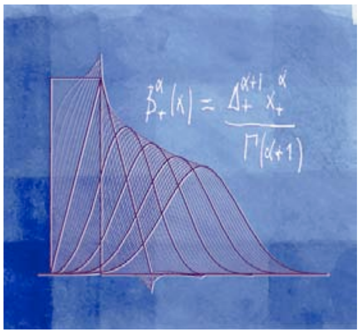
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Sparse stochastic processes

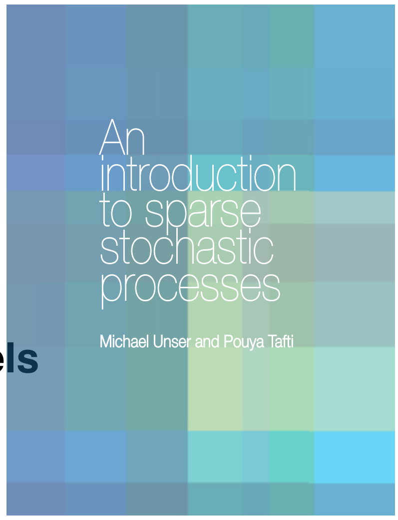
Part 1:
Theoretical Foundations

Prof. Michael Unser, LIB



EPFL Doctoral School EDEE, Course EE-726, Spring 2017

Table of content



1. **Introduction**
2. **Roadmap to the monograph**
3. *Mathematical context and background*
4. **Continuous-domain innovation models**
5. Operators and their inverses
6. Splines and wavelets
7. **Sparse stochastic processes**
8. **Sparse representations**
9. Infinite divisibility and transform-domain statistic
10. **Recovery of sparse signals**
11. Wavelet-domain methods

20th century statistical signal processing

Hypothesis: Signal = stationary **Gaussian** process

Karhunen-Loève transform (KLT) is optimal for compression

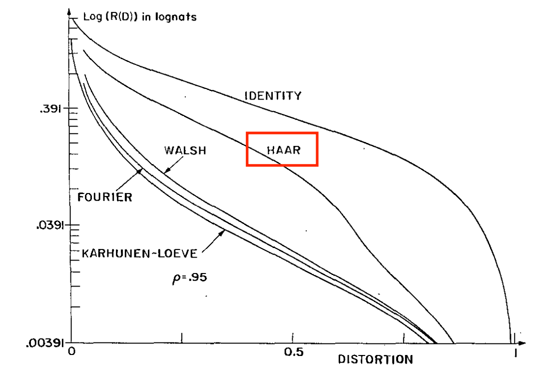


Fig. 5. Rate versus distortion for various transforms for a second-order Gauss–Markov process (ρ = 0.95, N = 256).

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

DCT asymptotically equivalent to KLT

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

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

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

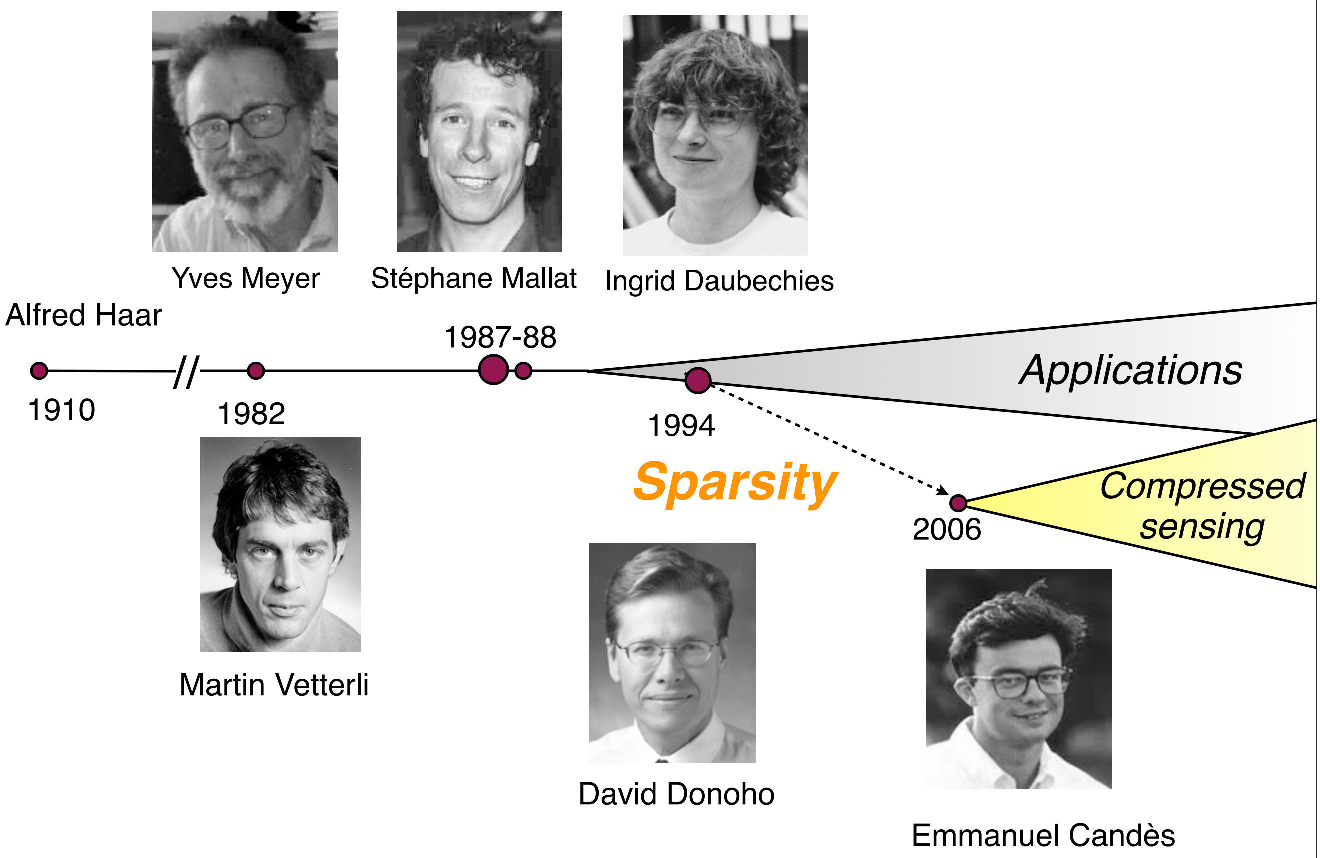
$\Updownarrow$ $\mathbf{L} = \mathbf{C}_s^{-1/2}$: Whitening filter

Wiener (LMMSE) solution = Gauss MMSE = Gauss MAP

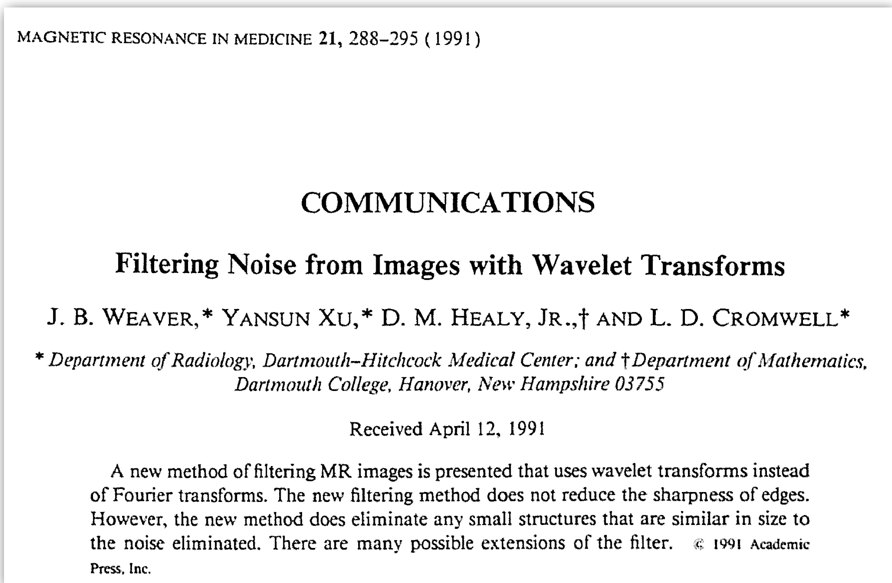
$$\mathbf{s}_{\mathrm{MAP}} = \arg\min_{\mathbf{s}} \underbrace{\frac{1}{\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2}_{\text{Data Log likelihood}} + \underbrace{\|\mathbf{C}_s^{-1/2}\mathbf{s}\|_2^2}_{\text{Gaussian prior likelihood}}$$

$\Leftrightarrow$ quadratic regularization (Tikhonov)

Then came wavelets ... and sparsity



Fact 1: Wavelets can outperform Wiener filter



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

COMMUNICATIONS

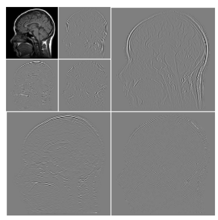
Filtering Noise from Images with Wavelet Transforms

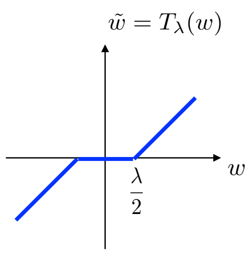
J. B. WEAVER,* YANSUN XU,* D. M. HEALY, JR.,† AND L. D. CROMWELL*

