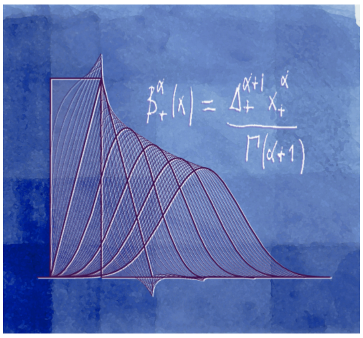
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Sparse stochastic processes and biomedical image reconstruction

Michael Unser
Biomedical Imaging Group
EPFL, Lausanne, Switzerland

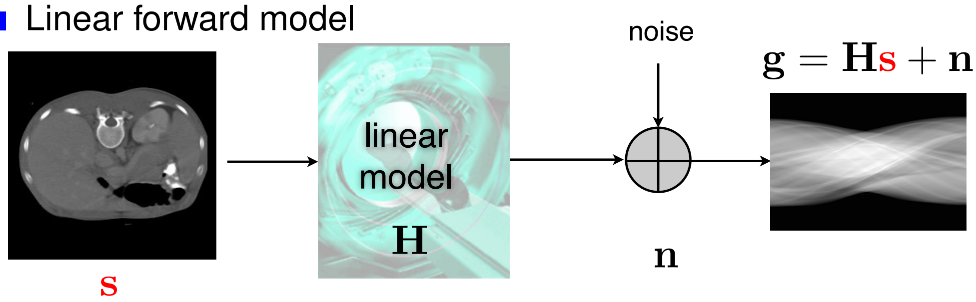
Joint work with P. Tafti, Q. Sun, A. Amini, M. Guerquin-Kern, E. Bostan, etc.



Tutorial presentation, FRUMAN, University Aix-Marseille, Feb. 4 2013.

Variational formulation of image reconstruction

- Linear forward model



Ill-posed inverse problem: recover $\mathbf{s}$ from noisy measurements $\mathbf{g}$

- Reconstruction as an optimization problem

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

Classical linear reconstruction

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

- Quadratic regularization (Tikhonov)

  $\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|^2$

  Formal linear solution: $\mathbf{s} = (\mathbf{H}^T\mathbf{H} + \lambda \mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T\mathbf{g} = \mathbf{R}_\lambda \cdot \mathbf{g}$

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

- Statistical formulation under Gaussian hypothesis

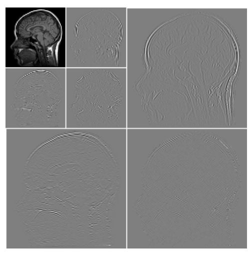
  Wiener (LMMSE) solution = Gauss MMSE = Gauss MAP

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

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

Sparsity-promoting reconstruction algorithms

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



- 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, FWISTA)

- ℓ_1 regularization (Total variation)

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

  Iterative reweighted least squares (IRLS) or FISTA

Key research questions

① Discretization of reconstruction problem

Continuous-domain formulation **Generalized sampling**

② Formulation of ill-posed reconstruction problem

Statistical modeling (beyond Gaussian)
supporting non-linear reconstruction schemes
(including CS)

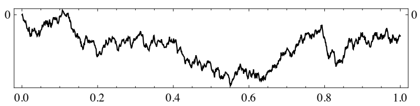
Sparse stochastic processes

③ Efficient implementation for large-scale imaging problem

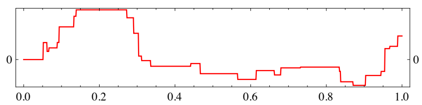
FISTA, ADM

Lévy processes

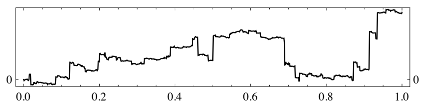
Constructed by Paul Lévy (circa 1930)



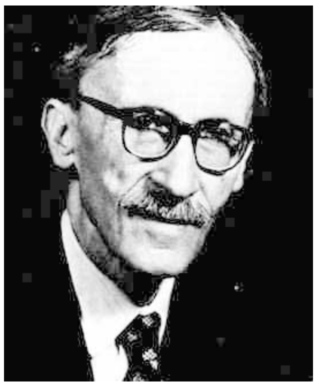
Gaussian: Brownian motion



Compound Poisson



SαS (Cauchy)



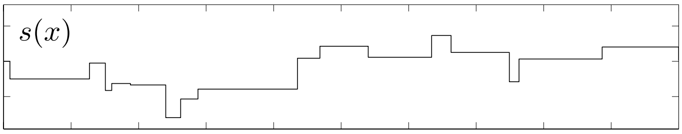
- Generalization of Brownian motion
- Independent increments: i.i.d. infinitely divisible (from Gaussian to heavy tailed)

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

⇒ **Archetype of a “sparse” random signal**

Haar wavelet analysis of a Lévy process

- Compound Poisson process = piecewise-constant signal



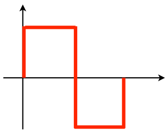
$\mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}x}$

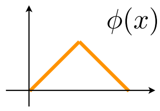
$\mathrm{D}^* = -\frac{\mathrm{d}}{\mathrm{d}x}$ (adjoint)

"Sparse derivative" property: $\mathrm{D}s(t) = \sum_n a_n \delta(x - x_n)$ with x_n jump locations

- Wavelet as a smoothed derivative

$$\psi_{\mathrm{Haar}}(x) = \mathrm{D}\phi(x)$$



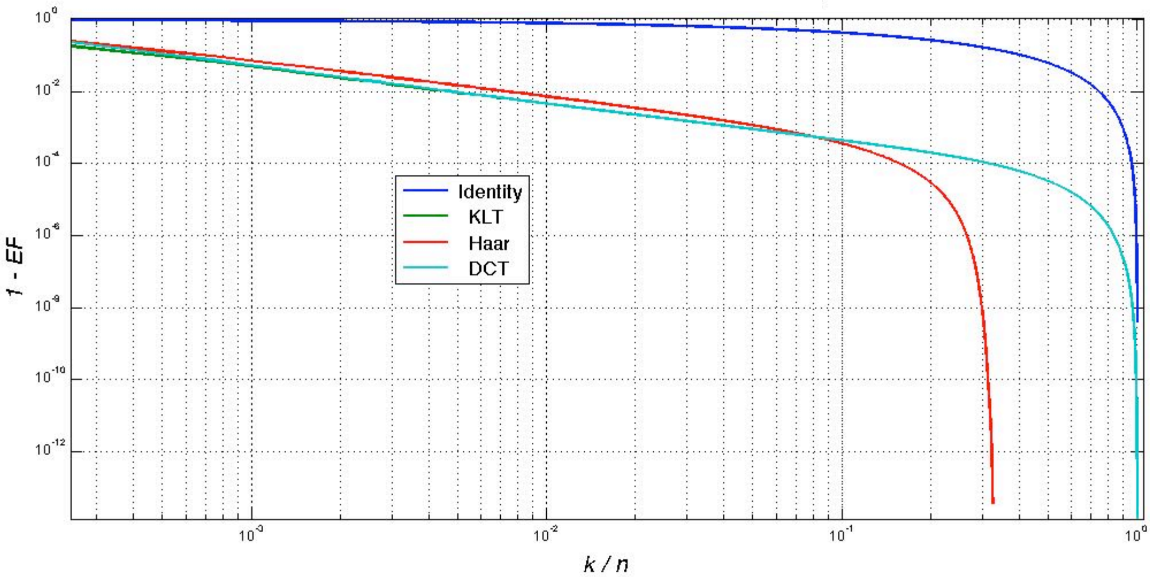


$$\Rightarrow \langle s, \psi_{\mathrm{Haar}}(\cdot - y_0)\rangle = \langle s, \mathrm{D}\phi(\cdot - y_0)\rangle = \langle \mathrm{D}^* s, \phi(\cdot - y_0)\rangle$$

sparse innovations (train of Dirac impulses)

