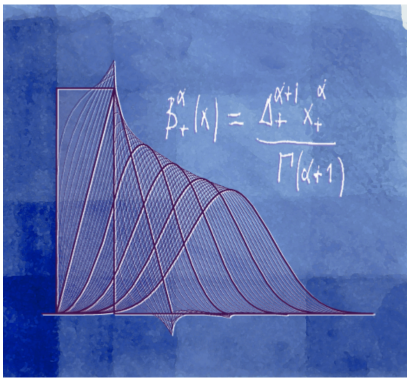
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Sparsity and optimality of splines: Deterministic vs. statistical justification

Michael Unser
Biomedical Imaging Group
EPFL, Lausanne, Switzerland



Mathematics and Image Analysis (MIA), 18-20 January 2016, Paris, France.

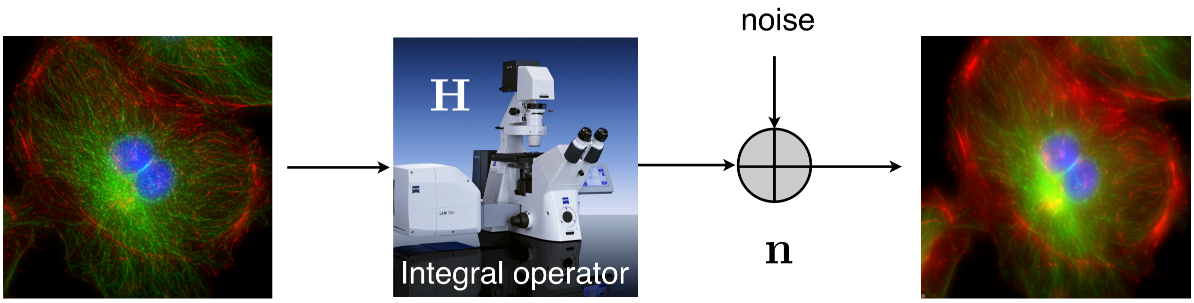
OUTLINE

- **Sparsity and signal reconstruction**
  - Inverse problems in bio-imaging
  - Compressed sensing: towards a continuous-domain formulation
- **Deterministic formulation**
  - Splines and operators
  - New optimality results for generalized TV
- **Statistical formulation**
  - Sparse stochastic processes
  - Derivation of MAP estimators

FNSNF Swiss National Science Foundation
erc

Inverse problems in bio-imaging

- Linear forward model $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}$



$\mathbf{s}$

Problem: recover $\mathbf{s}$ from noisy measurements $\mathbf{y}$

- Reconstruction as an optimization problem

$$\mathbf{s}_{\mathrm{rec}} = \arg\min \underbrace{\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2}_{\text{data consistency}} + \underbrace{\lambda \|\mathbf{L}\mathbf{s}\|_p^p}_{\text{regularization}}, \quad p = 1, 2$$

$-\log \mathrm{Prob}(\mathbf{s})$: prior likelihood

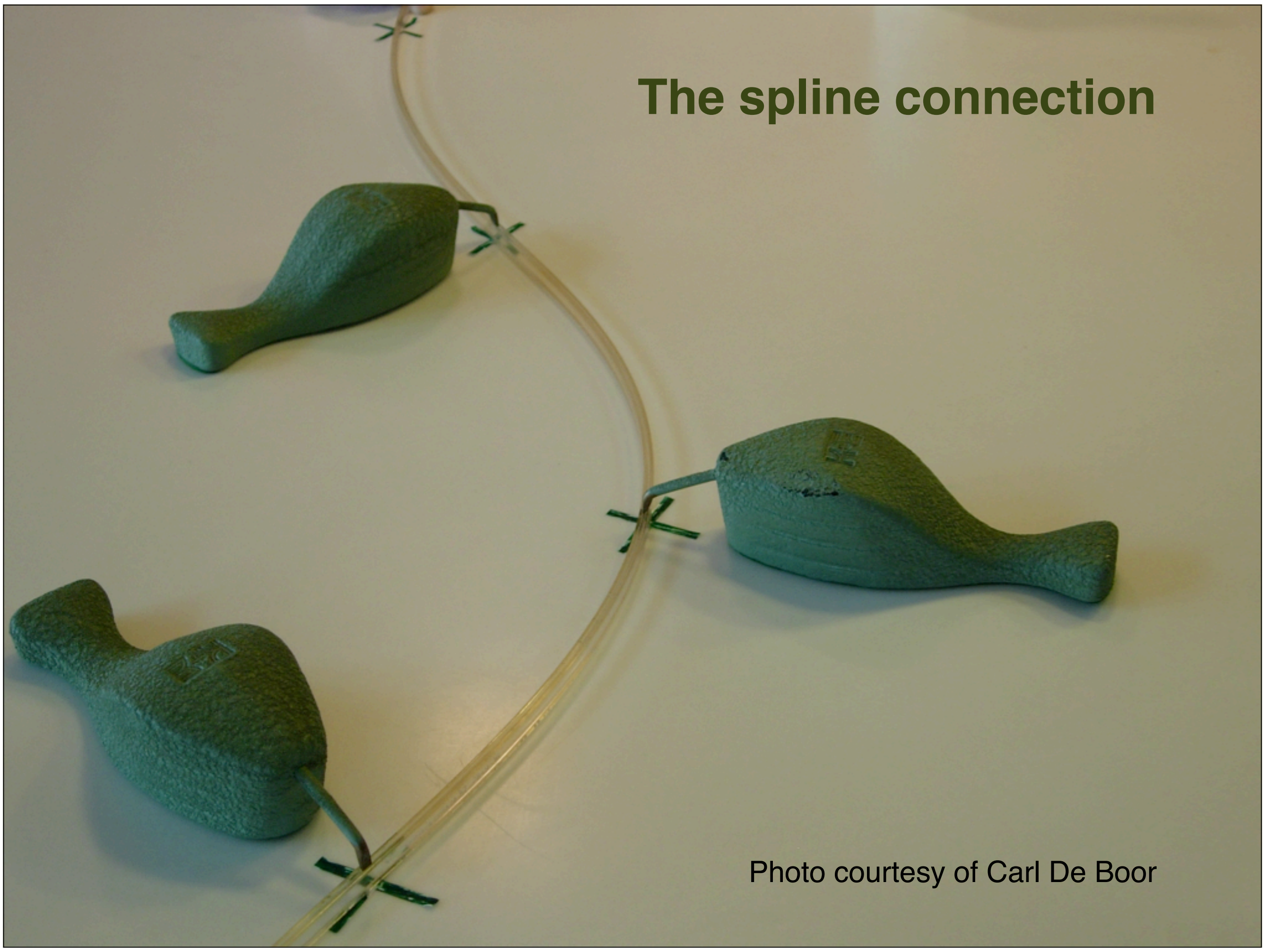
Inverse problems in imaging: Current status