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

Received April 12, 1991

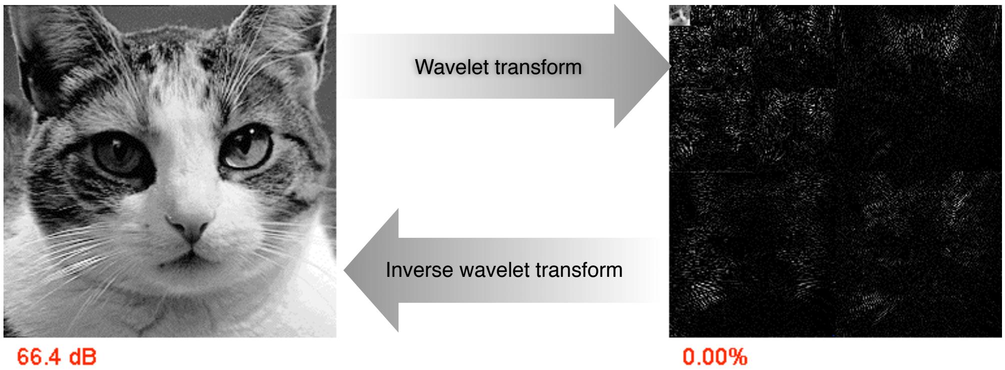
A new method of filtering MR images is presented that uses wavelet transforms instead of Fourier transforms. The new filtering method does not reduce the sharpness of edges. However, the new method does eliminate any small structures that are similar in size to the noise eliminated. There are many possible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





Fact 2: Wavelet coding can outperform jpeg

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



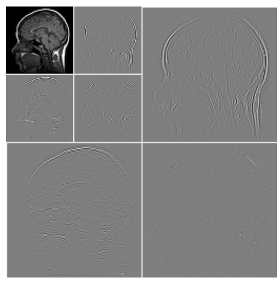
Discarding “small coefficients”

(Shapiro, *IEEE-IP* 1993)

the jpeg 2000 alliance

Fact 3: l_1 schemes can outperform l_2

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



- 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 $\mathbf{v} = \mathbf{W}^{-1}\mathbf{s}$

  Iterated shrinkage-thresholding algorithm (ISTA, FISTA)

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

- ℓ_1 regularization (Total variation)

  $\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_{\ell_1}$ with $\mathbf{L}$: gradient (Rudin-Osher, 1992)

  Iterative reweighted least squares (IRLS) or FISTA

1.2 SPARSE STOCHASTIC MODELS: The step beyond Gaussianity

Requirements for a comprehensive statistical framework

- Backward compatibility
- Continuous-domain formulation

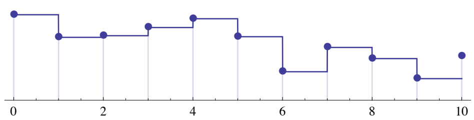
  piecewise-smooth signals, translation and scale-invariance, sampling . . .
- Predictive power

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

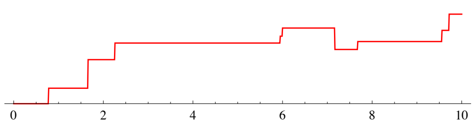
  Sparsity-promoting regularization, ℓ_1 norm minimization

Random spline: archetype of sparse signal

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

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

B-spline and derivative operator

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

Finite difference operator

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

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

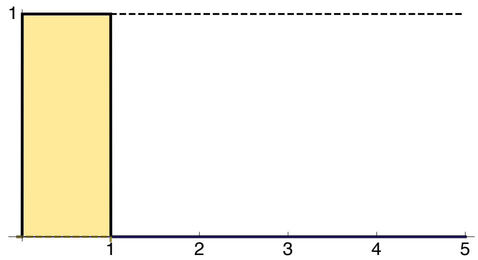
B-spline of degree 0

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

$$\updownarrow$$

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

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



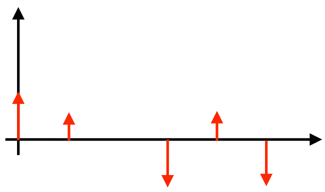
Compound Poisson process

- Stochastic differential equation

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

with boundary condition $s(0) = 0$

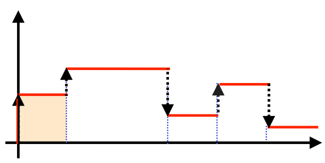
Innovation: $\quad w(t) = \sum_n a_n \delta(t - t_n)$



- Formal solution

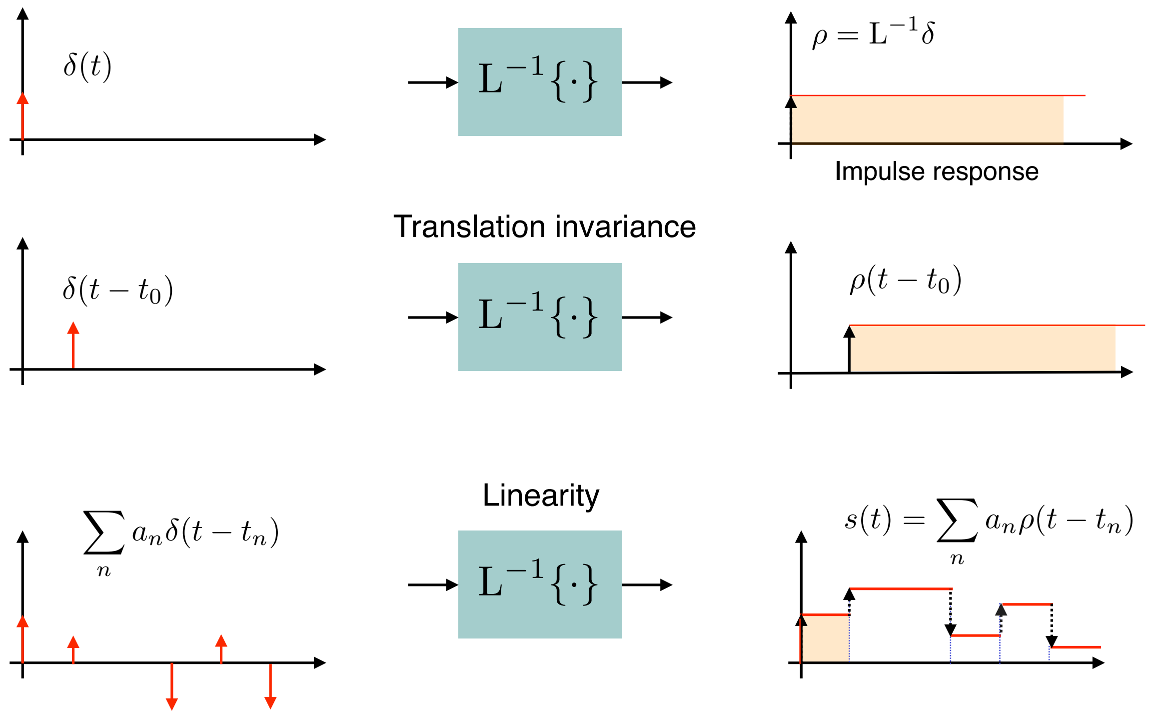
$$s(t) = \mathrm{D}^{-1}w(t) = \sum_n a_n \mathrm{D}^{-1}\{\delta(\cdot - t_n)\}(t)$$

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



Innovation-based synthesis

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

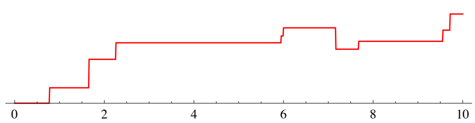


Compound Poisson process

- Stochastic differential equation

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

with boundary condition $s(0) = 0$



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

- Formal solution

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

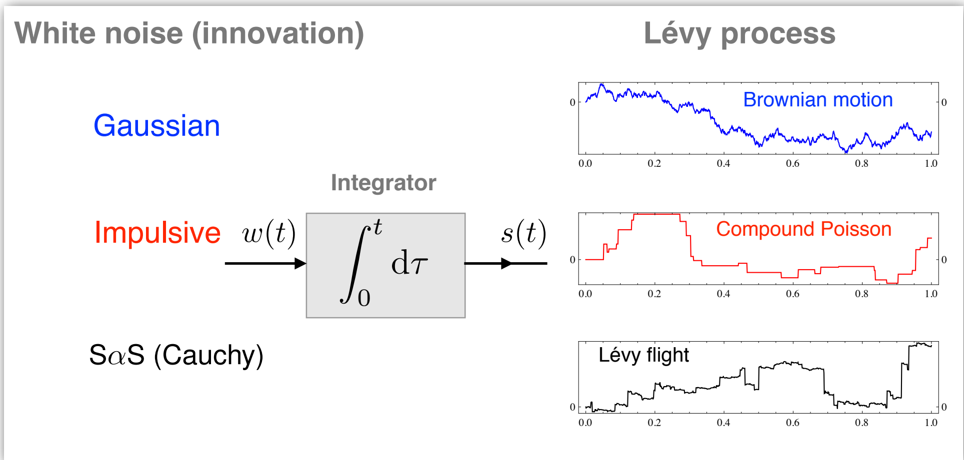
$$= \sum_n a_n \big(\mathbb{1}_+(t - t_n) - \mathbb{1}_+(-t_n)\big)$$

(impose boundary condition)

Lévy processes: all admissible brands of innovations

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

$$\mathrm{D}s = w \qquad \text{(perfect decoupling!)}$$



(Wiener 1923)

(Paul Lévy circa 1930)

