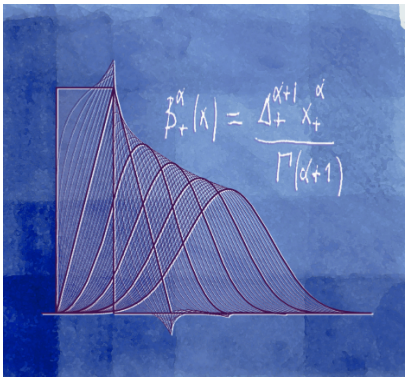
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Sparse stochastic processes

Part I: Introduction

Michael Unser
Biomedical Imaging Group
EPFL, Lausanne, Switzerland



Invited seminar, Korea Advanced Inst. of Science & Technology (KAIST), July 17-18, Seoul, Korea

20th century statistical signal processing

Hypothesis: Signal = stationary **Gaussian** process

Karhunen-Loève transform (KLT) is optimal for compression

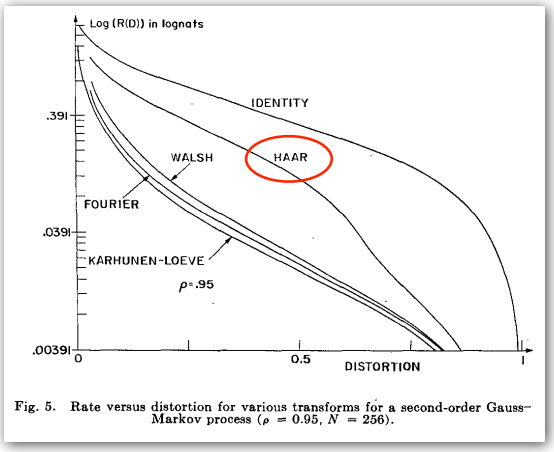


Fig. 5. Rate versus distortion for various transforms for a second-order Gauss–Markov process ($\rho = 0.95$, $N = 256$).

(Pearl et al., *IEEE Trans. Com* 1972)

DCT asymptotically equivalent to KLT

(Ahmed-Rao, 1975; U., 1984)

JPEG

20th century statistical signal processing

Hypothesis: Signal = **Gaussian** process

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}$$

Noise: i.i.d. Gaussian with variance σ^2

Signal covariance: $\mathbf{C}_s = \mathbb{E}\{\mathbf{s} \cdot \mathbf{s}^T\}$

Wiener filter is **optimal** for restoration/denoising

$$\mathbf{s}_{\mathrm{LMMSE}} = \mathbf{C}_s \mathbf{H}^T \left(\mathbf{H}\mathbf{C}_s\mathbf{H}^T + \sigma^2\mathbf{I}\right)^{-1} \mathbf{y} = \mathbf{F}_{\mathrm{Wiener}}\, \mathbf{y}$$

$\Updownarrow$ $\mathbf{L} = \mathbf{C}_s^{-1/2}$: Whitening filter

Wiener (LMMSE) solution = Gauss MMSE = Gauss MAP

$$\mathbf{s}_{\mathrm{MAP}} = \arg\min_{\mathbf{s}} \underbrace{\frac{1}{\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2}_{\text{Data Log likelihood}} + \underbrace{\|\mathbf{C}_s^{-1/2}\mathbf{s}\|_2^2}_{\text{Gaussian prior likelihood}}$$

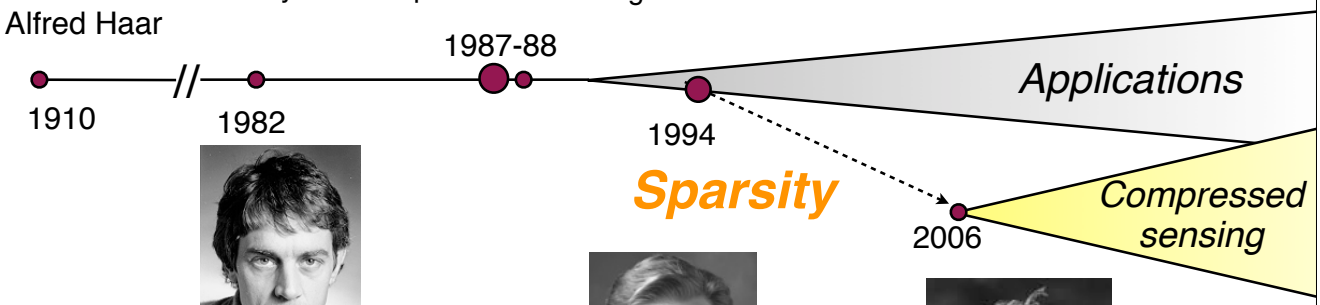
$\Leftrightarrow$ quadratic regularization (Tikhonov) $\lambda\|\mathbf{L}\mathbf{s}\|_2^2$

