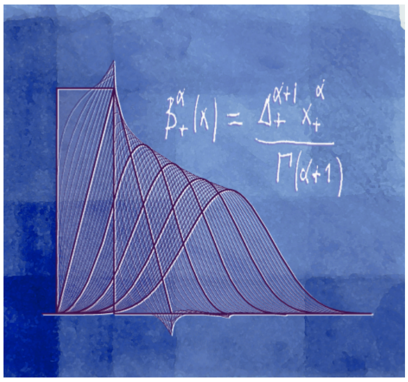
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Sparse stochastic processes: A statistical framework for compressed sensing and biomedical image reconstruction

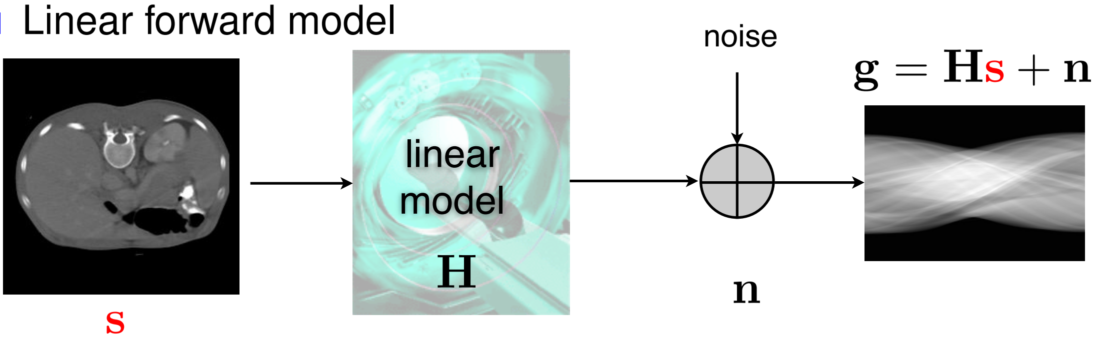
Michael Unser
Biomedical Imaging Group
EPFL, Lausanne, Switzerland



Tutorial, Inverse Problems and Imaging Conference, Institut Henri Poincaré, Paris, April 7-11, 2014.

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

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

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

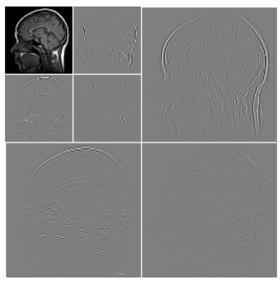
  Wiener (LMMSE) solution = Gauss MMSE = Gauss MAP

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

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)

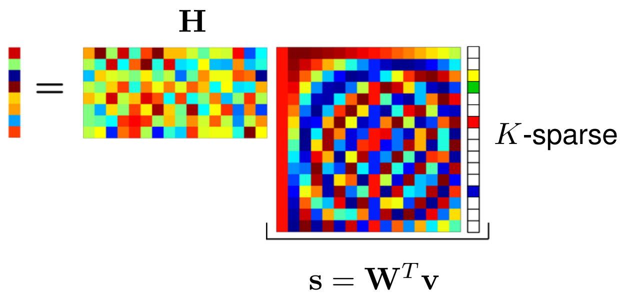
- ℓ_1 regularization (Total variation)

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

  Iterative reweighted least squares (IRLS) or FISTA

Theory of compressive sensing

- Generalized sampling setting (after discretization)
  - Linear inverse problem: $\mathbf{u} = \mathbf{H}\mathbf{s} + \mathbf{n}$
  - $N_u \times N_v$ system matrix : $\mathbf{A} = \mathbf{H}\mathbf{W}^T$



- Formulation of ill-posed recovery problem when $2K < N_u \ll N_v$

(P0) $\min_{\mathbf{v}} \|\mathbf{u} - \mathbf{A}\mathbf{v}\|_2^2$ subject to $\|\mathbf{v}\|_0 \leq K$

- Theoretical result

Under suitable conditions on $\mathbf{A}$ (e.g., restricted isometry), the solution is unique and the recovery problem (P0) is equivalent to:

(P1) $\min_{\mathbf{v}} \|\mathbf{u} - \mathbf{A}\mathbf{v}\|_2^2$ subject to $\|\mathbf{v}\|_1 \leq C_1$

[Donoho et al., 2005
Candès-Tao, 2006, ...]

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, ADMM

OUTLINE

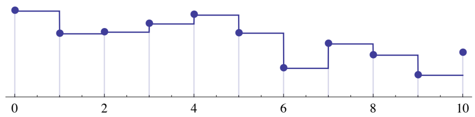
- Variational formulation of inverse problems ✔
- **Part I: Statistical modeling**
  An introduction to sparse stochastic processes
  - Generalized innovation model
  - Statistical characterization of signal
- **Part II: Recovery of sparse signals**
  Reconstruction of biomedical images
  - Discretization of inverse problem
  - Generic MAP estimator (iterative reconstruction algorithm)
  - Applications

Deconvolution microscopy
Magnetic resonance imaging
X-ray tomography
Phase-contrast tomography

An introduction to sparse stochastic processes

Splines and Legos revisited

- Cardinal spline of degree 0: piecewise-contant



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

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

Notion of D-spline:

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

B-spline and derivative operator

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

Finite difference operator

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

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

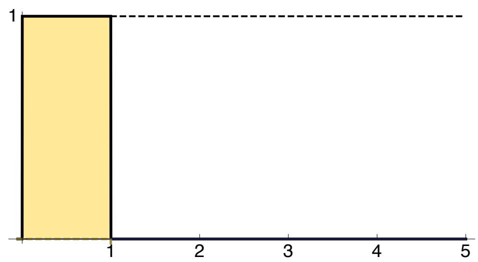
B-spline of degree 0

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

$$\updownarrow$$

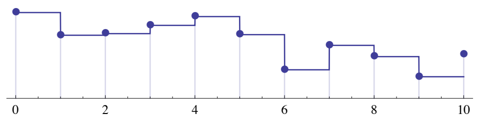
$$\hat{\beta}_+^0(\omega) = \frac{1 - \mathrm{e}^{-\mathrm{j}\omega}}{\mathrm{j}\omega}$$

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

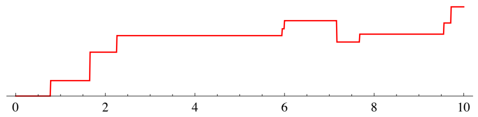


Random spline: archetype of sparse signal

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)

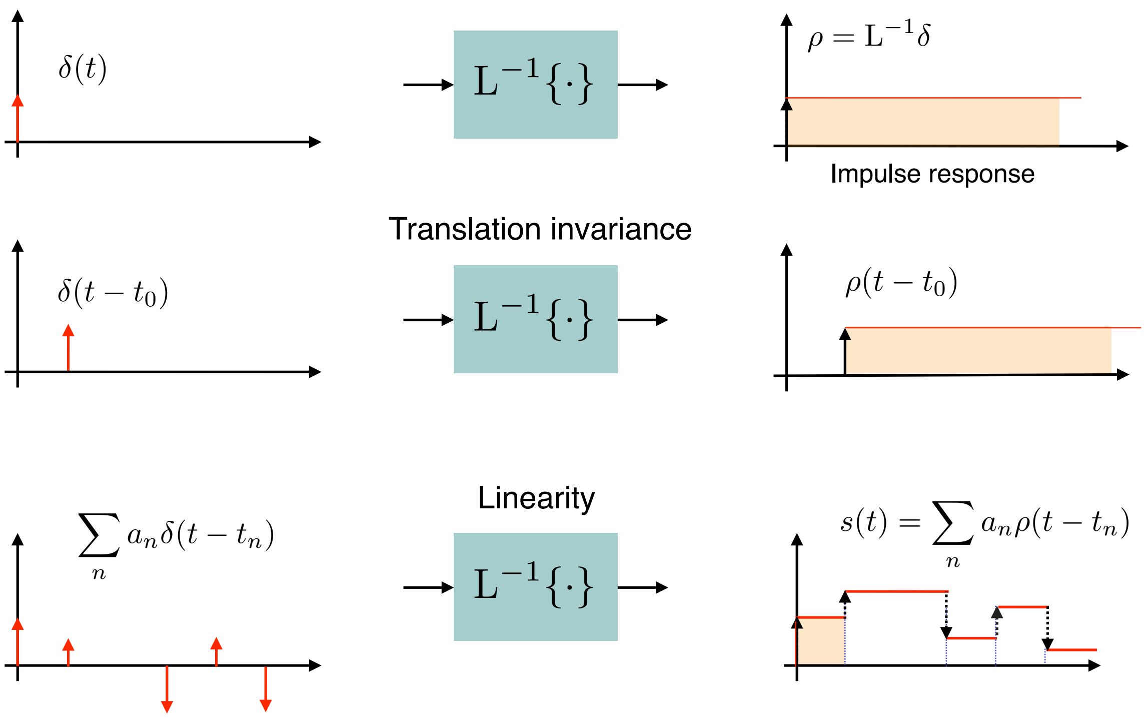
- Anti-derivative operators

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

Scale-invariant solution: $\mathrm{D}_0^{-1}\varphi(t) = \int_0^t \varphi(\tau)\mathrm{d}\tau$

Innovation-based synthesis

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

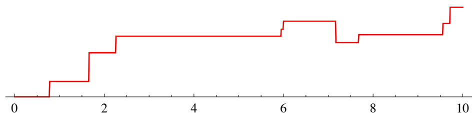


Compound Poisson process

- 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

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

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

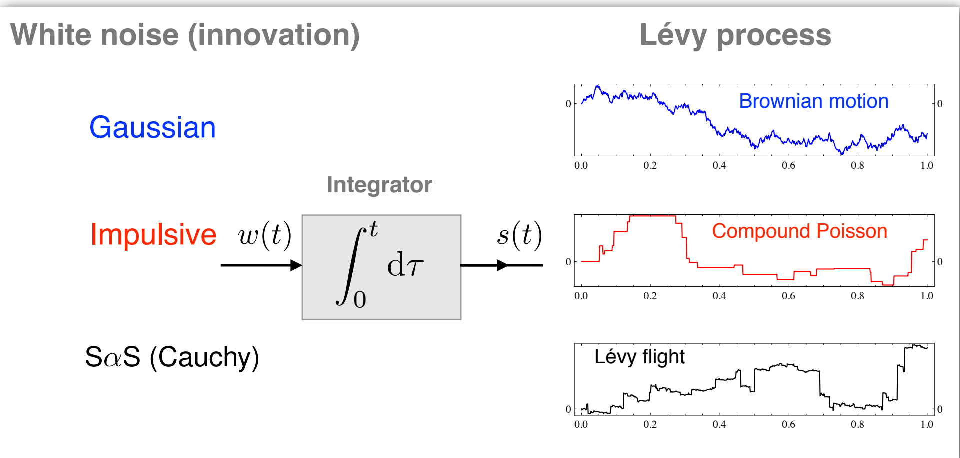
(impose boundary condition)

