EPFL

Image Processing II

Chapter 10
Computational Imaging

Prof. Michael Unser
EPFL Center for Imaging

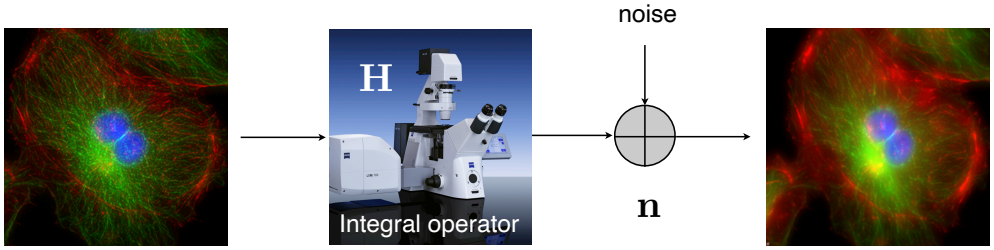
May 2025

OUTLINE

- **1. Imaging as an inverse problem**
  - Basic imaging operators
  - Discretization of the inverse problem
    - Appendix A: Magnetic Resonance Imaging
- **2. Classical reconstruction algorithms**
  - Vector calculus
  - Basic reconstruction: Least-squares solution
  - Tikhonov regularization
  - Iterative reconstruction
  - Wiener / LMSE solution
- **3. 2nd generation methods: the sparsity (re)evolution**
  - Sparsity and compressed sensing
  - Elements of convex analysis
  - Proximity operators
  - Image reconstruction under sparsity constraints
- **4. The learning revolution**
  - Image reconstruction using deep neural networks
  - iterative (PnP) schemes

Computational imaging = Image reconstruction

- 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

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

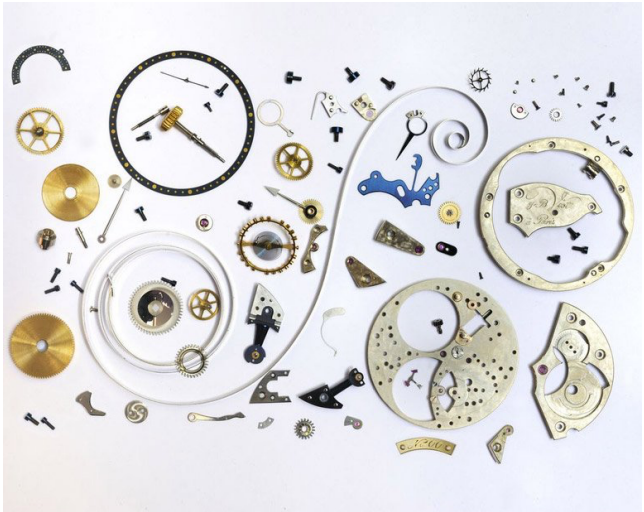
- Backprojection (

Basic limitations

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

Part 1:

Setting up the problem



1.1 BASIC IMAGING OPERATORS

- Fourier transform
- Windowing
- Convolution
- Radon transform
- Panorama of imaging modalities

Forward imaging model (noise-free)

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

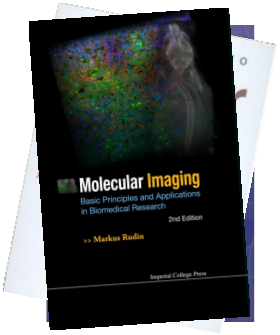
defined over a continuum in space-time

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

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)

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)

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

Foolproof usage:

$$\mathcal{F} : L_1(\mathbb{R}^d) \to C_0(\mathbb{R}^d) \triangleq \left\{ f : \mathbb{R}^d \to \mathbb{R} \text{ continuous with } \|f\|_{L_\infty} < \infty \text{ and } f(\boldsymbol{x}) \to 0 \text{ as } \|\boldsymbol{x}\| \to +\infty \right\}$$

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

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

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

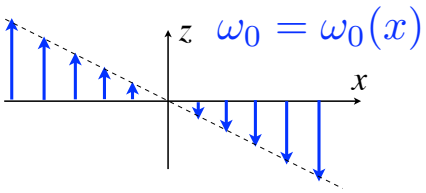
$$\mathrm{e}^{\mathrm{j}\langle \boldsymbol{\omega}_0, \boldsymbol{r} \rangle} f(\boldsymbol{r}) \quad \overset{\mathcal{F}}{\longleftrightarrow} \quad \hat{f}(\boldsymbol{\omega} - \boldsymbol{\omega}_0)$$

Application: Structured illumination microscopy (SIM)

Magnetic resonance imaging

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

Frequency encode:



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

- Linear forward model for MRI

$$\hat{s}(\boldsymbol{\omega}_m) = \int_{\mathbb{R}^3} s(\boldsymbol{r}) \mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}_m, \boldsymbol{r} \rangle} \mathrm{d}\boldsymbol{r}$$ (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: 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: $\hat{h}(\boldsymbol{\omega}) = \mathcal{F}\{h\}(\boldsymbol{\omega})$ with $\hat{h} \in L_\infty(\mathbb{R}^d)$

- Convolution as a frequency-domain product

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

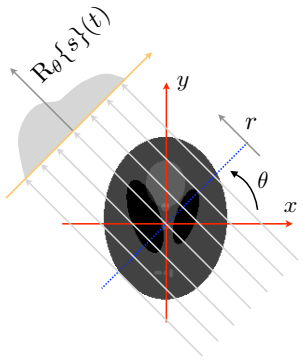
Basic operator: X-ray transform

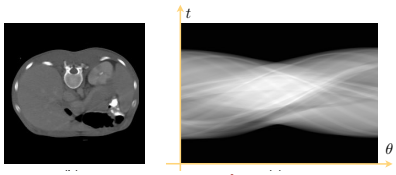
Projection geometry: $\boldsymbol{x} = t\boldsymbol{\theta} + r\boldsymbol{\theta}^{\perp}$ with $\boldsymbol{\theta} = (\cos\theta, \sin\theta)$

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

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	$y_i = \mathrm{H}\mathrm{W}_i x$ 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

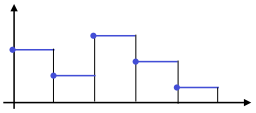
1.2 DISCRETIZATION

- Discretization: Finite dimensional formalism
- Exemples
  - Diffraction-limited convolution (Fluorescence microscopy)
  - MRI
- Appendix A: Magnetic Resonance Imaging

Discretization: Finite dimensional formalism

Selection of appropriate basis functions: $\beta_{\boldsymbol{k}} : \mathbb{R}^d \to \mathbb{R}$ with $\boldsymbol{k} \in \Omega \subset \mathbb{Z}^d$ and $\mathrm{Card}(\Omega) = K$

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



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

- Measurement model (image formation)

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

η_m: sampling/imaging function (mth detector)

$n[\cdot]$: additive noise

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

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

Example of basis functions

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

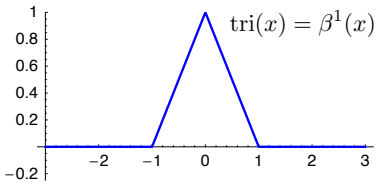
Separable generator: $\beta(\boldsymbol{x}) = \prod_{n=1}^{d} \beta(x_n)$

- Pixelated model

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

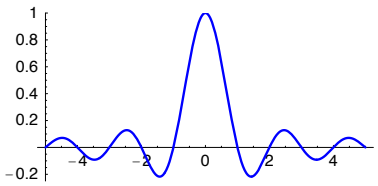
- Bilinear model

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



- Bandlimited representation

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



Example 1: Diffraction-limited convolution

Hypothesis: $\mathcal{F}\{h_{2\mathrm{D}}\}(\boldsymbol{\omega}) = \hat{h}_{2\mathrm{D}}(\boldsymbol{\omega}) = 0$ for $\|\boldsymbol{\omega}\| \geq \omega_0$ (Diffraction-limited optical transfer function)

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

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

  $\mathbf{H}$: convolution matrix diagonalized by discrete Fourier transform

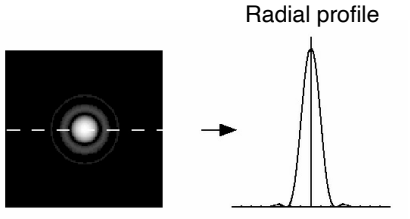
- Matrix-free implementation using the FFT

  The discretized forward model $\mathbf{s} \mapsto \mathbf{H}\mathbf{s}$ is implemented efficiently using the FFT

Diffraction-limited 2D model of a fluorescence microscope

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

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



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

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

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

Example 2: Magnetic resonance imaging (MRI)

- Physical image formation model (noise-free)

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

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

- Discretization in separable sinc basis

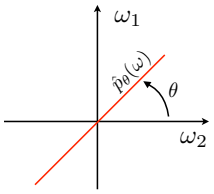
$$\begin{aligned} [\mathbf{H}]_{m,\boldsymbol{k}} &= \langle \eta_m, \mathrm{sinc}(\cdot - \boldsymbol{k})\rangle \\ &= \langle \mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}_m, \cdot\rangle}, \mathrm{sinc}(\cdot - \boldsymbol{k})\rangle = \mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}_m, \boldsymbol{k}\rangle}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

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

Example 3: Discretization of the Radon transform

Shift-invariant basis: $s(\boldsymbol{x}) = \sum_{\boldsymbol{n} \in \Omega} s[\boldsymbol{n}] \varphi(\boldsymbol{x} - \boldsymbol{n})$ with $\Omega \subset \mathbb{Z}^2$

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



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

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

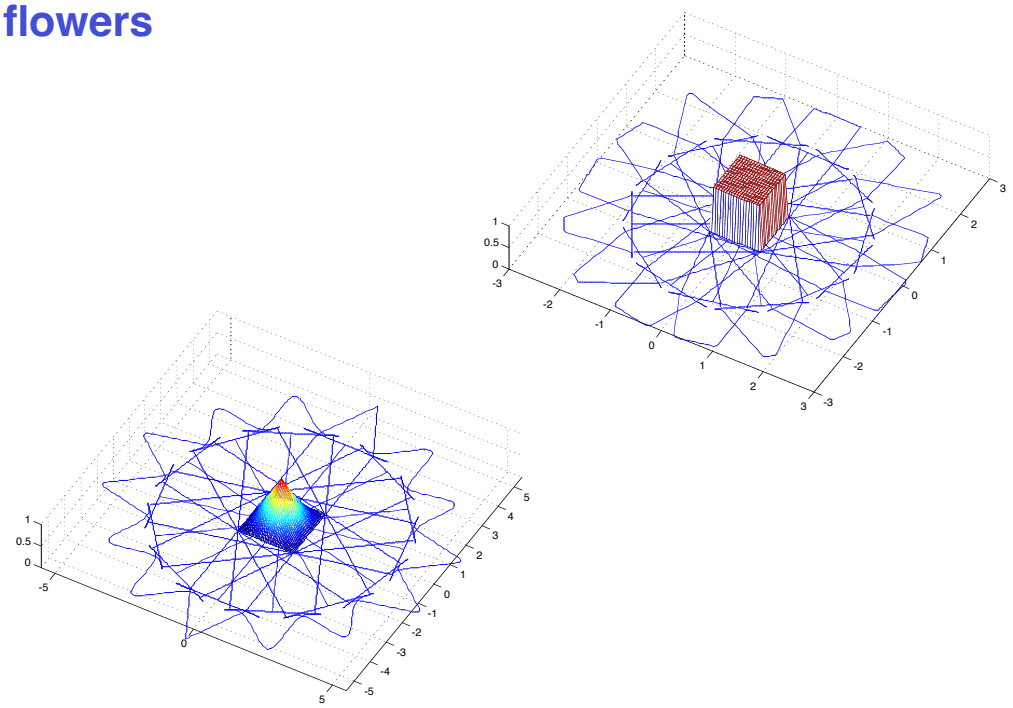
$$\mathrm{R}_\theta\{\varphi(\cdot - \boldsymbol{x}_0)\}(t) = \varphi_\theta(t - t_0)$$

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

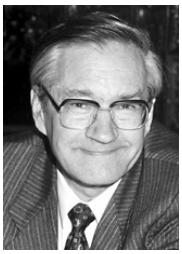
$$\varphi_\theta(t) = \left(\tfrac{1}{|\cos\theta|} \varphi_1\left(\tfrac{\cdot}{\cos\theta}\right) * \tfrac{1}{|\sin\theta|} \varphi_2\left(\tfrac{\cdot}{\sin\theta}\right) \right)(t).$$

$$[\mathbf{H}]_{(i,j),\boldsymbol{n}} = \mathrm{R}_{\theta_i}\{\varphi(\cdot - \boldsymbol{n})\}t - j) = \varphi_{\theta_i}(t_j - \langle \boldsymbol{n}, \boldsymbol{\theta}_i \rangle)$$