K-term approximation of Lévy processes

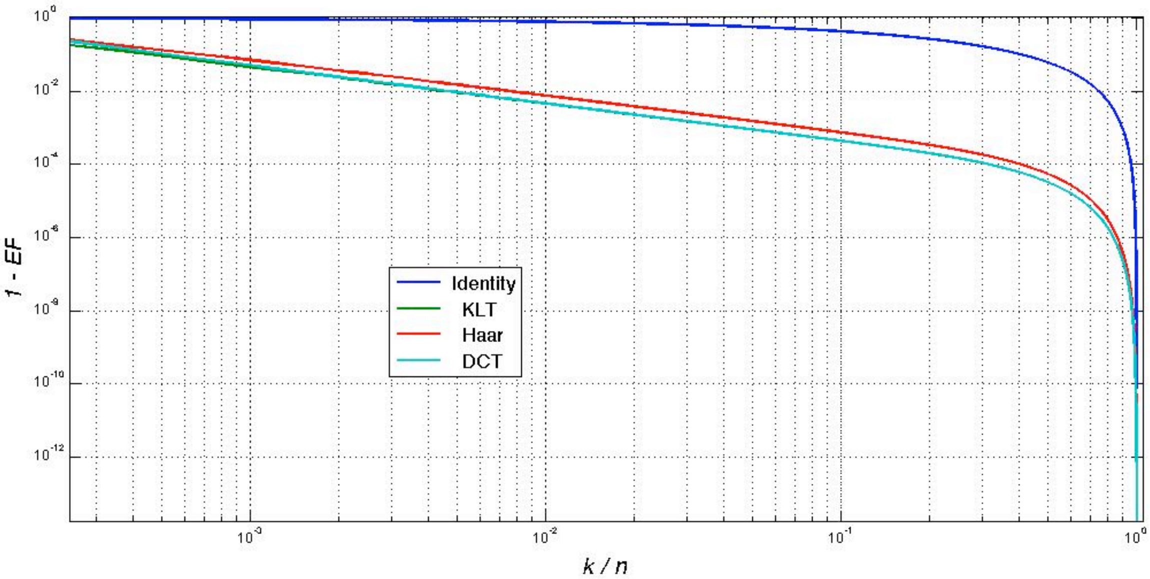
DCT→ Karhunen-Loève transform



Sparse: Compound Poisson

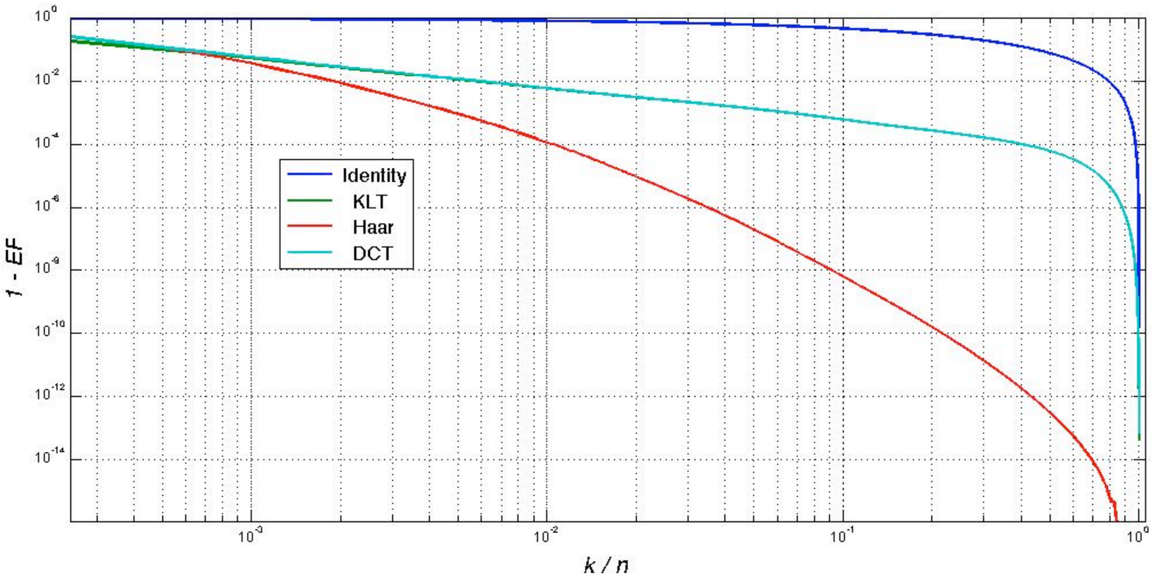
K-term approximation of Lévy processes

DCT→ Karhunen-Loève transform



Gaussian: Brownian motion

K-term approximation of Lévy processes



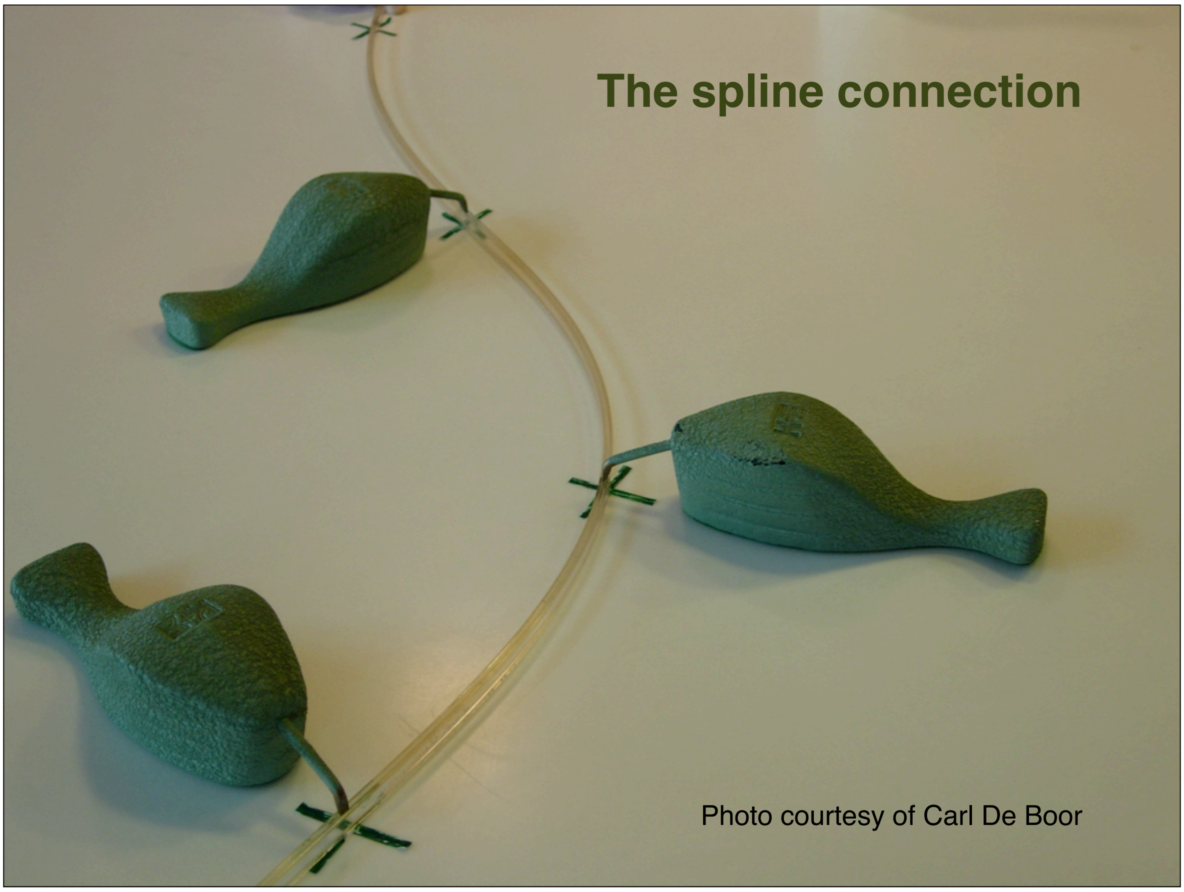
Heavy tailed: Lévy flight (Cauchy)

Arguments for continuous-domain formulation

- The real world is continuous
  - Input signal
  - Imaging physics
- Principled formulation
  - **Stochastic differential equations** (rather than reverse engineering)
  - Invariance to coordinate transformations
  - Specification of optimal estimators (MAP, MMSE)
- The power of continuous mathematics
  - Full backward compatibility with Gaussian theory, link with TV
  - Integral operators, characteristic form
  - Derivation of joint PDF in **any transformed domain** (wavelets, gradient, DCT)
  - Operational definition of "sparsity" based on existence considerations: **infinite divisibility** ⇒ processes are either Gaussian or **sparse**

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 $\qquad$ $\delta(x)$: Dirac distribution

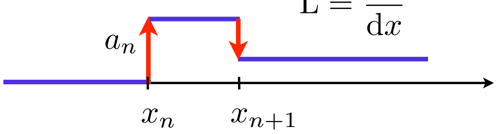
Definition

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)

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



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

(Vetterli et al., 2002)

Splines and Green's functions

Definition

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

$$\rho_{\mathrm{L}}(\boldsymbol{x}) \longrightarrow \boxed{\mathrm{L}\{\cdot\}} \longrightarrow \delta(\boldsymbol{x}) \quad \Rightarrow \quad \delta(\boldsymbol{x}) \longrightarrow \boxed{\mathrm{L}^{-1}\{\cdot\}} \longrightarrow \rho_{\mathrm{L}}(\boldsymbol{x})$$

(+ null-space component?)