- **Higher reconstruction quality**: Sparsity-promoting schemes almost systematically outperform the classical linear reconstruction methods in MRI, x-ray tomography, deconvolution microscopy, etc... (Lustig et al. 2007)
- **Increased complexity**: Resolution of linear inverse problems using ℓ_1 regularization requires more sophisticated algorithms (iterative and non-linear); efficient solutions (FISTA, ADMM) have emerged during the past decade. (Chambolle 2004; Figueiredo 2004; Beck-Teboule 2009; Boyd 2011)
- The paradigm is supported by the theory of **compressed sensing** (Candes-Romberg-Tao; Donoho, 2006)

- Outstanding research issues
  - Beyond ℓ_1 and TV: Connection with **statistical modeling & learning**
  - Beyond matrix algebra: **Continuous-domain** formulation
  - Guarantees of **optimality** (either deterministic or statistical)

Sparsity and continuous-domain modeling

- Compressed sensing (CS)
  - Generalized sampling and infinite-dimensional CS (Adcock-Hansen, 2011)
  - Xampling: CS of analog signals (Eldar, 2011)
- Splines and approximation theory
  - L_1 splines (Fisher-Jerome, 1975)
  - Locally-adaptive regression splines (Mammen-van de Geer, 1997)
  - Generalized TV (Steidl et al. 2005; Bredies et al. 2010)
- Statistical modeling
  - Sparse stochastic processes (Unser et al. 2011-2014)



Splines are intrinsically sparse

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

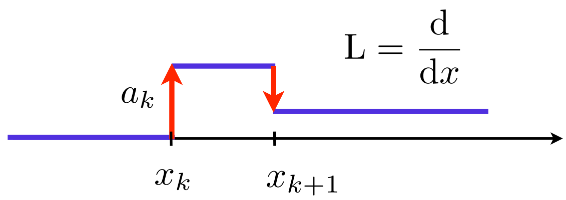
$\delta(\cdot - \boldsymbol{x}_0)$: Dirac impulse shifted by $\boldsymbol{x}_0 \in \mathbb{R}^d$

Definition

The function $s : \mathbb{R}^d \to \mathbb{R}$ is a (non-uniform) L-spline with knots $(\boldsymbol{x}_k)_{k=1}^{K}$ if

$$\mathrm{L}\{s\} = \sum_{k=1}^{K} a_k \delta(\cdot - \boldsymbol{x}_k) = w_\delta \quad : \text{ spline's innovation}$$

Spline theory: (Schultz-Varga, 1967)



- FIR signal processing: Innovation variables $(2K)$ (Vetterli et al., 2002)
  - Location of singularities (knots) : $\{x_k\}_{k=1}^{K}$
  - Strength of singularities (linear weights): $\{a_k\}_{k=1}^{K}$

Splines and operators

Definition

A linear operator $\mathrm{L} : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X} \supset \mathcal{S}(\mathbb{R}^d)$ and $\mathcal{Y}$ are appropriate subspaces of $\mathcal{S}'(\mathbb{R}^d)$, is called **spline-admissible** if

1. it is linear shift-invariant (LSI);
2. its null space $\mathcal{N}_\mathrm{L} = \{p \in \mathcal{X} : \mathrm{L}\{p\} = 0\}$ is finite-dimensional of size N_0;
3. there exists a function $\rho_\mathrm{L} : \mathbb{R}^d \to \mathbb{R}$ of slow growth (the Green's function of L) such that $\mathrm{L}\{\rho_\mathrm{L}\} = \delta$.

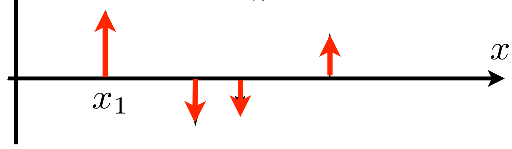
- Structure of null space: $\mathcal{N}_\mathrm{L} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$
  - Admits some basis $\boldsymbol{p} = (p_1, \cdots, p_{N_0})$
  - Only includes elements of the form $\boldsymbol{x}^{\boldsymbol{m}} \mathrm{e}^{\mathrm{j}\langle \boldsymbol{\omega}_0, \boldsymbol{x}\rangle}$ with $|\boldsymbol{m}| = \sum_{i=1}^{d} m_i \le n_0$

Spline synthesis: example

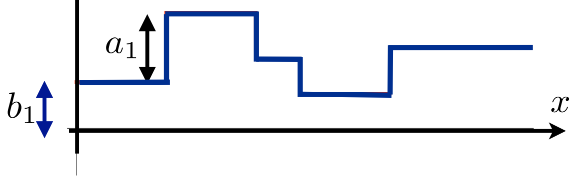
$$\mathrm{L} = \mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}x} \qquad \mathcal{N}_{\mathrm{D}} = \mathrm{span}\{p_1\}, \quad p_1(x) = 1$$

$\rho_{\mathrm{D}}(x) = \mathbb{1}_+(x)$: Heaviside function

$$w_\delta(x) = \sum_k a_k \delta(x - x_k)$$



$$s(x) = b_1 p_1(x) + \sum_k a_k \mathbb{1}_+(x - x_k)$$



Spline synthesis: generalization

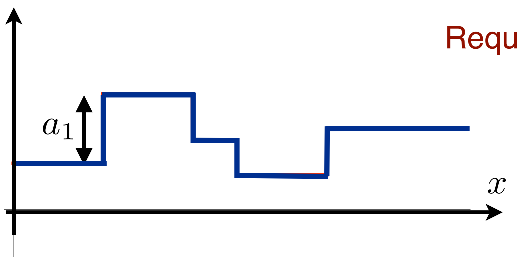
L: spline admissible operator (LSI)

$\rho_{\mathrm{L}}(\boldsymbol{x})$: Green's function of L $\qquad \mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$

