Pre-course, part I:
images, noise, regularization, and optimization

Luca Ratti
Department of Mathematics, University of Bologna, Italy

Mathematics & Machine Learning for image analysis
ERASMUS+ International PhD Summer School
Bologna, June 9, 2026

An introduction to the school via image denoising

Part I (14:30-15:45)

- Digital images
- Modelling noise
- Quality measures (MSE, PSNR, SSIM)
- Image denoising as a toy 'inverse' problem
- Bayesian formulation
- MAP estimators
- Regularisation
- Bits of optimisation: GD, proximal operator

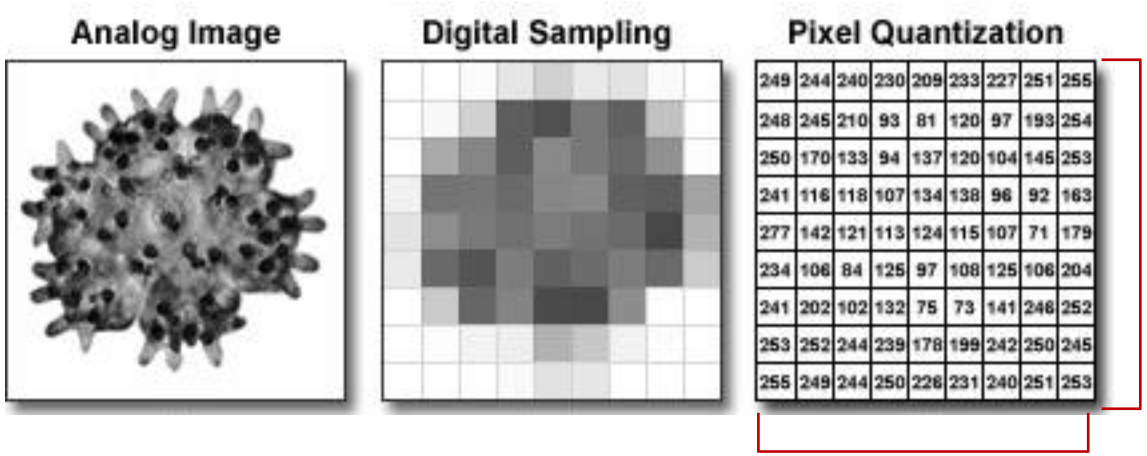
Slides: thanks to **Luca Calatroni**,
MaLGa, University of Genova

Part II (16:00-17:15)

- Image denoising as a regression problem
- Supervised/unsupervised learning
- Neural networks
- Bias-variance tradeoff
- Over-parametrization
- NNs for imaging: CNN, U-NET
- Training a NN
- NumPy/Matplotlib/Pytorch (tomorrow)

Digital images

Digital images are discrete representations of the continuous world we live in.

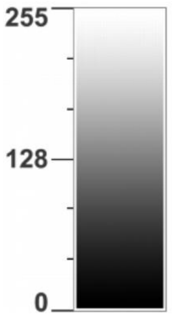


n_2

n_1

Sampling: allows to represent a continuous image into a finite (pixel) grid.

Quantisation: assigns a grey-level describing average brightness at each pixel.