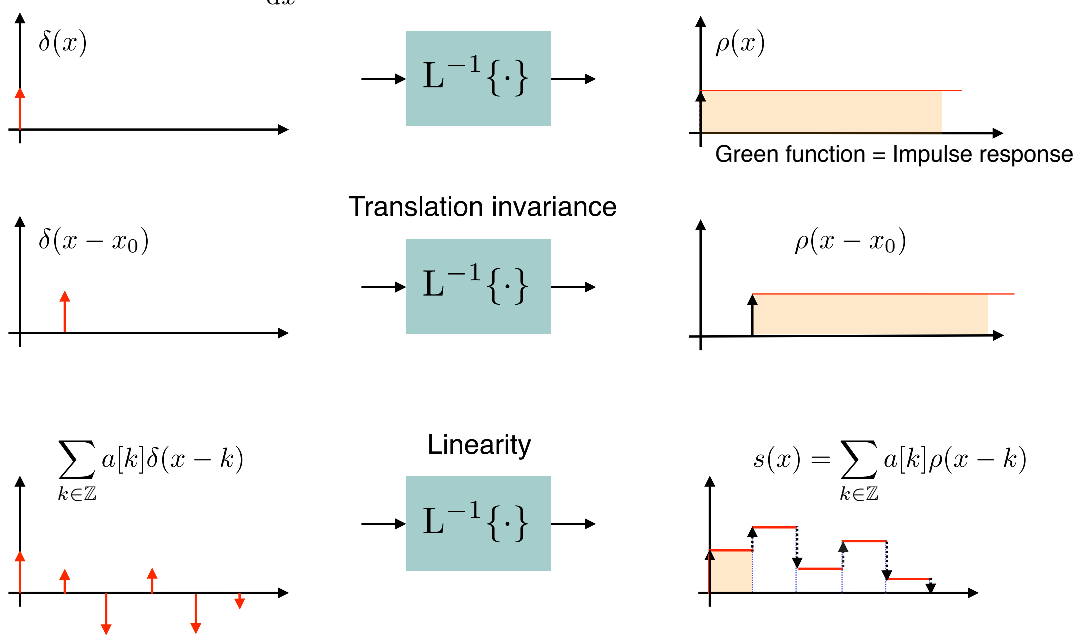
- General (non-uniform) L-spline: $\mathrm{L}\{s\}(\boldsymbol{x}) = \sum_{k\in\mathbb{Z}} a_k \delta(\boldsymbol{x} - \boldsymbol{x}_k)$

Formal integration

$$\sum_{k\in\mathbb{Z}} a_k \delta(\boldsymbol{x} - \boldsymbol{x}_k) \longrightarrow \boxed{\mathrm{L}^{-1}\{\cdot\}} \longrightarrow \quad s(\boldsymbol{x}) = p_{\mathrm{L}}(\boldsymbol{x}) + \sum_{k\in\mathbb{Z}} a_k \rho_{\mathrm{L}}(\boldsymbol{x} - \boldsymbol{x}_k)$$

Example of spline synthesis

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

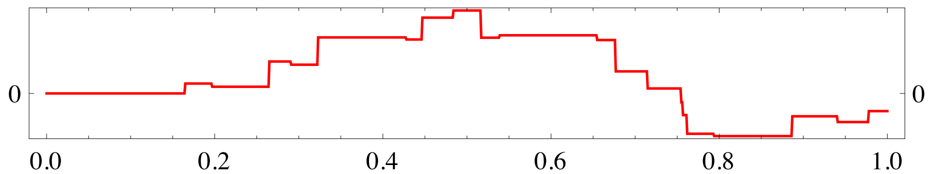


Compound Poisson process (sparse)

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

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

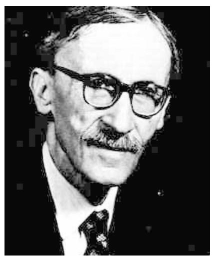
random stream of Diracs — **Compound Poisson process**



Random jumps with rate λ (Poisson point process)

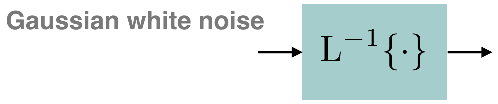
Jump size distribution: $a \backsim p(a)$

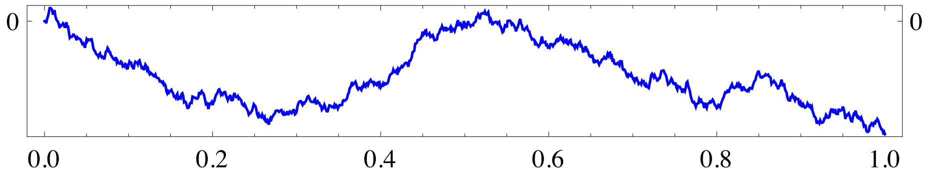
(Paul Lévy, 1934)



Brownian motion (Gaussian)

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

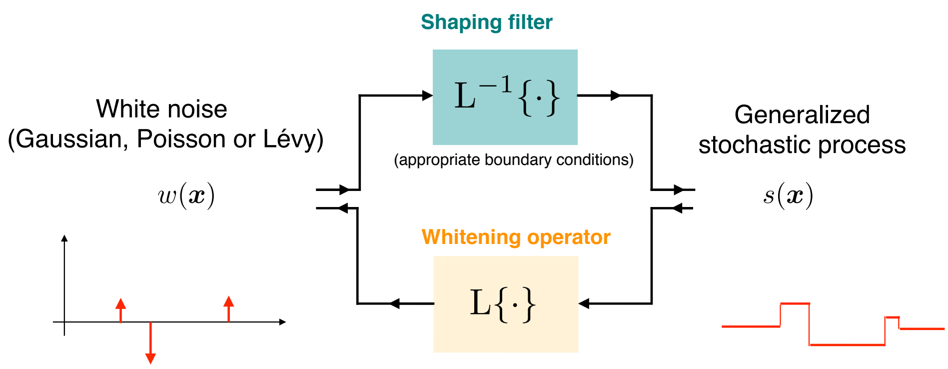




- Same higher-level properties as Compound Poisson process
  - Non-stationary
  - Self-similar: "$1/\omega$" spectral decay
  - Independent increments = defining property of Lévy processes

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

Except sparsity !



- 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.

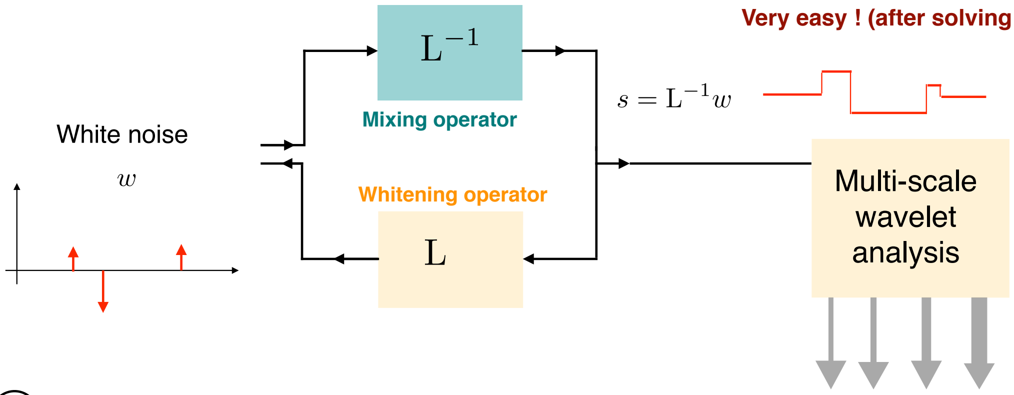
Road map for theory of sparse processes

(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

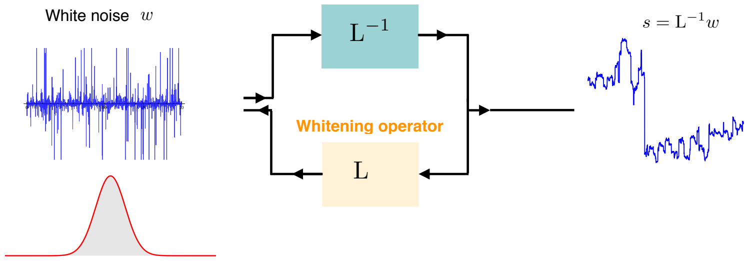
Gelfand's theory of *generalized stochastic processes*

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

  $\Rightarrow \quad X = \langle w, \varphi \rangle$ 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)

