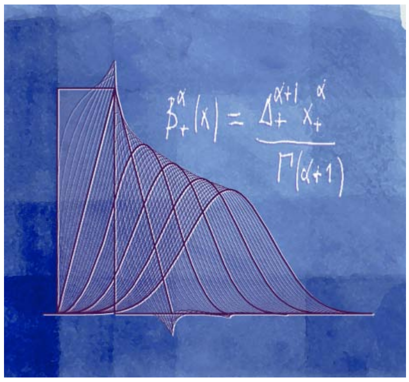
Biomedical image reconstruction

Michael Unser
Biomedical Imaging Group
EPFL, Lausanne, Switzerland



Tutorial Session, *European Molecular Imaging Meeting* (EMIM'17), 5-7 April 2017, Köln, Germany

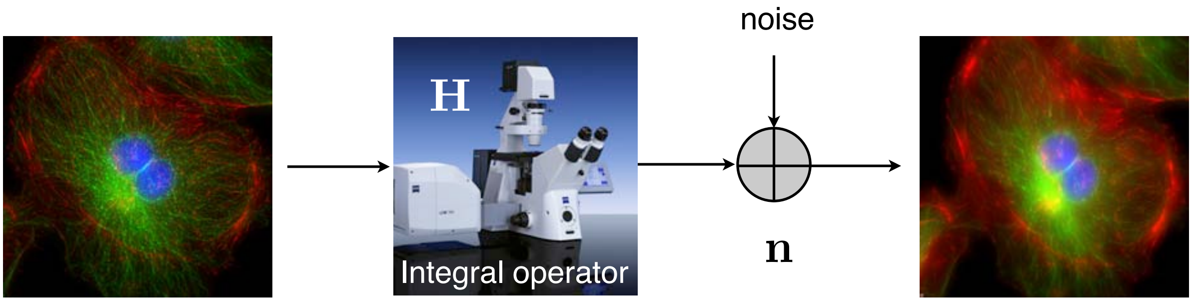
OUTLINE

- **1. Imaging as an inverse problem**
  - Basic imaging operators
  - Comparison of modalities
  - Discretization of the inverse problem
- **2. Classical reconstruction algorithms**
  - Backprojection
  - Tikhonov regularization
  - Wiener / LMSE solution
- **3. Modern methods: the sparsity (re)evolution**

  Specific examples:
  Magnetic resonance imaging
  Computed tomography
  Differential phase-contrast tomography
- **4. What's next: the learning revolution ?**

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 we

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

- 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**

Part 1:

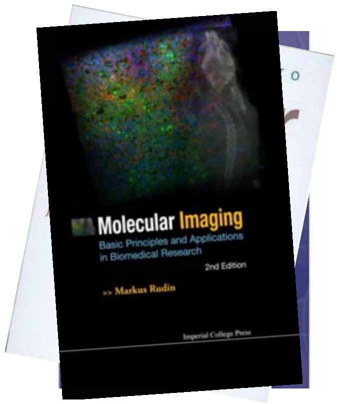
Setting up the problem

Forward imaging model (noise-free)

Unknown molecular/anatomical map: $s(\boldsymbol{r}), \boldsymbol{r} = (x, y, z, t) \in \mathbb{R}^d$

defined over a continuum in space-time

$s \in L_2(\mathbb{R}^d)$ (space of finite-energy functions)



Imaging operator $\mathrm{H} : s \mapsto \mathbf{y} = (y_1, \cdots, y_M) = \mathrm{H}\{s\}$

from continuum to discrete (finite dimensional)

$$\mathrm{H} : L_2(\mathbb{R}^d) \to \mathbb{R}^M$$

Linearity assumption: for all $s_1, s_2 \in L_2(\mathbb{R}^d)$, $\alpha_1, \alpha_2 \in \mathbb{R}$

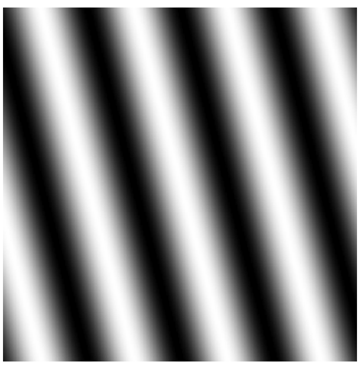
$$\mathrm{H}\{\alpha_1 s_1 + \alpha_2 s_2\} = \alpha_1 \mathrm{H}\{s_1\} + \alpha_2 \mathrm{H}\{s_2\}$$

impulse response of *m*th detector

$$\Rightarrow \quad [\mathbf{y}]_m = y_m = \langle \eta_m, s \rangle = \int_{\mathbb{R}^d} \eta_m(\boldsymbol{r}) s(\boldsymbol{r}) \mathrm{d}\boldsymbol{r}$$

(by the Riesz representation theorem)

Images are obviously made of sine waves ...



Basic operator: Fourier transform

$$\mathcal{F} : L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$$

$$\hat{f}(\boldsymbol{\omega}) = \mathcal{F}\{f\}(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} f(\boldsymbol{x})\mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}, \boldsymbol{x}\rangle} \mathrm{d}\boldsymbol{x}$$

Reconstruction formula (inverse Fourier transform)

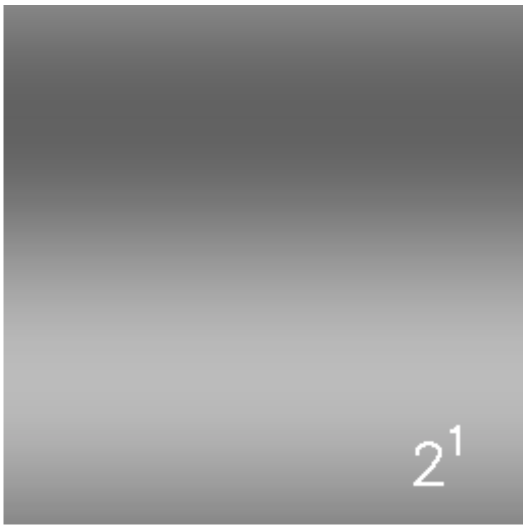
$$f(\boldsymbol{x}) = \mathcal{F}^{-1}\{f\}(\boldsymbol{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(\boldsymbol{\omega}) e^{j\langle \boldsymbol{\omega}, \boldsymbol{r}\rangle} \mathrm{d}\boldsymbol{\omega} \qquad \text{(a.e.)}$$

Equivalent analysis functions: $\eta_m(\boldsymbol{x}) = \mathrm{e}^{\mathrm{j}\langle \boldsymbol{\omega}_m, \boldsymbol{x}\rangle}$ (complex sinusoids)

2D Fourier reconstruction

Original image:

$f(\boldsymbol{x})$



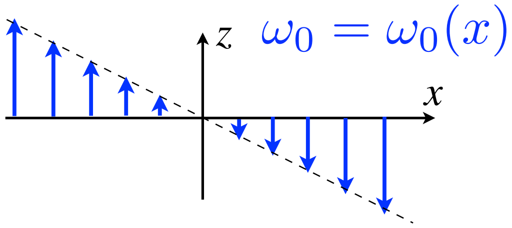
Reconstruction using N largest coefficients:

$$\tilde{f}(\boldsymbol{x}) = \frac{1}{(2\pi)^2} \sum_{\text{subset}} \hat{f}(\boldsymbol{\omega}) e^{j\langle \boldsymbol{x}, \boldsymbol{\omega}\rangle}$$

Magnetic resonance imaging

- Magnetic resonance: $\omega_0 = \gamma B_0$

  Frequency encode:



- Linear forward model for MRI

$\boldsymbol{r} = (x, y, z)$

$$\hat{s}(\boldsymbol{\omega}_m) = \int_{\mathbb{R}^3} s(\boldsymbol{r}) \mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}_m, \boldsymbol{r} \rangle} \mathrm{d}\boldsymbol{r}$$

(sampling of Fourier transform)

- Extended forward model with coil sensitivity

$$\hat{s}_w(\boldsymbol{\omega}_m) = \int_{\mathbb{R}^3} w(\boldsymbol{r}) s(\boldsymbol{r}) \mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}_m, \boldsymbol{r} \rangle} \mathrm{d}\boldsymbol{r}$$

Basic operator: Windowing

$$\mathrm{W} : L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$$

$$\mathrm{W}\{f\}(\boldsymbol{x}) = w(\boldsymbol{x}) f(\boldsymbol{x})$$

Positive window function (continuous and bounded): $w \in C_{\mathrm{b}}(\mathbb{R}^d), w(\boldsymbol{x}) \geq 0$

- Special case: modulation

$$w(\boldsymbol{r}) = \mathrm{e}^{\mathrm{j}\langle \boldsymbol{\omega}_0, \boldsymbol{r} \rangle}$$

$$\mathrm{e}^{\mathrm{j}\langle \boldsymbol{\omega}_0, \boldsymbol{r} \rangle} f(\boldsymbol{r}) \quad \xleftrightarrow{\mathcal{F}} \quad \hat{f}(\boldsymbol{\omega} - \boldsymbol{\omega}_0)$$

Application: Structured illumination microscopy (SIM)

Basic operator: Convolution

$$\mathrm{H}: L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$$

$$\mathrm{H}\{f\}(\boldsymbol{x}) = (h * f)(\boldsymbol{x}) = \int_{\mathbb{R}^d} h(\boldsymbol{x}-\boldsymbol{y}) f(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}$$

Impulse response: $h(\boldsymbol{x}) = \mathrm{H}\{\delta\}$

Equivalent analysis functions: $\eta_m(\boldsymbol{x}) = h(\boldsymbol{x}_m - \cdot)$

Frequency response: $\hat{h}(\boldsymbol{\omega}) = \mathcal{F}\{h\}(\boldsymbol{\omega})$

- Convolution as a frequency-domain product

$$(h * f)(\boldsymbol{x}) \quad \stackrel{\mathcal{F}}{\longleftrightarrow} \quad \hat{h}(\boldsymbol{\omega}) \hat{f}(\boldsymbol{\omega})$$

