

Problèmes inverses : optimisation et inférence

Jérôme IDIER

`idier@irccyn.ec-nantes.fr`

`http://www.irccyn.ec-nantes.fr/~idier`

Institut de Recherche en Communications et Cybernétique de Nantes
Équipe Analyse et décision en traitement du signal et de l'image

ÉCOLE D'ÉTÉ DE PEYRESQ

Problèmes inverses en traitement des signaux et des images
20-26 juillet 2009

Plan

- I. Généralités**
 - II. Déconvolution de train d'impulsions**
 - III. Restauration d'image et optimisation**
 - IV. Problèmes aveugles, méthodes autodidactes**
- Bibliographie**

Plan

I. Généralités

Problèmes directs et inverses ; exemples

Problèmes mal posés ; conditionnement

Méthodes naïves, régularisation

Approche pénalisée, interprétation bayésienne

II. Déconvolution de train d'impulsions

III. Restauration d'image et optimisation

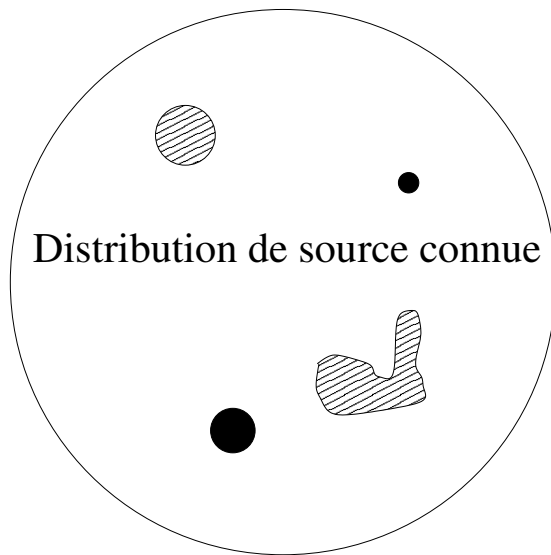
IV. Problèmes aveugles, méthodes autodidactes