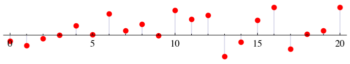
  Characteristic functional: $\widehat{\mathscr{P}}_w(\varphi) = \mathbb{E}\{e^{j\langle w, \varphi \rangle}\} = \int_{\mathcal{S}'} e^{j\langle s, \varphi \rangle} \mathscr{P}_w(ds),$ for any $\varphi \in \mathcal{S}$

- White noise property: independence at every point + stationarity

Defining Gaussian noise: discrete vs. continuous

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

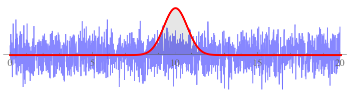
- Discrete white Gaussian noise



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

Characteristic function: $\hat{p}_X(\boldsymbol{\omega}) = g(\boldsymbol{\omega}) = \exp\left(\sum_{n=1}^{N} f(\omega_n)\right) = e^{-\frac{1}{2}\|\boldsymbol{\omega}\|^2}$

- Continuous-domain white Gaussian noise



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

Characteristic functional: $\widehat{\mathscr{P}}_s(\varphi) = G(\varphi) = e^{-\frac{1}{2}\|\varphi\|_{L_2}^2} = \exp\left(\int_{\mathbb{R}} f(\varphi(x))\mathrm{d}x\right)$

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

Characteristic form of white “noise” processes

- Functional characterization (Gelfand-Vilenkin)

The characteristic form

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

defines a white noise w over $\mathcal{S}'(\mathbb{R}^d)$

$\Leftrightarrow$ $f(\omega)$ is first-order conditionally-positive definite

(based on Schoenberg’s correspondence theorem
+ Minlos-Bochner theorem)

- Bottom line

- WNP uniquely specified by the function $f(\omega)$ (Lévy exponent)

Lévy exponent

- Definition

A continuous, complex-valued function $f : \mathbb{R} \to \mathbb{C}$ such that $f(0) = 0$ is a valid **Lévy exponent** if and only if $\hat{p}_{X_\tau}(\omega) = e^{\tau f(\omega)}$ is a valid characteristic function for any $\tau > 0$.

$\Leftrightarrow \quad \hat{p}_X(\omega) = e^{f(\omega)}$ is the characteristic function of an **infinitely divisible** random variable

Schoenberg's correspondence theorem
The function $e^{\tau f(\omega)}$ is positive-definite for any $\tau > 0$ if and only if $f(\omega)$ is **conditionally positive-definite of order one** ; i.e.,

$$\sum_{m=1}^{N}\sum_{n=1}^{N} f(\omega_m - \omega_n)\xi_m\overline{\xi}_n \geq 0$$

under the condition $\sum_{m=1}^{N} \xi_m = 0$ for every possible choice of $\omega_1, \dots, \omega_N \in \mathbb{R}$, $\xi_1, \dots, \xi_N \in \mathbb{C}$ and $N \in \mathbb{Z}^+$.

Example: $\quad f(\omega) = -|\omega|^\alpha, \quad 0 < \alpha \leq 2$

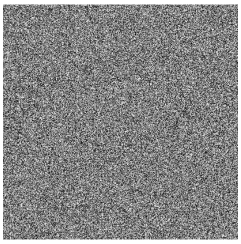
White noise: canonical distribution

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

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

- Interpretation: noise observation through a rectangular window

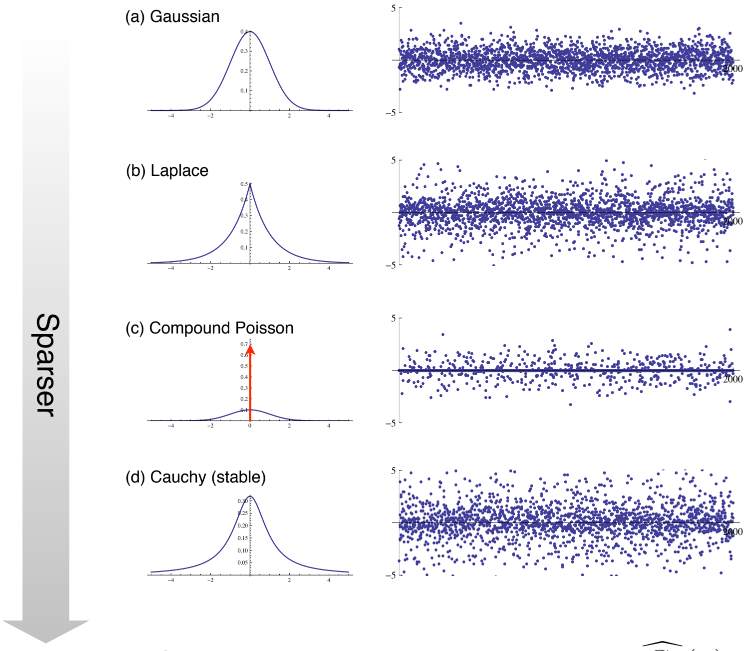
$\widehat{\mathscr{P}_w}(\omega\, \mathrm{rect}(\boldsymbol{x})) = e^{f(\omega)} \quad \Leftrightarrow \quad p_{\mathrm{id}}(x) = p_{X_{\mathrm{id}}}(x)$ with $X_{\mathrm{id}} = \langle \mathrm{rect}(\cdot - \boldsymbol{k}), w\rangle$ (i.i.d.)



- Special cases
  - $f(\omega) = -\frac{1}{2}|\omega|^2 \quad \Leftrightarrow \quad p_{\mathrm{id}}(x)$: normalized Gaussian
  - $f(\omega) = -|\omega|^\alpha$ with $\alpha \in (0, 2] \quad \Leftrightarrow \quad p_{\mathrm{id}}(x)$: Symmetric-$\alpha$-stable (S$\alpha$S)
  - 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: $\widehat{\mathscr{P}_w}(\varphi) = \exp\left(\int_{\mathbb{R}^d} f(\varphi(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}\right)$

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

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

where $b_1 \in \mathbb{R}$ and $b_2 \in \mathbb{R}^+$ are some constants, and where V is some (positive) Borel measure such that $\int_{\mathbb{R}} \min(a^2, 1) V(da) < \infty$.

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

Impulsive Poisson noise and random splines

- Impulsive Poisson noise

$$w_\delta(\boldsymbol{x}) = \sum_{k\in\mathbb{Z}} a_k \delta(\boldsymbol{x} - \boldsymbol{x}_k) \qquad \Rightarrow \quad \mathrm{L}^{-1} w_\delta \text{ is a L-spline with random knots}$$

$\boldsymbol{x}_k$: random point locations in $\mathbb{R}^d$ with Poisson density λ

a_k: i.i.d. random variables with amplitude pdf $p_A(a)$

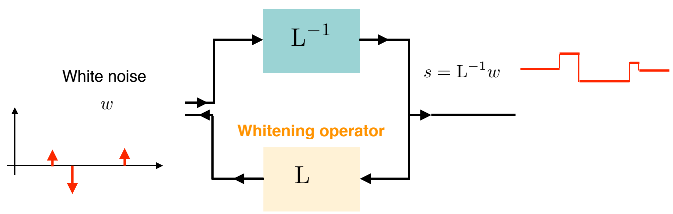
Theorem [U.-Tafti, IEEE-SP 2011]
The characteristic form of impulsive Poisson noise is

$$\widehat{\mathscr{P}}_{w_\delta}(\varphi) = \mathbb{E}\{e^{j\langle w_\delta, \varphi\rangle}\} = \exp\left(\int_{\mathbb{R}^d} f_{\mathrm{Poisson}}\big(\varphi(\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x}\right)$$

with Lévy exponent

$$f_{\mathrm{Poisson}}(\omega) = \lambda \int_{\mathbb{R}} (e^{ja\omega} - 1) p_A(a)\mathrm{d}a = \lambda(\hat{p}_A(\omega) - 1).$$

Steps 2 + 3: Characterization of sparse process



- Abstract formulation of innovation model

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

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

Mathematical conditions on L^{-1*} ?

Existence result

$\mathrm{T} = \mathrm{L}^{-1*}$

Theorem [U.-Tafti-Sun, preprint]
Let f is a valid Lévy exponent and T is a linear operator acting on $\varphi \in \mathcal{S}(\mathbb{R}^d)$ such that any one of the conditions below is met:

1. T is a continuous map from $\mathcal{S}(\mathbb{R}^d)$ into itself or, by extension, $\mathcal{R}(\mathbb{R}^d)$,
2. $|f(\omega)| + |\omega| \cdot |f'(\omega)| \leq B|\omega|^p$ for some $1 \leq p < \infty$ and for all $\omega \in \mathbb{R}$, and $\|\mathrm{T}\varphi\|_{L_p} \leq C\|\varphi\|_{L_p}$ for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$.

