European Molecular Imaging Meeting | EMIM 2020

24-28 August – virtual edition

Declaration of Financial Interests or Relationships

I have **no financial interests** or **relationships** to disclose with regard to the subject matter of this presentation.

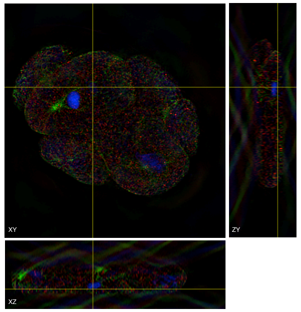
EPFL

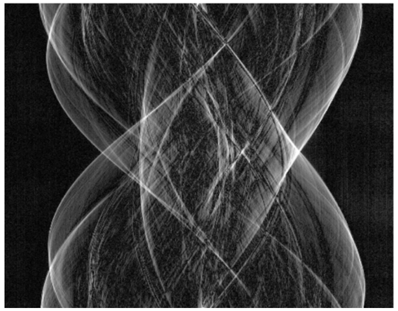
CIBM Center for Biomedical Imaging

Biomedical imaging as an inverse problem

Pol del Aguila Pla, Ph.D.
Center for Biomedical Imaging
Mathematical Imaging Section
Switzerland

Biomedical Imaging Group
EPFL, Lausanne, Switzerland





Slides mostly based on the work of Prof. Michaël Unser and the EPFL's Biomedical Imaging Group, in Lausanne, Switzerland.

Biomedical imaging as an inverse problem

Biomedically relevant quantity $s \mapsto \mathbf{s}$ ⟶ Process and/or instrument physics $\mathbf{H}$ ⟶ ⊕ ⟶ $\mathbf{y} = \mathbf{H}(\mathbf{s}) + \mathbf{n}$ Measurements / data

$\mathbf{n}$ noise statistical variability

- **1. Imaging as an inverse problem**
  - Basic imaging operators $\mathbf{H}$
  - Discretization of the inverse problem $s \mapsto \mathbf{s}$
- **2. A condensed history of image reconstruction**
  - Classical image reconstruction (1st gen.)
  - Sparsity-based image reconstruction (2nd gen.)
  - The learning revolution (3rd gen.)

EXAMPLE I

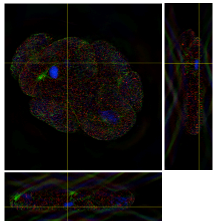
CIBM Center for Biomedical Imaging

3D widefield microscopy

$s(x_1, x_2, x_3, c)$ (3D + 3 fluorophores)

$\updownarrow$

$\mathbf{s}$

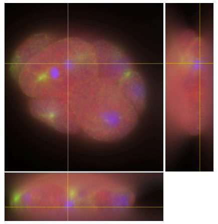


⟶

⟶ ⊕ ⟶

noise $\mathbf{n}$

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



Download the data from: http://bigwww.epfl.ch/deconvolution/bio/ (C. Elegans embryo)

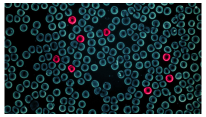
EXAMPLE II

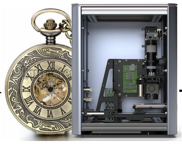
ELISpot and FluoroSpot immunoassays

$s(x_1, x_2, \sigma)$ (2D + time)

$\updownarrow$

$\mathbf{s}$

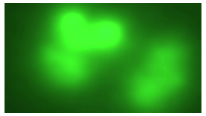




noise

$\mathbf{n}$

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



- Vaccine development (very timely)
- 2D $\mathbf{y} \to$ 3D $\mathbf{s}$

See more at youtu.be/lUHr0nVJ_H0

[del Aguila Pla, Jaldén, 2018] - IEEE TSP, Parts I and II

Download from: https://poldap.github.io/#/pubs

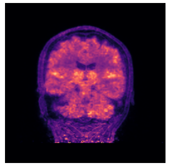
EXAMPLE III

Positron emission tomography (PET)

$s(\cdots)$ (2D/3D/+time)

$\updownarrow$

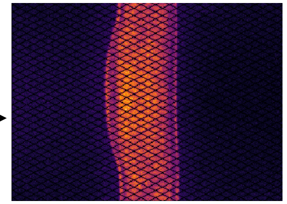
$\mathbf{s}$



noise

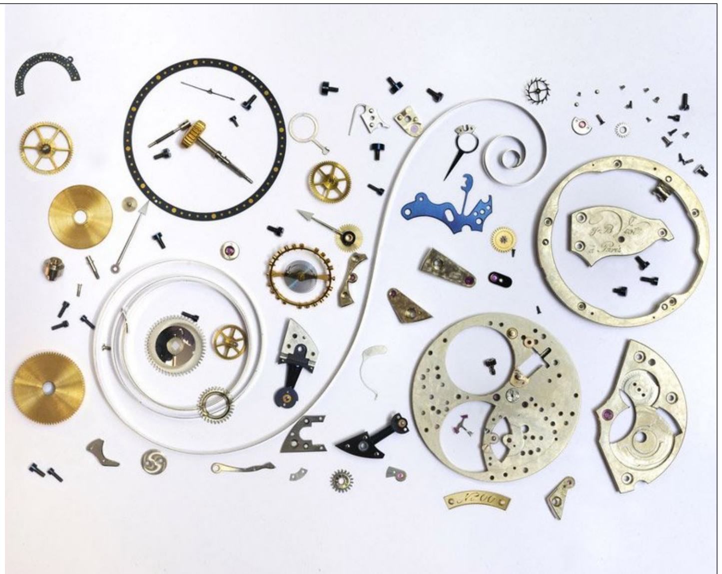
$\mathbf{n}$

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



Part 1

Imaging as an inverse problem



EPFL Biomedical Imaging Group

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 the mth detector

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

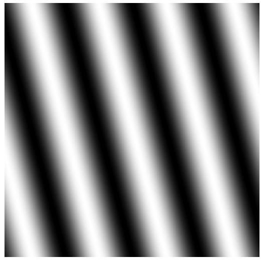
(by the Riesz representation theorem)

$\eta_m(\cdot)$: analysis function

Basic operators: Fourier transform

Images are obviously made of sine waves…

$\mathrm{e}^{\mathrm{j}\langle \boldsymbol{\omega}_0, \boldsymbol{x}\rangle}$



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

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

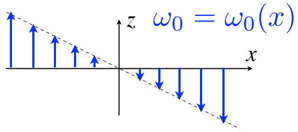
Reconstruction formula (inverse Fourier transform)

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

Ex.: Magnetic resonance imaging

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

Frequency encode:



- Linear forward model for MRI $\quad \boldsymbol{r} = (x, y, z)$

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

(sampling of Fourier transform)