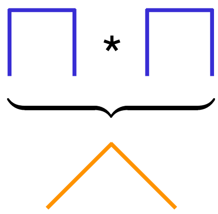
Decoupling Lévy processes: increments

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

Increment process is stationary with autocorrelation function

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

with $\quad (\beta_+^0)^\vee(t) = \beta_+^0(-t)$



Discrete increments

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

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

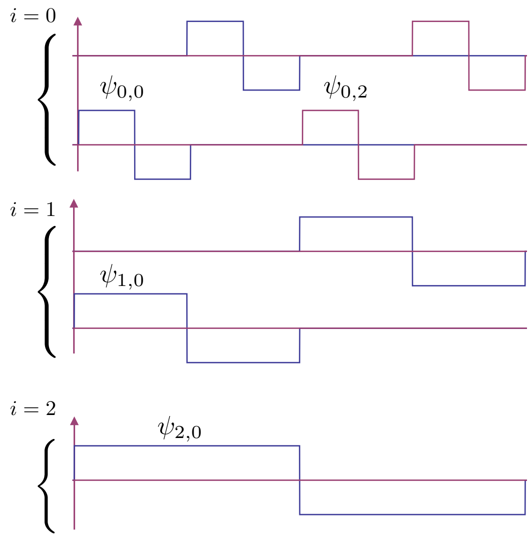
- $\{\beta_+^0(\cdot - k)\}$ are non-overlapping
- w is independent at every point (white noise)

Wavelet analysis of Lévy processes

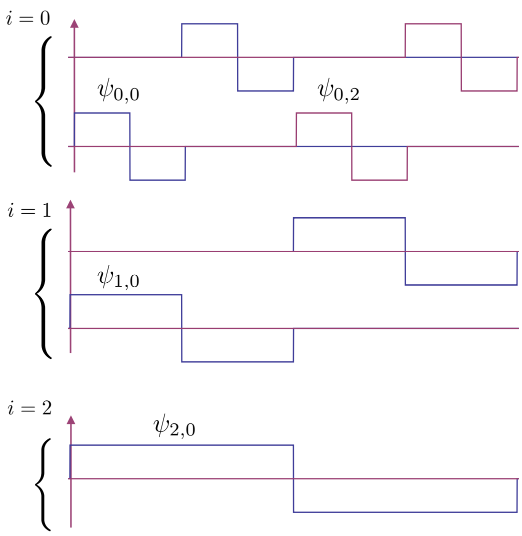
- Haar wavelets

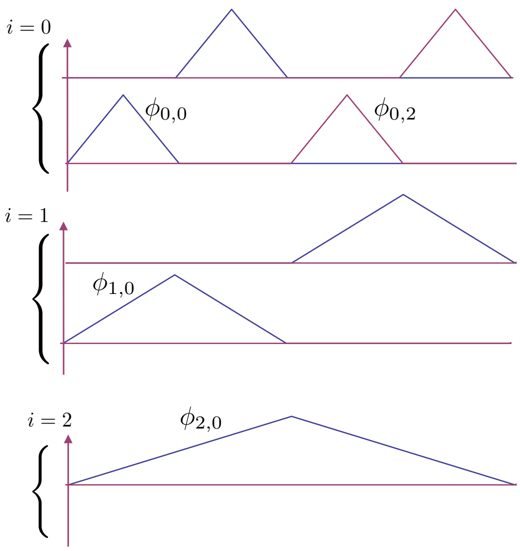
$$\psi_{\text{Haar}}(t) = \begin{cases} 1, & \text{for } 0 \le t < \frac{1}{2} \\ -1, & \text{for } \frac{1}{2} \le t < 1 \\ 0, & \text{otherwise.} \end{cases}$$

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



Wavelets as multi-scale derivatives





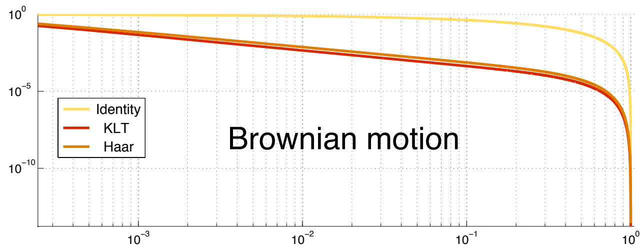
$$\psi_{i,k} = 2^{i/2-1}\mathrm{D}\phi_{i,k}$$
$$\mathrm{D}_0^{-1}\psi_{i,k} = 2^{i/2-1}\phi_{i,k}.$$

- Wavelet coefficients of Lévy process

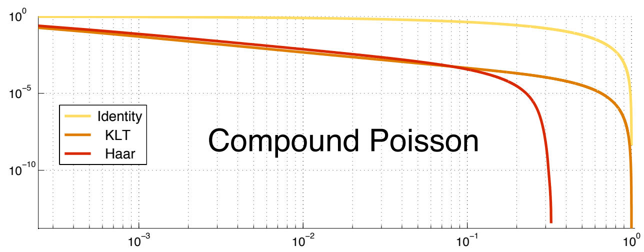
$$Y_{i,k} = \langle s, \psi_{i,k}\rangle \propto \langle s, \mathrm{D}\phi_{i,k}\rangle$$
$$\propto \langle \mathrm{D}^* s, \phi_{i,k}\rangle = -\langle w, \phi_{i,k}\rangle$$

M-term approximation: wavelets vs. KLT

Gaussian



Finite rate of innovation



Even sparser ...

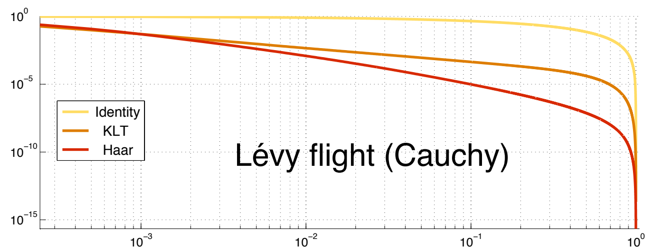
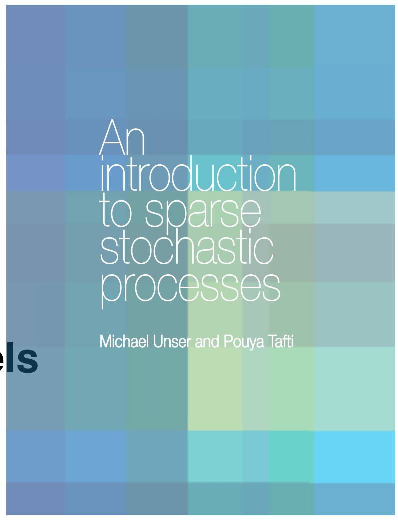
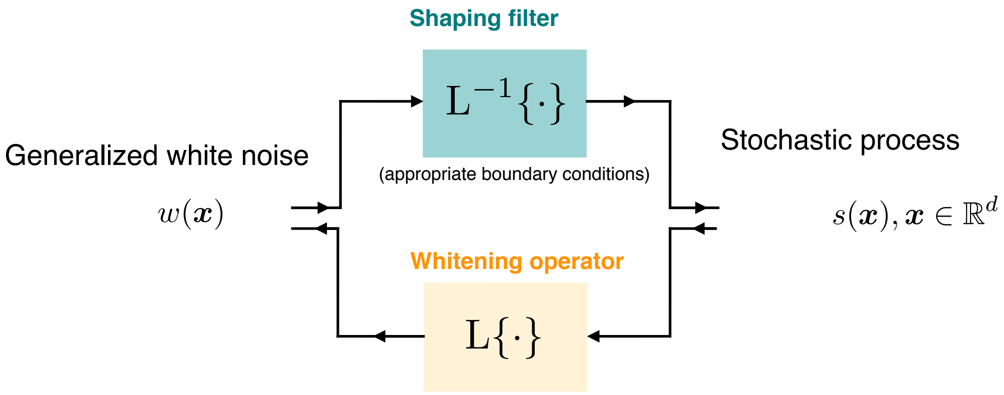


Table of content



1. **Introduction** ✔
2. **Roadmap to the monograph**
3. *Mathematical context and background*
4. **Continuous-domain innovation models**
5. Operators and their inverses
6. Splines and wavelets
7. **Sparse stochastic processes**
8. **Sparse representations**
9. Infinite divisibility and transform-domain statistic
10. **Recovery of sparse signals**
11. Wavelet-domain methods

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)

Short primer on probability theory

Random variable X

- Probability measure and density function (pdf)

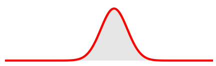
$$\mathrm{Prob}(X \in E) = \mathscr{P}_X(E) = \int_E p_X(x)\mathrm{d}x$$

Expectation: $$\mathbb{E}\{f(X)\} = \int_{\mathbb{R}} f(x)\mathscr{P}_X(\mathrm{d}x) = \int_{\mathbb{R}} f(x)p_X(x)\mathrm{d}x$$

- Characteristic function

$$\hat{p}_X(\omega) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega X}\} = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{j}\omega x} p_X(x)\mathrm{d}x$$

Example: Gaussian



$$p_X(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$$

$$\updownarrow \mathcal{F}$$

$$\hat{p}_X(\omega) = \mathrm{e}^{-\omega^2/2}$$

Bochner's theorem

Let $\widehat{p}_X : \mathbb{R} \to \mathbb{C}$ be a continuous, positive-definite function such that $\widehat{p}_X(0) = 1$.
Then, there exists a unique Borel probability measure $\mathscr{P}_X$ on $\mathbb{R}$, such that

$$\hat{p}_X(\omega) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{j}\omega x}\mathscr{P}_X(\mathrm{d}x) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{j}\omega x} p_X(x)\mathrm{d}x$$

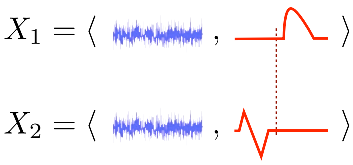
Generalized innovation process

- Difficulty 1: $w \neq w(\boldsymbol{x})$ is too rough to have a pointwise interpretation

- Difficulty 2: w is an infinite-dimensional random entity;
its "pdf" can be formally specified by a measure $\mathscr{P}_w(E)$ where $E \subseteq \mathcal{S}'(\mathbb{R}^d)$

- Axiomatic definition (Gelfand-Vilenkin 1964)

> w is a generalized innovation process (or continuous-domain white noise) in $\mathcal{S}'(\mathbb{R}^d)$ if
>
> 1. ***Observability*** : $X = \langle \varphi, w \rangle$ is a well-defined random variable for any test function $\varphi \in \mathcal{S}(\mathbb{R}^d)$.
> 2. ***Stationarity*** : $X_{\boldsymbol{x}_0} = \langle \varphi(\cdot - \boldsymbol{x}_0), w \rangle$ is identically distributed for all $\boldsymbol{x}_0 \in \mathbb{R}^d$.
> 3. ***Independent atoms*** : $X_1 = \langle \varphi_1, w \rangle$ and $X_2 = \langle \varphi_2, w \rangle$ are independent whenever φ_1 and φ_2 have non-intersecting support.