Modeling of optical systems

$f(x,y)$ $\longrightarrow$  $\longrightarrow$ $g(x,y) = (h * f)(x,y)$

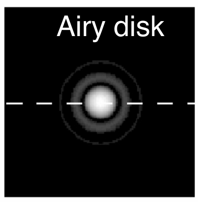
$h(x,y)$: Point Spread Function (PSF)

Diffraction-limited optics = LSI system

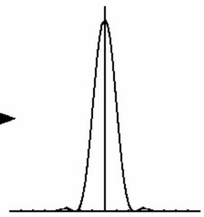
- Aberration-free point spread function (in focal plane)

$$h(x,y) = h(r) = C \cdot \left[\frac{2J_1(\pi r)}{\pi r}\right]^2$$

where $r = \sqrt{x^2+y^2}$ (radial distance)

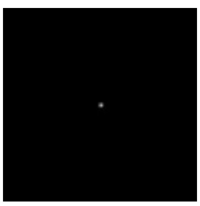


Radial profile

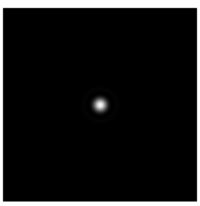


- Effect of misfocus

Point source



output



(in focus)

(defocus)

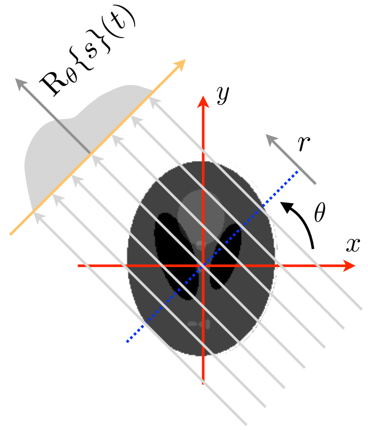
Basic operator: X-ray transform

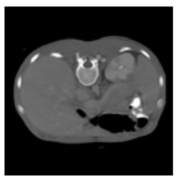
Projection geometry: $\boldsymbol{x} = t\boldsymbol{\theta} + r\boldsymbol{\theta}^{\perp}$ with $\boldsymbol{\theta} = (\cos\theta, \sin\theta)$

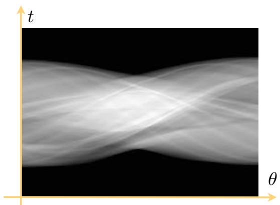
- Radon transform (line integrals)

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

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







sinogram

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

Central slice theorem

- Measurements of line integrals (Radon transform)

$$p_\theta(t) = \mathrm{R}_\theta\,\{f\}\,(t, \theta)$$

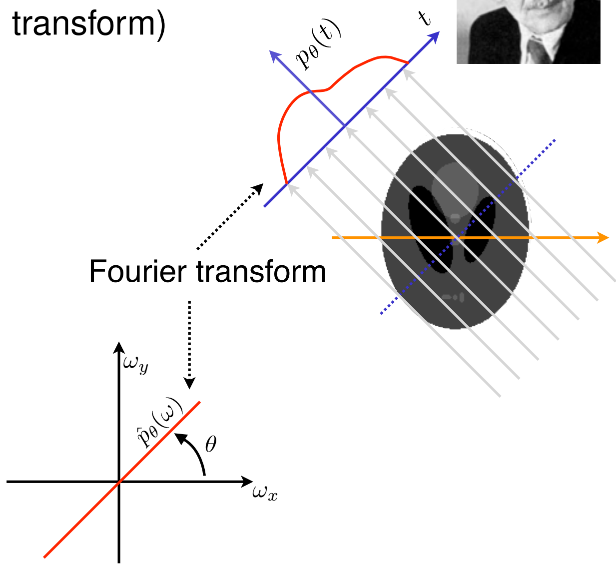
- 1D and 2D Fourier transforms

$$\hat{p}_\theta(\omega) = \mathcal{F}_{\mathrm{1D}}\{p_\theta\}(\omega)$$

$$\hat{f}(\boldsymbol{\omega}) = \mathcal{F}_{\mathrm{2D}}\{f\}(\boldsymbol{\omega}) = \hat{f}_{\mathrm{pol}}(\omega, \theta)$$

- Central-slice theorem

$$\hat{p}_\theta(\omega) = \hat{f}(\omega\cos\theta,\, \omega\sin\theta) = \hat{f}_{\mathrm{pol}}(\omega, \theta)$$



Proof: for $\theta = 0$

$$\hat{f}(\omega, 0) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)e^{-j\omega x}\,\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{+\infty} \underbrace{\left(\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y\right)}_{p_0(x)} e^{-j\omega x}\,\mathrm{d}x = \hat{p}_0(\omega)$$

then use rotation property of Fourier transform...

Modality	Radiation	Forward model	Variations
2D or 3D tomography	coherent x-ray	$y_i = \mathrm{R}_{\boldsymbol{\theta}_i} x$	parallel, cone beam, spiral sampling
3D deconvolution microscopy	fluorescence	$y = \mathrm{H}x$	brightfield, confocal, light sheet
structured illumination microscopy (SIM)	fluorescence	$y_i = \mathrm{HW}_i x$ H: PSF of microscope W_i: illumination pattern	full 3D reconstruction, non-sinusoidal patterns
Positron Emission Tomography (PET)	gamma rays	$y_i = \mathrm{H}_{\boldsymbol{\theta}_i} x$	list mode with time-of-flight
Magnetic resonance imaging (MRI)	radio frequency	$y = \mathrm{F}x$	uniform or non-uniform sampling in k space
Cardiac MRI (parallel, non-uniform)	radio frequency	$y_{t,i} = \mathrm{F}_t \mathrm{W}_i x$ W_i: coil sensitivity	gated or not, retrospective registration
Optical diffraction tomography	coherent light	$y_i = \mathrm{W}_i \mathrm{F}_i x$	with holography or grating interferometry

Discretization: Finite dimensional formalism

$$s(\boldsymbol{r}) = \sum_{\boldsymbol{k}\in\Omega} s[\boldsymbol{k}]\beta_{\boldsymbol{k}}(\boldsymbol{r})$$

Signal vector: $\mathbf{s} = \big(s[\boldsymbol{k}]\big)_{\boldsymbol{k}\in\Omega}$ of dimension K

- Measurement model (image formation)

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

η_m: sampling/imaging function (mth detector)

$n[\cdot]$: additive noise

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

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

Example of basis functions

Shift-invariant representation: $\beta_{\boldsymbol{k}}(\boldsymbol{x}) = \beta(\boldsymbol{x} - \boldsymbol{k})$

Separable generator: $\beta(\boldsymbol{x}) = \prod_{n=1}^{d} \beta(x_n)$

- Pixelated model

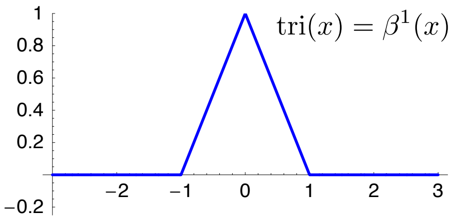
$$\beta(x) = \mathrm{rect}(x)$$

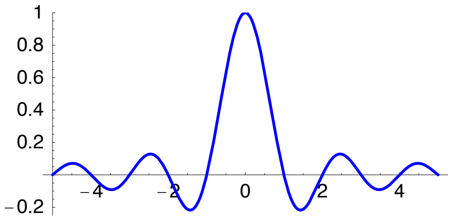
- Bilinear model

$$\beta(x) = (\mathrm{rect} * \mathrm{rect})(x) = \mathrm{tri}(x)$$

- Bandlimited representation

$$\beta(x) = \mathrm{sinc}(x)$$





Part 2:

Classical image

reconstruction

Discretized forward model: **y**=**Hs**+ **n**

Inverse problem: How to efficiently recover **s** from **y** ?

Vector calculus

- Scalar cost function $J(\mathbf{v}) : \mathbb{R}^N \to \mathbb{R}$

- Vector differentiation: $\dfrac{\partial J(\mathbf{v})}{\partial \mathbf{v}} = \begin{bmatrix} \partial J/\partial v_1 \\ \vdots \\ \partial J/\partial v_N \end{bmatrix} = \nabla J(\mathbf{v})$ (gradient)

- Necessary condition for an unconstrained optimum (minimum or maximum)

$$\frac{\partial J(\mathbf{v})}{\partial \mathbf{v}} = 0$$ (also sufficient if $J(\mathbf{v})$ is convex in $\mathbf{v}$)

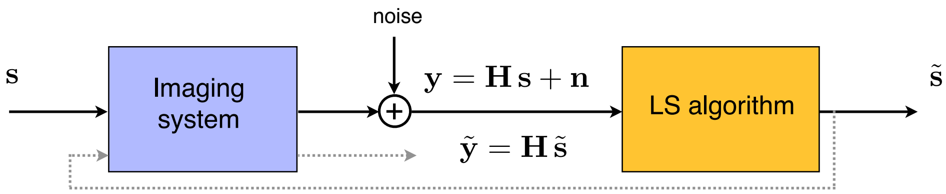
- Useful identities

$$\frac{\partial}{\partial \mathbf{v}}\left(\mathbf{a}^T\mathbf{v}\right) = \frac{\partial}{\partial \mathbf{v}}\left(\mathbf{v}^T\mathbf{a}\right) = \mathbf{a}$$

$$\frac{\partial}{\partial \mathbf{v}}\left(\mathbf{v}^T\mathbf{A}\mathbf{v}\right) = \left(\mathbf{A} + \mathbf{A}^T\right) \cdot \mathbf{v}$$

$$\frac{\partial}{\partial \mathbf{v}}\left(\mathbf{v}^T\mathbf{A}\mathbf{v}\right) = 2\mathbf{A} \cdot \mathbf{v}$$ if $\mathbf{A}$ is symmetric

Basic reconstruction: least-squares solution



- Least-squares fitting criterion: $J_{\mathrm{LS}}(\tilde{\mathbf{s}}, \mathbf{y}) = \|\mathbf{y} - \mathbf{H}\tilde{\mathbf{s}}\|^2$