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

- Extended forward model with coil sensitivity

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

Basic operators: Multiplication

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

$\mathrm{M}_w\{s\}(\boldsymbol{r}) = w(\boldsymbol{r}) s(\boldsymbol{r})$

Positive window function (cont. and bounded):
$w \in \mathbf{C}_{\mathrm{b}}(\mathbb{R}^d), w(\boldsymbol{r}) \geq 0$

- Special case: modulation

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

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

Applications:

- MRI with coil sensitivity
- Structured illumination microscopy (SIM)

Convolution

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

$$\mathrm{C}_h\{s\}(\boldsymbol{r}) = (h * s)(\boldsymbol{r}) = \int_{\mathbb{R}^d} h(\boldsymbol{r} - \tilde{\boldsymbol{r}}) s(\tilde{\boldsymbol{r}}) \mathrm{d}\tilde{\boldsymbol{r}}$$

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

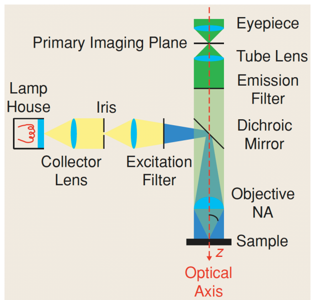
Frequency response: $\hat{h}(\boldsymbol{\omega}) = \mathcal{F}\{h\}(\boldsymbol{\omega})$

- Convolution as a frequency-domain product

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

$\mathrm{C}_h = \mathcal{F}^{-1} \circ \mathrm{M}_{\hat{h}} \circ \mathcal{F}$

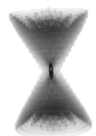
Ex.: 3D widefield microscopy

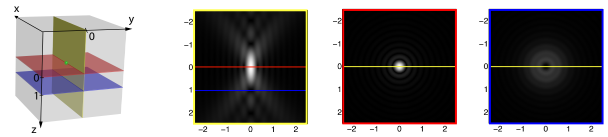


(Vonesch, Aguet, Vonesch, Unser, 2006)

- Diffraction-limited optics
  — Convolution / blur

$\mathbf{h}_{3\mathrm{D}}$





Generate PSFs for a specific microscope with: bigwww.epfl.ch/algorithms/psfgenerator/

Basic operators: X-ray / Radon transform

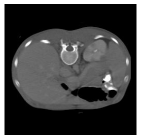
Projection geometry: $\boldsymbol{r} = t\boldsymbol{\theta} + q\boldsymbol{\theta}^{\perp}$ with $\boldsymbol{\theta} = [\cos\theta, \sin\theta]^{\mathrm{T}}$

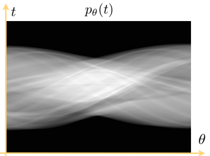
- Radon transform (line integrals) $\mathrm{R} : \mathcal{S}(\mathbb{R}^2) \to \mathcal{S}(\mathbb{R} \times S^1)$

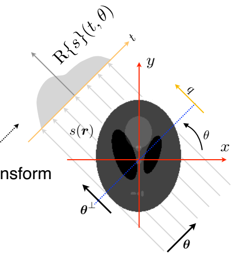
$$p_\theta(t) = \mathrm{R}\{s\}(t,\theta) = \int_{\mathbb{R}} s(t\boldsymbol{\theta} + q\boldsymbol{\theta}^{\perp})\mathrm{d}q = \int_{\mathbb{R}^2} s(\boldsymbol{r})\delta(t - \langle \boldsymbol{r}, \boldsymbol{\theta}\rangle)\mathrm{d}\boldsymbol{r}$$

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

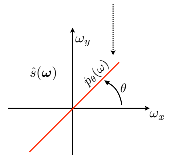

Johann Radon (1887-1956)







Fourier transform



- Central-slice theorem

$$\hat{p}_\theta(\omega) = (\mathcal{F}_t \circ \mathrm{R})\{s\}(\omega,\theta) = \mathcal{F}_{\boldsymbol{r}}\{s\}(\omega\boldsymbol{\theta}) = \hat{s}_{\mathrm{pol}}(\omega,\theta)$$

$\hat{s}_{\mathrm{pol}}$: 2D Fourier transform of s in polar coordinates.
$\mathcal{F}_t$ or $\mathcal{F}_{\boldsymbol{r}}$: Fourier transform along t or $\boldsymbol{r}$ (resp.)

What we have learned in the perspective of some common biomedical imaging modalities

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

But how do we represent $s \in L_2(\mathbb{R}^d)$?

Modality	Radiation	Analysis function η_m
2D or 3D tomography	coherent x-ray	$\delta(t_m - \langle \boldsymbol{r}, \boldsymbol{\theta}_m\rangle)$ $\boldsymbol{\theta}_m$: mth sampled direction
3D deconvolution microscopy	fluorescence	$h(\boldsymbol{r}_m - \boldsymbol{r})$ h: PSF of microscope
Structured illumination microscopy (SIM)	fluorescence	$h(\boldsymbol{r}_m - \boldsymbol{r})w_m(\boldsymbol{r})$ w_m: mth illumination pattern
Positron emission tomography (PET)	gamma rays	$\delta(t_m - \langle \boldsymbol{r}, \boldsymbol{\theta}_m\rangle)$ $(\boldsymbol{\theta}_m, t_m)$: mth LoR
Magnetic resonance imaging (MRI)	radio frequency	$\mathrm{e}^{\mathrm{j}\langle \boldsymbol{\omega}_m, \boldsymbol{r}\rangle}$
Cardiac MRI (parallel, non-uniform)	radio frequency	$w_m(\boldsymbol{r})\mathrm{e}^{\mathrm{j}\langle \boldsymbol{\omega}_m, \boldsymbol{r}\rangle}$ w_m: sensitivity of the mth coil

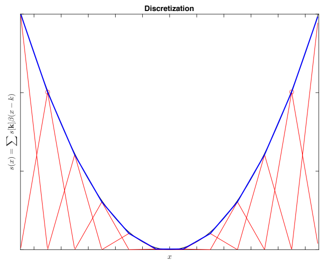
Discretization: Finite dimensional formalism

Represent the continuous signal on a finite combination of synthesis functions $\beta_{\boldsymbol{k}}(\boldsymbol{r})$

$s(\boldsymbol{r}) = \sum_{\boldsymbol{k}\in\Omega} s[\boldsymbol{k}]\beta_{\boldsymbol{k}}(\boldsymbol{r})$ with the signal vector (of coefficients): $\boldsymbol{s} = \big[s[\boldsymbol{k}]\big]_{\boldsymbol{k}\in\Omega}$ of dimension $K = |\Omega|$.