$\Omega = \{1, \dots, n_1\} \times \{1, \dots, n_2\}$ → image domain

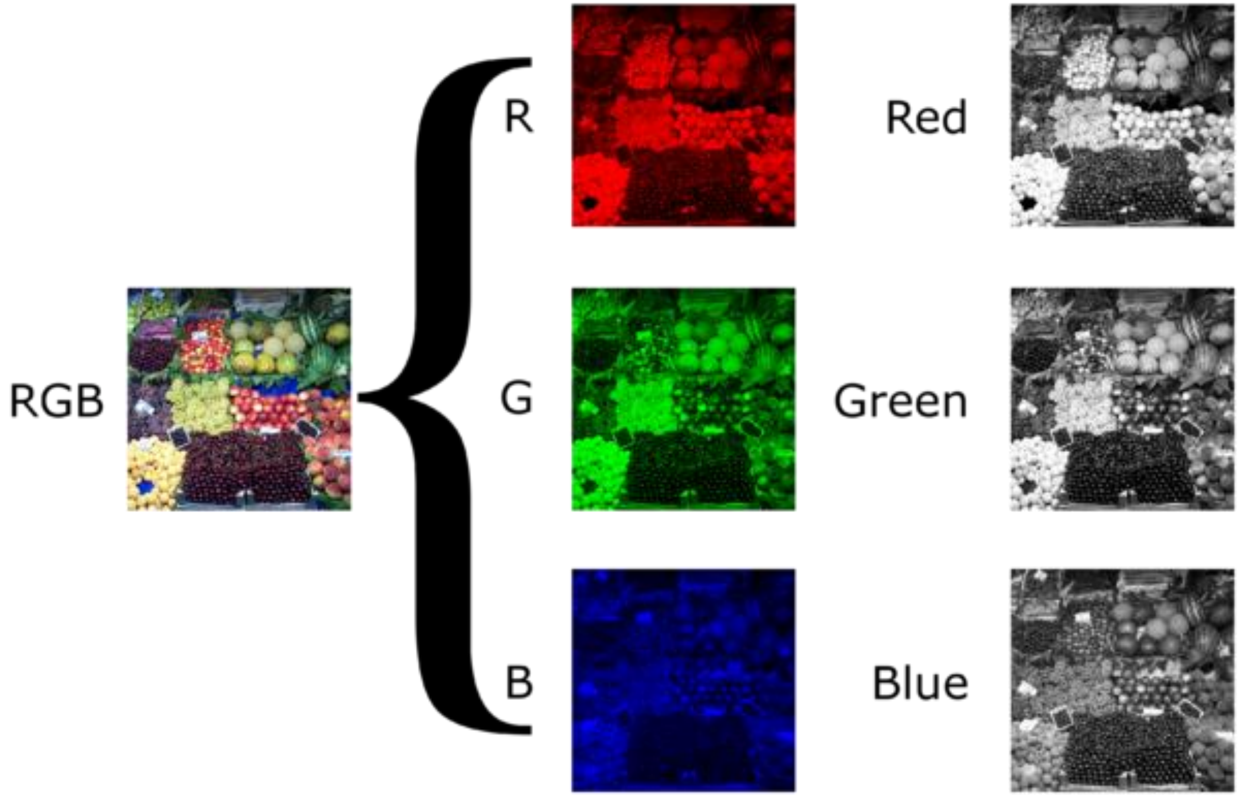
$\mathbf{X}: \Omega \to \{0, \dots, 255\}$

$\mathbf{X} = (x_{i,j}) \in \{0, \dots, 255\}^{n_1 \times n_2}$ → normalisation, adjustments.. $\mathbf{X} \in [0,1]^{n_1 \times n_2}$ $\quad \mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$

Upon vectorisation of the 2D image $\mathbf{X}$, consider a vector $\mathbf{x} \in \mathbb{R}^n$, with $n = n_1 n_2$.

Grayscale/RGB images

Natural scenes are not grayscale. Color is a combination of Red-Green-Blue channels.



The higher the intensity in the individual channel, the more represented is the color.

$$x_{i,j} = (r_{i,j}, g_{i,j}, b_{i,j}) \in \mathbb{R}^3$$

Hence, for RGB images:

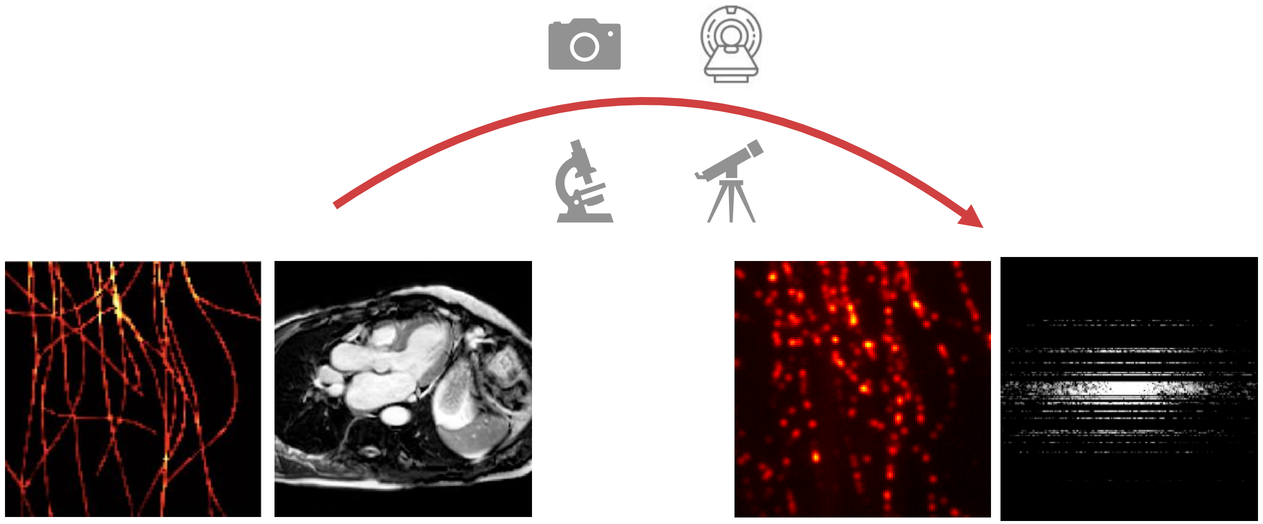
$$\mathbf{x} \in \mathbb{R}^{n\times 3} \quad \text{or} \quad \mathbb{R}^{3n}$$

- Other color spaces are possible (HSV, …)
- Often, color channels are processed separately.

