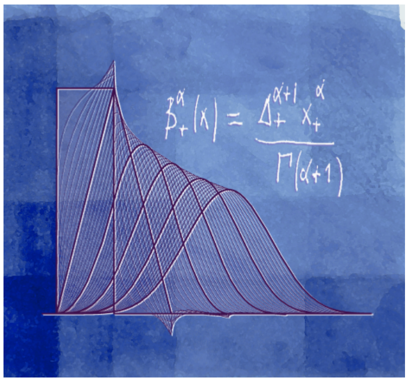
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Computational bioimaging
How to further reduce exposure and/or increase image quality

Michael Unser
Biomedical Imaging Group
EPFL, Lausanne, Switzerland



Plenary talk, Int. Conf. of the IEEE EMBS (EMBC’17), July 11-15, 2017, Jeju Island, Korea.

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

OUTLINE

- From classical to modern image reconstruction
  - The sparsity (r)evolution⇒ Compressed Sensing
- **Sparsity revisited**: the spline connection
- **Non-Gaussian** statistical modeling
  - Sparse stochastic processes
- **Iterative image reconstruction** (MAP formulation)

erc

Examples of biomedical image reconstruction

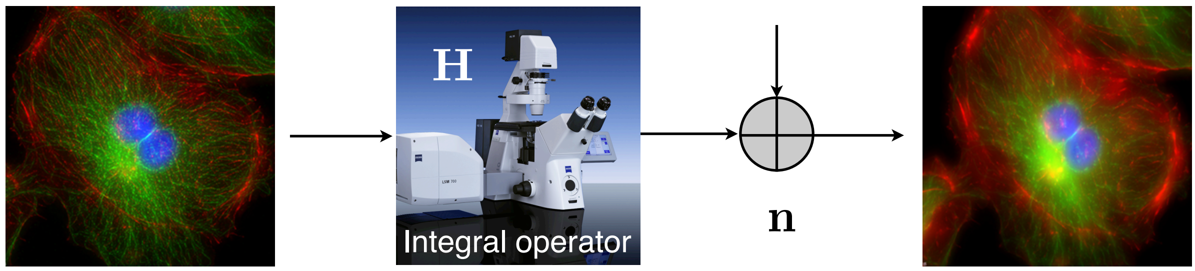
Deconvolution microscopy

Computed tomography

- **Learning**: Emergence of 3rd generation methods
  - SplineProx, FBPConvNet

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 scenari

Hypotheses: $\mathbf{H}$ is we

$\Rightarrow \quad \mathbf{s} \approx (\mathbf{H}^T\mathbf{H}$

- Backprojection (

Basic limitations

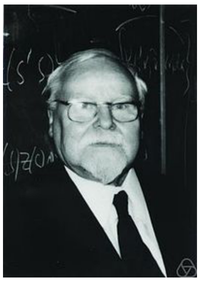
1) Inherent noise amplification
2) Difficulty to invert **H** (too large or non-square)
3) All interesting inverse problems are **ill-posed**

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)



Andrey N. Tikhonov (1906-1993)

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

- 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

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

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

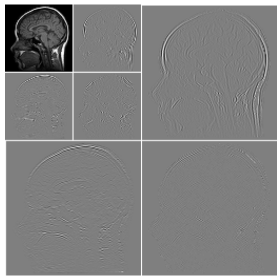
- Quadratic regularization (Tikhonov)

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

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)



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

- 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

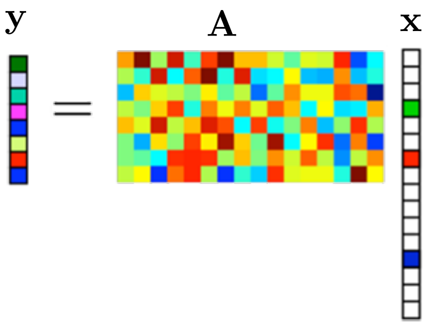
  - Wavelet-domain regularization (Figuereido et al., Daubechies et al. 2004)

    $\mathbf{v} = \mathbf{W}^{-1}\mathbf{s}$: wavelet expansion of s (typically, sparse)

    $\mathcal{R}(\mathbf{s}) = \|\mathbf{v}\|_{\ell_1}$

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

Compressive sensing (CS) and l_1 minimization



+ "noise"

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

Sparse representation of signal: $\mathbf{s} = \mathbf{W}\mathbf{x}$ with $\|\mathbf{x}\|_0 = K \ll N_x$

Equivalent $N_y \times N_x$ **sensing matrix** : $\mathbf{A} = \mathbf{H}\mathbf{W}$

Occam's razor: $\mathbf{x}_{\text{sparse}} = \min_{\mathbf{x}} \|\mathbf{x}\|_0$ subject to $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \leq \sigma^2$

- Relaxed (convex) formulation of recovery problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \leq \sigma^2$$

CS: Three fundamental ingredients

(Donoho, *IEEE T. Inf. Theo.* 2006) (Candès-Romberg, *Inv. Prob.* 2007)

1. Existence of sparsifying transform (**W** or **L**)
   - Wavelet basis
   - Dictionary
   - Differential operator (Gradient)

2. Incoherence of sensing matrix **A**
   - Restricted isometry; few linearly dependent columns (spark)
   - Quasi-random and delocalized structure:
     Gaussian matrix with i.i.d. entries,
     random sampling in Fourier domain

3. Non-linear signal recovery (l_1 minimization)

CS: Examples of applications in imaging

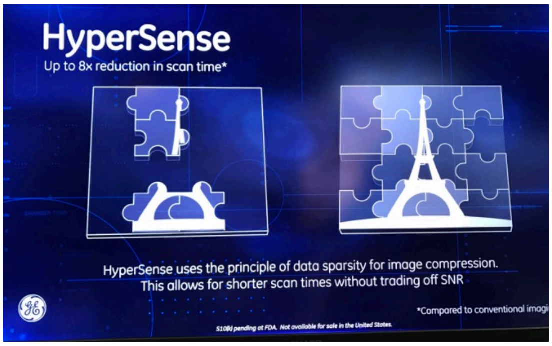
- Magnetic resonance imaging (MRI) (Lustig, *Mag. Res. Im.* 2007)
- Radio Interferometry (Wiaux, *Notic. R. Astro.* 2007)
- Teraherz Imaging (Chan, *Appl. Phys.* 2008)
- Digital holography (Brady, *Opt. Express* 2009; Marim 2010)
- Spectral-domain OCT (Liu, *Opt. Express* 2010)
- Coded-aperture spectral imaging (Arce, *IEEE Sig. Proc.* 2014)
- Localization microscopy (Zhu, *Nat. Meth.* 2012)
- Ultrafast photography (Gao, *Nature* 2014)