Spline flowers



Appendix A: Magnetic Resonance Imaging (MRI)



Richard R. Ernst (ETHZ)

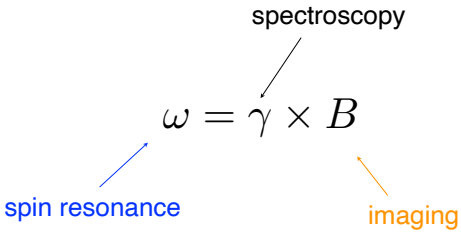
Nobel Prize in Chemistry 1991

Paul C. Lauterbur

Peter Mansfield

Nobel Prize in Medecine 2003

Exploits Lamor precession of nuclear spins:

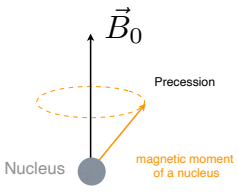


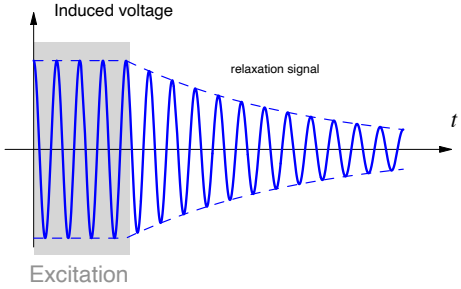
Physical principle of magnetic resonance

- Spins as signature of the atomic structure
  - The nuclei of individual atoms are characterized by a spin and a magnetic moment: they are equivalent to small magnets.
  - When exposed to a large external magnetic field $\vec{B}_0$, the nuclear spins tend to resonate and precess at the ***Larmor frequency***

    $\omega_0 = \gamma B_0$

    γ: gyromagnetic constant associated with a given nuclear type (e.g. ^{1}H ou ^{13}C)

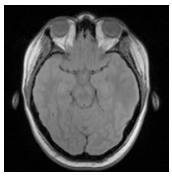




Example: $B_0 = 1.5$ Tesla corresponds to $f_0 = \frac{\omega_0}{2\pi} \approx 64$ MHz for Hydrogen nuclei

Magnetic Resonance Imaging (MRI)

"whole body" scanner (1.5 Tesla magnet)

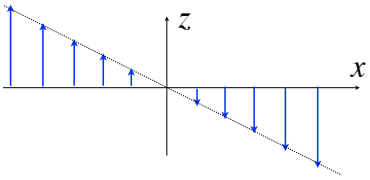


Proton density

- Controlled RF excitation
  - Signal induced by the spins of hydrogen nuclei (^{1}H=proton) exclusively
  - The relaxation decay during acquisition is negligeable
    $\rightarrow$ Elementary MNR responses are quasi-sinusoidal
  - One only considers a spin density $p(x)$ along the x axis (1D imaging)

Frequency encoding

$$\omega_0 = \omega_0(x)$$

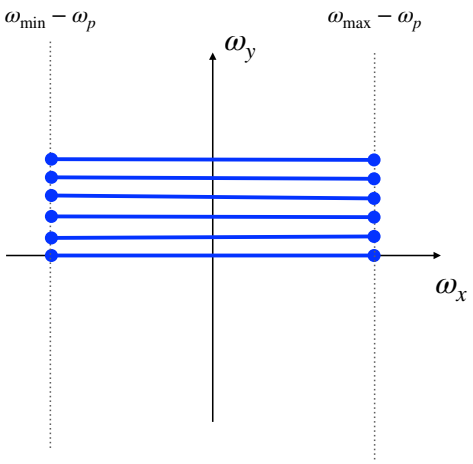


- Excitation via a bandlimited RF pulse: $\omega_0(x) \in (\omega_{\min}, \omega_{\max})$
- A gradient along x is applied during the measurement:
  $$B_z = B_z(x) = B_0 - \frac{G_x}{\gamma} x$$
- Measurements: the FID signal that is received back from the sample in the RF coil
- The signal is demodulated with its central frequency being mapped to 0.

- Principe of frequency encoding
  - Signal due to a point source of density p_0 at position x_0:
    $$s(t) = p_0 \mathrm{e}^{-\mathrm{j}(G_x x_0)t}$$ (pure sinusoid)
    The position x_0 is encoded in the frequency $\omega(x_0) = -G_x x_0$.
  - Signal due to a continuum of sources with density distribution $p(x)$
    $$s(t) = \int_{-\infty}^{\infty} p(x) \mathrm{e}^{-\mathrm{j}x(G_x t)} \mathrm{d}x$$
  - The spin density along x is recovered by inverse Fourier transform:
    $$p(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} s(t)|_{t=\omega/G_x} e^{j\omega x} \mathrm{d}\omega$$

Beyond one dimensional MRI

- Three-dimensional localization
  - Localization in z.
    A static magnetic field that is linear along z (gradient) is applied ***during*** the excitation. In this way, the RF pulse only excites a slice perpendicular to z: $\omega(z_0) \in (\omega_{\min}, \omega_{\max})$.
  - Localization in y by phase encoding.
    A linear gradient in y is applied during a suitable time **after** the excitation but **before** the acquisition. The process is iterated such as to sample the 2D Fourier transform along the ω_y axis.
  - Localization in x by frequency encoding.
    Measurement of the 2D Fourier transform of a slice along ω_x with $\omega_y = \mathrm{Const}$.



Part 2:

Classical image reconstruction

Discretized forward model: **y**=**H**s+ **n**

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

Vector calculus

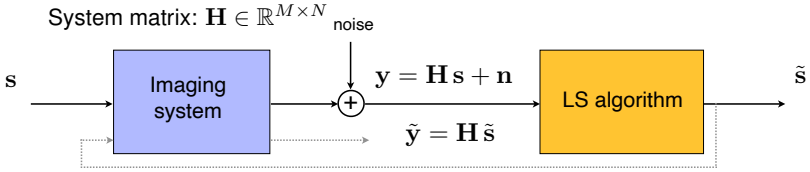
- Scalar cost function $J(\mathbf{v}) : \mathbb{R}^N \to \mathbb{R}$
- Vector differentiation: $\dfrac{\partial J(\mathbf{v})}{\partial \mathbf{v}} = \begin{bmatrix} \partial J/\partial v_1 \\ \vdots \\ \partial J/\partial v_N \end{bmatrix} = \nabla J(\mathbf{v})$ (gradient)
- Necessary condition for an unconstrained optimum (minimum or maximum)

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

- Useful identities

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

Basic reconstruction: least-squares solution



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

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

- Formal least-squares solution

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

  $\frac{\partial J_{\mathrm{LS}}(\mathbf{s},\mathbf{y})}{\partial \mathbf{s}} = 2\mathbf{H}^T\mathbf{H}\mathbf{s} - 2\mathbf{H}^T\mathbf{y} = 0 \quad \Rightarrow \quad \mathbf{s}_{\mathrm{LS}} = \arg\min\limits_{\mathbf{s}} J_{\mathrm{LS}}(\mathbf{s}, \mathbf{y}) = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}$

- Backprojection (poor man's solution): $\mathbf{s} \approx \mathbf{H}^T\mathbf{y}$

  OK if $\mathbf{H}$ is unitary $\Leftrightarrow$ $\mathbf{H}^{-1} = \mathbf{H}^T$

Data fit and maximum-likelihood estimation

- Using the knowledge of the noise distribution

  Hypothesis: independent, zero-mean measurement noise with known PDF $p(\mathbf{n})$

  Measurement equation $\Leftrightarrow$ $\mathbf{y} - \mathbf{H}\mathbf{s} = \mathbf{n}$

  Expected measurement: $\mathbb{E}\{\mathbf{y}\} = \mathbf{H}\mathbf{s}$ (detector-wise)

  Conditional probability density: $p(\mathbf{y}|\mathbf{s}) = \prod_{m=1}^{M} p(n_m)$

- Maximum-likelihood estimation (ML)

  $$J_{\mathrm{ML}}(\mathbf{s}, \mathbf{y}) = \log\left(p(\mathbf{y}|\mathbf{s})\right) = \sum_{m=1}^{M} \log\left(p(n_m)\right) = -J_{\mathrm{data}}(\mathbf{s}, \mathbf{y})$$

  Maximization of the ML criterion yields the most “likely” solution

- Special case: Gaussian white noise

  $$p(\mathbf{y}|\mathbf{s}) = p(\mathbf{n}) = \prod_{m=1}^{M} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_m - [\mathbf{H}\mathbf{s}]_m)^2}{2\sigma^2}\right)$$

  $$\Rightarrow J_{\mathrm{ML}}(\mathbf{s}, \mathbf{y}) = -\sum_{m=1}^{M} \left(\frac{(y_m - [\mathbf{H}\mathbf{s}]_m)^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right) \quad \propto \quad -J_{\mathrm{LS}}(\mathbf{s}, \mathbf{y}) + \mathrm{Const}$$

Making use of regularization

From a global perspective, ML deconvolution is ill-posed. It is better to constrain the solution by imposing explicit regularization constraints:

$$J(\mathbf{s}, \mathbf{y}) = J_{\mathrm{data}}(\mathbf{s}, \mathbf{y}) + \lambda \cdot J_{\mathrm{reg}}(\mathbf{s})$$

$\lambda \geq 0$: regularization factor (adjustable)

- Many possible choices of regularization
  - Quadratic or Tikhonov: $J_{\mathrm{reg}}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|^2$
    where $\mathbf{L}$ is a suitable differential operator (e.g., Laplacian) that penalizes oscillations
  - Maximum a posteriori or MAP (using Bayes rule):
    $J_{\mathrm{MAP}}(\mathbf{s}) \propto -\log(p(\mathbf{s}))$ (prior probability density)
  - Total variation: $J_{\mathrm{TV}}(s) = \|\nabla s\|_1$ (favors piecewise-constant solutions)
  - General non-linear: $J_{\mathrm{reg}}(\mathbf{s}) = \sum_{n=1}^{N} \Phi([\mathbf{L}\mathbf{s}]_n)$ where $\Phi(\cdot)$ is a symmetric increasing function
  - Wavelet regularization: $J_{\mathrm{reg}}(\mathbf{s}) = \|\mathbf{W}^{\mathsf{T}}\mathbf{s}\|_{\ell_1}$
    where $\mathbf{W}$ is a suitable wavelet transform (favors “sparse" solutions)

see Part III

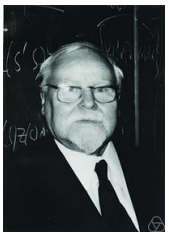
Linear inverse problems (20th century theory)

- 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

Iterative reconstruction algorithm

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

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

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

Steepest-descent algorithm

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

- Conjugate gradient
  - More efficient, but requires lot of storage

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

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 matrix-free 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

Statistical formulation (20th century)

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

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

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

Norbert Wiener (1894-1964)

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

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

$$-\log p_{\mathrm{Gauss}}(\mathbf{u}) = \tfrac{1}{2}\|\mathbf{u}\|_2^2 + \mathrm{Const.}$$

$\Updownarrow$ $\mathbf{L} = \mathbf{C}_s^{-1/2}$: Whitening filter $\qquad$ $\mathbf{u} = \mathbf{L}\mathbf{s}$: standardized Gaussian innovation

- Quadratic regularization (Tikhonov)

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

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

Appendix B: Convergence of iterative least squares

Basic Landweber/van Cittert iteration

$$\begin{aligned}\mathbf{s}^{(t+1)} &= \mathbf{s}^{(t)} + \gamma \mathbf{H}^{\mathsf{T}}(\mathbf{y} - \mathbf{H}\mathbf{s}^{(t)}) \\ &= \left(\mathbf{I} - \gamma \mathbf{H}^{\mathsf{T}}\mathbf{H}\right)\mathbf{s}^{(t)} + \gamma \mathbf{H}^{T}\mathbf{y}\end{aligned}$$

$$\Rightarrow \quad (\mathbf{s}^{(t+1)} - \mathbf{s}^{(t)}) = \left(\mathbf{I} - \gamma \mathbf{H}^{\mathsf{T}}\mathbf{H}\right)(\mathbf{s}^{(t)} - \mathbf{s}^{(t-1)})$$

Converges iff. $\mathbf{A} = \left(\mathbf{I} - \gamma \mathbf{H}^{\mathsf{T}}\mathbf{H}\right)$ is a **contraction map**; i.e., $\rho(\mathbf{A}) = \max_{\|\mathbf{e}\|=1} \|\mathbf{A}\mathbf{e}\| < 1$