Assume $\mathbf{x}$ is grayscale (extension is trivial)

Modeling degradation processes

Images are matrices/vectors. How to model acquisition processes?



$\mathbf{x} \in \mathbb{R}^n$
(unknown)

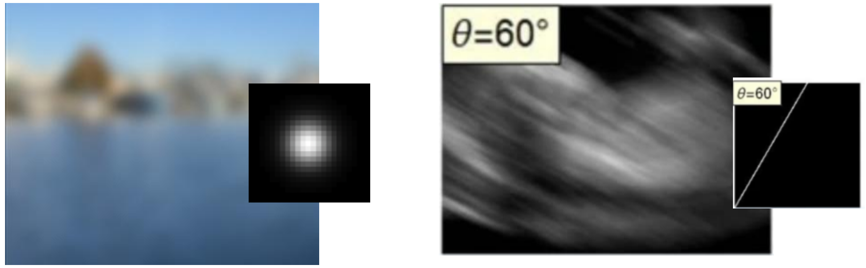
$\mathbf{y} \in \mathbb{R}^m$
(observed)

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

$$\mathbf{A} \in \mathbb{R}^{m \times n}$$

linear input ($\mathbf{x}$)-output ($\mathbf{y}$) relation

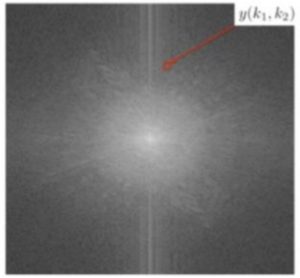
- **Convolution**: $\mathbf{A}\mathbf{x} \leftrightarrow h * \mathbf{X}$, h is a kernel



- **Masking**:

$$\mathbf{A} = \mathbf{M} \in \{0,1\}^{m \times n}$$

- **Fourier transform + Masking**:



$$\mathbf{A} \cdot = \mathbf{M}\mathcal{F}(\cdot)$$

Modeling noise

Acquisitions are never perfect. Interferences, errors, faults may happen.

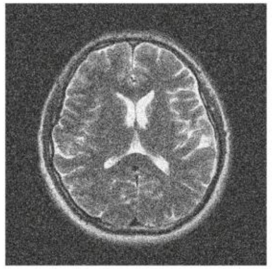
$$\mathbf{y} = \mathit{Noise}(\mathbf{A}\mathbf{x})$$

$\mathit{Noise}: \mathbb{R}^m \to \mathbb{R}^m$ codifies instrumental errors.

In the following: $\mathbf{A} = \mathbf{I} \in \mathbb{R}^{n \times n}$. How to model noise?

Gaussian noise:

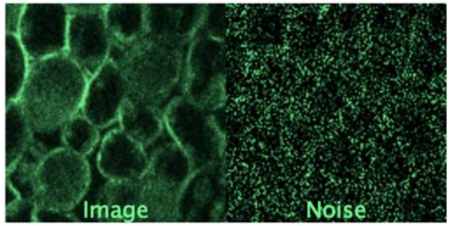
$$\mathit{Noise}(\mathbf{x}) = \mathbf{x} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$$



Mostly used due to CLT.
Models signal-independent electronic noise

Poisson noise:

$$\mathit{Noise}(\mathbf{x}) = \frac{1}{\gamma} \mathit{Pois}\,(\gamma \mathbf{x} + \boldsymbol{\beta}), \mathbf{x} \in \mathbb{R}^n_{\geq 0}, \boldsymbol{\beta} \in \mathbb{R}^n_{>0}$$



Used in low-photon imaging.
Astronomical, microscopy imaging.
Bertero, Boccacci, '98

Impulsive noise (salt and pepper):

$$\mathit{Noise}(\mathbf{x}) = (\mathbf{1} - \mathbf{s}) \odot \mathbf{x} + \mathbf{s} \odot \mathbf{c}$$

$$c_i = \mathcal{B}(1/2), s_i = \mathcal{B}(p), p \in [0,1]$$

Used to describe faulty detectors
and/or long time exposures under bad lighting.

Quality metrics: MSE, SNR, PSNR

How to assess image quality between two images using **pixel information**?