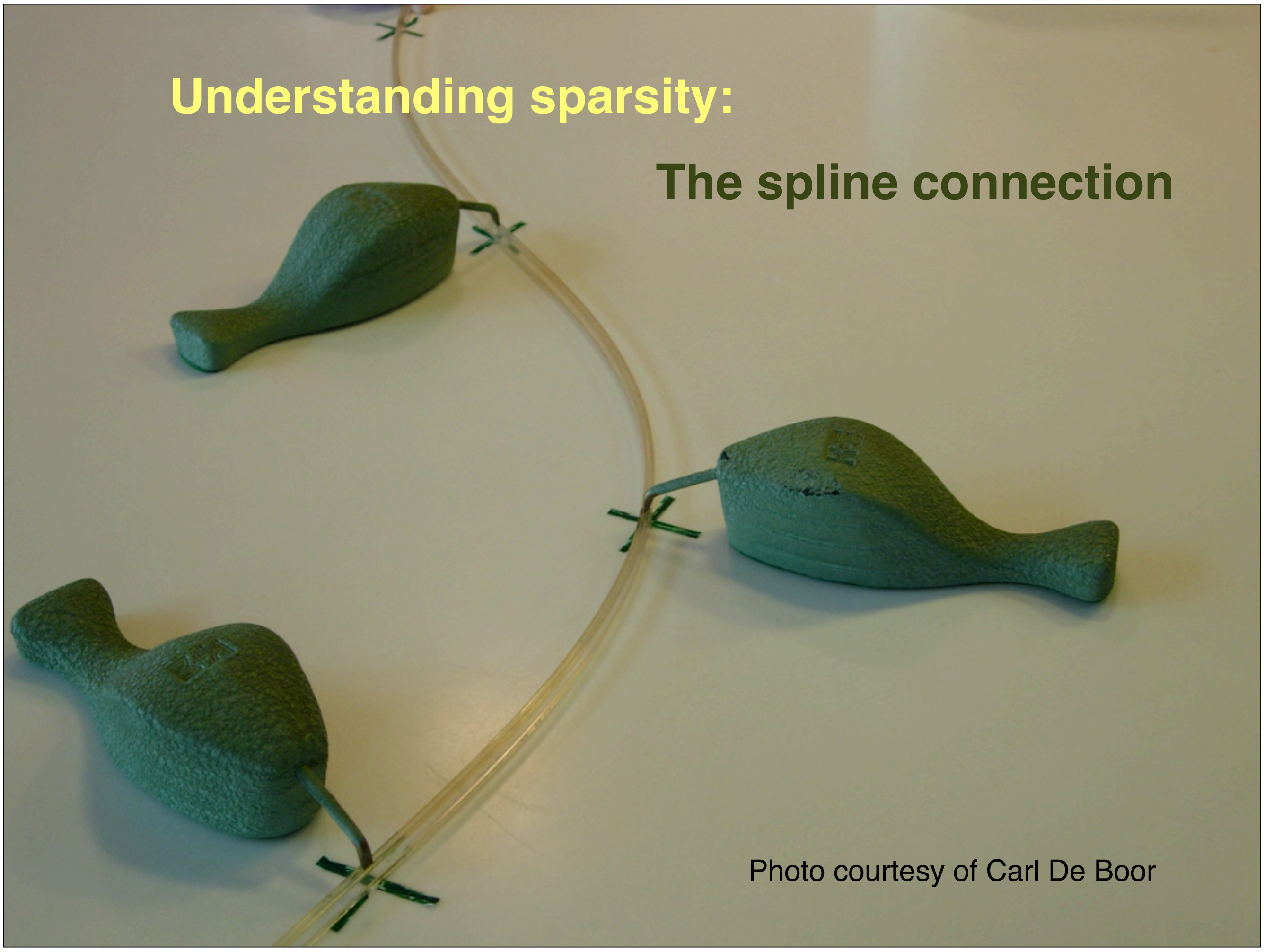
Ten years after: Impact in clinical imaging

Commercial implementation of **sparsity-based** iterative reconstruction algorithms

- Siemens: Compressed Sensing Cardiac Cine
  (FDA approval, Jan 2017)

- GE: HyperSense
  (FDA approval, April 2017)





Understanding sparsity:

The spline connection

Photo courtesy of Carl De Boor

Splines are analog and intrinsically sparse

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

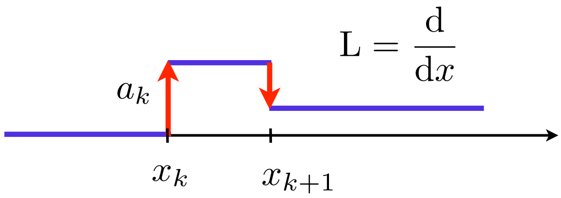
$\delta(\cdot - \boldsymbol{x}_0)$: Dirac impulse shifted by $\boldsymbol{x}_0 \in \mathbb{R}^d$

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)



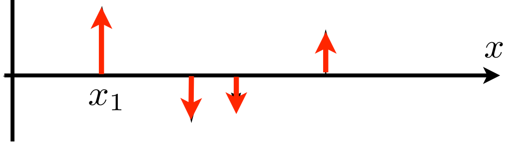
- FRI signal processing: Innovation variables $(2K)$ (Vetterli et al., 2002)
  - Location of singularities (knots) : $\{x_k\}_{k=1}^K$
  - Strength of singularities (linear weights): $\{a_k\}_{k=1}^K$

Spline synthesis: example

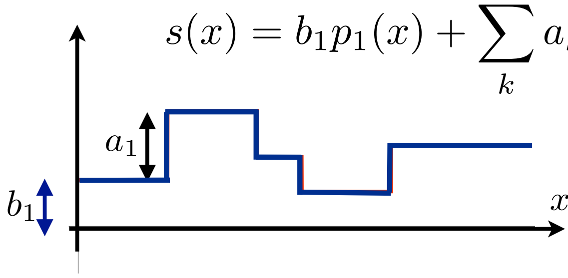
$\mathrm{L} = \mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}x}$ Null space: $\mathcal{N}_\mathrm{D} = \mathrm{span}\{p_1\}, \quad p_1(x) = 1$

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

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



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



Spline synthesis: generalization

L: spline admissible operator (LSI)

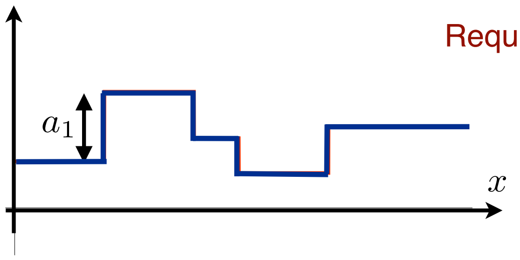
$\rho_{\mathrm{L}}(\boldsymbol{x}) = \mathrm{L}^{-1}\{\delta\}$: Green's function of L