Then came wavelets ... and sparsity

Yves Meyer

Stéphane Mallat

Ingrid Daubechies

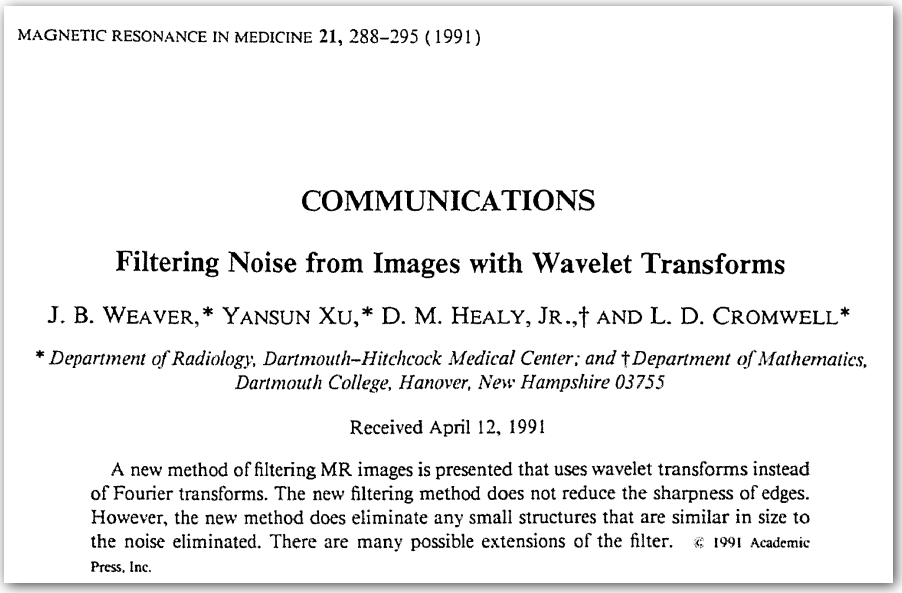


Martin Vetterli

David Donoho

Emmanuel Candès

Fact 1: Wavelets can outperform Wiener filter



MAGNETIC RESONANCE IN MEDICINE **21**, 288–295 (1991)

COMMUNICATIONS

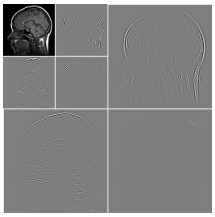
Filtering Noise from Images with Wavelet Transforms

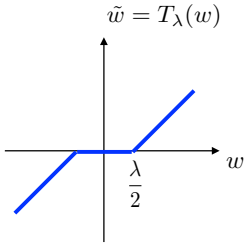
J. B. WEAVER,* YANSUN XU,* D. M. HEALY, JR.,† AND L. D. CROMWELL*

** Department of Radiology, Dartmouth–Hitchcock Medical Center; and † Department of Mathematics, Dartmouth College, Hanover, New Hampshire 03755*

Received April 12, 1991

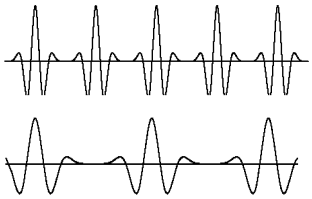
A new method of filtering MR images is presented that uses wavelet transforms instead of Fourier transforms. The new filtering method does not reduce the sharpness of edges. However, the new method does eliminate any small structures that are similar in size to the noise eliminated. There are many possible extensions of the filter. © 1991 Academic Press, Inc.

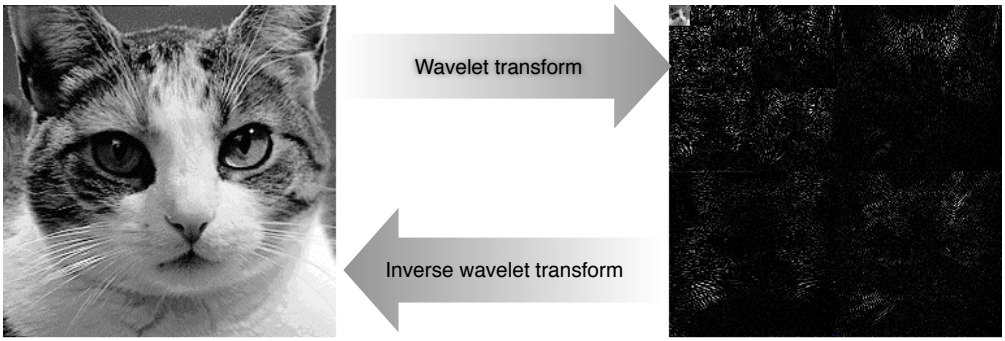




Fact 2: Wavelet coding can outperform jpeg

$$f(\boldsymbol{x}) = \sum_{i,\boldsymbol{k}} \psi_{i,\boldsymbol{k}}(\boldsymbol{x})\, w_{i,\boldsymbol{k}}$$