Lévy processes: all admissible brands of innovations

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

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

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



(Wiener 1923)

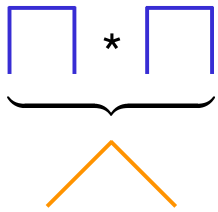
(Paul Lévy circa 1930)

Decoupling Lévy processes: increments

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

Increment process is stationary with autocorrelation function

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



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

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

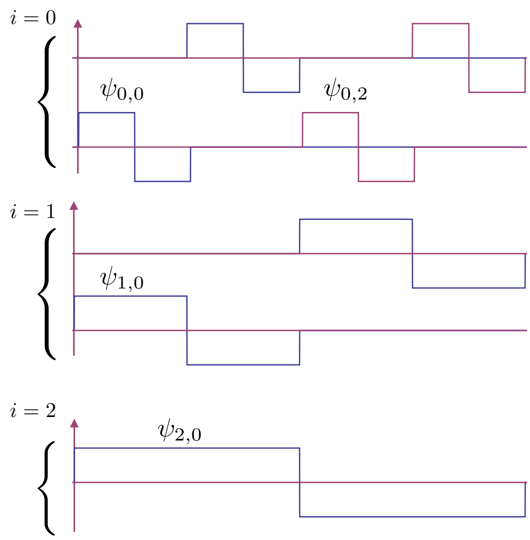
- $\{\beta_+^0(\cdot - k)\}$ are non-overlapping
- w is independent at every point (white noise)

Wavelet analysis of Lévy processes

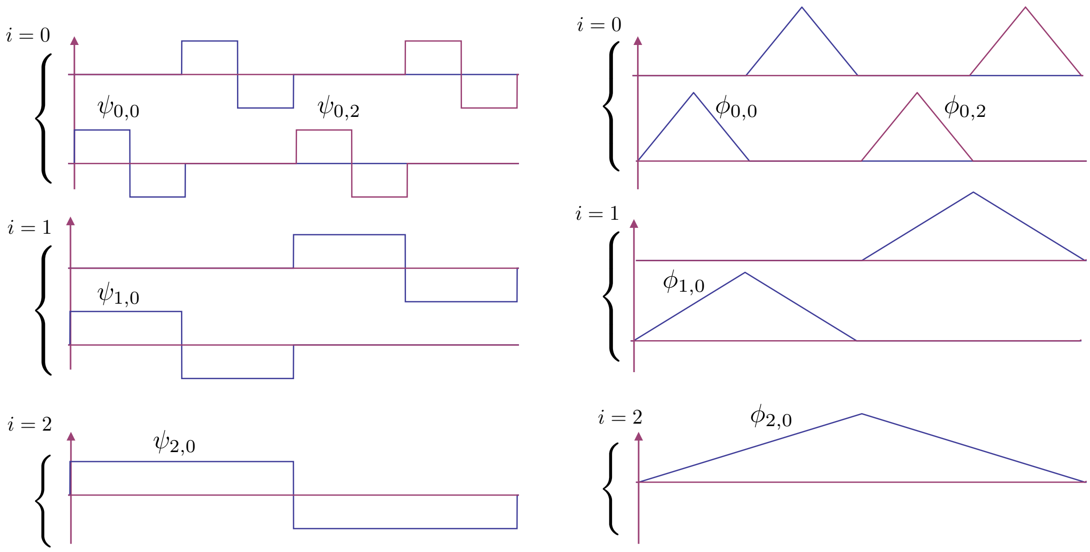
- Haar wavelets

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

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



Wavelets as multi-scale derivatives



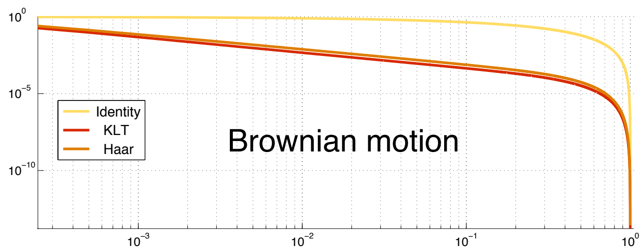
$$\psi_{i,k} = 2^{i/2-1}\mathrm{D}\phi_{i,k}$$
$$\mathrm{D}_0^{-1}\psi_{i,k} = 2^{i/2-1}\phi_{i,k}.$$

- Wavelet coefficients of Lévy process

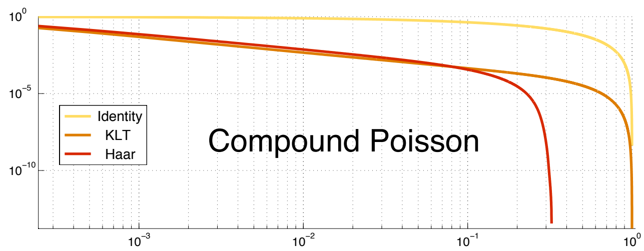
$$Y_{i,k} = \langle s, \psi_{i,k}\rangle \propto \langle s, \mathrm{D}\phi_{i,k}\rangle$$
$$\propto \langle \mathrm{D}^{*}s, \phi_{i,k}\rangle = -\langle w, \phi_{i,k}\rangle$$

M-term approximation: wavelets vs. KLT

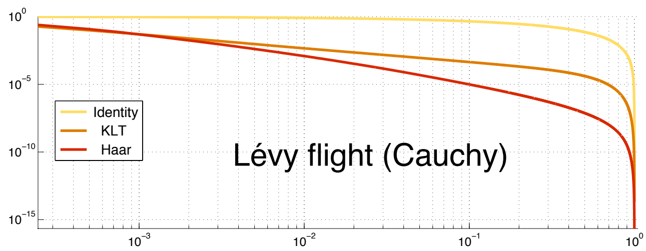
Gaussian



Finite rate of innovation



Even sparser ...



Generalized innovation model

Theoretical framework: Gelfand's theory of generalized stochastic processes

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

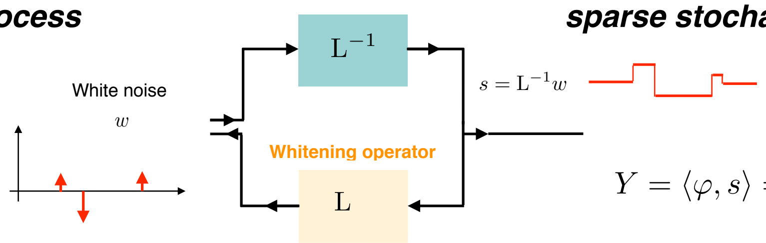
Solution of SDE

(2)

innovation process

(1)

$X = \langle \varphi, w \rangle$



sparse stochastic process

(3)

$Y = \langle \varphi, s \rangle = \langle \mathrm{L}^{-1*}\varphi, w \rangle$

Proper definition of
continuous-domain white noise

(4) Analysis step

Regularization operator vs. wavelet analysis

(Unser et *al*, *IEEE-IT* 2014)

Short primer on probability theory

Random variable X

- Probability measure and density function (pdf)

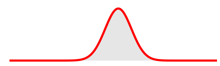
$$\mathrm{Prob}(X \in E) = \mathscr{P}_X(E) = \int_E p_X(x)\mathrm{d}x$$

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

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

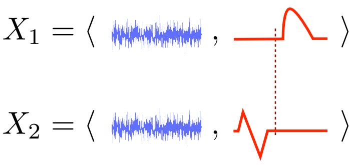
Generalized innovation process

- 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.

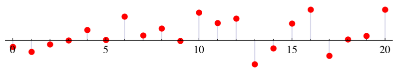
$\mathcal{S}$: Schwartz' space of smooth (infinitely differentiable) and rapidly decaying functions

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

Defining Gaussian noise: discrete vs. continuous

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

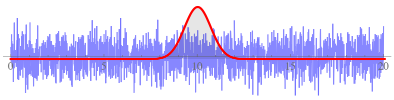
- **Discrete white Gaussian noise**



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

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

- **Continuous-domain white Gaussian noise**



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

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

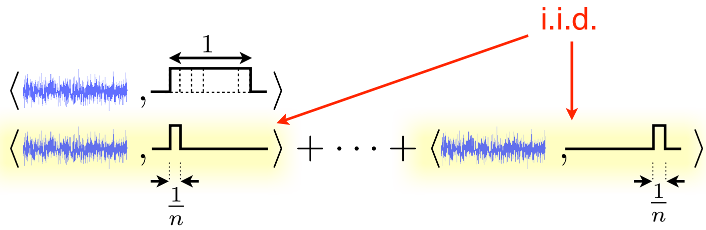
$$\hat{p}_{X_n}(\omega) = \mathbb{E}\{e^{j\omega\langle \varphi_n, w\rangle}\} = \mathbb{E}\{e^{j\langle \omega\varphi_n, w\rangle}\} = \widehat{\mathscr{P}}_w(\omega\varphi_n) = e^{-\frac{1}{2}\omega^2\|\varphi_n\|_{L_2}^2}$$

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 $X \overset{d}{=} X_1 + \cdots + X_N$.