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

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

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

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

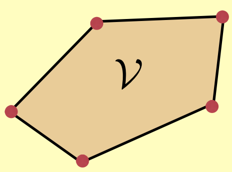
Theorem: The extremal points of the **generic linear-inverse** problem

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

are **non-uniform** L**-splines** of the form

$$f_{\mathrm{sparse}}(\boldsymbol{x}) = \sum_{k=1}^{K_{\mathrm{knots}}} a_k \rho_{\mathrm{L}}(\boldsymbol{x} - \boldsymbol{x}_k) + \sum_{n=1}^{N_0} b_n p_n(\boldsymbol{x})$$

with $K_{\mathrm{knots}} \le M - N_0$ and $\|\mathrm{L}\{f_{\mathrm{sparse}}\}\|_{\mathcal{M}} = \|\mathbf{a}\|_{\ell_1}$.



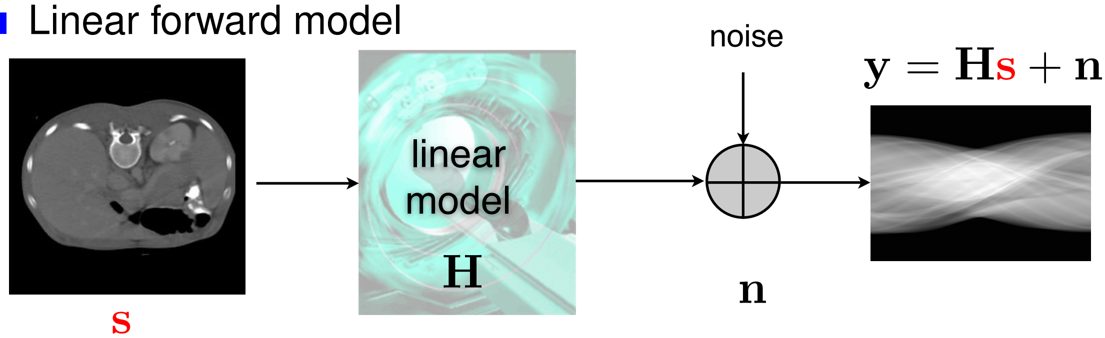
(U.-Fageot-Ward, ArXiv 2016; *SIAM Review* in press)

OUTLINE

- **From classical to modern image reconstruction** ✔
  - The sparsity (r)evolution⇒ Compressed Sensing
- **Sparsity revisited** ✔
  - Optimality of splines for gTV
- **Non-Gaussian statistical modeling**
  - Sparse stochastic processes
- **Iterative image reconstruction** (MAP formulation)
- **Learning**: Emergence of 3rd generation methods

Variational-MAP formulation of inverse problem

- Linear forward model



- Reconstruction as an optimization problem

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

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

An introduction to sparse stochastic processes

Random spline: archetype of sparse signal

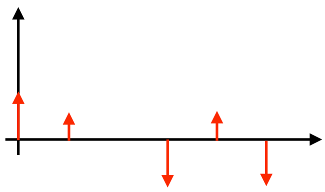
Random weights $\{a_n\}$ i.i.d. and random knots $\{t_n\}$ (Poisson with rate λ)

- Stochastic differential equation

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

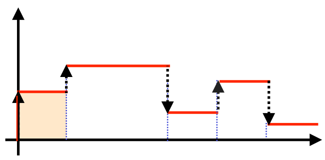
with boundary condition $s(0) = 0$

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



- Formal solution = Compound Poisson process

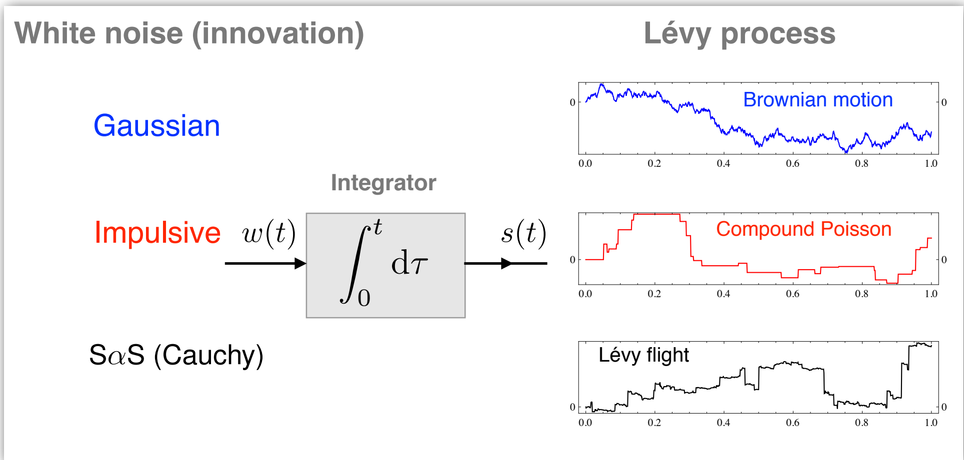
$$s(t) = \mathrm{D}^{-1}w(t) = \sum_n a_n \mathrm{D}^{-1}\{\delta(\cdot - t_n)\}(t)$$
$$= b_1 + \sum_n a_n \mathbb{1}_+(t - t_n)$$



Lévy processes: all admissible brands of innovations

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

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



(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) $\mathrm{L}s = w$

2

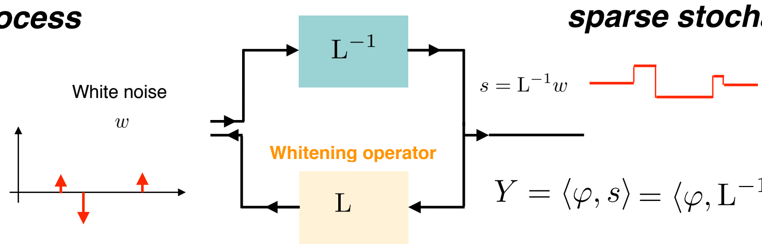
innovation process

sparse stochastic process

1

3

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



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

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)

Description of sparse stochastic process

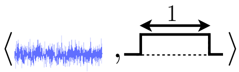
- Specification of spatial dependencies

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

- Specification of innovation (sparsity behavior)

  Canonical observation through a rectangular window

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

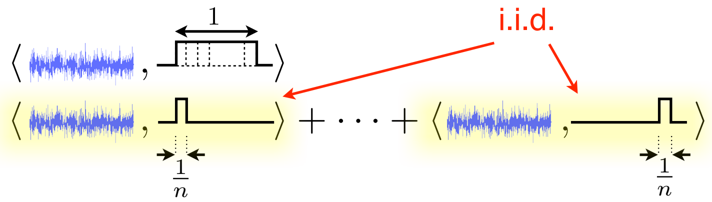


  w = white noise $\Rightarrow$ $X_{\mathrm{id}} = \langle w, \mathrm{rect} \rangle$ is **infinitely divisible**

  with **canonical Lévy exponent** $f(\omega) = \log \mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega X_{\mathrm{id}}}\}$.