- Applying the forward imaging model to such a signal

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



η_m: analysis function, $\beta_{\boldsymbol{k}}$: synthesis function
$n[\cdot]$: additive noise

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

Example of basis functions

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

Separable generator: $\beta(\boldsymbol{r}) = \prod_{n=1}^{d} \beta(\boldsymbol{r}[n])$

- Pixelated model

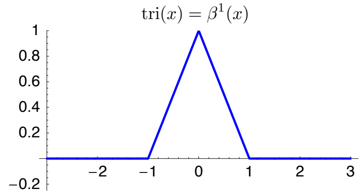
$$\beta(x) = \mathrm{rect}(x)$$

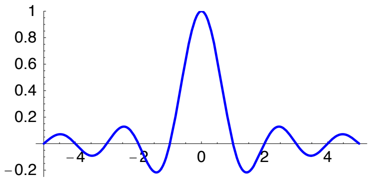
- Bilinear model

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

- Bandlimited representation

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





Part 2: A condensed history of image reconstruction

Classical image reconstruction (1st gen., 20th century)

Sparsity-based image reconstruction (2nd gen., 21st century)

The learning revolution (3rd gen.)

ICASSP2020 Barcelona

See more (2h 30min) at youtu.be/J6_5rPYnr_s

Classical image reconstruction (1st gen., 20th century)

- Easy scenario, well posedness:

$C_1\|\mathbf{s}\|_2 \le \|\mathbf{Hs}\|_2 \le C_2\|\mathbf{s}\|_2$ for all $\mathbf{s}\in\mathcal{X}$

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

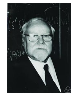
Backprojection: $\mathbf{s}\approx\mathbf{H}^{\mathrm{T}}\mathbf{y}$

Least-squares: $\min_{\tilde{\mathbf{s}}\in\mathbb{R}^M_+}\{\|\mathbf{y}-\mathbf{H}\tilde{\mathbf{s}}\|^2\}$

$\Rightarrow \boldsymbol{s} = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\boldsymbol{y}$

- **Limitations:**
  - Amplification of the noise
  - Interesting problems are ill-posed
  - Computationally expensive matrix inversions

$\mathbf{L}=\mathbf{C}_s^{-1/2}, \lambda=1$



Andrey N. Tikhonov (1906-1993)

$\mathbf{L}$ to promote smoothness

- Tikhonov / Wiener theory

$\min_{\tilde{\mathbf{s}}\in\mathbb{R}^K_+} \underbrace{\|\mathbf{y}-\mathbf{H}\tilde{\mathbf{s}}\|_2^2}_{\text{data consistency}} + \underbrace{\lambda\|\mathbf{L}\tilde{\mathbf{s}}\|_2^2}_{\text{regularization or Gaussian prior}}$

Norbert Wiener (1894-1964)

$s\sim\mathcal{N}(\mathbf{0},\mathbf{C}_s)$

"Filtered" backprojection

$\Rightarrow \mathbf{s} = (\mathbf{H}^T\mathbf{H}+\lambda\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T\mathbf{y}$

- **Limitations:**
  - Computationally expensive matrix inversions

- Efficient iterative algorithms $\min_{\tilde{\mathbf{s}}\in\mathbb{R}^K_+} J(\tilde{\mathbf{s}}, \boldsymbol{y})$

$\|\mathbf{y}-\mathbf{Hs}\|^2 = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} - 2\boldsymbol{s}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{s}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}\boldsymbol{s}$

Steepest-descent type: $\mathbf{s}^{(i+1)}\leftarrow \mathbf{s}^{(i)} - \gamma\nabla_{\mathbf{s}}\{J(\mathbf{s},\mathbf{y})\}(\mathbf{s}^{(i)})$

ex.: projected gradient descent, followed by $\mathbf{s}^{(i+1)}\leftarrow[\mathbf{s}^{(i+1)}]_+$

Classical image reconstruction (1st gen., 20th century)

- Easy scenario, well posedness:

$C_1\|\mathbf{s}\|_2 \le \|\mathbf{Hs}\|_2 \le C_2\|\mathbf{s}\|_2$ for all $\mathbf{s}\in\mathcal{X}$

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

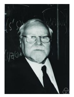
Backprojection: $\mathbf{s}\approx\mathbf{H}^{\mathrm{T}}\mathbf{y}$

Least-squares: $\min_{\tilde{\mathbf{s}}\in\mathbb{R}^M_+}\{\|\mathbf{y}-\mathbf{H}\tilde{\mathbf{s}}\|^2\}$

$\Rightarrow \boldsymbol{s} = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\boldsymbol{y}$

- **Limitations:**
  - Amplification of the noise
  - Interesting problems are ill-posed
  - Computationally expensive matrix inversions

$\mathbf{L}=\mathbf{C}_s^{-1/2}, \lambda=1$



Andrey N. Tikhonov (1906-1993)

$\mathbf{L}$ to promote smoothness

- Tikhonov / Wiener theory

$\min_{\tilde{\mathbf{s}}\in\mathbb{R}^K_+} \underbrace{\|\mathbf{y}-\mathbf{H}\tilde{\mathbf{s}}\|_2^2}_{\text{data consistency}} + \underbrace{\lambda\|\mathbf{L}\tilde{\mathbf{s}}\|_2^2}_{\text{regularization or Gaussian prior}}$

Norbert Wiener (1894-1964)

$s\sim\mathcal{N}(\mathbf{0},\mathbf{C}_s)$

"Filtered" backprojection

$\Rightarrow \mathbf{s} = (\mathbf{H}^T\mathbf{H}+\lambda\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T\mathbf{y}$

- **Limitations:**
  - Computationally expensive matrix inversions

- Efficient iterative algorithms $\min_{\tilde{\mathbf{s}}\in\mathbb{R}^K_+} J(\tilde{\mathbf{s}}, \boldsymbol{y})$

$\|\mathbf{y}-\mathbf{Hs}\|^2 = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} - 2\boldsymbol{s}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{s}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}\boldsymbol{s}$

Steepest-descent type: $\mathbf{s}^{(i+1)}\leftarrow \mathbf{s}^{(i)} - \gamma\nabla_{\mathbf{s}}\{J(\mathbf{s},\mathbf{y})\}(\mathbf{s}^{(i)})$

ex.: projected gradient descent, followed by $\mathbf{s}^{(i+1)}\leftarrow[\mathbf{s}^{(i+1)}]_+$

Least squares:

$\mathbf{s}^{(i+1)}\leftarrow\mathbf{s}^{(i)}+\tilde{\gamma}\left(\mathbf{s}_0-(\mathbf{H}^T\mathbf{H})\mathbf{s}^{(i)}\right)$ with $\mathbf{s}_0=\mathbf{H}^T\mathbf{y}$

Ground truth

Gaussian blur + additive noise

van Cittert animation

Classical image reconstruction (1st gen., 20th century)

- Easy scenario, well posedness:

$C_1\|\mathbf{s}\|_2 \le \|\mathbf{Hs}\|_2 \le C_2\|\mathbf{s}\|_2$ for all $\mathbf{s}\in\mathcal{X}$

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

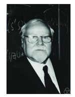
Backprojection: $\mathbf{s}\approx\mathbf{H}^{\mathrm{T}}\mathbf{y}$

Least-squares: $\min_{\tilde{\mathbf{s}}\in\mathbb{R}^M_+}\{\|\mathbf{y}-\mathbf{H}\tilde{\mathbf{s}}\|^2\}$

$\Rightarrow \boldsymbol{s} = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\boldsymbol{y}$

- **Limitations:**
  - Amplification of the noise
  - Interesting problems are ill-posed
  - Computationally expensive matrix inversions

$\mathbf{L}=\mathbf{C}_s^{-1/2}, \lambda=1$



Andrey N. Tikhonov (1906-1993)

$\mathbf{L}$ to promote smoothness

- Tikhonov / Wiener theory

$\min_{\tilde{\mathbf{s}}\in\mathbb{R}^K_+} \underbrace{\|\mathbf{y}-\mathbf{H}\tilde{\mathbf{s}}\|_2^2}_{\text{data consistency}} + \underbrace{\lambda\|\mathbf{L}\tilde{\mathbf{s}}\|_2^2}_{\text{regularization or Gaussian prior}}$

Norbert Wiener (1894-1964)

$s\sim\mathcal{N}(\mathbf{0},\mathbf{C}_s)$

"Filtered" backprojection

$\Rightarrow \mathbf{s} = (\mathbf{H}^T\mathbf{H}+\lambda\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T\mathbf{y}$

- **Limitations:**
  - Computationally expensive matrix inversions

- Efficient iterative algorithms $\min_{\tilde{\mathbf{s}}\in\mathbb{R}^K_+} J(\tilde{\mathbf{s}}, \boldsymbol{y})$

$\|\mathbf{y}-\mathbf{Hs}\|^2 = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} - 2\boldsymbol{s}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{s}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}\boldsymbol{s}$

Steepest-descent type: $\mathbf{s}^{(i+1)}\leftarrow \mathbf{s}^{(i)} - \gamma\nabla_{\mathbf{s}}\{J(\mathbf{s},\mathbf{y})\}(\mathbf{s}^{(i)})$

ex.: projected gradient descent, followed by $\mathbf{s}^{(i+1)}\leftarrow[\mathbf{s}^{(i+1)}]_+$

Tikhonov:

$\mathbf{s}^{(i+1)}\leftarrow\mathbf{s}^{(i)}+\tilde{\gamma}\left(\mathbf{s}_0-(\mathbf{H}^T\mathbf{H}+\lambda\mathbf{L}^T\mathbf{L})\mathbf{s}^{(i)}\right)$

Ground truth

Gaussian blur + additive noise

Not enough: λ=0.2

Optimal: λ=2

Too much: λ=20

Classical image reconstruction (1st gen., 20th century)

- Easy scenario, well posedness:

$C_1\|\mathbf{s}\|_2 \le \|\mathbf{H}\mathbf{s}\|_2 \le C_2\|\mathbf{s}\|_2$ for all $\mathbf{s}\in\mathcal{X}$

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

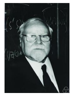
Backprojection: $\mathbf{s} \approx \mathbf{H}^{\mathrm{T}}\mathbf{y}$

Least-squares: $\min_{\tilde{\mathbf{s}}\in\mathbb{R}^M_+} \{\|\mathbf{y}-\mathbf{H}\tilde{\mathbf{s}}\|^2\}$

$\Rightarrow s = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{y}$

- **Limitations:**
  - Amplification of the noise
  - Interesting problems are ill-posed
  - Computationally expensive matrix inversions

$\mathbf{L} = \mathbf{C}_s^{-1/2}, \lambda = 1$



Andrey N. Tikhonov
(1906-1993)

L to promote smoothness

- Tikhonov / Wiener theory

$$\min_{\tilde{\mathbf{s}}\in\mathbb{R}^K_+} \underbrace{\|\mathbf{y}-\mathbf{H}\tilde{\mathbf{s}}\|_2^2}_{\text{data consistency}} + \underbrace{\lambda\|\mathbf{L}\tilde{\mathbf{s}}\|_2^2}_{\text{regularization or Gaussian prior}}$$

"Filtered" backprojection

$\Rightarrow \mathbf{s} = (\mathbf{H}^T\mathbf{H}+\lambda\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T\mathbf{y}$

Norbert Wiener
(1894-1964)

$s \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_s)$

- **Limitations:**
  - Computationally expensive matrix inversions

- Efficient iterative algorithms $\min_{\tilde{\mathbf{s}}\in\mathbb{R}^K_+} J(\tilde{\mathbf{s}}, \boldsymbol{y})$

$\|\mathbf{y}-\mathbf{H}\mathbf{s}\|^2 = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} - 2\boldsymbol{s}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{s}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}\boldsymbol{s}$

Steepest-descent type: $\mathbf{s}^{(i+1)} \leftarrow \mathbf{s}^{(i)} - \gamma\nabla_{\mathbf{s}}\{J(\mathbf{s},\mathbf{y})\}(\mathbf{s}^{(i)})$

ex.: projected gradient descent, followed by $\mathbf{s}^{(i+1)} \leftarrow [\mathbf{s}^{(i+1)}]_+$

Tikhonov:

$\mathbf{s}^{(i+1)} \leftarrow \mathbf{s}^{(i)} + \tilde{\gamma}(\mathbf{s}_0 - (\mathbf{H}^T\mathbf{H}+\lambda\mathbf{L}^T\mathbf{L})\mathbf{s}^{(i)})$

- Fast implementation
  - Diagonalization of convolution matrices $\Rightarrow$ FFT-based implementation
  - Applicable to:
    1. deconvolution microscopy (Wiener filter) $\rightarrow$ if $\mathbf{H}$ is a convolution matrix, so is $\mathbf{H}^{\mathrm{T}}\mathbf{H}$.
    2. parallel-ray computed tomography (FBP) $\rightarrow$ $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ is a convolution matrix.
    3. MRI

Sparsity-based image reconstruction (2nd gen., 21st century)

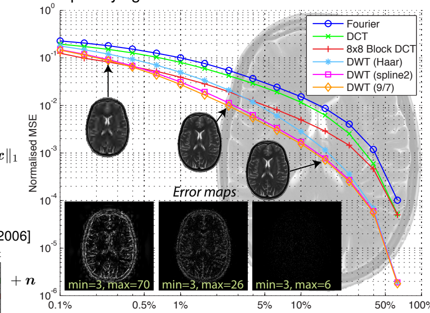
- Non-quadratic regularization

$$\min_{\mathbf{s}\in\mathbb{R}^K_+} \|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 + \lambda\mathcal{R}(\mathbf{s})$$