Spline's innovation: $\quad w_\delta(\boldsymbol{x}) = \sum_k a_k \delta(\boldsymbol{x} - \boldsymbol{x}_k)$

$$\Rightarrow \qquad s(\boldsymbol{x}) = \sum_k a_k \rho_{\mathrm{L}}(\boldsymbol{x} - \boldsymbol{x}_k) + \sum_{n=1}^{N_0} b_n p_n(\boldsymbol{x})$$

Requires specification of boundary conditions



Principled operator-based approach

- Biorthogonal basis of $\mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$
  - $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_{N_0})$ such that $\langle \phi_m, p_n \rangle = \delta_{m,n}$
  - $p = \sum_{n=1}^{N_0} \langle \phi_n, p \rangle p_n$ for all $p \in \mathcal{N}_{\mathrm{L}}$
- Operator-based spline synthesis
  - Boundary conditions: $\langle s, \phi_n \rangle = b_n, \; n = 1, \cdots, N_0$
  - Spline's innovation: $\mathrm{L}\{s\} = w_\delta = \sum_k a_k \delta(\cdot - \boldsymbol{x}_k)$

$$s(\boldsymbol{x}) = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w_\delta\}(\boldsymbol{x}) + \sum_{n=1}^{N_0} b_n p_n(\boldsymbol{x})$$

- Existence of $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ as a stable right-inverse of L ? (see **Theorem 1**)
  - $\mathrm{L}\mathrm{L}_{\boldsymbol{\phi}}^{-1} w = w$
  - $\langle \boldsymbol{\phi}, \mathrm{L}_{\boldsymbol{\phi}}^{-1} w \rangle = \mathbf{0}$

From Dirac impulses to Borel measures

$\mathcal{S}(\mathbb{R}^d)$: Schwartz's space of smooth and rapidly decaying test functions on $\mathbb{R}^d$

$\mathcal{S}'(\mathbb{R}^d)$: Schwartz's space of tempered distributions

- Space of real-valued, countably additive Borel measures on $\mathbb{R}^d$

$$\mathcal{M}(\mathbb{R}^d) = \big(C_0(\mathbb{R}^d)\big)' = \Big\{w \in \mathcal{S}'(\mathbb{R}^d) : \|w\|_{\mathrm{TV}} = \sup_{\varphi \in \mathcal{S}(\mathbb{R}^d): \|\varphi\|_\infty = 1} \langle w, \varphi \rangle < \infty \Big\},$$

where $w : \varphi \mapsto \langle w, \varphi \rangle = \int_{\mathbb{R}^d} \varphi(\boldsymbol{r}) \mathrm{d}w(\boldsymbol{r})$

- Equivalent definition of "total variation" norm

$$\|w\|_{\mathrm{TV}} = \sup_{\varphi \in C_0(\mathbb{R}^d): \|\varphi\|_\infty = 1} \langle w, \varphi \rangle$$

- Basic inclusions
  - $\delta(\cdot - \boldsymbol{x}_0) \in \mathcal{M}(\mathbb{R}^d)$ with $\|\delta(\cdot - \boldsymbol{x}_0)\|_{\mathrm{TV}} = 1$ for any $\boldsymbol{x}_0 \in \mathbb{R}^d$
  - $\|f\|_{\mathrm{TV}} = \|f\|_{L_1(\mathbb{R}^d)}$ for all $f \in L_1(\mathbb{R}^d) \quad \Rightarrow \quad L_1(\mathbb{R}^d) \subseteq \mathcal{M}(\mathbb{R}^d)$

Generalized Beppo-Levi spaces

L: spline-admissible operator

- Generalized "total variation" semi-norm (gTV)

$$\mathrm{gTV}(f) = \|\mathrm{L}\{f\}\|_{\mathrm{TV}}$$

- Generalized Beppo-Levi spaces

$$\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) = \big\{f : \mathbb{R}^d \to \mathbb{R} \;\big|\; \|\mathrm{L}f\|_{\mathrm{TV}} < \infty \big\}$$

$$f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \Leftrightarrow \mathrm{L}\{f\} \in \mathcal{M}(\mathbb{R}^d)$$

  - Classical Beppo-Levi spaces: $(\mathcal{M}(\mathbb{R}^d), \mathrm{L}) \to (L_p(\mathbb{R}), \mathrm{D}^n)$ (Deny-Lions, 1954)
  - Inclusion of non-uniform L-splines

$$s = \sum_k a_k \rho_{\mathrm{L}}(\cdot - \boldsymbol{x}_k) + \sum_{n=1}^{N_0} b_n p_n \quad \Rightarrow \quad \mathrm{L}\{s\} = \sum_k a_k \delta(\cdot - \boldsymbol{x}_k)$$

$$\mathrm{gTV}(s) = \|\mathrm{L}\{s\}\|_{\mathrm{TV}} = \sum_k |a_k| = \|\mathbf{a}\|_{\ell_1}$$

New optimality result: universality of splines

L: spline-admissible operator

$$\mathcal{M}_{\mathrm{L}}(\mathbb{R}) = \left\{ f : \mathrm{gTV}(f) = \|\mathrm{L}\{f\}\|_{\mathrm{TV}} = \sup_{\|\varphi\|_\infty \le 1} \langle \mathrm{L}\{f\}, \varphi \rangle < \infty \right\}$$

Generalized total variation : $\mathrm{gTV}(f) = \|\mathrm{L}\{f\}\|_{L_1}$ when $\mathrm{L}\{f\} \in L_1(\mathbb{R}^d)$

Linear measurement operator $\mathcal{M}_{\mathrm{L}}(\mathbb{R}) \to \mathbb{R}^M : f \mapsto \mathbf{z} = \boldsymbol{\nu}(f)$

$$\Leftrightarrow \quad z_m = \langle f, \nu_m \rangle$$

Theorem [U.-Fageot-Ward, 2015]: The **generic linear-inverse** problem

$$\min_{f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R})} \left(\|\mathbf{y} - \boldsymbol{\nu}(f)\|_2^2 + \lambda \|\mathrm{L}\{f\}\|_{\mathrm{TV}} \right)$$

admits a global solution of the form $f(\boldsymbol{x}) = \sum_{k=1}^{K} a_k \rho_{\mathrm{L}}(\boldsymbol{x} - \boldsymbol{x}_k) + \sum_{n=1}^{N_0} b_n p_n(\boldsymbol{x})$

with $K < M$, which is a **non-uniform** L**-spline** with knots $(\boldsymbol{x}_k)_{k=1}^{K}$.