$\min_{\tilde{\mathbf{s}}} \|\mathbf{y} - \tilde{\mathbf{y}}\|^2 = \min_{\mathbf{s}} J_{\mathrm{LS}}(\mathbf{s}, \mathbf{y})$ (maximum consistency with the data)

- Formal least-squares solution

$J_{\mathrm{LS}}(\mathbf{s}, \mathbf{y}) = \|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2 = \|\mathbf{y}\|^2 + \mathbf{s}^T\,\mathbf{H}^T\mathbf{H}\,\mathbf{s} - 2\,\mathbf{y}^T\mathbf{H}\,\mathbf{s}$

$\frac{\partial J_{\mathrm{LS}}(\mathbf{s},\mathbf{y})}{\partial \mathbf{s}} = 2\mathbf{H}^T\mathbf{H}\mathbf{s} - 2\mathbf{H}^T\mathbf{y}$

- Backprojection (poor m

OK if $\mathbf{H}$ is unitary $\Leftrightarrow$

Basic limitations

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

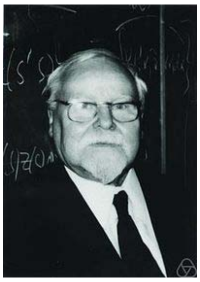
Linear inverse problems (20th century theory)

- 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)

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



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

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

$\Updownarrow$ $\quad \mathbf{L} = \mathbf{C}_s^{-1/2}$: 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 : $\mathbf{s} = (\mathbf{H}^T\mathbf{H} + \lambda\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T\mathbf{y} = \mathbf{R}_\lambda \cdot \mathbf{y}$

Iterative reconstruction algorithm

- Generic minimization problem: $\mathbf{s}_{\mathrm{opt}} = \arg\min_{\mathbf{s}} J(\mathbf{s}, \mathbf{y})$
- Steepest-descent solution

$$\mathbf{s}^{(k+1)} = \mathbf{s}^{(k)} - \gamma \nabla J\big(\mathbf{s}^{(k)}, \mathbf{y}\big)$$

- Iterative constrained least-squares reconstruction

$$J_{\mathrm{Tik}}(\mathbf{s}, \mathbf{y}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2 + \frac{\lambda}{2}\|\mathbf{L}\mathbf{s}\|^2$$

Gradient: $\dfrac{\partial J_{\mathrm{Tik}}(\mathbf{s}, \mathbf{y})}{\partial \mathbf{s}} = -\mathbf{s}_0 + (\mathbf{H}^T\mathbf{H} + \lambda \mathbf{L}^T\mathbf{L})\mathbf{s}$ with $\mathbf{s}_0 = \mathbf{H}^T\mathbf{y}$

Steepest-descent algorithm

$$\mathbf{s}^{(k+1)} = \mathbf{s}^{(k)} + \gamma\big(\mathbf{s}_0 - (\mathbf{H}^T\mathbf{H} + \lambda \mathbf{L}^T\mathbf{L})\tilde{\mathbf{s}}^{(k)}\big)$$

Positivity constraint (IC): $[\tilde{\mathbf{s}}^{(k+1)}]_i = \begin{cases} 0, & [\mathbf{s}^{(k+1)}]_i < 0 \\ [\mathbf{s}^{(k+1)}]_i, & \text{otherwise.} \end{cases}$ (projection on convex set)

Iterative deconvolution: unregularized case


Degraded image:
Gaussian blur + additive noise

van Cittert animation


Ground truth

Effect of regularization parameter

Degraded image:
Gaussian blur + additive noise

not enough: λ=0.02

not enough: λ=0.2

Optimal regularization: λ=2

too much: λ=20

too much: λ=200

Selecting the regularization operator

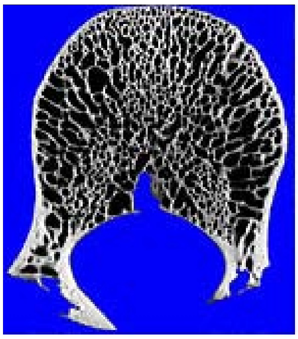
- Translation, rotation and scale-invariant operators
  - Laplacian: $\Delta s = (\boldsymbol{\nabla}^T \boldsymbol{\nabla}) s \quad \longleftrightarrow \quad -\|\boldsymbol{\omega}\|^2 \hat{s}(\boldsymbol{\omega})$
  - Modulus of gradient: $|\boldsymbol{\nabla} s|$
  - Fractional Laplacian: $(-\Delta)^{\frac{\gamma}{2}} \quad \longleftrightarrow \quad \|\boldsymbol{\omega}\|^{\gamma} \hat{s}(\boldsymbol{\omega})$
- TRS-invariant regularization functional

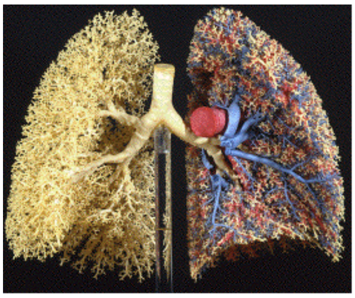
$$\|\boldsymbol{\nabla} s\|^2_{L_2(\mathbb{R}^d)} = \|(-\Delta)^{\frac{1}{2}} s\|^2_{L_2(\mathbb{R}^d)} \qquad \Rightarrow \qquad \mathbf{L}\text{: discrete version of gradient}$$

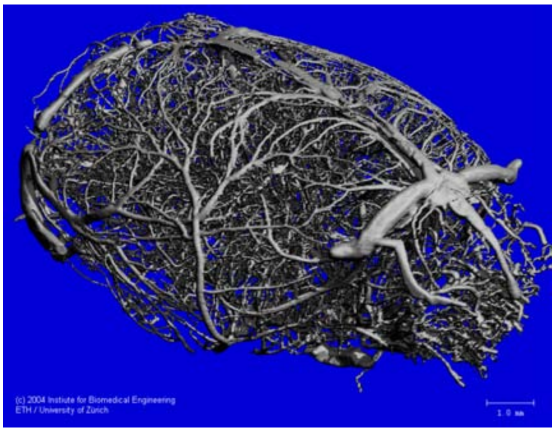
- Fractional Brownian motion field
  - Statistical decoupling/whitening: $(-\Delta)^{\frac{\gamma}{2}} s = w \quad \longleftrightarrow \quad \frac{1}{|\boldsymbol{\omega}|^{\gamma}}$ spectral decay

Relevance of self-similarity for bio-imaging

- Fractals and physiology







Designing fast reconstruction algorithms

Normal matrix: $\mathbf{A} = \mathbf{H}^T\mathbf{H}$ (symmetric)

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

Generic form of the iterator: $\mathbf{s}^{(k+1)} = \mathbf{s}^{(k)} + \gamma\big(\mathbf{s}_0 - (\mathbf{A} + \lambda\mathbf{L}^T\mathbf{L})\mathbf{s}^{(k)}\big)$

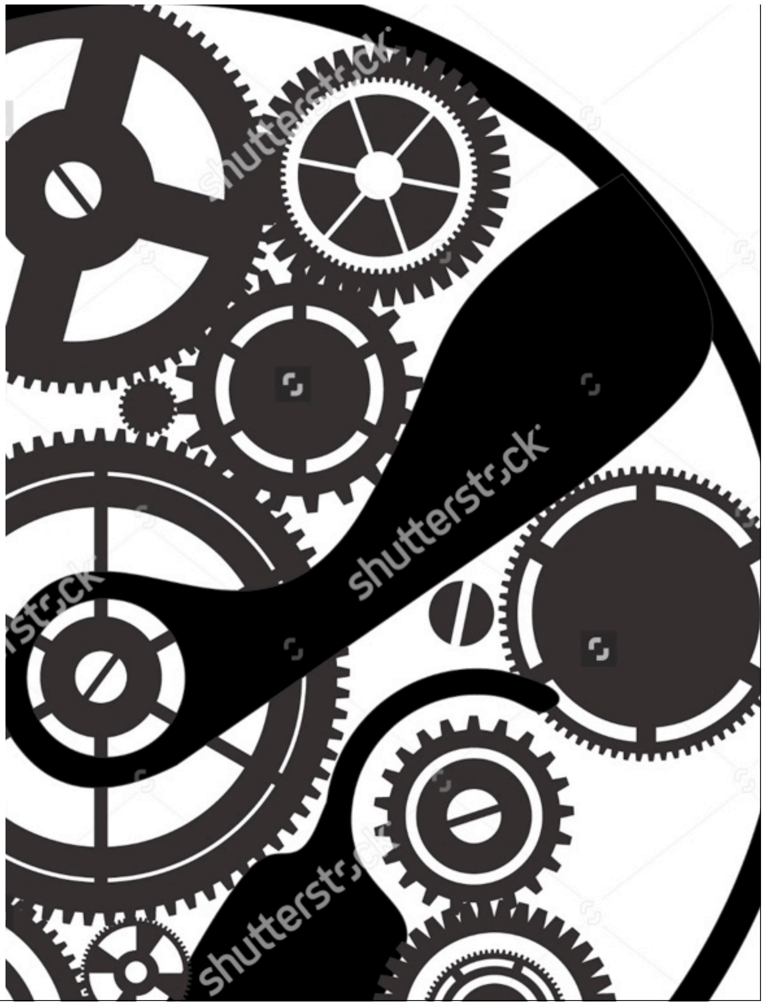
- Recognizing structured matrices
  - $\mathbf{L}$: convolution matrix $\Rightarrow$ $\mathbf{L}^T\mathbf{L}$: symmetric convolution matrix
  - $\mathbf{L}$, $\mathbf{A}$: convolution matrices $\Rightarrow$ $(\mathbf{A} + \lambda\mathbf{L}^T\mathbf{L})$: symmetric convolution matrix

- Fast implementation
  - Diagonalization of convolution matrices $\Rightarrow$ FFT-based implementation
  - Applicable to:
    - deconvolution microscopy (Wiener filter)
    - parallel rays computer tomography (FBP)
    - MRI, including non-uniform sampling of *k*-space

Part 3:

Modern image

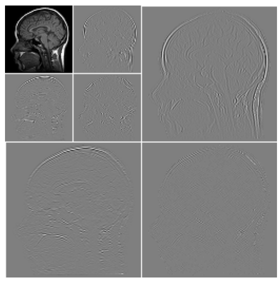
reconstruction



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)

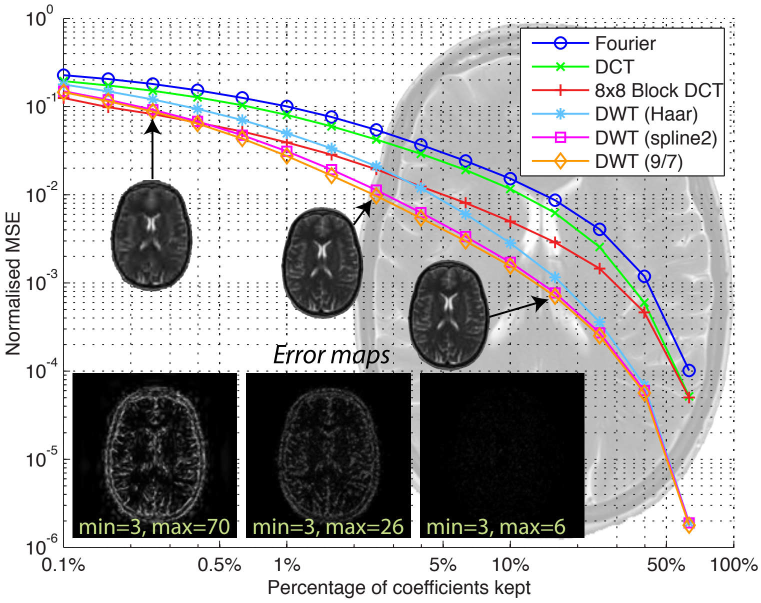
    $\mathbf{v} = \mathbf{W}^{-1}\mathbf{s}$: wavelet expansion of s (typically, sparse)

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

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

Sparsifying transforms

Biomedical images are well described by few basis coefficients



Prior = sparse representation

$$\mathcal{R}(\mathbf{s}) = \lambda \left\| \mathbf{W}^T \mathbf{s} \right\|_1$$

Advantages:
- convex
- favors sparse solutions
- Fast: WFISTA

(Guerquin-Kern *IEEE TMI* 2011)

Theory of compressive sensing

- Generalized sampling setting (after discretization)
  - Linear inverse problem: $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}$
  - Sparse representation of signal: $\mathbf{s} = \mathbf{W}\mathbf{x}$ with $\|\mathbf{x}\|_0 = K \ll N_x$
  - $N_y \times N_x$ system matrix : $\mathbf{A} = \mathbf{H}\mathbf{W}$
- Formulation of ill-posed recovery problem when $2K < N_y \ll N_x$

  (P0) $\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ subject to $\|\mathbf{x}\|_0 \leq K$
- Theoretical result

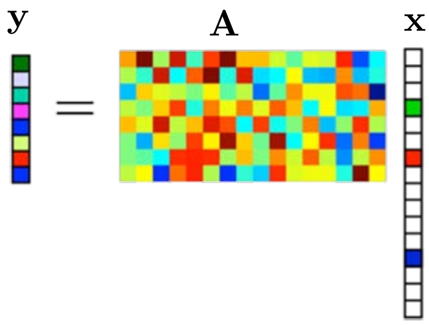
  Under suitable conditions on $\mathbf{A}$ (e.g., restricted isometry), the solution is unique and the recovery problem (P0) is equivalent to:

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

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

Compressive sensing (CS) and l_1 minimization

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



+ "noise"

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** : $\mathbf{A} = \mathbf{H}\mathbf{W}$

- Constrained (synthesis) formulation of recovery problem

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

Classical regularized least-squares estimator

- Linear measurement model:

  $y_m = \langle \mathbf{h}_m, \mathbf{x} \rangle + n[m], \quad m = 1, \dots, M$

- System matrix : $\mathbf{H} = [\mathbf{h}_1 \cdots \mathbf{h}_M]^T \in \mathbb{R}^{N \times N}$

$$\mathbf{x}_{\text{LS}} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_2^2$$

$$\Rightarrow \quad \mathbf{x}_{\text{LS}} = (\mathbf{H}^T\mathbf{H} + \lambda \mathbf{I}_N)^{-1}\mathbf{H}^T\mathbf{y}$$

$$= \mathbf{H}^T\mathbf{a} = \sum_{m=1}^{M} a_m \mathbf{h}_m \quad \text{where} \quad \mathbf{a} = (\mathbf{H}\mathbf{H}^T + \lambda \mathbf{I}_M)^{-1}\mathbf{y}$$

Interpretation: $\mathbf{x}_{\text{LS}} \in \text{span}\{\mathbf{h}_m\}_{m=1}^{M}$

Lemma

$(\mathbf{H}^T\mathbf{H} + \lambda \mathbf{I}_N)^{-1}\mathbf{H}^T = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T + \lambda \mathbf{I}_M)^{-1}$

Generalization: constrained l_2 minimization

- Discrete signal to reconstruct: $x = (x[n])_{n \in \mathbb{Z}}$
- Sensing operator $\mathrm{H} : \ell_2(\mathbb{Z}) \to \mathbb{R}^M$
  $x \mapsto \mathbf{z} = \mathrm{H}\{x\} = (\langle x, h_1 \rangle, \dots, \langle x, h_M \rangle)$ with $h_m \in \ell_2(\mathbb{Z})$
- Closed convex set in measurement space: $\mathcal{C} \subset \mathbb{R}^M$

Example: $\mathcal{C}_{\mathbf{y}} = \{\mathbf{z} \in \mathbb{R}^M : \|\mathbf{y} - \mathbf{z}\|_2^2 \leq \sigma^2\}$

Representer theorem for constrained ℓ_2 minimization

(P2) $\quad \min_{x \in \ell_2(\mathbb{Z})} \|x\|_{\ell_2}^2$ s.t. $\mathrm{H}\{x\} \in \mathcal{C}$

The problem (P2) has a unique solution of the form

$$x_{\mathrm{LS}} = \sum_{m=1}^{M} a_m h_m = \mathrm{H}^*\{\mathbf{a}\}$$

with expansion coefficients $\mathbf{a} = (a_1, \cdots, a_M) \in \mathbb{R}^M$.

(U.-Fageot-Gupta *IEEE Trans. Info. Theory*, Sept. 2016)

Constrained l_1 minimization $\Rightarrow$ sparsifying effect

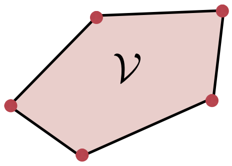
- Discrete signal to reconstruct: $x = (x[n])_{n \in \mathbb{Z}}$
- Sensing operator $\mathrm{H} : \ell_1(\mathbb{Z}) \to \mathbb{R}^M$
  $x \mapsto \mathbf{z} = \mathrm{H}\{x\} = (\langle x, h_1 \rangle, \dots, \langle x, h_M \rangle)$ with $h_m \in \ell_\infty(\mathbb{Z})$
- Closed convex set in measurement space: $\mathcal{C} \subset \mathbb{R}^M$

Representer theorem for constrained ℓ_1 minimization

(P1) $\quad \mathcal{V} = \arg \min_{x \in \ell_1(\mathbb{Z})} \|x\|_{\ell_1}$ s.t. $\mathrm{H}\{x\} \in \mathcal{C}$

is convex, weak*-compact with extreme points of the form

$$x_{\mathrm{sparse}}[\cdot] = \sum_{k=1}^{K} a_k \delta[\cdot - n_k] \quad \text{with} \quad K = \|x_{\mathrm{sparse}}\|_0 \leq M.$$



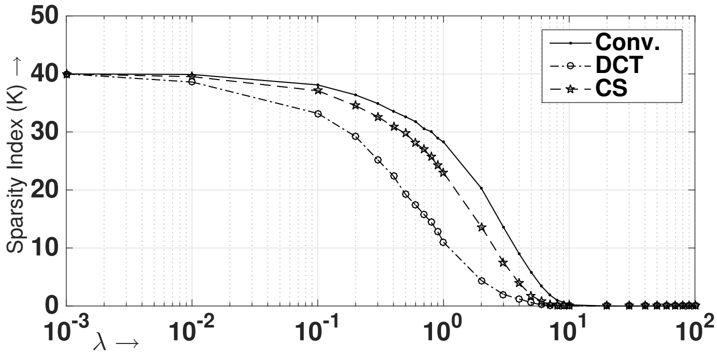
If CS condition is satisfied,
then solution is unique

(U.-Fageot-Gupta *IEEE Trans. Info. Theory*, Sept. 2016)

Controlling sparsity

Measurement model: $y_m = \langle h_m, x \rangle + n[m], \quad m = 1, \dots, M$

$$x_{\text{sparse}} = \arg \min_{x \in \ell_1(\mathbb{Z})} \left(\sum_{m=1}^{M} \left| y_m - \langle h_m, x \rangle \right|^2 + \lambda \|x\|_{\ell_1} \right)$$

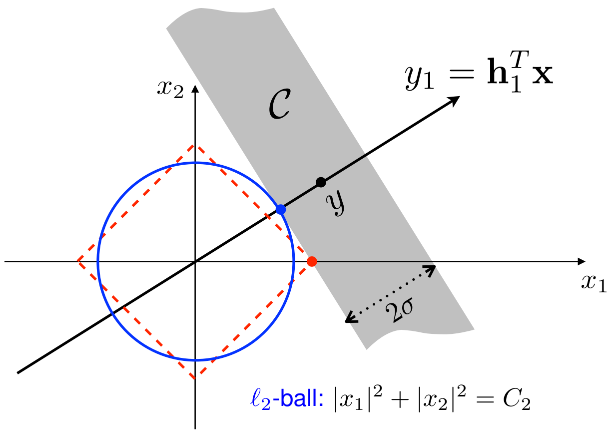


Geometry of l_2 vs. l_1 minimization

- Prototypical inverse problem

$$\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 + \lambda \|\mathbf{x}\|_{\ell_2}^2 \right\} \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{x}\|_{\ell_2} \text{ subject to } \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 \le \sigma^2$$

$$\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 + \lambda \|\mathbf{x}\|_{\ell_1} \right\} \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{x}\|_{\ell_1} \text{ subject to } \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 \le \sigma^2$$



