Inverse problems

Hervé CARFANTAN

Laboratoire d'Astrophysique de l'Observatoire Midi Pyrénées

Herve.Carfantan@obs-mip.fr

How to infer on parameters of interest from the observations ?

I Inverse problems

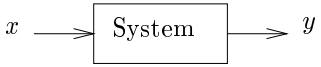
II Ill posed problems

III Regularisation

IV Data Analysis and estimation

TOSTISP Workshop — Oct. 2003

I Inverse Problems:



- **Forward problem:**

  Simulate data y, given input x

- **Inverse problem:**

  For a known system, estimate x from observations y (perturbations)

  Applications:

  deconvolution, image restoration, imaging...

  Also:

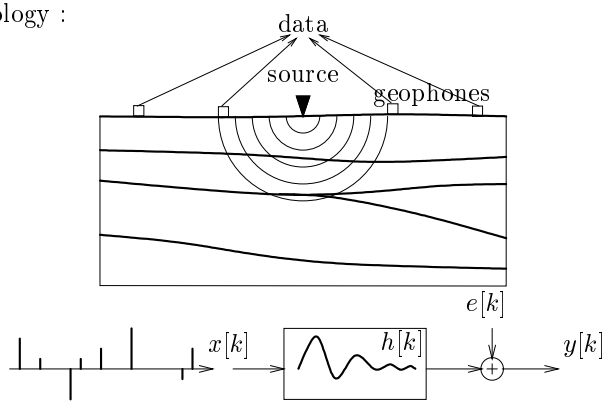
  spectral analysis, image segmentation...

1D convolution

Model : $y(t) = \int x(\tau)h(t-\tau)\mathrm{d}\tau + e(t) \quad \Rightarrow \quad y[k] = \sum_l x[l]h[k-l] + e[k]$

Examples : Integration time when using a CCD sensor,
Transmission through a communication channel...

Also seismology :

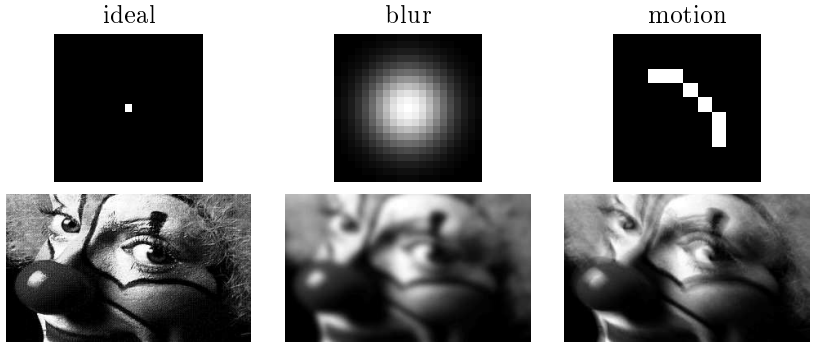


2D convolution

Model : $y(r,s) = \iint x(u,v)h(r-u,s-v)\mathrm{d}u\mathrm{d}v \quad (\ + \ e(r,s))$

$h(r,s)$: 2D kernel or point spread function (PSF)

2D kernel examples



Intuitives approaches for linear systems

Linear algebra: $\boldsymbol{y} = \mathbf{H}\boldsymbol{x}$

$$\Rightarrow \quad \text{Matrix inversion:} \quad \boldsymbol{x} = \mathbf{H}^{-1}\boldsymbol{y}$$

But $\mathbf{H}$ is not always square and invertible

1D convolution $\Rightarrow$ H Toeplitz
2D convolution $\Rightarrow$ H Toeplitz-Block Toeplitz

Fourier domain: (1D or 2D convolution)

$$\hat{y}(f) = \hat{h}(f)\hat{x}(f) \quad \text{with} \quad \hat{x} = \mathrm{FT}\{x\}$$

$$\Rightarrow \quad \text{Spectral division:} \quad \hat{x}(f) = \begin{cases} \hat{y}(f)/\hat{h}(f) & \text{if } \hat{h}(f) \neq 0 \\ 0 & \text{else} \end{cases}$$

Problem for $\hat{h}(f) = 0$!

What about noisy data?

II Ill-posed problems

Hadamard conditions (1923)

Let $H : X \to Y$, X and Y Hilbert spaces.