Bibliographie

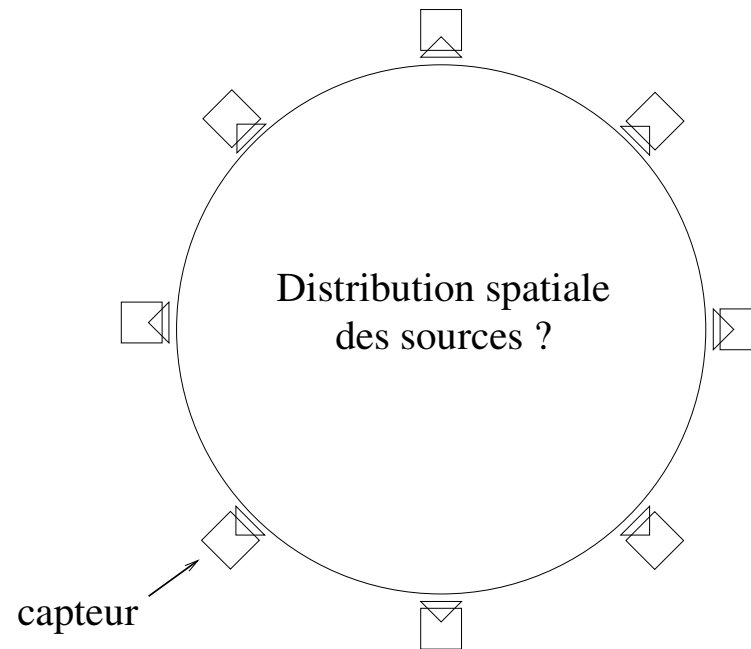
Problèmes *inverses*

■ Exemple 1 : tomographie

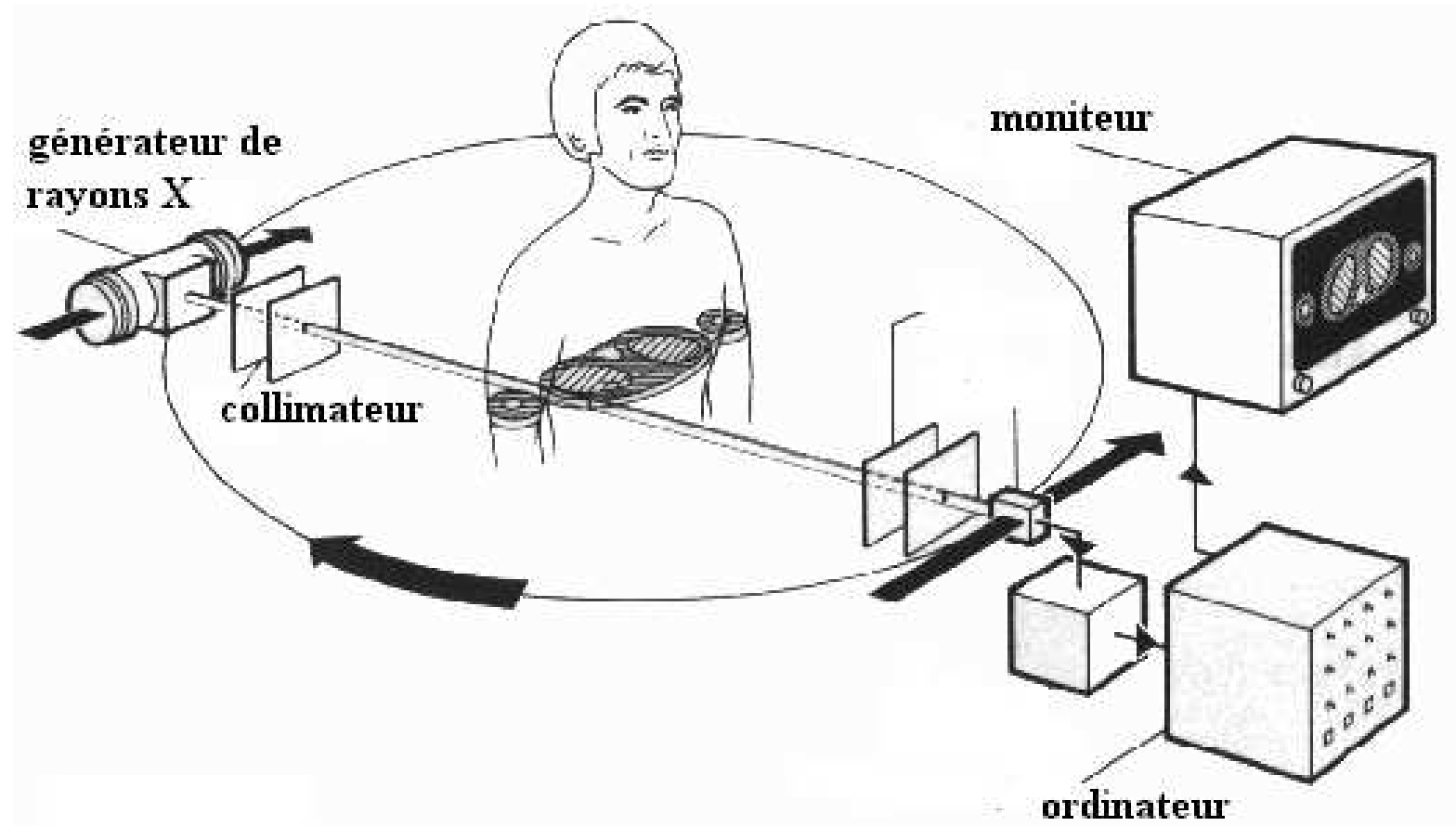
problème direct



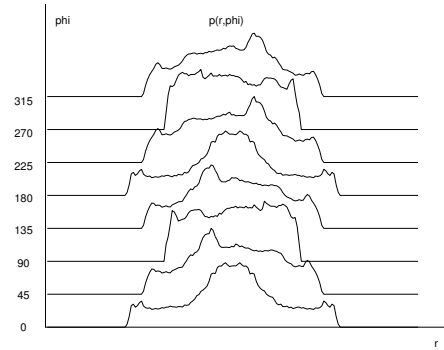
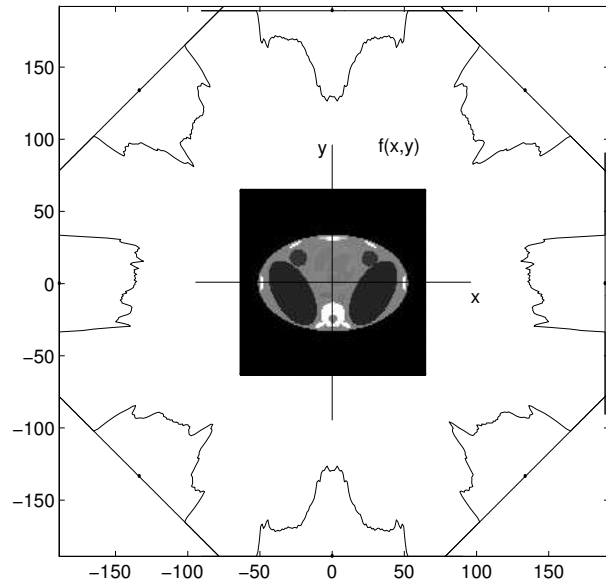
problème inverse