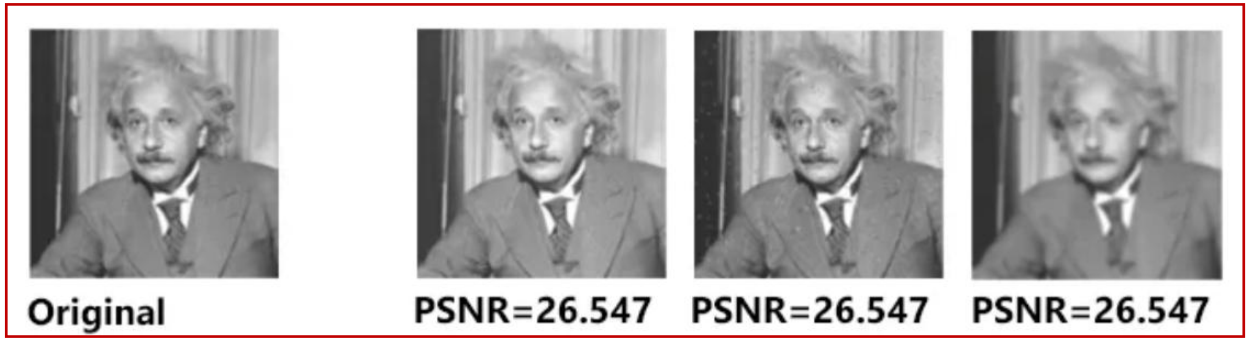
$$MSE(\mathbf{y},\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i|^2 = \frac{1}{n_1}\frac{1}{n_2}\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}|y_{i_1,i_2} - x_{i_1,i_2}|^2$$

$$SNR(\mathbf{y},\mathbf{x}) = 10\log_{10}\left(\frac{\|\mathbf{x}\|^2}{\|\mathbf{x}-\mathbf{y}\|^2}\right)$$

noise $\boldsymbol{\varepsilon}$

$$PSNR(\mathbf{y},\mathbf{x}) = 10\log_{10}\left(\frac{MAX^2}{MSE(\mathbf{y},\mathbf{x})}\right)$$

where MAX is the highest possible value (e.g., 255 or 1)



Sensitive to **pixel variations**

Is it a good quality metric for natural images?!

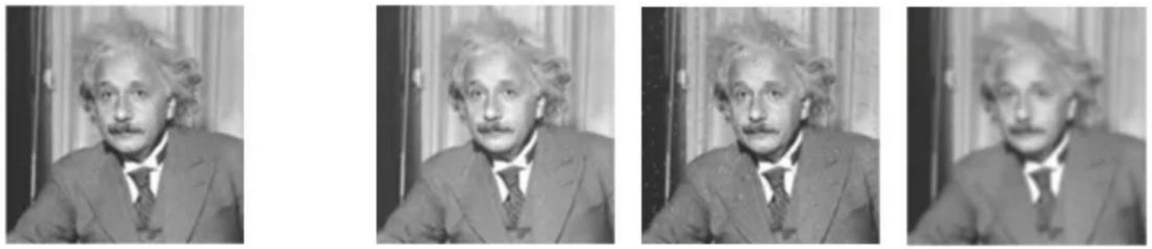
Quality metrics: SSIM

Wang, Bovik, Sheikh, Simoncelli, '04

$$SSIM(x, y) = \frac{(2\mu_{\mathbf{x}}\mu_{\mathbf{y}} + C_1)(2\sigma_{\mathbf{xy}} + C_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\mathbf{y}}^2 + C_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 + C_2)} \in [0,1]$$

- Based on image statistics: mean, variance, covariance + constant C_1, C_2 stabilising the division.

- Typically performed on small image patches + averaging



Original SSIM=1 | PSNR=26.547 SSIM=0.988 | PSNR=26.547 SSIM=0.840 | PSNR=26.547 SSIM=0.694

	RANGE	IDEAL	GOOD
MSE	$[0, MAX^2]$	0	< 100 (for images in [0,255])
SNR	$(-\infty, +\infty)$	$+\infty$	> 30
PSNR	$[0, +\infty)$	$+\infty$	> 30
SSIM	$[0,1]$	1	> 0.9

All such metrics are **supervised**. They depend on ground-truth $\mathbf{x}$.

Image denoising as a ‘toy’ inverse problem

Given $\mathbf{y} \in \mathbb{R}^n$, find $\mathbf{x} \in \mathbb{R}^n$ such that:

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$$

- No ill-posedness (operator to invert is $\mathbf{A} = \mathbf{I}$)
- Still challenging: the noise realization is unknown
- If noise is high (σ is big), image content can be lost

$\sigma = 0.002$ $\qquad$ $\sigma = 0.05$ $\qquad$ $\sigma = 0.3$

MSE →

← PSNR/SSIM

$\mathbf{x} \in [0,1]^{3n}$

How to tackle the **reconstruction** of x given y?

Bayesian formulation

Idea: model also $\mathbf{x}$ and $\mathbf{y}$ as realization of random variables $\mathcal{X} \sim \pi_{\mathcal{X}}, \mathcal{Y} \sim \pi_{\mathcal{Y}}$ + conditional laws.