The inversion of equation : $y = H(x)$ is **well posed** if the solution $\boldsymbol{x}(\boldsymbol{y})$ satisfies :

① **existence**

② **unicity**

③ **stability:** $\|\boldsymbol{y} - \boldsymbol{y}'\| \to 0 \;\Rightarrow\; \|\boldsymbol{x}(\boldsymbol{y}) - \boldsymbol{x}(\boldsymbol{y}')\| \to 0$

Remark : for a linear operator H:

① $\Leftrightarrow$ $y \in \mathrm{Im}\, H$;
② $\Leftrightarrow \mathrm{Ker}\, H = 0$.

The inversion is said to be **Ill posed** if one of these three conditions is not met.

Case of finite dimension

$$\boldsymbol{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{n}, \quad \boldsymbol{x} \in \mathbb{R}^N,\ \boldsymbol{y} \in \mathbb{R}^M,\ \mathbf{H}\text{: matrix } M \times N$$

① **Existence:**

If $\boldsymbol{y} \notin \mathrm{Im}\ \mathbf{H} \rightarrow$ projection of $\boldsymbol{y}$ on $\mathrm{Im}\ \mathbf{H}$

$\Leftrightarrow$ *Least squares solution:* $\boldsymbol{x}_{\mathrm{LS}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2$

② **Unicity:**

If $\mathrm{Ker}\, \mathbf{H} \neq \mathbf{0} \rightarrow$ minimum norm constraint

$\Leftrightarrow$ *generalised inverse:* $\boldsymbol{x}_{\mathrm{GI}}$ = projection of $\boldsymbol{x}_{\mathrm{LS}}$ on $(\mathrm{Ker}\, \mathbf{H})^{\perp}$

③ **Stability :** garanteed for finite dimension

Generalised inverse and SVD

Least squares: $\boldsymbol{x}_{\mathrm{LS}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2$

Normal equation: $\mathbf{H}^{\mathrm{t}}\boldsymbol{y} = (\mathbf{H}^{\mathrm{t}}\mathbf{H})\boldsymbol{x}$

Generalised inverse: $\boldsymbol{x}_{\mathrm{GI}}$ = projection of $\boldsymbol{x}_{\mathrm{LS}}$ on $(\mathrm{Ker}\, \mathbf{H})^{\perp}$

Singular values decomposition *(SVD)* :

- $\{\boldsymbol{u}_i\}_{i=1}^{M}, \{\boldsymbol{v}_j\}_{j=1}^{N}$ orthonormal bases of $\mathbb{R}^M, \mathbb{R}^N$,
- $\{\lambda_i\}_{i=1}^{\max(M,N)}$ decreasing positive sequence, such that:

$$\begin{cases} \mathbf{H}\boldsymbol{v}_j = \lambda_j \boldsymbol{u}_j \\ \mathbf{H}^{\mathrm{t}}\boldsymbol{u}_i = \lambda_i \boldsymbol{v}_i \end{cases} \Rightarrow \begin{cases} \mathbf{H}^{\mathrm{t}}\mathbf{H}\boldsymbol{v}_j = \lambda_j^2 \boldsymbol{v}_j \\ \mathbf{H}\mathbf{H}^{\mathrm{t}}\boldsymbol{u}_i = \lambda_i^2 \boldsymbol{u}_i \end{cases}$$

$$\boldsymbol{x}_{\mathrm{GI}} = \sum_{j=1}^{N} \alpha_j \boldsymbol{v}_j \quad \text{with} \quad \alpha_j = \begin{cases} \boldsymbol{u}_j^{\mathrm{t}}\boldsymbol{y}/\lambda_j & \text{si} \quad \lambda_j > 0 \\ 0 & \text{si} \quad \lambda_j = 0 \end{cases}$$

Generalised inverse and condition number

Error propagation:

If $\boldsymbol{y} \to \boldsymbol{y} + \delta\boldsymbol{y}$ then $\boldsymbol{x}_{\rm GI} \to \boldsymbol{x}_{\rm GI} + \delta\boldsymbol{x}_{\rm GI}$ and

$$\frac{\|\delta\boldsymbol{x}_{\rm GI}\|^2}{\|\boldsymbol{x}_{\rm GI}\|^2} \leqslant \frac{\|\delta\boldsymbol{y}\|^2}{\|\boldsymbol{y}\|^2}\left(\frac{\lambda_{\max}}{\lambda_{\min}}\right)^2 = \frac{\|\delta\boldsymbol{y}\|^2}{\|\boldsymbol{y}\|^2} C(\mathbf{H}^{\rm t}\mathbf{H})$$

Equality is reached for some couples $(\boldsymbol{y}, \delta\boldsymbol{y})$!