Optimality result for Dirac measures

- $\mathbf{F}$: linear continuous map $\mathcal{M}(\mathbb{R}^d) \to \mathbb{R}^M$
- $\mathcal{C}$: convex compact subset of $\mathbb{R}^M$
- Generic constrained TV minimization problem

$$\mathcal{V} = \arg \min_{w \in \mathcal{M}(\mathbb{R}^d) \,:\, \mathbf{F}(w) \in \mathcal{C}} \|w\|_{\mathrm{TV}}$$

Generalized Fisher-Jerome theorem

The solution set $\mathcal{V}$ is a **convex, weak$*$-compact** subset of $\mathcal{M}(\mathbb{R}^d)$ with **extremal points** of the form

$$w_\delta = \sum_{k=1}^{K} a_k \delta(\cdot - \boldsymbol{x}_k)$$

with $K \le M$ and $\boldsymbol{x}_k \in \mathbb{R}^d$.

Jerome-Fisher, 1975: Compact domain & scalar intervals

Existence of stable right-inverse operator

$$L_{\infty,n_0}(\mathbb{R}^d) = \{f : \mathbb{R}^d \to \mathbb{R} : \sup_{\boldsymbol{x}\in\mathbb{R}^d} (|f(\boldsymbol{x})|(1+\|\boldsymbol{x}\|)^{n_0}) < +\infty\}$$

Theorem 1 [U.-Fageot-Ward, preprint]
Let L be spline-admissible operator with a N_0-dimensional null space $\mathcal{N}_{\mathrm{L}} \subseteq L_{\infty,-n_0}(\mathbb{R}^d)$ such that $p = \sum_{n=1}^{N_0} \langle p, \phi_n \rangle p_n$ for all $p \in \mathcal{N}_{\mathrm{L}}$. Then, there exists a **unique and stable operator** $\mathrm{L}_{\boldsymbol{\phi}}^{-1} : \mathcal{M}(\mathbb{R}^d) \to L_{\infty,-n_0}(\mathbb{R}^d)$ such that, for all $w \in \mathcal{M}(\mathbb{R}^d)$,

- Right-inverse property: $\mathrm{L}\mathrm{L}_{\boldsymbol{\phi}}^{-1} w = w$,
- Boundary conditions: $\langle \boldsymbol{\phi}, \mathrm{L}_{\boldsymbol{\phi}}^{-1} w \rangle = \mathbf{0}$ with $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_{N_0})$.

Its generalized impulse response $g_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{\delta(\cdot - \boldsymbol{y})\}(\boldsymbol{x})$ is given by

$$g_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}) = \rho_{\mathrm{L}}(\boldsymbol{x} - \boldsymbol{y}) - \sum_{n=1}^{N_0} p_n(\boldsymbol{x}) q_n(\boldsymbol{y})$$

with ρ_{L} such that $\mathrm{L}\{\rho_{\mathrm{L}}\} = \delta$ and $q_n(\boldsymbol{y}) = \langle \phi_n, \rho_{\mathrm{L}}(\cdot - \boldsymbol{y}) \rangle$.

Characterization of generalized Beppo-Levi spaces

- Regularization operator $\mathrm{L} : \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d)$

$$f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \quad \Leftrightarrow \quad \mathrm{gTV}(f) = \|\mathrm{L}\{f\}\|_{\mathrm{TV}} < \infty$$

Theorem 2 [U.-Fageot-Ward, preprint]
Let L be a spline-admissible operator that admits a stable right-inverse $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ of the form specified by Theorem 1. Then, any $f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ has a unique representation as

$$f = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w + p,$$

where $w = \mathrm{L}\{f\} \in \mathcal{M}(\mathbb{R}^d)$ and $p = \sum_{n=1}^{N_0} \langle \phi_n, f \rangle p_n \in \mathcal{N}_{\mathrm{L}}$ with $\phi_n \in \big(\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)\big)'$. Moreover, $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \subseteq L_{\infty,-n_0}(\mathbb{R}^d)$ and is a Banach space equipped with the norm

$$\|f\|_{\mathcal{M}_{\mathrm{L}},\boldsymbol{\phi}} = \|\mathrm{L}f\|_{\mathrm{TV}} + \|\langle f, \boldsymbol{\phi} \rangle\|_2.$$

- Generalized Beppo-Levi space: $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) = \mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d) \oplus \mathcal{N}_{\mathrm{L}}$

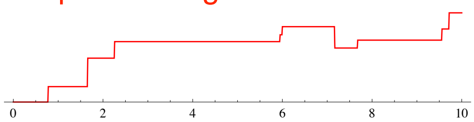
$$\mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d) = \big\{ f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) : \langle \boldsymbol{\phi}, f \rangle = \mathbf{0} \big\}$$

$$\mathcal{N}_{\mathrm{L}} = \big\{ p \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) : \mathrm{L}\{p\} = 0 \big\}$$

Link with sparse stochastic processes

Random spline: archetype of sparse signal

non-uniform spline of degree 0



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

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: $\mathrm{D}_{\phi_1}^{-1}\varphi(t) = \int_0^t \varphi(\tau)\mathrm{d}\tau$ (see **Theorem 1** with $\phi_1 = \delta$)

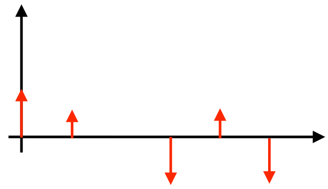
Compound Poisson process

- Stochastic differential equation

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

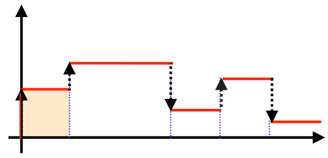
  with boundary condition $s(0) = \langle \phi_1, s \rangle = 0$ with $\phi_1 = \delta$