$\pi_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x})$ **Likelihood (what we know): given x, what is the most likely y?**

$\pi_{\mathcal{X}|\mathcal{Y}}(\mathbf{x}|\mathbf{y})$ **Posterior (what we would like to know): given y, what is the most likely x?**

$\pi_{\mathcal{Y}}(\mathbf{y})$ **Evidence term:** normally neglected, does not depend on $\mathbf{x}$

$\pi_{\mathcal{X}}(\mathbf{x})$ **Prior:** prior assumptions on the unknown quantity $\mathbf{x}$

$$\pi_{\mathcal{X}|\mathcal{Y}}(\mathbf{x}|\mathbf{y}) = \frac{\pi_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x})\pi_{\mathcal{X}}(\mathbf{x})}{\pi_{\mathcal{Y}}(\mathbf{y})}$$

Bayes' Theorem

In denoising:

$$\pi_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x}) = \pi_E(\mathbf{y} - \mathbf{x})$$

Bayesian formulation: Maximum A Posteriori estimator

$$\mathbf{x}^* \in argmax_{\mathbf{x}} \pi_{\mathcal{X}|\mathcal{Y}}(\mathbf{x}|\mathbf{y}) = argmax_{\mathbf{x}} \frac{\pi_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x})\pi_{\mathcal{X}}(\mathbf{x})}{\pi_{\mathcal{Y}}(\mathbf{y})} = argmax_{\mathbf{x}} \pi_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x})\pi_{\mathcal{X}}(\mathbf{x})$$

By taking the negative logarithm:

$$\mathbf{x}^* \in argmin_{\mathbf{x}}\{-\ln(\pi_{\mathcal{X}|\mathcal{Y}}(\mathbf{x}|\mathbf{y}))\} = argmin_{\mathbf{x}}\{-\ln(\pi_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x})\pi_{\mathcal{X}}(\mathbf{x}))\} = argmin_{\mathbf{x}}\{-\ln(\pi_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x})) - \ln(\pi_{\mathcal{X}}(\mathbf{x}))\}$$

log-likelihood   log-prior

Example: $\pi_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x}) = \pi_E(\mathbf{y}-\mathbf{x})$, $\pi_E(\boldsymbol{\varepsilon}) = \mathcal{N}(0,\sigma_{\boldsymbol{\varepsilon}})$, $\pi_{\mathcal{X}}(\mathbf{x}) = \mathcal{N}(\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})$

$$\mathbf{x}^* \in argmin_{\mathbf{x}}\left\{-\ln\left(\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma_{\boldsymbol{\varepsilon}}^2}}\exp\left(-\frac{|y_i - x_i|^2}{2\sigma_{\boldsymbol{\varepsilon}}^2}\right)\right) - \ln\left(\frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}_{\boldsymbol{x}}|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_{\boldsymbol{x}})^\top\boldsymbol{\Sigma}_{\boldsymbol{x}}{}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\boldsymbol{x}})\right)\right)\right\}$$

$$= argmin_{\mathbf{x}}\left\{\frac{1}{2\sigma_{\boldsymbol{\varepsilon}}^2}\|\mathbf{y}-\mathbf{x}\|^2 + \frac{1}{2}\|\mathbf{x}-\boldsymbol{\mu}_{\boldsymbol{x}}\|_{\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}}^2\right\}$$ + neglected constants

—-> towards optimization problem!

From a statistical to a variational perspective

MAP estimation: find $\mathbf{x}^*$ s. t.
$$\mathbf{x}^* \in argmin_{\mathbf{x}}\{-\ln(\pi_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x})) - \ln(\pi_{\mathcal{X}}(\mathbf{x}))\}$$

⟹

Variational regularization: find $\mathbf{x}^*$ s. t.
$$\mathbf{x}^* \in argmin_{\mathbf{x}}\, J(\mathbf{x}), \qquad J(\mathbf{x}) = D(\mathbf{x},\mathbf{y}) + \lambda R(\mathbf{x})$$

- Describe the **likelihood distribution** $\pi_{\mathcal{Y}|\mathcal{X}}$ based on the (degradation and) noise model
  ⟹ select a **data fidelity** term $D(\mathbf{x},\mathbf{y})$
- Express prior knowledge on x through the **prior distribution** $\pi_{\mathcal{X}}$
  ⟹ select a **regularization** term $R(\mathbf{x})$
- Tune some (hyper)parameters in $\pi_{\mathcal{Y}|\mathcal{X}}$ and $\pi_{\mathcal{X}}$ to reflect noise intensity, confidence in the prior
  ⟹ adjust the **regularization parameter** λ

How to choose good noise and image models?

How to model statistically some prior knowledge on $\mathbf{x}$?
How to enforce good mathematical properties on $J\colon \mathbb{R}^n \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$?
(e.g. *log-concave* probability distributions lead to ***convex*** optimization problems)

Data terms

$\mathbf{y} = \mathit{Noise}(\mathbf{x})$ ⇢ $\pi_{y|x}(\mathbf{y}|\mathbf{x})$ ⇢ ($-\ln$, + constants) $D(\mathbf{y}, \mathbf{x})$

Tailored distance function with observation related to noise assumptions?

