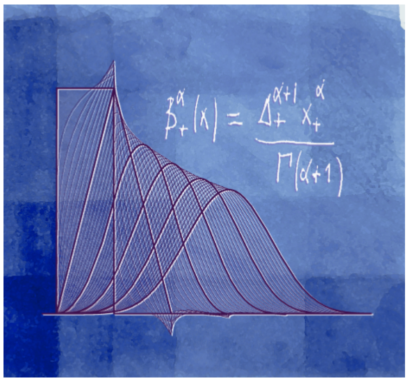
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Towards a theory of sparse stochastic processes: when Paul Lévy joins forces with Nobert Wiener

Michael Unser
Biomedical Imaging Group
EPFL, Lausanne, Switzerland

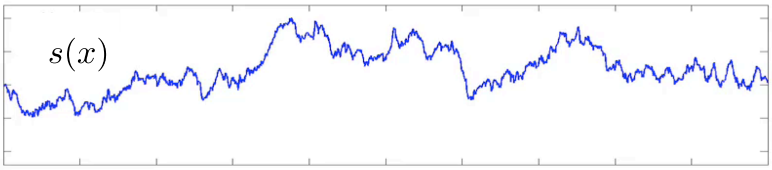
Joint work with P. Tafti, Q. Sun, T. Blu, M. Guerquin-Kern, E. Bostan, etc.



Mathematics and Image Analysis (MIA'12), Paris, 16-18 January 2012.

Brownian motion (a.k.a. Wiener process)

Mathematical construction by Wiener in 1923



- Gaussian process
- Non-stationary
- Self-similar: "$1/\omega$" spectral decay
- Independent increments

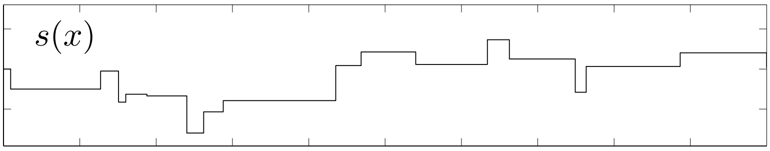
  $u[k] = s(k) - s(k-1)$: i.i.d. Gaussian

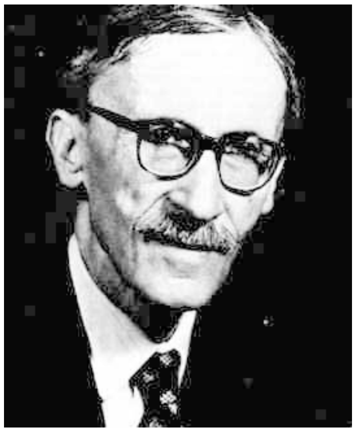
In-depth analysis of sample path properties in the 1930's by Paul Lévy

- Haar-related expansion whose coefficients are i.i.d. Gaussian

Lévy process

Constructed by Paul Lévy in the 1930's





- Non-Gaussian generalization of Wiener process
- Non-stationary
- Self-similar: "$1/\omega$" spectral decay
- Independent increments

  $u[k] = s(k) - s(k-1)$: i.i.d. infinitely divisible (heavy tailed)

Example: compound Poisson process (piecewise-constant, with random jumps)

$\Rightarrow$ Archetype of a "sparse" random signal

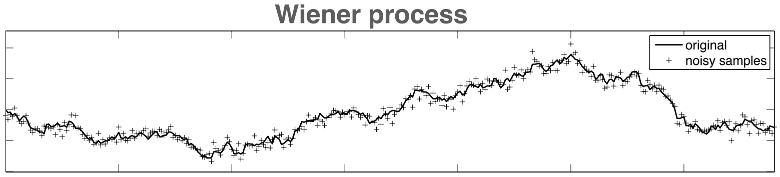
Simple denoising experiment

- Measurement model

  $$g[k] = s(k) + n[k]$$

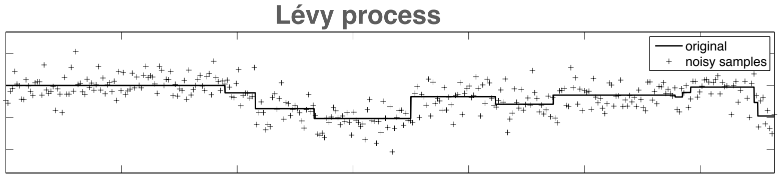
  - $s(x)$: Continuously-defined process
  - $n[k]$: Discrete white Gaussian noise

Wiener process



Brownian motion (Gaussian)

Lévy process



Compound Poisson process (Sparse)

- Controlled experiment
  - Matched 2nd-order statistics (correlation function)
  - Generalized spectrum $\sim \frac{1}{\omega}$

Three dominant paradigms

- Wiener solution (LMMSE) = Smoothing spline

$$\begin{aligned}\tilde{s}_{\text{spline}}(x) &= \arg\min_{s(x)}\left\{\sum_{k\in\mathbb{Z}}|g[k]-s(k)|^2+\mu\int_{\mathbb{R}}|\mathrm{D}s(x)|^2\mathrm{d}x\right\}\\ &= \arg\min_{s(x)}\left\{\|g-s\|_{\ell_2}^2+\mu\|\mathrm{D}s\|_{L_2(\mathbb{R})}^2\right\}\end{aligned}$$

Theoretical result: Continuous-domain LMMSE = piecewise-linear smoothing spline
Also provides MMSE and MAP for Brownian motion model (Blu-U., 2005)

- Wavelet solution = sparse signal recovery

$$\tilde{s}_{\text{wave}}(x)=\arg\min_{s(x)}\left\{\|g-s\|_{\ell_2}^2+\sum_i \mu_i\|w_i\|_{\ell_p}^p\right\}$$

with $w_i[k]=\langle s, 2^{i/2}\psi(x/2^i-k)\rangle_{L_2(\mathbb{R})}$

Theoretical result: Simple wavelet shrinkage algorithm (LASSO: Tibshirani, 1996)

Three dominant paradigms (cont'd)

- Total variation = non-quadratic regularization

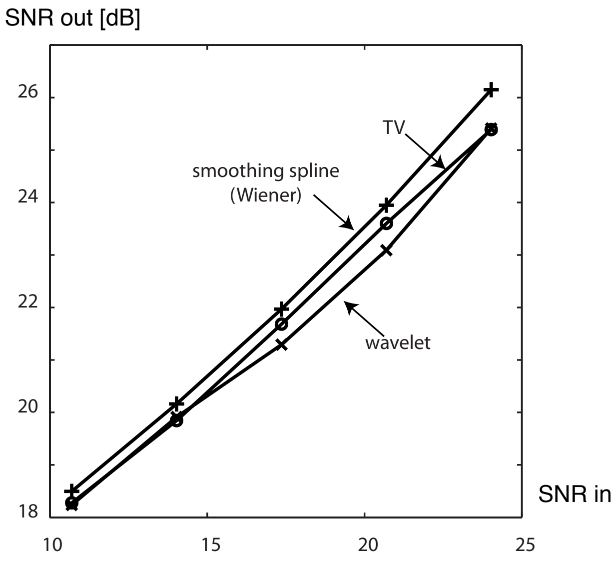
$$\tilde{s}_{\text{TV}}(x)=\arg\min_{s(x)}\left\{\|g-s\|_{\ell_2}^2+\mu\,\mathrm{TV}(s)\right\}$$

Theoretical result 1 [Mammen, *Annals of Statistics*, 1997]
Piecewise-constant spline with adaptive knots is a global minimizer

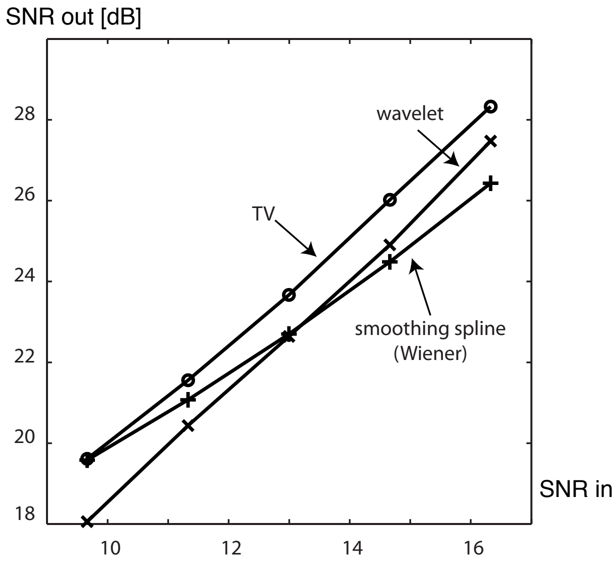
$$\tilde{s}_{\text{MAP}}(k)=\arg\min_{s}\left\{\|g-s\|_{\ell_2}^2+\mu\sum_k|s(k)-s(k-1)|\right\}$$

