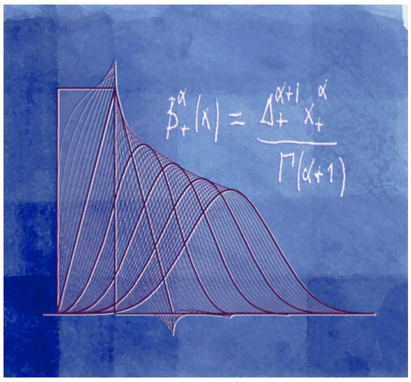
Sparsity and inverse problems: Think analog, act digital

Michael Unser
Biomedical Imaging Group
EPFL, Lausanne, Switzerland



Plenary talk, *Int. Conf. Acoust. Speech Sig. Proc.* (ICASSP), March 20-25, 2016, Shanghai, China

OUTLINE

- **Brief history of inverse problems in imaging**
  - Classical linear reconstruction methods
  - The sparsity (r)evolution
- **Think analog: the spline connection**
  gTV and new optimality results
- **Think analog & statistical**
  Introduction to **sparse stochastic processes**
- **Act digital: Algorithm design**
  Reconstruction of biomedical images

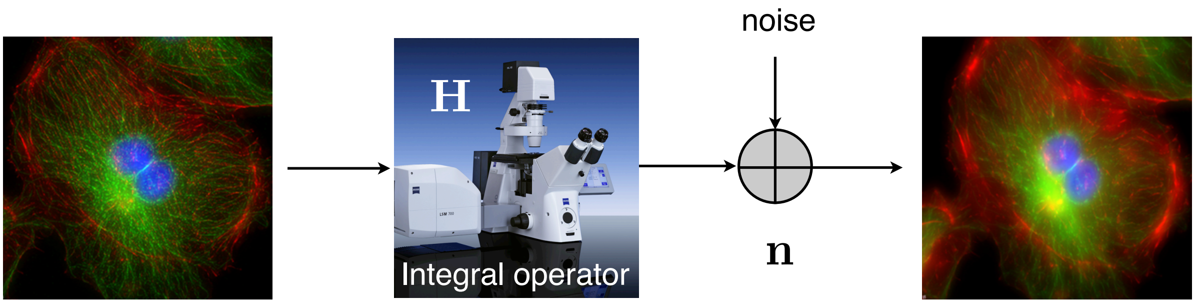
Specific examples:

Deconvolution microscopy
Computed tomography
Differential phase-contrast tomography

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 $\mathbf{y}$

- The easy scenario

Inverse problem is **well**

$\Rightarrow \quad \mathbf{s} \approx \mathbf{H}^{-1}\mathbf{y}$

- Backprojection (pc

Basic limitations

1) Inherent noise amplification
2) Difficulty to invert **H** (too large or non-square)
3) All interesting inverse problems are **ill-posed**

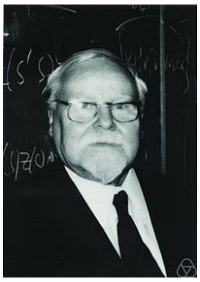
Linear inverse problems (20th century theory)

- Dealing with **ill-posed problems**: Tikhonov regularization

$\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_2^2$: regularization (or smoothness) functional

$\mathbf{L}$: regularization operator (i.e., Gradient)

$$\min_{\mathbf{s}} \|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 \quad \text{subject to} \quad \mathcal{R}(\mathbf{s}) \le C_0$$



Andrey N. Tikhonov (1906-1993)

- Equivalent variational problem

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

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

Interpretation: "filtered" backprojection

Statistical formulation (20th century)

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

$\mathbf{n}$: additive white Gaussian noise (i. i. d.)

$\mathbf{s}$: realization of Gaussian process with zero-mean and covariance matrix $\mathbb{E}\{\mathbf{s} \cdot \mathbf{s}^T\} = \mathbf{C}_s$

Norbert Wiener (1894-1964)

- Wiener (LMMSE) solution = Gauss MMSE = Gauss MAP

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

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

- Quadratic regularization (Tikhonov)

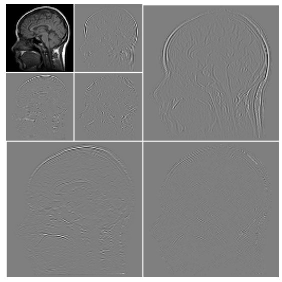
$$\mathbf{s}_{\mathrm{Tik}} = \arg\min_{\mathbf{s}} \left(\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 + \lambda\mathcal{R}(\mathbf{s})\right) \quad \text{with} \quad \mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_2^2$$

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

Linear inverse problems: The sparsity (r)evolution

(20th Century) $p = 2 \longrightarrow 1$ (21st Century)

$$\mathbf{s}_{\mathrm{rec}} = \arg\min_{\mathbf{s}} \left(\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 + \lambda \mathcal{R}(\mathbf{s}) \right)$$



- Non-quadratic regularization regularization

  $$\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_{\ell_2}^2 \longrightarrow \|\mathbf{L}\mathbf{s}\|_{\ell_p}^p \longrightarrow \|\mathbf{L}\mathbf{s}\|_{\ell_1}$$

  - Total variation (Rudin-Osher, 1992)
    $\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_{\ell_1}$ with $\mathbf{L}$: gradient
  - Wavelet-domain regularization (Figuereido et al., Daubechies et al. 2004)
    $\mathbf{v} = \mathbf{W}^{-1}\mathbf{s}$: wavelet expansion of s (typically, sparse)
    $\mathcal{R}(\mathbf{s}) = \|\mathbf{v}\|_{\ell_1}$