Condition number of a square matrix $\mathbf{A}$: $C(\mathbf{A}) = \frac{\lambda_{\max}}{\lambda_{\min}}$

$\mathbf{A}$ is said *ill conditionned* if $C(\mathbf{A})$ is large (*e.g.*, $C(\mathbf{A}) > 10^3$).

Systems obtained by discretisation
of an ill posed problem are ill conditionned

Condition number: illustration

$$\mathbf{H} = \begin{bmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{bmatrix},\ \boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},\ \boldsymbol{y} = \begin{bmatrix} 32 \\ 23 \\ 33 \\ 31 \end{bmatrix},$$

Perturbations:

$$\delta\boldsymbol{y} = \begin{bmatrix} 0,1 \\ -0,1 \\ 0,1 \\ -0,1 \end{bmatrix},\ (\|\delta\boldsymbol{y}\| \simeq 0,2) \ \Rightarrow\ \delta\boldsymbol{x} = \begin{bmatrix} 8,2 \\ -13,6 \\ 3,5 \\ 2,1 \end{bmatrix},\ (\|\delta\boldsymbol{x}\| \simeq 15)$$

Eigen values of $\mathbf{H}$: 30.3 ; 3.9 ; 0.8 ; 0.01 $\Rightarrow$ $C(\mathbf{H}) \simeq 3000$

$$\mathbf{H}^{-1} = \begin{bmatrix} 25 & -41 & 10 & -6 \\ -41 & 68 & -17 & 10 \\ 10 & -17 & 5 & -3 \\ -6 & 10 & -3 & 2 \end{bmatrix}$$

Links between SVD and spectral division with DFT

Circular convolution (with periodisation): $y[k] = \sum_{l=0}^{P-1} h[l]x[k-l] + e[k]$

$$\ldots \underbrace{x[-2]}_{x[N-2]},\ \underbrace{x[-1]}_{x[N-1]},\ \underbrace{x[0],\ x[1], \ldots\ x[N-1]}_{\boldsymbol{x}},\ \underbrace{x[N]}_{x[0]},\ \underbrace{x[N+1]}_{x[1]},\ \ldots$$

$\Rightarrow\ \boldsymbol{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{e}$ with $\mathbf{H} = \text{Circul}\{\boldsymbol{h}\}$, *circulant matrix*

$$H = \begin{bmatrix} h_0 & 0 & & \cdots & 0 & h_{P-1} & & \cdots & h_1 \\ h_1 & h_0 & 0 & \cdots & & 0 & h_{P-1} & \cdots & h_2 \\ & \ddots & \ddots & \ddots & & & \ddots & & \\ h_{P-1} & \cdots & & h_0 & 0 & & \cdots & & 0 \\ & & & & & & & & \\ \ddots & \ddots & \ddots & \ddots & & & \ddots & & \\ & & & & & & & & \\ 0 & & \cdots & & 0 & h_{P-1} & & \cdots & h_0 \end{bmatrix}$$

Diagonalisation of circulant matrices

$$\mathbf{H} = \frac{1}{N}\mathbf{W}^\dagger \text{diag}\{\hat{\boldsymbol{h}}\}\mathbf{W},\ \ \hat{\boldsymbol{h}} = \mathbf{W}\boldsymbol{h}\ \ (\boldsymbol{h}\ :\ 1^{\rm st}\ \text{line of}\ \mathbf{H}) \text{ and } \boldsymbol{h} \overset{\rm DFT}{\longrightarrow} \hat{\boldsymbol{h}}$$

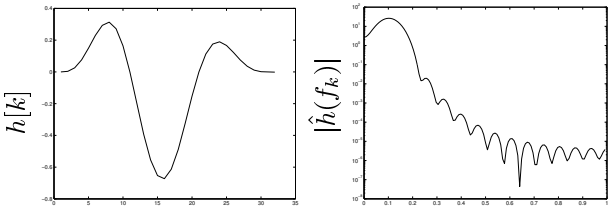
with $\mathbf{W} = \left\{e^{-2i\pi\frac{1}{N}(k-1)(l-1)}\right\}_{kl}$, DFT matrix, $\mathbf{W}^\dagger\mathbf{W} = \mathbf{W}\mathbf{W}^\dagger = N\mathbf{I}$