$\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_2^2 \quad \longrightarrow \|\mathbf{L}\mathbf{s}\|_p^p \quad \longrightarrow \mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_1$

- Total variation, $\mathbf{L}$: gradient [Rudin, Osher et al., 1992]
- Wavelet-domain regularization, $\boldsymbol{s} = \mathbf{W}\boldsymbol{x}$, $\mathcal{R}(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$
  equivalent to $\mathbf{L} = \mathbf{W}^{-1}$
  [Figuereido et al., 2004] [Daubechies et al., 2004]

- Sparsifying transforms $\mathbf{W}^{-1}$



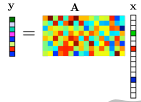
[Guerquin-Kern et al., 2011] - IEEE TMI

- Formalization: Compressed sensing [Donoho et al., 2005][Candès, Romberg, Tao, 2006]

Sparse representation of signal vector $\boldsymbol{x}\in\mathbb{R}^N$ with $\|\mathbf{x}\|_0 = S \ll N$

$M\times N$ system matrix: $\mathbf{A} = \mathbf{H}\mathbf{W}$ in the undersampled regime $M \ll N$

and $2S < M$



$$\min_{\boldsymbol{x}:\|\boldsymbol{x}\|_0\le K} \|\boldsymbol{y}-\mathbf{A}\boldsymbol{x}\|_2^2 \quad \Leftrightarrow \quad \min_{\boldsymbol{x}:\|\boldsymbol{x}\|_1\le C} \|\boldsymbol{y}-\mathbf{A}\boldsymbol{x}\|_2^2$$

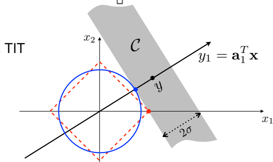
(nonconvex) (convex)

w. some conditions on $\mathbf{A}$

- A representer-theorem view [Unser, Fageot, Gupta, 2016] - IEEE TIT

$\min_{\mathbf{x}} \{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_2^2\} \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{x}\|_2$ subject to $\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 \le \sigma^2$

$\min_{\mathbf{x}} \{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1\} \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{x}\|_1$ subject to $\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 \le \sigma^2$



ℓ_2-ball: $|x_1|^2 + |x_2|^2 = C_2$

ℓ_1-ball: $|x_1| + |x_2| = C_1$

Sparsity-based image reconstruction (2nd gen., 21st century)

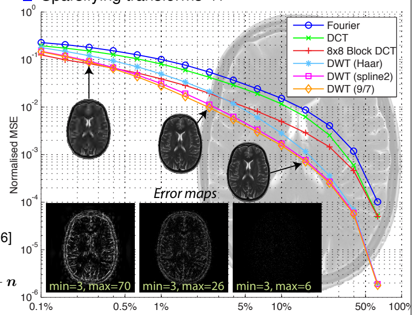
- Non-quadratic regularization

$$\min_{\mathbf{s}\in\mathbb{R}^K_+} \|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 + \lambda\mathcal{R}(\mathbf{s})$$

$\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_2^2 \quad \longrightarrow \|\mathbf{L}\mathbf{s}\|_p^p \quad \longrightarrow \mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_1$

- Total variation, $\mathbf{L}$: gradient [Rudin, Osher et al., 1992]
- Wavelet-domain regularization, $\boldsymbol{s} = \mathbf{W}\boldsymbol{x}$, $\mathcal{R}(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$
  equivalent to $\mathbf{L} = \mathbf{W}^{-1}$
  [Figuereido et al., 2004] [Daubechies et al., 2004]

- Sparsifying transforms $\mathbf{W}^{-1}$



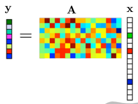
[Guerquin-Kern et al., 2011] - IEEE TMI

- Formalization: Compressed sensing [Donoho et al., 2005][Candès, Romberg, Tao, 2006]

Sparse representation of signal vector $\boldsymbol{x}\in\mathbb{R}^N$ with $\|\mathbf{x}\|_0 = S \ll N$

$M\times N$ system matrix: $\mathbf{A} = \mathbf{H}\mathbf{W}$ in the undersampled regime $M \ll N$

and $2S < M$



$$\min_{\boldsymbol{x}:\|\boldsymbol{x}\|_0\le K} \|\boldsymbol{y}-\mathbf{A}\boldsymbol{x}\|_2^2 \quad \Leftrightarrow \quad \min_{\boldsymbol{x}:\|\boldsymbol{x}\|_1\le C} \|\boldsymbol{y}-\mathbf{A}\boldsymbol{x}\|_2^2$$

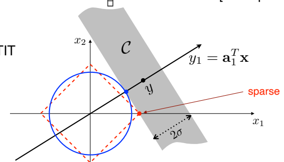
(nonconvex) (convex)

w. some conditions on $\mathbf{A}$

- A representer-theorem view [Unser, Fageot, Gupta, 2016] - IEEE TIT

$\min_{\mathbf{x}} \{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_2^2\} \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{x}\|_2$ subject to $\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 \le \sigma^2$

$\min_{\mathbf{x}} \{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1\} \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{x}\|_1$ subject to $\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 \le \sigma^2$



ℓ_2-ball: $|x_1|^2 + |x_2|^2 = C_2$

ℓ_1-ball: $|x_1| + |x_2| = C_1$

Sparsity-based image reconstruction (2nd gen., 21st century)

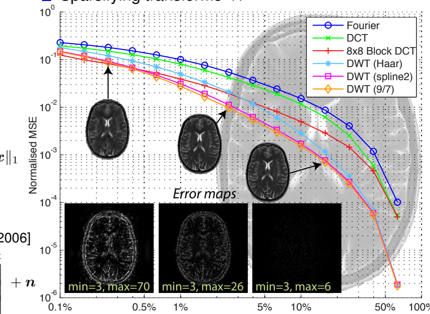
- Non-quadratic regularization

$$\min_{\mathbf{s}\in\mathbb{R}^K_+} \|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 + \lambda\mathcal{R}(\mathbf{s})$$

$\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_2^2 \quad \longrightarrow \|\mathbf{L}\mathbf{s}\|_p^p \quad \longrightarrow \mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_1$

- Total variation, $\mathbf{L}$: gradient [Rudin, Osher et al., 1992]
- Wavelet-domain regularization, $\boldsymbol{s} = \mathbf{W}\boldsymbol{x}$, $\mathcal{R}(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$
  equivalent to $\mathbf{L} = \mathbf{W}^{-1}$
  [Figuereido et al., 2004] [Daubechies et al., 2004]