Then, $\widehat{\mathscr{P}}_s(\varphi) = \mathbb{E}\{e^{j\langle s,\varphi\rangle}\} = \exp\left(\int_{\mathbb{R}} f\big(\mathrm{T}\varphi(t)\big)\mathrm{d}t\right)$ is the characteristic form of a generalized stochastic process over $\mathcal{S}'(\mathbb{R}^d)$.

- Implication for innovation model

Find an acceptable (L_p stable) inverse operator: $\mathrm{T} = \mathrm{L}^{-1*}$

Concrete example: (f)Brownian motion

$\mathrm{D}s = w$ (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: $\mathrm{D}_0^{-1*}\varphi(x) = \int_{\mathbb{R}} \frac{\hat{\varphi}(\omega) - \hat{\varphi}(0)}{-j\omega} e^{j\omega x} \frac{\mathrm{d}\omega}{2\pi}$

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

Stabilization ⇔ non-stationary behavior

$$\widehat{\mathscr{P}}_s(\varphi) = \exp\left(-\frac{1}{2}\|\mathrm{D}_0^{-1*}\varphi\|_{L_2}^2\right) = \exp\left(-\frac{1}{2}\int_{\mathbb{R}} \left|\frac{\hat{\varphi}(\omega) - \hat{\varphi}(0)}{j\omega}\right|^2 \frac{\mathrm{d}\omega}{2\pi}\right)$$

(by Parseval)

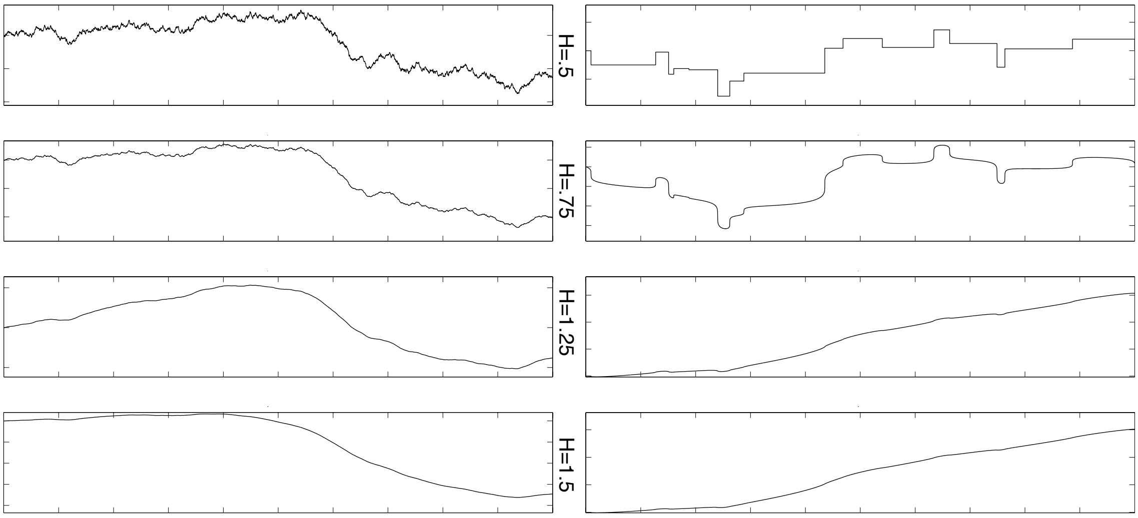
- Characteristic form of fractional Brownian motion

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

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

Self-similar processes (TS-invariant)

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

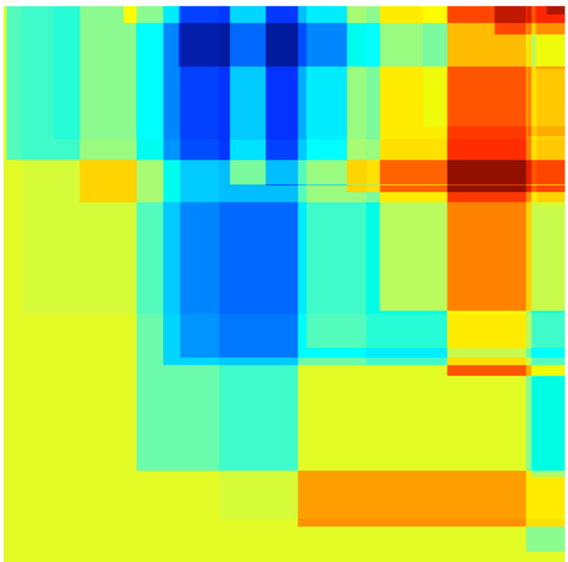


Gaussian — **Sparse (generalized Poisson)**

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

2D generalization: the Mondrian process

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



$\lambda = 30$

Scale- and rotation-invariant operators

Definition: An operator L is scale- and rotation-invariant iff.

$$\forall s(\boldsymbol{x}),\ \mathrm{L}\{s(\cdot)\}(\mathbf{R}_\theta \boldsymbol{x}/a) = C_a \cdot \mathrm{L}\{s(\mathbf{R}_\theta \cdot /a)\}(\boldsymbol{x})$$

where $\mathbf{R}_\theta$ is an arbitrary $d \times d$ unitary matrix and C_a a constant

- **Invariance theorem**

The complete family of real, scale- and rotation-invariant convolution operators is given by the fractional Laplacians

$$(-\Delta)^{\frac{\gamma}{2}} \quad \stackrel{\mathcal{F}}{\longleftrightarrow} \quad \|\boldsymbol{\omega}\|^\gamma$$

- Invariant Green functions (a.k.a. RBF) (Duchon, 1979)

$$\rho(\boldsymbol{x}) = \begin{cases} \|\boldsymbol{x}\|^{\gamma-d} \log \|\boldsymbol{x}\|, & \text{if } \gamma - d \text{ is even} \\ \|\boldsymbol{x}\|^{\gamma-d}, & \text{otherwise} \end{cases}$$

Inverse operators: fractional calculus

L	$\rho = \mathrm{L}^{-1}\delta$	$\mathrm{L}^{-1}\varphi$	$\mathrm{L}^{-1*}\varphi$	$0 < \gamma < 1 + d/2$
D^γ	$\frac{x_+^{\gamma-1}}{\Gamma(\gamma)}$	$\int_{\mathbb{R}} \frac{e^{j\omega x} - 1}{(j\omega)^\gamma} \hat{\varphi}(\omega) \frac{\mathrm{d}\omega}{2\pi}$	$\int_{\mathbb{R}} e^{j\omega x} \frac{\hat{\varphi}(\omega) - \hat{\varphi}(0)}{(-j\omega)^\gamma} \frac{\mathrm{d}\omega}{2\pi}$	
∂_τ^γ	$\frac{\lvert x\rvert^{\gamma-1}}{\Gamma(\gamma)} \left(A_{\gamma,\tau} + B_{\gamma,\tau}\mathrm{sign}(x)\right),\ \gamma \notin \mathbb{N}$	$\int_{\mathbb{R}} \frac{e^{j\omega x} - 1}{(-j\omega)^{\frac{\gamma}{2}-\tau}(j\omega)^{\frac{\gamma}{2}+\tau}} \hat{\varphi}(\omega) \frac{\mathrm{d}\omega}{2\pi}$	$\int_{\mathbb{R}} e^{j\omega x} \frac{\hat{\varphi}(\omega) - \hat{\varphi}(0)}{(j\omega)^{\frac{\gamma}{2}-\tau}(-j\omega)^{\frac{\gamma}{2}+\tau}} \frac{\mathrm{d}\omega}{2\pi}$	
$(-\Delta)^{\frac{\gamma}{2}}$	$C_\gamma \|\boldsymbol{x}\|^{\gamma-d}, \gamma - d \notin 2\mathbb{N}$	$\int_{\mathbb{R}^d} \frac{e^{j\langle \boldsymbol{x},\boldsymbol{\omega}\rangle} - 1}{\|\boldsymbol{\omega}\|^\gamma} \hat{\varphi}(\boldsymbol{\omega}) \frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^d}$	$\int_{\mathbb{R}^d} e^{j\langle \boldsymbol{x},\boldsymbol{\omega}\rangle} \frac{\hat{\varphi}(\boldsymbol{\omega}) - \hat{\varphi}(\mathbf{0})}{\|\boldsymbol{\omega}\|^\gamma} \frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^d}$	

(U.-Tafti, 2011)

Theorem (Generalized Riesz potentials)