Generalised inverse:

$$\mathbf{H}^\dagger\mathbf{H}\boldsymbol{x}_{\rm LS} = \mathbf{H}^\dagger\boldsymbol{y} \ \Leftrightarrow\ \text{diag}\{|\hat{\boldsymbol{h}}|^2\}\mathbf{W}\boldsymbol{x}_{\rm LS} = \text{diag}\{\hat{\boldsymbol{h}}^*\}\mathbf{W}\boldsymbol{y}$$

$$\Rightarrow \hat{\boldsymbol{x}}_{\rm GI}(f_k) = \begin{cases} \hat{y}(f_k)/\hat{h}(f_k) & \text{if} \quad \hat{h}(f_k) \neq 0, \\ 0 & \text{if} \quad \hat{h}(f_k) = 0 \end{cases} \qquad \begin{pmatrix} \text{spectral} \\ \text{division} \end{pmatrix}$$

Conditionment of H



III Regularisation

Ways to stabilise solution $\boldsymbol{x}_{\text{GI}}$

- Implicit prior information:
  - Troncated singular value decomposition
  - Iterative methods
  - Clean
- Explicit prior information:

  Adequation to the data is not sufficient... needs of prior information

  Several approaches:

  - constraint $\quad \boldsymbol{x}_{\text{c}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 \quad \text{s.c.} \quad \boldsymbol{x} \in \mathcal{X}$
  - parameterise $\quad \boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}(\boldsymbol{\theta})\|^2$
  - penalise $\quad \boldsymbol{x}_{\text{p}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 + \alpha\Omega(\boldsymbol{x})$

Troncated singular values decomposition (TSVD)

$$\boldsymbol{x}_{\text{TSVD}} = \sum_{k=0}^{K} \frac{\boldsymbol{u}_k^{\text{t}}\boldsymbol{y}}{\lambda_k}\boldsymbol{v}_k \quad \text{with} \quad K < \text{rank}(\mathbf{H})$$

λ_k decreasing sequence

$\Rightarrow$ Problem: very sensitive to choice of K

Iterative methods

- Bialy method: fixed point algorithm, to solve

$$\boldsymbol{x} = \alpha\boldsymbol{y} + (\mathbf{I} - \alpha\mathbf{H})\boldsymbol{x} \quad \text{with} \quad 0 < \alpha < \frac{2}{||H||}$$

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \alpha(\boldsymbol{y} - \mathbf{H}\boldsymbol{x}_n), \quad n = 0, 1, \ldots$$

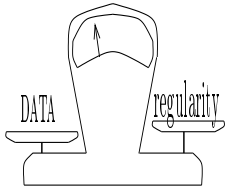
$\Rightarrow$ Stop iterating, otherwise gives $\boldsymbol{x}_{\text{LS}}$

- Landweber, Gerchberg-Papoulis-Van-Cittert...
- Minimisation algorithms: Conjugate gradient to minimise $\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2$

$\Rightarrow$ Stop iterating, otherwise gives $\boldsymbol{x}_{\text{LS}}$

$\Rightarrow$ Problem: choosing the number of iterations

Regularisation as a compromise



$$\boldsymbol{x}_\alpha(\boldsymbol{y}) = \arg\min_{\boldsymbol{x}} J_\alpha(\boldsymbol{x}),$$

$$J_\alpha(\boldsymbol{x}) = \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 + \alpha\Omega(\boldsymbol{x})$$

α : *regularisation parameter*

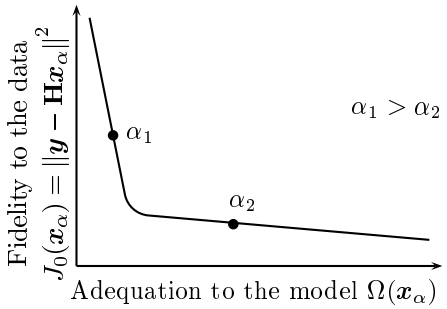
$\Omega(\boldsymbol{x})$ enforces desired properties of $\boldsymbol{x}$

$\alpha \to 0$	$\alpha \to \infty$
$\boldsymbol{x}_0 = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2$	$\boldsymbol{x}_\infty = \arg\min_{\boldsymbol{x}} \Omega(\boldsymbol{x})$
Least squares solution (not regularised)	prior solution (independant of $\boldsymbol{y}$)

$\Rightarrow$ Problems: $\begin{cases} \text{choice of } \Omega(\boldsymbol{x}) \\ \text{choice of the parameter } \alpha \end{cases}$

"L-CURVE"

$J_0(\boldsymbol{x}_\alpha)$ et $-\Omega(\boldsymbol{x}_\alpha)$ are increasing functions of α



QUADRATIC CASE: PTT APPROACH (PHILLIPS–TWOMEY–TIKHONOV)