ℓ_2-ball: $|x_1|^2 + |x_2|^2 = C_2$

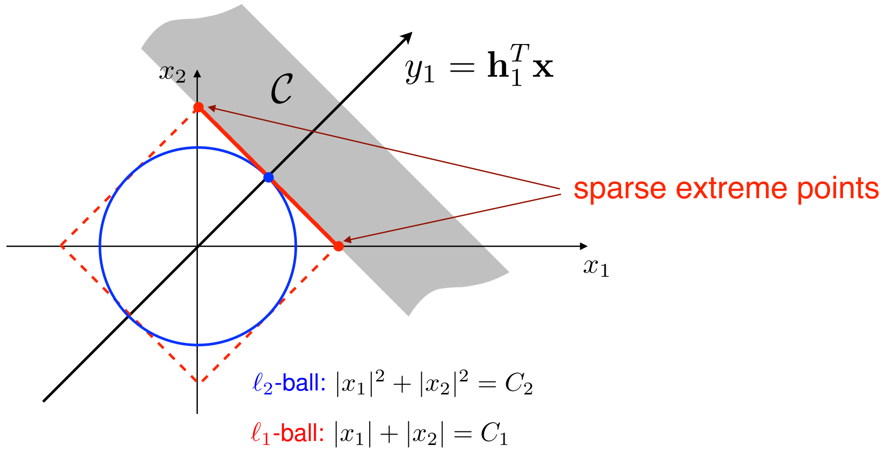
ℓ_1-ball: $|x_1| + |x_2| = C_1$

Geometry of l_2 vs. l_1 minimization

- Prototypical inverse problem

$$\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 + \lambda \|\mathbf{x}\|_{\ell_2}^2 \right\} \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{x}\|_{\ell_2} \text{ subject to } \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 \le \sigma^2$$

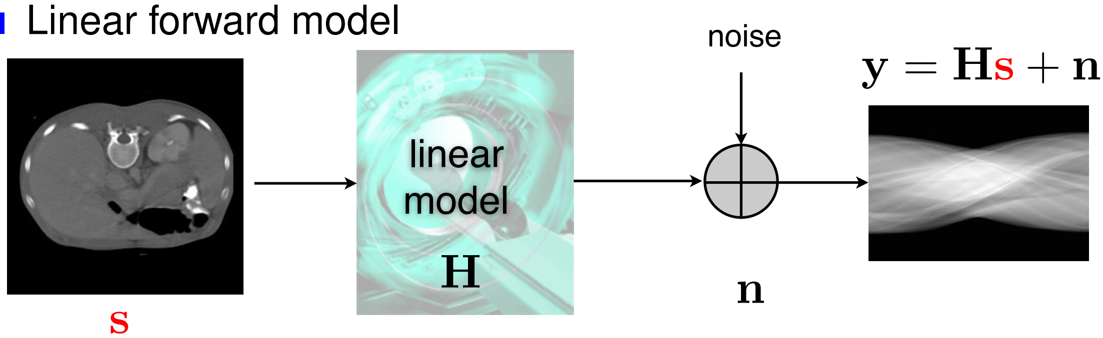
$$\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 + \lambda \|\mathbf{x}\|_{\ell_1} \right\} \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{x}\|_{\ell_1} \text{ subject to } \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 \le \sigma^2$$



Configuration for **non-unique** ℓ_1 solution

Variational-MAP formulation of inverse problem

- Linear forward model



- Reconstruction as an optimization problem

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

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

Discretization of reconstruction problem

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

- Statistical 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

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

- Physical model: image formation and acquisition

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

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

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

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)

Statistical decoupling

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

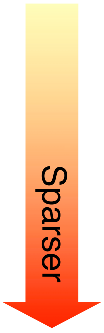
- Additive white Gaussian noise scenario (AWGN)

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

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

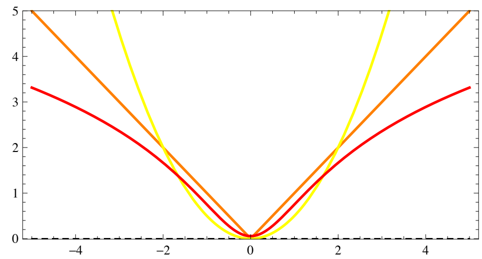
General form of MAP estimator

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



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

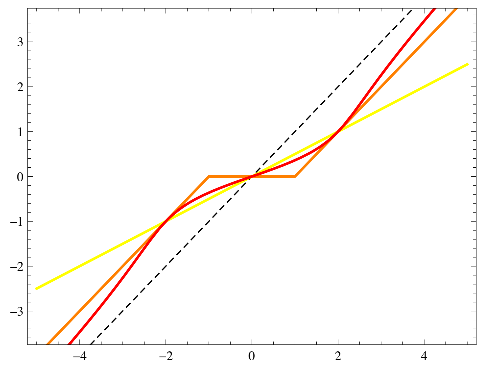
Potential: $\Phi_U(x) = -\log p_U(x)$



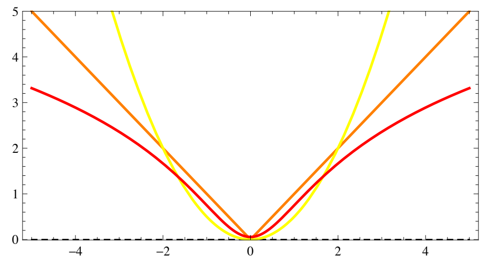
Proximal operator: pointwise denoiser

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

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



$\sigma^2\Phi_U(u)$



- 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: $\boldsymbol{\alpha}$

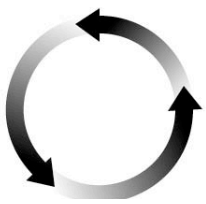
$$\mathcal{L}_{\mathcal{A}}(\mathbf{s},\mathbf{u},\boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 + \sigma^2 \sum_n \Phi_U([\mathbf{u}]_n) + \boldsymbol{\alpha}^T(\mathbf{L}\mathbf{s}-\mathbf{u}) + \frac{\mu}{2}\|\mathbf{L}\mathbf{s}-\mathbf{u}\|_2^2$$

(Bostan et al. *IEEE TIP* 2013)

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



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

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

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

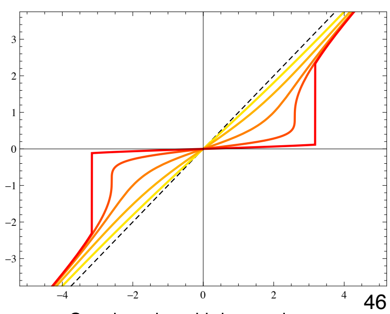
Linear inverse problem: $\mathbf{s}^{k+1} = \left(\mathbf{H}^T\mathbf{H} + \mu\mathbf{L}^T\mathbf{L}\right)^{-1}\left(\mathbf{H}^T\mathbf{y} + \mathbf{z}^{k+1}\right)$

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

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

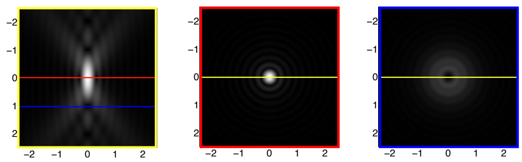


Cauchy prior with increasing s_0

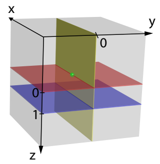
Deconvolution of fluorescence micrographs

- 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

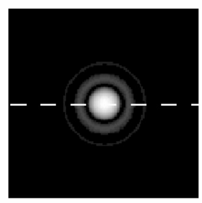
- λ: wavelength (emission)
- M: magnification factor
- f_0: focal length
- $P(\omega_1,\omega_2) = \mathbb{1}_{\|\boldsymbol{\omega}\|<R_0}$: pupil function
- $\mathrm{NA} = n \sin\theta = R_0/f_0$: numerical aperture

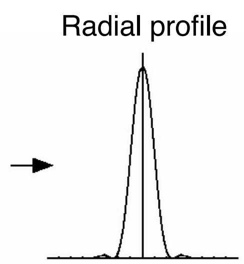
2-D convolution model

Thin specimen $s(x,y)$

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

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





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

- Modulation transfer function

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

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

2-D deconvolution: numerical set-up

- Discretization

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

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

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

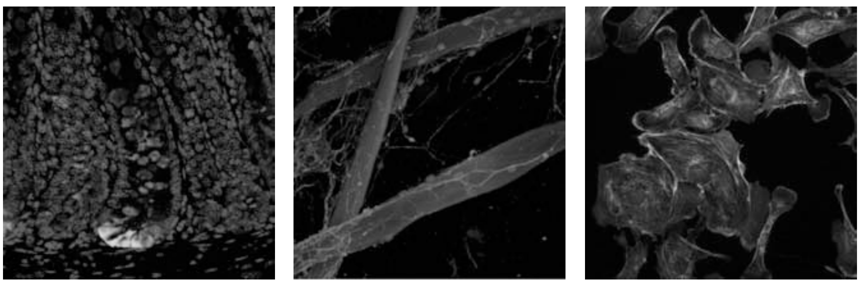
  $\mathbf{H}$ and $\mathbf{L}$: convolution matrices diagonalized by discrete Fourier transform

- Linear step of ADMM algorithm implemented using the FFT

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

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

Deconvolution experiments



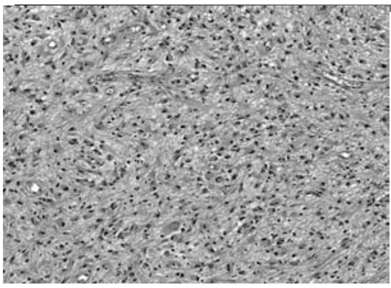
(a) (b) (c)

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

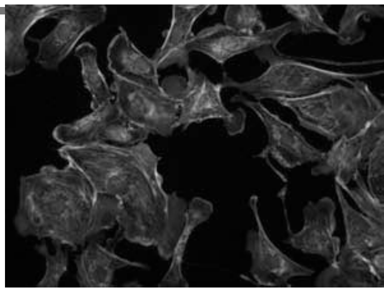
Table 10.2 Deconvolution performance of MAP estimators based on different prior distributions.