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



- Formal solution

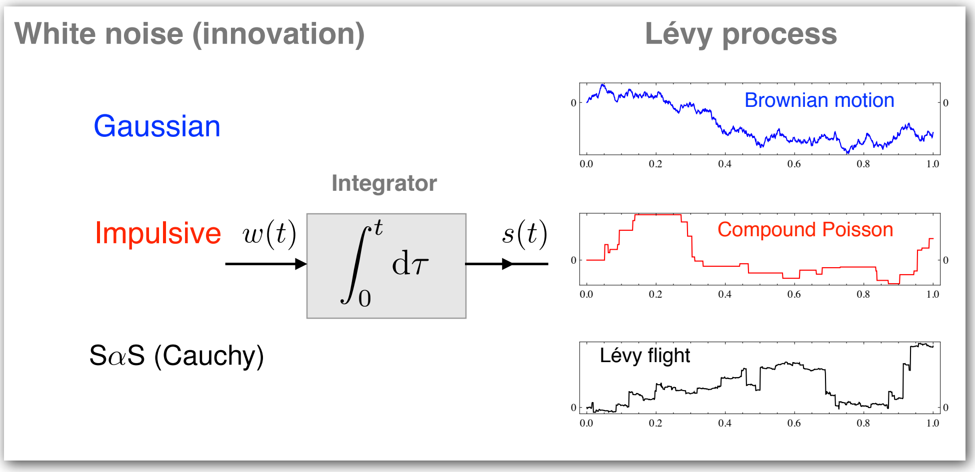
$$s(t) = \mathrm{D}_{\phi_1}^{-1} w(t) = \sum_n a_n \mathrm{D}_{\phi_1}^{-1}\{\delta(\cdot - t_n)\}(t)$$
$$= b_1 + \sum_n a_n \mathbb{1}_+(t - t_n)$$



Lévy processes: all admissible brands of innovations

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

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



(Wiener 1923)

(Paul Lévy circa 1930)

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 (general operator)

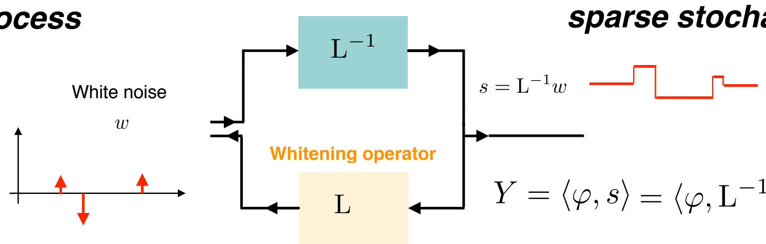
②

innovation process

sparse stochastic process

①

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



③

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

④ Approximate decoupling

Proper definition of ***continuous-domain*** white noise

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

Regularization operator vs. wavelet analysis

Main feature: inherent sparsity
(few significant coefficients)

From Dirac impulses to innovation processes

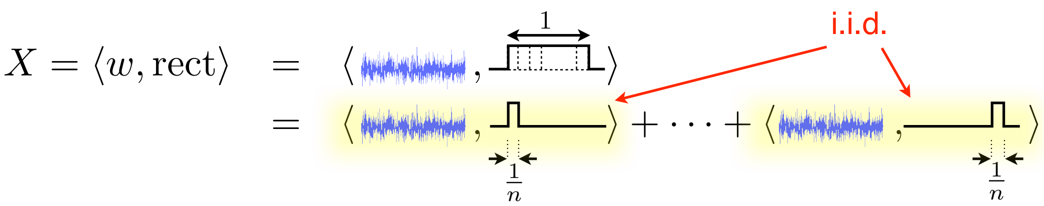
w is a generalized innovation process (or continuous-domain white noise) in $\mathcal{S}'(\mathbb{R}^d)$ if

1. ***Observability***: $X = \langle \varphi, w \rangle$ is an ordinary random variable for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$.
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.

Theorem (under mild technical conditions) (Amini-U., *IEEE-IT* 2014)
w is an innovation process in $\mathcal{S}'(\mathbb{R}^d)$

$\Rightarrow \quad X = \langle \varphi, w \rangle$ is well defined and **infinitely divisible** for any $\varphi \in L_p(\mathbb{R}^d)$

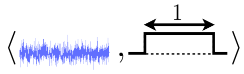
Definition: A random variable X with generic pdf $p_{\mathrm{id}}(x)$ is **infinitely divisible** (id) iff., for any $N \in \mathbb{Z}^+$, there exist i.i.d. random variables $X_1, \dots, X_N$ such that $X \overset{\mathrm{d}}{=} X_1 + \cdots + X_N$.



Probability laws of innovations are infinite divisible

- Canonical observation through a rectangular test function

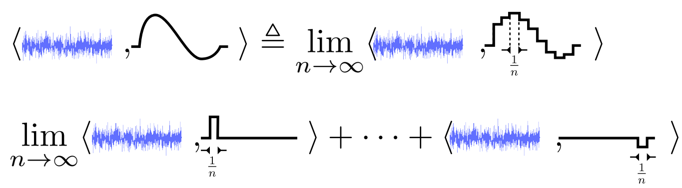
$$X_{\rm id} = \langle w, \mathrm{rect}\rangle \quad = \quad$$



w innovation process $\Leftrightarrow$ $X_{\rm id} = \langle w, \mathrm{rect}\rangle$ infinitely divisible
with **canonical Lévy exponent** $f(\omega) = \log \widehat{p}_{\rm id}(\omega)$

- Statistical description of white Lévy noise w (innovation)
  - Generic observation: $X = \langle \varphi, w\rangle$ with $\varphi \in L_p(\mathbb{R}^d)$