- Fonctionnal framework

$$\Omega(\boldsymbol{x}) = \int \sum_k \omega_k(t) \left(\frac{\mathrm{d}^k x}{\mathrm{d}t^k}(t) \right)^2 \mathrm{d}t$$

- Finite dimension ($\boldsymbol{x} \in \mathbb{R}^N$)

$$\begin{array}{rcl} & & \Omega(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{x}_\infty)^{\mathrm{t}} \mathbf{Q} (\boldsymbol{x} - \boldsymbol{x}_\infty) \\ \boldsymbol{x}_{\mathrm{PTT}} & = & (\mathbf{H}^{\mathrm{t}}\mathbf{H} + \alpha\mathbf{Q})^{-1} (\alpha \mathbf{Q}\boldsymbol{x}_\infty + \mathbf{H}^{\mathrm{t}}\boldsymbol{y}) \\ & = & \boldsymbol{x}_\infty + (\mathbf{H}^{\mathrm{t}}\mathbf{H} + \alpha\mathbf{Q})^{-1} \mathbf{H}^{\mathrm{t}} (\boldsymbol{y} - \mathbf{H}\boldsymbol{x}_\infty) \end{array}$$

$$\text{Example : } \mathbf{D} = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & \mathbf{0} & \\ & & \ddots & \ddots & \\ & \mathbf{0} & & & \\ & & & -1 & 1 \end{bmatrix}, \quad \mathbf{Q} = \sum_k (\mathbf{D}^k)^{\mathrm{t}} \mathrm{diag}\{\omega_k\} \mathbf{D}^k$$

- Stabilisation of $\boldsymbol{x}_{\mathrm{GI}}$ ($\boldsymbol{x}$ being low energy):

$$J_\alpha(\boldsymbol{x}) = \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 + \alpha \underbrace{\sum_{k=1}^{N} x_k^2}_{\|\boldsymbol{x}\|^2}$$

$$\boldsymbol{x}_{\mathrm{PTT}} = (\mathbf{H}^{\mathrm{t}}\mathbf{H} + \alpha\mathbf{I})^{-1}\mathbf{H}^{\mathrm{t}}\boldsymbol{y}$$

- Smoothness of the desired solution :

$$J_\alpha(\boldsymbol{x}) = \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 + \alpha \underbrace{\sum_{k=1}^{N-1} (x_k - x_{k-1})^2}_{\|\mathbf{D}\boldsymbol{x}\|^2}$$

$$\boldsymbol{x}_{\mathrm{PTT}} = (\mathbf{H}^{\mathrm{t}}\mathbf{H} + \alpha\mathbf{D}^{\mathrm{t}}\mathbf{D})^{-1}\mathbf{H}^{\mathrm{t}}\boldsymbol{y}$$

- **Computation of the solution (approximated)**

① Circular convolution: let $\boldsymbol{v}^{[L]} = [\boldsymbol{v}^{\mathrm{t}}, 0, ..., 0]^{\mathrm{t}}$ be of dimension L.

If $L \geqslant N + 2P - 2$, then $\boldsymbol{h}^{[L]} \circledast \boldsymbol{x}^{[L]} = [\ldots, (\mathbf{H}\boldsymbol{x})^{\mathrm{t}}, \ldots]^{\mathrm{t}}$.

② Fast convolution:

$$\begin{array}{rcccc} \mathbf{H}\boldsymbol{x} & \rightsquigarrow & \boldsymbol{h}^{[L]} \circledast \boldsymbol{x}^{[L]} & \xrightarrow{\mathrm{DFT}} & \hat{\boldsymbol{h}}^{[L]} \hat{\boldsymbol{x}}^{[L]} \\ \text{Moreover, } \mathbf{H}^{\mathrm{t}}\boldsymbol{y} & \rightsquigarrow & (\boldsymbol{h}^{-})^{[L]} \circledast \boldsymbol{y}^{[L]} & \xrightarrow{\mathrm{DFT}} & (\hat{\boldsymbol{h}}^{\,*})^{[L]} \hat{\boldsymbol{y}}^{[L]} \text{ with } \boldsymbol{h}^{-}[n] = \boldsymbol{h}[-n] \end{array}$$