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

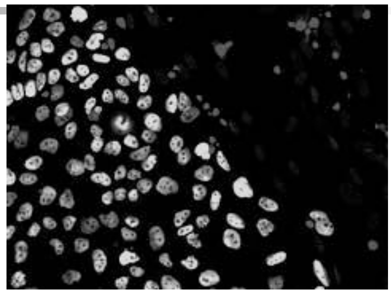
2D deconvolution experiment



Astrocytes cells



bovine pulmonary artery cells



human embryonic stem cells

Disk shaped PSF (7x7)

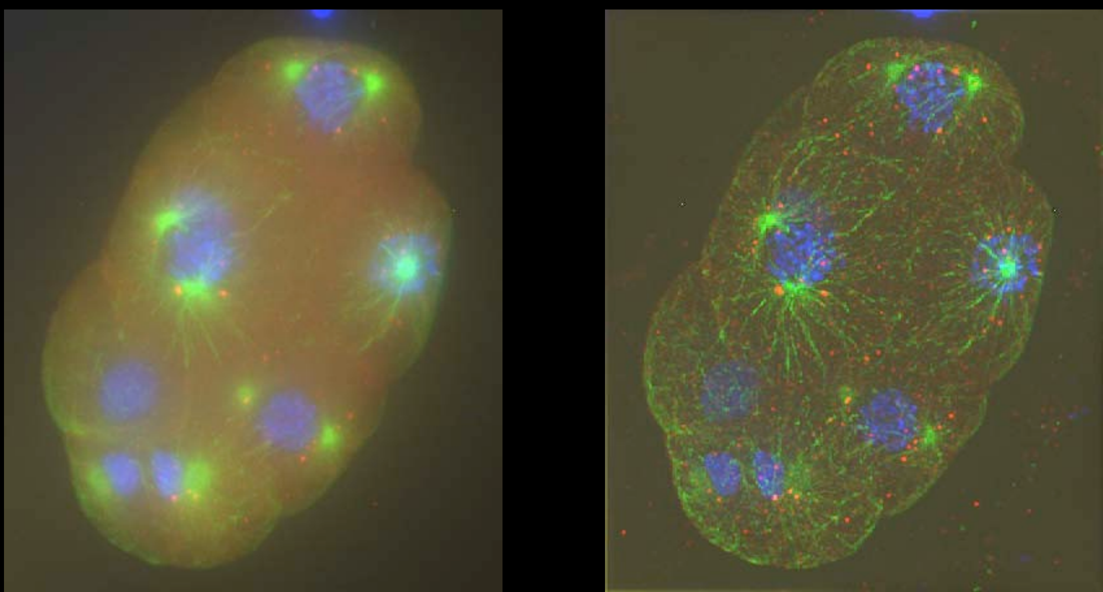
L : gradient

Deconvolution results in dB

Optimized parameters

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

3D deconvolution of widefield stack



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;
wavelet regularization (Haar), 3 decomposition levels for X-Y, 2 decomposition levels for Z.

(Vonesch-U., *IEEE TIP* 2009)

Magnetic resonance imaging (MRI)

- Physical image formation model (noise-free)

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

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

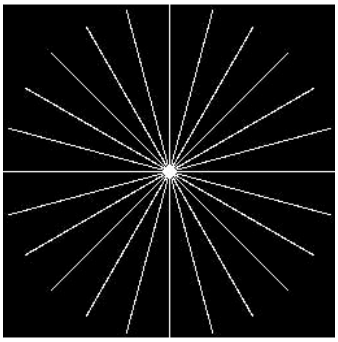
- Discretization in separable sinc basis

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

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

MRI: Shepp-Logan phantom

Original SL Phantom



Fourier Sampling Pattern
12 Angles

Laplace prior (TV)

Student prior (log)

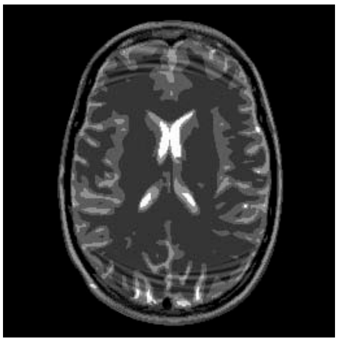
L : gradient
Optimized parameters

MRI phantom: Spiral sampling in k-space

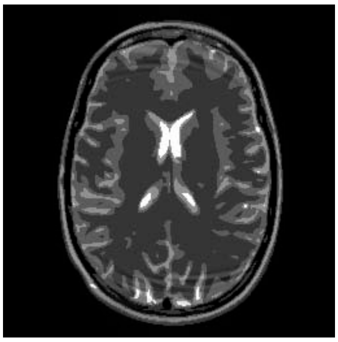
L : gradient

Optimized parameters

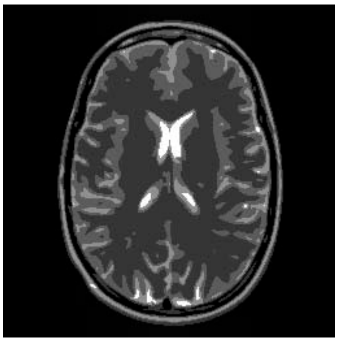
Original Phantom
(Guerquin-Kern TMI 2012)



Gaussian prior (Tikhonov)
SER =17.69 dB



Laplace prior (TV)
SER = 21.37 dB



Student prior
SER = 27.22 dB

MRI reconstruction experiments

(a) (b) (c)

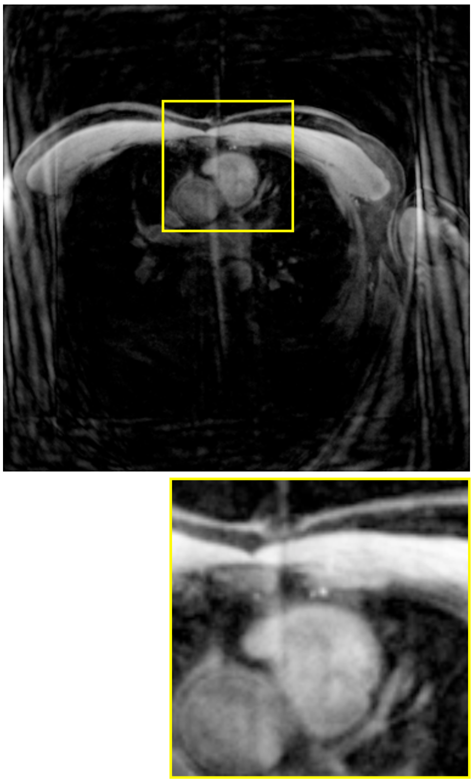
Figure 10.4 Data used in MR reconstruction experiments. (a) Cross section of a wrist. (b) Angiography image. (c) k-space sampling pattern along 40 radial lines.

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

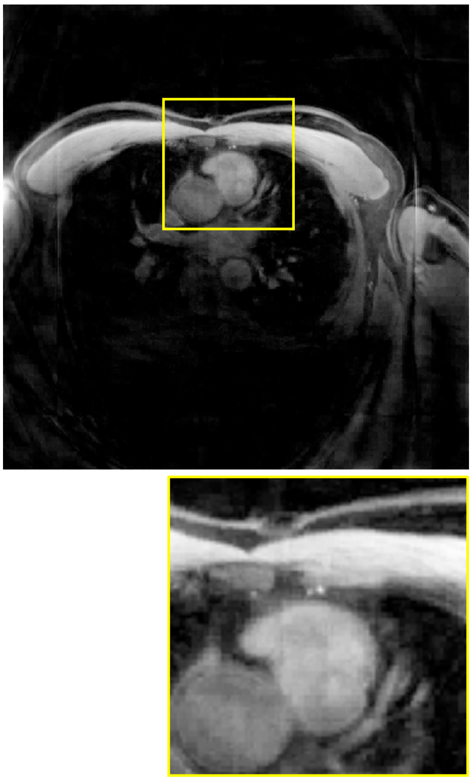
	Radial lines	Estimation performance (SNR in dB) Gaussian	Laplace	Student's
Wrist	20	8.82	**11.8**	5.97
	40	11.30	**14.69**	13.81
Angiogram	20	4.30	9.01	**9.40**
	40	6.31	14.48	**14.97**

ISMRM reconstruction challenge

L_2 regularization (Laplacian)



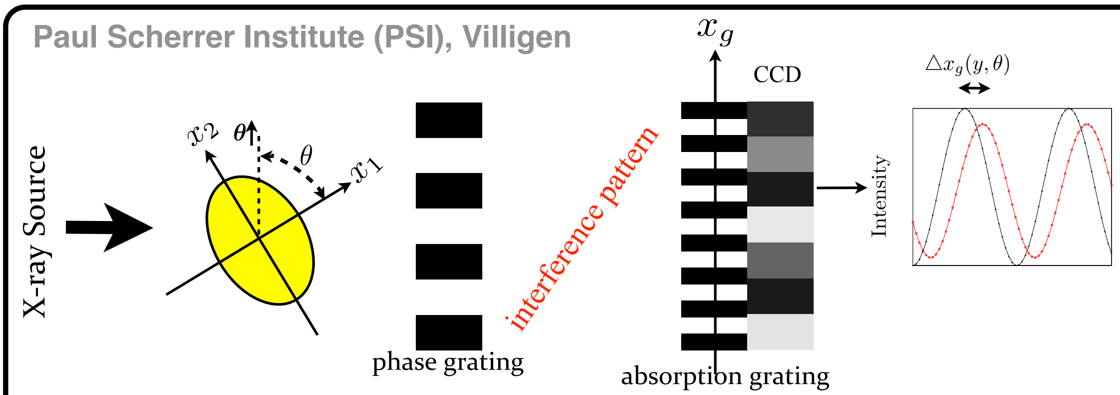
ℓ_1 wavelet regularization



(Guerquin-Kern *IEEE TMI* 2011)