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

$X = \langle w, \mathrm{rect} \rangle$



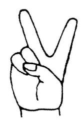
☒ Probability laws of sparse processes are id

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

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

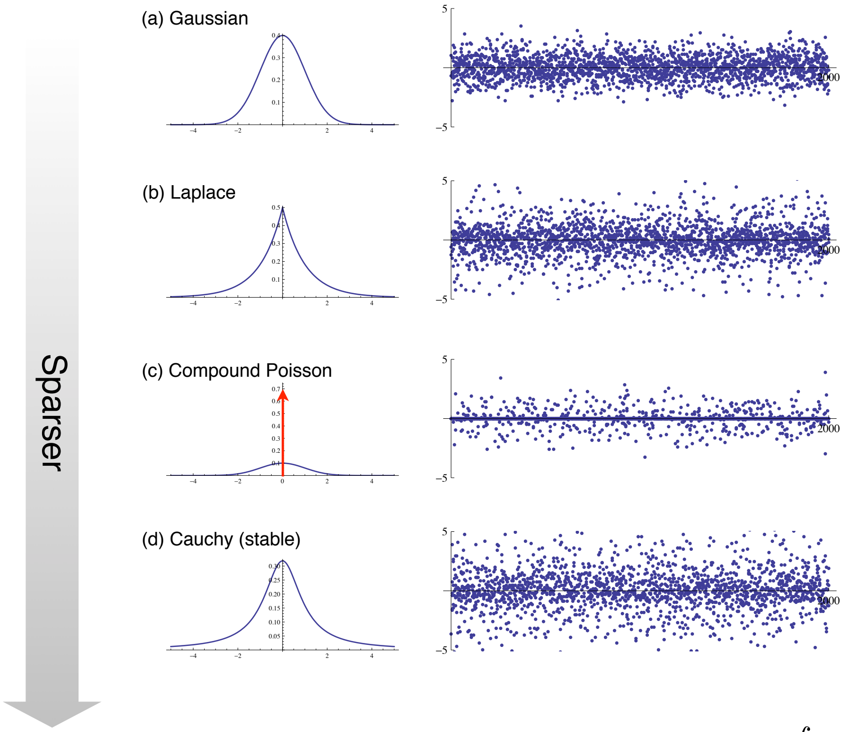


= explicit form of pdf

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

Examples of infinitely divisible laws

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



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

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

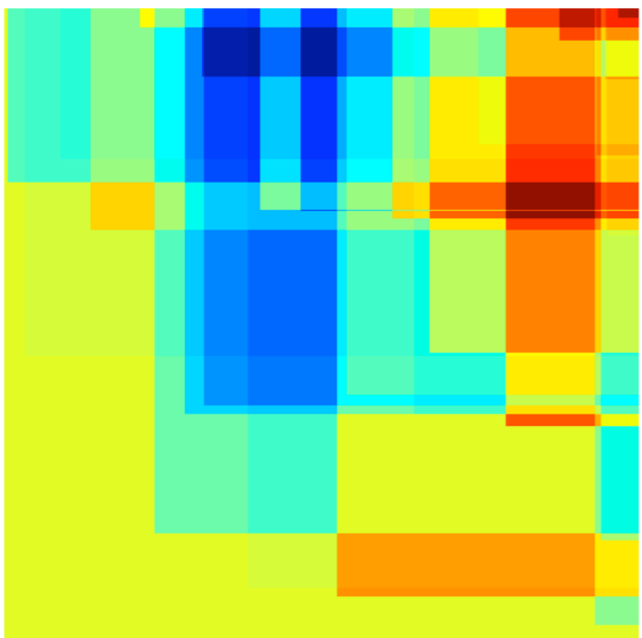
$$p_{\text{Poisson}}(x) = \mathcal{F}^{-1}\{\mathrm{e}^{\lambda(\hat{p}_A(\omega)-1)}\}$$

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

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

Aesthetic sparse signal: the Mondrian process

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



$\lambda = 30$

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

- **From classical to modern image reconstruction** ✔
  - The sparsity (r)evolution⇒ Compressed Sensing
- **Sparsity revisited** ✔
  - Optimality of splines for gTV
- **Non-Gaussian statistical modeling** ✔
  - Sparse stochastic processes
- **Iterative image reconstruction** (MAP formulation)
- **Learning**: Emergence of 3rd generation methods

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 $\Rightarrow$ $\mathbf{u} = \mathbf{L}\mathbf{s}$ (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

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

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

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

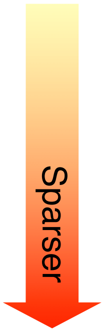
- Additive white Gaussian noise scenario (AWGN)

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

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

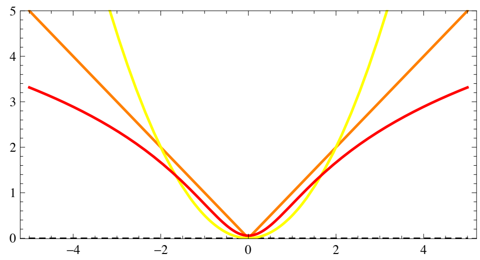
General form of MAP estimator

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



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

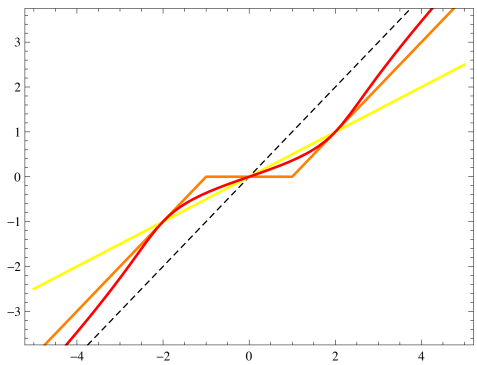
Potential: $\Phi_U(x) = -\log p_U(x)$



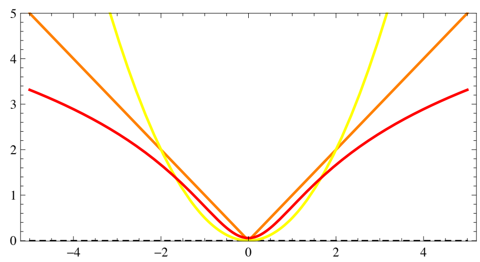
Proximal operator: pointwise denoiser

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

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



$\sigma^2\Phi_U(u)$



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

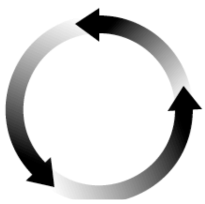
Lagrange multipler vector: $\boldsymbol{\alpha}$