66.4 dB

0.00%

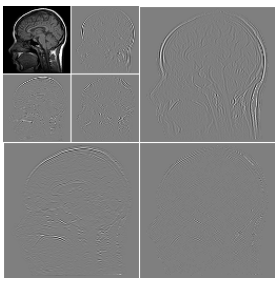
Discarding "small coefficients"

(Shapiro, *IEEE-IP* 1993)

the jpeg 2000 alliance

Fact 3: l_1 schemes can outperform l_2 optimization

$$\mathbf{s}^\star = \text{argmin} \underbrace{\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2}_{\text{data consistency}} + \underbrace{\lambda \mathcal{R}(\mathbf{s})}_{\text{regularization}}$$



- Wavelet-domain regularization

  Wavelet expansion: $\mathbf{s} = \mathbf{W}\mathbf{v}$ (typically, sparse)

  Wavelet-domain sparsity-constraint: $\mathcal{R}(\mathbf{s}) = \|\mathbf{v}\|_{\ell_1}$ with $\mathbf{v} = \mathbf{W}^{-1}\mathbf{s}$

  (Nowak et al., Daubechies et al. 2004)

- ℓ_1 regularization (Total variation) (Rudin-Osher, 1992)

  $\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_{\ell_1}$ with $\mathbf{L}$: gradient

- Compressed sensing/sampling (Candes-Romberg-Tao; Donoho, 2006)

SPARSE STOCHASTIC MODELS: The step beyond Gaussianity

Requirements for a comprehensive statistical framework

- Backward compatibility
- Continuous-domain formulation

  piecewise-smooth signals, translation and scale-invariance, sampling ...
- Predictive power

  Can wavelets really outperform sinusoidal transforms (KLT) ?
- Statistical justification and refinement of current algorithms

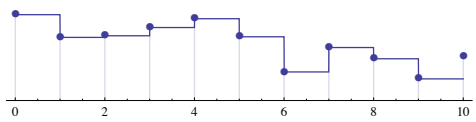
  Sparsity-promoting regularization, ℓ_1 norm minimization

1.3 FROM SPLINES TO STOCHASTIC PROCESSES

- Splines and Legos revisited
- Higher-order polynomial splines
- Random splines, innovations, Lévy processes
- Wavelet analysis of Lévy processes
- Lévy's synthesis of Brownian motion

Splines and Legos revisited

- Cardinal spline of degree 0: piecewise-contant



$$f_1(t) = \sum_{k\in\mathbb{Z}} f_1[k]\beta_+^0(t-k)$$

$$\beta_+^0(t) = \begin{cases} 1, & \text{for } 0 \le t < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Notion of D-spline:

$$\mathrm{D}f_1(t) = \sum_{k\in\mathbb{Z}} a_1[k]\delta(t-k)$$

B-spline and derivative operator

Derivative $\quad \mathrm{D}f(t) = \dfrac{\mathrm{d}f(t)}{\mathrm{d}t} \qquad \mathrm{D} \overset{\mathcal{F}}{\longleftrightarrow} \mathrm{j}\omega$

Finite difference operator

$$\mathrm{D_d}f(t) = f(t) - f(t-1) \qquad \mathrm{D_d} \overset{\mathcal{F}}{\longleftrightarrow} 1 - \mathrm{e}^{-\mathrm{j}\omega}$$

$$= (\beta_+^0 * \mathrm{D}f)(t)$$

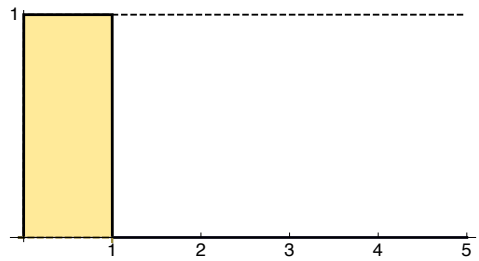
B-spline of degree 0

$$\beta_+^0(t) = \mathrm{D_d}\mathrm{D}^{-1}\delta(t) = \mathrm{D_d}\mathbb{1}_+(t)$$

$$\updownarrow$$

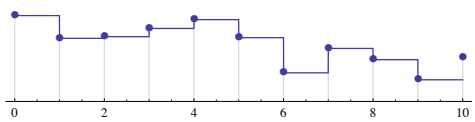
$$\hat{\beta}_+^0(\omega) = \frac{1 - \mathrm{e}^{-\mathrm{j}\omega}}{\mathrm{j}\omega}$$

$$\beta_+^0(t) = \mathbb{1}_+(t) - \mathbb{1}_+(t-1)$$

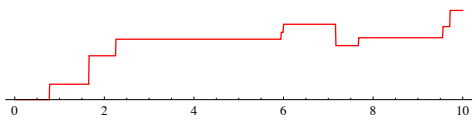


Random splines and innovations

cardinal



non-uniform



$$\mathrm{D}s(t) = \sum_n a_n \delta(t - t_n) = w(t)$$

Random weights $\{a_n\}$ i.i.d. and random knots $\{t_n\}$ (Poisson with rate λ)