- Spectral analysis

  $\mathbf{A}$ and $\mathbf{H}^{\mathsf{T}}\mathbf{H}$ are symmetric matrices that share the same eigenvectors $\mathbf{u}_n$

  $$(\mathbf{H}^{\mathsf{T}}\mathbf{H})\mathbf{u}_n = \lambda_n \mathbf{u}_n \quad \Leftrightarrow \quad \mathbf{A}\mathbf{u}_n = \underbrace{(1 - \gamma\lambda_n)}_{\gamma_n}\mathbf{u}_n$$

- Condition for convergence

  $$\rho(\mathbf{A}) = \max_n |\gamma_n| < 1 \quad \Rightarrow \quad 0 < \gamma < \frac{2}{\lambda_{\max}\{\mathbf{H}^{\mathsf{T}}\mathbf{H}\}} \quad \text{(since } \lambda_n \geq 0 \text{ for all } n\text{)}$$

Part 3:

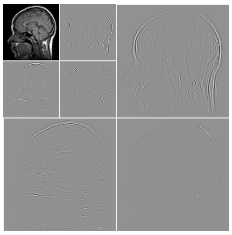
Sparsity-based image reconstruction

(2nd generation)

Linear inverse problems: Sparsity

(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)

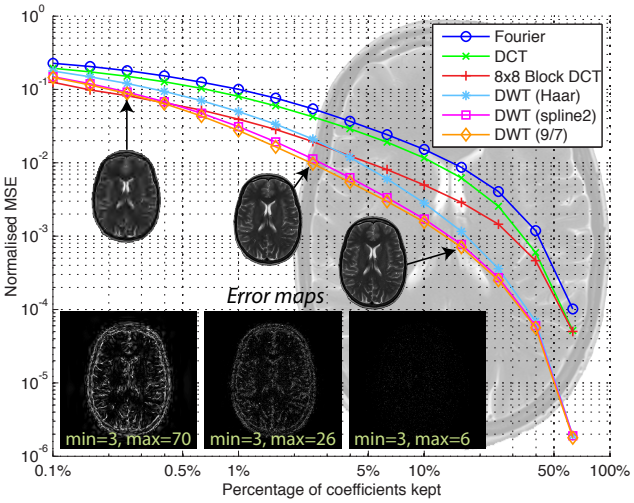
    $\mathbf{v} = \mathbf{W}^{-1}\mathbf{s}$: wavelet expansion of $\mathbf{s}$ (typically, sparse)

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

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

Sparsifying transforms

Biomedical images are well described by few basis coefficients



Prior = sparse representation

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

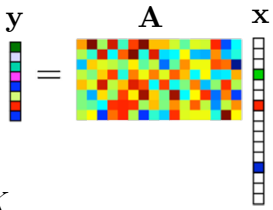
Advantages:
- convex
- favors sparse solutions
- Fast: WFISTA

(Guerquin-Kern *IEEE TMI* 2011)

Theory of compressed sensing

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

- 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 $\|\mathbf{x}\|_0 = K_0 \ll K$
  - $M \times N$ system matrix : $\mathbf{A} = \mathbf{H}\mathbf{W}$



+ "noise"

- Formulation of ill-posed recovery problem when $2K_0 < M \ll K$

  (P0) $\min_{\mathbf{x}\in\mathbb{R}^K} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ subject to $\|\mathbf{x}\|_0 \leq K_0$

- Theoretical result

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

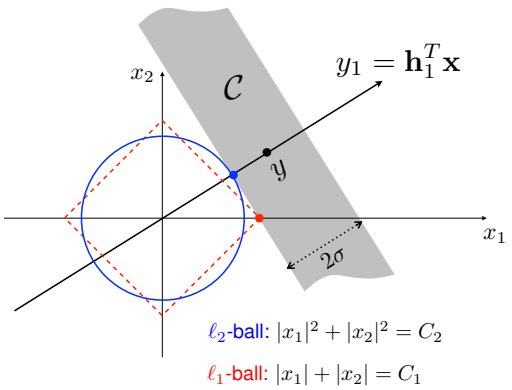
  (P1) $\min_{\mathbf{x}\in\mathbb{R}^K} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ subject to $\|\mathbf{x}\|_1 \leq C_0$

Geometry of l_2 vs. l_1 minimization

- Prototypical inverse problem

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

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

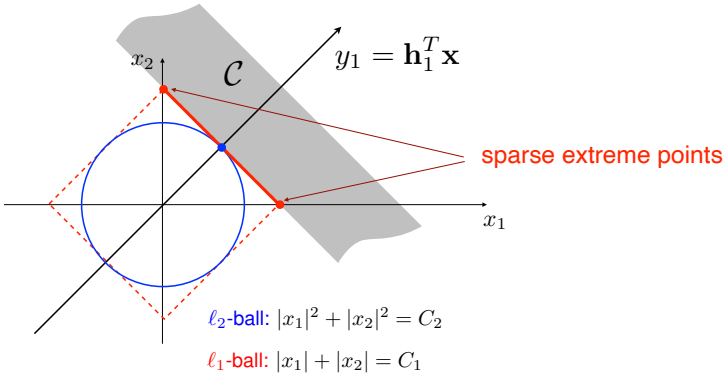


Geometry of l_2 vs. l_1 minimization

- Prototypical inverse problem

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

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



Configuration for **non-unique** ℓ_1 solution

Elements of convex analysis

Definitions

Extended real line: $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$. A function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ is said to be:

(i) **proper** if there exists at least one $\mathbf{x}_0$ such that $f(\mathbf{x}_0) < +\infty$ (by default).

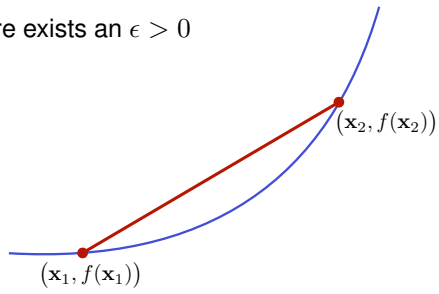
(ii) **coercive** if $f(\mathbf{x}) \to +\infty$ as $\|\mathbf{x}\| \to +\infty$.

(iii) **lower semi-continuous** (l.s.c.) at a point $\mathbf{x}_0$ if, for every $y < f(\mathbf{x}_0)$, there exists an $\epsilon > 0$ such that $y < f(\mathbf{x})$ for every $\mathbf{x} \in B_\epsilon(\mathbf{x}_0) = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x} - \mathbf{x}_0\|_2 < \epsilon\}$.

(iv) **convex** if, for all $\lambda \in (0, 1)$ and all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d$ such that $\mathbf{x}_1 \neq \mathbf{x}_2$,

$$f\big(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2\big) \leq \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2).$$

(v) **strictly convex** if $f\big(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2\big) < \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2)$.



(ii) + (iii) on $\mathbb{R}^d \Rightarrow f(\mathbf{x})$ is bounded from below and it reaches its infimum = **existence of a minimizer**

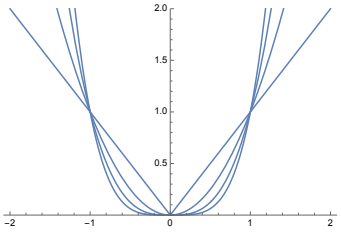
(iv) $\Rightarrow$ any local minimum is also global

(ii) + (iii) + (v) $\Rightarrow f(\mathbf{x})$ admits a **unique minimizer** over $\mathbb{R}^d$

Examples of convex functions

Set of admissible functions: $\Gamma_0(\mathbb{R}^d) = \left\{ f : \mathbb{R}^d \to \overline{\mathbb{R}} \text{ s.t. } f \text{ proper, convex and l.s.c. on } \mathbb{R}^d \right\}$

- Univariate functions
  - If f is differentiable: f is convex $\Leftrightarrow$ $f'(x) = \frac{\mathrm{d}f(x)}{\mathrm{d}x}$ is non-decreasing
  - If f is twice differentiable: f is convex $\Leftrightarrow$ $f''(x) = \frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} \geq 0$
  - Example: $\psi_p(x) = |x|^p$ is continuous (and a fortiori l.s.c.) convex for $p \geq 1$.
    It is strictly convex for $p > 1$.



- Sum of (weighted) convex functions
  - Let $f_1, f_2 \in \Gamma_0(\mathbb{R}^d)$. Then, $\alpha_1 f_1 + \alpha_2 f_2 \in \Gamma_0(\mathbb{R}^d)$ for any $\alpha_1, \alpha_2 \in \mathbb{R}^+$.
  - Example: $f_p(\mathbf{x}) = \|\mathbf{x}\|_p^p = \sum_{i=1}^{d} \psi_p(x_i) \in \Gamma_0(\mathbb{R}^d)$. $\|\mathbf{x}\|_p^p$ is strictly convex for $p > 1$.
- Indicator functions
  - Let $C \subseteq \mathbb{R}^d$ be a **closed convex subset** of $\mathbb{R}^d$. Then, the barrier function

$$i_C(\mathbf{x}) = \begin{cases} 0, & \text{if } \mathbf{x} \in C \\ +\infty, & \text{otherwise} \end{cases}$$ is l.s.c. and convex.

Proximity operator

(Moreau 1962)

Set of admissible functions: $\Gamma_0(\mathbb{R}^d) = \left\{ f : \mathbb{R}^d \to \overline{\mathbb{R}} \text{ s.t. } f \text{ proper, convex and l.s.c. on } \mathbb{R}^d \right\}$

Definition

The proximity operator of $f \in \Gamma_0(\mathbb{R}^d)$ is the (multivariate) function $\mathrm{prox}_f : \mathbb{R}^d \to \mathbb{R}^d$ defined by

$$\mathrm{prox}_f(\mathbf{x}) \triangleq \arg\min_{\mathbf{z} \in \mathbb{R}^d} \tfrac{1}{2}\|\mathbf{x} - \mathbf{z}\|^2 + f(\mathbf{z}).$$

NB: If $f \in \Gamma_0(\mathbb{R}^d)$, then the minimizer is guaranteed to exist and to be unique.

- Properties
  - Both $\mathbf{x} \mapsto \mathrm{prox}_f(\mathbf{x})$ and $\mathbf{x} \mapsto \mathbf{x} - \mathrm{prox}_f(\mathbf{x})$ are firmly non-expansive.
  - The fixed point of $\mathbf{x}^{t+1} = \mathrm{prox}_f(\mathbf{x}^t)$ is the minimizer of $f(\mathbf{x})$
- Projection on a convex set
  - Let i_C be the barrier function associated with the closed convex subset $C \subset \mathbb{R}^d$. Then,

$$\mathrm{prox}_{i_C}(\mathbf{x}) = \arg\min_{\mathbf{z} \in C} \|\mathbf{x} - \mathbf{z}\|_2 = \mathrm{Proj}_C(\mathbf{x})$$

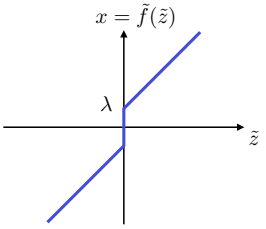
Proximity operator of $\|\cdot\|_{\ell_1}$-potential

Let $\Phi : \mathbb{R}^N \to \overline{\mathbb{R}}$ be a **separable additive** potential with $\Phi(\mathbf{x}) = \sum_{n=1}^{N} \phi_n(x_n)$ where $\phi_n \in \Gamma_0(\mathbb{R})$.

Then, $\Phi \in \Gamma_0(\mathbb{R}^N)$ (i.e., convex l.s.c.) and

$$\mathrm{prox}_{\Phi}(\mathbf{x}) = \begin{pmatrix} \mathrm{prox}_{\phi_1}(x_1) \\ \vdots \\ \mathrm{prox}_{\phi_N}(x_N) \end{pmatrix} \qquad \text{(layer of \textbf{pointwise nonlinearities})}$$

where $\mathrm{prox}_{\phi_n}(x) = \arg\min_{z\in\mathbb{R}} \left(\frac{1}{2}(x-z)^2 + \phi_n(z)\right)$.