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



  - X is ***infinitely divisible*** with (modified) Lévy exponent

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

☒ Probability laws of sparse processes are id

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

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

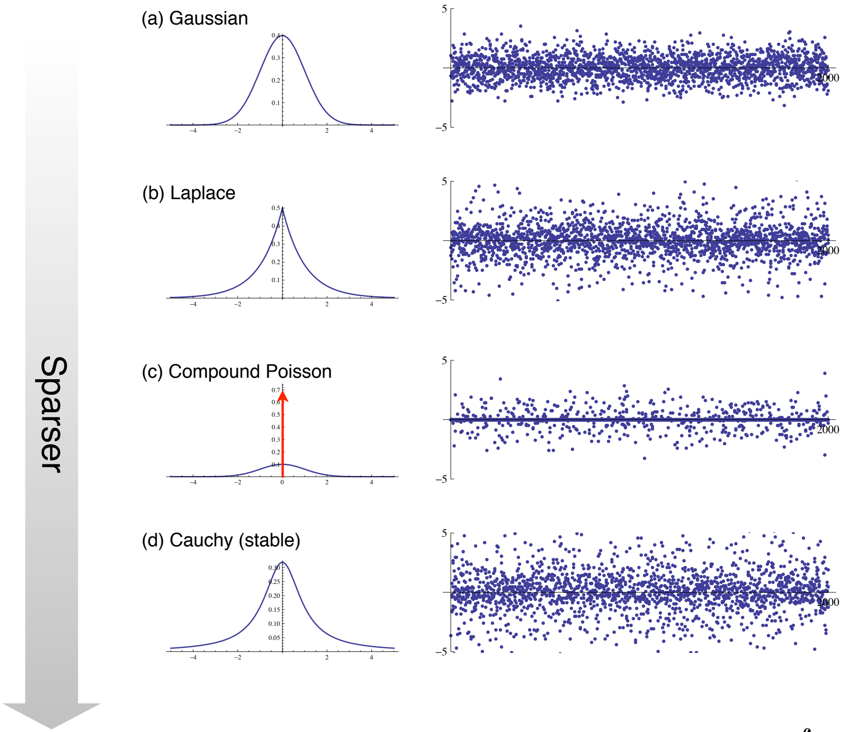
$$\Rightarrow \quad p_Y(y) = \mathcal{F}^{-1}\{\mathrm{e}^{f_\phi(\omega)}\}(y) = \int_{\mathbb{R}} \mathrm{e}^{f_\phi(\omega) - \mathrm{j}\omega y}\frac{\mathrm{d}\omega}{2\pi}$$

= explicit form of pdf

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

Examples of infinitely divisible laws

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



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

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

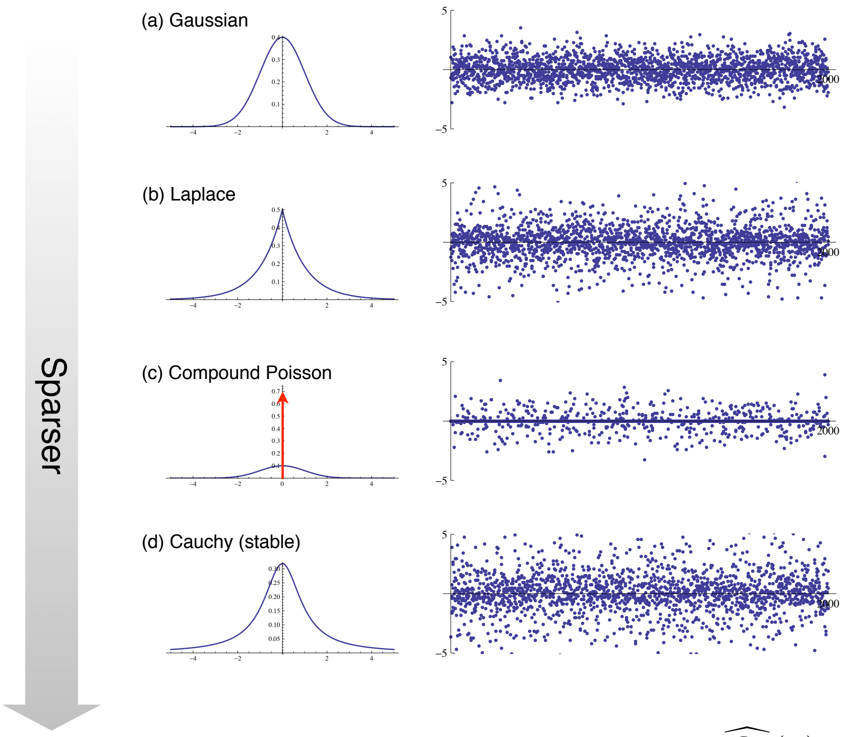
$$p_{\mathrm{Poisson}}(x) = \mathcal{F}^{-1}\{\mathrm{e}^{\lambda(\hat{p}_A(\omega)-1)}\}$$

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

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

Examples of id noise distributions

$p_{\mathrm{id}}(x)$ Observations: $X_n = \langle w, \mathrm{rect}(\cdot - n)\rangle$



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

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

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

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

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

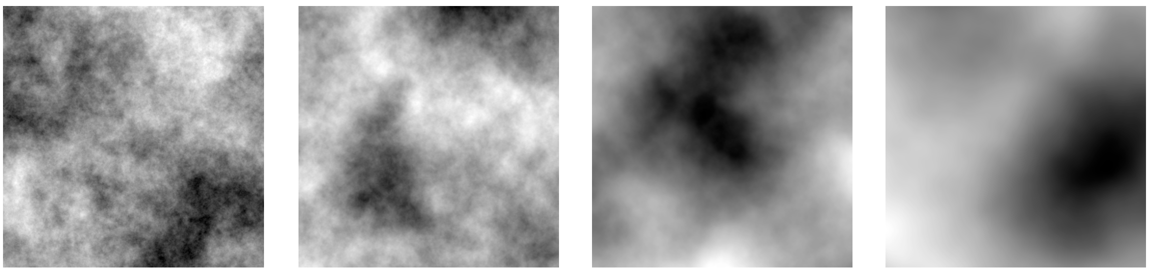
Aesthetic sparse signal: the Mondrian process

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

Scale- and rotation-invariant processes

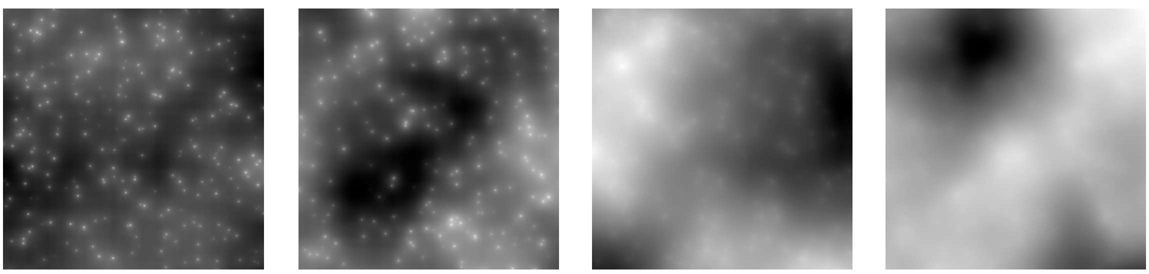
Stochastic partial differential equation : $(-\Delta)^{\frac{H+1}{2}} s(\boldsymbol{x}) = w(\boldsymbol{x})$

