Inverse problem and optimization

Laurent Condat, Nelly Pustelnik

CNRS, Gipsa-lab

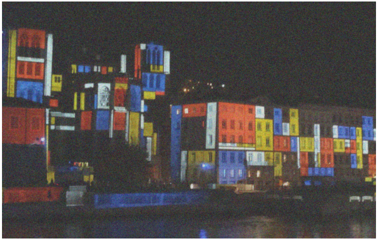
CNRS, Laboratoire de Physique de l'ENS de Lyon

Decembre, 15th 2016

Plan

1. Examples of inverse problems
2. Direct model: linear operator, noise
3. Ill-posed problem
4. Inversion
5. Regularization
6. Gradient descent

Example: denoising



Observations

Denoised image

Example: denoising

Example: denoising

Example: motion blur

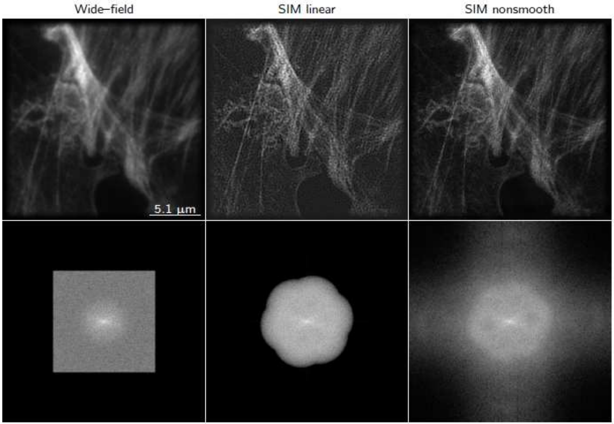
Observations

Restored image

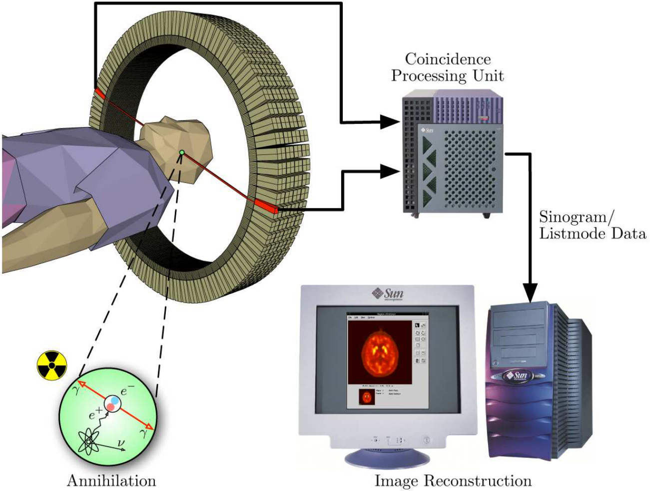
Example: motion blur

Example: motion blur

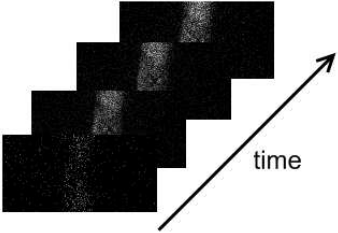
Example: structured illumination microscopy



Example: positron emission tomography

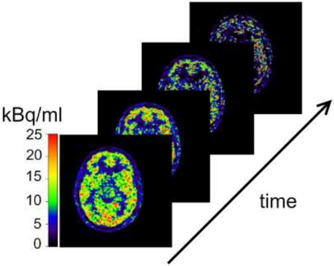


Example: positron emission tomography



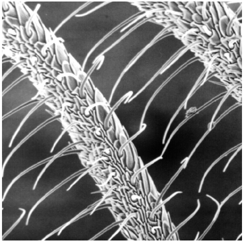
Measurements

?



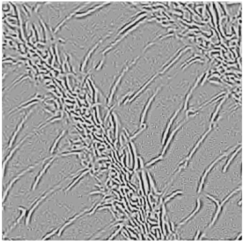
Reconstructed images

Example: cartoon-texture decomposition

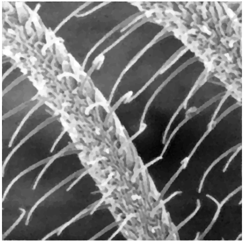


Observations

$\xrightarrow{?}$



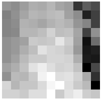
Texture



Cartoon

Notations

- Image $x \in \mathbb{R}^{N_1 \times N_2}$



$$x = \left(x_{n_1,n_2}\right)_{1 \le n_1 \le N_1, 1 \le n_2 \le N_2}$$

- Vector consisting of the values of the image of size $N = N_1 \times N_2$ arranged column-wise $x \in \mathbb{R}^N$ (with $N = N_1 \times N_2$)

