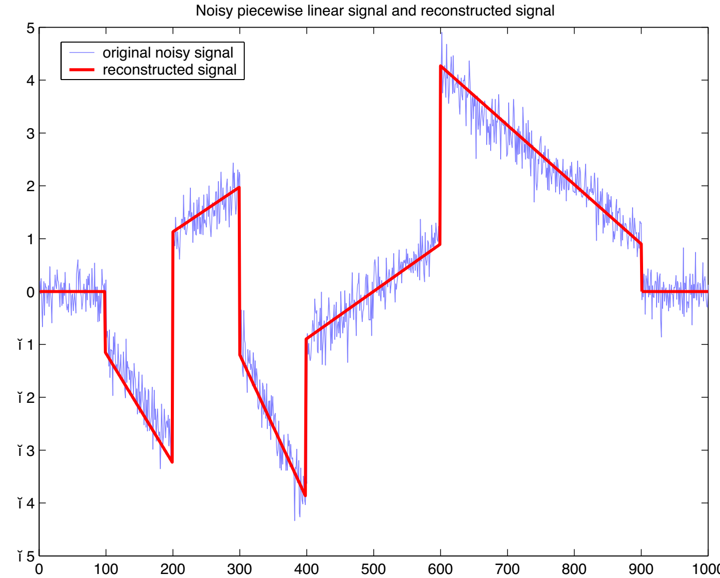
Variations:

- finite length signals, local kernels
- Two-dimensions

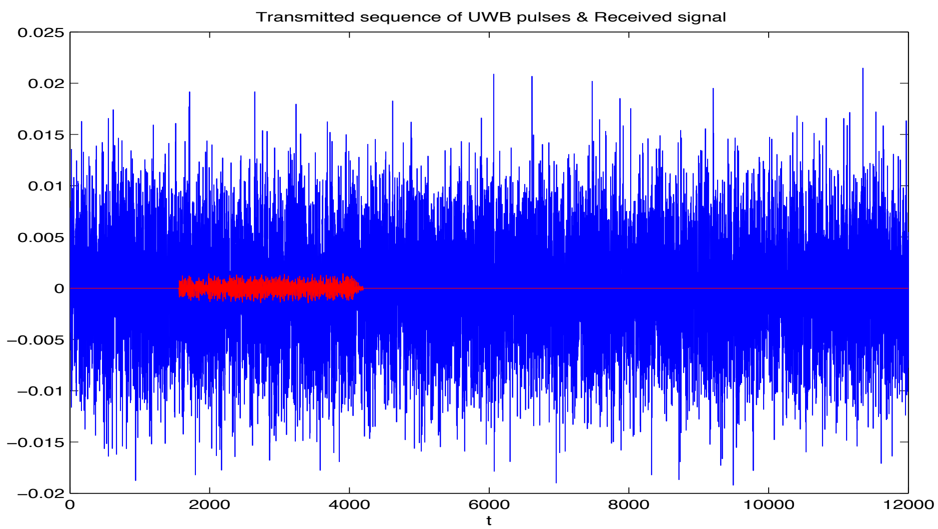
and the noisy case....

Use subspace methods (I.Maravic)



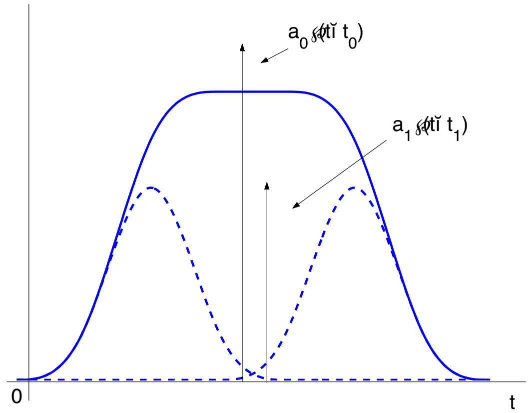


Application example: UWB (low rate of innovation...but lots of noise!)