- Rectangular test function

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



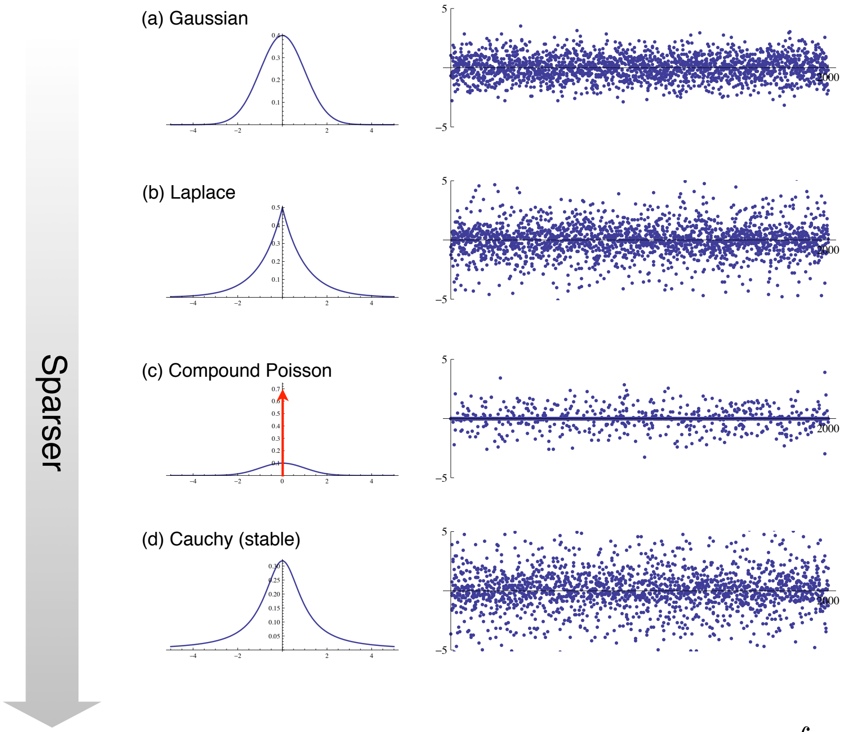
Proposition

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

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



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

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

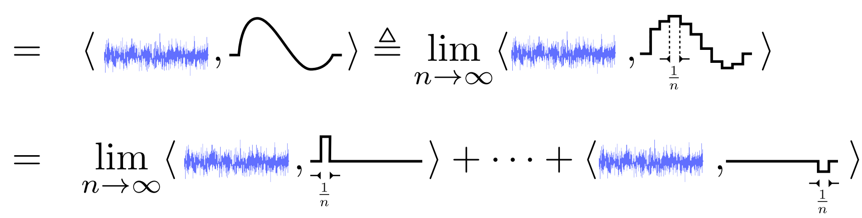
$$p_{\mathrm{Poisson}}(x) = \mathcal{F}^{-1}\{\mathrm{e}^{\lambda(\hat{p}_A(\omega)-1)}\}$$

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

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

Characterization of generalized innovation

$X_\varphi = \langle w, \varphi \rangle$



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)

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

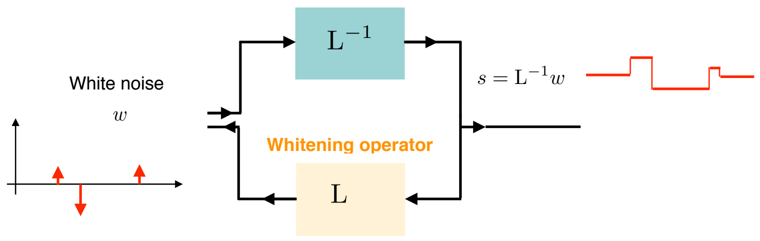
$$\widehat{p}_{\mathrm{id}}(\omega) = \int_{\mathbb{R}} p_{\mathrm{id}}(x)\mathrm{e}^{\mathrm{j}\omega x}\mathrm{d}x = \exp\left(f(\omega)\right)$$

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}.

Steps 2 + 3: Characterization of sparse process



- Abstract formulation of innovation model

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

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

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)

Innovation model: statistical implications (id)

- Statistical description of white Lévy noise w (innovation)
  - Characterized by canonical (p-admissible) Lévy exponent $f(\omega)$
  - Generic observation: $X = \langle \varphi, w \rangle$ with $\varphi \in L_p(\mathbb{R}^d)$
  - X is ***infinitely divisible*** with (modified) Lévy exponent

$$f_\varphi(\omega) = \log \widehat{p}_X(\omega) = \int_{\mathbb{R}^d} f\big(\omega\varphi(\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x}$$

- Linear observation of generalized stochastic process

$$s = \mathrm{L}^{-1}w \quad \Leftrightarrow \quad \langle \psi, s \rangle = \langle \psi, \mathrm{L}^{-1}w \rangle = \langle \underbrace{\mathrm{L}^{-1*}\psi}, w \rangle$$

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

Example 1: (f)Brownian motion

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

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

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

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

Stabilization ⇔ non-stationary behavior

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

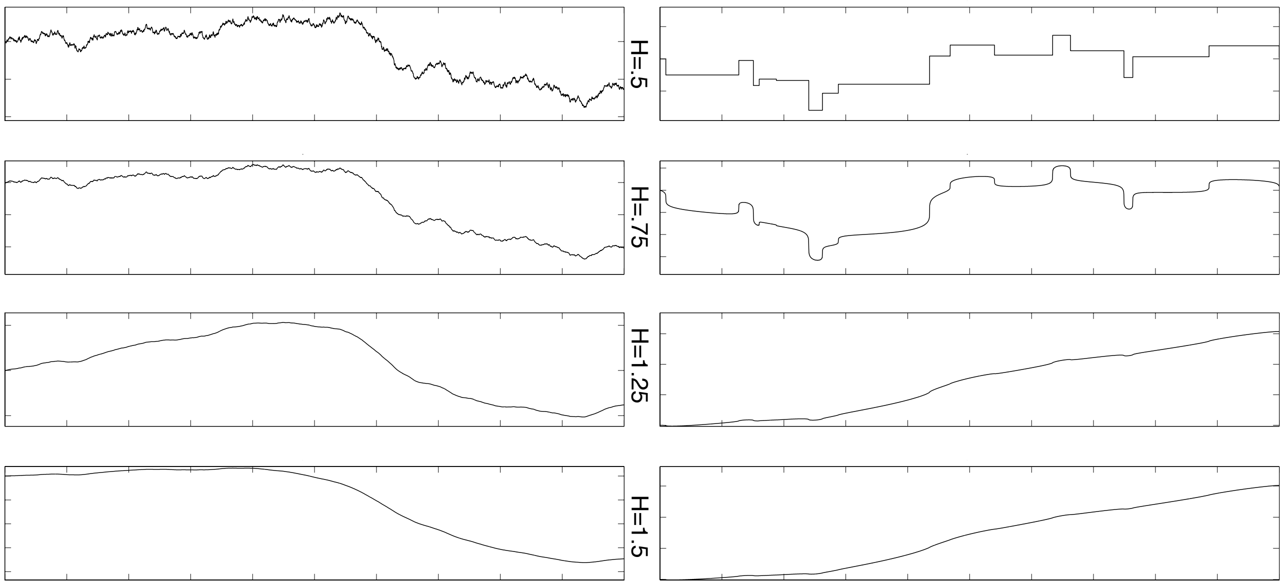
- Characteristic form of fractional Brownian motion

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

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

Example in 1D: Self-similar processes

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

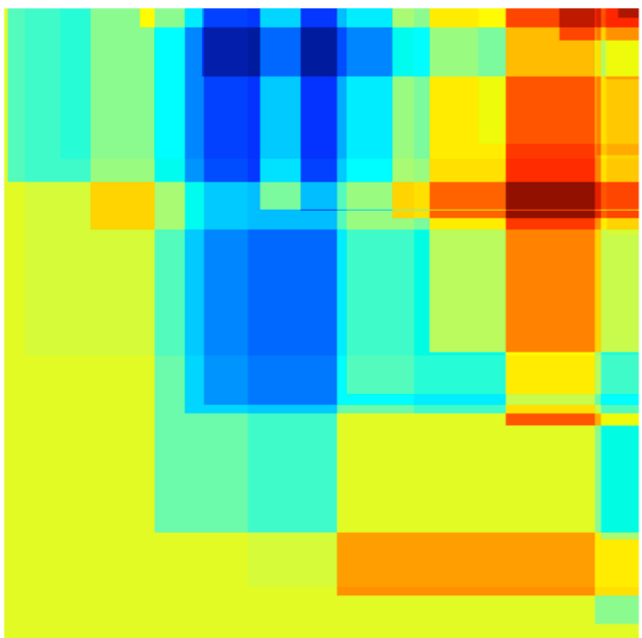


Gaussian | **Sparse (generalized Poisson)**

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

Aesthetic sparse signal: the Mondrian process

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

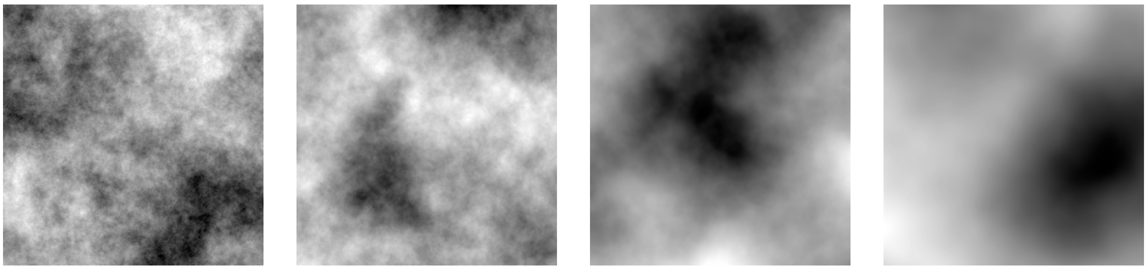


$\lambda = 30$

Scale- and rotation-invariant processes

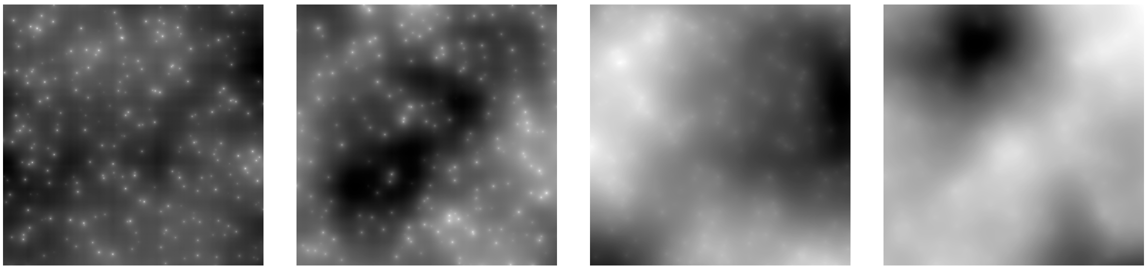
Stochastic partial differential equation : $(-\Delta)^{\frac{H+1}{2}} s(\boldsymbol{x}) = w(\boldsymbol{x})$