Theoretical result 2 [Unser et al., *IEEE Trans. Sig. Proc.* 2010]
MAP solution for specific Lévy process (continuous-domain model)

Denoising results



Wiener process (Gaussian)

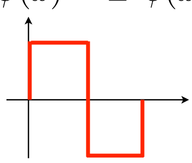


Poisson process (Sparse)

Commonalities $\Rightarrow$ innovation modeling

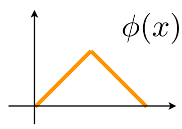
- Central role of derivative operator
  - Quadratic spline energy: $\|\mathrm{D}s\|^2_{L_2(\mathbb{R})}$
  - TV as an L_1 norm: $s \in W_1^1 \quad \Leftrightarrow \quad \mathrm{TV}(s) = \|\mathrm{D}s\|_{L_1(\mathbb{R})}$
  - Wavelet as a smoothed derivative: $\psi_{\mathrm{Haar}}(x) = \mathrm{D}\phi(x)$

$\psi(x) = \mathrm{L}^*\phi(x)$



$\mathrm{L}^* = \frac{\mathrm{d}}{\mathrm{d}x}$

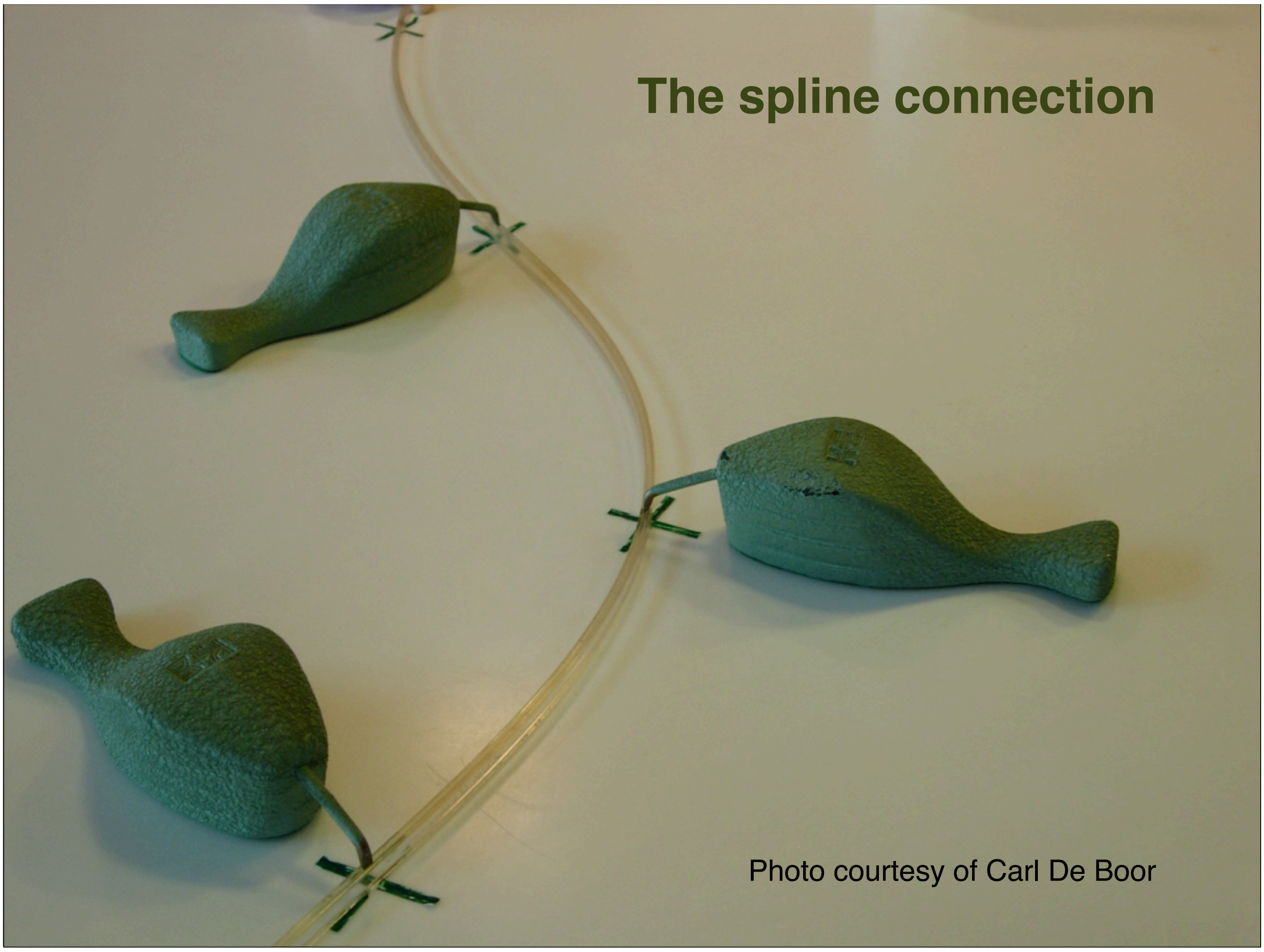
$\phi(x)$



$\Rightarrow \langle f, \psi(\cdot - x_0)\rangle = \mathrm{L}(f * \phi^*)(x_0) = -\frac{\mathrm{d}}{\mathrm{d}x}\,(f*$  $)\,(x_0)$

OUTLINE

- Gaussian (Wiener) vs. sparse (Lévy) signals ✔
- The spline connection
  - *L*-splines and signals with finite rate of innovation
  - Green functions as elementary building blocks
- Sparse stochastic processes
  - Generalized innovation model
  - Gelfand's theory of generalized stochastic processes
  - Statistical characterization of sparse stochastic processes
- Implications of innovation model
  - Link with regularization
  - Wavelet representation of sparse processes
  - Determination of transform-domain statistics
- Sparse processes and signal reconstruction
  - MAP estimator
  - MRI examples



Splines: signals with finite rate of innovation

$\mathrm{L}\{\cdot\}$: differential operator

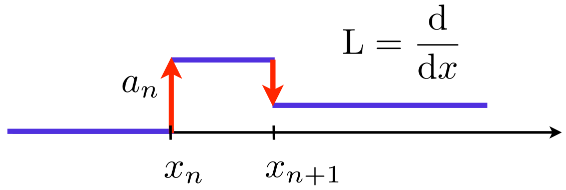
$\delta(x)$: Dirac distribution

Definition

The function $s(x)$ is a (non-uniform) L-spline with knots $\{x_n\}$ iff.

$$\mathrm{L}\{s\}(x) = \sum_{n=1}^{N} a_n \delta(x - x_n)$$

Spline theory: (Schultz-Varga, 1967)

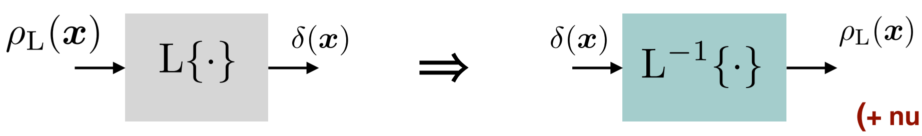


- FIR signal processing: Innovation variables $(2N)$ (Vetterli et al., 2002)
  - Location of singularities (knots) : $\{x_n\}_{n=1,\dots,N}$
  - Strength of singularities (linear weights): $\{a_n\}_{n=1,\dots,N}$

Splines and Green's functions

Definition

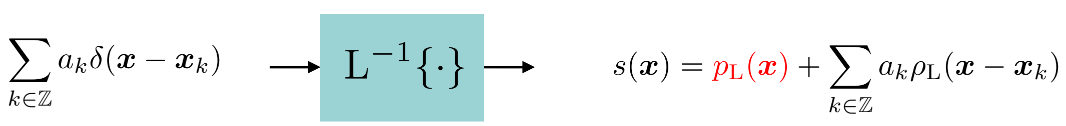
$\rho_{\mathrm{L}}(\boldsymbol{x})$ is a Green function of the shift-invariant operator L iff $\mathrm{L}\{\rho_{\mathrm{L}}\} = \delta$



(+ null-space component?)