- Special case: $\Phi(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$ with parameter $\lambda \in \mathbb{R}^+$ and $\phi_n(x) = \lambda|x|$.

$$g(x,z) = \tfrac{1}{2}(z-x)^2 + \lambda|z|$$

Optimum at $\tilde{z}$ with $\frac{\partial}{\partial z} g(x,\tilde{z}) = (\tilde{z}-x) + \lambda\,\mathrm{sign}(\tilde{z}) = 0$

$$\Rightarrow \quad x = \tilde{z} + \lambda\,\mathrm{sign}(\tilde{z}) = \tilde{f}(\tilde{z})$$

$$\Rightarrow \quad \tilde{z} = \mathrm{prox}_{\phi_n}(x) = \tilde{f}^{-1}(x) = \begin{cases} x-\lambda, & x > \lambda \\ 0, & x \in [-\lambda,\lambda] \\ x+\lambda, & x < -\lambda \end{cases} \qquad \text{(soft-threshold)}$$

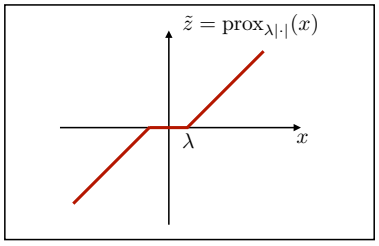
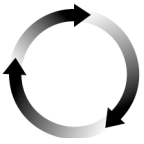


Image reconstruction under sparsity constraints $p = 1$

- Convex optimization problem with non-smooth regularization

(1) $\quad \mathbf{s}_{\mathrm{sparse}} = \arg\min_{\mathbf{s}} \left(\frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 + g(\mathbf{s})\right)$ with $\quad g(\mathbf{s}) = \lambda\|\mathbf{L}\mathbf{s}\|_{\ell_1}$ (regularization)

- Solution by forward-backward splitting (Combettes-Wajs, 2005)



Repeat

$$\mathbf{z}^{(n)} = \mathbf{s}^{(n-1)} + \gamma\left(\mathbf{H}^{\mathsf{T}}\mathbf{y} - \mathbf{H}^{\mathsf{T}}\mathbf{H}\mathbf{s}^{(n-1)}\right)$$

$$\mathbf{s}^{(n)} = \mathrm{prox}_{\gamma g}\left(\mathbf{z}^{(n)}\right)$$

until stop criterion

Guarantee of convergence: $\gamma \leq \frac{2}{\lambda_{\max}(\mathbf{H}^{\mathsf{T}}\mathbf{H})}$

Linear step (consistency with imaging physics)

Proximal step (regularization = prior constraints)

Proximal operator: $\quad \mathrm{prox}_g(\mathbf{z}) = \arg\min_{\mathbf{s}} \left(\frac{1}{2}\|\mathbf{z}-\mathbf{s}\|_2^2 + g(\mathbf{s})\right)$

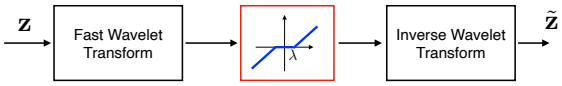
Interpretation: Same as (1) with $\mathbf{H} = \mathbf{I}$ $\quad \Rightarrow \quad$ "denoising" of current estimate $\mathbf{z}$

Efficient proximal denoising: wavelet-domain soft thresholding

- Regularization: Promote sparsity in an orthogonal wavelet basis

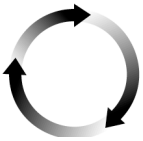
$g(\mathbf{s}) = \lambda \|\mathbf{W}^{\mathsf{T}}\mathbf{s}\|_{\ell_1}$ with $\mathbf{W}^{\mathsf{T}}\mathbf{W} = \mathbf{I}$ (Orthonormality)

Proximal step: $\tilde{\mathbf{z}} = \text{prox}_g(\mathbf{z}) = \arg\min_{\mathbf{s}} \left(\frac{1}{2}\|\mathbf{z}-\mathbf{s}\|_2^2 + \lambda\|\mathbf{W}^{\mathsf{T}}\mathbf{s}\|_{\ell_1}\right)$



- Iterative Soft-Thresholding Algorithm (ISTA) (Figueiredo-Nowak 2003)

Repeat



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

$$\mathbf{s}^{(n)} = \text{WaveletThreshold}\left(\mathbf{z}^{(n)}; \lambda/L\right)$$

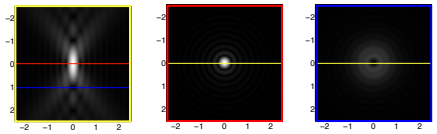
until stop criterion

$L \geq \lambda_{\max}(\mathbf{H}^{\mathsf{T}}\mathbf{H})$: Lipschitz constant of data term

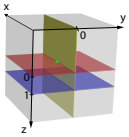
Deconvolution of fluorescence micrographs

- Physical model of a diffraction-limited microscope

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



3-D point spread function (PSF)



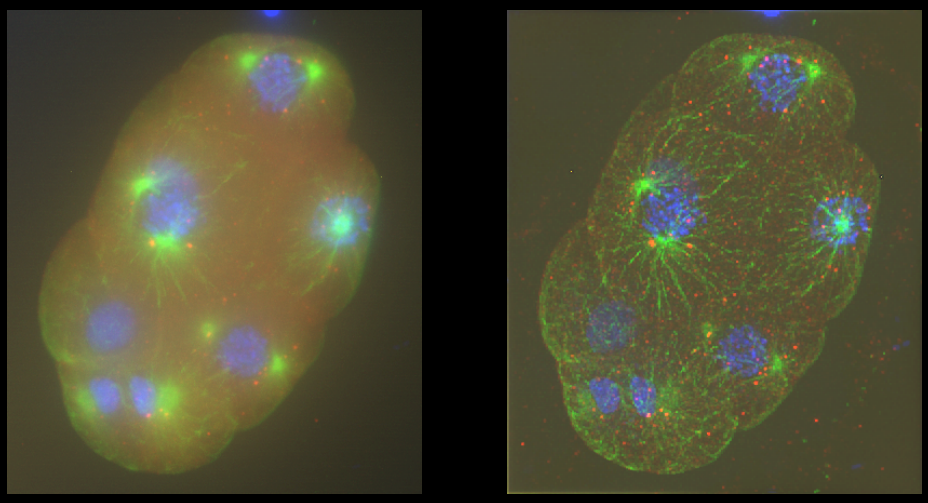
$$h_{3\text{D}}(x,y,z) = I_0 \left| p_\lambda\left(\tfrac{x}{M}, \tfrac{y}{M}, \tfrac{z}{M^2}\right)\right|^2$$

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

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
- $\text{NA} = n\sin\theta = R_0/f_0$: numerical aperture

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)

Extended penalized least-squares formulation

$$\text{(1)} \quad \mathbf{s}_\lambda = \arg\min_{\mathbf{s}\in\mathbb{R}^K} \left(\tfrac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 + \lambda \sum_{n=1}^{N} \Phi\big([\mathbf{L}\mathbf{s}]_n\big) \right)$$

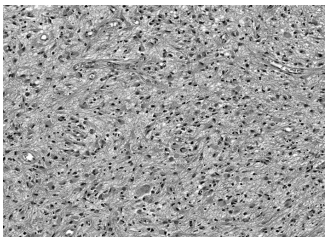
- System matrix: $\mathbf{H} \in \mathbb{R}^{M\times K}$ - Regularization operator: $\mathbf{L} \in \mathbb{R}^{N\times K}$ (not necessarily invertible)
- (Weakly-)convex potential $\Phi : \mathbb{R} \to \mathbb{R}$ (e.g., $\Phi(x) = |x|^p$)

- Bayesian interpretation: MAP estimator
  - i.i.d. Gaussian measurement noise: $\mathbf{n} = \mathbf{y} - \mathbf{s} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}) = \frac{1}{(\sqrt{2\pi}\sigma)^M} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{n}\|^2\right)$
  - Statistical signal model (prior): $p(\mathbf{s}) = \exp\left(-\frac{1}{2}\sum_{n=1}^{N} \Phi([\mathbf{L}\mathbf{s}]_n\right)$ (Gibbs distribution)

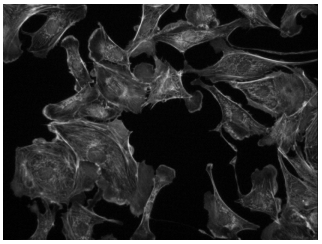
  $\Rightarrow$ Maximum a posteriori (MAP) estimator is given by (1) with $\lambda = \sigma^2$.

- Variants of ISTA
  - Inclusion of an inner loop for the iterative evaluation of $\mathrm{prox}_{J_{\mathrm{reg}}}(\mathbf{x})$.
  - Acceleration via the use of momentum: FISTA = Fast ISTA (Beck-Teboule 2009)
  - Alternating direction method of multipliers (ADMM) involving an auxiliary variable $\mathbf{u} = \mathbf{L}\mathbf{x}$ (Ramani-Fessler 2010)
    $\rightarrow$ alternation between some gradient-like updates and a separable prox_Φ on $\mathbf{u}$
  - Incorporation of positivity constraints (barrier functional); splitting, primal-dual formulation (Condat-Vu 2013)

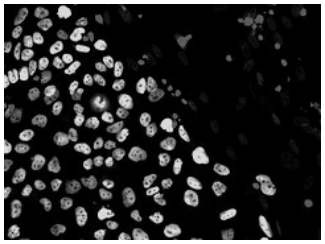
2D deconvolution experiment



Astrocytes cells



Bovine pulmonary artery cells



Human embryonic stem cells

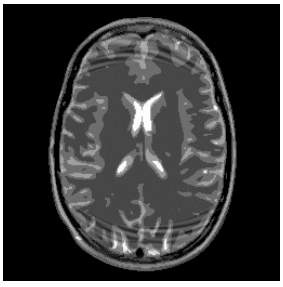
Disk-shaped PSF (7×7), L: gradient (TV-like), optimized parameters

Deconvolution results (SNR in dB)

	Tikhonov	l_1 prior (TV)	relaxed l_0 prior
Astrocytes cells	**12.18**	10.48	10.52
Pulmonary cells	16.9	**19.04**	18.34
Stem cells	15.81	20.19	**20.5**

MRI phantom: Spiral sampling in k-space

Original Phantom
(Guerquin-Kern TMI 2012)

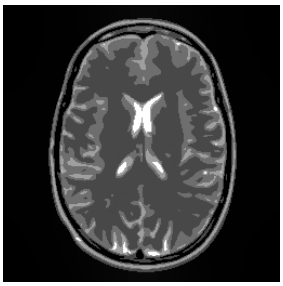


ℓ_2 prior (Tikhonov)
SER =17.69 dB

L : gradient

Optimized parameters

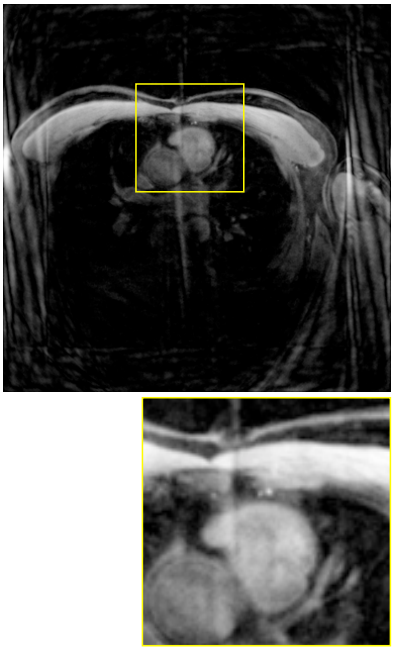
ℓ_1 prior (TV)
SER = 21.37 dB



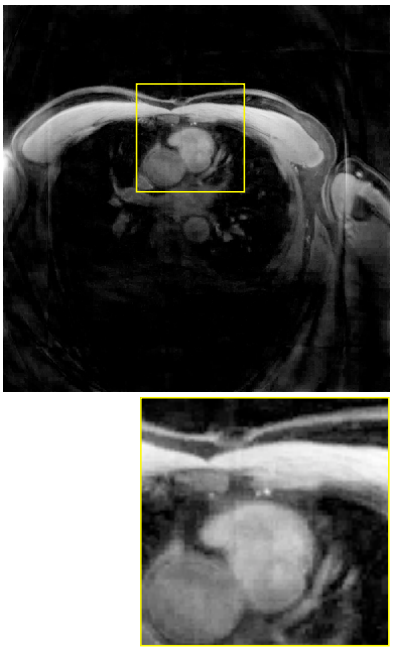
Relaxed ℓ_0 prior
SER = 27.22 dB

ISMRM reconstruction challenge

L_2 regularization (Laplacian)