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

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



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

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

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

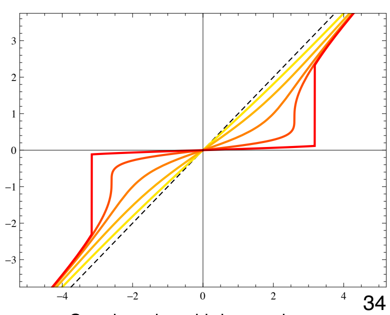
Linear inverse problem: $\mathbf{s}^{k+1} = \left(\mathbf{H}^T\mathbf{H} + \mu\mathbf{L}^T\mathbf{L}\right)^{-1}\left(\mathbf{H}^T\mathbf{y} + \mathbf{z}^{k+1}\right)$

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

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

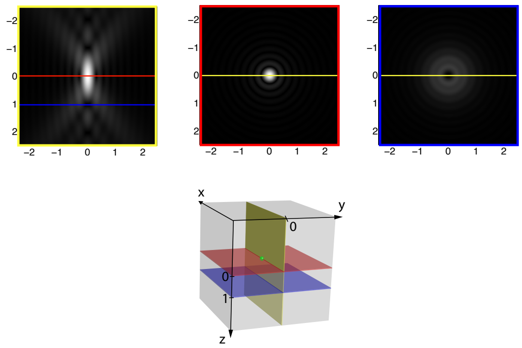


Cauchy prior with increasing s_0

Deconvolution of fluorescence micrographs

- Physical model of a diffraction-limited microscope

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



3-D point spread function (PSF)

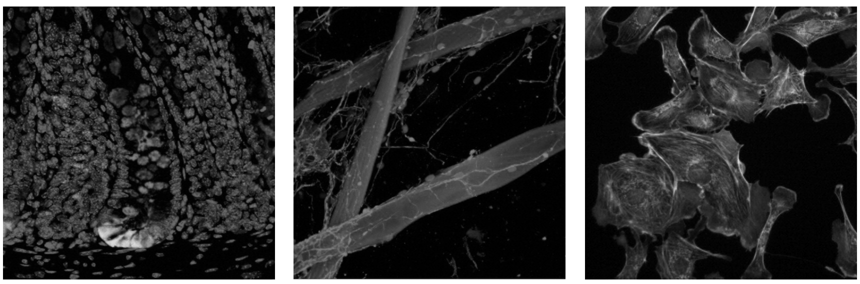
$$h_{3D}(x, y, z) = I_0 \left| p_\lambda \left(\frac{x}{M}, \frac{y}{M}, \frac{z}{M^2} \right) \right|^2$$

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

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)

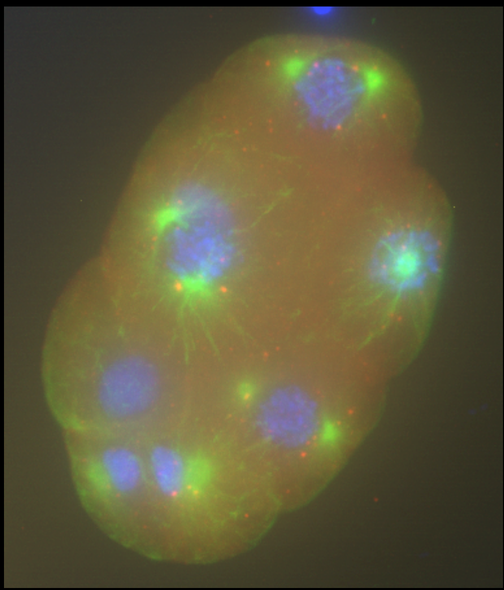
Figure 10.3 Images used in deconvolution experiments. (a) Stem cells surrounded by goblet cells. (b) Nerve cells growing around fibers. (c) Artery cells.

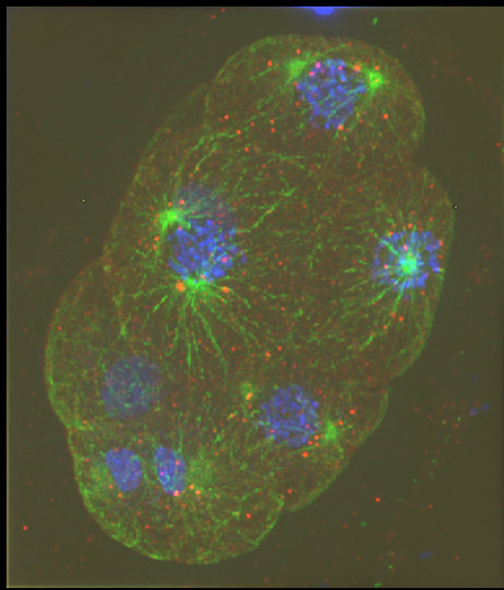
Table 10.2 Deconvolution performance of MAP estimators based on different prior distributions.

		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

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)

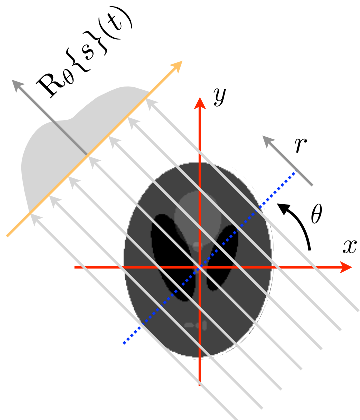
Computed tomography (straight rays)

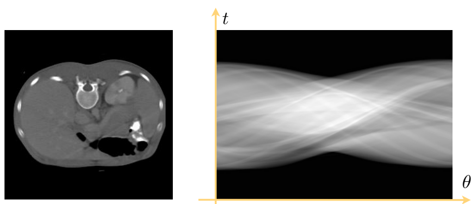
Projection geometry: $\boldsymbol{x} = t\boldsymbol{\theta} + r\boldsymbol{\theta}^{\perp}$ with $\boldsymbol{\theta} = (\cos\theta, \sin\theta)$

- Radon transform (line integrals)

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

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



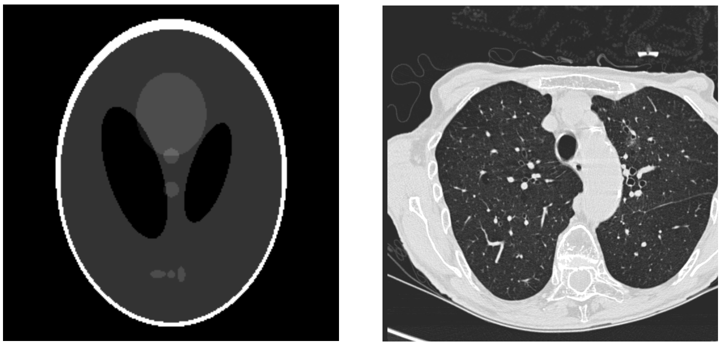


sinogram

(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.