Gaussian



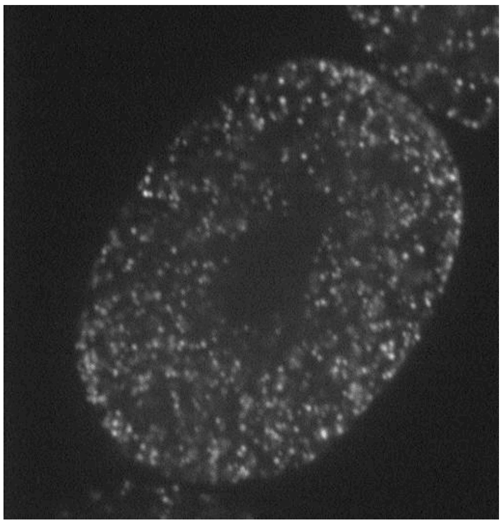
H=.5 H=.75 H=1.25 H=1.75

Sparse (generalized Poisson)



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

Powers of ten: from astronomy to biology



PART II: RECOVERY OF SPARSE SIGNALS

- **10.1 Discretization of inverse problem**
- **10.2 MAP estimation and regularization**
- **10.3 MAP reconstruction of biomedical images**
  - Deconvolution of fluorescent micrographs
  - Magnetic resonance imaging
  - X-ray tomography
- **10.4 Quest for minimum error solution**

Generalized innovation model

Theoretical framework: Gelfand's theory of generalized stochastic processes

(2) Specification of inverse operator

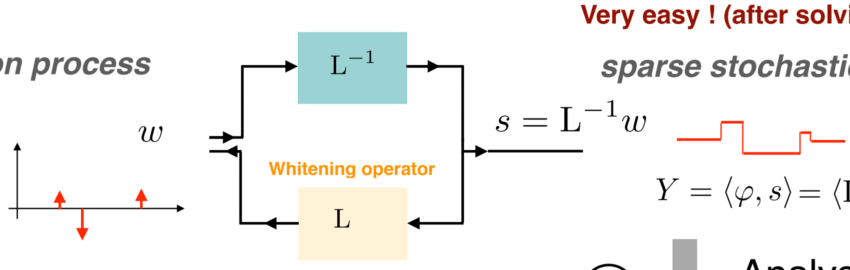
Functional analysis solution of SDE

(3)

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

innovation process *sparse stochastic process*

$X = \langle \varphi, w \rangle$



$Y = \langle \varphi, s \rangle = \langle \mathrm{L}^{-1*} \varphi, w \rangle$

(4) Analysis step

Lévy exponent: $f(\omega)$

(1) Characterization of continuous-domain white noise

Higher mathematics: generalized functions (Schwartz)

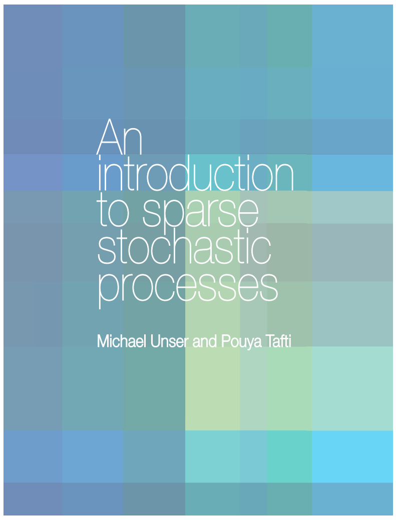
measures on topological vector spaces

Minlos-Bochner and Lévy-Khintchine theorems

Regularization operator vs. wavelet analysis

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

Table of content

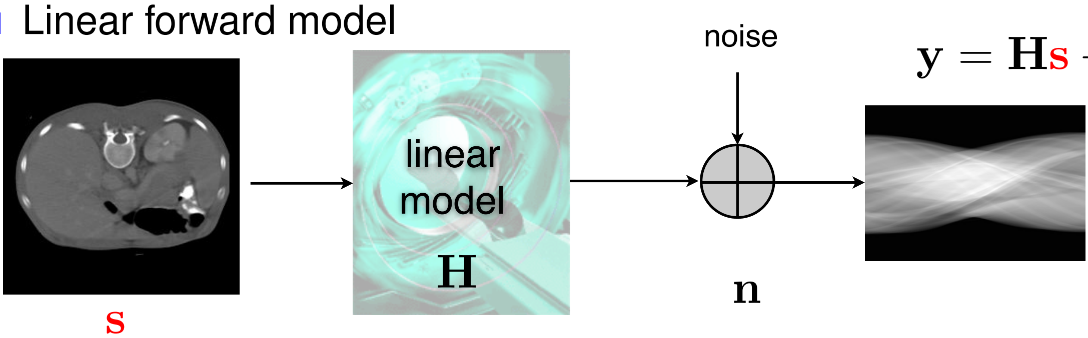


http://www.sparseprocesses.org/

Resolution of linear inverse problems

- Linear forward model

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



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

- Reconstruction as an optimization problem

Bayesian prior $-\log p_S(\mathbf{s})$

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

Gaussian-noise likelihood $-\log p_{Y|S}(\mathbf{n})$

Recap on probability model

- Continuous-domain model: $s = \mathrm{L}^{-1} w$

  $w = \mathrm{L}s$: generalized white noise process with Lévy exponent $f(\omega)$

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

- Discretization: $s(\boldsymbol{k}), \boldsymbol{k} \in \mathbb{Z}^d$ (sampled values)

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

  Discrete increment process:

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

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

- Statistical properties
  - $u[\boldsymbol{k}]$ are identically distributed and approximately independent
  - **Infinitely divisible** with Lévy exponent $f_U(\omega) = \log \widehat{p}_U(\omega) = \log \widehat{\mathscr{P}}_w(\omega\beta_{\mathrm{L}})$

Discretization of reconstruction problem

Spline-like reconstruction model: $s(\boldsymbol{r}) = \sum_{\boldsymbol{k}\in\Omega} s[\boldsymbol{k}]\beta_{\boldsymbol{k}}(\boldsymbol{r}) \quad \longleftrightarrow \quad \mathbf{s} = (s[\boldsymbol{k}])_{\boldsymbol{k}\in\Omega}$

- Innovation model

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

Discretization

$\mathbf{u} = \mathbf{L}\mathbf{s}$ (matrix notation)

p_U is part of **infinitely divisible** family

- Physical model: image formation and acquisition

$$y_m = \int_{\mathbb{R}^d} s_1(\boldsymbol{x})\eta_m(\boldsymbol{x})\mathrm{d}\boldsymbol{x} + n[m] = \langle s_1, \eta_m\rangle + n[m], \quad (m = 1, \dots, M)$$

$$\mathbf{y} = \mathbf{y}_0 + \mathbf{n} = \mathbf{H}\mathbf{s} + \mathbf{n}$$

$\mathbf{n}$: i.i.d. noise with pdf p_N

$$[\mathbf{H}]_{m,\boldsymbol{k}} = \langle \eta_m, \beta_{\boldsymbol{k}}\rangle = \int_{\mathbb{R}^d} \eta_m(\boldsymbol{r})\beta_{\boldsymbol{k}}(\boldsymbol{r})\mathrm{d}\boldsymbol{r}: \quad (M \times K) \text{ system matrix}$$

Posterior probability distribution

$$p_{S|Y}(\mathbf{s}|\mathbf{y}) = \frac{p_{Y|S}(\mathbf{y}|\mathbf{s})p_S(\mathbf{s})}{p_Y(\mathbf{y})} = \frac{p_N(\mathbf{y}-\mathbf{Hs})p_S(\mathbf{s})}{p_Y(\mathbf{y})} \qquad \text{(Bayes' rule)}$$

$$= \frac{1}{Z}p_N(\mathbf{y}-\mathbf{Hs})p_S(\mathbf{s})$$

$$\mathbf{u} = \mathbf{Ls} \qquad \Rightarrow \qquad p_S(\mathbf{s}) \propto p_U(\mathbf{Ls})$$

$$p_{S|Y}(\mathbf{s}|\mathbf{y}) \propto p_N(\mathbf{y}-\mathbf{Hs})p_U(\mathbf{Ls}) \approx p_N(\mathbf{y}-\mathbf{Hs}) \prod_{\boldsymbol{k}\in\Omega} p_U([\mathbf{Ls}]_{\boldsymbol{k}})$$

(decoupling simplication)

- Additive white Gaussian noise scenario (AWGN)

$$p_{S|Y}(\mathbf{s}|\mathbf{y}) \propto \exp\left(-\frac{\|\mathbf{y}-\mathbf{Hs}\|^2}{2\sigma^2}\right) \prod_{\boldsymbol{k}\in\Omega} p_U([\mathbf{Ls}]_{\boldsymbol{k}})$$

Statement of MAP reconstruction problem

- Hypotheses

  $\mathbf{y} = \mathbf{Hs} + \mathbf{n}$ where $\mathbf{n}$ AWGN with variance σ^2

  $\mathbf{Ls} = \mathbf{u}$: i.i.d. with pdf p_U and id potential function $\Phi_U(x) = -\log p_U(x)$

- Maximum a posteriori (MAP) estimator

$$\mathbf{s}_{\mathrm{MAP}} = \arg\min_{\mathbf{s}\in\mathbb{R}^K} \left(\frac{1}{2}\|\mathbf{y}-\mathbf{Hs}\|_2^2 + \sigma^2 \sum_{\boldsymbol{k}\in\Omega} \Phi_U([\mathbf{Ls}]_{\boldsymbol{k}})\right)$$