- Sparsifying transforms $\mathbf{W}^{-1}$



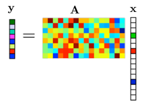
[Guerquin-Kern et al., 2011] - IEEE TMI

- Formalization: Compressed sensing [Donoho et al., 2005][Candès, Romberg, Tao, 2006]

Sparse representation of signal vector $\boldsymbol{x}\in\mathbb{R}^N$ with $\|\mathbf{x}\|_0 = S \ll N$

$M\times N$ system matrix: $\mathbf{A} = \mathbf{H}\mathbf{W}$ in the undersampled regime $M \ll N$

and $2S < M$



$$\min_{\boldsymbol{x}:\|\boldsymbol{x}\|_0\le K} \|\boldsymbol{y}-\mathbf{A}\boldsymbol{x}\|_2^2 \quad \Leftrightarrow \quad \min_{\boldsymbol{x}:\|\boldsymbol{x}\|_1\le C} \|\boldsymbol{y}-\mathbf{A}\boldsymbol{x}\|_2^2$$

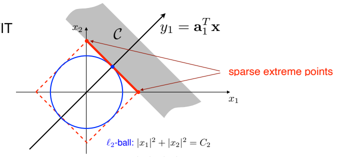
(nonconvex) (convex)

w. some conditions on $\mathbf{A}$

- A representer-theorem view [Unser, Fageot, Gupta, 2016] - IEEE TIT

$\min_{\mathbf{x}} \{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_2^2\} \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{x}\|_2$ subject to $\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 \le \sigma^2$

$\min_{\mathbf{x}} \{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1\} \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{x}\|_1$ subject to $\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 \le \sigma^2$



ℓ_2-ball: $|x_1|^2 + |x_2|^2 = C_2$

ℓ_1-ball: $|x_1| + |x_2| = C_1$

Sparsity-based image reconstruction (2nd gen., 21st century)

■ Non-quadratic regularization

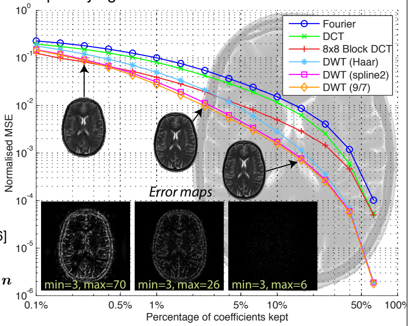
$$\min_{\mathbf{s}\in\mathbb{R}^K_+} \|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 + \lambda\mathcal{R}(\mathbf{s})$$

$\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_2^2 \quad \longrightarrow \|\mathbf{L}\mathbf{s}\|_p^p \quad \longrightarrow \mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_1$

■ Total variation, $\mathbf{L}$: gradient [Rudin, Osher et al., 1992]

■ Wavelet-domain regularization, $\boldsymbol{s} = \mathbf{W}\boldsymbol{x}$, $\mathcal{R}(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$
equivalent to $\mathbf{L} = \mathbf{W}^{-1}$
[Figuereido et al., 2004] [Daubechies et al., 2004]

■ Sparsifying transforms $\mathbf{W}^{-1}$

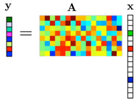


[Guerquin-Kern et al., 2011] - IEEE TMI

■ Formalization: Compressed sensing [Donoho et al., 2005][Candès, Romberg, Tao, 2006]

Sparse representation of signal vector $\boldsymbol{x}\in\mathbb{R}^N$ with $\|\mathbf{x}\|_0 = S \ll N$

$M\times N$ system matrix: $\mathbf{A} = \mathbf{H}\mathbf{W}$ in the undersampled regime $M \ll N$ and $2S < M$



$$\min_{\boldsymbol{x}:\|\boldsymbol{x}\|_0\le K} \|\boldsymbol{y}-\mathbf{A}\boldsymbol{x}\|_2^2 \quad \Leftrightarrow \quad \min_{\boldsymbol{x}:\|\boldsymbol{x}\|_1\le C} \|\boldsymbol{y}-\mathbf{A}\boldsymbol{x}\|_2^2$$

(nonconvex) (convex)

w. some conditions on $\mathbf{A}$

■ A representer-theorem view [Unser, Fageot, Gupta, 2016] - IEEE TIT

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

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

■ Sparse stochastic processes

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

There is an operator L that "whitens" the continuous signal, i.e., $w = \mathrm{L}\{s\}$
$\Rightarrow \mathbf{u} = \mathbf{L}\mathbf{s}$ where $\mathbf{L}$ discretizes L, and $p_U(\boldsymbol{u})$ is inifinitely divisible

$$p_{S|Y}(\mathbf{s}|\mathbf{y}) \propto \exp\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{s}\|^2}{2\sigma^2}\right)\prod_{k\in\Omega} p_U([\mathbf{L}\mathbf{s}]_k) \qquad \|\cdot\|_1 \Leftrightarrow U \text{ is Laplace}$$

Sparsity-based image reconstruction (2nd gen., 21st century)

■ Proximal optimization algorithms: The workhorse of 2nd generation methods

■ Ex: Altenating direction method of multipliers (ADMM), [Boyd et al., 2011]

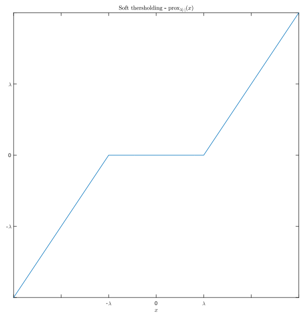
$$\min_{\mathbf{s},\mathbf{u}:\mathbf{u}=\mathbf{L}\mathbf{s}} \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 + \lambda\mathcal{R}(\mathbf{u}) \qquad \mathcal{L}_{\mathcal{A}}(\mathbf{s},\mathbf{u},\boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 + \lambda\mathcal{R}(\mathbf{u}) + \boldsymbol{\alpha}^T(\mathbf{L}\mathbf{s}-\mathbf{u}) + \frac{\mu}{2}\|\mathbf{L}\mathbf{s}-\mathbf{u}\|_2^2$$

Linear inverse problem: $\quad \mathbf{s}^{k+1} \leftarrow \left(\mathbf{H}^T\mathbf{H} + \mu\mathbf{L}^T\mathbf{L}\right)^{-1}\left(\mathbf{H}^T\mathbf{y} + \mathbf{z}^{k+1}\right)$ with $\mathbf{z}^{k+1} \leftarrow \mathbf{L}^T\left(\mu\mathbf{u}^k - \boldsymbol{\alpha}^k\right)$

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

Nonlinear denoising: $\quad \mathbf{u}^{k+1} \leftarrow \mathrm{prox}_{\frac{\lambda}{\mu}\mathcal{R}}\left(\mathbf{L}\mathbf{s}^{k+1} + \frac{1}{\mu}\boldsymbol{\alpha}^{k+1}\right)$