- Compressed sensing/sampling (Candes-Romberg-Tao; Donoho, 2006)

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 nonlinear); 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)

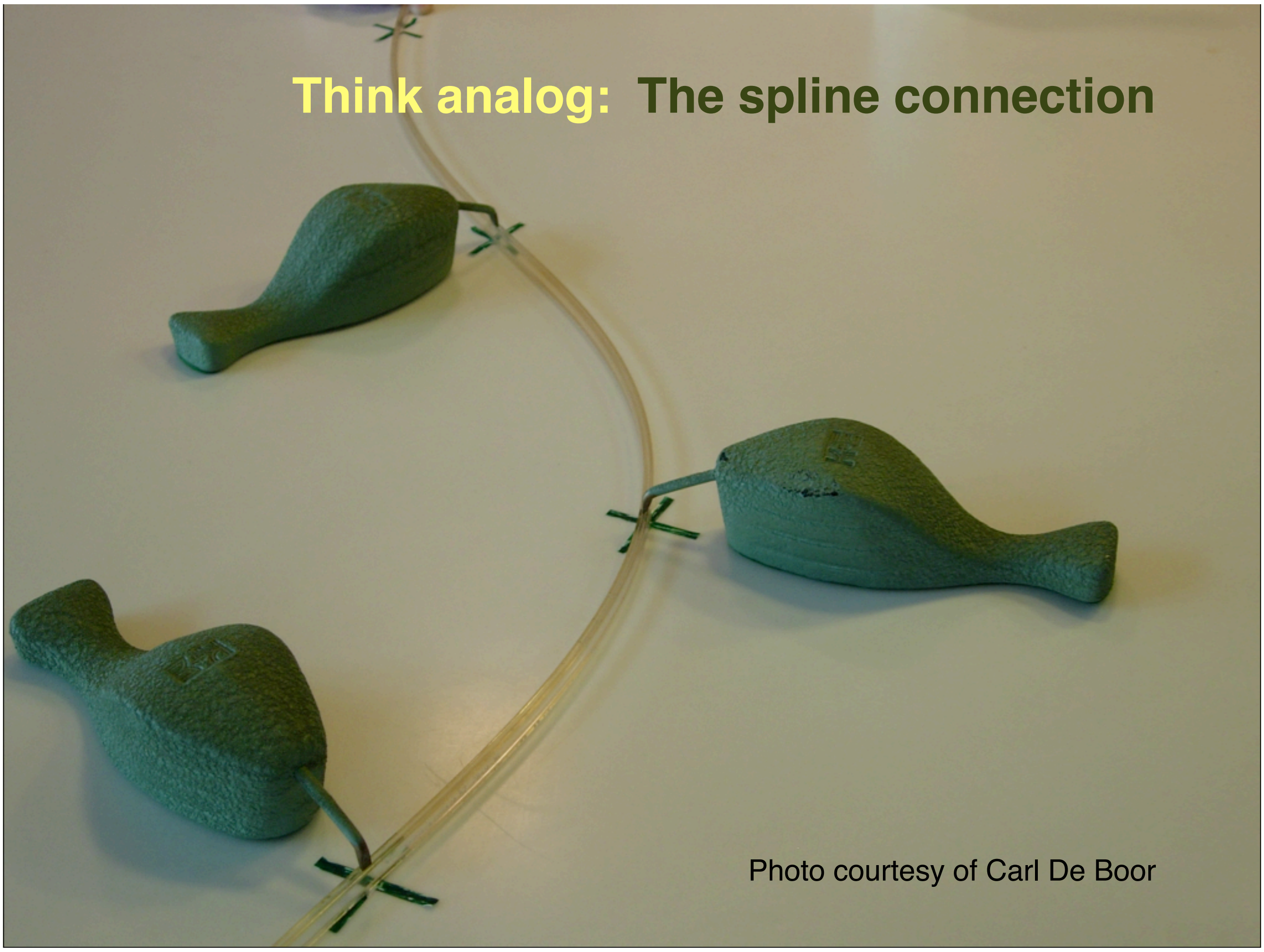
- Outstanding research issues
  - Beyond ℓ_1 and TV: Connection with **statistical modeling & learning**
  - Beyond matrix algebra: **Continuous-domain** formulation
  - Guarantees of **optimality** (either deterministic or statistical)

Sparsity: Continuous-domain formulation

- 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 analog and intrinsically sparse

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

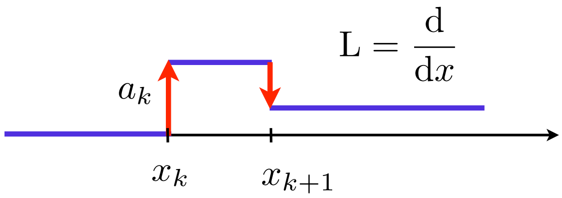
$\delta(\cdot - \boldsymbol{x}_0)$: Dirac impulse shifted by $\boldsymbol{x}_0 \in \mathbb{R}^d$

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)