Differential phase-contrast tomography

Paul Scherrer Institute (PSI), Villigen



(Pfeiffer, *Nature* 2006)

Mathematical model

$$y(t,\theta) = \frac{\partial}{\partial t}\mathrm{R}_\theta\{s\}(t)$$

$$\mathbf{y} = \mathbf{H}\,\mathbf{s}$$

$$[\mathbf{H}]_{(i,j),\mathbf{k}} = \frac{\partial}{\partial t}\mathrm{P}_{\theta_j}\beta_{\mathbf{k}}(t_j)$$

Properties of Radon transform

- Projected translation invariance

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

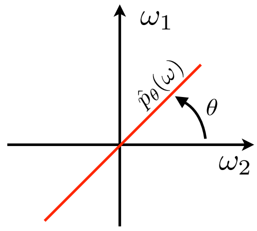
- Pseudo-distributivity with respect to convolution

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

- Fourier central-slice theorem

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

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



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

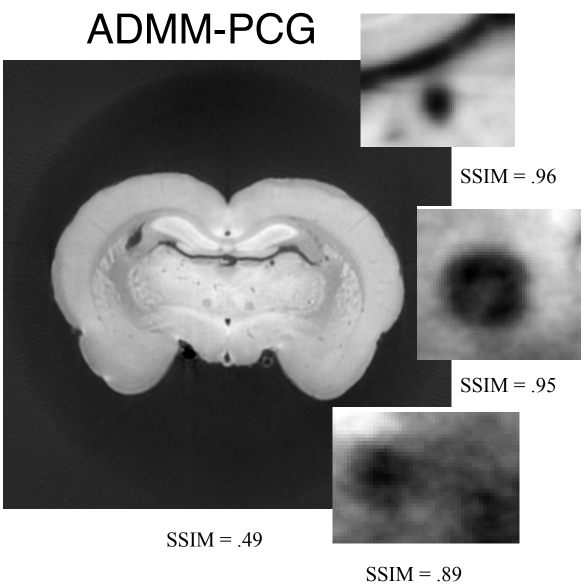
$$\mathrm{R}_\theta\{\varphi(\cdot - \boldsymbol{x}_0)\}(t) = \varphi_\theta(t - t_0)$$

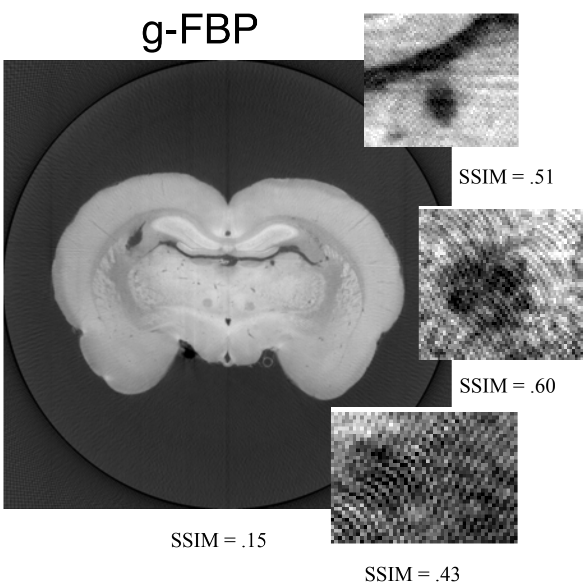
where $t_0 = \langle \boldsymbol{x}_0, \boldsymbol{\theta}\rangle$ and

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

Reducing the numbers of views

Rat brain reconstruction with 181 projections



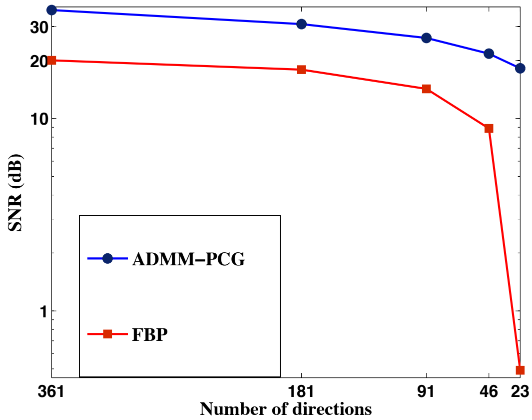


Collaboration: Prof. Marco Stampanoni, TOMCAT PSI / ETHZ

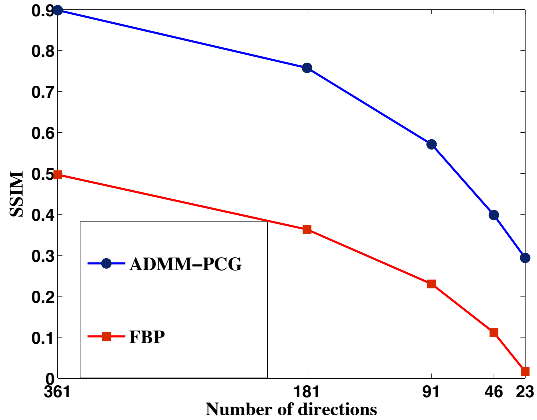
(Nichian et al. *Optics Express* 2013)

Performance evaluation

Goldstandard: high-quality iterative reconstruction with 721 views



(a)



(b)

⇒ Reduction of acquisition time by a factor 10 (or more) ?

Physical model — Statistical model of signal

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

Schematic structure of reconstruction algorithm:

N_{iter}

Repeat

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

$\mathbf{u}^{(\mathbf{n})} = \arg\min_{\mathbf{u}} J(\mathbf{x}^{(n)}, \mathbf{u})$: Statistical or "denoising" step

until stop criterion

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)

- **Faster imaging, reduced radiation exposure**: Reconstruction from a lesser number of measurements supported by **compressed sensing**. (Candes-Romberg-Tao; Donoho, 2006)

- **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)

- Outstanding research issues
  - Beyond ℓ_1 and TV: Connection with **statistical modeling & learning**
  - Beyond matrix algebra: **Continuous-domain** formulation

Part 4:

Short guess about the future:
The (deep) learning revolution (??)

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**
  (based of sparsity or statistics/ICA) $\Rightarrow$ "optimal" $\mathbf{L}$

  (Elad 2006, Ravishankar 2011, Mairal 2012)

- Learning of non-linearities / Proximal operators

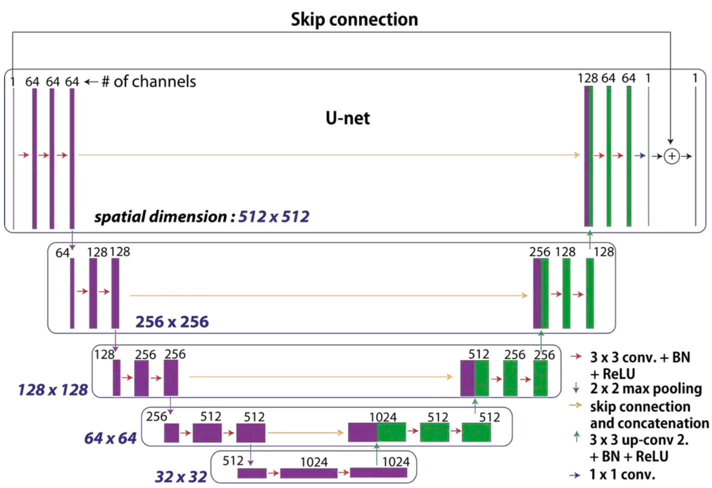
  CNN-type parametrization, backpropagation $\Rightarrow$ "optimal" potential Φ

  (Chen-Pock 2015-2016, Kamilov 2016)

Recent appearance of Deep Conv Neural Nets

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

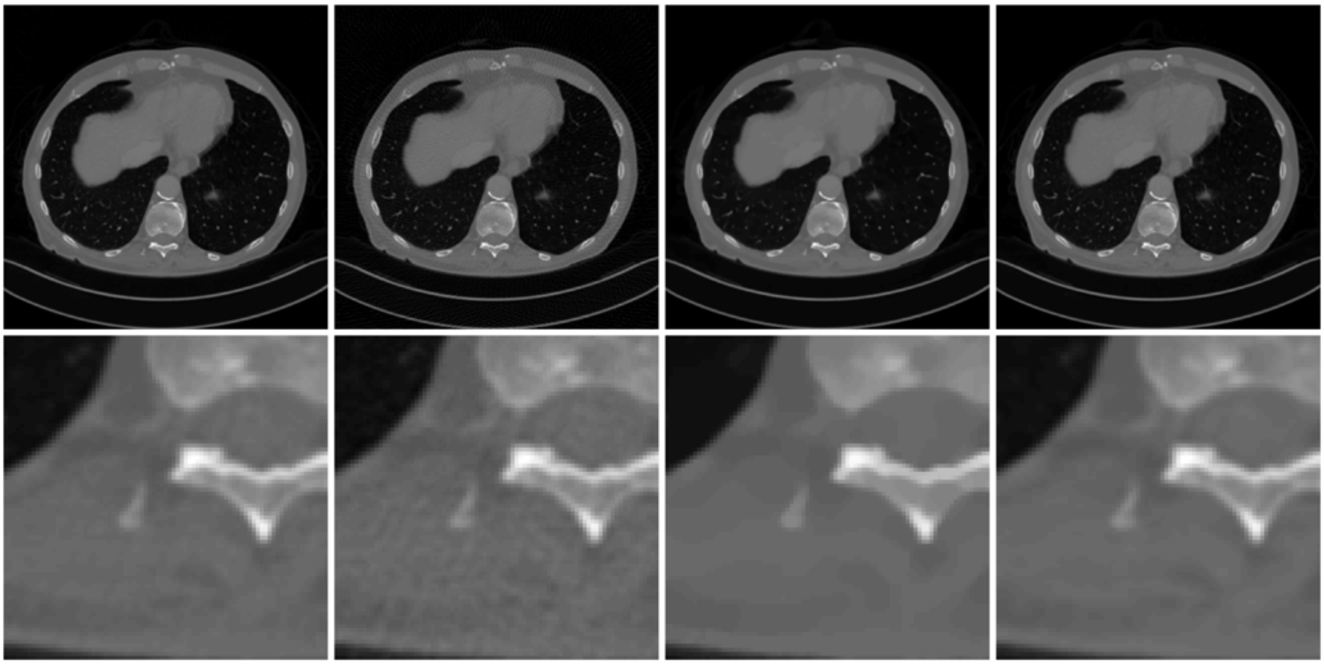
- CT reconstruction based on Deep ConvNets
  - Input: Sparse view FBP reconstruction
  - Training: Set of 500 high-quality full-view CT reconstructions
  - Architecture: U-Net with skip connection (Jin et al., arXiv:1611.03679)