$\mathrm{prox}_{\lambda\mathcal{R}}(\mathbf{y}) = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{y}-\mathbf{u}\|_2^2 + \lambda\mathcal{R}(\mathbf{u})$ (often separable)



- Example: $\mathbf{L}$ wavelet, $\mathcal{R}(\mathbf{u}) = \|\mathbf{u}\|_1$ (Laplace model) $\rightarrow$ wavelet thresholding denoiser
- Separation in linear step (problem specific) and statistical step (denoiser for prior distribution)

■ **Advantages:**
- Higher reconstruction quality [Lustig et al., 2007]
- Low-dose, faster imaging (CS)
- Re-use of fast implementations for 1st gen. methods

■ **Limitations:**
- Increased complexity (despite efficient solutions)
- Challenging to implement efficiently

erc GlobalBioIm

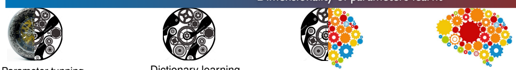
Download from: github.com/Biomedical-Imaging-Group/GlobalBioIm

See more at bit.ly/GlobalBioIm-example
Explained with all detail at youtu.be/J6_5rPYnr_s?t=6240

The learning revolution (3rd gen.)

■ Overarching topic: How do we leverage what we already know?

Dimensionality of parameters learnt



Parameter tunning
- Learning λ
- Calibrate forward model

Dictionary learning
- An "optimal" $\mathbf{L}$

[Elad 2006]
[Ravishankar 2011]
[Mairal 2012]

Learning non-linear functions
- An "optimal" $\mathcal{R}$
- CNN parametrization
- Backpropagation

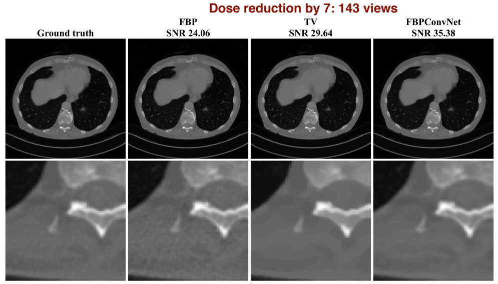
[Chen, Pock, 2015-2016]
[Kamilov 2016]

Task-specific learning (CNNs)
- Artifact removal for 1st gen. reconstructions
- Super-resolution in 3D imaging
- Speeding up well-performing 2nd gen. methods
- Low-dose / short-scan regime

[Jin et al., 2016] → 2019 Best Paper Award IEEE Signal Processing Society
[Weigert et al., 2017]
[Hammernik et al., 2018]

Dose reduction by 7: 143 views

CT data



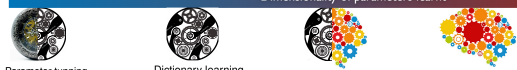
1000 views FBP

[Jin et al., 2016]

The learning revolution (3rd gen.)

■ Overarching topic: How do we leverage what we already know?

Dimensionality of parameters learnt



Parameter tunning
- Learning λ
- Calibrate forward model

Dictionary learning
- An "optimal" $\mathbf{L}$

[Elad 2006]
[Ravishankar 2011]
[Mairal 2012]

Learning non-linear functions
- An "optimal" $\mathcal{R}$
- CNN parametrization
- Backpropagation

[Chen, Pock, 2015-2016]
[Kamilov 2016]

Task-specific learning (CNNs)
- Artifact removal for 1st gen. reconstructions
- Super-resolution in 3D imaging
- Speeding up well-performing 2nd gen. methods
- Low-dose / short-scan regime

[Jin et al., 2016] → 2019 Best Paper Award IEEE Signal Processing Society
[Weigert et al., 2017]
[Hammernik et al., 2018]

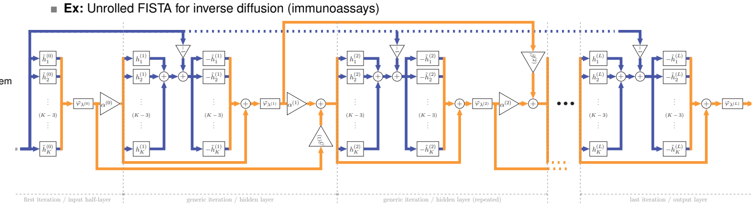
■ Going too far to the right
- Ignoring the physics, learning them instead

[Zhu et al., 2018] - Nature

■ Guiding principles
- Learn image-to-image maps, either
  1. within the image (sparsified or not) domain, or
  2. within the measurement domain.
- Need for very large training datasets with gold-standard reconstructions. But exploit the problem structure.
- Data augmentation.

■ Unrolling as a motivation for deep architechtures [Gregor, LeCun, 2010]
- Use 2nd generation methods as a starting point
- Motivate the architecture using optimization

■ **Ex:** Unrolled FISTA for inverse diffusion (immunoassays)

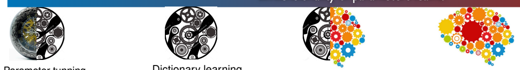


[del Aguila Pla, Saxena, Jaldén, 2019]

The learning revolution (3rd gen.)

- Overarching topic: How do we leverage what we already know?

Dimensionality of parameters learnt



Parameter tunning
- Learning λ
- Calibrate forward model

Dictionary learning
- An "optimal" $\mathbf{L}$

[Elad 2006]
[Ravishankar 2011]
[Mairal 2012]

Learning non-linear functions
- An "optimal" $\mathcal{R}$
- CNN parametrization
- Backpropagation

[Chen, Pock, 2015-2016]
[Kamilov 2016]

Task-specific learning (CNNs)
- Artifact removal for 1st gen. reconstructions
- Super-resolution in 3D imaging
- Speeding up well-performing 2nd gen. methods
- Low-dose / short-scan regime

[Jin et al., 2016] → 2019 Best Paper Award IEEE Signal Processing Society
[Weigert et al., 2017]
[Hammernik et al., 2018]

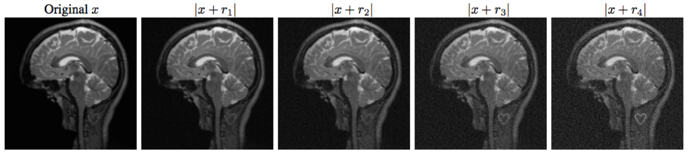
- Going too far to the right
  - Ignoring the physics, learning them instead

  [Zhu et al., 2018] - Nature

- Guiding principles
  - Learn image-to-image maps, either
    1. within the image (sparsified or not) domain, or
    2. within the measurement domain.
  - Need for very large training datasets with gold-standard reconstructions. But exploit the problem structure.
  - Data augmentation.