- Characteristic functional ($\omega \rightarrow \varphi$)

$$\widehat{\mathscr{P}}_w(\varphi) = \mathbb{E}\{e^{\mathrm{j}\langle \varphi, w \rangle}\} = \int_{\mathcal{S}'} e^{\mathrm{j}\langle \varphi, g \rangle} \mathscr{P}_w(\mathrm{d}g)$$

finite-dimensional	infinite-dimensional
random variable X in $\mathbb{R}^N$	generalized stochastic process s in $\mathscr{S}'$
probability measure $\mathscr{P}_X$ on $\mathbb{R}^N$ $\mathscr{P}_X(E) = \mathrm{Prob}(X \in E) = \int_E p_X(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ (p_X is a generalized [i.e., hybrid] pdf) for suitable subsets $E \subset \mathbb{R}^N$	probability measure $\mathscr{P}_s$ on $\mathscr{S}'$ $\mathscr{P}_s(E) = \mathrm{Prob}(s \in E) = \int_E \mathscr{P}_s(\mathrm{d}g)$ for suitable subsets $E \subset \mathscr{S}'$
characteristic function $\widehat{\mathscr{P}}_X(\boldsymbol{\omega}) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle \boldsymbol{\omega}, X \rangle}\} = \int_{\mathbb{R}^N} \mathrm{e}^{\mathrm{j}\langle \boldsymbol{\omega}, \boldsymbol{x} \rangle} p_X(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$, $\boldsymbol{\omega} \in \mathbb{R}^N$	characteristic functional $\widehat{\mathscr{P}}_s(\varphi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle \varphi, s \rangle}\} = \int_{\mathscr{S}'} \mathrm{e}^{\mathrm{j}\langle \varphi, g \rangle} \mathscr{P}_s(\mathrm{d}g)$, $\varphi \in \mathscr{S}$

Table 3.2 Comparison of notions of finite-dimensional statistical calculus with the theory of generalized stochastic processes. See Sections 3.4 for an explanation.

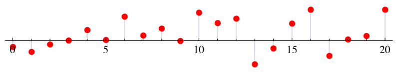
$\mathcal{S}$: Schwartz' space of smooth (infinitely differentiable) and rapidly decaying functions

$\mathcal{S}'$: Schwartz' space of tempered distributions (generalized functions)

Defining Gaussian noise: discrete vs. continuous

Lévy exponent: $\log \hat{p}_X(\omega) = f(\omega) = -\frac{1}{2}\omega^2$

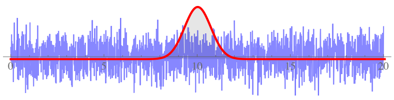
- **Discrete white Gaussian noise**



$X = (X_1, \dots, X_N)$ with X_n i.i.d standardized Gaussian

Characteristic function: $\hat{p}_X(\boldsymbol{\omega}) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle \boldsymbol{\omega}, X\rangle}) = \exp\left(\sum_{n=1}^{N} f(\omega_n)\right) = \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{\omega}\|^2}$

- **Continuous-domain white Gaussian noise**



Infinite-dimensional entity w with generic observations $X_n = \langle \varphi_n, w\rangle$

Characteristic functional: $\widehat{\mathscr{P}_w}(\varphi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle \varphi, w\rangle}\} = \mathrm{e}^{-\frac{1}{2}\|\varphi\|_{L_2}^2} = \exp\left(\int_{\mathbb{R}} f(\varphi(x))\mathrm{d}x\right)$

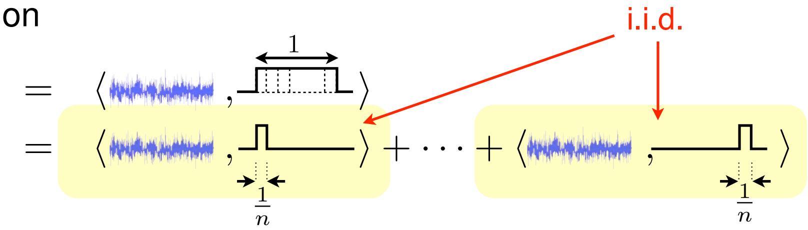
$\hat{p}_{X_n}(\omega) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega\langle \varphi_n, w\rangle}\} = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle \omega\varphi_n, w\rangle}\} = \widehat{\mathscr{P}_w}(\omega\varphi_n) = \mathrm{e}^{-\frac{1}{2}\omega^2\|\varphi_n\|_{L_2}^2}$

Infinite divisibility and Lévy exponents

Definition: A random variable X with generic pdf $p_{\mathrm{id}}(x)$ is **infinitely divisible** (id) iff., for any $N \in \mathbb{Z}^+$, there exist i.i.d. random variables $X_1, \dots, X_N$ such that $X \overset{\mathrm{d}}{=} X_1 + \cdots + X_N$.

- **Rectangular test function**

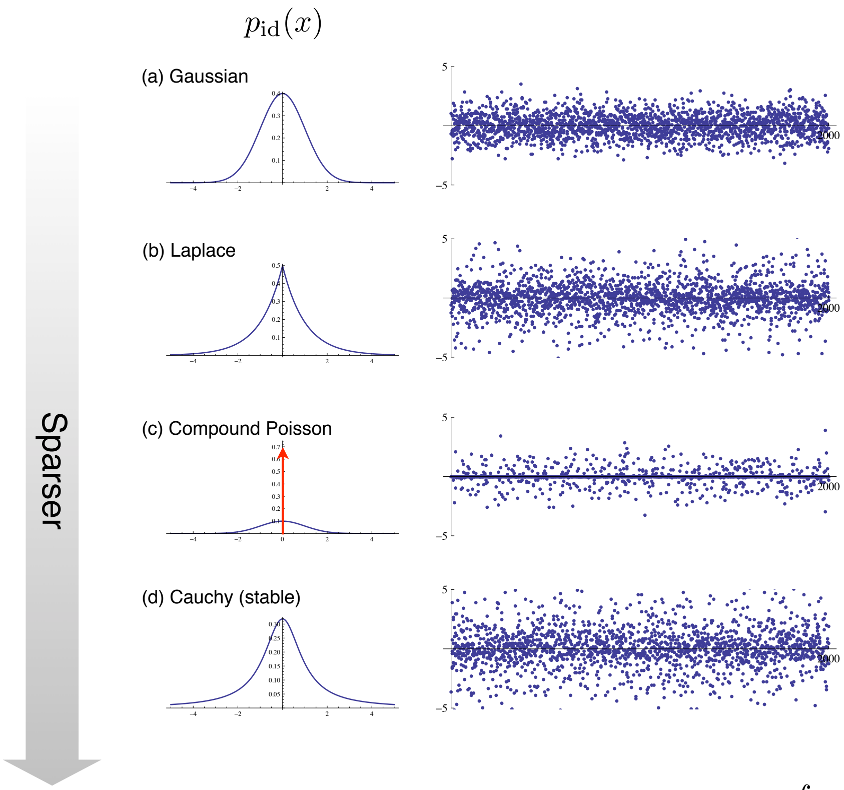
$X_{\mathrm{id}} = \langle w, \mathrm{rect}\rangle$



Proposition
The random variable $X_{\mathrm{id}} = \langle w, \mathrm{rect}\rangle$ where w is a generalized innovation process is infinitely divisible. It is uniquely characterized by its **Lévy exponent** $f(\omega) = \log \hat{p}_{\mathrm{id}}(\omega)$.

Bottom line: There is a one-to-one correspondence between Lévy exponents and infinitely divisible distributions and, by extension, innovation processes.

Examples of infinitely divisible laws



$$p_{\mathrm{Gauss}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{x^2}{2\sigma^2}}$$

$$p_{\mathrm{Laplace}}(x) = \frac{\lambda}{2} \mathrm{e}^{-\lambda|x|}$$

$$p_{\mathrm{Poisson}}(x) = \mathcal{F}^{-1}\{\mathrm{e}^{\lambda(\hat{p}_A(\omega)-1)}\}$$

$$p_{\mathrm{Cauchy}}(x) = \frac{1}{\pi\,(x^2+1)}$$

Characteristic function: $\widehat{p}_{\mathrm{id}}(\omega) = \int_{\mathbb{R}} p_{\mathrm{id}}(x) \mathrm{e}^{\mathrm{j}\omega x} \mathrm{d}x = \mathrm{e}^{f(\omega)}$

Canonical Lévy-Khintchine representation

Definition

A (positive) measure μ_v on $\mathbb{R}\backslash\{0\}$ is called a ***Lévy measure*** if it satisfies

$$\int_{\mathbb{R}} \min(a^2, 1)\mu_v(\mathrm{d}a) = \int_{\mathbb{R}} \min(a^2, 1) v(a) \mathrm{d}a < \infty.$$

The corresponding ***Lévy density*** $v : \mathbb{R} \to \mathbb{R}^+$ is such that $\mu_v(\mathrm{d}a) = v(a)\mathrm{d}a$.