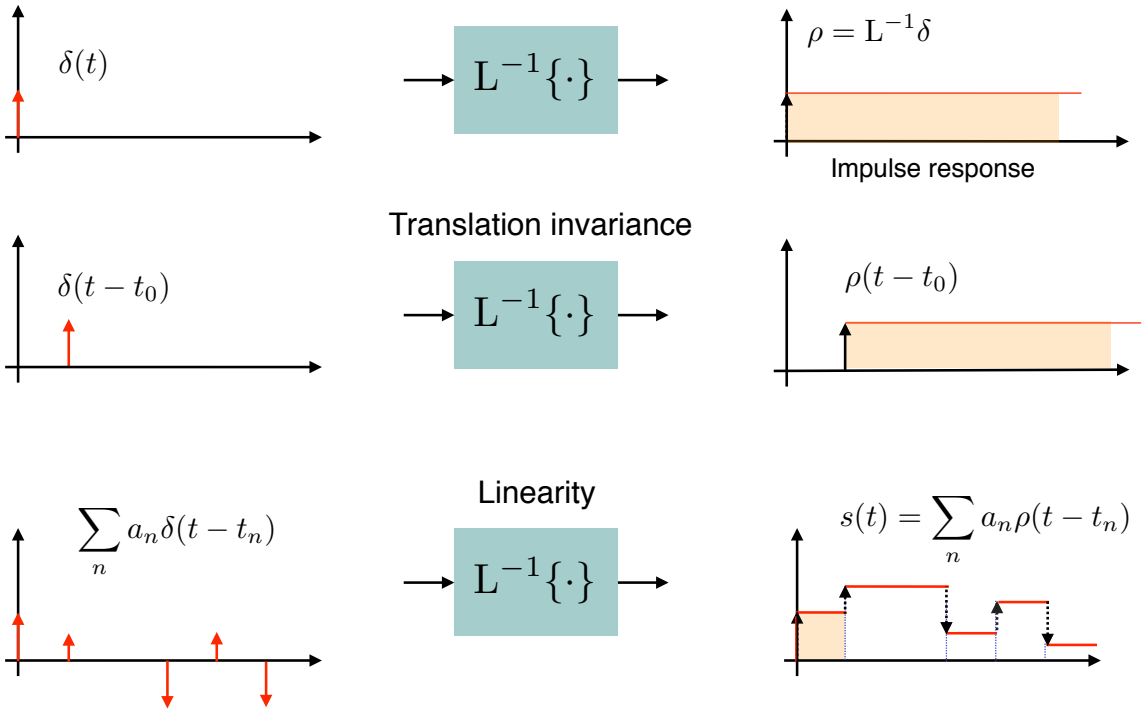
- Anti-derivative operators

Shift-invariant solution: $\mathrm{D}^{-1}\varphi(t) = (\mathbb{1}_+ * \varphi)(t) = \displaystyle\int_{-\infty}^{t} \varphi(\tau)\mathrm{d}\tau$

Scale-invariant solution: $\mathrm{D}_0^{-1}\varphi(t) = \displaystyle\int_{0}^{t} \varphi(\tau)\mathrm{d}\tau$

Innovation-based synthesis

$$\mathrm{L} = \frac{\mathrm{d}}{\mathrm{d}t} = \mathrm{D} \quad \Rightarrow \quad \mathrm{L}^{-1}\text{: integrator}$$

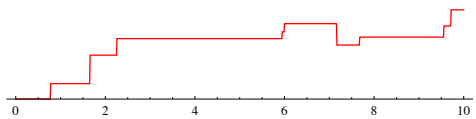


Compound Poisson process

- Stochastic differential equation

$$\mathrm{D}s(t) = w(t)$$

with boundary condition $s(0) = 0$



Innovation: $\quad w(t) = \sum_n a_n \delta(t - t_n)$

- Formal solution

$$s(t) = \mathrm{D}_0^{-1} w(t) = \sum_n a_n \mathrm{D}_0^{-1}\{\delta(\cdot - t_n)\}(t)$$