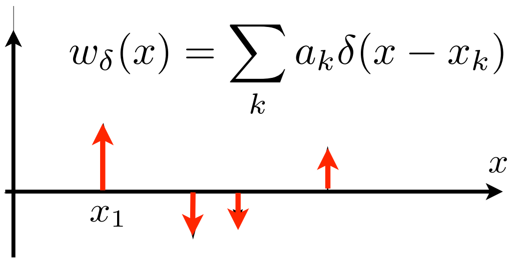


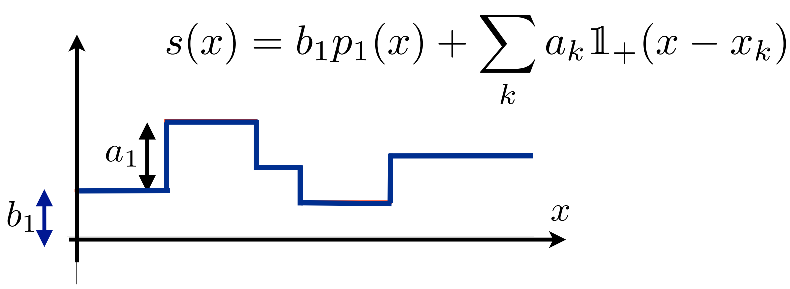
- FRI signal processing: Innovation variables $(2K)$ (Vetterli et al., 2002)
  - Location of singularities (knots) : $\{x_k\}_{k=1}^K$
  - Strength of singularities (linear weights): $\{a_k\}_{k=1}^K$

Spline synthesis: example

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

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





Spline synthesis: generalization

L: spline admissible operator (LSI)

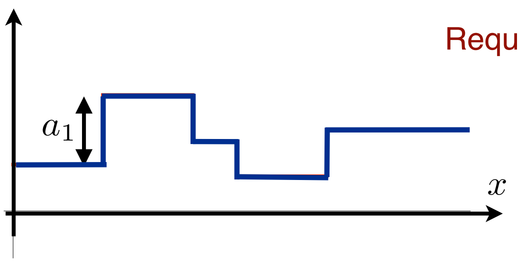
$\rho_{\mathrm{L}}(\boldsymbol{x}) = \mathrm{L}^{-1}\{\delta\}$: Green's function of L

Finite-dimensional null space: $\mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$

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

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

Requires specification of boundary conditions



New optimality result: universality of splines

L: spline-admissible operator

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

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

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

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

Theorem: The **generic linear-inverse** problem

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

admits global solution(s) 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 $K \le M - N_0$, which is a **non-uniform** L**-spline** with knots $(\boldsymbol{x}_k)_{k=1}^{K}$.

(U.-Fageot-Ward, ArXiv 2016)

OUTLINE

- Brief history of inverse problems in imaging ✔
- **Think analog & deterministic** ✔
  Optimality of splines for gTV
- **Think analog & statistical**
  Introduction to sparse stochastic processes
- **Act digital: Algorithm design**
  Reconstruction of biomedical images

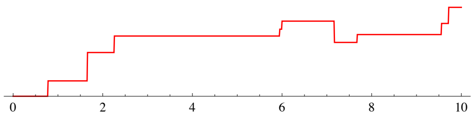
Specific examples:

Deconvolution microscopy
Computed tomography
Differential phase-contrast tomography

An introduction to 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: $\mathrm{D}_0^{-1}\varphi(t) = \int_0^t \varphi(\tau)\mathrm{d}\tau$

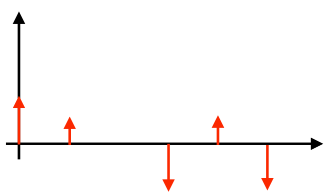
Compound Poisson process

- Stochastic differential equation

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

with boundary condition $s(0) = 0$

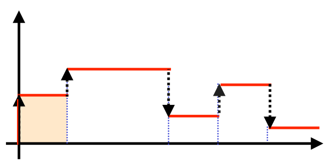
Innovation: $w(t) = \sum_n a_n \delta(t - t_n)$



- Formal solution

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

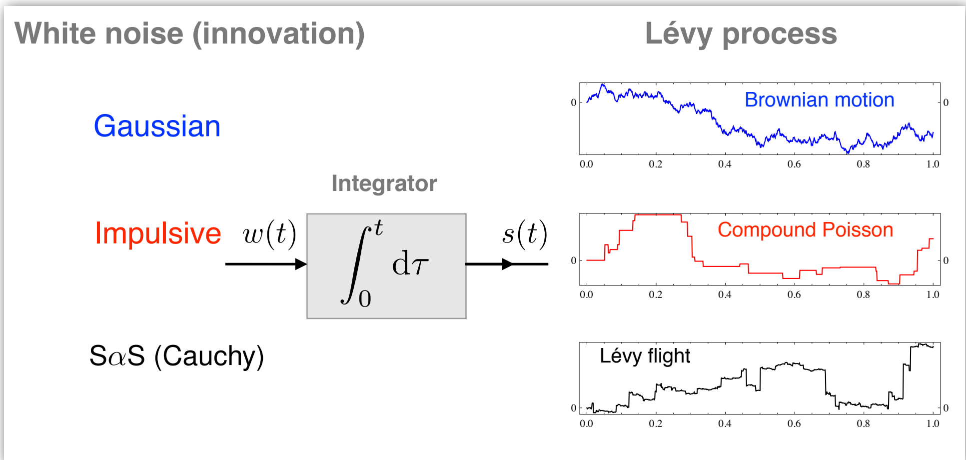
$$= \sum_n a_n \mathbb{1}_+(t - t_n)$$



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)

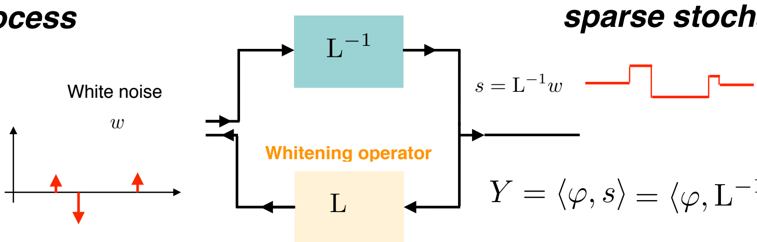
2

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)