Theorem (Lévy-Khintchine)

A probability distribution p_{id} is **infinitely divisible** (id) iff. its characteristic function can be written as

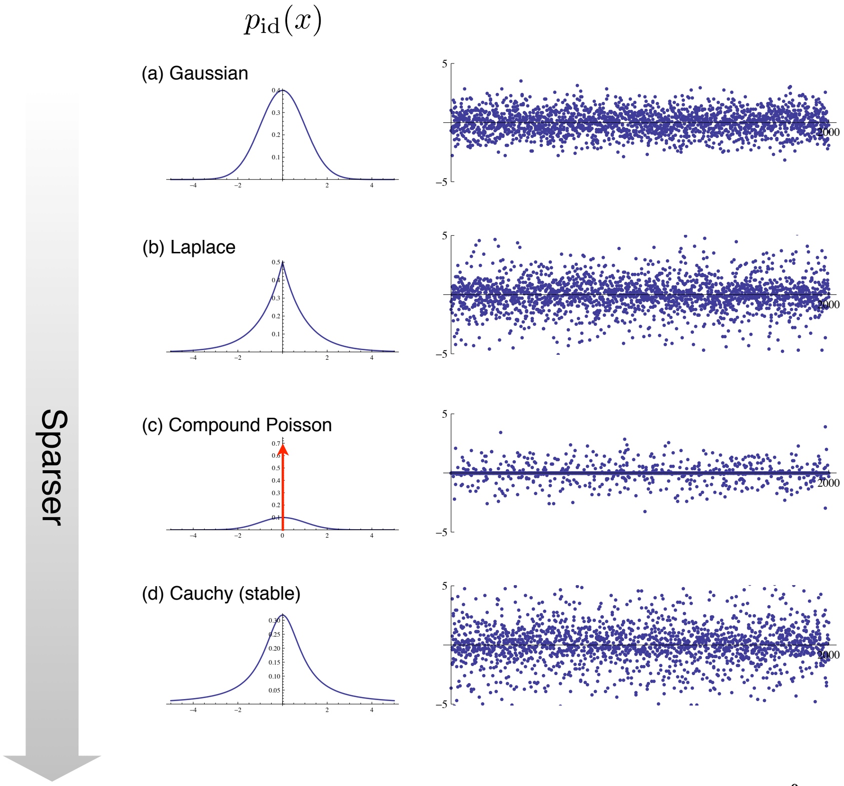
$$\widehat{p}_{\mathrm{id}}(\omega) = \int_{\mathbb{R}} p_{\mathrm{id}}(x) \mathrm{e}^{\mathrm{j}\omega x} \mathrm{d}x = \exp\left(f(\omega)\right)$$

with

$$f(\omega) = \log \widehat{p}_{\mathrm{id}}(\omega) = \mathrm{j}b_1'\omega - \frac{b_2\omega^2}{2} + \int_{\mathbb{R}\backslash\{0\}} \left(\mathrm{e}^{\mathrm{j}a\omega} - 1 - \mathrm{j}a\omega \mathbb{1}_{|a|<1}(a)\right) v(a)\mathrm{d}a$$

where $b_1' \in \mathbb{R}$ and $b_2 \in \mathbb{R}^+$ are some arbitrary constants, and where v is an admissible Lévy density. The function f is called the **Lévy exponent** of p_{id}.

Examples of infinitely divisible laws



(a) $f(\omega) = -\frac{\sigma_0^2}{2}|\omega|^2$

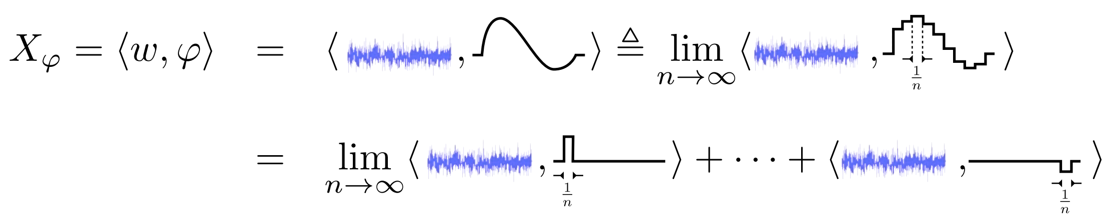
(b) $f(\omega) = \log\left(\frac{1}{1+\omega^2}\right)$

(c) $f(\omega) = \lambda \int_{\mathbb{R}} (\mathrm{e}^{\mathrm{j}x\omega} - 1) p_A(x) \mathrm{d}x$

(d) $f(\omega) = -s_0|\omega|$

Characteristic function: $\widehat{p}_{\mathrm{id}}(\omega) = \int_{\mathbb{R}} p_{\mathrm{id}}(x) \mathrm{e}^{\mathrm{j}\omega x} \mathrm{d}x = \mathrm{e}^{f(\omega)}$

Characterization of generalized innovation



Theorem

Let w be a generalized stochastic process such that $X_{\mathrm{id}} = \langle w, \mathrm{rect}\rangle$ is well-defined. Then, w is a generalized innovation (white noise) in $\mathcal{S}'(\mathbb{R}^d)$ **if and only if** its characteristic form is given by

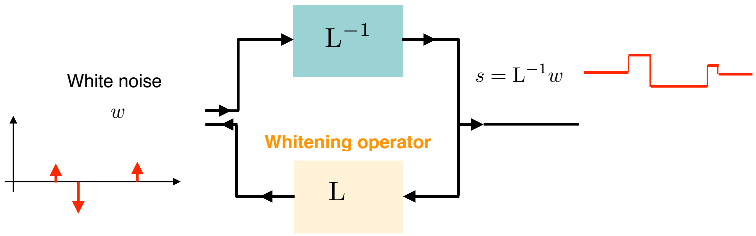
$$\widehat{\mathscr{P}}_w(\varphi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle w,\varphi\rangle}\} = \exp\left(\int_{\mathbb{R}^d} f(\varphi(\boldsymbol{r}))\mathrm{d}\boldsymbol{r}\right)$$

where $f(\omega)$ is a **valid Lévy exponent** (in fact, the Lévy exponent of X_{id}).
Moreover, the random variables $X_\varphi = \langle w, \varphi\rangle$ are all **infinitely divisible** with modified Lévy exponent

$$f_\varphi(\omega) = \int_{\mathbb{R}^d} f(\omega\varphi(\boldsymbol{r}))\mathrm{d}\boldsymbol{r}$$

(Gelfand-Vilenkin 1964; Amini-U. IEEE-IT 2014)

Steps 2 + 3: Characterization of sparse process



- Abstract formulation of innovation model

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

$$\Rightarrow \quad \widehat{\mathscr{P}}_s(\varphi) = \mathbb{E}\{\mathrm{e}^{j\langle s,\varphi\rangle}\} = \widehat{\mathscr{P}}_w(\mathrm{L}^{-1*}\varphi) = \exp\left(\int_{\mathbb{R}^d} f\big(\mathrm{L}^{-1*}\varphi(\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x}\right)$$

Sufficient condition for existence:
L^{-1*} continuous operator: $\mathcal{S}(\mathbb{R}^d) \to L_p(\mathbb{R}^d)$

(U.-Tafti-Sun, IEEE-IT 2014)

☒ Probability laws of sparse processes are id

- Analysis: go back to **innovation process**: $w = \mathrm{L}s$
  - Generic random observation: $X = \langle \varphi, w\rangle$ with $\varphi \in \mathcal{S}(\mathbb{R}^d)$ or $\varphi \in L_p(\mathbb{R}^d)$ (by extension)
  - Linear functional: $Y = \langle \psi, s\rangle = \langle \psi, \mathrm{L}^{-1}w\rangle = \langle \overbrace{\mathrm{L}^{-1*}\psi}, \ w\rangle$

  If $\phi = \mathrm{L}^{-1*}\psi \in L_p(\mathbb{R}^d)$ then $Y = \langle \psi, s\rangle = \langle \phi, w\rangle$ is ***infinitely divisible***
  with Lévy exponent $f_\phi(\omega) = \int_{\mathbb{R}^d} f\big(\omega\phi(\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x}$

$$\Rightarrow \quad p_Y(y) = \mathcal{F}^{-1}\{\mathrm{e}^{f_\phi(\omega)}\}(y) = \int_{\mathbb{R}} \mathrm{e}^{f_\phi(\omega) - \mathrm{j}\omega y}\frac{\mathrm{d}\omega}{2\pi}$$

= explicit form of pdf

Unser and Tafti | An Introduction to Sparse Stochastic Processes | Cambridge

Operators: fundamental invariance properties

Definition
An operator T is **shift-invariant** iff., for any function φ in its domain and any $\boldsymbol{r}_0 \in \mathbb{R}^d$,

$$\mathrm{T}\{\varphi(\cdot - \boldsymbol{r}_0)\}(\boldsymbol{r}) = \mathrm{T}\{\varphi\}(\boldsymbol{r} - \boldsymbol{r}_0).$$

Definition
An operator T is **scale-invariant** of order γ iff., for any function φ in its domain,

$$\mathrm{T}\{\varphi\}(\boldsymbol{r}/a) = |a|^\gamma \mathrm{T}\{\varphi(\cdot/a)\}(\boldsymbol{r}),$$

where $a \in \mathbb{R}^+$ is the dilation factor.

$$\mathrm{T}\{\varphi(a\cdot)\}(\boldsymbol{r}) = |a|^\gamma \mathrm{T}\{\varphi\}(a\boldsymbol{r})$$

Definition
An operator T is scalarly **rotation-invariant** iff., for any function φ in its domain,

$$\mathrm{T}\{\varphi\}(\mathbf{R}^T\boldsymbol{r}) = \mathrm{T}\{\varphi(\mathbf{R}^T\cdot)\}(\boldsymbol{r}),$$

where $\mathbf{R}$ is any orthogonal matrix in $\mathbb{R}^{d\times d}$.