Sparser

  - Gaussian: $p_U(x) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-x^2/(2\sigma_0^2)} \qquad \Rightarrow \qquad \Phi_U(x) = \frac{1}{2\sigma_0^2}x^2 + C_1$
  - Laplace: $p_U(x) = \frac{\lambda}{2}e^{-\lambda|x|} \qquad \Rightarrow \qquad \Phi_U(x) = \lambda|x| + C_2$
  - Student: $p_U(x) = \frac{1}{B\left(r, \frac{1}{2}\right)}\left(\frac{1}{x^2+1}\right)^{r+\frac{1}{2}} \qquad \Rightarrow \qquad \Phi_U(x) = \left(r+\frac{1}{2}\right)\log(1+x^2) + C_3$

$p_X(x)$	$\Phi_X(x) = -\log p_X(x)$ as $x \to 0$	$\Phi_X(x)$ as $x \to \pm\infty$	Smooth	Convex
Gaussian	$a_0 + \frac{x^2}{2\sigma^2}$	$a_0 + \frac{x^2}{2\sigma^2}$	Yes	Yes
Laplace ($\lambda \in \mathbb{R}^+$)	$a_0 + \lambda\lvert x\rvert$	$a_0 + \lambda\lvert x\rvert$	No	Yes
Sym Gamma $r \in \mathbb{R}^+$	$\begin{cases} \log(a'_0 + a'_r\lvert x\rvert^{2r-1} + O(x^2)), & r < 3/2 \\ a_0 + \frac{x^2}{4r-6} + O(\lvert x\rvert^{\min(4,2r-1)}), & r > 3/2 \end{cases}$	$b_0 + \lvert x\rvert - (r-1)\log\lvert x\rvert$	No	No
Hyperbolic secant	$a_0 + \frac{\pi^2 x^2}{8\sigma_0^2} + O(x^4)$	$-\log\sigma_0 + \frac{\pi}{2\sigma_0}\lvert x\rvert$	Yes	Yes
Meixner $r, s \in \mathbb{R}^+$	$a_0 + \frac{\psi^{(1)}(r/2)}{4s^2}x^2 + O(x^4)$	$b_0 + \frac{\pi}{2s}\lvert x\rvert - (r-1)\log\lvert x\rvert$	Yes	No
Cauchy $s \in \mathbb{R}^+$	$a_0 + \frac{x^2}{s^2} + O(x^4)$	$b_0 - \log s + 2\log\lvert x\rvert$	Yes	No
Sym Student $r \in \mathbb{R}^+$	$a_0 + \left(r + \frac{1}{2}\right)x^2 + O(x^4)$	$b_0 + (2r+1)\log\lvert x\rvert$	Yes	No
SαS, $\alpha \in (0,2]$, $s \in \mathbb{R}^+$	$a_0 + \frac{\Gamma(\frac{3}{\alpha})}{2s^2\Gamma(\frac{1}{\alpha})}x^2 + O(x^4)$	$b_0 - \alpha\log s + (\alpha+1)\log\lvert x\rvert$	Yes	No

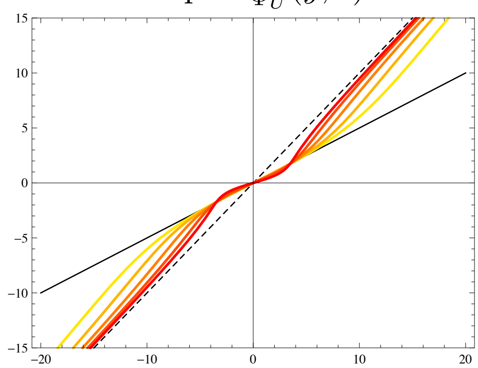
$\Gamma(z)$ and $\psi^{(1)}(r)$ are Euler's gamma and first-order poly-gamma functions, respectively (see Appendix C).

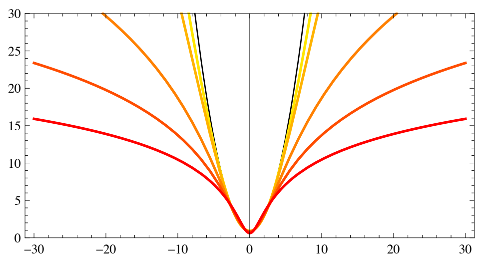
Table 10.1 Asymptotic behavior of the potential function $\Phi_X(x)$ for the infinite-divisible distributions in Table 4.1.

Proximal operators

$$\mathrm{prox}_{\Phi_U}(y;\sigma^2) = \arg\min_{u\in\mathbb{R}} \frac{1}{2}|y-u|^2 + \sigma^2\Phi_U(u)$$

$$\tilde{u} = \mathrm{prox}_{\Phi_U}(y;1)$$





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

Solution of functional equation: $-y + \tilde{u} + \sigma^2\Phi'_U(\tilde{u}) = 0$

One-to-one mapping $y \mapsto \tilde{u}$ when $1 + \sigma^2\Phi''_U(u) \geq 0$

$$\Rightarrow \quad \tilde{u}(y) = (\mathrm{Id} + \sigma^2\Phi'_U)^{-1}(y)$$

Proximal operators (Cont'd)

- Special cases: $\Phi_1(u) = \lambda|u|$ (Laplace) and $\Phi_2(u) = |u|^2/(2\sigma_0^2)$ (Gaussian)

$$T_1(y) = \mathrm{prox}_{\Phi_1}(y;\sigma^2) = \mathrm{sign}(y)\left(|y| - \lambda\sigma^2\right)_+ \qquad \textbf{(soft-threshold)}$$

$$T_2(y) = \mathrm{prox}_{\Phi_2}(y;\sigma^2) = \frac{\sigma_0^2}{\sigma_0^2+\sigma^2}y \qquad \textbf{(linear attenuation)}$$

- Asymptotics

Around origin when Φ_U is twice differentiable

$$\Phi_U'(u) = \Phi_U''(0)u + O(u^2) \quad \Rightarrow \quad \mathrm{prox}_{\Phi_U}(y;\sigma^2) = \frac{y}{1+\sigma^2\Phi_U''(0)} \text{ as } y \to 0$$

Exponential and sub-exponential category: $\lim_{u\to\infty} \Phi_U'(u) = b_1 + O(1/u)$

$$\mathrm{prox}_{\Phi_U}(y;\sigma^2) \sim y - \sigma^2 b_1 \text{ as } y \to +\infty \qquad \textbf{(shrinkage)}$$

Heavy-tailed category: $\lim_{u\to\infty} \Phi_U'(u) = b_2/u$

$$\mathrm{prox}_{\Phi_U}(y;\sigma^2) \sim y \text{ as } y \to \infty \qquad \textbf{(identity)}$$

Maximum a posteriori (MAP) estimation

- Constrained optimization formulation

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

$$\mathbf{s}_{\mathrm{MAP}} = \arg\min_{\mathbf{s}\in\mathbb{R}^K}\left(\frac{1}{2}\|\mathbf{y} - \mathbf{Hs}\|_2^2 + \sigma^2\sum_n \Phi_U([\mathbf{u}]_n)\right) \text{ subject to } \mathbf{u} = \mathbf{Ls}$$

- Augmented Lagrangian method

Quadratic penalty term: $\frac{\mu}{2}\|\mathbf{Ls} - \mathbf{u}\|_2^2$

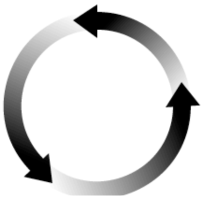
Lagrange multipler vector: $\boldsymbol{\alpha}$

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

Alternating direction method of multipliers (ADMM)

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

Sequential minimization



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

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

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

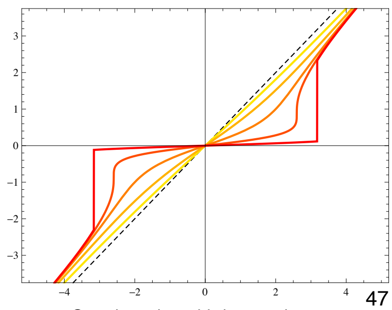
Linear inverse problem: $\mathbf{s}^{k+1} = \left(\mathbf{H}^T\mathbf{H} + \mu\mathbf{L}^T\mathbf{L}\right)^{-1}\left(\mathbf{H}^T\mathbf{y} + \mathbf{z}^{k+1}\right)$

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

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

- Proximal operator taylored to stochastic model

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



Cauchy prior with increasing s_0

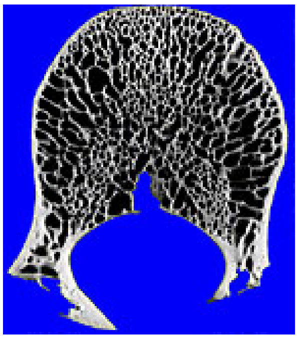
10.3 RECONSTRUCTION OF BIOMEDICAL IMAGES

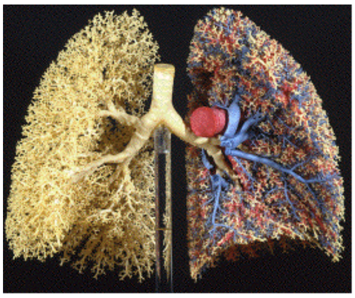
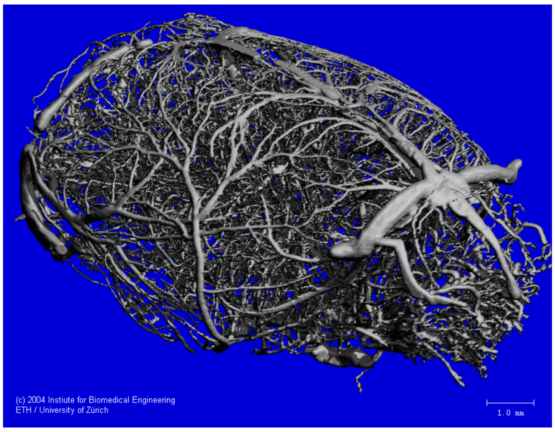
- Common image model and numerical set-up
  - $\frac{1}{|\boldsymbol{\omega}|^\gamma}$ spectral decay $\longleftrightarrow$ $(-\Delta)^{\frac{\gamma}{2}}s = w$ (self-similar image model)
  - Robust localization/decoupling $\mathbf{L}$: discrete gradient magnitude (rotation invariant)
  - Three flavors of potentials:

    $|x|^2$ (Gaussian), $|x|$ (Laplacian), $\log(x^2+\epsilon)$ (Student)
- Deconvolution of fluorescent micrographs
- Magnetic resonance imaging
- X-ray tomography

Relevance of self-similariy for bio-imaging

- Fractals and physiology

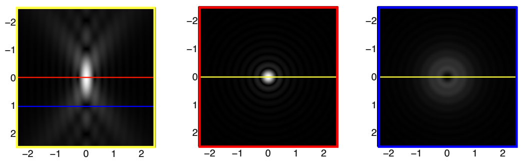



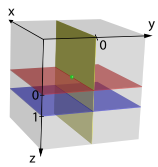
Deconvolution of fluorescence micrographs

- Physical model of a diffraction-limited microscope

$$g(x,y,z) = (h_{3\mathrm{D}} * s)(x,y,z)$$



3-D point spread function (PSF)



$$h_{3\mathrm{D}}(x,y,z) = I_0 \left| p_\lambda \left(\frac{x}{M}, \frac{y}{M}, \frac{z}{M^2} \right) \right|^2$$

$$p_\lambda(x,y,z) = \int_{\mathbb{R}^2} P(\omega_1,\omega_2) \exp\left(\mathrm{j}2\pi z \frac{\omega_1^2+\omega_2^2}{2\lambda f_0^2} \right) \exp\left(-\mathrm{j}2\pi \frac{x\omega_1 + y\omega_2}{\lambda f_0} \right) \mathrm{d}\omega_1 \mathrm{d}\omega_2$$

Optical parameters

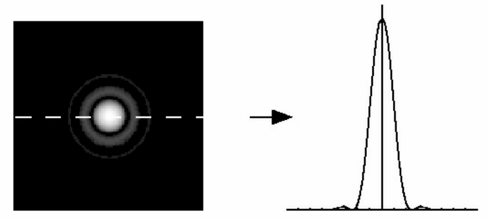
- λ: wavelength (emission)
- M: magnification factor
- f_0: focal length
- $P(\omega_1,\omega_2) = \mathbb{1}_{\|\boldsymbol{\omega}\| < R_0}$: pupil function
- $\mathrm{NA} = n \sin\theta = R_0/f_0$: numerical aperture

2-D convolution model

Thin specimen $s(x,y)$ →  → $g(x,y) = (h_{2D} * s)(x,y)$

- Airy disk: $h_{2D}(x,y) = I_0 \left| 2\frac{J_1(r/r_0)}{r/r_0} \right|^2$

Radial profile



with $r = \sqrt{x^2+y^2}$, $r_0 = \frac{\lambda f_0}{2\pi R_0}$, $J_1(r)$: first-order Bessel function.

- Modulation transfer function

$$\hat{h}_{2D}(\boldsymbol{\omega}) = \begin{cases} \frac{2}{\pi}\left(\arccos\left(\frac{\|\boldsymbol{\omega}\|}{\omega_0}\right) - \frac{\|\boldsymbol{\omega}\|}{\omega_0}\sqrt{1-\left(\frac{\|\boldsymbol{\omega}\|}{\omega_0}\right)^2}\right), & \text{for } 0 \leq \|\boldsymbol{\omega}\| < \omega_0 \\ 0, & \text{otherwise} \end{cases}$$

Cut-off frequency (Rayleigh): $\omega_0 = \frac{2R_0}{\lambda f_0} = \frac{\pi}{r_0} \approx \frac{2\mathrm{NA}}{\lambda}$

2-D deconvolution: numerical set-up

- Discretization

$\omega_0 \leq \pi$ and representation in (separable) sinc basis $\{\mathrm{sinc}(\boldsymbol{x}-\boldsymbol{k})\}_{\boldsymbol{k}\in\mathbb{Z}^2}$

Analysis functions: $\eta_{\boldsymbol{m}}(x,y) = h_{2D}(x-m_1, y-m_2)$

$$\begin{aligned}[\mathbf{H}]_{\boldsymbol{m},\boldsymbol{k}} &= \langle \eta_{\boldsymbol{m}}, \mathrm{sinc}(\cdot - \boldsymbol{k})\rangle \\ &= \langle h_{2D}(\cdot - \boldsymbol{m}), \mathrm{sinc}(\cdot - \boldsymbol{k})\rangle \\ &= \big(\mathrm{sinc} * h_{2D}\big)(\boldsymbol{m}-\boldsymbol{k}) = h_{2D}(\boldsymbol{m}-\boldsymbol{k}).\end{aligned}$$

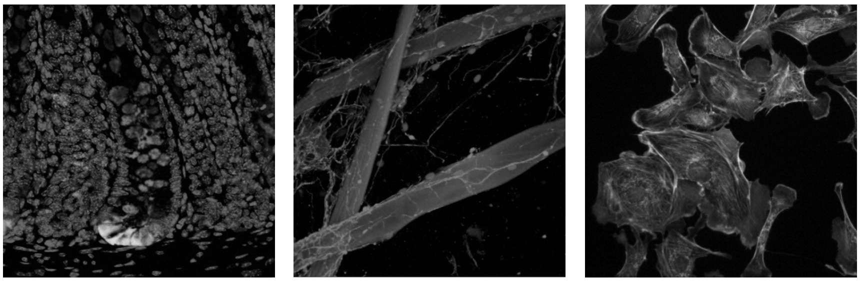
$\mathbf{H}$ and $\mathbf{L}$: convolution matrices diagonalized by discrete Fourier transform

- Linear step of ADMM algorithm implemented using the FFT

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

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

Deconvolution experiments



(a) (b) (c)

Figure 10.3 Images used in deconvolution experiments. (a) Stem cells surrounded by goblet cells. (b) Nerve cells growing around fibers. (c) Artery cells.

Table 10.2 Deconvolution performance of MAP estimators based on different prior distributions.

		Estimation performance (SNR in dB)		
	BSNR (dB)	Gaussian	Laplace	Student's
Stem cells	20	**14.43**	13.76	11.86
	30	**15.92**	15.77	13.15
	40	**18.11**	**18.11**	13.83
Nerve cells	20	13.86	**15.31**	14.01
	30	15.89	**18.18**	15.81
	40	18.58	**20.57**	16.92
Artery cells	20	14.86	**15.23**	13.48
	30	16.59	**17.21**	14.92
	40	18.68	**19.61**	15.94

Magnetic resonance imaging (MRI)

- Physical image formation model (noise-free)

$$\hat{s}(\boldsymbol{\omega}_m) = \int_{\mathbb{R}^2} s(\boldsymbol{r}) \mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}_m, \boldsymbol{r} \rangle} \mathrm{d}\boldsymbol{r}$$

(sampling of Fourier transform)

Equivalent analysis function: $\eta_m(\boldsymbol{r}) = \mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}_m, \boldsymbol{r} \rangle}$

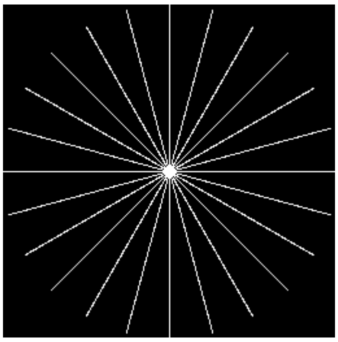
- Discretization in separable sinc basis

$$[\mathbf{H}]_{m,\boldsymbol{n}} = \langle \eta_m, \mathrm{sinc}(\cdot - \boldsymbol{n}) \rangle$$
$$= \langle \mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}_m, \cdot \rangle}, \mathrm{sinc}(\cdot - \boldsymbol{n}) \rangle = \mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}_m, \boldsymbol{n} \rangle}$$

Property: $\mathbf{H}^T\mathbf{H}$ is circulant (FFT-based implementation)

MRI: Shepp-Logan phantom

Original SL Phantom



Fourier Sampling Pattern
12 Angles

Laplace prior (TV)

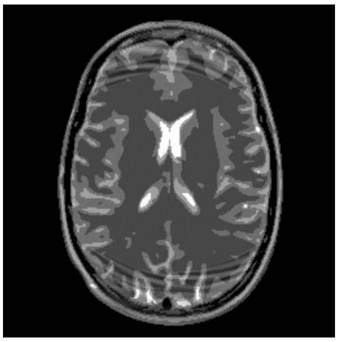
Student prior (log)

L : gradient

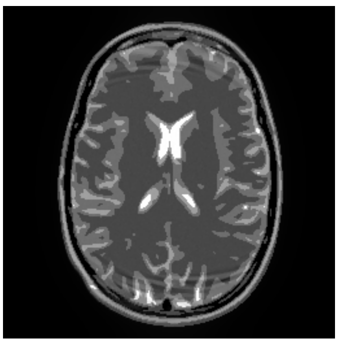
Optimized parameters

MRI phantom: Spiral sampling in k-space

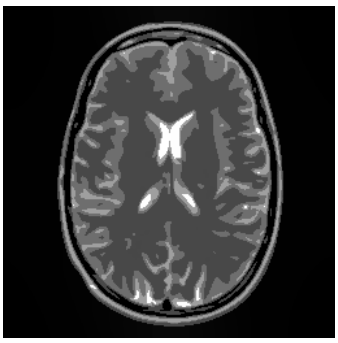
Original Phantom
(Guerquin-Kern TMI 2012)



Gaussian prior (Tikhonov)
SER =17.69 dB



Laplace prior (TV)
SER = 21.37 dB



Student prior
SER = 27.22 dB

L : gradient

Optimized parameters

MRI reconstruction experiments

(a) (b) (c)

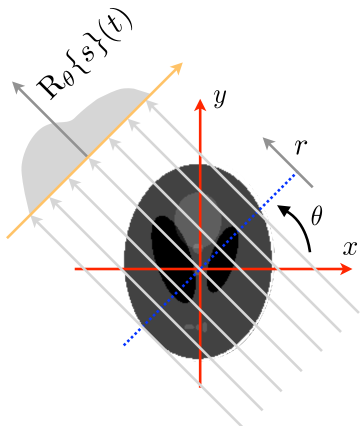
Figure 10.4 Data used in MR reconstruction experiments. (a) Cross section of a wrist. (b) Angiography image. (c) k-space sampling pattern along 40 radial lines.