- Unrolling as a motivation for deep architechtures [Gregor, LeCun, 2010]
  - Use 2nd generation methods as a starting point
  - Motivate the architecture using optimization

- A word of warning: Instability in deep learning

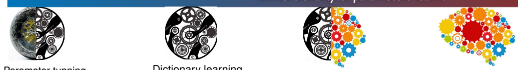


[Antun et al., 2020] On instabilities of deep learning in image reconstruction - Does AI come at a cost?

The learning revolution (3rd gen.)

- Overarching topic: How do we leverage what we already know?

Dimensionality of parameters learnt



Parameter tunning
- Learning λ
- Calibrate forward model

Dictionary learning
- An "optimal" $\mathbf{L}$

[Elad 2006]
[Ravishankar 2011]
[Mairal 2012]

Learning non-linear functions
- An "optimal" $\mathcal{R}$
- CNN parametrization
- Backpropagation

[Chen, Pock, 2015-2016]
[Kamilov 2016]

Task-specific learning (CNNs)
- Artifact removal for 1st gen. reconstructions
- Super-resolution in 3D imaging
- Speeding up well-performing 2nd gen. methods
- Low-dose / short-scan regime

[Jin et al., 2016] → 2019 Best Paper Award IEEE Signal Processing Society
[Weigert et al., 2017]
[Hammernik et al., 2018]

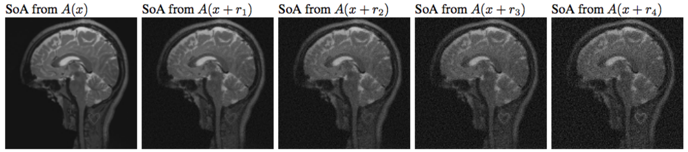
- Going too far to the right
  - Ignoring the physics, learning them instead

  [Zhu et al., 2018] - Nature

- Guiding principles
  - Learn image-to-image maps, either
    1. within the image (sparsified or not) domain, or
    2. within the measurement domain.
  - Need for very large training datasets with gold-standard reconstructions. But exploit the problem structure.
  - Data augmentation.

- Unrolling as a motivation for deep architechtures [Gregor, LeCun, 2010]
  - Use 2nd generation methods as a starting point
  - Motivate the architecture using optimization

- A word of warning: Instability in deep learning

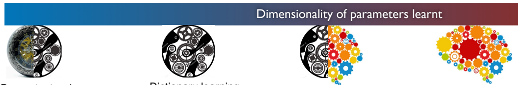


[Antun et al., 2020] On instabilities of deep learning in image reconstruction - Does AI come at a cost?

The learning revolution (3rd gen.)

- Overarching topic: How do we leverage what we already know?

Dimensionality of parameters learnt



Parameter tunning
- Learning λ
- Calibrate forward model

Dictionary learning
- An "optimal" $\mathbf{L}$

[Elad 2006]
[Ravishankar 2011]
[Mairal 2012]

Learning non-linear functions
- An "optimal" $\mathcal{R}$
- CNN parametrization
- Backpropagation

[Chen, Pock, 2015-2016]
[Kamilov 2016]

Task-specific learning (CNNs)
- Artifact removal for 1st gen. reconstructions
- Super-resolution in 3D imaging
- Speeding up well-performing 2nd gen. methods
- Low-dose / short-scan regime

[Jin et al., 2016] → 2019 Best Paper Award IEEE Signal Processing Society
[Weigert et al., 2017]
[Hammernik et al., 2018]

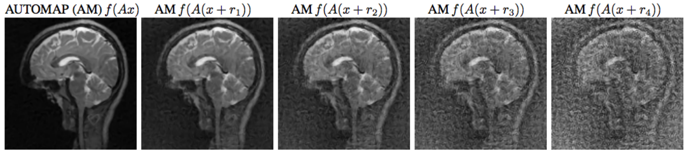
- Going too far to the right
  - Ignoring the physics, learning them instead

  [Zhu et al., 2018] - Nature

- Guiding principles
  - Learn image-to-image maps, either
    1. within the image (sparsified or not) domain, or
    2. within the measurement domain.
  - Need for very large training datasets with gold-standard reconstructions. But exploit the problem structure.
  - Data augmentation.

- Unrolling as a motivation for deep architechtures [Gregor, LeCun, 2010]
  - Use 2nd generation methods as a starting point
  - Motivate the architecture using optimization

- A word of warning: Instability in deep learning   Kernel (nullspace) awareness



[Antun et al., 2020] On instabilities of deep learning in image reconstruction - Does AI come at a cost?

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 3, 2019.
- 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*, vol. 59, no. 4, pp. 769-793, December 2017.

Algorithms and imaging applications

- E. Soubies, F. Soulez, M.T. McCann, T.-a. Pham, L. Donati, T. Debarre, D. Sage, M. Unser, "**Pocket Guide to Solve Inverse Problems with GlobalBioIm**," *Inverse Problems*, vol. 35, no. 10, paper no. 104006, pp. 1-20, October 2019.
- 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, 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, 2011.
- E. Soubies, M. Unser, "Computational Super-Sectioning for Single-Slice Structured-Illumination Microscopy," *IEEE Trans. Computational Imaging*, vol. 5, no. 2, pp. 24
- 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.

References (Cont'd)

- P. del Aguila Pla and J. Jaldén, "Cell detection by functional inverse diffusion and non-negative group sparsity—Part I: Modeling and Inverse Problems", *IEEE Transactions on Signal Processing*, vol. 66, no. 20, pp. 5407–5421, 2018
- P. del Aguila Pla and J. Jaldén, "Cell detection by functional inverse diffusion and non-negative group sparsity—Part II: Proximal optimization and Performance evaluation", *IEEE Transactions on Signal Processing*, vol. 66, no. 20, pp. 5422–5437, 2018

Deep neural networks

- K.H. Jin, M.T. McCann, E. Froustey, M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Trans. Image Processing*, vol. 26, no. 9, pp. 4509-4522, 2017.
- M.T. McCann, K.H. Jin, M. Unser, "Convolutional Neural Networks for Inverse Problems in Imaging—A Review," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 85-95, 2017.
- H. Gupta, K.H. Jin, H.Q. Nguyen, M.T. McCann, M. Unser, "CNN-Based Projected Gradient Descent for Consistent CT Image Reconstruction," *IEEE Trans. Medical Imaging*, vol. 37, no. 6, pp. 1440-1453, 2018.
- M. Unser, "A representer theorem for deep neural networks," Journal of Machine Learning Research, vol. 20, no. 110, pp. 1-30, 2019.
- P. del Aguila Pla, V. Saxena, J. Jaldén, "SpotNet - Learned iterations for cell detection in image-based immunoassays," 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, April 8-11, 2019.

Acknowledgments

Many thanks to current and former members of the EPFL's Biomedical Imaging Group


Prof. Michaël Unser

EPFL CIBM Center for Biomedical Imaging

Preprints, papers, and demos: bigwww.epfl.ch