Fractional-order operators

- Liouville's fractional derivative

$$\mathrm{D}^\gamma \varphi(r) = \int_{\mathbb{R}} (\mathrm{j}\omega)^\gamma \hat{\varphi}(\omega) \mathrm{e}^{\mathrm{j}\omega r} \frac{\mathrm{d}\omega}{2\pi}$$

Proposition [U.-Blu, 2007]
The complete family of 1-D scale-invariant convolution operators of order $\gamma \in \mathbb{R}$ reduces to the fractional derivative ∂_τ^γ whose Fourier-based definition is

$$\partial_\tau^\gamma \varphi(r) = \int_{\mathbb{R}} (\mathrm{j}\omega)^{\frac{\gamma}{2}+\tau} (-\mathrm{j}\omega)^{\frac{\gamma}{2}-\tau} \hat{\varphi}(\omega) \mathrm{e}^{\mathrm{j}\omega r} \frac{\mathrm{d}\omega}{2\pi}$$

Order of differentiation: γ Phase factor: $\tau \in \mathbb{R}$

- Semi-group property

$$\partial_\tau^\gamma \partial_{\tau'}^{\gamma'} = \partial_{\tau+\tau'}^{\gamma+\gamma'}, \qquad \text{for } \gamma', \gamma+\gamma' \in (-1, +\infty) \text{ and } \tau, \tau' \in \mathbb{R}$$

Special cases: $\mathrm{D}^\gamma = \partial_{\gamma/2}^\gamma$, $\mathcal{H}_\tau = \partial_\tau^0$ (fractional Hilbert transform)

Invariance properties: definitions

for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$

Translation of $\phi \in \mathcal{S}'(\mathbb{R}^d)$ by $\boldsymbol{r}_0$: $\langle \varphi, \phi(\cdot - \boldsymbol{r}_0)\rangle = \langle \varphi(\cdot + \boldsymbol{r}_0), \phi\rangle$

A generalized stochastic process s is **stationary** if it has the same probability laws as its translated version $s(\cdot - \boldsymbol{r}_0)$ for any $\boldsymbol{r}_0 \in \mathbb{R}^d$.

$$\Leftrightarrow \qquad \widehat{\mathscr{P}}_s(\varphi) = \widehat{\mathscr{P}}_s\big(\varphi(\cdot + \boldsymbol{r}_0)\big)$$

Affine transformation of $\phi \in \mathcal{S}'(\mathbb{R}^d)$: $\langle \varphi, \phi(\mathbf{T}^{-1}\cdot)\rangle = |\det(\mathbf{T})|\,\langle \varphi(\mathbf{T}\cdot), \phi\rangle$

A generalized stochastic process s is **isotropic** if it has the same probability laws as its rotated version $s(\mathbf{R}^T\cdot)$ for any $(d \times d)$ rotation matrix $\mathbf{R}$.

$$\Leftrightarrow \qquad \widehat{\mathscr{P}}_s(\varphi) = \widehat{\mathscr{P}}_s\big(\varphi(\mathbf{R}\cdot)\big)$$

A generalized stochastic process s is **self-similar** of scaling order H if it has the same probability laws as any of its scaled and renormalized version $a^H s(\cdot/a)$.

$$\Leftrightarrow \qquad \widehat{\mathscr{P}}_s(\varphi) = \widehat{\mathscr{P}}_s\big(a^{H+d}\varphi(a\cdot)\big)$$

Duality relation: $\langle \varphi, a^H s(\cdot/a)\rangle = \langle a^{H+d}\varphi(a\cdot), s\rangle$

H: Hurst exponent

Invariance properties of innovation model

Theorem

The high-level statistical properties of $s = \mathrm{L}^{-1}w$ are tightly linked to the invariance properties of L^{-1} (or, equivalently, L^{-1*}) described by its generalized impulse response

$$h(\cdot, \boldsymbol{r}') = \mathrm{L}^{-1}\{\delta(\cdot - \boldsymbol{r}')\} \in \mathcal{S}'(\mathbb{R}^d \times \mathbb{R}^d).$$

1. If L^{-1} is **linear shift-invariant**, then s is **stationary** and $h(\boldsymbol{r}, \boldsymbol{r}') = h(\boldsymbol{r} - \boldsymbol{r}', 0) = \rho_{\mathrm{L}}(\boldsymbol{r} - \boldsymbol{r}')$ where $\rho_{\mathrm{L}} = \mathrm{L}^{-1}\{\delta\}$ is the Green's function of L.
2. If L^{-1} is **translation- and rotation-invariant**, then s is **stationary isotropic** and $h(\boldsymbol{r}, \boldsymbol{r}') = \rho_{\mathrm{L}}(|\boldsymbol{r} - \boldsymbol{r}'|)$ where $\rho_{\mathrm{L}}(|\boldsymbol{r}|) = \mathrm{L}^{-1}\{\delta\}(\boldsymbol{r})$ is a purely radial function.
3. If L^{-1*} is **scale-invariant** of order $(-\gamma)$ and $\sigma_w^2 = -f''(0) < \infty$, then s is **wide-sense self-similar** with Hurst exponent $H = \gamma - d/2$.
4. If L^{-1*} is **scale-invariant** of order $(-\gamma)$ and f is homogeneous of degree $0 < \alpha \le 2$, then s is **self-similar** with Hurst exponent $H = \gamma - d + d/\alpha$.

7.4 Lévy processes and extensions

Classical definition

The stochastic process $W = \{W(t) : t \in \mathbb{R}^+\}$ is a Lévy process if it fulfills the following requirements:

1. $W(0) = 0$ almost surely.
2. Given $0 \le t_1 < t_2 < \ldots < t_n$, the increments $W(t_2) - W(t_1)$, $W(t_3) - W(t_2)$, $\ldots$, $W(t_n) - W(t_{n-1})$ are mutually independent.
3. For any given step T, the increment process $\delta_T W(t)$, where δ_T is the operator that associates $W(t)$ to $\big(W(t) - W(t-T)\big)$, is stationary.

- Equivalent generalized process: solution of unstable SDE

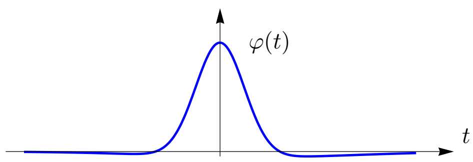
$\mathrm{D}W = \dot{W} = w$ subject to boundary condition $W(0) = 0$

$$\Rightarrow \quad W(t) = \langle \mathbb{1}_{(0,t]}, w\rangle = \int_0^t w(\tau)\mathrm{d}\tau = \int_0^t \mathrm{d}W(\tau)$$

Stabilizing the anti-derivative operator

D: scale-invariant operator with $\gamma = 1$

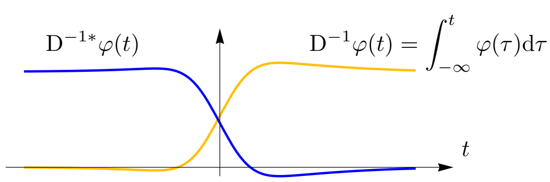
... but the system is no longer BIBO stable



Adjoint inverse operator (LSI):

$$\mathrm{D}^{-1*}\varphi(t) = \int_t^{+\infty} \varphi(\tau)\mathrm{d}\tau = (\mathbb{1}_+^\vee * \varphi)(t)$$

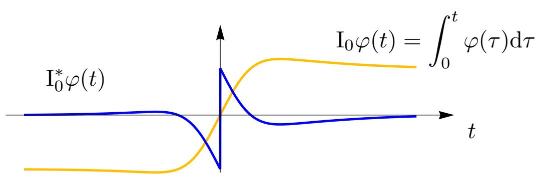
$\notin L_p(\mathbb{R})$



Modified anti-derivative operators:

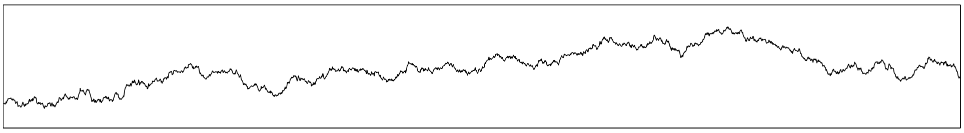
$$\mathrm{I}_0^*\varphi(t) = \mathrm{D}^{-1*}\varphi(t) - (\mathrm{D}^{-1*}\varphi)(-\infty)\ \mathbb{1}_+^\vee(t)$$