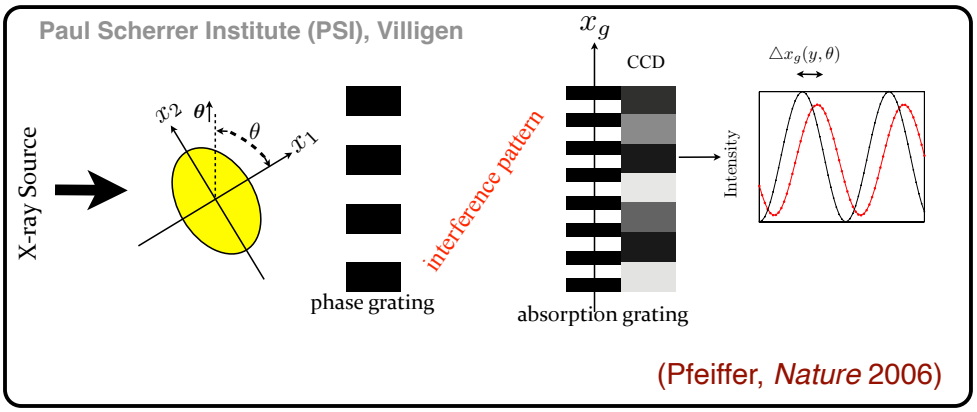


ℓ_1 wavelet regularization



(Guerquin-Kern *IEEE TMI* 2011)

Differential phase-contrast tomography



(Pfeiffer, *Nature* 2006)

Mathematical model

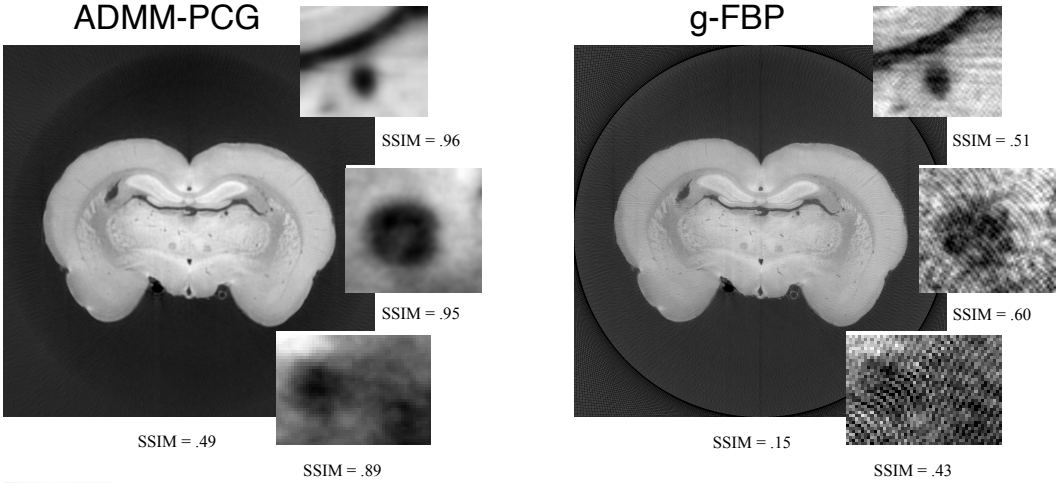
$$y(t,\theta) = \frac{\partial}{\partial t} \mathrm{R}_\theta\{s\}(t)$$

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

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

Reducing the numbers of views

Rat brain reconstruction with 181 projections

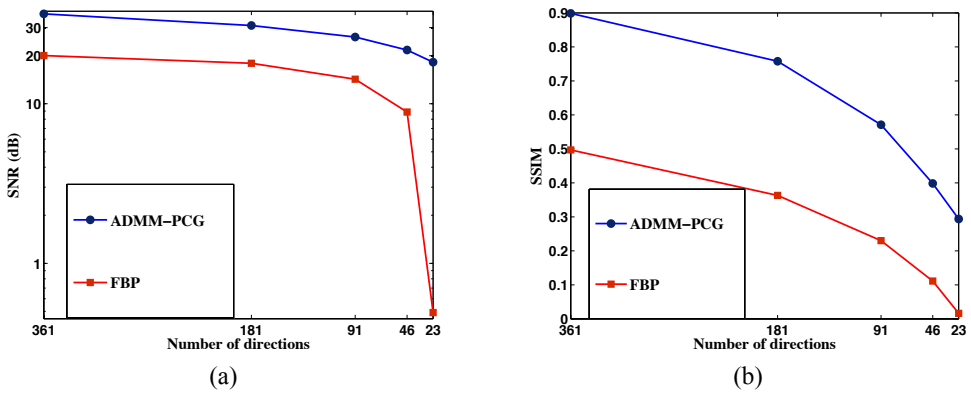


Collaboration: Prof. Marco Stampanoni, TOMCAT PSI / ETHZ

(Nichian et al. *Optics Express* 2013)

Performance evaluation

Goldstandard: high-quality iterative reconstruction with 721 views



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

Compressed sensing: Applications in imaging

- Magnetic resonance imaging (MRI) (Lustig, *Mag. Res. Im.* 2007)

GE Healthcare PHILIPS SIEMENS

- 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)

Part 4:

The (deep) learning revolution
⇒ Rise of data-driven methods

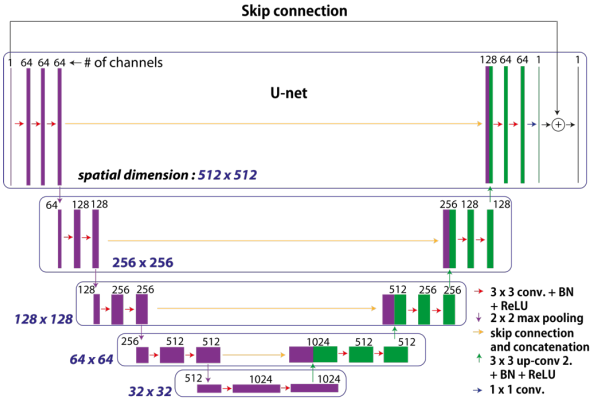
- First generation of **deep CNN**-based methods for image reconstruction
- **Iterative** reconstruction using **stable**—**and not so deep**—CNNs

Appearance of Deep ConvNets

(Jin et al. 2016; Adler-Öktem 2017; Chen et al. 2017; ...)

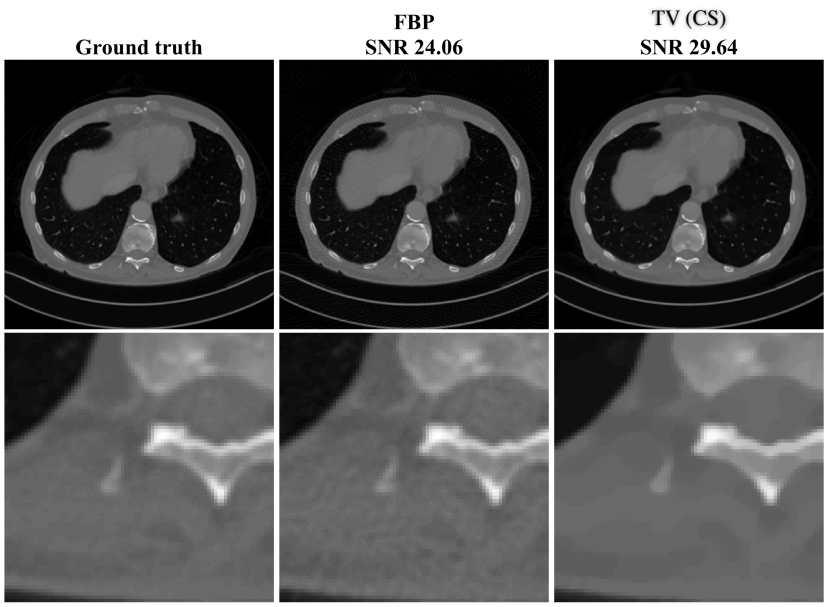
(Jin et al., IEEE TIP 2017)

- 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



CT data

Dose reduction by 7: 143 views

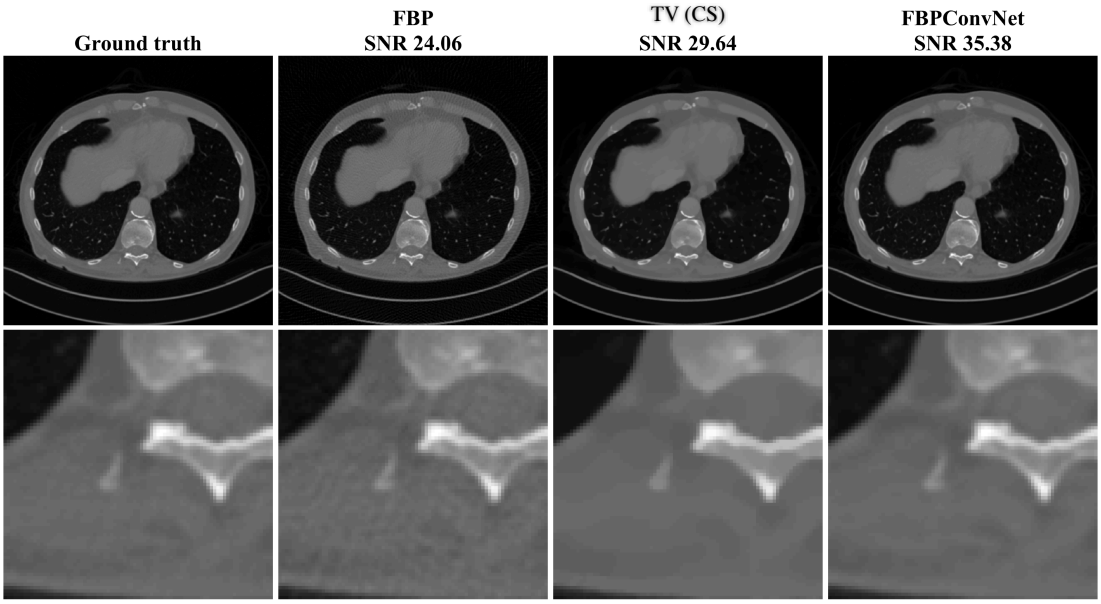


Reconstructed from
from 1000 views

MAYO CLINIC

CT data

Dose reduction by 7: 143 views



Reconstructed from
from 1000 views

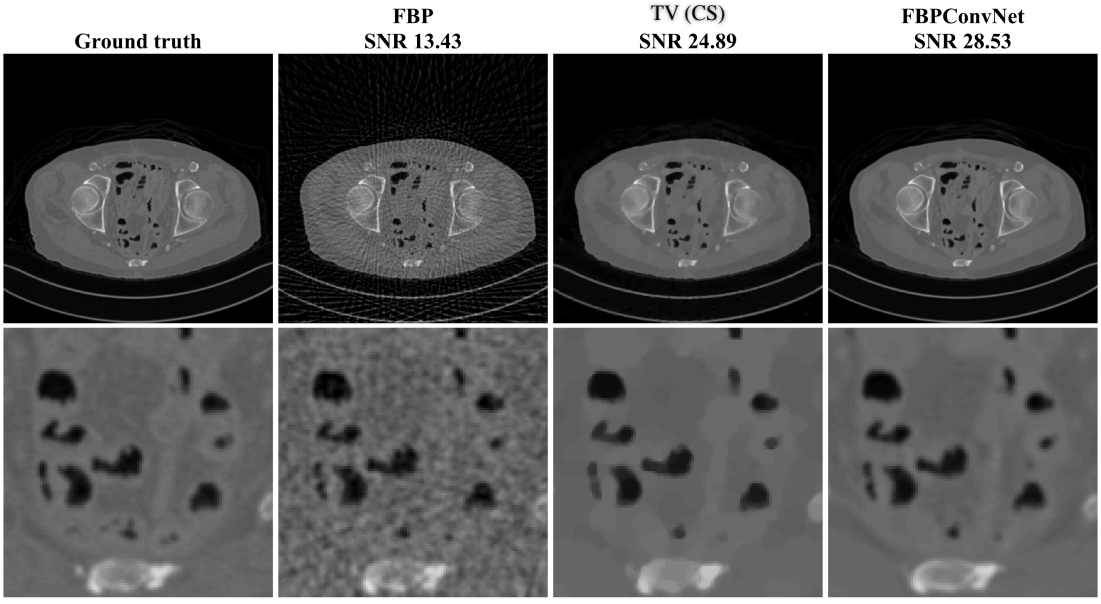
MAYO CLINIC

(Jin et al, *IEEE Trans. Im Proc.*, 2017)

IEEE
2019 Best Paper Award

CT data

Dose reduction by 20: 50 views



Reconstructed from
from 1000 views

MAYO CLINIC

(Jin et al., *IEEE Trans. Im Proc.*, 2017)