Probability laws of innovations are infinite divisible

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

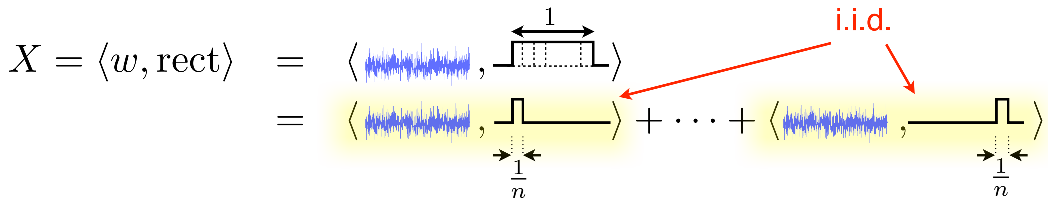
1. ***Observability***: $X = \langle \varphi, w \rangle$ is an ordinary random variable for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$.
2. ***Stationarity*** : $X_{\boldsymbol{x}_0} = \langle \varphi(\cdot - \boldsymbol{x}_0), w \rangle$ is identically distributed for all $\boldsymbol{x}_0 \in \mathbb{R}^d$.
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 $\mathcal{S}'(\mathbb{R}^d)$

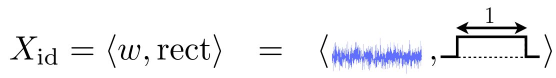
$\Rightarrow$ $X = \langle \varphi, w \rangle$ is well defined and **infinitely divisible** for any $\varphi \in L_p(\mathbb{R}^d)$

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



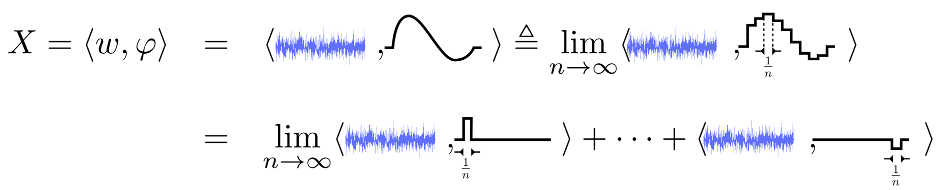
Probability laws of innovations are infinite divisible

- Canonical observation through a rectangular test function



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

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



  - X 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 $f_\phi(\omega) = \int_{\mathbb{R}^d} f\big(\omega\phi(\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x}$

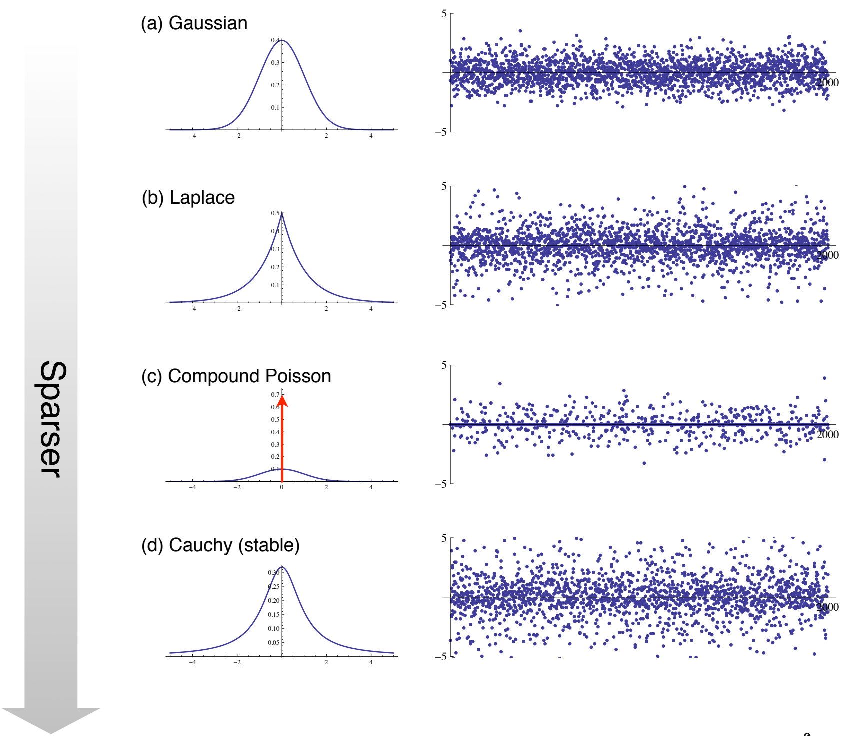
$$\Rightarrow \quad p_Y(y) = \mathcal{F}^{-1}\{\mathrm{e}^{f_\phi(\omega)}\}(y) = \int_{\mathbb{R}} \mathrm{e}^{f_\phi(\omega) - \mathrm{j}\omega y} \frac{\mathrm{d}\omega}{2\pi}$$