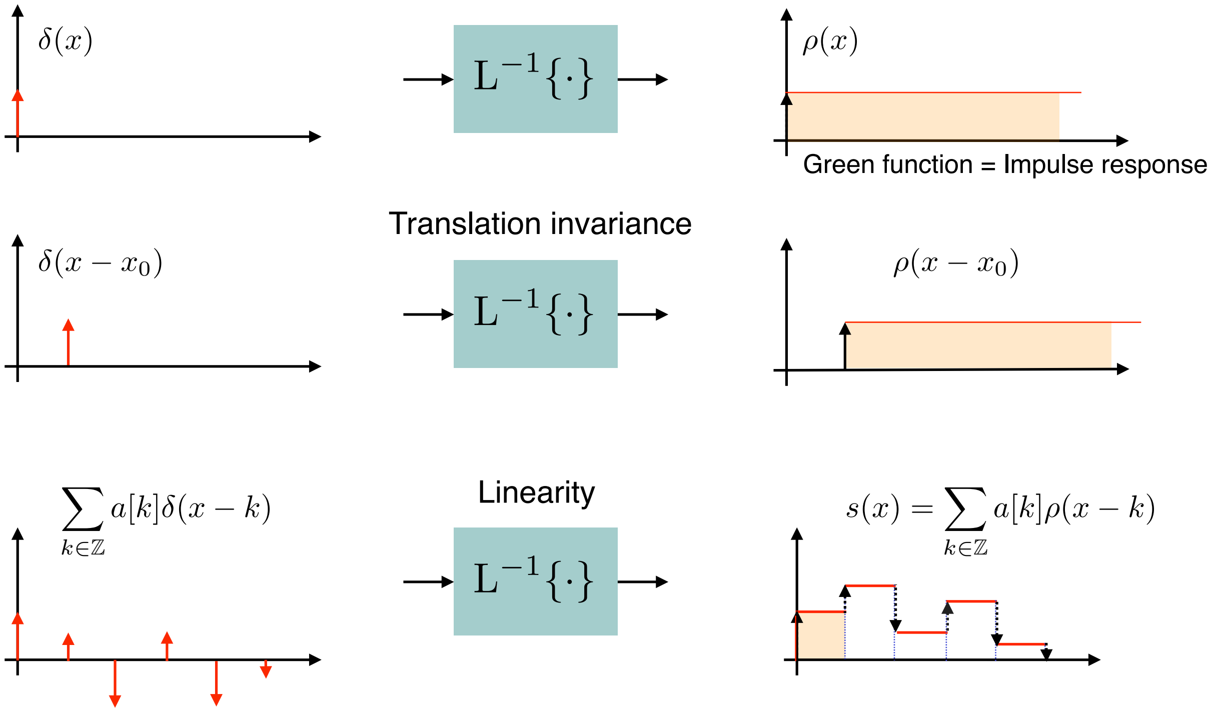
- General (non-uniform) L-spline: $\mathrm{L}\{s\}(\boldsymbol{x}) = \sum_{k\in\mathbb{Z}} a_k \delta(\boldsymbol{x} - \boldsymbol{x}_k)$

Formal integration



Example of spline synthesis

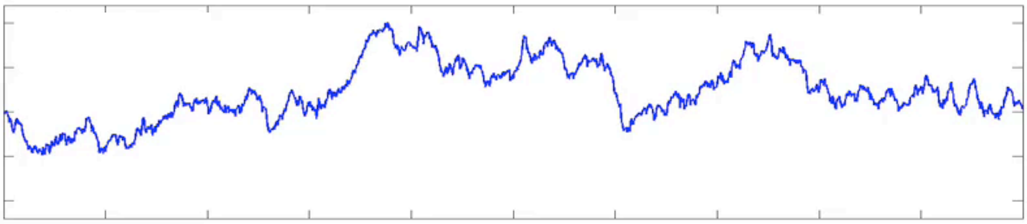
$$\mathrm{L} = \frac{\mathrm{d}}{\mathrm{d}x} = \mathrm{D} \;\Rightarrow\; \mathrm{L}^{-1}\text{: integrator}$$



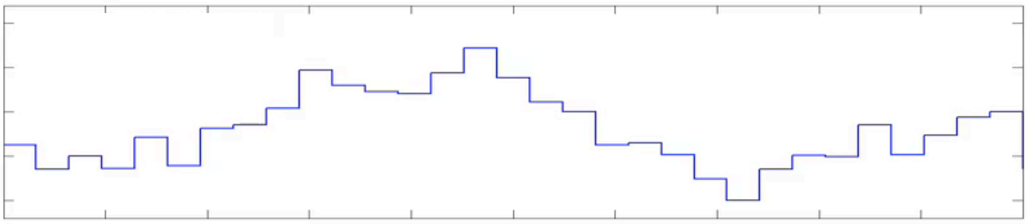
Brownian motion vs. spline synthesis

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

white noise
or
stream of Diracs $\longrightarrow$ $\boxed{\mathrm{L}^{-1}\{\cdot\}}$ $\longrightarrow$



Brownian motion



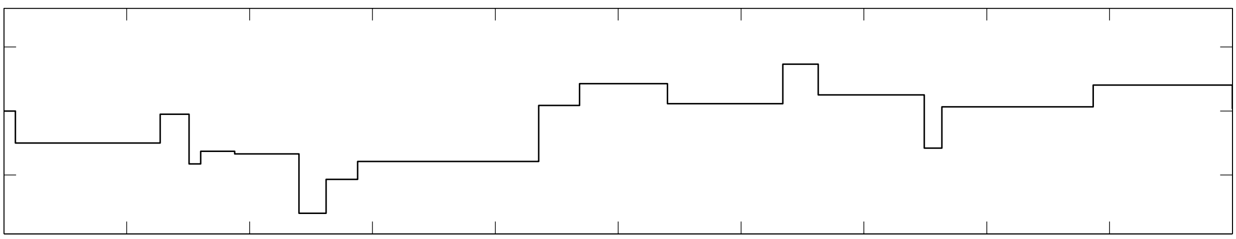
Cardinal spline (Schoenberg, 1946)

Compound Poisson process (sparse)

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

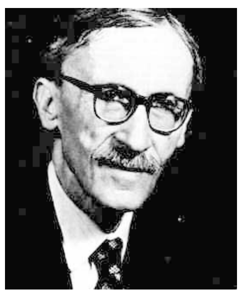
$$r(x) = \sum_k a_k \delta(x - x_k) \longrightarrow \boxed{\mathrm{L}^{-1}\{\cdot\}} \longrightarrow s(x)$$

random stream of Diracs **Compound Poisson process**



Random jumps with rate λ (Poisson point process)

Jump size distribution: $a \backsim p(a)$



(Paul Lévy, 1934)

Generalized stochastic processes

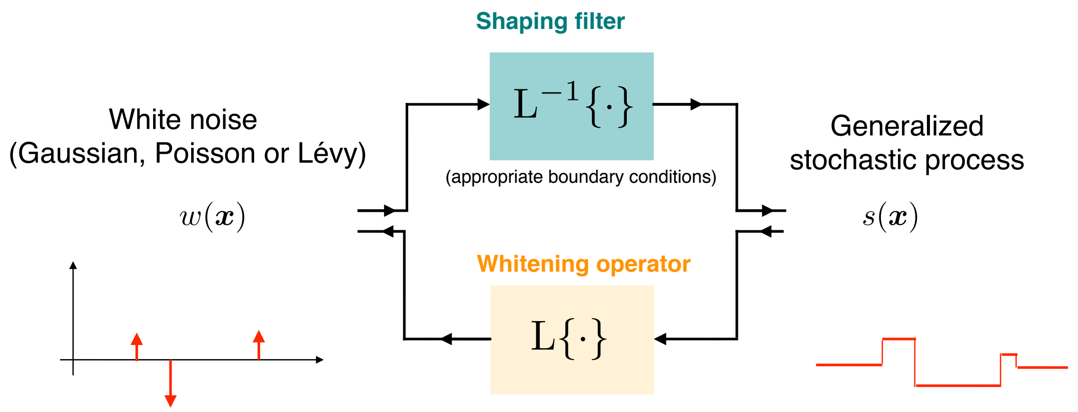
Splines are in direct correspondence with stochastic processes (stationary or fractals) that are solution of the same partial differential equation, but with a random driving term.