$$= \sum_n a_n \big(\mathbb{1}_+(t - t_n) - \mathbb{1}_+(-t_n)\big)$$

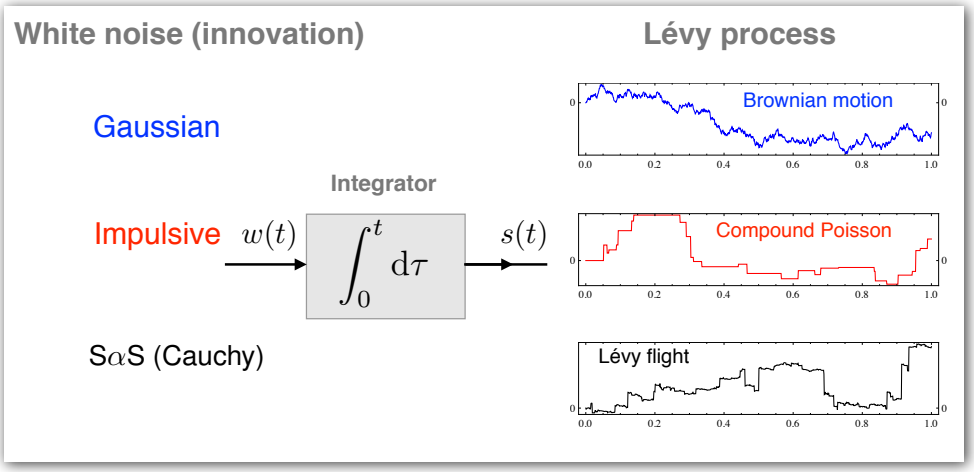
(impose boundary condition)

Generalization: Lévy processes

Generalized innovations : white Lévy noise with $\mathbb{E}\{w(t)w(t')\} = \sigma_w^2 \delta(t-t')$

$\mathrm{D}s = w$ (unstable SDE !)

$$s = \mathrm{D}_0^{-1} w \quad \Leftrightarrow \quad \forall \varphi \in \mathcal{S}, \quad \langle \varphi, s \rangle = \langle \mathrm{D}_0^{-1*} \varphi, w \rangle$$



(Wiener 1923)

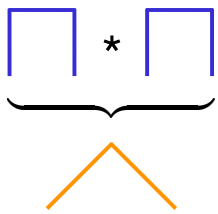
(Paul Lévy circa 1930)

Decoupling Lévy processes: increments

Increment process: $u(t) = \mathrm{D_d} s(t) = \mathrm{D_d} \mathrm{D}_0^{-1} w(t) = (\beta_+^0 * w)(t).$

Increment process is stationary with autocorrelation function

$$R_u(\tau) = \mathbb{E}\{u(t)u(t+\tau)\} = \left(\beta_+^0 * (\beta_+^0)^\vee * R_w\right)(y)$$
$$= \sigma_w^2 \beta_+^1(\tau - 1)$$



Discrete increments

$$u[k] = s(k) - s(k-1) = \langle w, \beta_+^0(\cdot - k) \rangle.$$

$u[k]$ are i.i.d. because

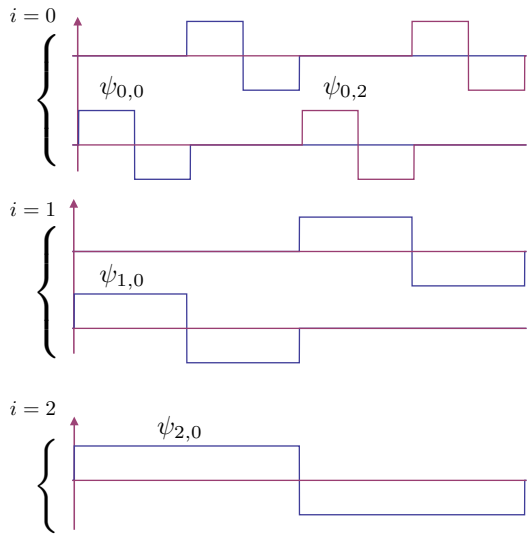
- $\{\beta_+^0(\cdot - k)\}$ are non-overlapping
- w is independent at every point (white noise)