$$x = \left(x_n\right)_{1 \le n \le N}$$

Direct model

$$\boxed{z = \mathcal{D}_\alpha(H\overline{x})}$$

- $\overline{x} = (\overline{x}_n)_{1 \leq n \leq N} \in \mathbb{R}^N$: vector consisting of the (unknown) values of the original image of size $N = N_1 \times N_2$.
- $z = (z_j)_{1 \leq j \leq M} \in \mathbb{R}^M$: vector containing the observed values of size $M = M_1 \times M_2$.
- $H \in \mathbb{R}^{M \times N}$: matrix associated to a linear degradation operator.
- $\mathcal{D}_\alpha \colon \mathbb{R}^M \to \mathbb{R}^M$: models other degradations such as nonlinear ones or the effect of the noise, parameterized by α (e.g. additive noise with variance α, Poisson noise with scaling parameter α).

Direct model

$$\boxed{z = \mathcal{D}_\alpha(H\overline{x})}$$

- $\overline{x} = (\overline{x}_n)_{1 \leq n \leq N} \in \mathbb{R}^N$: vector consisting of the (unknown) values of the original image of size $N = N_1 \times N_2$.
- $z = (z_j)_{1 \leq j \leq M} \in \mathbb{R}^M$: vector containing the observed values of size $M = M_1 \times M_2$.
- $H \in \mathbb{R}^{M \times N}$: matrix associated to a linear degradation operator.
- $\mathcal{D}_\alpha \colon \mathbb{R}^M \to \mathbb{R}^M$: models other degradations such as nonlinear ones or the effect of the noise, parameterized by α (e.g. additive noise with variance α, Poisson noise with scaling parameter α).

Direct model: convolution

$$\boxed{z = \mathcal{D}_\alpha(H\overline{x})} \quad \Rightarrow \quad z = \mathcal{D}_\alpha(h * \overline{x})$$

where $\{h * \overline{x}\}$: convolution product with the Point Spread Function (PSF) h of size $Q_1 \times Q_2$.

Link between h and H:
Under zero-end conditions,

- unknown image $\overline{x}$ is zero outside its domain $[0, N_1 - 1] \times [0, N_2 - 1]$,
- kernel h is zero outside its domain $[0, Q_1 - 1] \times [0, Q_2 - 1]$.

Direct model: convolution

Link between h and H:
Zero padding of $\overline{x}$ and h: extended image $\overline{x}^{\mathrm{e}}$ and kernel h^{e} of size $M_1 \times M_2$ as

$$\overline{x}^{\mathrm{e}}_{i_1,i_2} = \begin{cases} \overline{x}_{i_1,i_2} & \text{if } 0 \leq i_1 \leq N_1 - 1 \text{ and } 0 \leq i_2 \leq N_2 - 1 \\ 0 & \text{if } N_1 \leq i_1 \leq M_1 - 1 \text{ and } N_2 \leq i_2 \leq M_2 - 1, \end{cases}$$

$$h^{\mathrm{e}}_{i_1,i_2} = \begin{cases} h_{i_1,i_2} & \text{if } 0 \leq i_1 \leq Q_1 - 1 \text{ and } 0 \leq i_2 \leq Q_2 - 1 \\ 0 & \text{if } Q_1 \leq i_1 \leq M_1 - 1 \text{ and } Q_2 \leq i_2 \leq M_2 - 1, \end{cases}$$

This yields to

$$(H\overline{x})_{j_1,j_2} = \sum_{i_1=0}^{M_1-1} \sum_{i_2=0}^{M_2-1} h^{\mathrm{e}}_{j_1-i_1,j_2-i_2} \overline{x}^{\mathrm{e}}_{i_1,i_2}$$

where $j_1 \in \{0, \dots, M_1 - 1\}$ and $j_2 \in \{0, \dots, M_2 - 1\}$.