③ $\boldsymbol{x}_{\mathrm{PTT}}{}^{[L]} = (\mathbf{H}^\dagger\mathbf{H} + \alpha\mathbf{Q})^{-1}\mathbf{H}^\dagger\boldsymbol{y} \overset{\mathrm{DFT}^{-1}}{\longleftarrow} \hat{\boldsymbol{x}}_{\mathrm{PTT}}{}^{[L]} = \dfrac{(\hat{\boldsymbol{h}}^{\,*})^{[L]} \hat{\boldsymbol{y}}^{[L]}}{\left|\hat{\boldsymbol{h}}^{[L]}\right|^2 + \alpha \left|\hat{\boldsymbol{c}}^{[L]}\right|^2}$ if $\boldsymbol{x}_\infty = \mathbf{0}$

Exact result for $M = N = P$ if

$$\mathbf{H} = \mathrm{Circul}\{\boldsymbol{h}\} \text{ and } \mathbf{Q} = \mathrm{Circul}\{\boldsymbol{c}\}.$$

Wiener filter (1940)

- Let $X = \{X_n\}_{\mathbb{Z}}$, and $Y = \{Y_n\}_{\mathbb{Z}}$ be jointly stationnary and centered with autocorrelations C_x, C_y and intercorrelation C_{xy}.

> Let g be a filter such that $C_{xy} = g * C_y$. Then, for all k, $\overset{o}{X} = g * Y$ minimises the mean square error $\mathrm{E}\left[\left(\overset{o}{X}_k - X_k\right)^2\right]$.

Application : $\boldsymbol{y} = \boldsymbol{h} * \boldsymbol{x} + \boldsymbol{e}$, with (X, E) decorrelated :

$$\forall f,\ \ \hat{g}(f) = \frac{\hat{h}^*(f) S_x(f)}{\left|\hat{h}(f)\right|^2 S_x(f) + S_e(f)} = \frac{1}{\hat{h}(f)} \frac{\left|\hat{h}(f)\right|^2 S_x(f)}{\left|\hat{h}(f)\right|^2 S_x(f) + S_e(f)}$$

Often presented as a smoothing of the data before applying the inverse filter

- Classical use:

① X and E, being white of power σ_x^2 and σ_e^2:

$$\forall f,\ \hat{g}(f) = \frac{\hat{h}^*(f)}{\left|\hat{h}(f)\right|^2 + \sigma_e^2/\sigma_x^2}$$

② Finite dimension application with periodisation:

$$\hat{\boldsymbol{x}}_{\mathrm{W}}{}^{[L]} = \frac{\hat{\boldsymbol{h}}^{*[L]}}{\left|\hat{\boldsymbol{h}}^{[L]}\right|^2 + \sigma_e^2/\sigma_x^2} \hat{\boldsymbol{y}}^{[L]}$$

$\Leftrightarrow$ PTT approach for $\boldsymbol{Q} = \boldsymbol{I}$ and $\alpha = \sigma_n^2/\sigma_x^2$!

Maximum entropy

- Information theory (Shannon):

the least informative signal maximises the entropy

$$\mathcal{H}(\boldsymbol{x}) = -\sum_{k=1}^{N} p[k] \log(p[k]) \quad \text{with} \quad p[k] = \frac{x[l]}{\sum_{l=1}^{N} x[l]}$$

- Maximum entropy regularisation: (different alternative)

$$\boldsymbol{x}_{\mathrm{ME}} \text{ maximise } \mathcal{H}(x) \ \text{ under constraint } \frac{1}{N}\left\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\right\|^2 \leqslant \sigma_b^2$$

Minimisation of a penalised criterion: $J_\alpha(\boldsymbol{x}) = \left\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\right\|^2 + \alpha \sum_{k=1}^{N} p[k] \log(p[k])$

$\Rightarrow$ α : Lagrange parameter

Remark: positiveness ensured

Penalisation for "sparse peaks" signals

- "sparse peaks" signals:

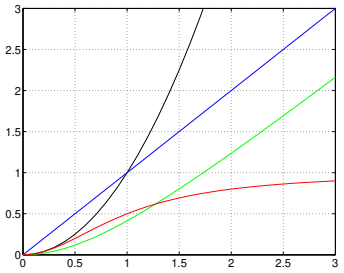
Low number of high value points

Example : reflectivity in seismology, line spectra in spectral analysis...

- Penalised criterion:

$$J_\alpha(\boldsymbol{x}) = \left\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\right\|^2 + \alpha \sum_{k=1}^{N} \phi(x[k])$$

$$\phi(x) = \begin{cases} x^2 & \text{quadratic} \\ |x|^p, \ 1 < p \leqslant 2 & L_p \text{ norm} \\ \sqrt{s^2 + x^2} & \text{hyperbolic} \\ \frac{x^2}{s^2 + x^2} & \text{non convex} \end{cases}$$



Penalisation for "images with contours"

- "Images with contours":

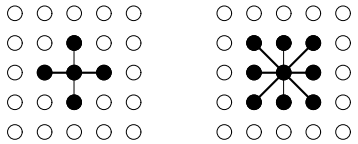
  Piecewise smooth signals with rare discontinuities

  Example : images but also in 1D...