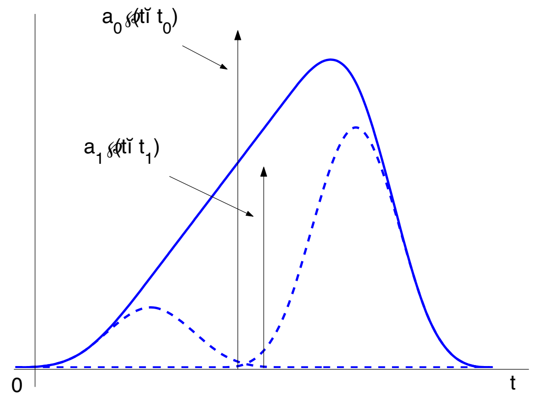


A local algorithm for FRI sampling [DVB:04]

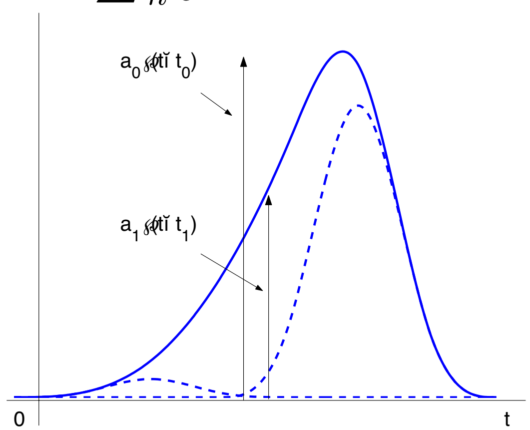
The return of Strang-Fix!



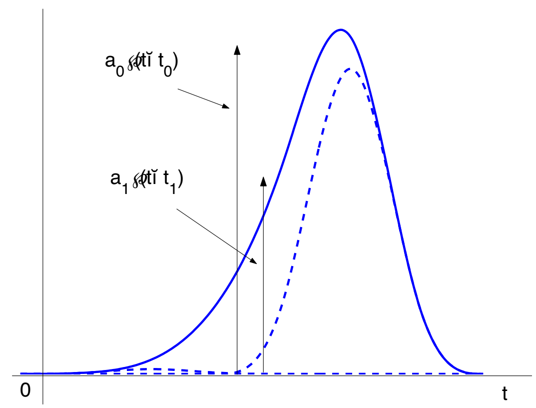
$$\sum_n y_n = a_0 + a_1$$



$$\sum_n n y_n = a_0 t_0 + a_1 t_1$$



$$\sum_n n^2 y_n = a_0 t_0^2 + a_1 t_1^2$$



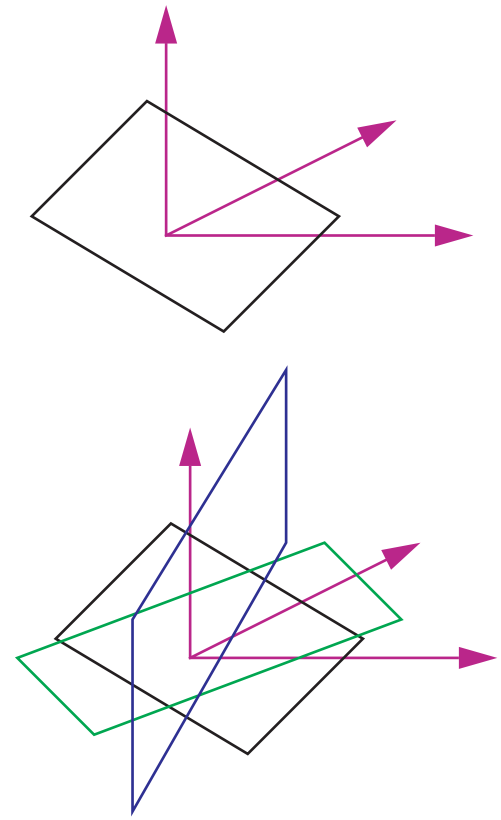
$$\sum_n n^3 y_n = a_0 t_0^3 + a_1 t_1^3$$

local, polynomial complexity reconstruction, for diracs and piecewise polynomials

Conclusions

Wavelets and the French revolution

- too early to say?
- from smooth to piecewise smooth functions



Sparsity and the Art of Motorcycle Maintenance

- sparsity as a key feature with many applications
- denoising, inverse problems, compression

LA versus NLA:

- approximation rates can be vastly different!

To first order, operational, high rate, D(R)

- improvements still possible
- low rate analysis difficult

Two-dimensions:

- really harder! and none used in JPEG2000...
- approximation starts to be understood, compression mostly open
- contourlet leads to efficient algorithms

Beyond subspaces:

- FRI results on sampling, many open questions!