Direct model: convolution

Link between h and H:

$$H = \begin{bmatrix} \tilde{H}_0 & \tilde{H}_{M_1-1} & \tilde{H}_{M_1-2} & \cdots & \tilde{H}_1 \\ \tilde{H}_1 & \tilde{H}_0 & \tilde{H}_{M_1-1} & \cdots & \tilde{H}_2 \\ \tilde{H}_2 & \tilde{H}_1 & \tilde{H}_0 & \cdots & \tilde{H}_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{H}_{M_1-1} & \tilde{H}_{M_1-2} & \tilde{H}_{M_1-3} & \cdots & \tilde{H}_0 \end{bmatrix}.$$

where $\tilde{H}_{j_1}$ denotes a circulant matrix with M_2 columns such that

$$\tilde{H}_{j_1} = \begin{bmatrix} h^{\mathrm{e}}_{j_1,0} & h^{\mathrm{e}}_{j_1,M_2-1} & h^{\mathrm{e}}_{j_1,M_2-2} & \cdots & h^{\mathrm{e}}_{j_1,1} \\ h^{\mathrm{e}}_{j_1,1} & h^{\mathrm{e}}_{j_1,0} & h^{\mathrm{e}}_{j_1,M_2-1} & \cdots & h^{\mathrm{e}}_{j_1,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h^{\mathrm{e}}_{j_1,M_2-1} & h^{\mathrm{e}}_{j_1,M_2-2} & h^{\mathrm{e}}_{j_1,M_2-3} & \cdots & h^{\mathrm{e}}_{j_1,0} \end{bmatrix} \in \mathbb{R}^{M_2 \times M_2}.$$

Direct model: convolution

If H is a block-circulant matrix with circulant blocks, then

$$\boxed{H = U^* D U}$$

where

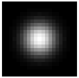
- D: diagonal matrix,
- U: unitary matrix (i.e, $U^* = U^{-1}$) representing the discrete Fourier transform,
- $\cdot^*$ denotes here the transpose conjugate.

Efficient computation of $H\overline{x}$:

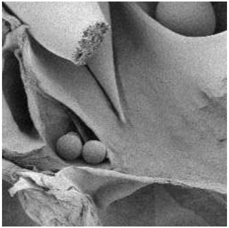
$$\begin{aligned} H\overline{x} &= U^* D U (U^* U)\overline{x} \\ &= U^* D \overline{X} \end{aligned}$$

where $\overline{X}$ denotes the Fourier transform of $\overline{x}$.