- Penalised criterion:

$$J_\alpha(\boldsymbol{x}) = \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 + \alpha \sum_{k \sim l} \phi(x[k] - x[l])$$

4-neighborhood 8-neighborhood



Minimisation of penalised criterion

- Fidelity to data:

$$\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 \text{ quadratic}$$

- Prior model: $\phi(x)$
  - ϕ quadratic

    $\Rightarrow$ J_α quadratic : solution of a linear system of equations
  - ϕ convex

    $\Rightarrow$ J_α convex: "easy" minimisation
  - ϕ non convex

    $\Rightarrow J_\alpha$ non convex: "difficult" minimisation

    possible local minima

IV Data analysis and estimation

- **Noise on the data**

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{x}, \boldsymbol{e}), \quad \text{(for example: } \boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{e})$$

$$\boldsymbol{e} \text{ badly known} \quad \Rightarrow \quad \begin{cases} \text{statistical model:} \\ \text{e.g. density function } \boldsymbol{e} \mapsto p_{\boldsymbol{E}} \end{cases}$$

- **Likelihood**

(linear system with additive noise)

$$L(\boldsymbol{x}; \boldsymbol{y}) = p(\boldsymbol{y} \,|\, \boldsymbol{x}) \quad (= p_E(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}))$$

neg-log-likelihood: $V(\boldsymbol{x}; \boldsymbol{y}) = -\ln L(\boldsymbol{x}; \boldsymbol{y})$

Maximum likelihood (ML) estimator :

$$\boxed{\boldsymbol{x}_{\text{ML}} = \arg\max_{\boldsymbol{x}} L(\boldsymbol{x}; \boldsymbol{y}) = \arg\min_{\boldsymbol{x}} V(\boldsymbol{x}; \boldsymbol{y})}$$

ML : Linear Gaussian case

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{e}, \quad \boldsymbol{E} \sim \mathcal{N}(\boldsymbol{m}_{\boldsymbol{e}}, \mathbf{R}_{\boldsymbol{e}})$$

$$\boldsymbol{x}_{\text{ML}} = \arg\min_{\boldsymbol{x}} (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} - \boldsymbol{m}_{\boldsymbol{e}})^{\text{t}} \mathbf{R}_{\boldsymbol{e}}^{-1} (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} - \boldsymbol{m}_{\boldsymbol{e}})$$

$$\rightsquigarrow \text{ Normal equation: } (\boldsymbol{H}^{\text{t}} \mathbf{R}_{\boldsymbol{e}}^{-1} \boldsymbol{H}) \boldsymbol{x}_{\text{MV}} = \boldsymbol{H}^{\text{t}} \mathbf{R}_{\boldsymbol{e}}^{-1} (\boldsymbol{y} - \boldsymbol{m}_{\boldsymbol{e}})$$

Particular cases:

- $\boldsymbol{m}_{\boldsymbol{e}} = \mathbf{0}$, $\mathbf{R}_{\boldsymbol{e}} \propto \boldsymbol{I}$ $\Rightarrow \boldsymbol{x}_{\text{ML}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 = \boldsymbol{x}_{\text{LS}}$
- $\boldsymbol{m}_{\boldsymbol{e}} = \mathbf{0}$, $\mathbf{R}_{\boldsymbol{e}} = \text{diag}[\sigma_1^2, \ldots, \sigma_M^2]$ $\Rightarrow$ Weighted LS

$$\boldsymbol{x}_{\text{WLS}} = \arg\min \sum_{k=1}^{M} \frac{(y_k - [\boldsymbol{H}\boldsymbol{x}]_k)^2}{\sigma_k^2}$$

Other models for noise : examples

Additive Laplacian noise:

$$\boldsymbol{y} = \psi(\boldsymbol{x}) + \boldsymbol{e} \quad \text{and} \quad p_{\boldsymbol{E}}(\boldsymbol{n}) = \prod_{k=1}^{M} \frac{\sigma}{2} \exp(-\sigma|e_k|)$$

$$\Rightarrow \quad \mathrm{L}^1 \text{ estimator: } \boldsymbol{x}_{\mathrm{ML}} = \arg\min \sum_{k=1}^{M} |y_k - [\psi(\boldsymbol{x})]_k|$$