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

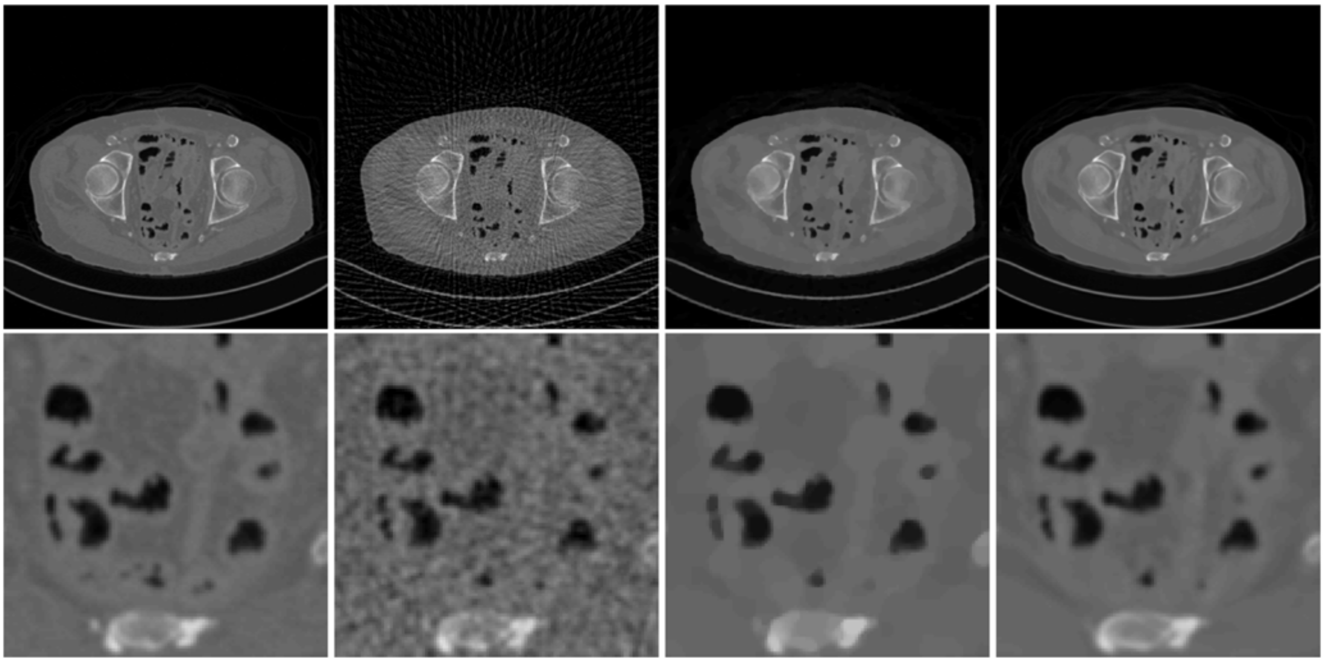
(Jin et al., arXiv:1611.03679)

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

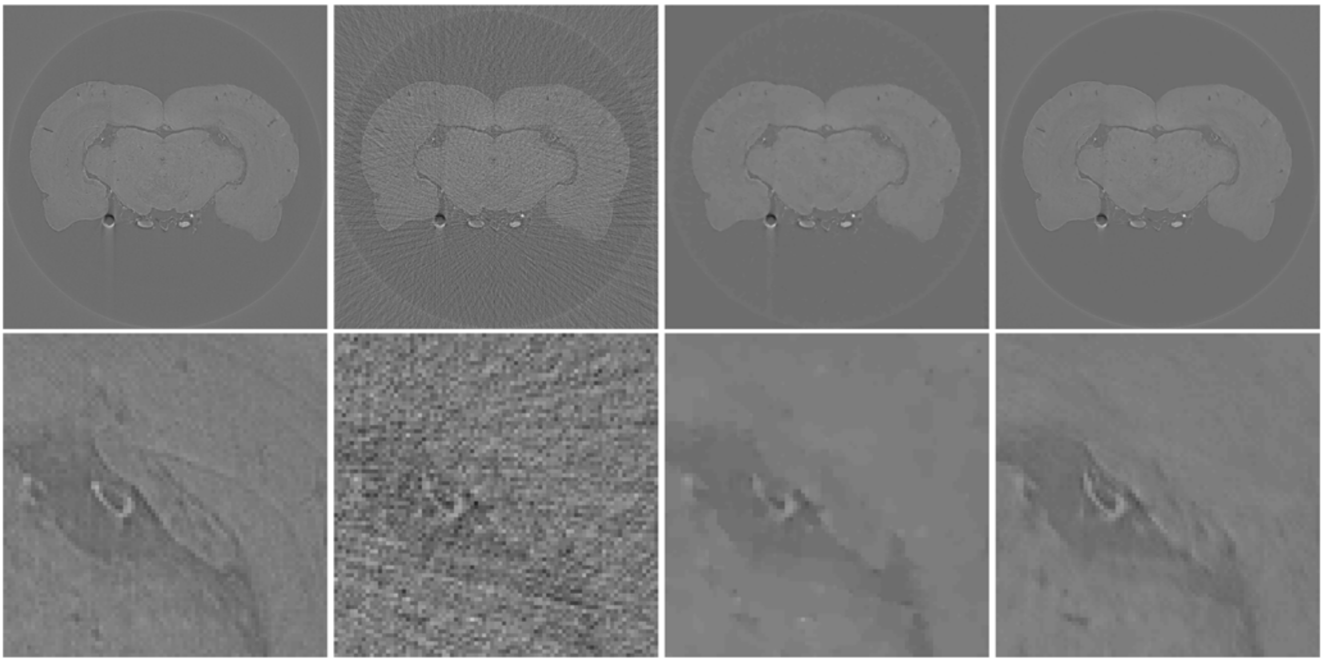
(Jin et al., arXiv:1611.03679)

MAYO CLINIC

Dose reduction by 14: 51 views

μCT data

Ground truth	FBP SNR 3.265	TV SNR 7.481	FBPConvNet SNR 9.003



Reconstructed from
from 721 views

COMPARISON OF SNR BETWEEN DIFFERENT RECONSTRUCTION ALGORITHMS FOR EXPERIMENTAL DATASET.

Metrics \ Methods		FBP	TV [13]	Proposed
avg. SNR (dB)	145 views (x5)	5.38	8.25	11.34
	51 views (x14)	3.29	7.25	8.85

PAUL SCHERRER INSTITUT

Challenges for deep learning methods

- Fundamental change of paradigm

  Requires availability of **extensive sets of representative training data** together with **gold-standards** = desired high-quality reconstruction

- Research challenges/opportunities
  - How does one assess **reconstruction quality** ?
    Should be “task oriented”!!!
  - Use of CNN to **correct artifacts** of current methods
  - Reconstruction from **fewer measurements**
    (trained on high-quality full-view data sets).
  - Use of CNN to **emulate/speedup** some well-performing, but “slow”, reference reconstruction methods
  - Development of more **realistic simulators**
    both “ground truth” images + physical forward model
  - **True 3D** CNN toolbox (still missing)

Can we trust the results ?

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].

- 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.
  - M.T. McCann, M. Nilchian, M. Stampanoni, M. Unser, "Fast 3D Reconstruction Method for Differential Phase Contrast x-Ray CT," *Optics Express*, vol. 24, no. 13, pp. 14564-14581, 2016.
  - K.H. Jin, M.T. McCann, E. Froustey, M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," arXiv:1611.03679 [cs.CV].

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/

General convex problems with gTV regularization

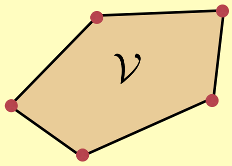
$$\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) = \{s : \mathrm{gTV}(s) = \|\mathrm{L}\{s\}\|_{\mathcal{M}} = \sup_{\|\varphi\|_\infty \le 1} \langle \mathrm{L}\{s\}, \varphi\rangle < \infty\}$$

- **Linear** measurement operator $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \to \mathbb{R}^M : f \mapsto \mathbf{z} = \mathrm{H}\{f\}$
- $\mathcal{C}$: **convex** compact subset of $\mathbb{R}^M$
- Finite-dimensional **null space** $\mathcal{N}_{\mathrm{L}} = \{q \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) : \mathrm{L}\{q\} = 0\}$ with basis $\{p_n\}_{n=1}^{N_0}$

Admissibility of regularization: $\mathrm{H}\{q_1\} = \mathrm{H}\{q_2\} \Leftrightarrow q_1 = q_2$ for all $q_1, q_2 \in \mathcal{N}_{\mathrm{L}}$

Representer theorem for gTV regularization

The extremal points of the constrained minimization problem



$$\mathcal{V} = \arg \min_{f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)} \|\mathrm{L}\{f\}\|_{\mathcal{M}} \quad \text{s.t.} \quad \mathrm{H}\{f\} \in \mathcal{C}$$

are necessarily 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 $K \le M - N_0$; that is, **non-uniform** L**-splines** with knots at the $\boldsymbol{x}_k$ and $\|\mathrm{L}\{f\}\|_{\mathcal{M}} = \sum_{k=1} |a_k|$. The full solution set is the **convex hull** of those extremal points.

(U.-Fageot-Ward, SIAM Review, in Press)