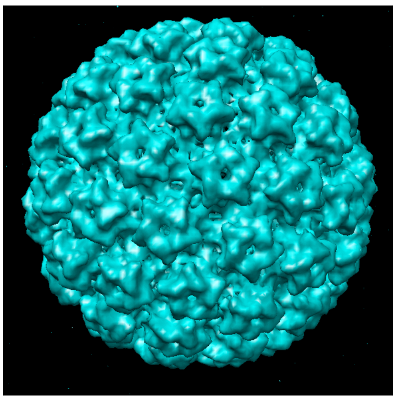
Table 10.4 Reconstruction results of X-ray computed tomography using different estimators.

	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

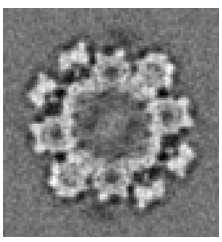
$\mathbf{L}$: discrete gradient

Cryo-electron tomography (real data)

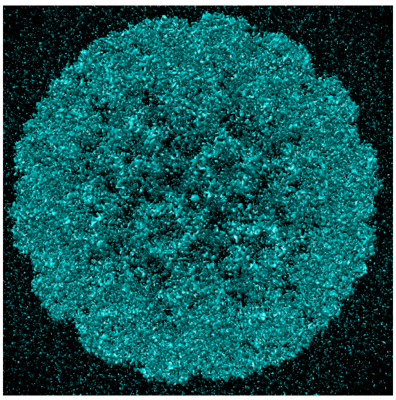
Standard Fourier-based reconstruction

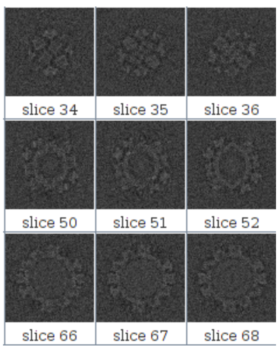


6.185 Å

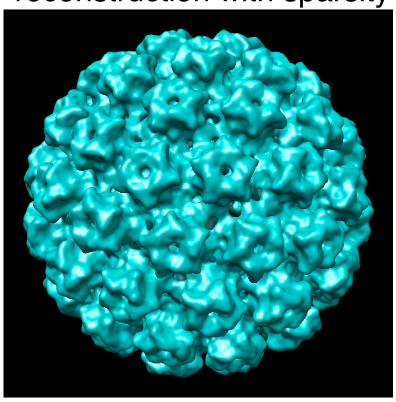


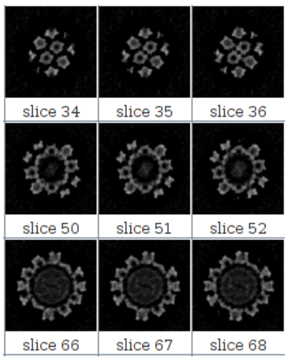
High-resolution Fourier-based reconstruction





High-resolution reconstruction with sparsity





Conceptual summary of 2nd generation methods

Physical model — Statistical model of signal

$$J(\mathbf{x},\mathbf{u}) = \underbrace{\frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2}_{\text{consistency}} + \underbrace{\lambda R(\mathbf{u})}_{\text{prior constraints}} + \underbrace{\mu\|\mathbf{L}\mathbf{x}-\mathbf{u}\|_2^2}_{\text{algorithmic coupling}}$$

Schematic structure of reconstruction algorithm:

Repeat (N_{iter})

- $\mathbf{x}^{(n)} = \arg\min_{\mathbf{x}} J(\mathbf{x},\mathbf{u}^{(n-1)})$: Linear step (problem specific)
- $\mathbf{u}^{(n)} = \arg\min_{\mathbf{u}} J(\mathbf{x}^{(n)},\mathbf{u})$: Statistical or “denoising” step

until stop criterion

Part 3:

The (deep) learning (r)evolution

⇒ Emergence of 3rd generation methods

Learning within the current paradigm

- Data-driven tuning of parameters: λ, calibration of forward model

  Semi-blind methods, sequential optimization

- Improved decoupling/representation of the signal

  Data-driven **dictionary learning** $\Rightarrow$ "optimal" $\mathbf{L}$
  (based of sparsity or statistics/ICA)

  (Elad 2006, Ravishankar 2011, Mairal 2012)

- Learning of non-linearities / Proximal operators $\Rightarrow$ "optimal" potential Φ

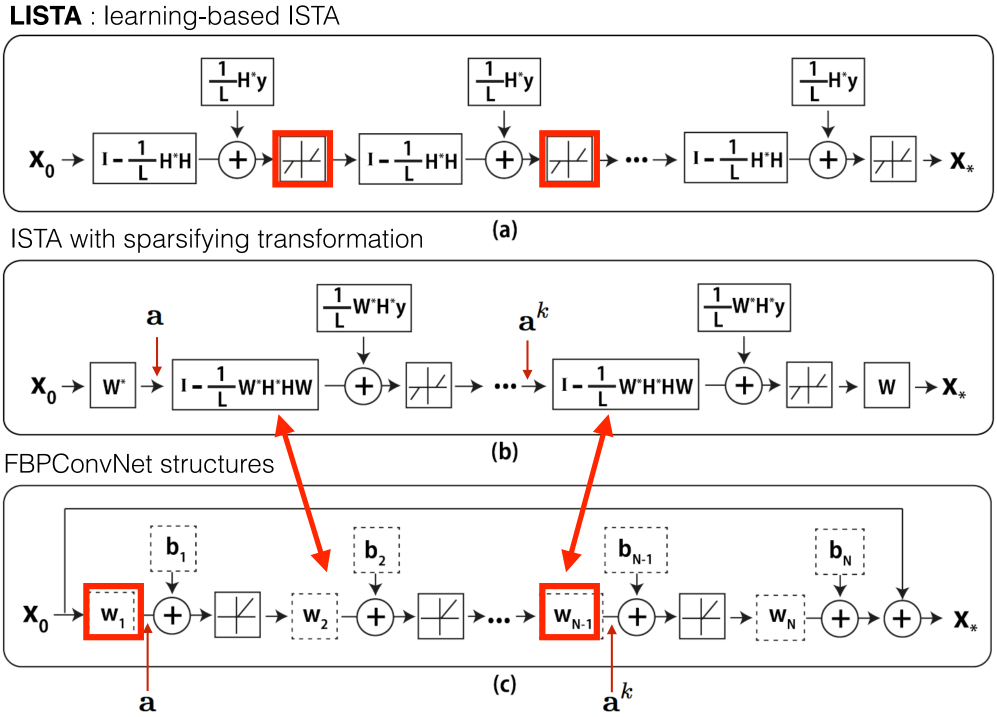
  CNN-type parametrization, backpropagation