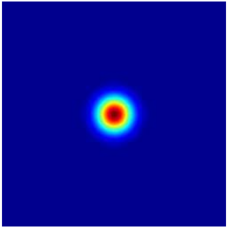
Direct model: convolution



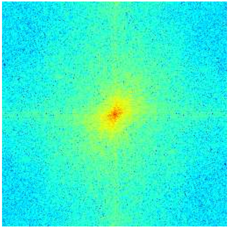
Gaussian filter h



Original image $\overline{x}$



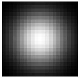
$\mathcal{F}(h)$



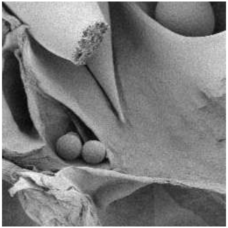
$\mathcal{F}(\overline{x})$

$\{h * \overline{x}\}$

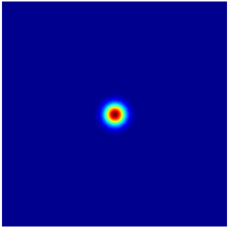
Direct model: convolution



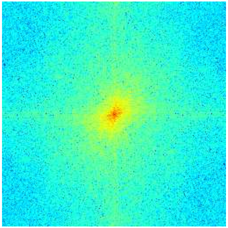
Gaussian filter h



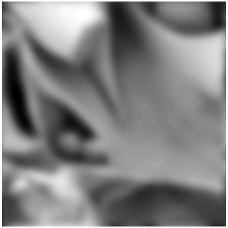
Original image $\overline{x}$



$\mathcal{F}(h)$

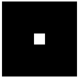


$\mathcal{F}(\overline{x})$

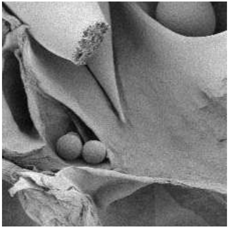


$\{h * \overline{x}\}$

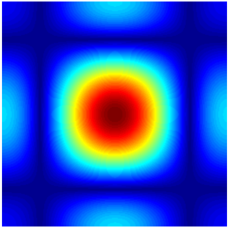
Direct model: convolution



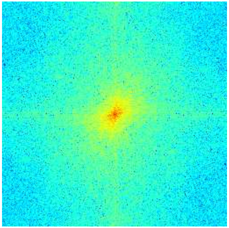
Uniform filter h



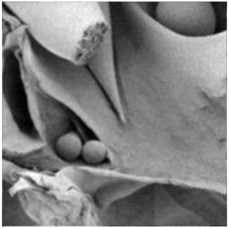
Original image $\overline{x}$



$\mathcal{F}(h)$

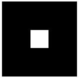


$\mathcal{F}(\overline{x})$

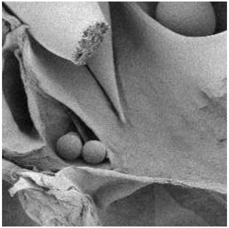


$\{h * \overline{x}\}$

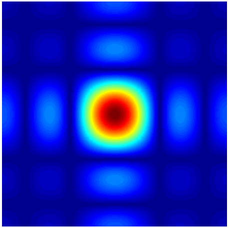
Direct model: convolution



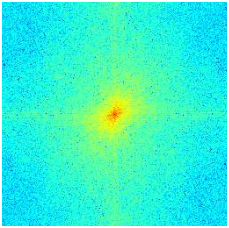
Uniform filter h



Original image $\overline{x}$



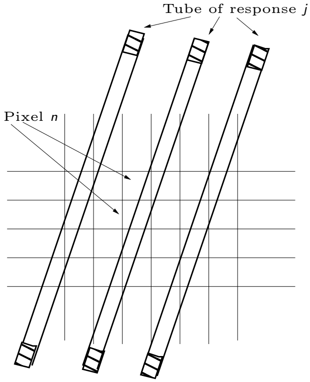
$\mathcal{F}(h)$



$\mathcal{F}(\overline{x})$

$\{h * \overline{x}\}$

Direct model: tomography (without noise)

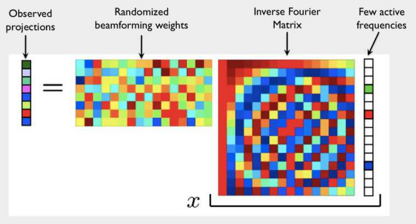


$$\boxed{z = H\overline{x}}$$

- $\overline{x} = (\overline{x}_n)_{1 \leq n \leq N} \in \mathbb{R}^N$: vector consisting of the (unknown) values of the original image of size $N = N_1 \times N_2$.
- $H = (H_{ji})_{1 \leq j \leq M, 1 \leq n \leq N}$: probability to detect an event in the tube/line of response.
- $z = (z_j)_{1 \leq j \leq M} \in \mathbb{R}^M$: vector containing the observed values (sinogram).

Direct model: compressed sensing

$$\boxed{z = H\overline{x}}$$



- $\overline{x} = (\overline{x}_n)_{1\leq n\leq N} \in \mathbb{R}^N$: vector consisting of the (unknown) values of the original image of size $N = N_1 \times N_2$.
- $z = (z_j)_{1\leq j\leq M} \in \mathbb{R}^M$: vector containing the observed values (size $M \ll N$).
- $H = (H_{ji})_{1\leq j\leq M, 1\leq j\leq N}$: random measurement matrix (size $M \times N$).

Direct model: super-resolution

$$(\forall b \in \{1, \ldots, B\}) \quad \boxed{z_b = \mathrm{D}_b \mathrm{T} \mathrm{W}\, \overline{x} + \varepsilon_b}$$

- z: B multicomponent images at low-resolution (size M),
- $\overline{\mathrm{x}}$: (high-resolution) image to be recovered (size N),
- D_b: downsampling matrix (size $M \times N$ such that $M < N$),
- T: matrix associated to the blur (size $N \times N$),
- W: warp matrix (size $N \times N$),
- $\varepsilon_b \sim \mathcal{N}(0, \sigma^2 \mathrm{Id}_K)$: noise often assumed to be a zero-mean white Gaussian additive noise.