Wavelet analysis of Lévy processes

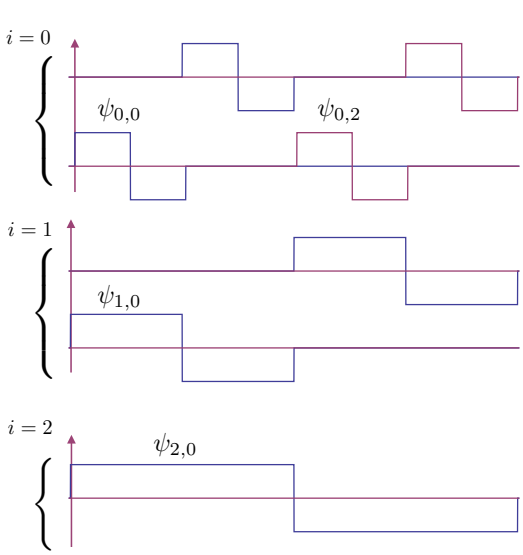
- Haar wavelets

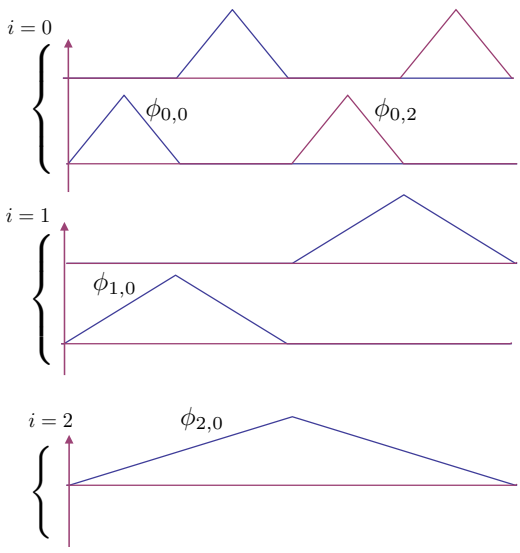
$$\psi_{\mathrm{Haar}}(t) = \begin{cases} 1, & \text{for } 0 \le t < \frac{1}{2} \\ -1, & \text{for } \frac{1}{2} \le t < 1 \\ 0, & \text{otherwise.} \end{cases}$$

$$\psi_{i,k}(t) = 2^{-i/2}\psi_{\mathrm{Haar}}\left(\frac{t - 2^i k}{2^i}\right)$$



Wavelets as multi-scale derivatives





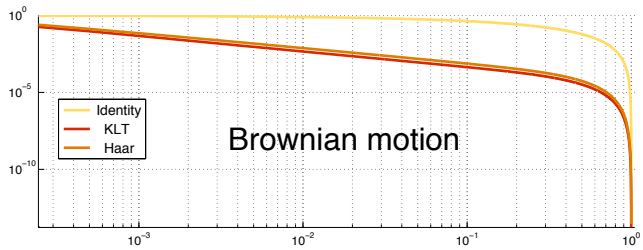
$$\psi_{i,k} = 2^{i/2-1}\mathrm{D}\phi_{i,k}$$
$$\mathrm{D}_0^{-1}\psi_{i,k} = 2^{i/2-1}\phi_{i,k}.$$

- Wavelet coefficients of Lévy process

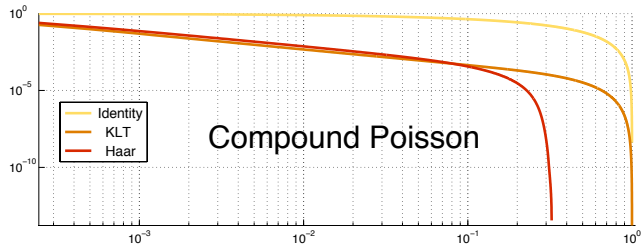
$$Y_{i,k} = \langle s, \psi_{i,k}\rangle \propto \langle s, \mathrm{D}\phi_{i,k}\rangle$$
$$\propto \langle \mathrm{D}^* s, \phi_{i,k}\rangle = -\langle w, \phi_{i,k}\rangle$$

M-term approximation: wavelets vs. KLT

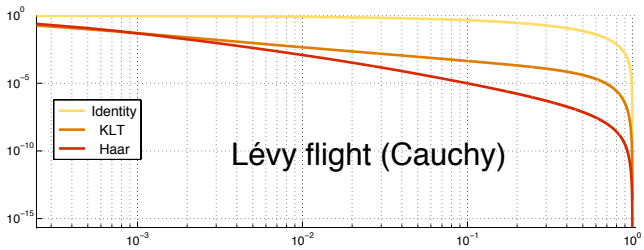
Gaussian



Finite rate of innovation



Even sparser ...



Wavelet-based synthesis of Brownian motion

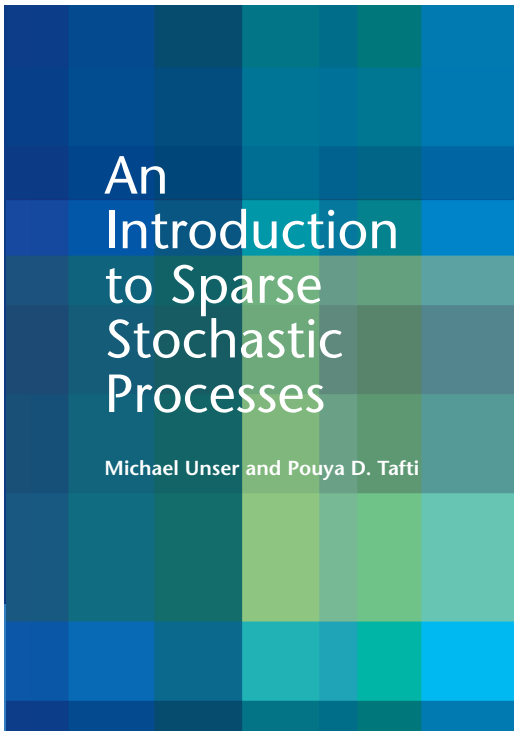
- White Gaussian noise

$$w = \sum_{i\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} Z_{i,k}\psi_{i,k} \text{ with } Z_{i,k} = \langle w, \psi_{i,k}\rangle$$