Application: less sensitive to aberrant data than Gaussian case

Poisson noise: $p(\boldsymbol{y} \,|\, \boldsymbol{x}) = \prod_{k=1}^{M} \frac{(\tilde{y}_k)^{y_k}}{y_k!} \exp \tilde{y}_k$ with $\tilde{y}_k = [\psi(\boldsymbol{x})]_k$ and $\boldsymbol{y} \in \mathbb{N}^M$

Applications : Positon emission tomography, photon counting

Multiplicative noise: $y_k = [\psi(\boldsymbol{x})]_k e_k, \ 1 \leqslant k \leqslant M$

For a Gaussian noise $\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$: $V(\boldsymbol{x}; \boldsymbol{y}) = \sum_{k=1}^{M} \frac{y_k^2}{2\sigma_k^2 [\psi(\boldsymbol{x})]_k^2} + \ln |[\psi(\boldsymbol{x})]_k^2|$

Applications : speckle noise

Bayesian Estimation

Bayes rule

$p(\boldsymbol{x})$: **prior law** of $\boldsymbol{x}$, summarizes knownledge on $\boldsymbol{x}$ **before** measurement

$$p(\boldsymbol{x} \,|\, \boldsymbol{y}) = \frac{p(\boldsymbol{y} \,|\, \boldsymbol{x}) p(\boldsymbol{x})}{p(\boldsymbol{y})} \propto p(\boldsymbol{y} \,|\, \boldsymbol{x}) p(\boldsymbol{x})$$

$p(\boldsymbol{x} \,|\, \boldsymbol{y})$: **posterior law** of $\boldsymbol{x}$, summarizes knowledge on $\boldsymbol{x}$ **after** measurement

Decision $\quad p(\boldsymbol{x} \,|\, \boldsymbol{y}) \overset{?}{\rightsquigarrow} \overset{o}{\boldsymbol{x}}$

- Maximum *a posteriori*: $\boldsymbol{x}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{x}} p(\boldsymbol{x} \,|\, \boldsymbol{y})$
- *Posterior* Mean: $\boldsymbol{x}_{\mathrm{PM}} = \mathrm{E}[\boldsymbol{X} \,|\, \boldsymbol{Y} = \boldsymbol{y}]$
- Marginal MAP: $\boldsymbol{x}_{\mathrm{MMAP}k} = \arg\max_{x_k} p(x_k \,|\, \boldsymbol{y}), \ 1 \leqslant k \leqslant N$

Link with likelihood framework

- $\boldsymbol{x}_{\mathrm{ML}} = \boldsymbol{x}_{\mathrm{MAP}}$ for a uniform prior law
- $\boldsymbol{x}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{x}} p(\boldsymbol{x} \,|\, \boldsymbol{y}) = \arg\min_{\boldsymbol{x}} V(\boldsymbol{y} \,|\, \boldsymbol{x}) - \ln p(\boldsymbol{x})$
  Penalised neg-log-Likelihood

The MAP estimator is a regularised solution obtained by penalisation

Additional tools (statistical framework)

- Error analysis (estimator's variance)
- Monte-Carlo optimisation (global minimisation)
- Monte-Carlo Markov Chain (MCMC):
  Estimate $\boldsymbol{x}_{\mathrm{PM}}$ using samples of $p(\boldsymbol{x} \,|\, \boldsymbol{y})$
  usefull when $p(\boldsymbol{x} \,|\, \boldsymbol{y})$ is very difficult to maximise

Extension to even more complex problems

- Myopic deconvolution : $p(\boldsymbol{x}, \boldsymbol{h} \,|\, \boldsymbol{y})$
- Data fusion : $p(\boldsymbol{x} \,|\, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots \boldsymbol{y}_N)$

References

- *Inverse problem theory: Methods for data fitting and model parameter estimation,*
  A. Tarantola, Elsevier Science Publishers, 1987
- *Image reconstruction and restoration: Overview of common estimation structures and problems,*
  G. Demoment, IEEE Tran. on Acousticic Speech and Signal Processing, 37, 1989.
- *Probabilistic Approach to Inverse Problems,*
  K. Mosegaard and A. Tarantola, International Handbook of Earthquake & Engineering Seismology, Part A., p 237-265, Academic Press, 2002.
  `http://web.ccr.jussieu.fr/tarantola/`
- *Approche bayésienne pour les problèmes inverses,*
  sous la direction de J. Idier, Traité Information – Commande – Communication, Hermes Science, 2001