Direct model

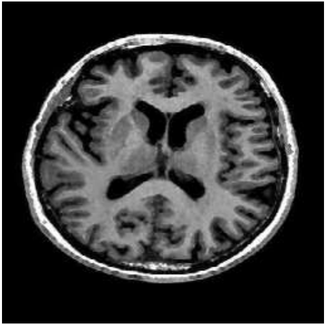
$$\boxed{z = \mathcal{D}_\alpha(H\overline{x})}$$

- $\overline{x} = (\overline{x}_n)_{1 \le n \le N} \in \mathbb{R}^N$: vector consisting of the (unknown) values of the original image of size $N = N_1 \times N_2$.
- $z = (z_j)_{1 \le j \le M} \in \mathbb{R}^M$: vector containing the observed values of size $M = M_1 \times M_2$.
- $H \in \mathbb{R}^{M \times N}$: matrix associated to a linear degradation operator
- $\mathcal{D}_\alpha \colon \mathbb{R}^M \to \mathbb{R}^M$: models other degradations such as nonlinear ones or the effect of the noise, parameterized by α (e.g. additive noise with variance α, Poisson noise with scaling parameter α)

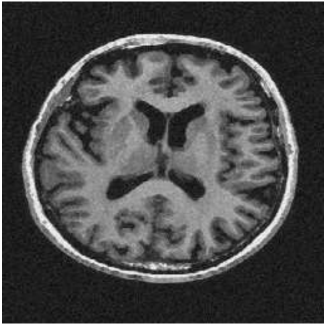
Direct model: Gaussian noise

$$\boxed{z = \mathcal{D}_\alpha(H\overline{x})} \quad \Rightarrow \quad z = H\overline{x} + b$$

where b: white additive Gaussian noise with variance $\alpha = \sigma^2$.



Original image

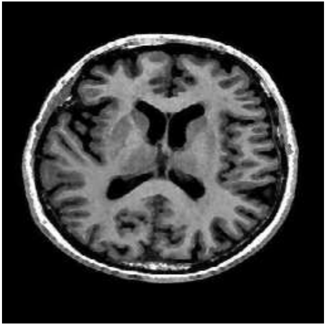


Degraded image with $\sigma = 10$

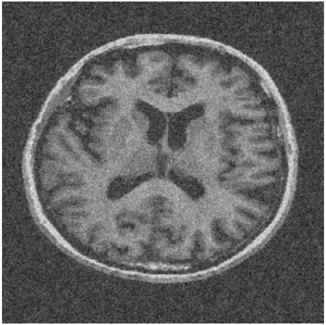
Direct model: Gaussian noise

$$\boxed{z = \mathcal{D}_\alpha(H\overline{x})} \quad \Rightarrow \quad z = H\overline{x} + b$$

where b: white additive Gaussian noise with variance $\alpha = \sigma^2$.



Original image

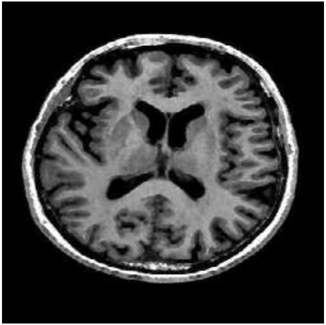


Degraded image $\sigma = 30$

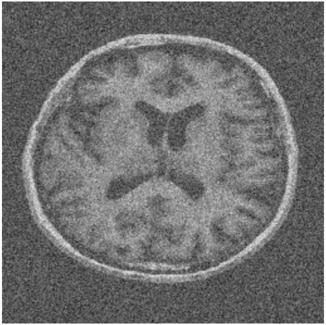
Direct model: Gaussian noise

$$\boxed{z = \mathcal{D}_\alpha(H\overline{x})} \quad \Rightarrow \quad z = H\overline{x} + b$$

where b: white additive Gaussian noise with variance $\alpha = \sigma^2$.



Original image

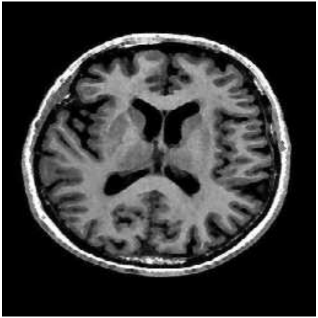


Degraded image $\sigma = 50$

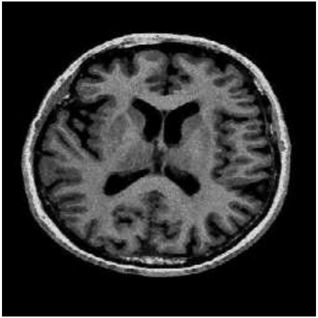
Direct model: Poisson noise

$$\boxed{z = \mathcal{D}_\alpha(H\overline{x})}$$

where $\mathcal{D}_\alpha$: Poisson noise with scaling parameter α
$\Rightarrow$ noise variance varies with image intensity.