  (Chen-Pock 2015-2016, Kamilov 2016)

Connection with deep neural networks

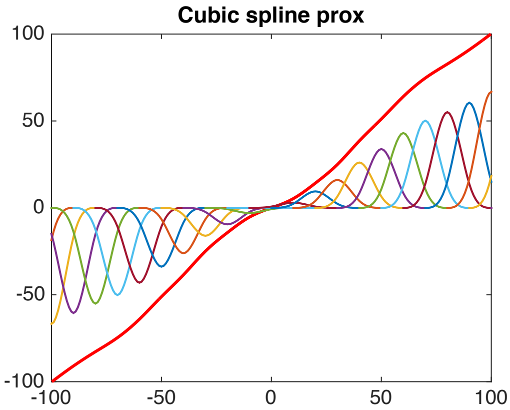
(Gregor-LeCun 2010)

Unrolled Iterative Shrinkage Thresholding Algorithm (ISTA)



SplineProx: learning shrinkage functions

$$\text{SplineProx}(x|\mathbf{a}) = \sum_{k\in\mathbb{Z}} a_k \beta^3(x/T - k)$$



Pointwise nonlinearity parametrized by B-spline coefficients $\mathbf{a}$

Mathematical constraints: ***firmly nonexpansive*** vs. ***monotonic***

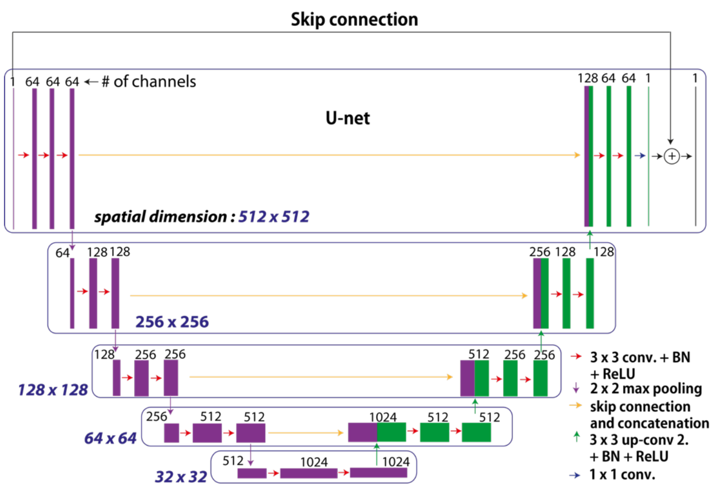
(Kamilov *IEEE SPL* 2016; Nguyen et al., arXiv:1705.05591 [cs.LG])

Recent appearance of Deep ConvNets

(Jin et al. 2016; Chen et al. 2017; ...)

- CT reconstruction based on Deep ConvNets
  - Input: Sparse view FBP reconstruction
  - Training: Set of 500 high-quality full-view CT reconstructions
  - Architecture: U-Net with skip connection

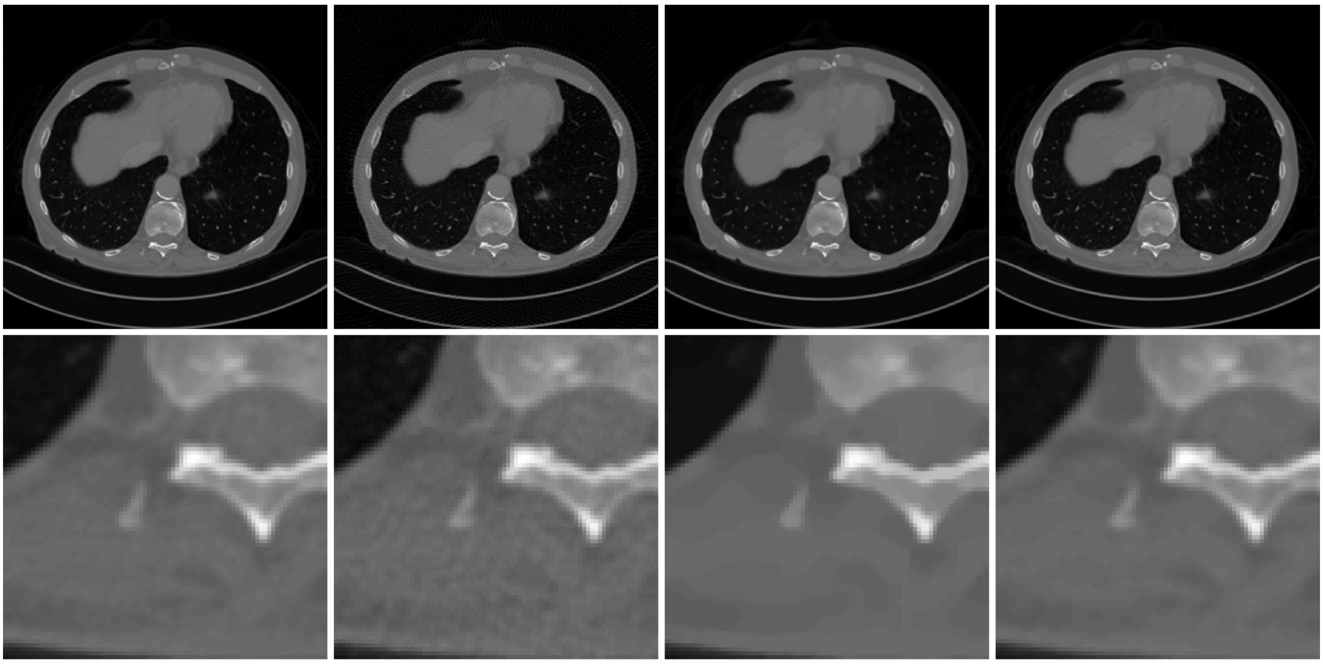
(Jin et al., arXiv:1611.03679)