I_0^*: continuous operator $\mathcal{S}(\mathbb{R}) \to \mathcal{R}(\mathbb{R})$

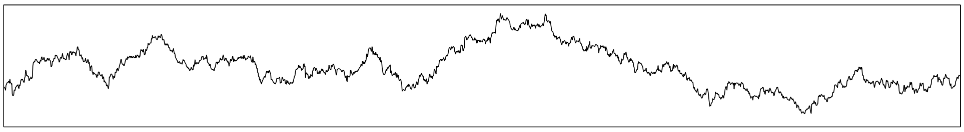


I_0: imposes vanishing boundary condition at $t = 0$

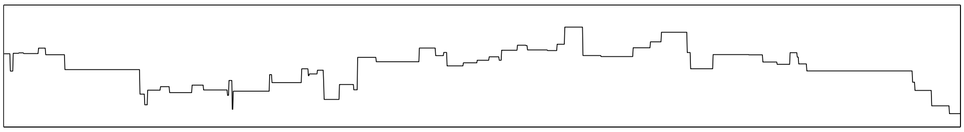
From Brownian motion to Lévy flights



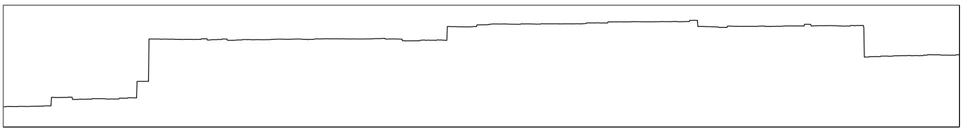
(a): Gaussian



(b): Laplace

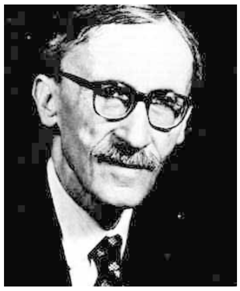


(c): Compound Poisson



(d): Cauchy

Nobert Wiener



Paul Lévy

Ordinary differential systems

- First-order operator: $\mathrm{P}_\alpha = \mathrm{D} - \alpha\mathrm{Id}$ with $\mathrm{Re}(\alpha) \neq 0$

$$\rho_\alpha(r) = \mathcal{F}^{-1}\left\{\frac{1}{\mathrm{j}\omega - \alpha}\right\}(r) = \begin{cases} \mathrm{e}^{\alpha r}\mathbb{1}_{[0,\infty)}(r) & \text{if } \mathrm{Re}(\alpha) < 0, \\ -\mathrm{e}^{\alpha r}\mathbb{1}_{(-\infty,0]}(r) & \text{if } \mathrm{Re}(\alpha) > 0. \end{cases}$$

Inverse operator: $\mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$

$$\mathrm{P}_\alpha^{-1}\varphi = \rho_\alpha * \varphi$$

Adjoint: $\mathrm{P}_\alpha^* = -\mathrm{P}_{-\alpha}$

$$\Rightarrow \mathrm{P}_\alpha^{-1*}\varphi = -\rho_{-\alpha} * \varphi = \rho_\alpha^\vee * \varphi$$

- Higher-order operators with $\mathrm{Re}(\alpha_n) \neq 0$ and $N > M$

$$\mathrm{P}_{\alpha_1} \cdots \mathrm{P}_{\alpha_N}\{s\}(r) = q_M(\mathrm{D})\{w\}(r)$$

Inverse operator $\mathrm{L}^{-1} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$

$$\mathrm{L}^{-1*} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$$

$$\mathrm{L}^{-1} = \mathrm{P}_{\alpha_N}^{-1} \cdots \mathrm{P}_{\alpha_1}^{-1}\, q_M(\mathrm{D})$$

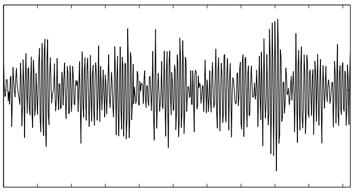
$\mathrm{L}^{-1}\varphi = \rho_\mathrm{L} * \varphi$ with $\rho_\mathrm{L} \in \mathcal{R}(\mathbb{R})$ (exponential decay)

Application: signal modeling (Audio)

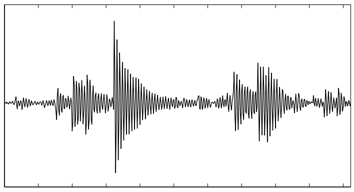
- Sparse, bandpass processes

$$\mathrm{L} = \frac{\mathrm{d}^n}{\mathrm{d}t^n} + a_{n-1}\frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}} + \cdots + a_1\frac{\mathrm{d}}{\mathrm{d}t} + a_0\mathrm{I}$$

poles = [−.05 + jπ/2, −.05 − jπ/2], zeros = []



(a) Gaussian



(b) Alpha stable **α**=1.2

- Mixed sparse processes: $s_{\mathrm{mix}} = s_1 + \cdots + s_M$

$$\widehat{\mathscr{P}}_{s_{\mathrm{mix}}}(\varphi) = \prod_{m=1}^{M} \widehat{\mathscr{P}}_{s_m}(\varphi) = \exp\left(\int_{\mathbb{R}} \sum_{m=1}^{M} f_m\left(\mathrm{L}_m^{-1*}\varphi(t)\right)\mathrm{d}t\right)$$

Gaussian (Am)

generalized Lévy (Am, S**α**S)

(f)Brownian motion revisited

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

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

L_2-stable anti-derivative: $\mathrm{I}_0^*\varphi(t) = \int_{\mathbb{R}} \frac{\hat{\varphi}(\omega) - \hat{\varphi}(0)}{-\mathrm{j}\omega} \mathrm{e}^{\mathrm{j}\omega t} \frac{\mathrm{d}\omega}{2\pi}$

- Characteristic form of Brownian motion (a.k.a. Wiener process)

Stabilization ⇔ non-stationary behavior

$$\widehat{\mathscr{P}}_W(\varphi) = \exp\left(-\frac{1}{2}\|\mathrm{I}_0^*\varphi\|_{L_2}^2\right)$$

$$= \exp\left(-\frac{1}{2}\int_{\mathbb{R}} \left|\frac{\hat{\varphi}(\omega) - \hat{\varphi}(0)}{-\mathrm{j}\omega}\right|^2 \frac{\mathrm{d}\omega}{2\pi}\right) \qquad \text{(by Parseval)}$$

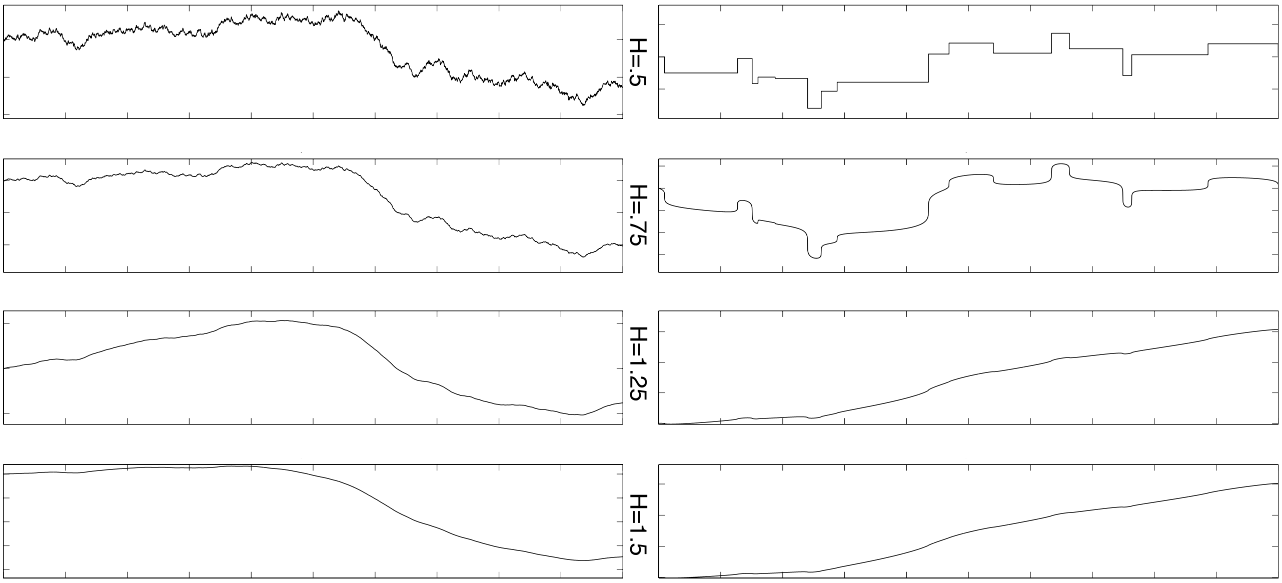
- Characteristic form of fractional Brownian motion

$$\widehat{\mathscr{P}}_s(\varphi) = \exp\left(-\frac{1}{2}\int_{\mathbb{R}} \left|\frac{\hat{\varphi}(\omega) - \hat{\varphi}(0)}{|\omega|^{\gamma}}\right|^2 \frac{\mathrm{d}\omega}{2\pi}\right)$$

(Blu-U., *IEEE-SP* 2007)

Example in 1D: Self-similar processes

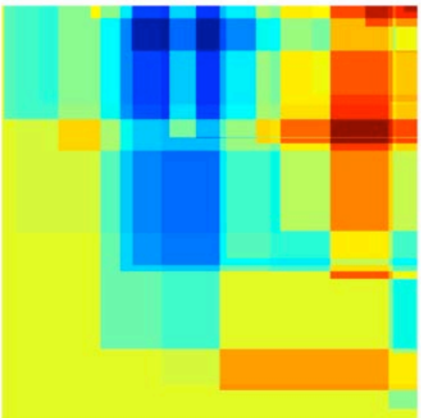
$$\mathrm{L} \quad \stackrel{\mathcal{F}}{\longleftrightarrow} \quad (j\omega)^{H+\frac{1}{2}} \quad \Rightarrow \quad \mathrm{L}^{-1}\text{: fractional integrator}$$



Gaussian | **Sparse (generalized Poisson)**

Fractional Brownian motion (Mandelbrot, 1968) | (U.-Tafti, *IEEE-SP* 2010)

Equations of a screen saver



Mondrian process

$$\frac{\partial^2 s(\boldsymbol{x})}{\partial x_1 \partial x_2} = \sum_k A_k \delta(\boldsymbol{x} - \boldsymbol{x}_k)$$

$$\Rightarrow \qquad s(\boldsymbol{x}) = a_0 + \sum_k A_k (\boldsymbol{x} - \boldsymbol{x}_k)_+^0$$

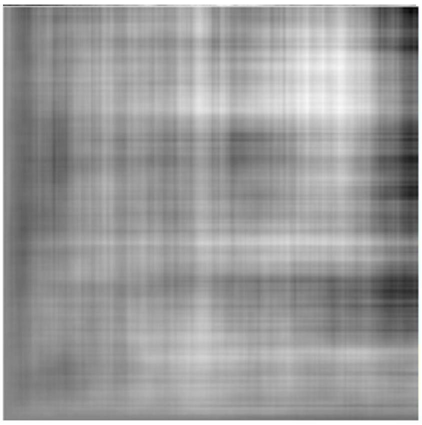
where $\boldsymbol{x}_k$ are Poisson distributed with rate λ
and A_k i.i.d. Gaussian with characteristic function $\hat{p}_A$.