Gaussian



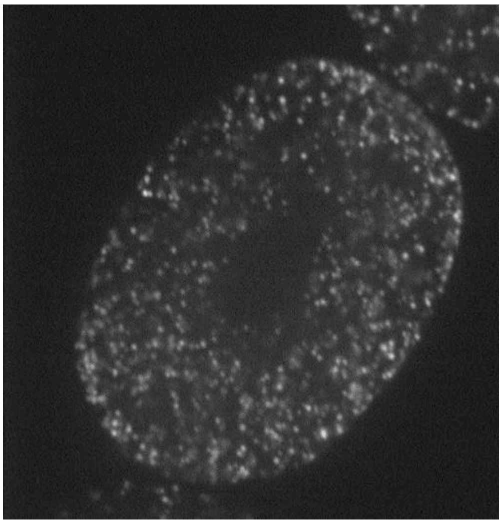
H=.5 H=.75 H=1.25 H=1.75

Sparse (generalized Poisson)



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

Powers of ten: from astronomy to biology



High-level properties of SSP

- **Infinite divisible probability laws**: broadest class of distributions preserved through linear transformation.
- **Explicit calculations**: Analytical determination of transform-domain statistics (including, joint pdfs).
- **Unifying framework**: includes all traditional families of stochastic processes (ARMA, fBm), as well as their non-Gaussian generalizations.
- **Sparsifying transforms / ICA**: SSP are (approximately) decoupled in a matched operator-like wavelet basis. (Pad-U., *IEEE-SP* 2015)
- N**-term approximation properties**: SSP are truly "sparse" as described by their inclusion in (weighted) Besov spaces. (Fageot et al., *ACHA* 2015)

Unser and Tafti An Introduction to Sparse Stochastic Processes CAMBRIDGE

STATISTICAL SIGNAL RECONSTRUCTION

- Discretization of reconstruction problem
- Signal reconstruction algorithm (MAP)

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

$\mathbf{u} = \mathbf{Ls}$ (matrix notation)

p_U is part of **infinitely divisible** family

- Physical model: image formation and acquisition

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

$$\mathbf{y} = \mathbf{y}_0 + \mathbf{n} = \mathbf{Hs} + \mathbf{n}$$

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

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

Posterior probability distribution

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

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

$$\mathbf{u} = \mathbf{L}\mathbf{s} \qquad \Rightarrow \qquad p_S(\mathbf{s}) \propto p_U(\mathbf{L}\mathbf{s}) \approx \prod_{\boldsymbol{k}\in\Omega} p_U([\mathbf{L}\mathbf{s}]_{\boldsymbol{k}})$$

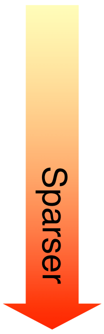
- Additive white Gaussian noise scenario (AWGN)

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

... and then take the log and maximize ...

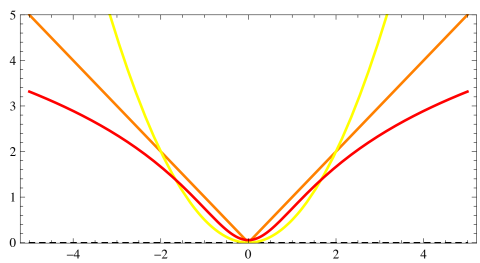
General form of MAP estimator

$$\mathbf{s}_{\text{MAP}} = \operatorname{argmin}\left(\tfrac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 + \sigma^2 \sum_n \Phi_U([\mathbf{L}\mathbf{s}]_n)\right)$$

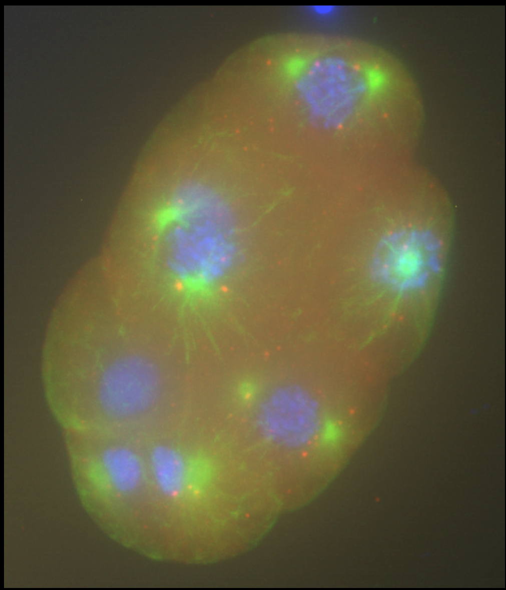


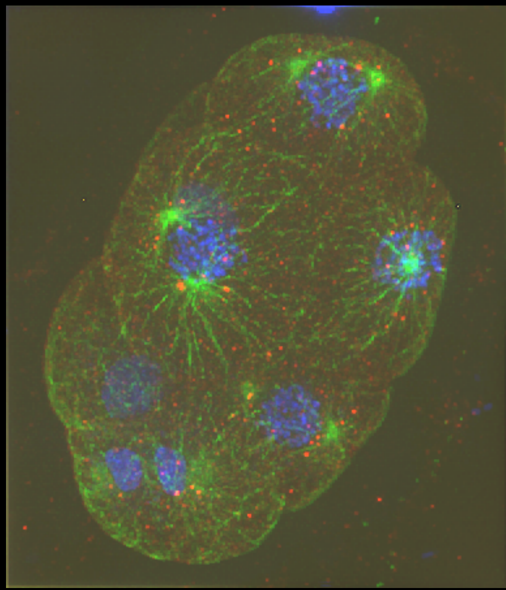
- Gaussian: $p_U(x) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-x^2/(2\sigma_0^2)}$ $\Rightarrow$ $\Phi_U(x) = \frac{1}{2\sigma_0^2}x^2 + C_1$
- Laplace: $p_U(x) = \frac{\lambda}{2} e^{-\lambda|x|}$ $\Rightarrow$ $\Phi_U(x) = \lambda|x| + C_2$
- Student: $p_U(x) = \frac{1}{B\left(r, \frac{1}{2}\right)}\left(\frac{1}{x^2+1}\right)^{r+\frac{1}{2}}$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