Gaussian noise:

$\mathit{Noise}(\mathbf{x}) = \mathbf{x} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{I})$ ⇢ $D(\mathbf{y}, \mathbf{x}) = D(\mathbf{y} - \mathbf{x}) = \frac{1}{2\sigma_{\boldsymbol{\varepsilon}}^2} \| \mathbf{y} - \mathbf{x} \|_2^2$

Poisson noise: (without background)

$\mathit{Noise}(\mathbf{x}) = \frac{1}{\gamma} \mathit{Pois}(\gamma \mathbf{x}), \mathbf{x} \in \mathbb{R}_{\geq 0}^n, \gamma > 0$ ⇢ $D(\mathbf{y}, \mathbf{x}) = D(\mathbf{y}, \mathbf{x}) = \mathit{KL}(\mathbf{y}, \mathbf{x}) = \sum_{i=1}^{n} x_i - y_i \ln(x_i)$

Impulsive noise (salt and pepper):

$\mathit{Noise}(\mathbf{x}) = (\mathbf{1} - \mathbf{s}) \odot \mathbf{x} + \mathbf{s} \odot \mathbf{c}$ ⇢ $D(\mathbf{y}, \mathbf{x}) = D(\mathbf{y} - \mathbf{x}) = \frac{1}{\tau_{\boldsymbol{\varepsilon}}} \| \mathbf{y} - \mathbf{x} \|_1$

$c_i = \mathcal{B}(1/2), s_i = \mathcal{B}(p), p \in [0,1]$

$\approx \mathbf{x} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim \mathcal{L}(\mathbf{0}, \tau_{\boldsymbol{\varepsilon}} \mathbf{I})$ (Laplace r.v.)

noise “sparsity” (the residual is 0 only in few pixels)

Regularisation terms

$$\pi_\mathcal{X}(\mathbf{x}) \xrightarrow[\text{+ constants}]{-\ln} R(\mathbf{x})$$

codify *a-priori* information

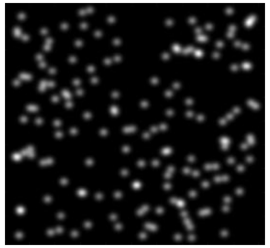
Small deviations from the mean:

$$\pi_\mathcal{X}(\mathbf{x}) = \mathcal{N}(\mu_\mathbf{x}, \boldsymbol{\Sigma}_x) \dashrightarrow R(\mathbf{x}) = \frac{1}{2} \| \mathbf{x} - \mu_\mathbf{x} \|^2_{\boldsymbol{\Sigma}_x^{-1}}$$

Tikhonov

Sparsity:

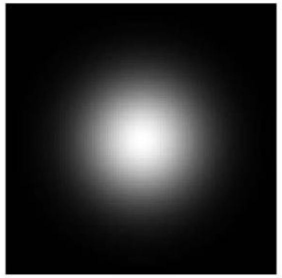
$$\pi_\mathcal{X}(\mathbf{x}) = \mathcal{L}(\mathbf{0}, \tau\mathbf{I}) \dashrightarrow R(\mathbf{x}) = \frac{1}{\tau} \| \mathbf{x} \|_1$$



Sparsity promotion
Compressed sensing

Smoothness (let $q_i = \| (\nabla\mathbf{x})_i \|_2$)

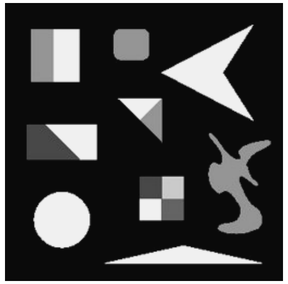
$$\pi_\mathcal{X}(\mathbf{x}) = \pi_Q(\mathbf{q}) = \mathcal{N}(\mathbf{0}, \sigma_\mathbf{q}^2\mathbf{I}) \dashrightarrow R(\mathbf{x}) = \frac{1}{2\sigma_\mathbf{q}^2} \| \nabla\mathbf{x} \|_2^2$$



Tikhonov/
Sobolev

Piece-wise constancy: (let $q_i = \| (\nabla\mathbf{x})_i \|_2$)

$$\pi_\mathcal{X}(\mathbf{x}) = \pi_Q(\mathbf{q}) = \mathcal{L}(\mathbf{0}, \tau\mathbf{I}) \dashrightarrow R(\mathbf{x}) = \frac{1}{\tau} TV(\mathbf{x}) = \frac{1}{\tau} \| \nabla\mathbf{x} \|_{2,1}$$