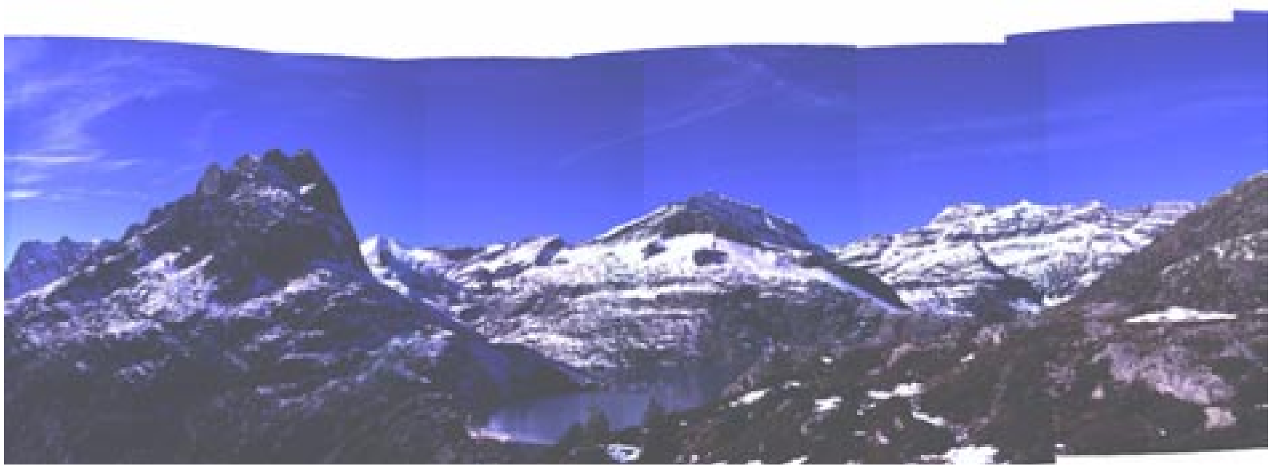
Outlook

Do we understand image representation/compression better?

- high rate, high resolution: there is promise
- low rate: room at the bottom?

New images

- plenoptic functions (set of all possible images)



- non BL images (FRI?)
- manifolds, structure of natural images

Distributed images

- interactive approximation/compression
- SW, WZ, DKLT...

Why Image Representation Remains a Fascinating Topic...

A lone student standing
in front of four tanks.

Publications

For overviews:

- D.Donoho, M.Vetterli, R.DeVore and I.Daubechies, Data Compression and Harmonic Analysis, IEEE Tr. on IT, Oct.1998.
- M. Vetterli, Wavelets, approximation and compression, IEEE Signal Processing Magazine, Sept. 2001

Coming up:

- M.Vetterli, J.Kovacevic and V.Goyal, Fourier and Wavelets: Theory, Algorithms and Applications, Prentice-Hall, 200X ;)

For more details, Theses

- C.Weidmann, Oligoquantization in low-rate lossy source coding, PhD Thesis, EPFL, 2000.
- M. N. Do, Directional Multiresolution Image Representations , Ph.D. Thesis, EPFL, 2001.
- P. L. Dragotti, Wavelet Footprints and Frames for Signal Processing and Communications, PhD Thesis, EPFL, 2002.
- R.Shukla, Rate-distortion optimized geometrical image processing, PhD Thesis, EPFL, 2004.

Papers:

- A.Cohen, I.Daubechies, O.Gulieruz and M.Orchard, On the importance of combining wavelet-based non-linear approximation with coding strategies, IEEE Tr. on IT, 2002
- P. L. Dragotti, M. Vetterli. Wavelets footprints: theory, algorithms and applications, IEEE Transactions on Signal Processing, May 2003.
- R. Shukla, P. L. Dragotti, M. N. Do and M. Vetterli, Rate-distortion optimized tree structured compression algorithms for piecewise smooth images, IEEE Transactions Image Processing, 2004.
- M. N. Do and M. Vetterli, Framing pyramids. IEEE Transactions on Signal Processing, Sept. 2003.
- M. N. Do and M. Vetterli, Contourlets. in Beyond Wavelets, J. Stoeckler and G. V. Welland eds., Academic Press, 2003.
- M.N.Do and M.Vetterli, Contourlets: A computational framework for directional multiresolution image representation, submitted, 2003.
- C.Weidmann, M.Vetterli, Rate-distortion behavior of sparse sources, IEEE Tr. on IT, under revision.
- M. Vetterli, P. Marziliano and T. Blu, Sampling signals with finite rate of innovation, IEEE Transactions on SP, June 2002.
- I. Maravic and M. Vetterli, Sampling and Reconstruction of Signals with Finite Rate of Innovation in the Presence of Noise, IEEE Transactions on SP, 2004, submitted.