Original image

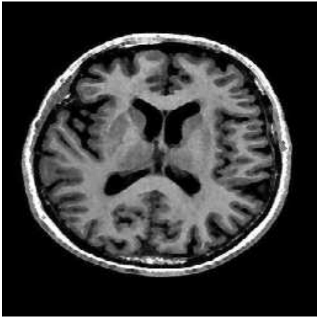


Poisson noise $\alpha = 1$

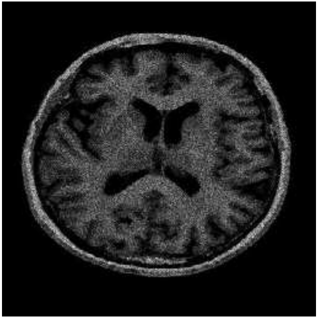
Direct model: Poisson noise

$$\boxed{z = \mathcal{D}_\alpha(H\overline{x})}$$

where $\mathcal{D}_\alpha$: Poisson noise with scaling parameter α
$\Rightarrow$ noise variance varies with image intensity.



Original image

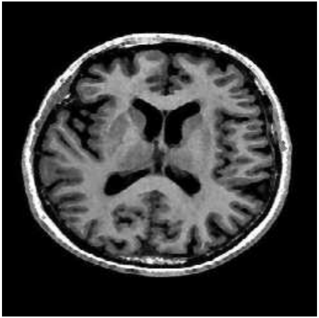


Poisson noise $\alpha = 0.1$

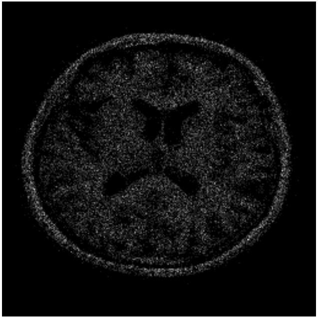
Direct model: Poisson noise

$$\boxed{z = \mathcal{D}_\alpha(H\overline{x})}$$

where $\mathcal{D}_\alpha$: Poisson noise with scaling parameter α
$\Rightarrow$ noise variance varies with image intensity.



Original image

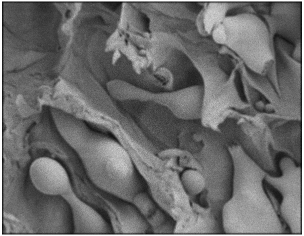


Poisson noise $\alpha = 0.01$

Inverse problem

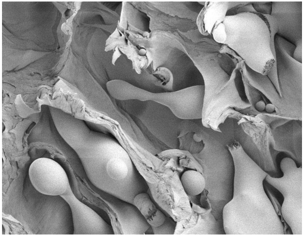
Inverse problem :
Find $\widehat{x}$ the closest from $\overline{x}$ from observations

$$z = \mathcal{D}_\alpha(H\overline{x})$$



Observations $z \in \mathbb{R}^M$

$\xrightarrow{?}$



Restored image $\widehat{x} \in \mathbb{R}^N$

Hadamard conditions

The problem $z = H\overline{x}$ is said to be well-posed if it fulfills the Hadamard conditions (1902)

1. existence of a solution,

   i.e. the range $\mathrm{ran}\, H$ of H is equal to $\mathbb{R}^M$,

2. uniqueness of the solution,

   i.e. the nullspace $\ker H$ of H is equal to $\{0\}$,

3. stability of the solution $\widehat{x}$ relatively to the observation,

   i.e. $\big(\forall (z, z') \in \big(\mathbb{R}^M\big)^2\big)$

$$\|z - z'\| \to 0 \quad \Rightarrow \quad \|\widehat{x}(z) - \widehat{x}(z')\| \to 0.$$

Hadamard conditions

The problem $z = H\overline{x}$ is said to be well-posed if it fulfills the Hadamard conditions

1. existence of a solution,

   i.e. every vector z *in* $\mathbb{R}^M$ *is the image of a vector* x *in* $\mathbb{R}^N$,

2. uniqueness of the solution,

   i.e. if $\widehat{x}(z)$ *and* $\widehat{x}'(z)$ *are two solutions, then they are necessarily equal since* $\widehat{x}(z) - \widehat{x}'(z)$ *belongs to* $\ker A$,

3. stability of the solution $\widehat{x}$ relatively to the observation,

   i.e. ensure that a small perturbation of the observed image leads to a slight variation of the recovered image.