Unique left-inverse of $\mathrm{L}^* = (-\Delta)^{\frac{\gamma}{2}}$ that is L_p-stable and scale-invariant:

$$\mathrm{I}_p^\gamma \varphi(\boldsymbol{x}) = \int_{\mathbb{R}^d} e^{j\langle \boldsymbol{x},\boldsymbol{\omega}\rangle} \frac{\hat{\varphi}(\boldsymbol{\omega}) - \sum_{|\boldsymbol{k}|=0}^{\lfloor \gamma - d + \frac{d}{p} \rfloor} \hat{\varphi}^{(\boldsymbol{k})}(\mathbf{0}) \frac{\boldsymbol{\omega}^{\boldsymbol{k}}}{\boldsymbol{k}!}}{\|\boldsymbol{\omega}\|^\gamma} \frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^d} = \mathrm{L}^{-1*}\varphi(\boldsymbol{x})$$

$$\forall \varphi \in \mathcal{S}(\mathbb{R}^d), \qquad \|\mathrm{I}_p^\gamma \varphi\|_{L_p(\mathbb{R}^d)} < C \cdot \|\varphi\|_{L_p(\mathbb{R}^d)}$$

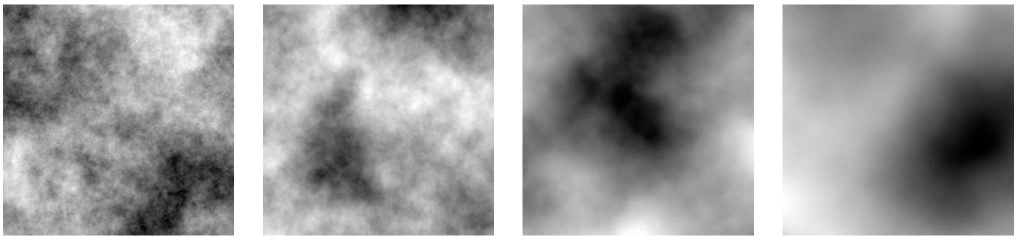
for $\gamma \in \mathbb{R}^+ \backslash \mathbb{Z}^+$ and $1 \le p \le +\infty$.

(Sun-U., 2012)

Scale- and rotation-invariant processes

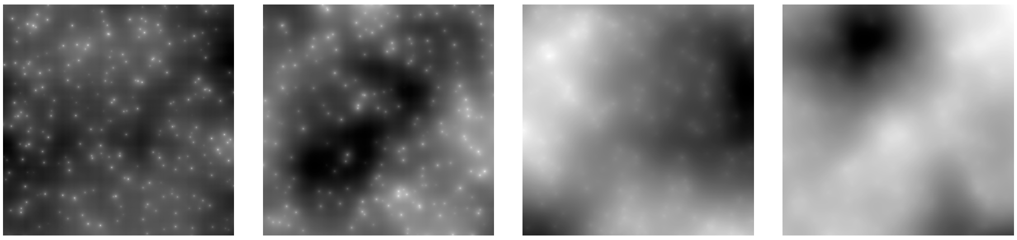
Stochastic partial differential equation : $(-\Delta)^{\frac{H+1}{2}} s(\boldsymbol{x}) = w(\boldsymbol{x})$

Gaussian



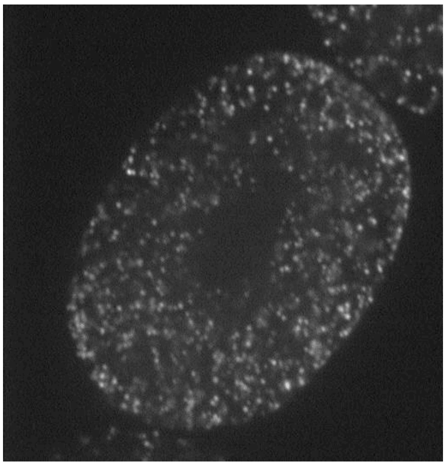
H=.5 H=.75 H=1.25 H=1.75

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
- Wavelet analysis
- Link with regularization
- Signal reconstruction algorithm (MAP)

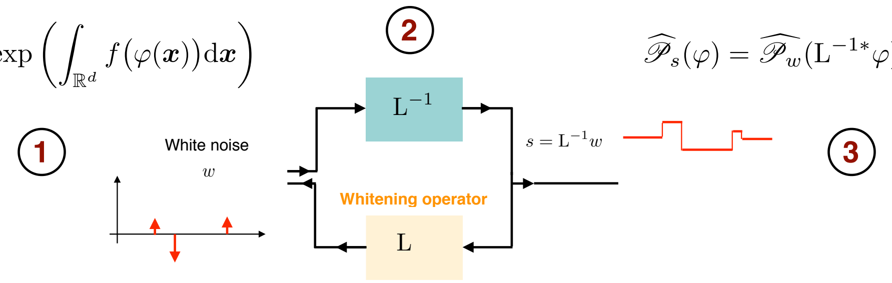
Recap on infinite-dimensional innovation model

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

Solution of SDE

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

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



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 $\left(\beta_{\mathrm{L}} \in L_1(\mathbb{R}^d)\right)$ 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

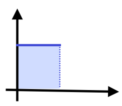
$$\begin{aligned} \mathrm{D}s(x) &= w(x) \\ s(x) &= \int_0^x w(y)\mathrm{d}y \end{aligned}$$

- 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

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

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

$$\Longrightarrow \quad \begin{aligned} \widehat{\mathscr{P}}_u(\varphi) &= \widehat{\mathscr{P}}_w(\beta_{\mathrm{L}}^{\vee} * \varphi) \\ &= \exp\left(\int_{\mathbb{R}^d} f\big(\beta_{\mathrm{L}}^{\vee} * \varphi(\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x}\right) \end{aligned}$$

$f(\omega)$: Lévy exponent of innovation process

Infinite-divisibility of discrete innovation

- Signal decoupling: discrete version of operator

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

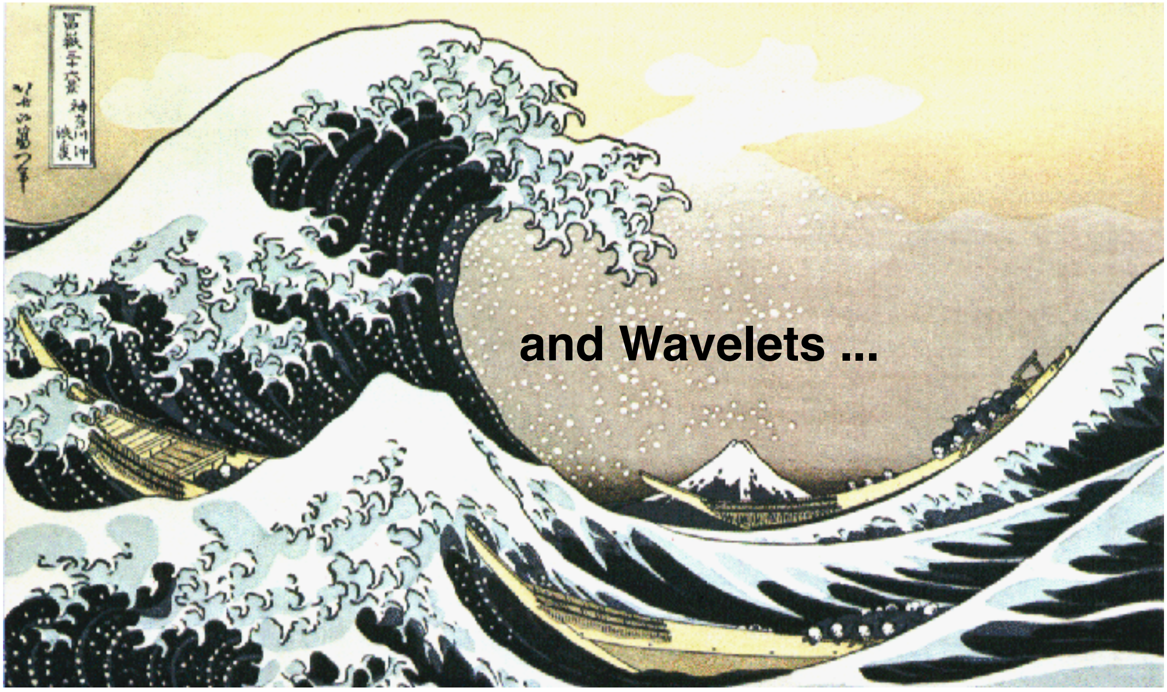
- Characteristic form of $u = \beta_{\mathrm{L}} * w$

$$\mathbb{E}\{e^{j\langle u, \varphi\rangle}\} = \widehat{\mathscr{P}}_u(\varphi) = \exp\left(\int_{\mathbb{R}^d} f\big(\beta_{\mathrm{L}}^{\vee} * \varphi(\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x}\right)$$