Defining operator equation: $\mathrm{L}\{s(\cdot)\}(\boldsymbol{x}) = r(\boldsymbol{x})$

- Specific driving terms
  - $r(\boldsymbol{x}) = \delta(\boldsymbol{x}) \quad \Rightarrow \quad s(\boldsymbol{x}) = \mathrm{L}^{-1}\{\delta\}(\boldsymbol{x})$: Green function
  - $r(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} a[\boldsymbol{k}]\delta(\boldsymbol{x}-\boldsymbol{k}) \quad \Rightarrow \quad s(\boldsymbol{x})$: Cardinal L-spline
  - $r(\boldsymbol{x})$: white noise $\quad \Rightarrow \quad s(\boldsymbol{x})$: generalized stochastic process

⚠ non-empty null space of L, boundary conditions

References: stationary proc. (U.-Blu, *IEEE-SP* 2006), fractals (Blu-U., *IEEE-SP* 2007), sparse processes (U.-Tafti, *IEEE-SP* 2010)



- What is white noise ?
  - **The problem**: Continuous-domain white noise does not have a pointwise interpretation.
  - **Standard stochastic calculus.** Statisticians circumvent the difficulty by working with *random measures* ($\mathrm{d}W(x) = w(x)\mathrm{d}x$) and *stochastic integrals*; i.e, $s(x) = \int_{\mathbb{R}} \rho(x,x')\mathrm{d}W(x')$ where $\rho(x,x')$ is a suitable kernel.
  - **Innovation model.** The white noise interpretation is more appealing for engineers (cf. Papoulis), but it requires a proper distributional formulation and operator calculus.

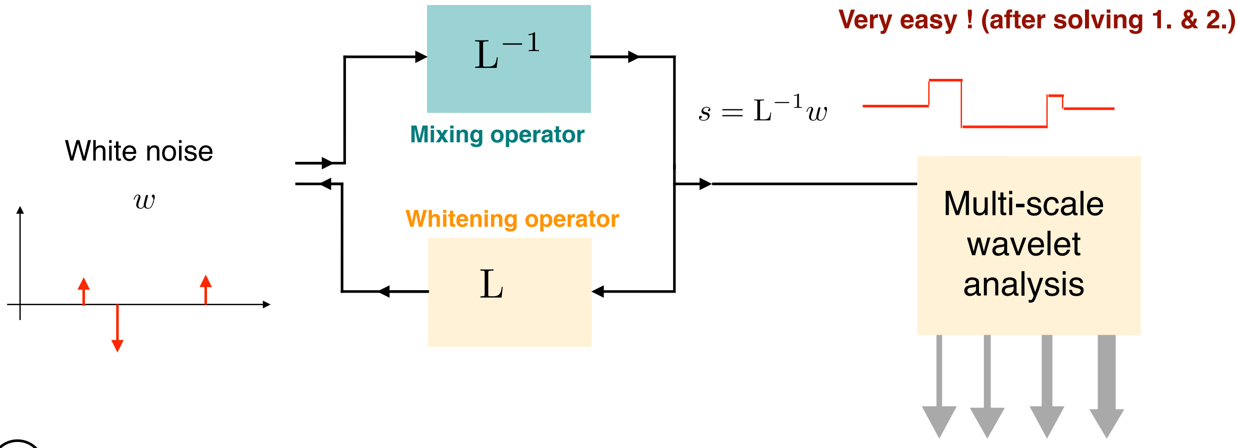
Road map for theory of sparse processes

(2) Specification of inverse operator

Functional analysis solution of SDE

(3) Characterization of generalized stochastic process

Very easy ! (after solving 1. & 2.)



(1) Characterization of continuous-domain white noise

Higher mathematics: generalized functions (Schwartz)

measures on topological vector spaces

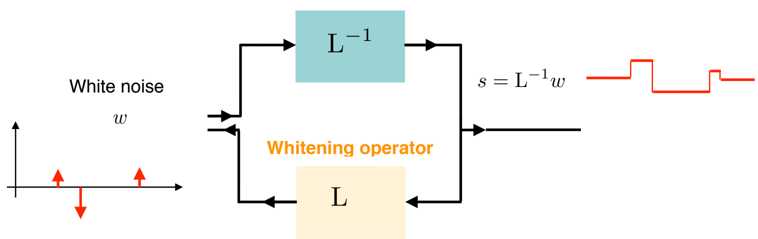
Gelfand's theory of *generalized stochastic processes*

Infinite divisibility (Lévy-Khintchine formula)

(4) Characterization of transform-domain statistics

Easy when: $\psi_i = \mathrm{L}^* \phi_i$

Step 1: White noise characterization



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

  $\Rightarrow \quad s_\varphi = \langle w, \varphi \rangle$ for any $\varphi \in \mathcal{S}$

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

- Infinite-dimensional counterpart of characteristic function (Gelfand, 1955)

  Characteristic functional: $\widehat{\mathscr{P}_w}(\varphi) = \mathbb{E}\{e^{j\langle w,\varphi\rangle}\} = \int_{\mathcal{S}'} e^{j\langle s,\varphi\rangle} \mathscr{P}_w(ds),$ for any $\varphi \in \mathcal{S}$

- White noise property: independence at every point

Characteristic form of white "noise" processes

- Definition: Independence at every point

$\widehat{\mathscr{P}_w}(\varphi_1+\varphi_2)=\widehat{\mathscr{P}_w}(\varphi_1)\times\widehat{\mathscr{P}_w}(\varphi_2)$ whenever $\varphi_1\times\varphi_2=0$ (disjoint support)

- Functional characterization (Gelfand-Vilenkin)

The characteristic form $\widehat{\mathscr{P}_w}(\varphi)=\exp\left(\int_{\mathbb{R}^d} f(\varphi(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}\right)$ defines a white noise w over $\mathcal{S}'(\mathbb{R}^d)$

$\Leftrightarrow$ $f:\mathbb{R}\rightarrow\mathbb{C}$ is a conditionally positive-definite function of order one

$\Leftrightarrow$ $f(\omega)$ is a valid Lévy exponent

$\Leftrightarrow$ $\hat{p}_{\mathrm{id}}(\omega)=e^{f(\omega)}$ is an infinitely-divisible (id) characteristic function

- Bottom line

WNP uniquely specified by $f(\omega)$ $\Leftrightarrow$ $p_{\mathrm{id}}(x)=\mathcal{F}^{-1}\{e^{f(\omega)}\}(x)$ (canonical id pdf)

Example of usage: $X=w(\varphi_0)=\langle w,\varphi_0\rangle$

$\Rightarrow$ $p_X(x)=\mathcal{F}^{-1}\{\hat{p}_X(\omega)\}$ where $\hat{p}_X(\omega)=\mathbb{E}\{e^{j\omega\langle w,\varphi_0\rangle}\}=\widehat{\mathscr{P}_w}(\omega\varphi_0)$

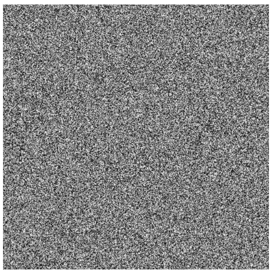
White noise: canonical distribution

Continuous-domain white noise is **highly singular**; its points values are undefined

A given brand uniquely specified by $p_{\mathrm{id}}(x)=\mathcal{F}^{-1}\{e^{f(\omega)}\}(x)$

- Interpretation: noise observation through a rectangular window

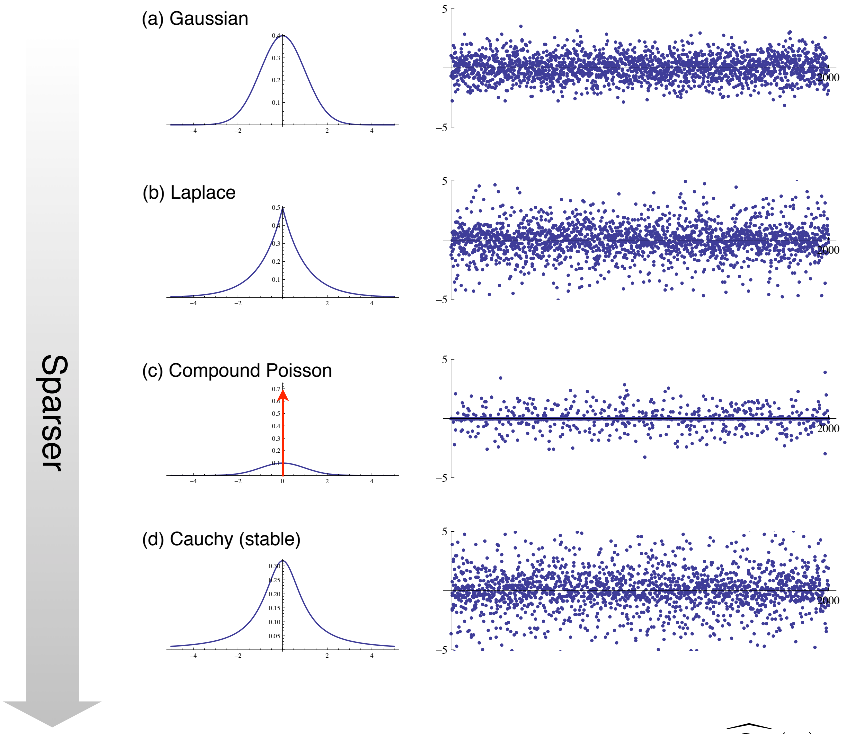
$\widehat{\mathscr{P}_w}(\omega\,\mathrm{rect}(\boldsymbol{x}))=e^{f(\omega)}$ $\Leftrightarrow$ $p_{\mathrm{id}}(x)=p_{X_{\mathrm{id}}}(x)$ with $X_{\mathrm{id}}=\langle\mathrm{rect}(\cdot-\boldsymbol{k}),w\rangle$ (i.i.d.)