Total
Variation

The regularisation parameter

Example: $\pi_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x}) = \pi_E(\mathbf{y}-\mathbf{x}), \pi_{\mathcal{X}}(\mathbf{x}) = \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{x}}^2\mathbf{I})$

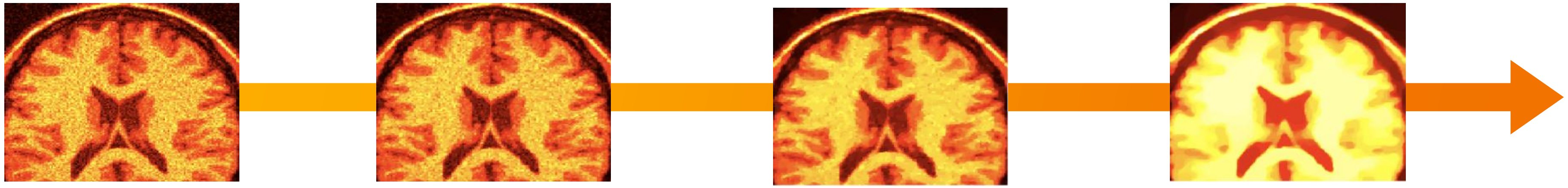
$$argmin_{\mathbf{x}} \frac{1}{2\sigma_\varepsilon^2} \| \mathbf{y}-\mathbf{x} \|^2 + \frac{1}{2\sigma_x^2} \| \mathbf{x} \|_2^2 = argmin_{\mathbf{x}} \| \mathbf{y}-\mathbf{x} \|^2 + \frac{\sigma_\varepsilon^2}{\sigma_x^2} \| \mathbf{x} \|^2 = argmin_{\mathbf{x}} \| \mathbf{y}-\mathbf{x} \|^2 + \lambda \| \mathbf{x} \|^2$$

Ratio between noise level (*can be estimated*) and image statistical features (*hard to estimate*).

More in general, the regularization parameter $\lambda > 0$ weights data fit against regularization.

$$argmin_{\mathbf{x}} D(\mathbf{y},\mathbf{x}) + \lambda R(\mathbf{x})$$

- Small λ: low regularization, trust in the data, noise overfit
- High λ: strong regularization, noise removal, artefacts induced by R



$\lambda \to +\infty$

Examples

$$\mathbf{y} = Noise(\mathbf{x})$$

Gaussian noise + piece-wise constant image: reference model for many applications.

$$argmin_{\mathbf{x}} \frac{1}{2} \| \mathbf{y} - \mathbf{x} \|_2^2 + \lambda TV(\mathbf{x})$$

Poisson noise + sparse signal: used in microscopy/astronomical imaging

$$argmin_{\mathbf{x} \geq 0} KL(\mathbf{y}, \mathbf{x} + \boldsymbol{\beta}) + \lambda \| \mathbf{x} \|_1$$

Gaussian noise + Tikhonov-type regularisation: often used when $\mathbf{A} \neq \mathbf{I}$ for applications

$$argmin_{\mathbf{x}} \frac{1}{2} \| \mathbf{y} - \mathbf{x} \|_2^2 + \frac{\lambda}{2} \| \mathbf{L}\mathbf{x} \|_2^2 \quad \mathbf{L} \in \mathbb{R}^{d \times n}$$

Solving the problem: nods on optimisation

$argmin_{\mathbf{x}} J(\mathbf{x}) := D(\mathbf{y}, \mathbf{x}) + \lambda R(\mathbf{x})$

variational formulation of the image denoising problem

$J: \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0} \cup \{+\infty\}$ is proper

$$dom(J) = \{\mathbf{x}: J(\mathbf{x}) < +\infty\} \neq \emptyset$$

J is L-smooth, i.e. has L-Lipschitz (Gâteaux) gradient:

$$\exists L > 0: \| \nabla J(\mathbf{x}_1) - \nabla J(\mathbf{x}_2) \|_2 \leq L \| \mathbf{x}_1 - \mathbf{x}_2 \|_2$$

$$\Longleftrightarrow$$

$$J(\mathbf{x}_1) \leq J(\mathbf{x}_2) + \langle \nabla J(\mathbf{x}_2), \mathbf{x}_2 - \mathbf{x}_1 \rangle + \frac{L}{2} \| \mathbf{x}_2 - \mathbf{x}_1 \|_2^2$$

under convexity

J is convex:

$$(\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n), (\forall \alpha \in [0,1]): J(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leq \alpha J(\mathbf{x}_1) + (1-\alpha) J(\mathbf{x}_2)$$

Solving the problem: nods on optimisation

$argmin_{\mathbf{x}} J(\mathbf{x}) := D(\mathbf{y}, \mathbf{x}) + \lambda R(\mathbf{x})$

J is proper, convex, L-smooth and coercive.