- Brownian motion

$$\begin{aligned} s(t) &= \mathrm{D}_0^{-1} w \\ &= \sum_{i\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} Z_{i,k} \mathrm{D}_0^{-1} \psi_{i,k}(t) \\ &= \sum_{i\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} 2^{i/2-1} Z_{i,k} \phi_{i,k}(t) \end{aligned}$$

CAMBRIDGE UNIVERSITY PRESS



ebook: web preprint

http://www.sparseprocesses.org

Extension to fractional orders: B-splines

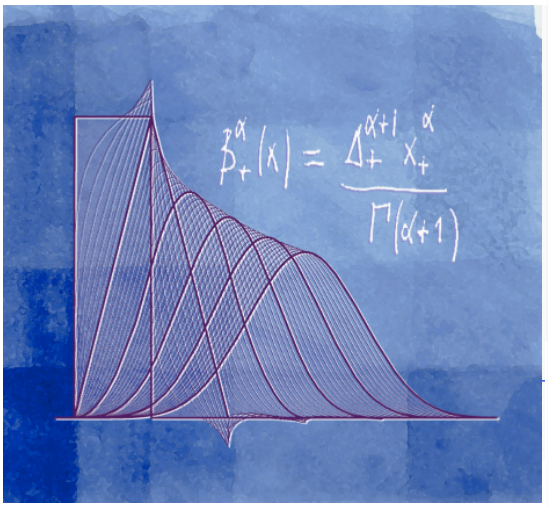
$\mathrm{L} = \mathrm{D}^\gamma$ (fractional derivative) Fourier multiplier: $(\mathrm{j}\omega)^\gamma$

$\mathrm{L_d} = \Delta_+^\gamma$ (fractional differences) Fourier multiplier: $(1 - \mathrm{e}^{-\mathrm{j}\omega})^\gamma$

$$\beta_+^0(r) = \Delta_+ r_+^0 \quad \overset{\mathcal{F}}{\longleftrightarrow} \quad \frac{1 - \mathrm{e}^{-\mathrm{j}\omega}}{\mathrm{j}\omega}$$

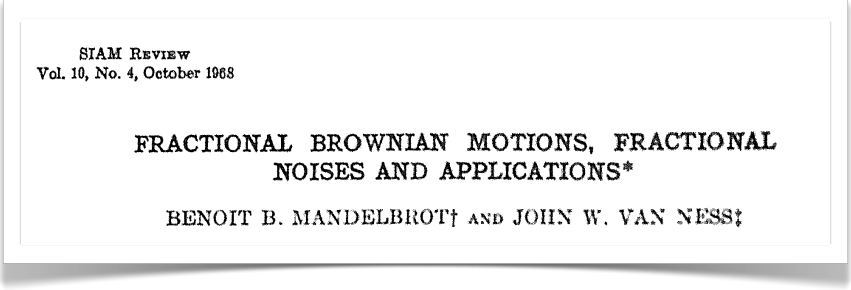
$$\vdots \qquad\qquad \vdots$$

$$\beta_+^\alpha(r) = \frac{\Delta_+^{\alpha+1} r_+^\alpha}{\Gamma(\alpha+1)} \quad \overset{\mathcal{F}}{\longleftrightarrow} \quad \left(\frac{1 - \mathrm{e}^{-\mathrm{j}\omega}}{\mathrm{j}\omega}\right)^{\alpha+1}$$



One-sided power function: $r_+^\alpha = \begin{cases} r^\alpha, & r \geq 0 \\ 0, & r < 0 \end{cases}$ Degree $\alpha = \gamma - 1$