= explicit form of pdf

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

Examples of infinitely divisible laws

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



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

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

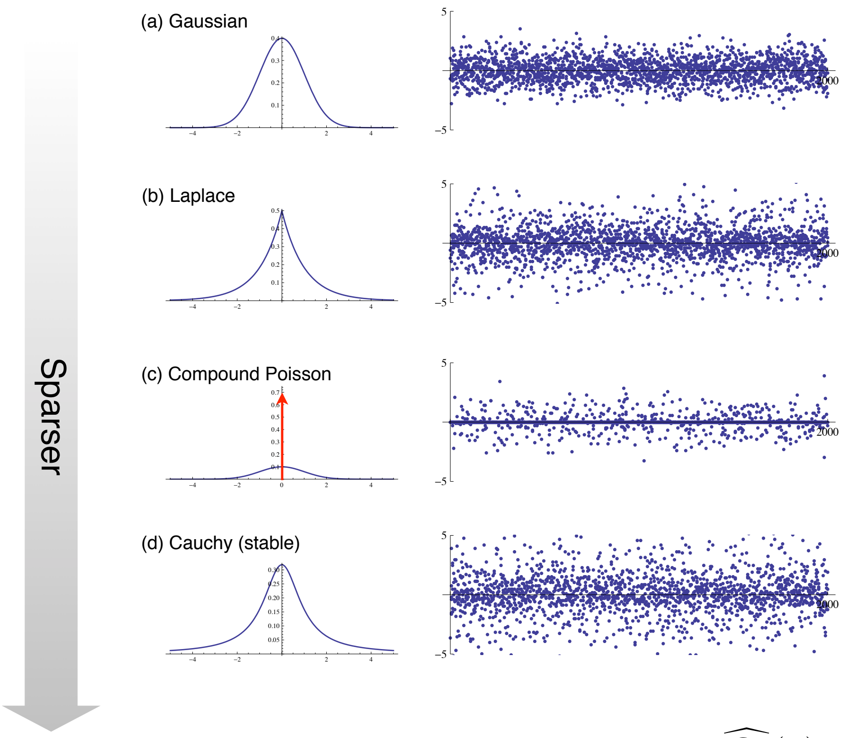
$$p_{\mathrm{Poisson}}(x) = \mathcal{F}^{-1}\{\mathrm{e}^{\lambda(\hat{p}_A(\omega)-1)}\}$$

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

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

Examples of id noise distributions

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



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

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

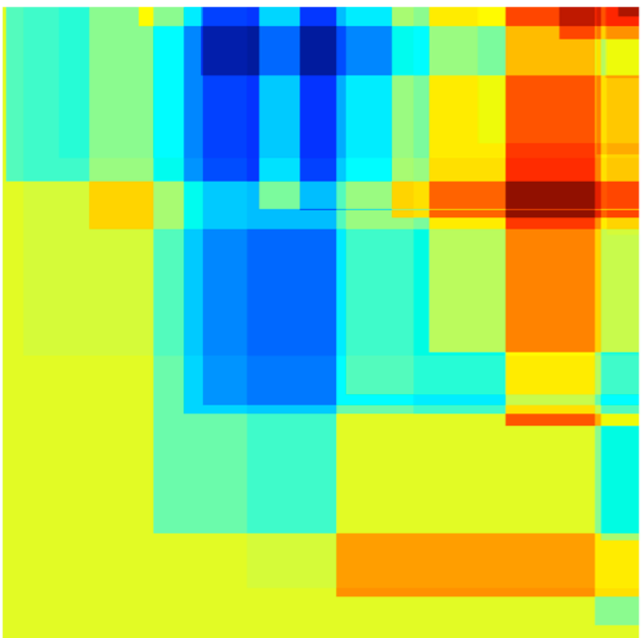
$f(\omega) = \lambda \int_{\mathbb{R}} (e^{jx\omega} - 1) p(x) \mathrm{d}x$

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

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

Aesthetic sparse signal: the Mondrian process

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

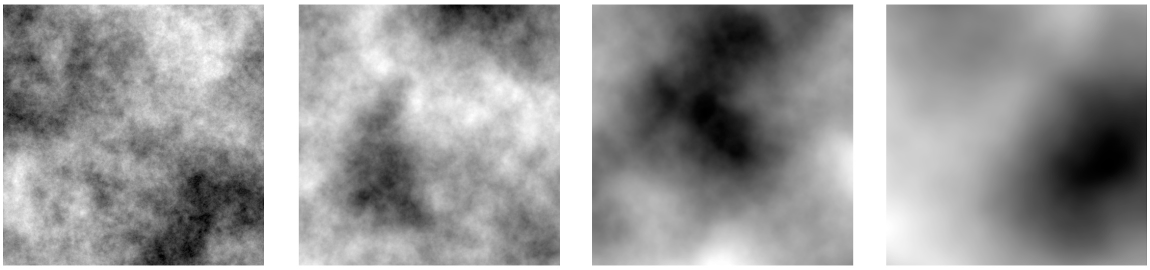


$\lambda = 30$

Scale- and rotation-invariant processes

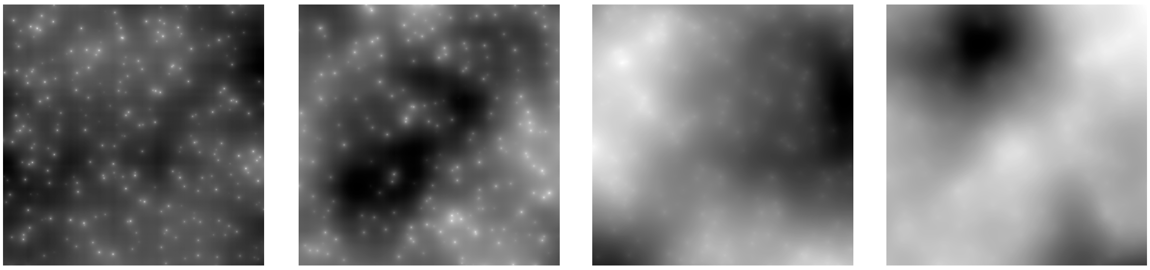
Stochastic partial differential equation : $(-\Delta)^{\frac{H+1}{2}} s(\boldsymbol{x}) = w(\boldsymbol{x})$