- Statistical implications
  - $u = \mathrm{L_d}s$ is stationary and **infinitely divisible**
  - Characteristic function of $X = \langle u, \varphi\rangle$ with $\varphi = \delta$:

$$\hat{p}_X(\omega) = \mathbb{E}\{e^{j\omega X}\} = \widehat{\mathscr{P}}_u(\omega\delta) = \exp\left(\int_{\mathbb{R}^d} f\big(\omega\beta_{\mathrm{L}}(-\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x}\right)$$

  - Quality of decoupling depends upon support of B-spline $\beta_{\mathrm{L}}(\boldsymbol{x})$



Wavelet analysis of sparse processes

- Innovation model (SDE)

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

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

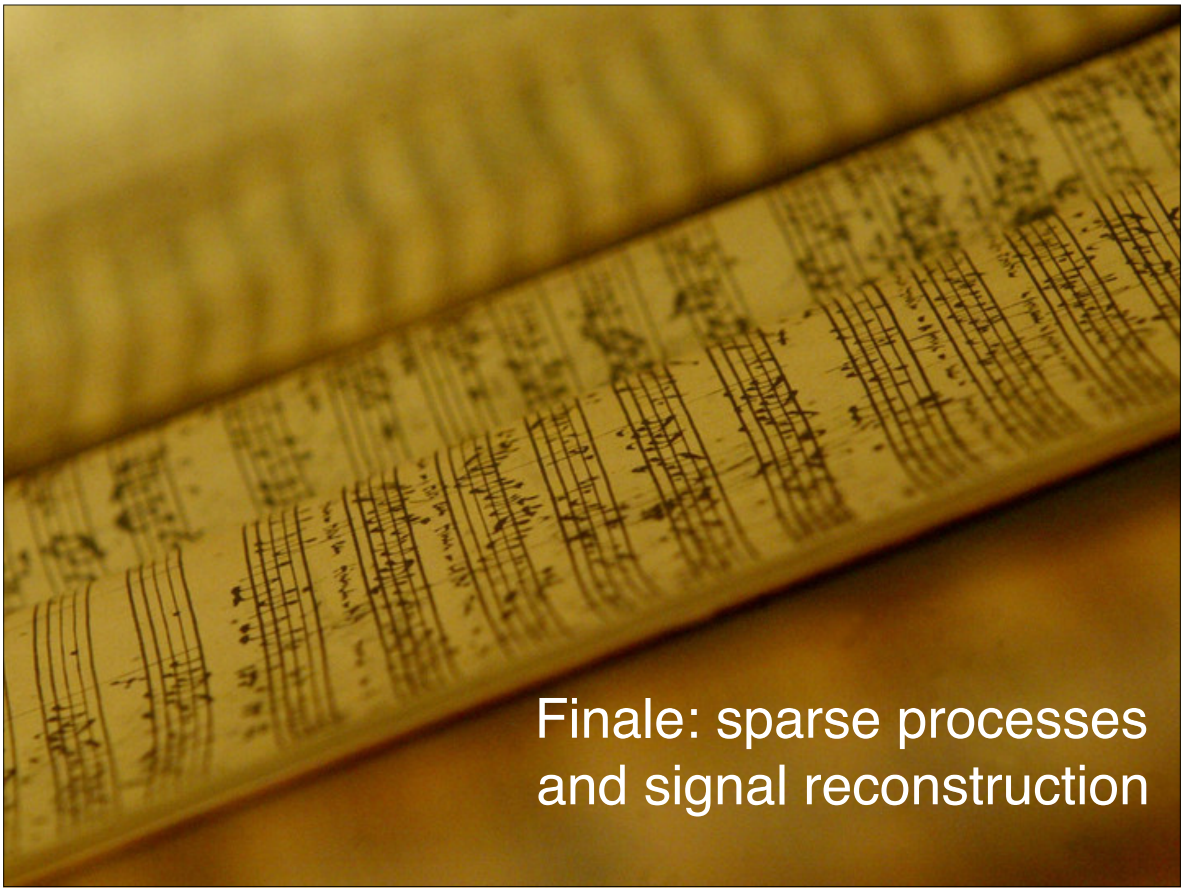
  ϕ_i: smoothing kernel at wavelet scale i

- 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



Signal reconstruction: MAP formulation

- Innovation model

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

Discretization → $\mathbf{u} = \mathbf{L}\mathbf{s}$ (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

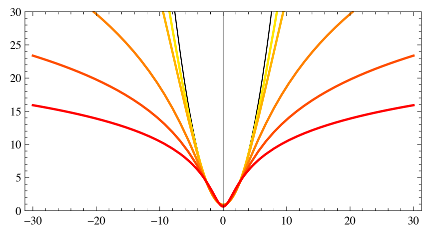
$$\mathbf{s}^{\star} = \arg\min \left(\frac{1}{2}\|\mathbf{g} - \mathbf{H}\mathbf{s}\|_2^2 + \sigma^2 \sum_n \Phi_X([\mathbf{L}\mathbf{s}]_n)\right)$$

MAP estimator: special cases

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

Sparser

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



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

Reconstruction algorithms

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

- Constrained optimization formulation

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

$$\mathbf{s}^* = \underset{\mathbf{s}}{\mathrm{argmin}}\left(\frac{1}{2}\|\mathbf{g}-\mathbf{H}\mathbf{s}\|_2^2 + \sigma^2 \sum_n \Phi_X([\mathbf{u}]_n)\right) \quad \text{s.t.} \quad \mathbf{u} = \mathbf{L}\mathbf{s}$$

- Augmented Lagrangian — AL/ADM (Ramani-Fessler TMI 2011)

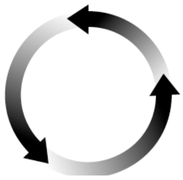
  Innovation variable: $\mathbf{u} = \mathbf{L}\mathbf{s}$

$$\mathcal{L}_{\mathcal{A}}(\mathbf{s}, \mathbf{u}, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{g}-\mathbf{H}\mathbf{s}\|_2^2 + \lambda\left(\sum_n \Phi_X([\mathbf{u}]_n) + \boldsymbol{\alpha}^T(\mathbf{L}\mathbf{s}-\mathbf{u}) + \frac{\mu}{2}\|\mathbf{L}\mathbf{s}-\mathbf{u}\|_2^2\right)$$

Alternating direction method (ADM)

$$\mathcal{L}_{\mathcal{A}}(\mathbf{s},\mathbf{u},\boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{g}-\mathbf{Hs}\|_2^2 + \lambda\left(\sum_n \Phi_X([\mathbf{u}]_n) + \boldsymbol{\alpha}^T(\mathbf{Ls}-\mathbf{u}) + \frac{\mu}{2}\|\mathbf{Ls}-\mathbf{u}\|_2^2\right)$$

Sequential minimization



$$\mathbf{s}^{k+1} \leftarrow \arg\min_{\mathbf{s}\in\mathbb{R}^N} \mathcal{L}_{\mathcal{A}}(\mathbf{s},\mathbf{u}^k,\boldsymbol{\alpha}^k)$$

$$\mathbf{u}^{k+1} \leftarrow \arg\min_{\mathbf{u}\in\mathbb{R}^N} \mathcal{L}_{\mathcal{A}}(\mathbf{s}^{k+1},\mathbf{u},\boldsymbol{\alpha}^k)$$

$$\boldsymbol{\alpha}^{k+1} = \boldsymbol{\alpha}^k - \mu\left(\mathbf{u}^{k+1} - \mathbf{Ls}^{k+1}\right)$$

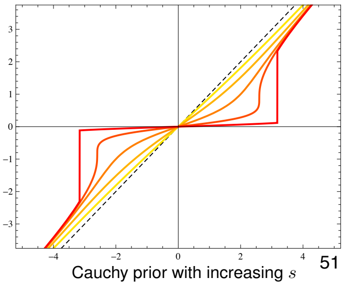
Linear inverse problem: $\mathbf{s}^{k+1} = \left(\mathbf{H}^T\mathbf{H} + \mu\mathbf{L}^T\mathbf{L}\right)^{-1}\left(\mathbf{H}^T\mathbf{y} + \mathbf{z}^{k+1}\right)$

with $\mathbf{z}^{k+1} = \mathbf{L}^T\left(\mu\mathbf{u}^{k+1} - \boldsymbol{\alpha}\right)$

Nonlinear denoising: $\mathbf{u}^{k+1} = \mathrm{prox}_{\Phi_X}\left(\mathbf{Ls}^k + \boldsymbol{\alpha}^k; \lambda\mu^{-1}\right)$