Table 10.3 MR reconstruction performance of MAP estimators based on different prior distributions.

	Radial lines	Estimation performance (SNR in dB)		
		Gaussian	Laplace	Student's
Wrist	20	8.82	**11.8**	5.97
	40	11.30	**14.69**	13.81
Angiogram	20	4.30	9.01	**9.40**
	40	6.31	14.48	**14.97**

X-ray tomography

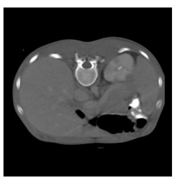
Projection geometry: $\boldsymbol{x} = t\boldsymbol{\theta} + r\boldsymbol{\theta}^{\perp}$ with $\boldsymbol{\theta} = (\cos\theta, \sin\theta)$

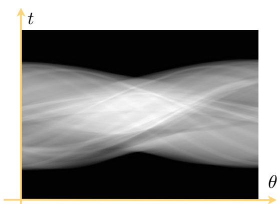


- Radon transform (line integrals)

$$\mathrm{R}_\theta\{s(\boldsymbol{x})\}(t) = \int_{\mathbb{R}} s(t\boldsymbol{\theta} + r\boldsymbol{\theta}^{\perp})\mathrm{d}r$$

$$= \int_{\mathbb{R}^2} s(\boldsymbol{x})\delta(t - \langle \boldsymbol{x}, \boldsymbol{\theta}\rangle)\mathrm{d}\boldsymbol{x}$$





sinogram

Equivalent analysis functions: $\eta_m(\boldsymbol{x}) = \delta\big(t_m - \langle \boldsymbol{x}, \boldsymbol{\theta}_m\rangle\big)$

Properties of Radon transform

- Projected translation invariance

$$\mathrm{R}_\theta\{\varphi(\cdot - \boldsymbol{x}_0)\}(t) = \mathrm{R}_\theta\{\varphi\}(t - \langle \boldsymbol{x}_0, \boldsymbol{\theta}\rangle)$$

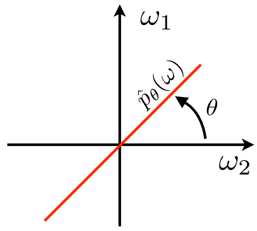
- Pseudo-distributivity with respect to convolution

$$\mathrm{R}_\theta\{\varphi_1 * \varphi_2\}(t) = \left(\mathrm{R}_\theta\{\varphi_1\} * \mathrm{R}_\theta\{\varphi_2\}\right)(t)$$

- Fourier central-slice theorem

$$\int_{\mathbb{R}} \mathrm{R}_\theta\{\varphi\}(t)\mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}t = \hat{\varphi}(\boldsymbol{\omega})\big|_{\boldsymbol{\omega}=\omega\boldsymbol{\theta}}$$

$$\hat{p}_\theta(\omega) = \widehat{\mathrm{R}_\theta\{\varphi\}}(\omega) = \hat{\varphi}(\omega\cos\theta, \omega\sin\theta)$$



Proposition: Consider the separable function $\varphi(\boldsymbol{x}) = \varphi_1(x)\varphi_2(y)$. Then,

$$\mathrm{R}_\theta\{\varphi(\cdot - \boldsymbol{x}_0)\}(t) = \varphi_\theta(t - t_0)$$

where $t_0 = \langle \boldsymbol{x}_0, \boldsymbol{\theta}\rangle$ and

$$\varphi_\theta(t) = \left(\tfrac{1}{|\cos\theta|}\varphi_1\left(\frac{\cdot}{\cos\theta}\right) * \tfrac{1}{|\sin\theta|}\varphi_2\left(\frac{\cdot}{\sin\theta}\right)\right)(t).$$

Discretization using polynomial B-splines

- Separable B-spline reconstruction model

$$s(\boldsymbol{x}) = \sum_{\mathbf{k}} s[\mathbf{k}]\beta^n(\boldsymbol{x} - \boldsymbol{k}) \quad \text{with} \quad \beta^n(\boldsymbol{x}) = \beta^n(x)\beta^n(y)$$

Centered polynomial B-spline: $\displaystyle \beta^n(x) = \sum_{k=0}^{n+1}(-1)^k\binom{n+1}{k}\frac{\left(x - k + \frac{n+1}{2}\right)_+^n}{n!}$

- Radon transform of B-spline

$$\mathrm{R}_\theta\left\{\beta^n(x)\beta^n(y)\right\}(t) = \beta^n_\theta(t)$$

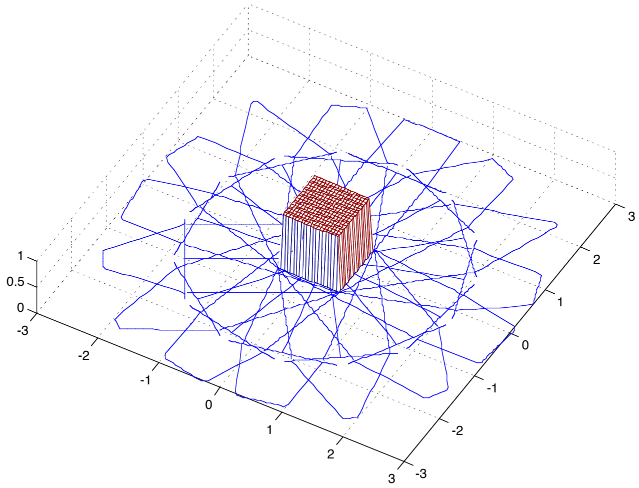
$$= \sum_{k=0}^{n+1}\sum_{k'=0}^{n+1}(-1)^{k+k'}\binom{n+1}{k}\binom{n+1}{k'}\frac{\left(t + \left(\frac{n+1}{2} - k\right)\cos\theta + \left(\frac{n+1}{2} - k'\right)\sin\theta\right)_+^{2n+1}}{|\cos\theta|^{n+1}\,|\sin\theta|^{n+1}\,(2n+1)!}$$

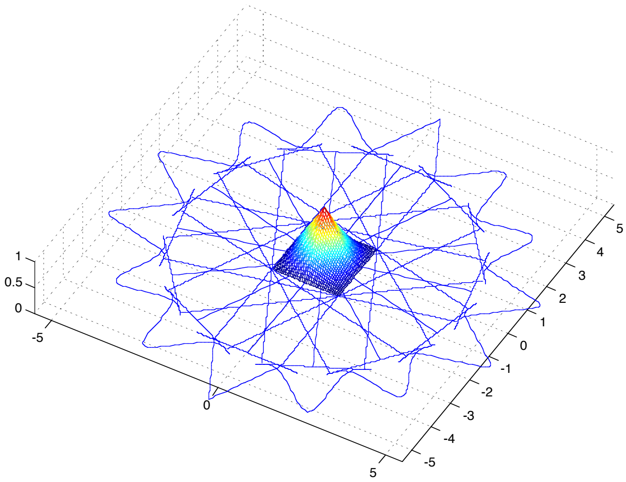
Justification: $\frac{t_+^{n_1}}{n_1!} * \frac{t_+^{n_2}}{n_2!} = \frac{t_+^{n_1+n_2+1}}{(n_1+n_2+1)!}$

- System matrix

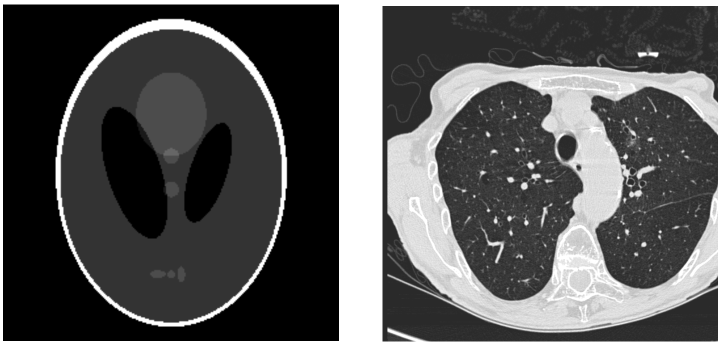
$$[\mathbf{H}]_{m,\boldsymbol{k}} = \langle\delta(t_m - \langle\cdot, \boldsymbol{\theta}_m\rangle), \beta^n(\cdot - \boldsymbol{k})\rangle$$
$$= \mathrm{R}_{\theta_m}\left\{\beta^n(\cdot - \boldsymbol{k})\right\}(t_m) = \beta^n_{\theta_m}(t_m - \langle\boldsymbol{k}, \boldsymbol{\theta}_m\rangle)$$

Spline flowers





X-ray tomography reconstruction results



(a) (b)

Figure 10.6 Images used in X-ray tomographic reconstruction experiments. (a) The Shepp-Logan (SL) phantom. (b) Cross section of a human lung.

Table 10.4 Reconstruction results of X-ray computed tomography using different estimators.

	Directions	Estimation performance (SNR in dB)		
		Gaussian	Laplace	Student's
SL Phantom	120	16.8	17.53	**18.76**
SL Phantom	180	18.13	18.75	**20.34**
Lung	180	**22.49**	21.52	21.45
Lung	360	**24.38**	22.47	22.37

10.4 QUEST FOR MINIMUM ERROR SOLUTION

How suitable are MAP estimators ?