Mandelbrot's fractional Brownian motion



SIAM Review
Vol. 10, No. 4, October 1968

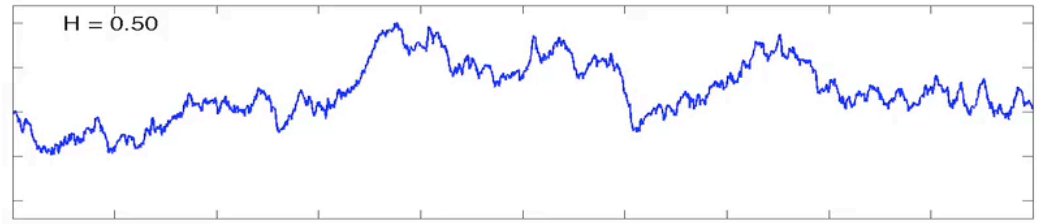
FRACTIONAL BROWNIAN MOTIONS, FRACTIONAL NOISES AND APPLICATIONS*

BENOIT B. MANDELBROT† AND JOHN W. VAN NESS‡

Solution of fractional SDE (Blu-U., 2007)

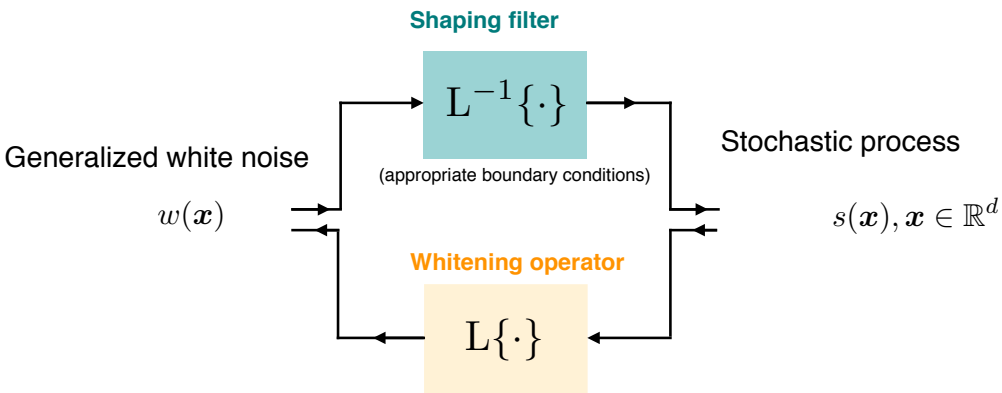
$\mathrm{D}^\gamma s = w$ where w is **Gaussian** white noise $\Rightarrow \quad s = \mathrm{I}_2^\gamma w$

I_2^γ: scale-invariant, right-inverse of D^γ



Hurst exponent: $H = \gamma - \frac{1}{2}$

Continuous-domain innovation model



Main outcome: non-Gaussian solutions are **necessarily** sparse (***infinitely divisible***)

Why? ... as will explained in next chapters ...

(invoking powerful theorems in functional analysis: Bochner-Minlos, Gelfand, Schoenberg & Lévy-Khinchine)

2.1 DECOUPLING OF SPARSE PROCESSES

$$s = \mathrm{L}^{-1} w \qquad \Leftrightarrow \qquad w = \mathrm{L} s$$

- Discrete approximation of operator
- Operator-like wavelet analysis

Decoupling: Linear combination of samples

Input: $s(\boldsymbol{k}), \boldsymbol{k} \in \mathbb{Z}^d$ (sampled values) $\qquad s = \mathrm{L}^{-1} w$

Discrete approximation of whitening operator: L_d

$$\mathrm{L}_\mathrm{d} \delta(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} d_\mathrm{L}[\boldsymbol{k}] \delta(\boldsymbol{x} - \boldsymbol{k})$$

Discrete increment process:

$$u[\boldsymbol{k}] = \mathrm{L}_\mathrm{d} s(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{k}} = (\beta_\mathrm{L} * w)(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{k}} = \langle \underbrace{\beta_\mathrm{L}^\vee(\cdot - \boldsymbol{k})}_{\varphi}, w \rangle$$

Generalized B-spline:

$$\beta_\mathrm{L}(\boldsymbol{x}) = \mathrm{L}_\mathrm{d} \mathrm{L}^{-1} \delta(\boldsymbol{x})$$

A-to-D translator

Decoupling: Wavelet analysis

$$\mathrm{L}s = w$$

Generalized operator-like wavelets:

$$\psi_i(\boldsymbol{x}) = \mathrm{L}^* \phi_i(\boldsymbol{x})$$

(Khalidov-U. 2006, Ward-U. JFAA 2013)

Operator-like wavelet analysis of sparse process:

$$\begin{aligned}\langle \psi_i(\cdot - \boldsymbol{x}_0), s\rangle =& \langle \mathrm{L}^* \phi_i(\cdot - \boldsymbol{x}_0), s\rangle \\ =& \langle \phi_i(\cdot - \boldsymbol{x}_0), \mathrm{L}s)\rangle \\ =& \langle \underbrace{\phi_i(\cdot - \boldsymbol{x}_0)}_{\varphi}, w\rangle = (\phi_i^\vee * w)(\boldsymbol{x}_0)\end{aligned}$$