- Special cases
  - $f(\omega)=\frac{1}{2}|\omega|^2$ $\Leftrightarrow$ $p_{\mathrm{id}}(x)$: normalized Gaussian
  - $f(\omega)=|\omega|^\alpha$ with $\alpha\in(0,2]$ $\Leftrightarrow$ $p_{\mathrm{id}}(x)$: Symmetric-α-stable (SαS)
  - Also allowed: compound Poisson, Beta, Student, Cauchy, etc. (typically heavy tailed)

Examples of id noise distributions

$p_{\mathrm{id}}(x)$



(a) $f(\omega) = \frac{1}{2\sigma_0^2}|\omega|^2$

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

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

(d) $f(\omega) = s|\omega|$

Complete mathematical characterization: $\widehat{\mathscr{P}_w}(\varphi) = \exp\left(\int_{\mathbb{R}^d} f\big(\varphi(\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x}\right)$

Complete characterization of id distributions

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 has the same distribution as $X_1 + \cdots + X_N$.

Lévy-Khinchine theorem

$p_{\mathrm{id}}(x)$ is an infinitely divisible (id) PDF iff. its characteristic function can be written as

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

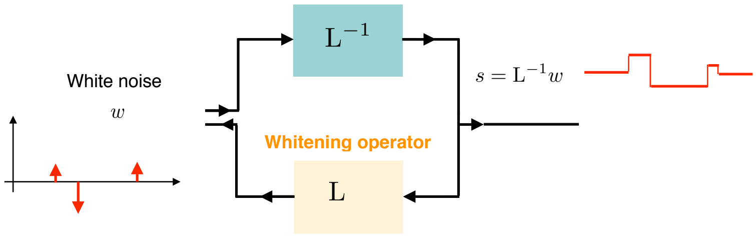
with Lévy exponent

$$f(\omega) = jb_1\omega - \frac{b_2\omega^2}{2} + \int_{\mathbb{R}\setminus\{0\}} \left(e^{ja\omega} - 1 - ja\omega \mathbb{I}_{|a|<1}(a)\right) v(a)\mathrm{d}a$$

where $b_1 \in \mathbb{R}$ and $b_2 \in \mathbb{R}^+$ are some constants, and where $v(a) \geq 0$ is some positive function (density Lévy) such that $\int_{\mathbb{R}} \min(a^2, 1) v(a)\mathrm{d}a < \infty$.

Theoretical relevance: one-to-one correspondence between a "classical" id PDF and a white noise processes

Steps 2 + 3: Characterization of sparse process



- Abstract formulation of innovation model

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

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

- Technical aspect: functional analysis

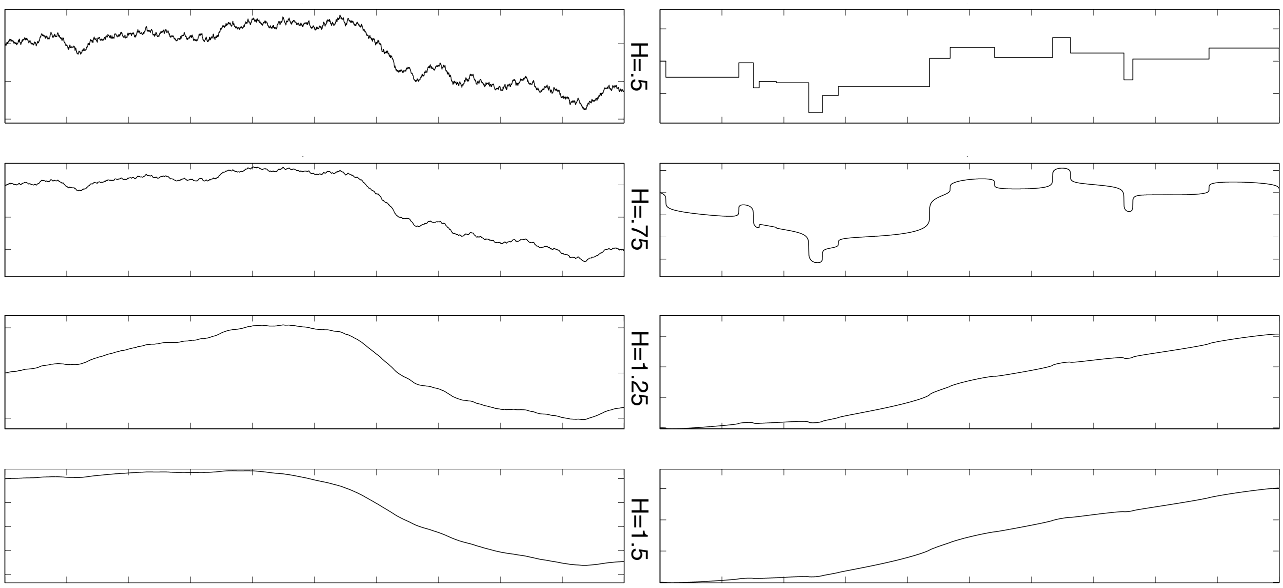
Find an acceptable inverse of L such that the adjoint operator L^{-1*} is well-defined over Schwartz's class of test functions

Ideally: $\mathrm{L}^{-1*}\varphi \in \mathcal{S}$

or at least $\|\mathrm{L}^{-1*}\varphi\|_{L_p} < C\,\|\varphi\|_{L_p}$ (continuity)

Self-similar processes (TS-invariant)

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

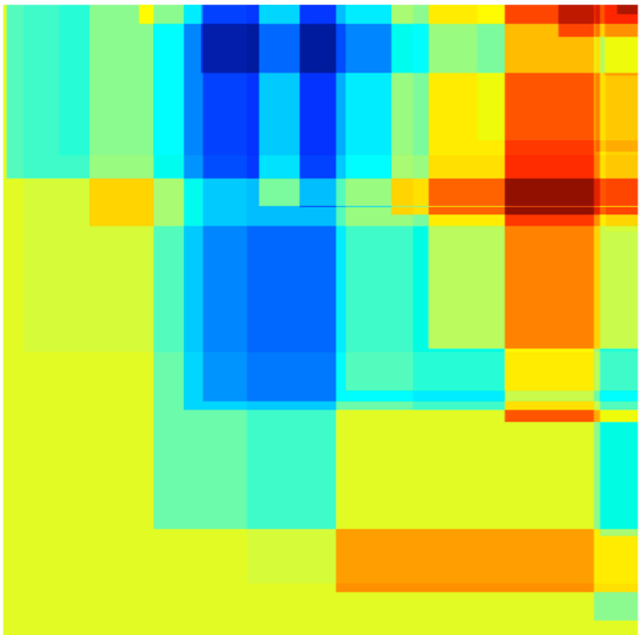


Gaussian — Fractional Brownian motion (Mandelbrot, 1968)

Sparse (generalized Poisson) — (U.-Tafti, *IEEE-SP* 2010)

2D generalization: the Mondrian process

$$\mathrm{L} = \mathrm{D}_x \mathrm{D}_y \quad \stackrel{\mathcal{F}}{\longleftrightarrow} \quad (j\omega_x)(j\omega_y)$$

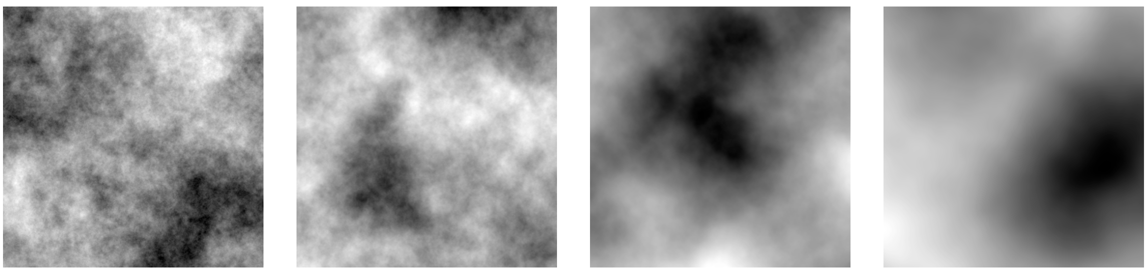


$\lambda = 30$

Scale- and rotation-invariant processes

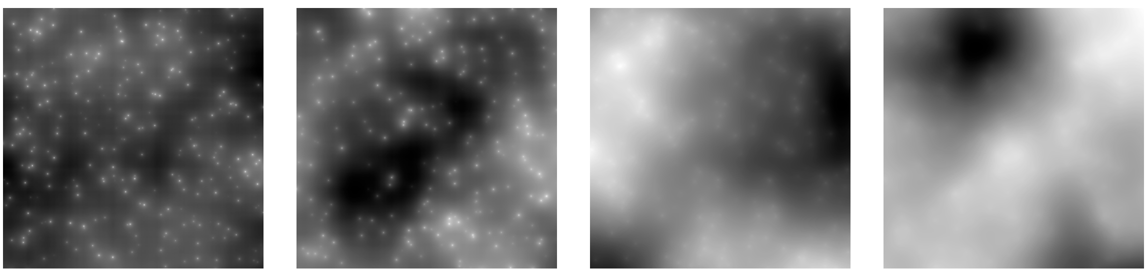
Stochastic partial differential equation : $(-\Delta)^{\frac{H+1}{2}} s(\boldsymbol{x}) = w(\boldsymbol{x})$