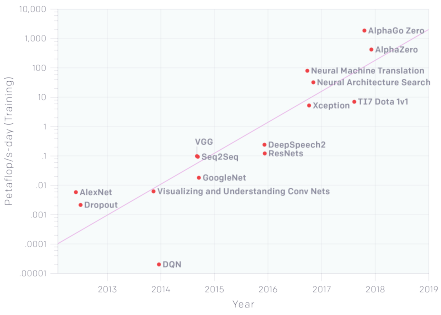
Deep CNNs for bioimage reconstruction images

- X-ray tomography	(Jin···Unser, *IEEE TIP* 2017) (Chen···Wang, *Biomed Opt. Exp* 2017)
- Magnetic resonance imaging (MRI)	(Hammernik···Pock, *Mag Res Med* 2018) (Tezcan···Konukoglu, *IEEE TMI* 2018)
- Dynamic MRI (cardial imaging)	(Schlemper···Rueckert, *IEEE TMI 2018*) (Hauptmann···Arridge, *Mag Res Med* 2019)
- 2D microscopy	(Rivenson···Ozcan, *Optica* 2017)
- 3D fluorescence microscocopy	(Weigert···Jug, Myers, *Nature Meth. 2018*)
- Super-resolution microscopy	(Nehme···Shechtman, *Optica* 2018)
- Diffraction tomography	(Sun···Kamilov, *Optics Express* 2018)
- Ultrasound	(Yoon···Ye, *IEEE TMI* 2019)

But CNN-based methods also have their weaknesses

- They require **lots of training data**
  - Medical imaging: limited access to patient data
  - Lack of gold standards (except for compressed sensing scenarios)
  - Training for (3D) medical imaging is **extremely computer intensive**

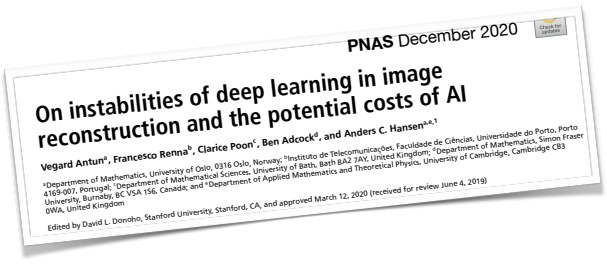
AI Moore's law: doubling every 3.4 month !



source: openAI

- They are **hard to tune**
  - Many design parameters: depth, width, number of channels
  - Use of ad hoc modules: batch normalization

- They **lack robustness**
  - Adversarial attacks
  - Unpredictable results

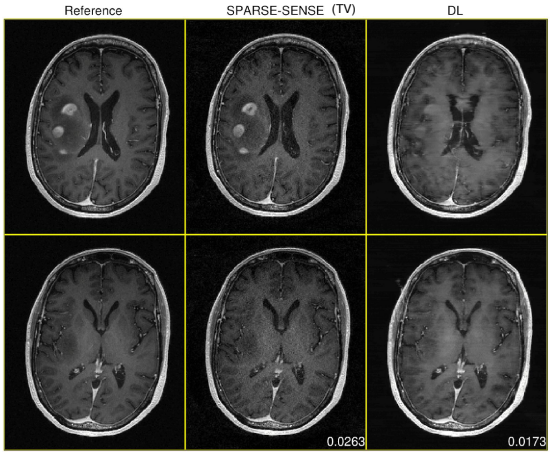

PNAS December 2020

On instabilities of deep learning in image reconstruction and the potential costs of AI

Vegard Antun[a], Francesco Renna[b], Clarice Poon[c], Ben Adcock[d], and Anders C. Hansen[a,e,1]

[a]Department of Mathematics, University of Oslo, 0316 Oslo, Norway; [b]Instituto de Telecomunicações, Faculdade de Ciências, Universidade do Porto, Porto 4169-007, Portugal; [c]Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, United Kingdom; [d]Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada; and [e]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, United Kingdom

Edited by David L. Donoho, Stanford University, Stanford, CA, and approved March 12, 2020 (received for review June 4, 2019)



(variant of FBPConvNet)

Figure 3: Reconstructions in a case of anaplastic astrocytoma, a rare malignant brain tumor. SPARSE-SENSE and DL reconstructions are from the same 4x-accelerated retrospectively undersampled acquisition. DL achieves lower whole-volume MAE than SPARSE-SENSE, but fails to properly reconstruct regions near the tumor.

G. Nataraj and R. Otazo. "Investigating robustness to unseen pathologies in model-free deep multicoil reconstruction." ISMRM 2020 Workshop on Data Sampling & Image Reconstruction

PnP-FBS variant of Iterative Image Reconstruction

Schematic structure of iterative reconstruction algorithm : $\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left(\frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + g(\mathbf{x}) \right)$

Repeat (N_{iter})

$\mathbf{z}^{(n)} = \mathbf{x}^{(n-1)} + \alpha \left(\mathbf{H}^{\mathsf{T}}\mathbf{y} - \mathbf{H}^{\mathsf{T}}\mathbf{H}\mathbf{x}^{(n-1)} \right)$ Linear step (consistency with imaging physics)

$\mathbf{x}^{(n)} = \text{prox}_{\alpha g}\left(\mathbf{z}^{(n)}\right)$ Proximal or "denoising" step (regularization)

until stop criterion

Proximal operator: $\text{prox}_{g}\left(\mathbf{z}\right) = \arg\min_{\mathbf{x}} \left(\frac{1}{2} \|\mathbf{z} - \mathbf{x}\|^2 + g(\mathbf{x}) \right)$

Plug-and-Play variant (Venkatakrishnan-Bouman 2013)

Repeat (N_{iter})

$\mathbf{z}^{(n)} = \mathbf{x}^{(n-1)} + \alpha \left(\mathbf{H}^{\mathsf{T}}\mathbf{y} - \mathbf{H}^{\mathsf{T}}\mathbf{H}\mathbf{x}^{(n-1)} \right)$ Linear step (consistency with imaging physics)

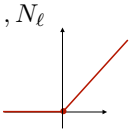
$\mathbf{x}^{(n)} = \left((1-\beta)\text{Id} + \beta f_{\boldsymbol{\theta}} \right)\left(\mathbf{z}^{(n)}\right)$ Suitable nonlinear map (e.g., CNN)

until stop criterion

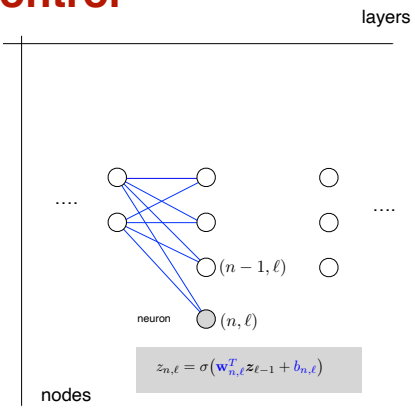
Requirement for convergence: $\|f_{\boldsymbol{\theta}}\|_{\text{Lip}} \leq 1$ (Non-expansive operator) (Bauschke-Combettes 2017, Hertrich et al. 2021)

Deep neural network with stability control

- Layers: $\ell = 1, \ldots, L$
- Deep structure descriptor: $(N_0, N_1, \cdots, N_L)$
- Neuron or node index: $(n, \ell), \quad n = 1, \cdots, N_\ell$
- Activation function: $\sigma : \mathbb{R} \to \mathbb{R}$ (ReLU)



- Linear step: $\mathbb{R}^{N_{\ell-1}} \to \mathbb{R}^{N_\ell}$
  $\boldsymbol{f}_\ell : \boldsymbol{x} \mapsto \boldsymbol{f}_\ell(\boldsymbol{x}) = \mathbf{W}_\ell \boldsymbol{x} + \mathbf{b}_\ell$
- Nonlinear step: $\mathbb{R}^{N_\ell} \to \mathbb{R}^{N_\ell}$
  $\boldsymbol{\sigma}_\ell : \boldsymbol{x} \mapsto \boldsymbol{\sigma}_\ell(\boldsymbol{x}) = \big(\sigma(x_1), \ldots, \sigma(x_{N_\ell})\big)$



Learned

$$\boldsymbol{f}_{\text{deep}}(\boldsymbol{x}) = (\boldsymbol{\sigma}_L \circ \boldsymbol{f}_L \circ \boldsymbol{\sigma}_{L-1} \circ \cdots \circ \boldsymbol{\sigma}_2 \circ \boldsymbol{f}_2 \circ \boldsymbol{\sigma}_1 \circ \boldsymbol{f}_1)(\boldsymbol{x})$$

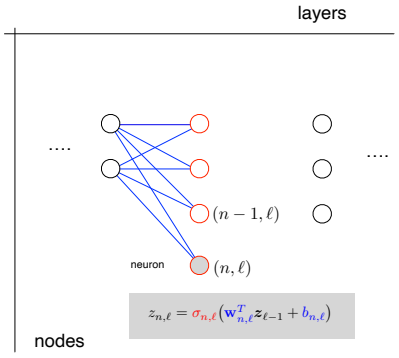
Controlling Lipschitz stability

$$\|\boldsymbol{f}_{\text{deep}}\|_{\text{Lip}} = \sup_{\boldsymbol{x} \neq \boldsymbol{x}'} \frac{\|\boldsymbol{f}_{\text{deep}}(\boldsymbol{x}) - \boldsymbol{f}_{\text{deep}}(\boldsymbol{x}')\|}{\|\boldsymbol{x} - \boldsymbol{x}'\|} \leq \prod_{\ell=1}^{L} \underbrace{\|\boldsymbol{\sigma}_\ell\|_{\text{Lip}}}_{1} \underbrace{\|\mathbf{W}_\ell\|_{\mathcal{S}_\infty}}_{=1}$$

Spectral normalization (Miyato ICLR 2018)

Neural nets with free-form activations and stability

- Layers: $\ell = 1, \ldots, L$
- Deep structure descriptor: $(N_0, N_1, \cdots, N_L)$
- Neuron or node index: $(n, \ell), \quad n = 1, \cdots, N_\ell$
- Activation function $\sigma_{n,\ell} : \mathbb{R} \to \mathbb{R}$ (free-form)
- Linear step: $\mathbb{R}^{N_{\ell-1}} \to \mathbb{R}^{N_\ell}$
  $\boldsymbol{f}_\ell : \boldsymbol{x} \mapsto \boldsymbol{f}_\ell(\boldsymbol{x}) = \mathbf{W}_\ell \boldsymbol{x} + \mathbf{b}_\ell$
- Nonlinear step: $\mathbb{R}^{N_\ell} \to \mathbb{R}^{N_\ell}$
  $\boldsymbol{\sigma}_\ell : \boldsymbol{x} \mapsto \boldsymbol{\sigma}_\ell(\boldsymbol{x}) = \big(\sigma_{n,\ell}(x_1), \ldots, \sigma_{N_\ell,\ell}(x_{N_\ell})\big)$



$$\boldsymbol{f}_{\text{deep}}(\boldsymbol{x}) = (\boldsymbol{\sigma}_L \circ \boldsymbol{f}_L \circ \boldsymbol{\sigma}_{L-1} \circ \cdots \circ \boldsymbol{\sigma}_2 \circ \boldsymbol{f}_2 \circ \boldsymbol{\sigma}_1 \circ \boldsymbol{f}_1)(\boldsymbol{x})$$

Joint learning / training

Stability control: $\|\boldsymbol{f}_{\text{deep}}\|_{\text{Lip}} \leq \prod_{\ell=1}^{L} \underbrace{\|\boldsymbol{\sigma}_\ell\|_{\text{Lip}}}_{1} \underbrace{\|\mathbf{W}_\ell\|_{\mathcal{S}_\infty}}_{1} = 1$

Lip-1 splines

spectral normalization vs. Parseval frame

Learning activation functions / pointwise

Finding the "optimal" pointwise nonlinearity $\sigma : \mathbb{R} \to \mathbb{R}$

Infinite-dimensional optimization problem is that is inherently ill-posed

- Incorporating a **regularization**
  - Should not penalize simple solutions (e.g., identity or linear scaling)
  - Should impose diffentiability (for DNN to be trainable via backpropagation)
  - Should favor simplest CPWL solutions; i.e., with "sparse 2nd derivatives"

  $\Rightarrow$ minimizing/constraining $\mathrm{TV}^{(2)}(\sigma) \triangleq \|\mathrm{D}^2\sigma\|_{\mathcal{M}}$ (Second-order total-variation)

- Controlling **stability**: $\mathrm{Lip}(\sigma) \triangleq \sup_{x\in\mathbb{R}} |\mathrm{D}\sigma(x)| \leq 1$

- Search space: $\mathrm{BV}^{(2)}(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} : \|\mathrm{D}^2 f\|_{\mathcal{M}} < \infty\} \subset \mathrm{Lip}(\mathbb{R})$

TV(2) regularization with slope constraints

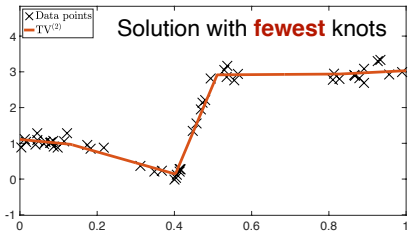
Generic loss functional $E : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ (strictly convex) Slope parameters: $s_{\min} < s_{\max}$

(TV2-SC) $$S = \arg \min_{f\in \mathrm{BV}^{(2)}(\mathbb{R})} \left(\sum_{m=1}^{M} \mathrm{E}(f(x_m), y_m) + \lambda \mathrm{TV}^{(2)}(f) \right),$$
$$\text{s.t.} \quad s_{\min} \leq f'(x) \leq s_{\max}, \quad \forall x \in \mathbb{R}$$

Theorem (new improved: for Stéphane Mallat's birthday)

The solution set of (TV2-SC) is a non-empty, weak*-compact subset of $\mathrm{BV}^{(2)}(\mathbb{R})$, and **all its extreme points** are **adaptive piecewise-linear splines** with a most $(M-2)$ knots.