Potential: $\Phi_U(x) = -\log p_U(x)$



3D deconvolution with sparsity constraints

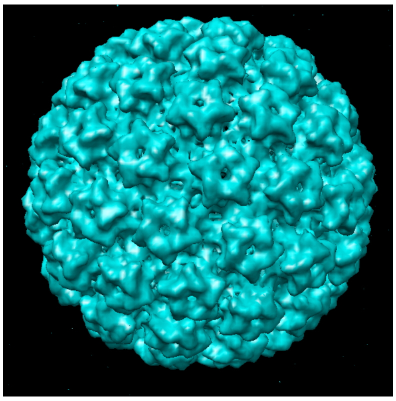
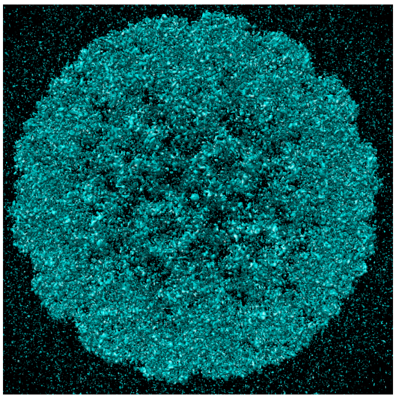
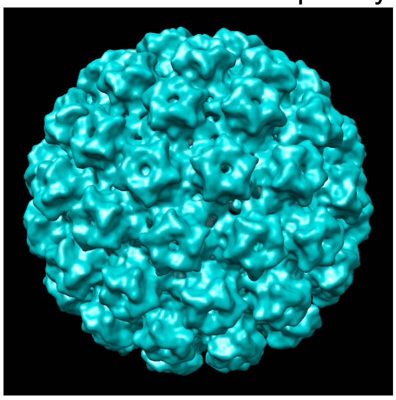




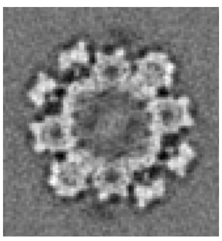
Maximum intensity projections of 384×448×260 image stacks;
Leica DM 5500 widefield epifluorescence microscope with a 63× oil-immersion objective;
C. Elegans embryo labeled with Hoechst, Alexa488, Alexa568;

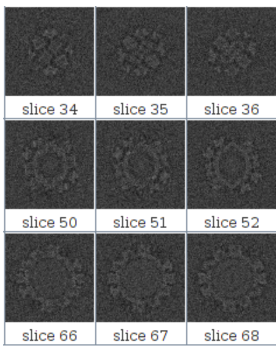
(Vonesch-U. *IEEE Trans. Im. Proc.* 2009)

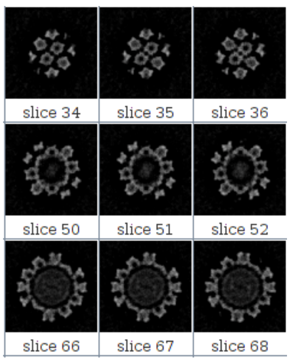
Cryo-electron tomography (real data)

Standard Fourier-based reconstruction	High-resolution Fourier-based reconstruction	High-resolution reconstruction with sparsity
		

6.185 Å







SUMMARY: Sparsity in infinite dimensions

- Continuous-domain formulation $s \in \mathcal{X}$
  - Linear measurement model $s \mapsto \mathbf{y} = \mathrm{H}\{s\}$
  - Linear signal model: PDE $\mathrm{L}s = w$
  - L-splines = signals with "sparsest" innovation $\Rightarrow \quad s = \mathrm{L}^{-1}w$
- Deterministic optimality result $\mathrm{gTV}(s) = \|\mathrm{L}s\|_{\mathrm{TV}}$
  - gTV **regularization**: favors "sparse" innovations
  - Non-uniform L-splines: ***universal*** solutions of linear inverse problems
- Statistical model that supports sparsity
  - Statistical **decoupling**:
    Gaussian vs. **sparse** innovations (Poisson, student, SαS)
  - Unifying framework: "sparse stochastic processes" $s = \mathrm{L}^{-1}w$
  - MAP enforces sparsity through non-quadratic regularization

Acknowledgments

Many thanks to (former) members of EPFL's Biomedical Imaging Group

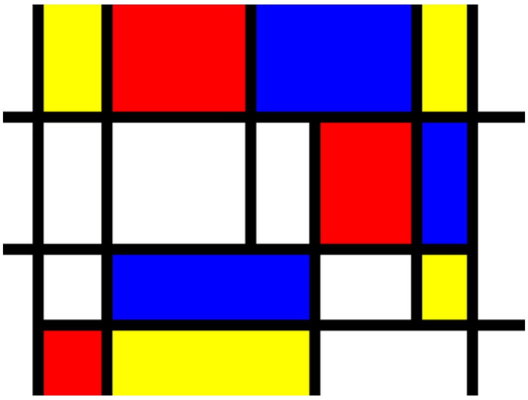
- Dr. Pouya Tafti
- Prof. Arash Amini
- Dr. John-Paul Ward
- Julien Fageot
- Emrah Bostan
- Dr. Masih Nilchian
- Dr. Ulugbek Kamilov
- Dr. Cédric Vonesch
- ....

and collaborators ...

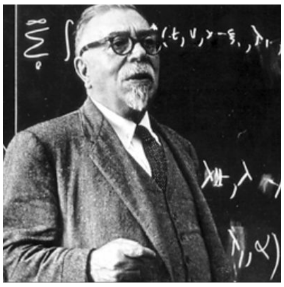
- Prof. Demetri Psaltis
- Prof. Marco Stampanoni
- Prof. Carlos-Oscar Sorzano
- Dr. Arne Seitz
- ....

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

Gaussian vs. Sparse

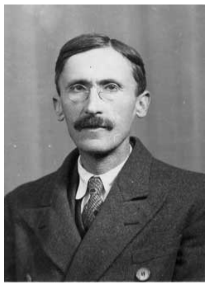


Fourier analysis Splines Wavelet analysis



Norbert Wiener

Isaac Schoenberg



Paul Lévy

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. Nilchian, C. Vonesch, S. Lefkimmiatis, P. Modregger, M. Stampanoni, M. Unser, "Constrained Regularized Reconstruction of X-Ray-DPCI Tomograms with Weighted-Norm," *Optics Express*, vol. 21, no. 26, pp. 32340-32348, 2013.
  - U.S. Kamilov, I.N. Papadopoulos, M.H. Shoreh, A. Goy, C. Vonesch, M. Unser, D. Psaltis, "Learning Approach to Optical Tomography," *Optica*, vol. 2, no. 6, pp. 517-522, June 2015.