Gaussian



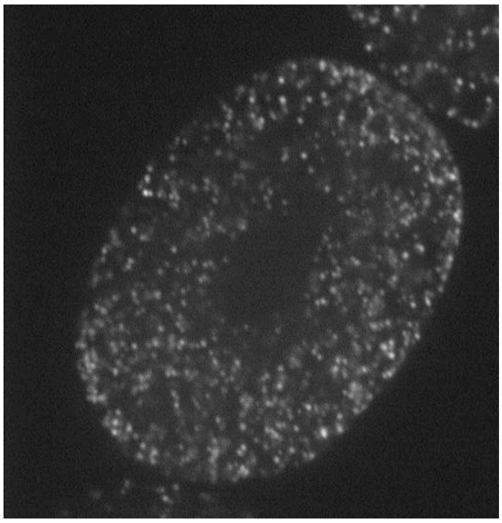
H=.5 H=.75 H=1.25 H=1.75

Sparse (generalized Poisson)



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

Powers of ten: from astronomy to biology



IMPLICATION OF INNOVATION MODEL

- Optimized analysis tools = B-splines
- Decoupling sparse processes
- Wavelet analysis of sparse processes
- Determination of transform-domain statistics
- Signal reconstruction algorithm (MAP)

Recap on infinite-dimensional innovation model

Generic test function $\varphi \in \mathcal{S}$ plays the role of index variable

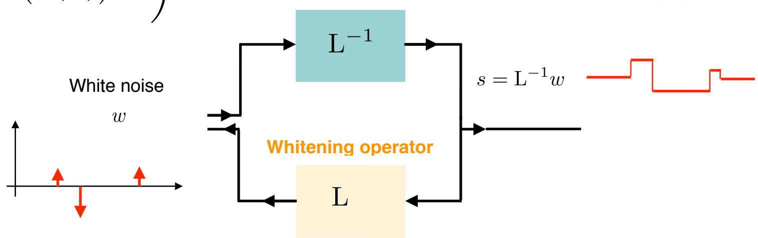
Solution of SDE

(2)

$$\widehat{\mathscr{P}_w}(\varphi) = \exp\left(\int_{\mathbb{R}^d} f\big(\varphi(\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x}\right)$$

$$\widehat{\mathscr{P}_s}(\varphi) = \widehat{\mathscr{P}_w}(\mathrm{L}^{-1*}\varphi)$$

(1)



(3)

White "noise" signature: $p_{\mathrm{id}}(x) = \mathcal{F}^{-1}\{e^{f(\omega)}\}(x)$

(4) Analysis step

Regularization operator vs. wavelet analysis

Optimized analysis tools = B-splines

- Whitening operator L

  Green function $\rho_{\mathrm{L}}(\boldsymbol{x})$ such that $\mathrm{L}\rho_{\mathrm{L}} = \delta$

$$\begin{aligned} \mathrm{L}s &= w \\ s &= \mathrm{L}^{-1}w \end{aligned}$$

- Discrete version of operator

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

- Generalized B-spline

$$\beta_{\mathrm{L}}(\boldsymbol{x}) = \mathrm{L_d}\mathrm{L}^{-1}\delta(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} d[\boldsymbol{k}]\rho_{\mathrm{L}}(\boldsymbol{x}-\boldsymbol{k})$$

  Quality of discrete approximation:

$$\mathrm{L_d}s(\boldsymbol{x}) = \mathrm{L_d}\underbrace{\mathrm{L}^{-1}\mathrm{L}}_{\mathrm{Id}}s(\boldsymbol{x}) = \beta_{\mathrm{L}} * \mathrm{L}s(\boldsymbol{x})$$

$\Rightarrow$ β_{L} should be well-defined $\big(\beta_{\mathrm{L}} \in L_1(\mathbb{R}^d)\big)$ and maximally localized (short support)

Optimized analysis tools: introductory example

- Whitening operator D

  Green function $\rho_{\mathrm{D}}(x) = 1_+(x)$ (unit step)

SDE for Lévy process

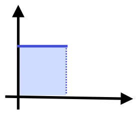
$$\begin{aligned} \mathrm{D}s(x) &= w(x) \\ s(x) &= \int_0^x w(y)\mathrm{d}y \end{aligned}$$

- Finite difference operator

$$\mathrm{D_d}s(x) = s(x) - s(x-1)$$

- Piecewise-constant B-spline

$$\beta_{(0)}(x) = 1_+(x) - 1_+(x-1) = \mathrm{rect}(x - \frac{1}{2})$$



B-spline of minimal support: $\beta_{(0)}(x) \in L_p(\mathbb{R})$ for $p > 0$

Decoupling sparse processes

- Innovation model (SDE)

$$\begin{aligned} \mathrm{L}s &= w \\ s &= \mathrm{L}^{-1}w \end{aligned}$$

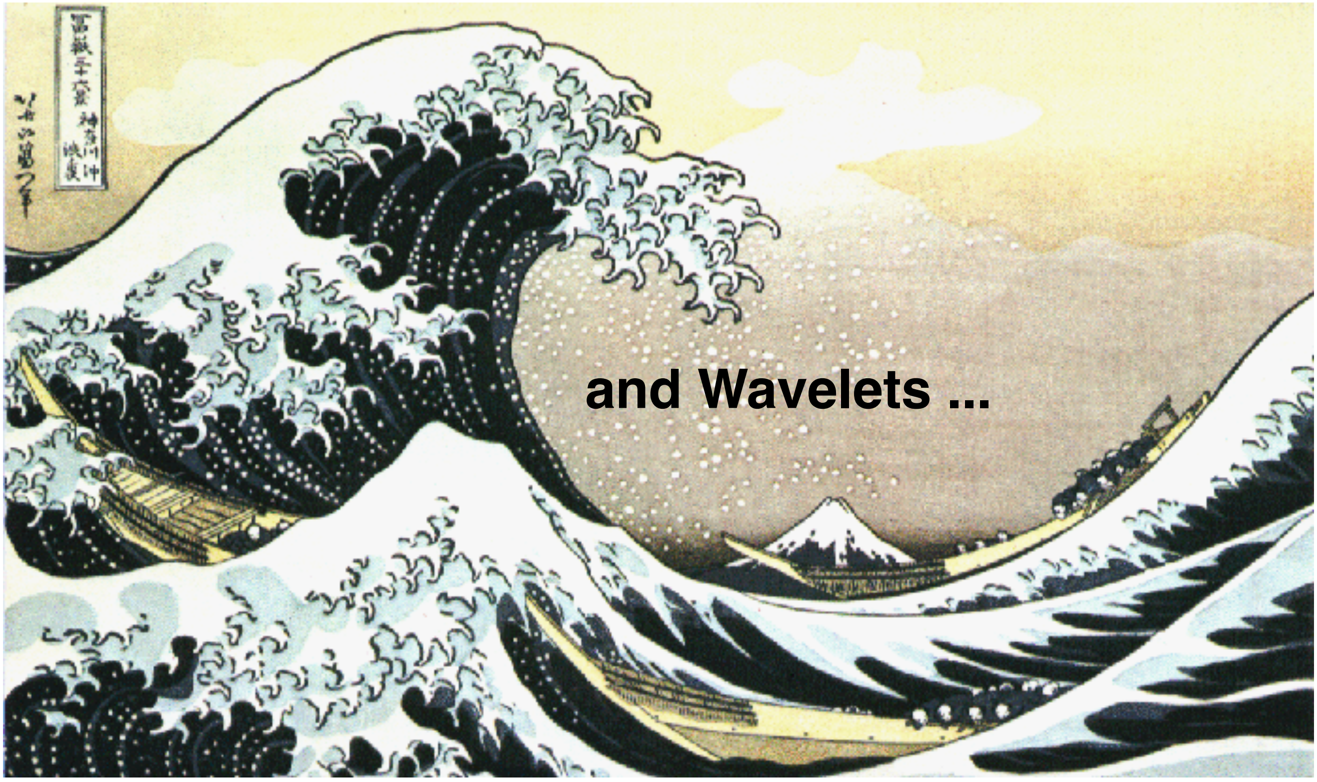
- Generalized increment process

$$u = \mathrm{L_d}s = \mathrm{L_d}\mathrm{L}^{-1}w = \beta_{\mathrm{L}} * w$$

$$\langle u, \varphi \rangle = \langle \beta_{\mathrm{L}} * w, \varphi \rangle = \langle w, \beta_{\mathrm{L}}^{\vee} * \varphi \rangle \quad \text{with} \quad \beta_{\mathrm{L}}^{\vee}(\boldsymbol{x}) = \beta_{\mathrm{L}}(-\boldsymbol{x})$$