Gaussian



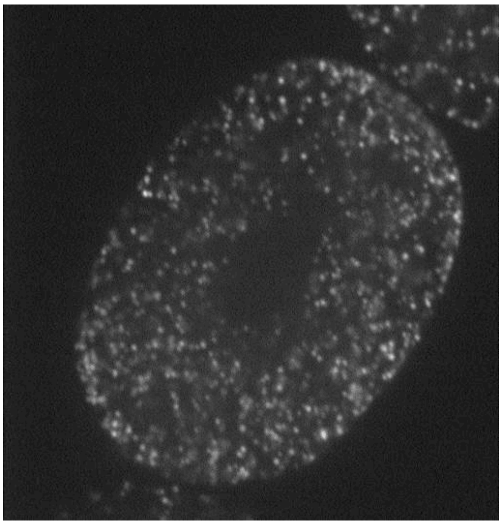
H=.5 H=.75 H=1.25 H=1.75

Sparse (generalized Poisson)



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

Powers of ten: from astronomy to biology



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

OUTLINE

- Brief history of inverse problems in imaging ✔
- **Think analog & deterministic** ✔
  Optimality of splines for gTV
- **Think analog & statistical** ✔
  Sparse stochastic processes
- **Act digital: Reconstruction of biomedical images**
  - Discretization of reconstruction problem
  - Signal reconstruction algorithm (MAP)
  - Examples of image reconstruction

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

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

Posterior probability distribution

$$\begin{aligned} p_{S|Y}(\mathbf{s}|\mathbf{y}) &= \frac{p_{Y|S}(\mathbf{y}|\mathbf{s})p_S(\mathbf{s})}{p_Y(\mathbf{y})} = \frac{p_N(\mathbf{y} - \mathbf{H}\mathbf{s})p_S(\mathbf{s})}{p_Y(\mathbf{y})} \\ &= \frac{1}{Z}p_N(\mathbf{y} - \mathbf{H}\mathbf{s})p_S(\mathbf{s}) \end{aligned}$$

(Bayes' rule)

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

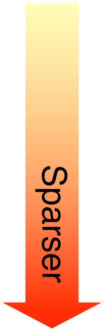
- Additive white Gaussian noise scenario (AWGN)

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

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

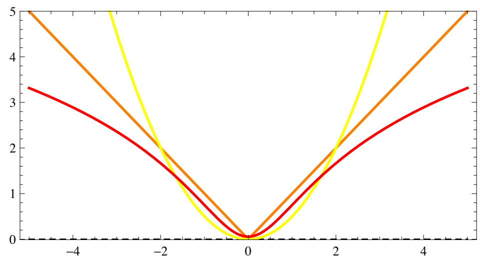
General form of MAP estimator

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



- 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: $p_U(x) = \dfrac{1}{B\left(r,\frac{1}{2}\right)}\left(\dfrac{1}{x^2+1}\right)^{r+\frac{1}{2}}$ $\Rightarrow$ $\Phi_U(x) = \left(r+\dfrac{1}{2}\right)\log(1+x^2) + C_3$

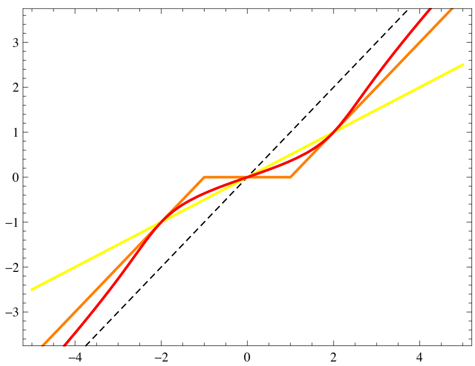
Potential: $\Phi_U(x) = -\log p_U(x)$



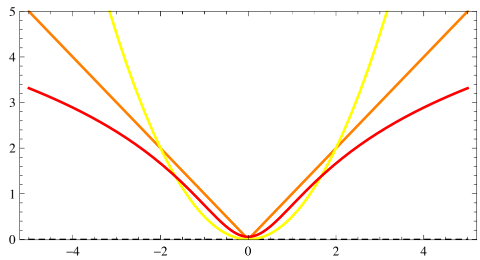
Proximal operator: pointwise denoiser

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

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



$\sigma^2\Phi_U(u)$



- linear attenuation — ℓ_2 minimization
- soft-threshold — ℓ_1 minimization
- shrinkage function $\approx$ ℓ_p relaxation for $p \to 0$

Maximum a posteriori (MAP) estimation

- Constrained optimization formulation

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

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

- Augmented Lagrangian method

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

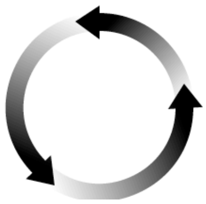
Lagrange multipler vector: $\boldsymbol{\alpha}$

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

Alternating direction method of multipliers (ADMM)

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

Sequential minimization



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

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

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

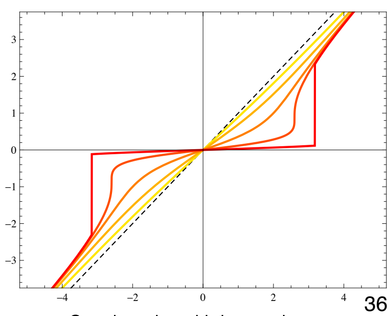
Linear inverse problem: $\mathbf{s}^{k+1} = \left(\mathbf{H}^T\mathbf{H} + \mu\mathbf{L}^T\mathbf{L}\right)^{-1}\left(\mathbf{H}^T\mathbf{y} + \mathbf{z}^{k+1}\right)$

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

Nonlinear denoising: $\mathbf{u}^{k+1} = \mathrm{prox}_{\Phi_U}\left(\mathbf{L}\mathbf{s}^{k+1} + \frac{1}{\mu}\boldsymbol{\alpha}^{k+1}; \frac{\sigma^2}{\mu}\right)$