- Proximal operator taylored to stochastic model

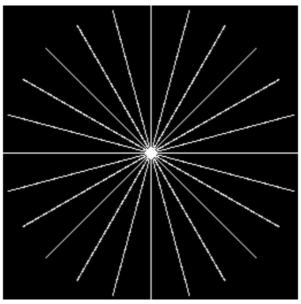
$$\mathrm{prox}_{\Phi_X}(y;\lambda) = \arg\min_u \frac{1}{2}|y-u|^2 + \lambda\Phi_X(u)$$



Cauchy prior with increasing s

MRI: Shepp-Logan phantom

Original SL Phantom



Fourier Sampling Pattern
12 Angles

Laplace prior (TV)

Student prior (log)

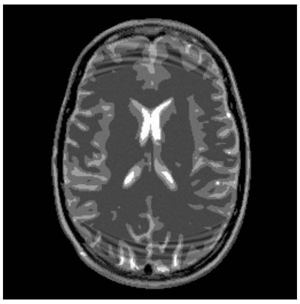
L : gradient
Optimized parameters

MRI phantom: Spiral sampling in k-space

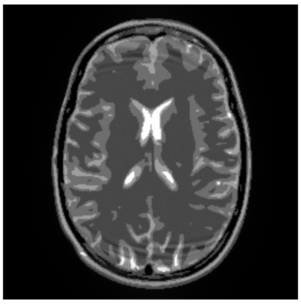
L : gradient

Optimized parameters

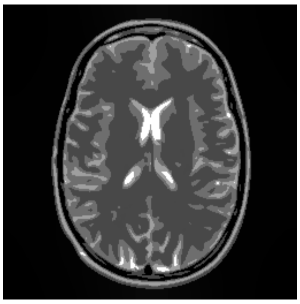
Original Phantom
(Guerquin-Kern TMI 2012)



Gaussian prior (Tikhonov)
SER =17.69 dB

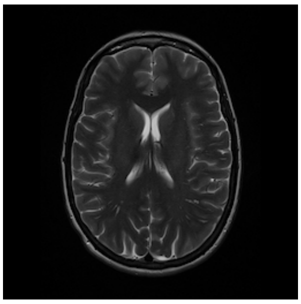


Laplace prior (TV)
SER = 21.37 dB



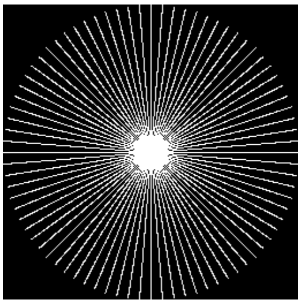
Student prior
SER = 27.22 dB

MRI reconstruction



Real T2 Brain Image

MR Angiography Image



k-space sampling pattern

40 radial lines

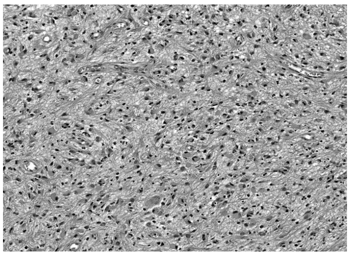
L : gradient

Optimized parameters

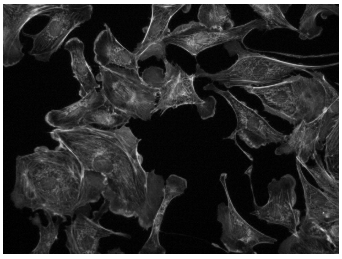
Reconstruction results in dB

	Gaussian Estimator	Laplace Estimator	Student's Estimator
T2 brain Image	8.71	**16.08**	15.79
MR Angiography Image	6.31	14.48	**14.97**

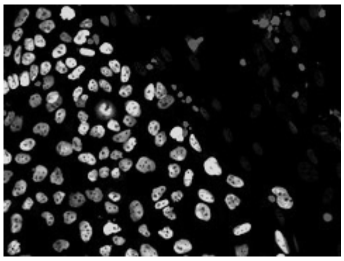
2D deconvolution experiment



Astrocytes cells



bovine pulmonary artery cells



human embryonic stem cells

Disk shaped PSF (7x7)

L : gradient

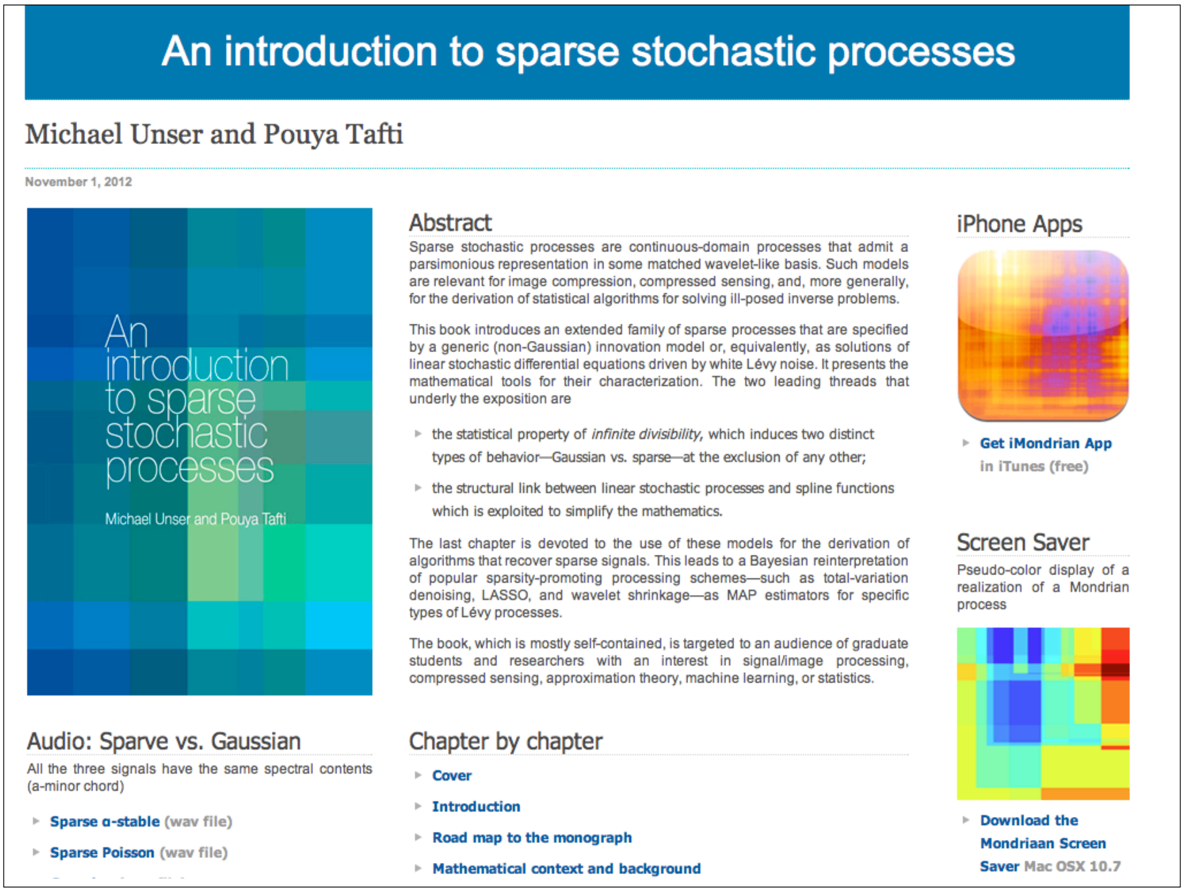
Deconvolution results in dB

Optimized parameters

	Gaussian Estimator	Laplace Estimator	Student's Estimator
Astrocytes cells	**12.18**	10.48	10.52
Pulmonary cells	16.90	**19.04**	18.34
Stem cells	15.81	20.19	**20.50**

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
  - Central role of B-spline
  - Sparsification via “operator-like” behavior
- Theoretical framework for sparse signal recovery
  - New statistically-founded sparsity priors
  - Analytical determination of PDF in **any** transformed domain
  - Derivation of optimal estimators (MAP, MMSE)
  - Guide for the development of novel algorithms



References

- Theory of generalized stochastic processes

  - M. Unser and P. Tafti, *An introduction to sparse stochastic processes*, e-book at `http://www.sparseprocesses.org`.
  - 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.
  - Q. Sun, M. Unser, "Left-Inverses of Fractional Laplacian and Sparse Stochastic Processes," *Advances in Computational Mathematics*, vol. 36, no. 3, pp. 399-441, April 2012.
  - 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

  - E. Bostan, U.S. Kamilov, M. Nilchian, M. Unser, "Sparse Stochastic Processes and Discretization of Linear Inverse Problems," preprint, available at http://arXiv:1210.5839
  - 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.

Acknowledgments

Many thanks to

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

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

erc

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