$$\Longrightarrow \quad \widehat{\mathscr{P}}_u(\varphi) = \widehat{\mathscr{P}}_w(\beta_{\mathrm{L}}^{\vee} * \varphi)$$

- Statistical implications
  - $u = \mathrm{L_d}s$ is stationary with characteristic function $\widehat{\mathscr{P}}_w(\omega\beta_{\mathrm{L}}^{\vee})$
  - Quality of decoupling depends upon support of B-spline $\beta_{\mathrm{L}}(\boldsymbol{x})$



Wavelet analysis of sparse processes

- Innovation model (SDE)

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

- Operator-like wavelet: $\psi_i = \mathrm{L}^*\phi_i$

  ϕ_i: smoothing kernel at wavelet scale i

- Wavelet analysis

$$v_i(\boldsymbol{x}) = \langle \psi_i(\cdot - \boldsymbol{x}), s\rangle = \langle \mathrm{L}^*\phi_i(\cdot - \boldsymbol{x}), \mathrm{L}^{-1}w\rangle = \langle \phi_i(\cdot - \boldsymbol{x}), w\rangle$$

$$\Longrightarrow \quad \widehat{\mathscr{P}}_{v_i}(\varphi) = \widehat{\mathscr{P}}_w(\phi_i * \varphi)$$

- Statistical implications
  - Wavelet coefficients v_i are stationary with characteristic function $\widehat{\mathscr{P}}_w(\omega\phi_i)$
  - Quality of decoupling depends upon support of wavelet/smoothing kernel ϕ_i

Determination of transform-domain statistics

- Common white noise analysis framework

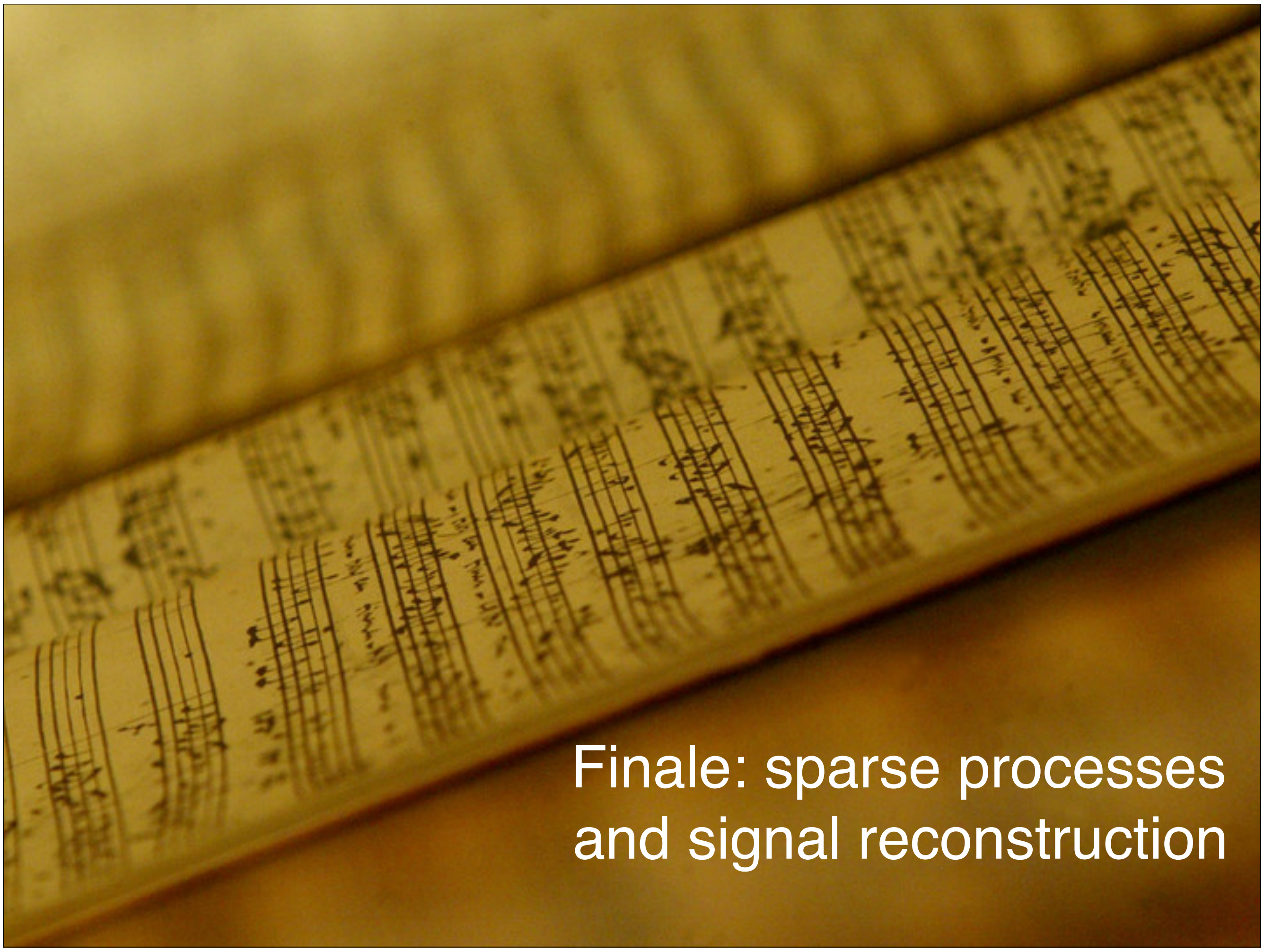
$$X(\varphi) = \langle w, \varphi \rangle \quad \text{for suitable } \varphi \quad (\text{e.g., } \beta_{\mathrm{L}}, \phi_i, \mathrm{rect})$$

- Explicit form of characteristic function

$$\hat{p}_{X(\varphi)}(\omega) = \widehat{\mathscr{P}_w}(\omega\varphi) = e^{f_\varphi(\omega)} \quad \text{with} \quad f_\varphi(\omega) = \int_{\mathbb{R}^d} f\big(\omega\varphi(\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x}$$

- General properties
  - $\int_{\mathbb{R}} |x|^p p_{\mathrm{id}}(x)\mathrm{d}x < \infty \quad \Rightarrow \quad p_{X(\varphi)}(x)$ well-defined for all $\varphi \in L_p(\mathbb{R}^d)$
  - $p_{\mathrm{id}}(x)$ is symmetric, unimodal $\quad \Rightarrow \quad p_{X(\varphi)}(x)$ idem
  - $p_{\mathrm{id}}(x) = O(1/|x|^p)$ with $p > 1$ (heavy tailed) $\quad \Rightarrow \quad p_{X(\varphi)}(x)$ idem
  - $p_{\mathrm{id}}(x)$ (non-Gaussian) $\quad \Rightarrow \quad p_{X(\varphi)}(x)$ is sparse

Decay: $O(e^{-\gamma|x|})$ (exponential), $x^p e^{-\gamma|x|}$, or $O(1/|x|^p)$ (algebraic)



Signal reconstruction: MAP formulation

- Innovation model of the signal

$$\begin{aligned} \mathrm{L}s &= w \\ s &= \mathrm{L}^{-1}w \end{aligned}$$

- Signal decoupling: discrete version of operator

$$u(\boldsymbol{x}) = \mathrm{L_d}\, s(\boldsymbol{x}) \quad \Leftrightarrow \quad \mathbf{u} = \mathbf{Ls} \quad \text{(matrix notation)}$$