Inversion

Inverse filtering (if $M = N$ et H est inversible)

$$\begin{aligned}\widehat{x} &= H^{-1}z \\ &= H^{-1}(H\overline{x} + b) \quad \text{if additive noise } b \in \mathbb{R}^M \\ &= \overline{x} + H^{-1}b\end{aligned}$$

Remark :

$\rightarrow$ Closed form expression but noise amplification if H ill-conditioned
(*ill-posed problem*).

Inversion

Inverse filtering (if $M \geq N$ and rank of H is N)

$$\begin{aligned} \widehat{x} &= (H^* H)^{-1} H^\top z \\ &= (H^* H)^{-1} H^* (H\overline{x} + b) \quad \text{if additive noise } b \in \mathbb{R}^M \\ &= \overline{x} + (H^* H)^{-1} H^* b \end{aligned}$$

Remark :

$\rightarrow$ Closed form expression but noise amplification if H ill-conditioned
(*ill-posed problem*).

Regularization

Variational approach

$$\boxed{\widehat{x} \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} \; \|z - Hx\|_2^2 + \lambda \Omega(x)}$$

where

- $\|z - Hx\|_2^2$: data-term,
- $\Omega(x)$: regularization term (e.g. $\Omega(x) = \|x\|_2^2$),
- $\lambda \geq 0$: regularization parameter.

Remarks

$\rightarrow$ If $\lambda = 0$: inverse filtering,

Maximum A Posteriori (MAP)

Let x and z be random vector realizations X and Z.
Maximum A Posteriori (MAP)

$$\hat{x} \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmax}}\, \mu_{X|Z=z}(x)$$

$\rightarrow$ find x that maximizes the posterior $\mu_{X|Z=z}(x)$

Maximum A Posteriori (MAP)

Let x and z be random vector realizations X and Z.
Maximum A Posteriori (MAP)

$$\hat{x} \in \underset{x \in \mathbb{R}^N}{\text{Argmax}}\ \mu_{X|Z=z}(x)$$

$\rightarrow$ find x that maximizes the posterior $\mu_{X|Z=z}(x)$

Bayes rule

$$\begin{aligned}
\max_{x \in \mathbb{R}^N} \mu_{X|Z=z}(x) &\Leftrightarrow \max_{x \in \mathbb{R}^N} \mu_{Z|X=x}(z) \cdot \mu_X(x) \\
&\Leftrightarrow \min_{x \in \mathbb{R}^N} \big\{ -\log(\mu_{Z|X=x}(z)) - \log(\mu_X(x)) \big\}
\end{aligned}$$

Maximum A Posteriori (MAP)

Let x and z be random vector realizations X and Z.
Maximum A Posteriori (MAP)

$$\hat{x} \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmax}}\ \mu_{X|Z=z}(x)$$

$\rightarrow$ find x that maximizes the posterior $\mu_{X|Z=z}(x)$

Bayes rule

$$\begin{aligned}
\max_{x \in \mathbb{R}^N} \mu_{X|Z=z}(x) &\Leftrightarrow \max_{x \in \mathbb{R}^N} \mu_{Z|X=x}(z) \cdot \mu_X(x) \\
&\Leftrightarrow \min_{x \in \mathbb{R}^N} \Big\{ \underbrace{-\log(\mu_{Z|X=x}(z))}_{\text{Data-term}} \underbrace{-\log(\mu_X(x))}_{\text{A priori}} \Big\} \\
&\Leftrightarrow \min_{x \in \mathbb{R}^N} f_1(x) + f_2(x)
\end{aligned}$$

Data-term: Gaussian noise

$$
\boxed{(\forall x \in \mathbb{R}^N) \quad f_1(x) = -\log(\mu_{Z|X=x}(z))}
$$

- Let $z = H\overline{x} + b$ with $b \sim \mathcal{N}(0, \alpha)$
- Gaussian likelihood:

$$
\mu_{Z|X=x}(z) = \prod_{i=1}^{M} \frac{1}{\sqrt{2\pi\alpha}} \exp\left(\frac{((Hx)^{(i)} - z^{(i)})^2}{2\alpha}\right)
$$

- Data-term:

$$
f_1(x) = \sum_{i=1}^{M} \frac{1}{2\alpha}((Hx)_i - z_i)^2
$$

Data-term: Poisson noise

$$\boxed{(\forall x \in \mathbb{R}^N) \quad f_1(x) = -\log(\mu_{Z|X=x}(z))}$$

- Let $z = \mathcal{D}_\alpha(H\overline{x})$ where $\mathcal{D}_\alpha$ Poisson noise with parameter α.