➠ A detailed investigation of simpler denoising problem

- MMSE estimators for first-order processes
- Direct solution by belief propagation
- MMSE vs. MAP denoising of Lévy processes

MMSE estimators for first-order processes

Task: Recovery of non-Gaussian AR(1) and Lévy processes from noisy samples

- Measurement model: $p_{(Y_1:Y_N|X_1:X_N)}(\mathbf{y}|\mathbf{x}) = \prod_{n=1}^{N} \underbrace{p_{Y|X}(y_n|x_n)}$
  independent noise contributions

- Discrete innovation model: $u_n = x_n - a_1 x_{n-1}$, u i.i.d. with pdf p_U

- Posterior distribution of signal

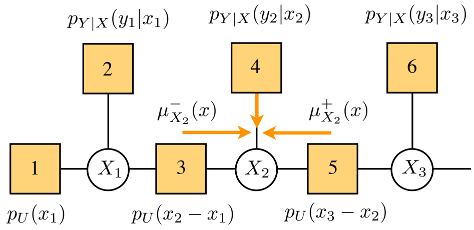
$$p_{(X_1:X_N|Y_1:Y_N)}(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} \prod_{n=1}^{N} p_{Y|X}\left(y_n|x_n\right) \prod_{n=1}^{N} p_U\big(\underbrace{x_n - a_1 x_{n-1}}_{u_n}\big)$$

- Signal estimators

$$\mathbf{x}_{\mathrm{MAP}}(\mathbf{y}) = \arg\max_{\mathbf{x}\in\mathbb{R}^N} \left\{p_{(X_1:X_N|Y_1:Y_N)}(\mathbf{x}|\mathbf{y})\right\}$$

$$\mathbf{x}_{\mathrm{MMSE}}(\mathbf{y}) = \mathbb{E}\{\mathbf{x}|\mathbf{y}\} = \int_{\mathbb{R}^N} \mathbf{x}\, p_{(X_1:X_N|Y_1:Y_N)}(\mathbf{x}|\mathbf{y})\mathrm{d}\mathbf{x}$$: optimal MMSE solution

Introduction to BP algorithm



- Estimation of a three-point Lévy process

$$p_{(X_1:X_3|Y_1:Y_3)}(\mathbf{x}|\mathbf{y})$$
$$\propto p_U(x_1)\, p_{Y|X}(y_1|x_1)\, p_U(x_2 - x_1)\, p_{Y|X}(y_2|x_2)\, p_U(x_3 - x_2)\, p_{Y|X}(y_3|x_3)$$

- Recursive evaluation of marginal distributions

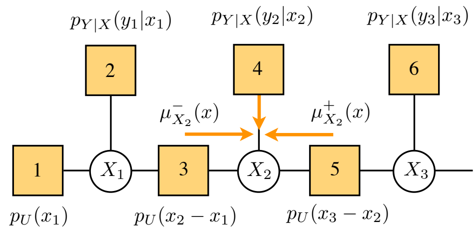
$$p_{(X_2|Y_1:Y_3)}(x_2|\mathbf{y}) = \int_{\mathbb{R}}\int_{\mathbb{R}} p_{(X_1:X_3|Y_1:Y_3)}(\mathbf{x}|\mathbf{y})\mathrm{d}x_1\mathrm{d}x_3$$
$$\propto \underbrace{\int_{\mathbb{R}} \overbrace{p_U(x_1)}^{\mu^-_{X_1}(x_1)}\, p_{Y|X}(y_1|x_1)\, p_U(x_2 - x_1)\mathrm{d}x_1}_{\mu^-_{X_2}(x_2)} \cdot p_{Y|X}(y_2|x_2) \cdot \underbrace{\int_{\mathbb{R}} p_U(x_3 - x_2)\, p_{Y|X}(y_3|x_3) \cdot \overbrace{1}^{\mu^+_{X_3}(x_3)}\, \mathrm{d}x_3}_{\mu^+_{X_2}(x_2)}$$

$$p_{(X_3|Y_1:Y_3)}(x_3|\mathbf{y}) = \int_{\mathbb{R}}\int_{\mathbb{R}} p_{(X_1:X_3|Y_1:Y_3)}(\mathbf{x}|\mathbf{y})\mathrm{d}x_1\mathrm{d}x_2$$
$$\propto \underbrace{\int_{\mathbb{R}} \mu^-_{X_2}(x_2)\, p_{Y|X}(y_2|x_2)\, p_U(x_3 - x_2)\mathrm{d}x_2}_{\mu^-_{X_3}(x_3)} \cdot p_{Y|X}(y_3|x_3) \cdot \underbrace{1}_{\mu^+_{X_3}(x_3)} .$$

Direct MMSE solution by belief propagation

Factorized representation: $p_{(X_n|Y_1:Y_N)}(x_n|\mathbf{y}) = \mu^-_{X_n}(x_n) \cdot p_{Y|X}(y_n|x_n) \cdot \mu^+_{X_n}(x_n)$

Auxiliary **belief** functions $\mu^-_{X_n}(x)$ and $\mu^+_{X_n}(x)$



- BP for Lévy and non-Gaussian AR(1) processes

– *Initialization*: Set

$$\mu^-_{X_1}(x) = p_U(x)$$
$$\mu^+_{X_N}(x) = 1$$

– *Forward message recursion*: For $n = 2$ to N, compute

$$\mu^-_{X_n}(x) \propto \int_{\mathbb{R}} \mu^-_{X_{n-1}}(z)\, p_{Y|X}(y_{n-1}|z)\, p_U(x - a_1 z)\, \mathrm{d}z$$

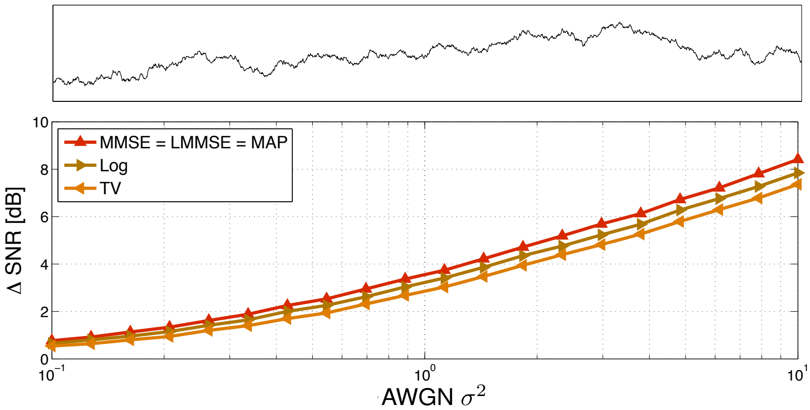
– *Backward message recursion*: For $n = (N-1)$ down to 1, compute

$$\mu^+_{X_n}(x) \propto \int_{\mathbb{R}} p_U(z - a_1 x)\, p_{Y|X}(y_{n+1}|z)\, \mu^+_{X_{n+1}}(z)\, \mathrm{d}z$$

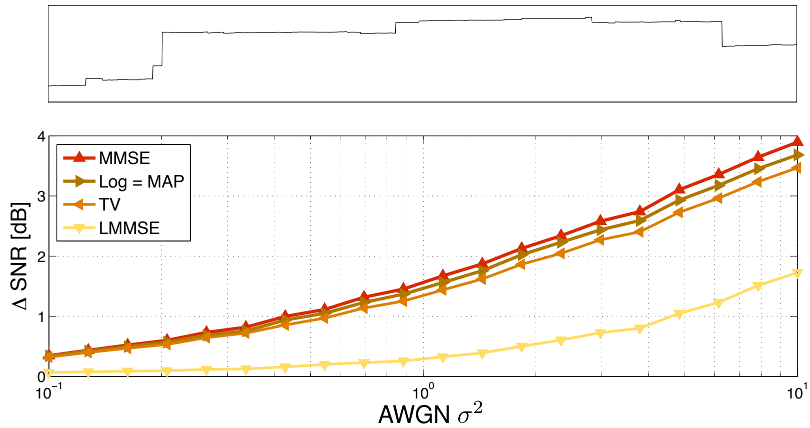
– *Results*: For $n = 1$ to N, compute

$$p_{(X_n|Y_1:Y_N)}(x|\mathbf{y}) \propto \mu^-_{X_n}(x) \cdot p_{Y|X}(y_n|x) \cdot \mu^+_{X_n}(x)$$
$$[\mathbf{x}_{\mathrm{MMSE}}]_n = \int_{\mathbb{R}} x\, p_{(X_n|Y_1:Y_N)}(x|\mathbf{y})\, \mathrm{d}x$$

Experiment: Denoising of Lévy processes



Gaussian



Cauchy

CONCLUSION

- Unifying continuous-domain stochastic model
  - Backward compatibility with classical Gaussian theory
  - Operator-based formulation: Lévy-driven SDEs or SPDEs
  - **Gaussian** vs. **sparse** (generalized Poisson, student, SαS)
- Regularization
  - Sparsification via “operator-like” behavior (whitening)
  - Specific family of id potential functions (typ., non-convex)
- Conceptual framework for sparse signal recovery
  - New statistically-founded sparsity priors
  - Derivation of optimal estimators (MAP, MMSE)
  - Principled approach for the development of novel algorithms
- Challenges
  - Calculation of MMSE solution (belief propagation ?)
  - Fast algorithms for solving large scale inverse problems with (more or less) structure

References

- Theory of sparse stochastic processes

  - M. Unser and P. Tafti, ***An Introduction to Sparse Stochastic Processes***, Cambridge University Press, 2014; preprint, available at `http://www.sparseprocesses.org`.
  - 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, P. Tafti, and Q. Sun, "A unified formulation of Gaussian vs. sparse stochastic processes—Part I: Continuous-domain theory," *IEEE Trans. Information Theory*, vol. 60, no. 3, pp. 1945-1962, March 2014.

- Algorithms and imaging applications

  - E. Bostan, U.S. Kamilov, M. Nilchian, M. Unser, "Sparse Stochastic Processes and Discretization of Linear Inverse Problems," *IEEE Trans. Image Processing*, vol. 22, no. 7, pp. 2699-2710, 2013.
  - C. Vonesch, M. Unser, "A Fast Multilevel Algorithm for Wavelet-Regularized Image Restoration," *IEEE Trans. Image Processing*, vol. 18, no. 3, pp. 509-523, March 2009.
  - 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, 2011.
  - U. Kamilov, P. Pad, A. Amini, M. Unser, "MMSE Estimation of Sparse Lévy Processes," *IEEE Trans. Signal Processing*, vol. 61, no. 1, pp. 137-147, January 1, 2013.

Acknowledgments

Many thanks to

- Dr. Pouya Tafti
- Prof. Qiyu Sun
- Prof. Arash Amini
- Dr. Cédric Vonesch
- Dr. Matthieu Guerquin-Kern
- Emrah Bostan
- Ulugbek Kamilov
- Masih Nilchian

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

erc

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