The **sparsest spline** solution is identifiable using a variant of Debarre's algorithm.



(Debarre JCAM 2022)

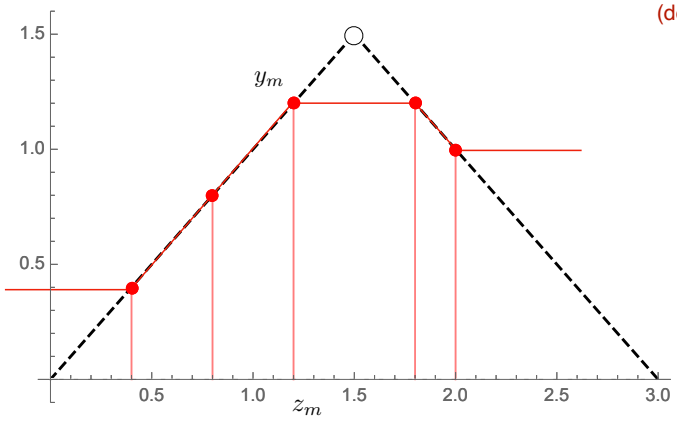
- Special cases of $(s_{\min}, s_{\max})$
  - $(-1, 1)$: Lipschitz-1 splines (Aziznejad, IEEE OJSP 2022)
  - $(0, 1)$: firmly non-expansive = prox of a convex potential
  - $(0, +\infty)$: monotone splines = derivative of a convex potential
  - $(-\rho, +\infty)$ with $0 < \rho$ (small): weakly-monotone splines = derivative of a ρ-weakly-convex potential

$(s_{\min}, s_{\max}) = \mathbb{R}$ (unconstrained)

Comparison of linear interpolators

$$\arg\min_{f\in H^1(\mathbb{R})} \int_{\mathbb{R}} |\mathrm{D}f(x)|^2 \mathrm{d}x \quad \text{s.t.} \quad f(x_m) = y_m,\ m = 1, \ldots, M$$

(de Boor 1966)



$\mathcal{M}$-norm = weak extension of L_1 norm:

$$\|f\|_{\mathcal{M}} = \sup_{\varphi\in\mathcal{S}(\mathbb{R}):\ \|\varphi\|_{L_\infty}\leq 1} \langle f, \varphi\rangle$$

$$\arg\min_{f\in \mathrm{BV}^{(2)}(\mathbb{R})} \|\mathrm{D}^2 f\|_{\mathcal{M}} \quad \text{s.t.} \quad f(x_m) = y_m,\ \ m = 1, \ldots, M$$

(U. JMLR 2019; Lemma 2)

Representer theorem for stable, free-form deep neural

Theorem (Optimality of Lipschitz-1 deep spline networks) (Unser et al., ACHA 2025)

- neural network $\mathbf{f} : \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$ with **deep structure** $(N_0, N_1, \ldots, N_L)$
  $$\boldsymbol{x} \mapsto \boldsymbol{f}_{\mathrm{deep}}(\boldsymbol{x}) = (\boldsymbol{\sigma}_L \circ \boldsymbol{f}_L \circ \boldsymbol{\sigma}_{L-1} \circ \cdots \circ \boldsymbol{f}_2 \circ \boldsymbol{\sigma}_1 \circ \boldsymbol{f}_1)(\boldsymbol{x})$$
- linear transformations $\boldsymbol{f}_\ell : \mathbb{R}^{N_{\ell-1}} \to \mathbb{R}^{N_\ell}, \boldsymbol{x} \mapsto \mathbf{W}_\ell \boldsymbol{x}$ with $\mathbf{W}_\ell \in \mathbb{R}^{N_\ell \times N_{\ell-1}}$
- **free-form** activations $\boldsymbol{\sigma}_\ell = (\sigma_{1,\ell}, \ldots, \sigma_{N_\ell,\ell}) : \mathbb{R}^{N_\ell} \to \mathbb{R}^{N_\ell}$ with $\sigma_{1,\ell}, \ldots, \sigma_{N_\ell,\ell} \in \mathrm{BV}^{(2)}(\mathbb{R})$

Given a series data points $(\boldsymbol{x}_m, \boldsymbol{y}_m)$ $m = 1, \ldots, M$, we then define the training problem

$$\arg\min_{(\mathbf{W}_\ell),(\sigma_{n,\ell}\in \mathrm{BV}^{(2)}(\mathbb{R}))} \left(\sum_{m=1}^{M} E\big(\boldsymbol{y}_m, \boldsymbol{f}_{\mathrm{deep}}(\boldsymbol{x}_m)\big) + \lambda \sum_{\ell=1}^{L} \sum_{n=1}^{N_\ell} \mathrm{TV}^{(2)}(\sigma_{n,\ell}) \right)$$
$$\text{s.t.} \quad \mathrm{Lip}(\sigma_{n,\ell}), \|\mathbf{W}_\ell\|_{\mathcal{S}_\infty} \leq 1, \quad (n = 1, \ldots, N_\ell,\ \ \ell = 1, \cdots, L) \qquad (1)$$

$$\Rightarrow \mathrm{Lip}(\boldsymbol{f}_{\mathrm{deep}}) \leq 1$$

where $E : \mathbb{R}^{N_L} \times \mathbb{R}^{N_L} \to \mathbb{R}^+$ is an arbitrary convex loss function.

The solution of (1) exists and is achieved by a **deep spline network** with activations of the form

$$\sigma_{n,\ell}(x) = b_{1,n,\ell} + b_{2,n,\ell} x + \sum_{k=1}^{K_{n,\ell}} a_{k,n,\ell}\, \mathrm{ReLU}(x - \tau_{k,n,\ell}),$$

with adaptive parameters $K_{n,\ell} \leq M - 2$, $\tau_{1,n,\ell}, \ldots, \tau_{K_{n,\ell},n,\ell} \in \mathbb{R}$, and $b_{1,n,\ell}, b_{2,n,\ell}, a_{1,n,\ell}, \ldots, a_{K_{n,\ell},n,\ell} \in \mathbb{R}$.

Precursor without stability: (Unser, JMLR 2019)

Outcome of representer theorem: deep splines

Each neuron $\big($fixed index $(n, \ell)\big)$ is a **piecewise-linear spline**, characterized by

- its number $0 \leq K_{n,\ell}$ of knots (ideally, much smaller than M);
- the location $\{\tau_k = \tau_{k,n,\ell}\}_{k=1}^{K_{n,\ell}}$ of these knots (ReLU biases);
- the expansion coefficients $\mathbf{b}_{n,\ell} = (b_{1,n,\ell}, b_{2,n,\ell}) \in \mathbb{R}^2$, $\boldsymbol{a}_{n,\ell} = (a_{1,n,\ell}, \ldots, a_{K,n,\ell}) \in \mathbb{R}^K$.

These parameters (including the number of knots) are **data-dependent** and adjusted automatically during training.

- Link with ℓ_1 minimization techniques

$$\mathrm{TV}^{(2)}(\sigma_{n,\ell}) = \sum_{k=1}^{K_{n,\ell}} |a_{k,n,\ell}| = \|\mathbf{a}_{n,\ell}\|_1 \qquad \text{and} \qquad \mathrm{Lip}(\sigma_{n,\ell}) = \sup_{K \in \{1,\ldots,K_{n,\ell}\}} \left| \sum_{k=1}^{K} a_{k,n,\ell} \right|$$

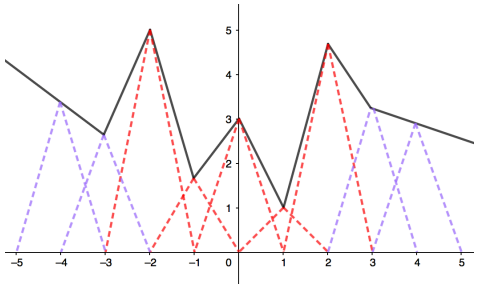
Implementation: Lip-1 spline CNN (trained for

$$\boldsymbol{f}_{\mathrm{deep}}(\boldsymbol{x}) = (\boldsymbol{\sigma}_L \circ \boldsymbol{f}_L \circ \boldsymbol{\sigma}_{L-1} \circ \cdots \circ \boldsymbol{\sigma}_2 \circ \boldsymbol{f}_2 \circ \boldsymbol{\sigma}_1 \circ \boldsymbol{f}_1)(\boldsymbol{x})$$

Learnable linear spline nonlinearities (pointwise)

Convolutional layer

$\sigma_{n,\ell}(x)$



Knot spacing: T, Number of knots: K

- Linear B-spline basis
  - Compact support
  - Efficient forward & backward pass
  - Easy to compute Lipschitz constant (max. absolute derivative)

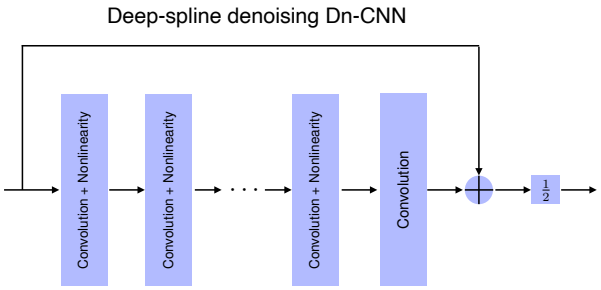
(Bohra et al. *IEEE Open JSP* 2020)

- Constrain Lipschitz constant of each layer to be no greater than one
  - Convolutional layer: Lip-1 projector (spectral normalization vs. Parseval frame)
  - Linear spline layer: Lip-1 spline projector

(Ducotterd et al. JMLR 2024)

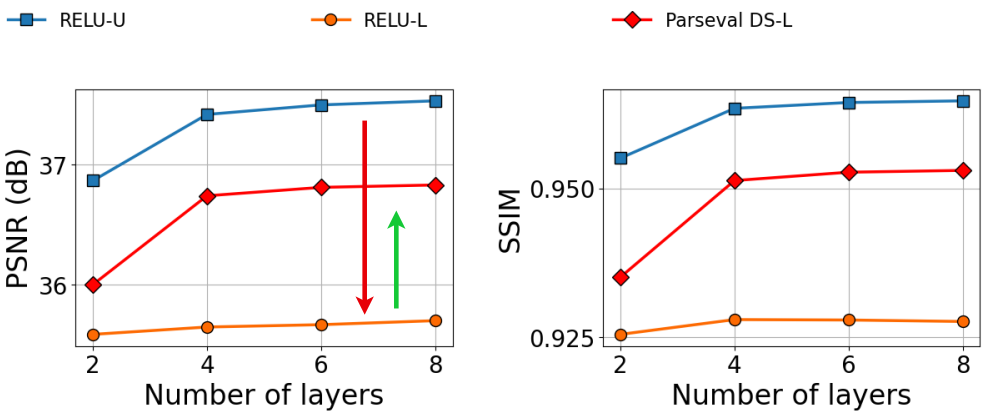
PnP image reconstruction: Experimental set-up

- Training of Gaussian denoiser
  - 240K examples of 40×40 patches from BSD500 dataset
  - Additive Gaussian noise with $\sigma = 5/255$
  - 3×3 convolution kernels, 32 channels
  - Deep spline activations with $T = 0.1$, $K = 51$
  - Number of layers = 3, 5, 7, 9
- Compressed sensing MRI
  - 256×256 ground-truth images
  - Subsampling ratio = 0.3
  - Gaussian additive noise with $\sigma = 10/255$
  - Number of layers of denoising CNN = 5