- Poisson likelihood:

$$\mu_{Z|X=x}(z) = \prod_{i=1}^{M} \frac{\exp\big(-\alpha(Hx)^{(i)}\big)}{z^{(i)}!} \big(\alpha(Hx)^{(i)}\big)^{z^{(i)}}$$

- Data-term: $f_1(x) = \sum_{i=1}^{M} \Psi_i\big((Hx)^{(i)}\big)$

$$(\forall v \in \mathbb{R}) \quad \Psi_i(v) = \begin{cases} \alpha v - z^{(i)} \ln(\alpha v) & \text{if } z^{(i)} > 0 \text{ and } v > 0, \\ \alpha v & \text{si } z^{(i)} = 0 \text{ and } v \geq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Prior: Tikhonov /TV

$$(\forall x \in \mathbb{R}^N) \quad f_2(x) = -\log(\mu_X(x))$$

- Tikhonov [Tikhonov, 1963]

$$f_2(x) = \|Lx\|^2$$
$$= \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \{l * x\}_{i,j}^2 \quad \text{avec} \quad l = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Minimisation problem

Variational approach

$$\boxed{\widehat{x} \in \underset{x \in \mathbb{R}^N}{\text{Argmin}}\ \mathcal{L}(x) + \lambda \Omega(x)}$$

where

- $\mathcal{L}(x) = f(z, Hx)$: data-term,
- $\Omega(x)$: regularization term,
- $\lambda \geq 0$: regularization parameter.

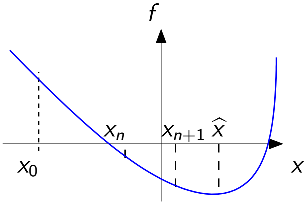
Convex optimization ?

Let $f : \mathbb{R}^N \to]-\infty, +\infty]$.
An optimization problem consists in solving:

$$\boxed{\widehat{x} \in \underset{x \in \mathbb{R}^N}{\mathrm{Argmin}}\ f(x)}$$

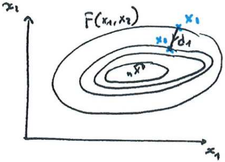
- $\widehat{x}$ is a global solution if for every $x \in \mathbb{R}^N$, $f(\widehat{x}) \leq f(x)$,

Course objectif: Build a sequence $(x_n)_{n \in \mathbb{Z}}$ that converges to $\widehat{x}$.



Gradient descent

- Illustration:



- Remarks:
  - f is displayed with its level lines,
  - Fermat rule:

$$\widehat{x} \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} f(x) \Leftrightarrow \nabla f(\widehat{x}) = 0$$

  - Solve a problem with N equations and N unknowns.
  - Closed form expression for very few f.
  - If no closed form expression possible $\Rightarrow$ iterative method.

Solving mean square problem

- Find

$$\widehat{x} \in \operatorname*{Argmin}_{x \in \mathbb{R}^N} \|Hx - z\|_2^2 \quad \text{with} \quad \begin{cases} H \in \mathbb{R}^{M \times N} \\ z \in \mathbb{R}^M \end{cases}$$

- Optimality conditions

$$\begin{aligned} \nabla f(\widehat{x}) = 0 \quad & \Leftrightarrow \quad H^*(H\widehat{x} - z) = 0 \\ & \Leftrightarrow \quad \widehat{x} = (H^* H)^{-1} H^* z \end{aligned}$$

- Difficulty: invert $H^* H$.

Solving logistic regression

- Find

$$\widehat{x} \in \underset{x\in\mathbb{R}}{\mathrm{Argmin}} \log(1+\exp(-yx)) \quad \text{with} \quad y \in \mathbb{R}$$

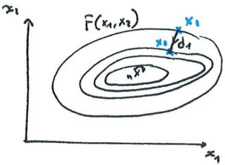
- Optimality conditions

$$\nabla f(\widehat{x}) = 0 \quad \Leftrightarrow \quad \frac{-y\exp(-y\widehat{x})}{1+\exp(-y\widehat{x})} = 0$$

- Difficulty: no closed form expression.

Gradient descent

- Illustration:



- Iterations:

$$(\forall k \in \mathbb{N}) \quad x^{[k+1]} = x^{[k]} + \gamma^{[k]} d^{[k]}$$

where

$$\begin{cases} d^{[k]} \in \mathbb{R}^N\text{: descent direction,} \\ \gamma^{[k]} > 0\text{: step-size.} \end{cases}$$