CT data

Dose reduction by 7: 143 views

Ground truth	FBP SNR 24.06	TV SNR 29.64	FBPConvNet SNR 35.38



Reconstructed from
from 1000 views

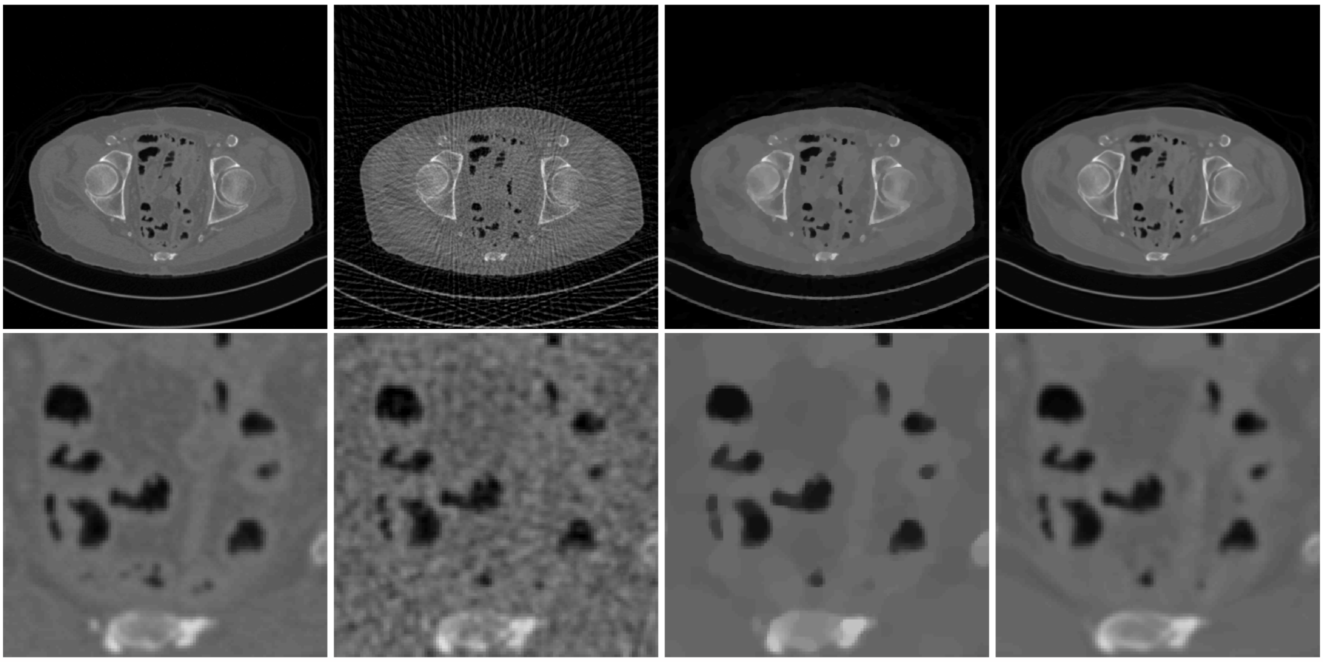
(Jin et al., *IEEE Trans. Im Proc.*, in press)

MAYO CLINIC

CT data

Dose reduction by 20: 50 views

Ground truth	FBP SNR 13.43	TV SNR 24.89	FBPConvNet SNR 28.53



Reconstructed from
from 1000 views

(Jin et al., *IEEE Trans. Im Proc.*, in press)

MAYO CLINIC

Primary reconstruction methods: **typology**

Methods	*Theories*	
1. **Classical (FBP)**	Tikhonov	Gaussian estimation
2. **Sparsity-driven**	Compressed sensing	**gTV-splines**
3. **Extended modeling**	**Sparse stochastic processes**	Bi-level optimization (joint statistical estimation)
4. **Deep learning**	Machine learning	Data science

degree of understanding

New learning era

Reconstruction methods: **tuning parameters**

1. **Classical (FBP)**	L or power spectrum $\longrightarrow$	λ
2. **Sparsity-driven**	L or sparsifying transform $\longrightarrow$	λ
3. **Extended modeling**	L or sparsifying transform + Lévy exponent $\longrightarrow$	Dictionary + Shrinkage functions
4. **Deep learning**	weights of neural network (**millions**)	

complexity

Comparison chart

	1. **Classical reconstruction**	2. **Sparsity driven**	3. **Extended modeling**	4. **Deep learning**
Speed	+++	++	+	--- / +++ training
Reconstruction quality	+	++	++(+)	+++
Full views	++	++(+)	+++	??? no way to train
Few views (CS)	+	++	++(+)	
Theoretical guarantees (optimality, worst case...)	+++	++(+)	+(+) ?	
Robustness	++	+++	++(+)	-- need to retrain for each configuration
Training data requirement	Modest	Modest	Average	**Enormous**

Conclusion

- Can we further reduce exposure time/improve image quality ?

  Preliminary results with **FBPConvNet** suggest that there is still room considerable improvement using learning: $\times 4$ reduction or more ?

- Key to success

  - Realistic forward model (**physics**) with autocalibration
  - Better **signal modeling**
  - Reducing the number of parameters to tune

- The future: **Computational imaging** with **educated learning**

  Find a "safe" compromize between **principled approaches** (robust, with guarantees of performance) and purely **data-driven approaches** (top performers) whose functioning is not yet understood.

Important requirement for bioimaging

- Looking "good" is not enough ⇒ task-oriented evaluation
- Difficulty of having "goldstandard" for training CNNs
- Need worst-case guarantees; understanding of bias/limitations

Selected references

- Theoretical foundations

- M. Unser and P. Tafti, ***An Introduction to Sparse Stochastic Processes***, Cambridge University Press, 2014; preprint, available at `http://www.sparseprocesses.org`.
- M. Unser, J. Fageot, H. Gupta, "Representer Theorems for Sparsity-Promoting ℓ_1 Regularization," *IEEE Trans. Information Theory*, vol. 62, no. 9, pp. 5167-5180, September 2016.
- M. Unser, J. Fageot, J.P. Ward, "Splines Are Universal Solutions of Linear Inverse Problems with Generalized-TV Regularization," *SIAM Review* (in press), arXiv:1603.01427 [math.FA].
- H.Q. Nguyen, E. Bostan, M. Unser, "Learning Convex Regularizers for Optimal Bayesian Denoising," arXiv:1705.05591 [cs.LG]

- 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 Transactions on Medical Imaging*, vol. 30, no. 9, pp. 1649-1660, September 2011.
- 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.
- K.H. Jin, M.T. McCann, E. Froustey, M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Trans. Image Processing*, in press.

Acknowledgments

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

- Prof. Arash Amini
- Dr. Emrah Bostan
- Dr. Julien Fageot
- Harshit Gupta
- Dr. Kyong Jin
- Prof. Ulugbek Kamilov
- Prof. Mathieu Querquin-Kern
- Dr. Michael McCann
- Dr. Ha Nguyen
- Dr. Masih Nilchian
- Dr. Cédric Vonesch
- Prof. John-Paul Ward

  ....

and collaborators ...

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

erc²

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