- Proximal operator taylored to stochastic model

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

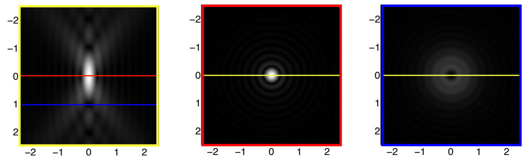


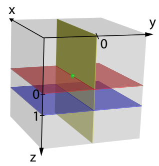
Cauchy prior with increasing s_0

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)

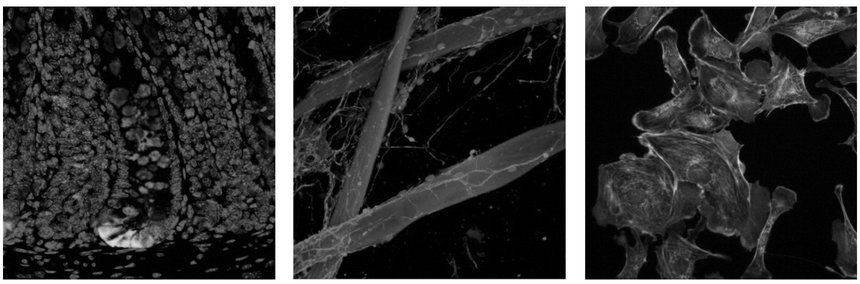
$$h_{3\mathrm{D}}(x,y,z) = I_0 \left| p_\lambda \left(\frac{x}{M}, \frac{y}{M}, \frac{z}{M^2} \right) \right|^2$$

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

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

Deconvolution experiments



(a) (b) (c)

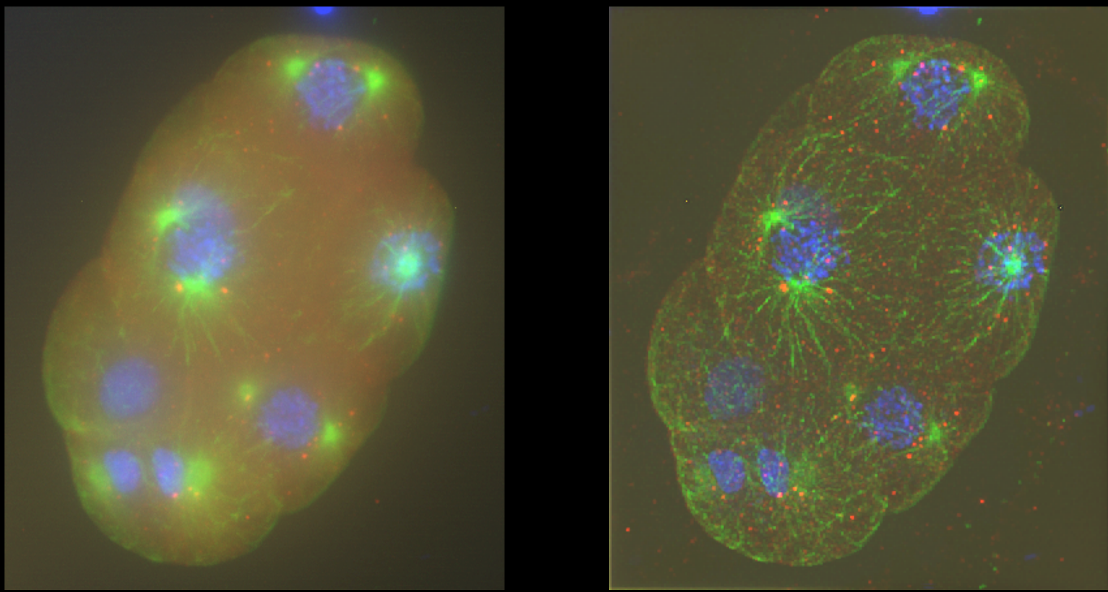
Figure 10.3 Images used in deconvolution experiments. (a) Stem cells surrounded by goblet cells. (b) Nerve cells growing around fibers. (c) Artery cells.

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

$\mathbf{L}$: discrete gradient

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)

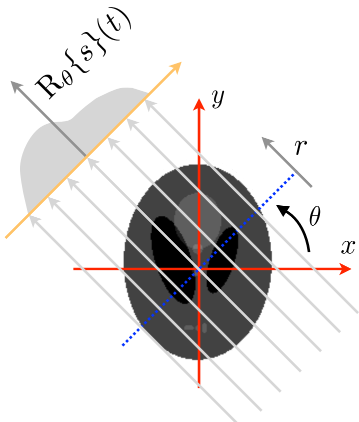
Computed tomography (straight rays)

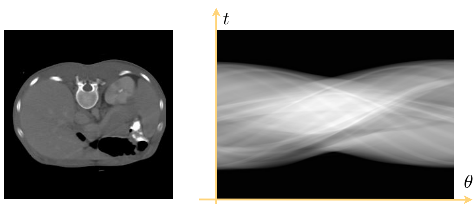
Projection geometry: $\boldsymbol{x} = t\boldsymbol{\theta} + r\boldsymbol{\theta}^{\perp}$ with $\boldsymbol{\theta} = (\cos\theta, \sin\theta)$

- Radon transform (line integrals)

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

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



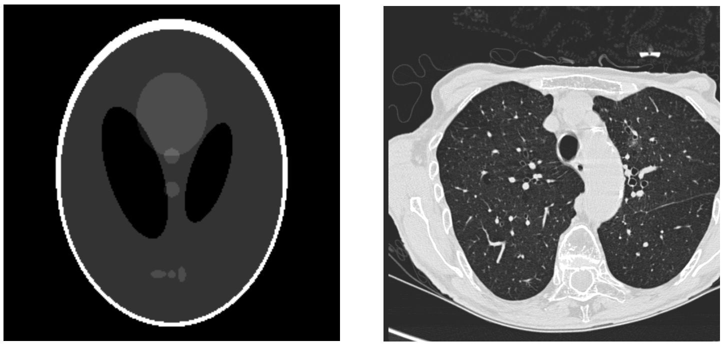


sinogram

(applicable to tomographic phase microscopy with plane wave illumination)

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

Computed 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.