$\lim_{\|\mathbf{x}\| \to +\infty} J(\mathbf{x}) = +\infty$

Theorem

There exists a minimizer for J. All local minimizers are global minimizers.
For all $\mathbf{x}^* \in Argmin_{\mathbf{x}} J(\mathbf{x})$, there holds $\nabla J(\mathbf{x}^*) = \mathbf{0}$.

Algorithm (gradient descent): for $\mathbf{x}_0 \in dom(J), \tau \in \left(0, \frac{2}{L}\right), k \geq 0$:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \tau \nabla J(\mathbf{x}_k)$$

Theorem

There holds $\mathbf{x}_k \to \mathbf{x}^*$ and for the function values: $J(\mathbf{x}_k) - J(\mathbf{x}^*) \leq \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|^2}{2\tau k}$.

Example on Tikhonov regularisation

$$argmin_{\mathbf{x}} J(\mathbf{x}) := \frac{1}{2} \| \mathbf{y} - \mathbf{x} \|_2^2 + \lambda \| \mathbf{L}\mathbf{x} \|_2^2, \mathbf{L} \in \mathbb{R}^{d \times n}$$

Examples: $\mathbf{L} \in \{\mathbf{I}, \nabla, \nabla^2\}$

Remark: the problem is quadratic, it can be solved by looking at the optimality condition:

$$(\mathbf{x}^* - \mathbf{y}) + \lambda \mathbf{L}^T \mathbf{L} \mathbf{x}^* = \mathbf{0} \Rightarrow (\mathbf{I} + \lambda \mathbf{L}^T \mathbf{L}) \mathbf{x}^* = \mathbf{y}$$

and solving the linear system, e.g., using DFT. Also, faster iterative methods exploiting further regularity (strong convexity, C^2) can be employed.

Hansen, Nagy, O'leary, '06, Nesterov, '83

$$\nabla J(\mathbf{x}) = (\mathbf{x} - \mathbf{y}) + \lambda \mathbf{L}^T \mathbf{L}\mathbf{x}, L = 1 + \lambda \| \mathbf{L}^T \mathbf{L} \|_*, \mathbf{x}_0 \in \mathbb{R}^n, \tau \in (0, 2/L)$$

```
while not converging
```

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \tau\big((\mathbf{x}_k - \mathbf{y}) + \lambda \mathbf{L}^T \mathbf{L} \mathbf{x}_k\big)$$

```
end
```

Image denoisers and proximal operators

For general (possibly non-smooth) regularisation functionals $R: \mathbb{R}^n \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$, note that:

$$argmin_{\mathbf{x}} \frac{1}{2} \| \mathbf{y} - \mathbf{x} \|^2 + \lambda R(\mathbf{x}) =: prox_{\lambda R}(\mathbf{y})$$

where $prox_{\lambda R}: \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is single-valued if R is convex and multi-valued (multiple minimisers) otherwise.

Examples:

$R(\mathbf{x}) = \iota_C(\mathbf{x}), C$ is convex and closed. $\quad prox_{\iota_C}(\mathbf{y}) = \mathrm{proj}_C(\mathbf{y})$

$R(\mathbf{x}) = \| \mathbf{x} \|_1 \quad prox_{\lambda \|\cdot\|_1}(\mathbf{y}) = SoftThreshold(\mathbf{y}; \lambda)$

$R(\mathbf{x}) = TV(\mathbf{x}) \quad prox_{\lambda TV(\cdot)}(\mathbf{y})?$

Non-smooth functionals (∇R not defined)

Proximal operators are widely used as implicit variants of gradients for non-smooth optimization:

$$\mathbf{x}_{k+1} = prox_{\tau\lambda R}(\mathbf{x}_k - \tau \nabla_{\mathbf{x}} D(\mathbf{y}, \mathbf{A}\mathbf{x}_k))$$

→ Gradient-descent on data term + denoising

This observation stands at the very basis of Plug & Play approaches where $prox_{\tau\lambda R}$ → $\mathcal{D}_\varsigma$

Kamilov, Bouman, Buzzard, Wuhlberg, ‘23

References

J. Kaipio, E. Somersalo, *Statistical and computational inverse problems*, Springer, 2005.

Tony F. Chan, J. Shen, *Image Processing and Analysis: variational, PDE, wavelet and stochastic methods*, SIAM, 2005.

M. Pragliola, L. Calatroni, A. Lanza, F. Sgallari, *On and beyond Total Variation in imaging: the role of space variance*, SIAM Review, 65 (3), (2023).

S. Arridge, P. Maas, O. Öktem, C.B. Schönlieb, *Solving inverse problems using data-driven models*, Acta Numerica, 2019

A. Chambolle, T. Pock, *An introduction to continuous optimization for imaging*, Acta Numerica, 2016

A. Beck, *First-order methods in optimization*, Volume 25, MOS-SIAM series on Optimization, 2017.