- Statistical characterization
  - $X = [\mathbf{u}]_n$ identically distributed (approx. independent)
  - Probability density function: $p_X(x) = \mathcal{F}^{-1}\{\widehat{\mathscr{P}_w}(\omega \beta_{\mathrm{L}}^{\vee})\}(x)$
  - Potential function: $\Phi_X(x) = -\log p_X(x)$

- Maximum a posteriori (MAP) estimator for AWN

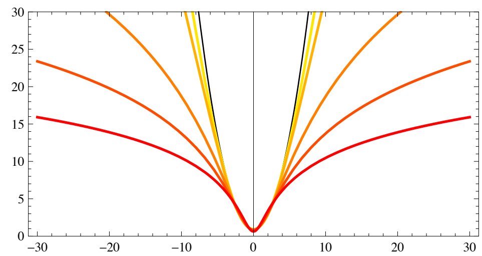
$$\mathbf{s}^{\star} = \mathrm{argmin} \left(\tfrac{1}{2} \|\mathbf{g} - \mathbf{Hs}\|_2^2 + \sigma^2 \textstyle\sum_n \Phi_X([\mathbf{Ls}]_n) \right)$$

MAP estimator: special cases

$$\mathbf{s}^{\star} = \mathrm{argmin} \left(\tfrac{1}{2} \|\mathbf{g} - \mathbf{Hs}\|_2^2 + \sigma^2 \textstyle\sum_n \Phi_X([\mathbf{Ls}]_n) \right)$$

- Gaussian: $p_X(x) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-x^2/(2\sigma_0^2)} \quad \Rightarrow \quad \Phi_X(x) = \frac{1}{2\sigma_0^2} x^2$
- Laplace: $p_X(x) = \frac{\lambda}{2} e^{-\lambda |x|} \quad \Rightarrow \quad \Phi_X(x) = \lambda |x|$
- Student: $p_X(x) = \dfrac{1}{B\left(r, \frac{1}{2}\right)} \left(\dfrac{1}{x^2+1} \right)^{r+\frac{1}{2}} \quad \Rightarrow \quad \Phi_X(x) = \left(r + \dfrac{1}{2}\right) \log(1 + x^2)$

Sparser



Student potentials: $r = 2, 4, 8, 32$ (fixed variance)

Reconstruction algorithms

FWISTA (Guerquin-Kern TMI 2011), IRWL1 (Candès), AL (Ramani TMI 2011)

- Variable splitting with quadratic penalty

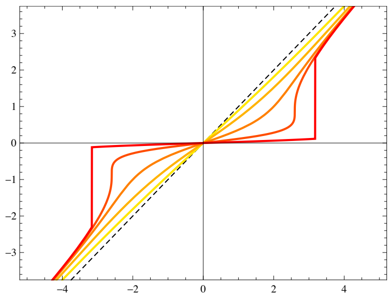
Auxiliary **innovation** variable: $\mathbf{u} = \mathbf{L}\mathbf{s}$

$$(\mathbf{s}^*, \mathbf{u}^*) = \underset{\mathbf{s},\mathbf{u}\in\mathbb{R}^N}{\mathrm{argmin}} \left(\frac{1}{2}\|\mathbf{g} - \mathbf{H}\mathbf{s}\|_2^2 + \sigma^2 \sum_n \Phi_X([\mathbf{u}]_n) + \frac{\mu}{2}\|\mathbf{L}\mathbf{s} - \mathbf{u}\|_2^2 \right)$$

ADM: Alternating minimization over s (linear problem) and $\mathbf{u}$ (non-linear)

- Proximal operator taylored to stochastic model

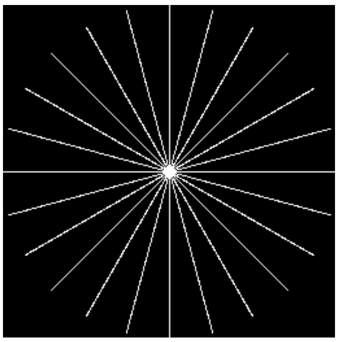
$$\mathrm{prox}_{\Phi_X}(y;\lambda) = \arg\min_u \frac{1}{2}|y-u|^2 + \lambda\Phi_X(u)$$



Cauchy prior with increasing s

MRI: Shepp-Logan phantom

Original SL Phantom



Fourier Sampling Pattern
12 Angles

Laplace prior (TV)

Student prior (log)

L : gradient

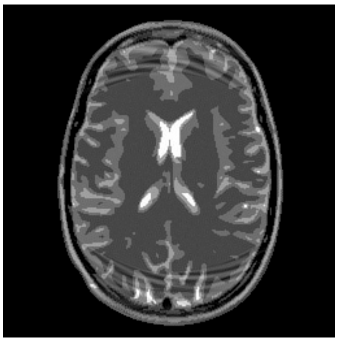
Optimized parameters

MRI: Spiral sampling in k-space

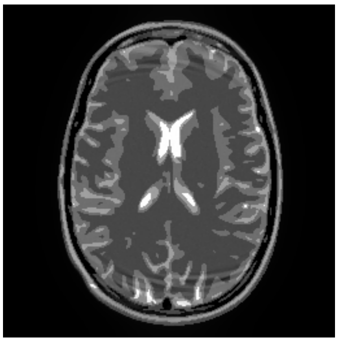
L : gradient

Optimized parameters

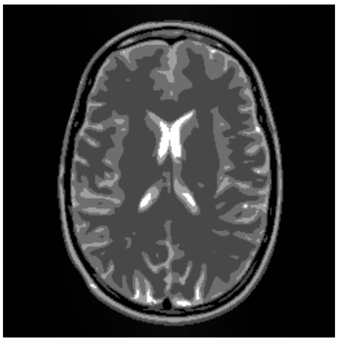
Original Phantom
(Guerquin-Kern et al.)



Gaussian prior (Tikhonov)
SER =17.69 dB



Laplace prior (TV)
SER = 21.37 dB



Student prior
SER = 27.22 dB

CONCLUSION

- Unifying continuous-domain innovation model
  - Backward compatibility with classical Gaussian theory
  - Operator-based formulation: Lévy-driven SDEs or SPDEs
  - **Gaussian** vs. **sparse** (generalized Poisson, student, SαS)
  - Focus on unstable SDEs ⇒ non-stationary, self-similar processes
- Regularization vs. wavelet analysis
  - Central role of B-spline
  - Sparsification via “operator-like” behavior
    - Discrete approximation of whitening operator
    - Multi-resolution: wavelets
- Theoretical framework for sparse signal recovery
  - New statistically-founded sparsity priors
  - Derivation of optimal estimators (MAP, MMSE)
  - Guide for the development of novel algorithms

References

- Theory of generalized stochastic processes

  - M. Unser, P. Tafti, and Q. Sun, "A unified formulation of Gaussian vs. sparse stochastic processes—Part I: Continuous-domain theory," preprint, available at http://arxiv.org/abs/1108.6150.
  - M. Unser, P.D. Tafti, "Stochastic models for sparse and piecewise-smooth signals", *IEEE Trans. Signal Processing*, vol. 59, no. 3, pp. 989-1006, March 2011.
  - M. Unser and T. Blu, "Generalized Smoothing Splines and the Optimal Discretization of the Wiener Filter," *IEEE Transactions on Signal Processing*, vol. 53, no. 6, pp. 2146–2159, June 2005. 1256–1267, April 2006.
  - T. Blu, M. Unser, "Self-Similarity: Part II—Optimal Estimation of Fractal Processes," *IEEE Trans. Signal Processing*, vol. 55, no. 4, pp. 1364-1378, April 2007.
  - P.D. Tafti, D. Van De Ville, M. Unser, "Invariances, Laplacian-Like Wavelet Bases, and the Whitening of Fractal Processes," *IEEE Trans. Image Processing*, vol. 18, no. 4, pp. 689-702, April 2009.

- Algorithms and imaging applications

  - M. Guerquin-Kern, M. Häberlin, K.P. Pruessmann, M. Unser, "A Fast Wavelet-Based Reconstruction Method for Magnetic Resonance Imaging," *IEEE Trans. Medical Imaging*, vol. 30, no. 9, pp. 1649-1660, September 2011.
  - S. Ramani and J.A. Fessler, "Parallel MR Image Reconstruction Using Augmented Lagrangian Methods," *IEEE Trans. Medical Imaging*, vol. 30, no. 3, pp. 694-706, August 2011.

Acknowledgments

Many thanks to

- Dr. Pouya Tafti
- Prof. Qiyu Sun
- Prof. Thierry Blu
- Dr. Arash Amini
- Dr. Hagai Kirshner
- Matthieu Guerquin-Kern
- Emrah Bostan

- **Members of EPFL's Biomedical Imaging Group**

erc

- Preprints and demos: http://bigwww.epfl.ch/