Learned Lip-1 filters = Parseval CNN

Results: Gaussian denoising with Parseval CNN

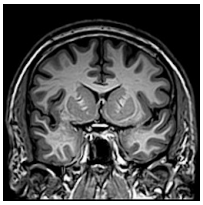


- Drop in performance for constrained ReLU nets
- DS-L performs better than ReLU-L even with fewer parameters

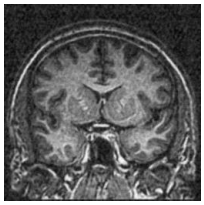
Compressed Sensing MRI

Subsampling mask	Random		Radial		Cartesian	
Image type	Brain	Bust	Brain	Bust	Brain	Bust
Zero-filling	23.72	25.88	22.99	23.92	21.34	23.03
ReLU-L	30.70	30.59	29.60	30.09	23.70	26.87
Parseval DS-L	33.19	33.88	31.68	33.15	24.97	28.68

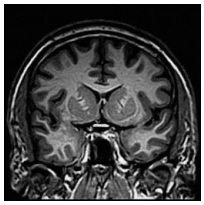
Random sampling pattern



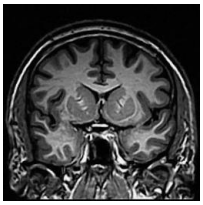
Ground truth



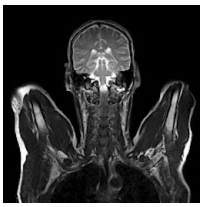
Zero-filled reconstruction



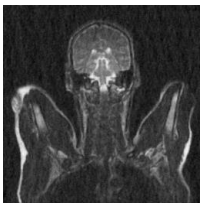
ReLU-L



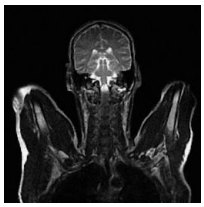
Parseval DS-L



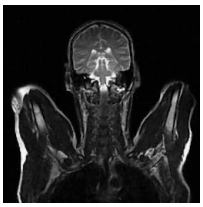
Ground truth



Zero-filled reconstruction



ReLU-L



Parseval DS-L

WCRR variant: Learnable Weakly-Convex Ridge

$$\min_{\mathbf{x}\in\mathbb{R}^N} \left(\frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 + \sum_{i=1}^{I_{\text{chan}}} \langle \mathbf{1}, \boldsymbol{\Phi}_i(\mathbf{W}_i\mathbf{x})\rangle \right)$$

Weakly-convex extension of FoE (Chen-Pock 2014)

- System matrix: $\mathbf{H} \in \mathbb{R}^{M\times N}$
- Learnable filters (CNN) : $\mathbf{W}_i \in \mathbb{R}^{N\times N}, \quad i = 1,\dots,I_{\text{chan}}$
- Shared free-form potentials : $\boldsymbol{\Phi}_i(\mathbf{u}) = \big(\Phi_i(u_1),\dots,\Phi_i(u_N)\big)$ with $\Phi_i(u) = \int_{-\infty}^{u} \phi_i(x)\mathrm{d}x$

- Iterative reconstruction

  Recurrent neural network (steepest descent)

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} - \alpha \left(\sum_{i=1}^{I_{\text{chan}}} \mathbf{W}_i^{\mathsf{T}} \boldsymbol{\phi}_i(\mathbf{W}_i\mathbf{x}^{(n)}) + \mathbf{H}^{\mathsf{T}}\big(\mathbf{H}\mathbf{x}^{(n)} - \mathbf{y}\big) \right) \quad \text{with} \quad \boldsymbol{\phi}_i = \boldsymbol{\Phi}_i'$$

(Goujon, SIAM J. Im. Sci 2025)

- Training on denoising problem
  - Parametrization of the slope: $\phi_i = \Phi_i' : \mathbb{R} \to \mathbb{R}$
    s.t. weak-monotonicity constraint and penalty on $\mathrm{TV}^{(2)}(\phi_i)$ (sparsity) $\Rightarrow$ **linear splines**
  - Deep equilibrium training of variational denoiser where the ϕ_i are expanded in a B-spline basis.

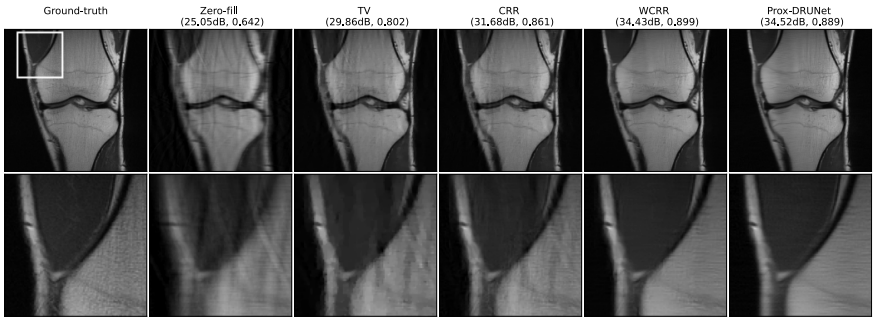
Table 4.1
PSNR and SSIM values for both reconstruction experiments.

Metric	PSNR	SSIM
Zero-fill	27.92	0.711
TV[5]	32.03	0.7922
CRR-NN [19]	33.14	0.842
WCRR-NN	34.55	0.858
Prox-DRUNet [23]	35.09	0.864

(a) MRI

Metric	PSNR	SSIM	Param.
TV	31.57	0.852	1
ACR [37]	31.58	0.848	$6 \cdot 10^5$
CRR-NN	32.87	0.862	$5 \cdot 10^3$
AR [34]	33.62	0.875	$2 \cdot 10^7$
WCRR-NN	34.06	0.895	$2 \cdot 10^4$
Prox-DRUNet	34.20	0.901	$2 \cdot 10^7$

(b) CT



10-81

but, PSNR (or SSIM) is not the whole story

Theoretical guarantees : convergence, consistency, stability

convex handcrafted | convex learned | weakly convex learned | "state-of-the-art" CNN

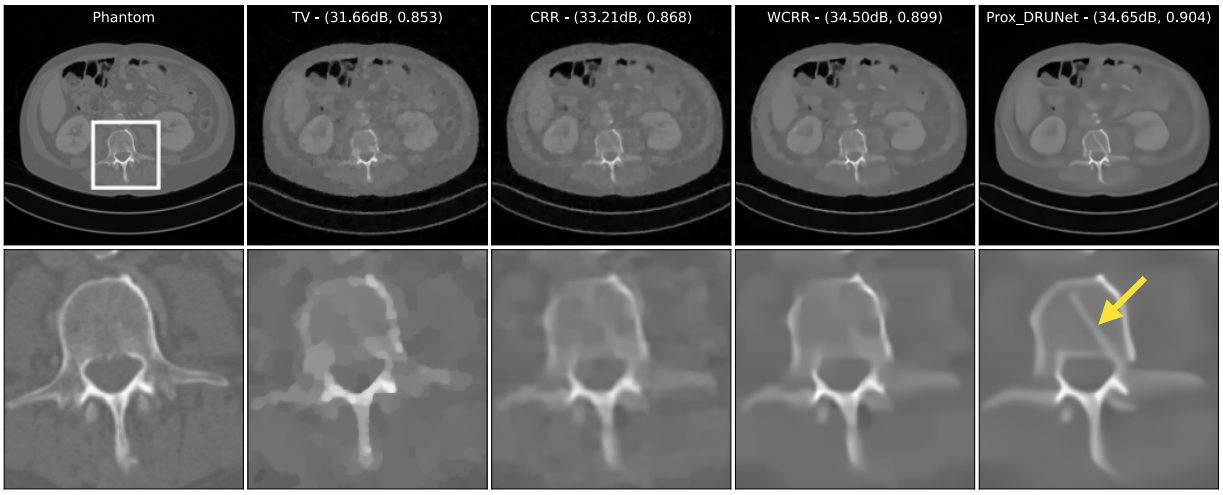
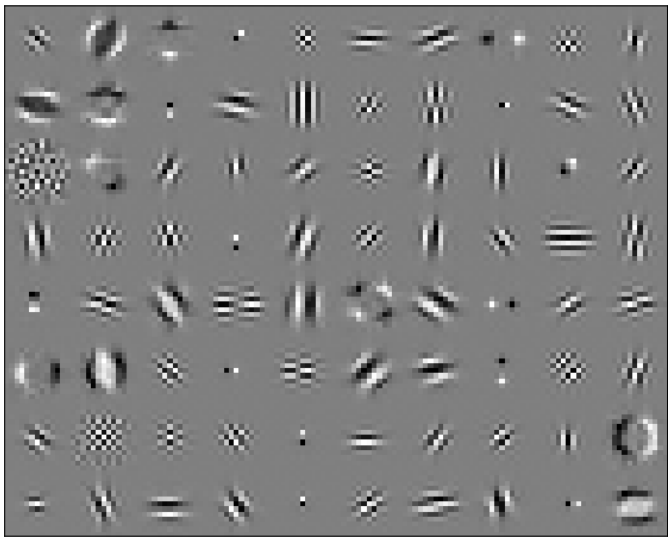


Figure 4.2. *Reconstructions for the sparse-view CT experiment. The reported metrics are PSNR and SSIM.*

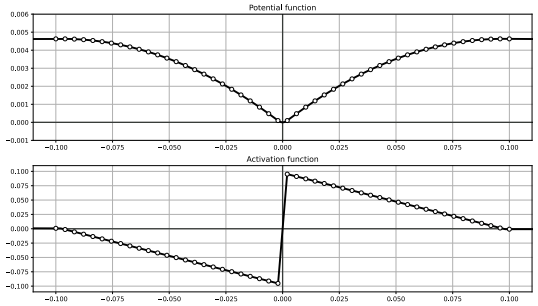
10-82

Learned filters and nonlinearities

80 channels



Un panorama d'ondelettes



Nonlinearities are shared up to a channel-wise scaling factor

et une "spline" bien dimensionnée

Conclusion: Current status of computational imaging

- Classical reconstruction algorithms
  - Are typically linear and have a fast implementation (e.g. filtered backprojection)
  - Can be derived from the minimization of a quadratic cost functional
  - Use regularization to deal with ill-posedness and to avoid amplification of noise
  - Are well understood and come with theoretical guarantees (stability, etc.)
- 2nd generation methods / compressed sensing
  - Derived from the minimization of a convex (sparsity-promoting) cost functional
  - Iterative reconstruction (Gradient-based or proximal)
  - Can handle more extreme scenarios (lower dose, less measurements)
- Work in progress: The learning revolution
  - Amazing reconstruction results / state-of-the-art in current challenges
  - Still poorly understood, requires lots of training
  - Lack of robustness, tendency to hallucinate

Stability/trust vs. performance ?

Conclusion: Cont'd

- What benefits all approaches
  - Deep understanding of the imaging physics $\Rightarrow$ accurate forward model
  - Proper discretization via the use of appropriate basis functions
  - Fast, matrix-free implementation of forward model/normal matrix (whenever possible)

- How the newer CNN-based methods profit from the older ones
  - CNN variant of "classical" PnP iterative reconstruction pipeline
  - Enforcing stability (Lip-1) to guarantee convergence
  - Producing reconstruction that are **consistent** with the measurements
  - Search for shallower architecture for faster training and inference

References

- Foundations
  - M.T. McCann, M. Unser, **Biomedical Image Reconstruction: From the Foundations to Deep Neural Networks**, *Foundations and Trends in Signal Processing*, vol. 13, no. 3, pp. 280-359, December 2019.

- Examples of algorithm design and imaging modalities
  - 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, 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.

- A glimpse at on-going research: Deep spline framework
  - M. Unser, "A Representer Theorem for Deep Neural Networks," *Journal of Machine Learning Research*, vol. 20, no. 110, pp. 1-30, 2019.
  - P. Bohra, J. Campos, H. Gupta, S. Aziznejad, M. Unser, "Learning Activation Functions in Deep (Spline) Neural Networks," *IEEE Open Journal of Signal Processing*, Vol. 1 , pp. 295-309, 2020.
  - A. Goujon, S. Neumayer, P. Bohra, S. Ducotterd, M. Unser, "A Neural-Network-Based Convex Regularizer for Image Reconstruction," *IEEE Transactions on Computational Imaging*, vol. 9, pp. 781–795, 2023.