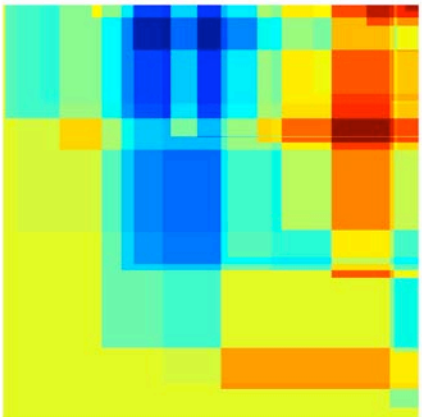
Complete mathematical description (characteristic form)

$\forall \varphi \in \mathcal{S}(\mathbb{R}^2)$ (Schwartz's space of smooth and rapidly-decaying test functions):

$$\mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle s, \varphi\rangle}\} = \exp\left(\lambda \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{p}_A\left(\int_{x_1}^{\infty}\int_{x_2}^{\infty} \varphi(x_1', x_2')\mathrm{d}x_1'\mathrm{d}x_2'\right)\mathrm{d}x_1\mathrm{d}x_2 - \lambda\right)$$

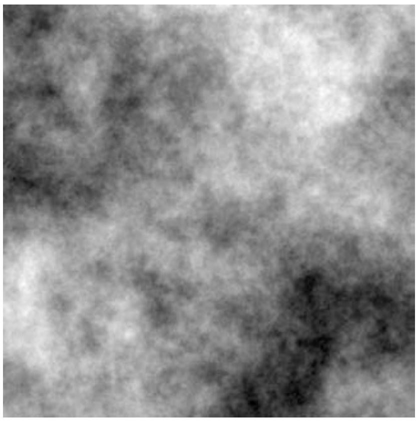
with $\hat{p}_A(\omega) = e^{-\frac{\omega^2}{2}}$





Gaussian

$$\mathrm{L} = \mathrm{D}_{r_1}\mathrm{D}_{r_2} \quad \stackrel{\mathcal{F}}{\longleftrightarrow} \quad (\mathrm{j}\omega_1)(\mathrm{j}\omega_2)$$

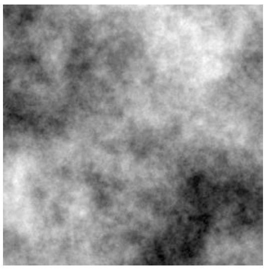


$$\mathrm{L} = (-\Delta)^{1/2} \quad \stackrel{\mathcal{F}}{\longleftrightarrow} \quad \|\boldsymbol{\omega}\|$$

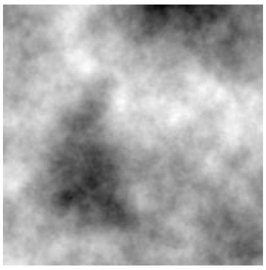
Scale- and rotation-invariant processes

Stochastic partial differential equation : $(-\Delta)^{\frac{H+1}{2}} s(\boldsymbol{x}) = w(\boldsymbol{x})$

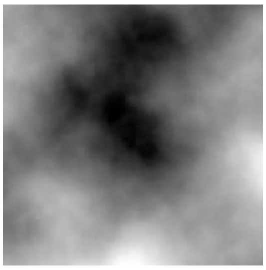
Gaussian



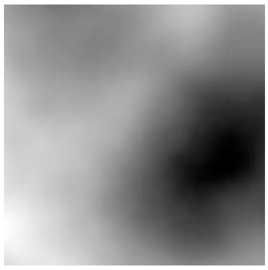
H=.5



H=.75

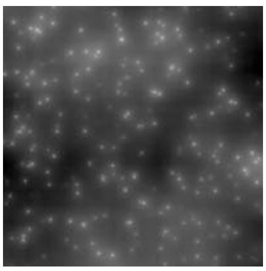


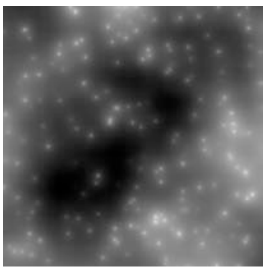
H=1.25

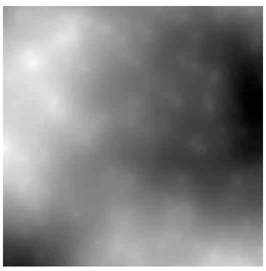


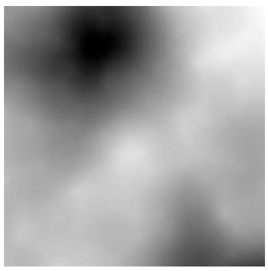
H=1.75

Sparse (generalized Poisson)



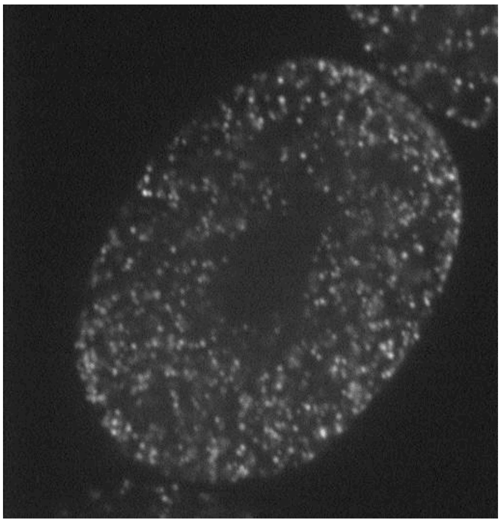






(U.-Tafti, *IEEE-SP* 2010)

Powers of ten: from astronomy to biology



2.1 DECOUPLING OF SPARSE

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

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

Discrete approximation of whitening operator: $\mathrm{L_d}$

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

Discrete increment process:

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

Generalized B-spline:

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

A-to-D translator

Decoupling: Wavelet analysis

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

Generalized operator-like wavelets:

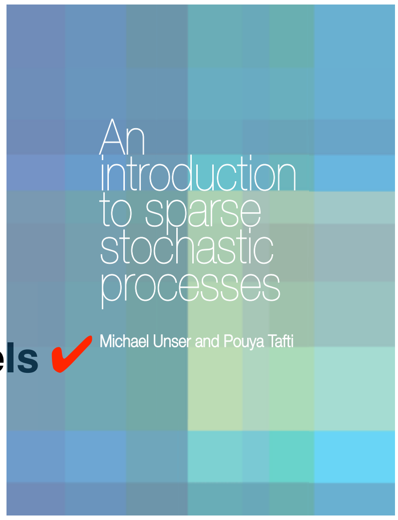
$$\psi_i(\boldsymbol{x}) = \mathrm{L}^*\phi_i(\boldsymbol{x})$$

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

Operator-like wavelet analysis of sparse process:

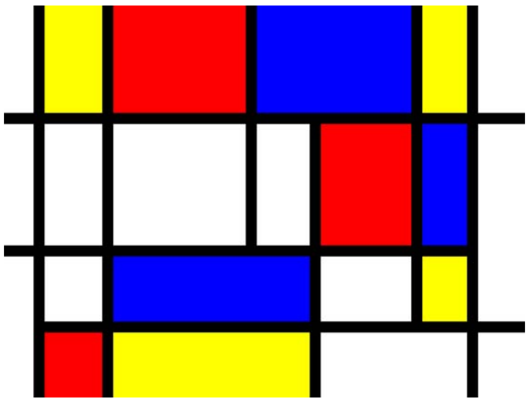
$$\begin{aligned}\langle \psi_i(\cdot - \boldsymbol{x}_0), s\rangle &= \langle \mathrm{L}^*\phi_i(\cdot - \boldsymbol{x}_0), s\rangle \\ &= \langle \phi_i(\cdot - \boldsymbol{x}_0), \mathrm{L}s\rangle \\ &= \langle \underbrace{\phi_i(\cdot - \boldsymbol{x}_0)}_{\varphi}, w\rangle = (\phi_i^{\vee} * w)(\boldsymbol{x}_0)\end{aligned}$$

Table of content



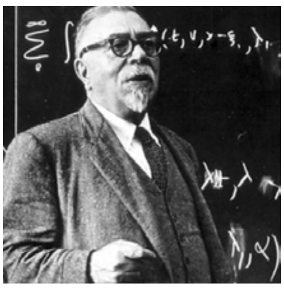
1. **Introduction** ✔
2. **Roadmap to the monograph** ✔
3. *Mathematical context and background*
4. **Continuous-domain innovation models** ✔
5. Operators and their inverses
6. Splines and wavelets
7. **Sparse stochastic processes** ✔
8. **Sparse representations**
9. Infinite divisibility and transform-domain statistic
10. **Recovery of sparse signals**
11. Wavelet-domain methods

Gaussian vs. Sparse



Fourier analysis

Wavelet analysis



Norbert Wiener

Paul Lévy