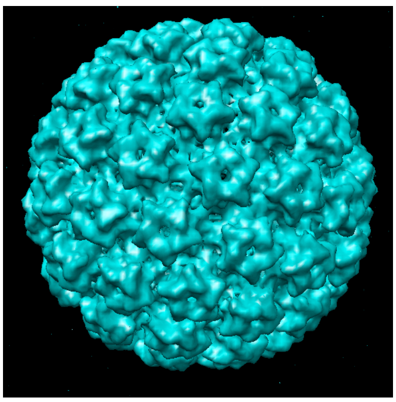
Table 10.4 Reconstruction results of X-ray computed tomography using different estimators.

	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

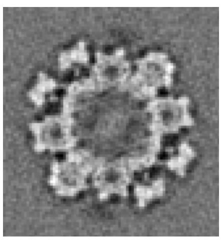
$\mathbf{L}$: discrete gradient

Cryo-electron tomography (real data)

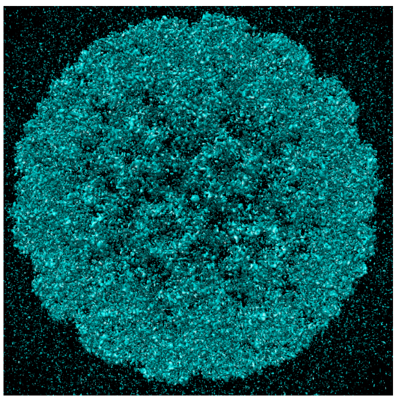
Standard Fourier-based reconstruction

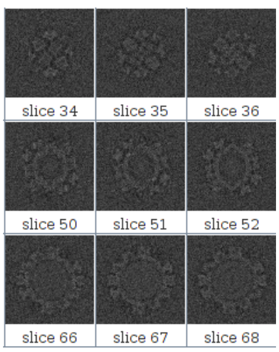


6.185 Å

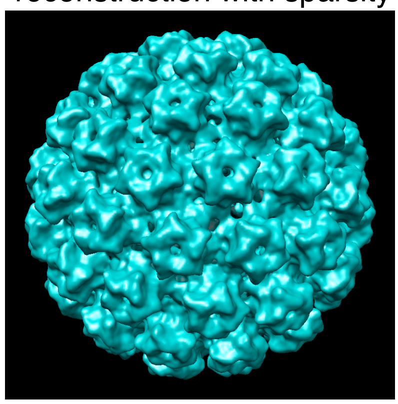


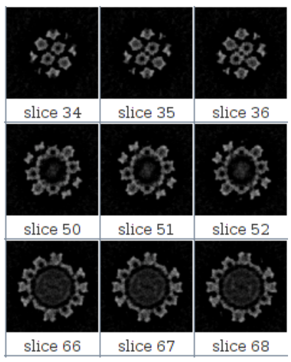
High-resolution Fourier-based reconstruction





High-resolution reconstruction with sparsity





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

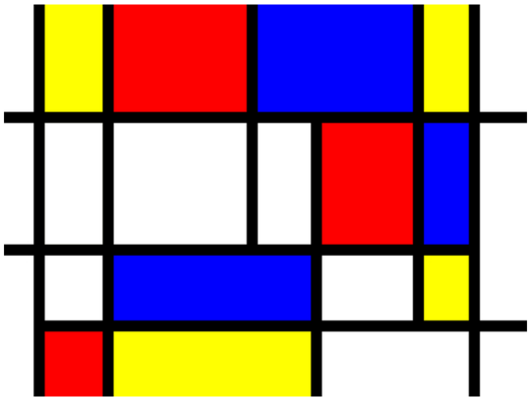
- Dr. Pouya Tafti
- Prof. Arash Amini
- Dr. John-Paul Ward
- Julien Fageot
- Dr. Emrah Bostan
- Dr. Masih Nilchian
- Dr. Ulugbek Kamilov
- Dr. Cédric Vonesch
- ....

and collaborators ...

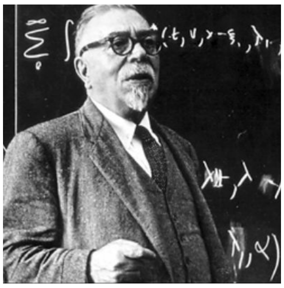
- Prof. Demetri Psaltis
- Prof. Marco Stampanoni
- Prof. Carlos-Oscar Sorzano
- Dr. Arne Seitz
- ....

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

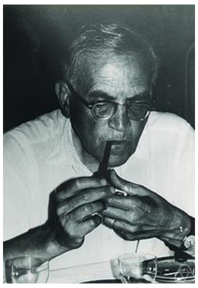
Gaussian vs. Sparse



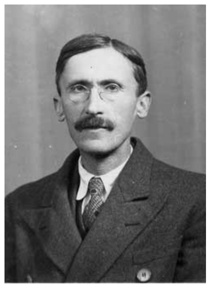
Fourier analysis   Splines   Wavelet analysis



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. 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.

- L_1-optimality of splines for inverse problems

  - M. Unser, J. Fageot, J.P. Ward, "Splines Are Universal Solutions of Linear Inverse Problems with Generalized-TV Regularization," arXiv:1603.01427 [math.FA].

- 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.