- Tomographie à rayons X (imagerie médicale)

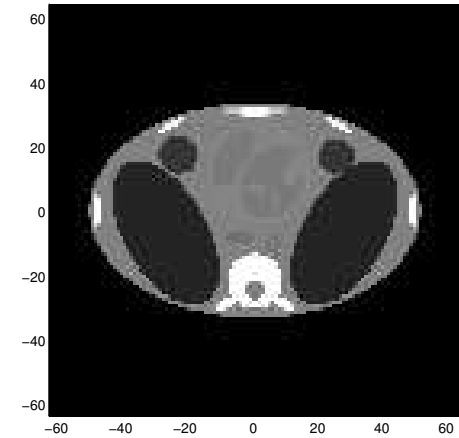


↪ Reconstruction 2D à partir de projections (problème direct linéaire) :

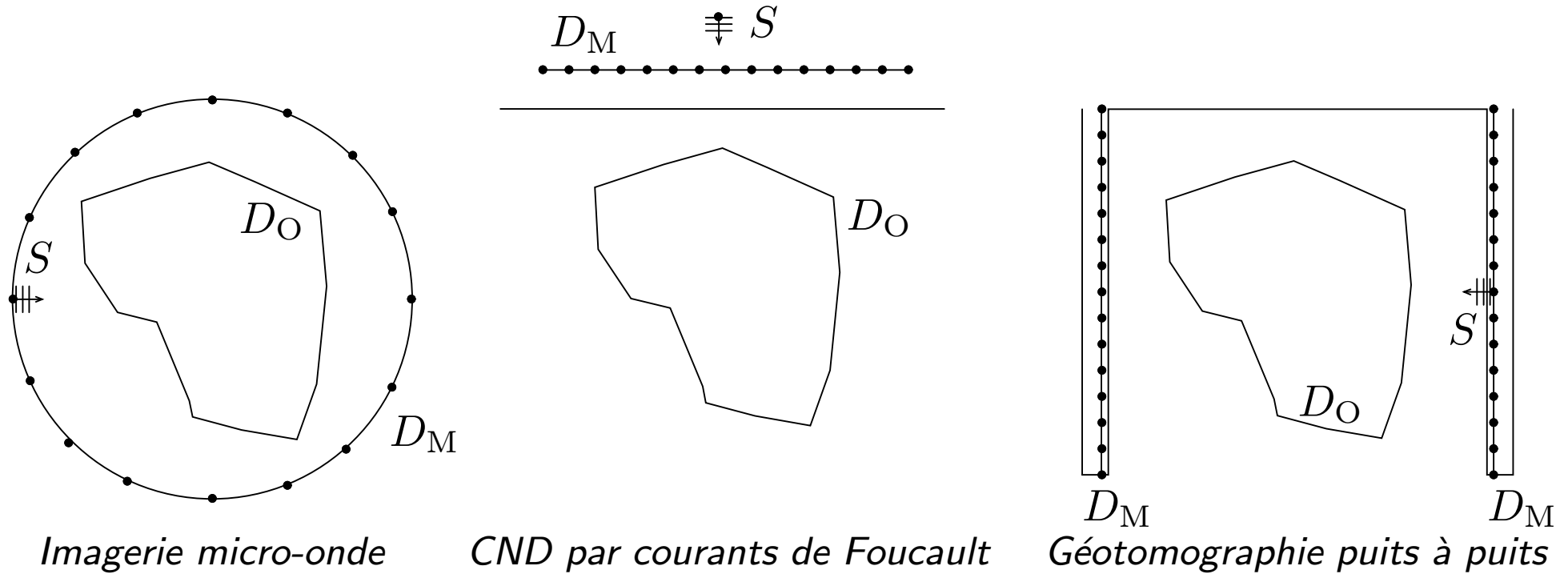


?

⇒

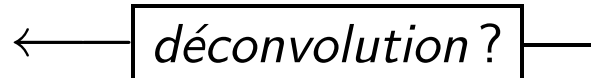
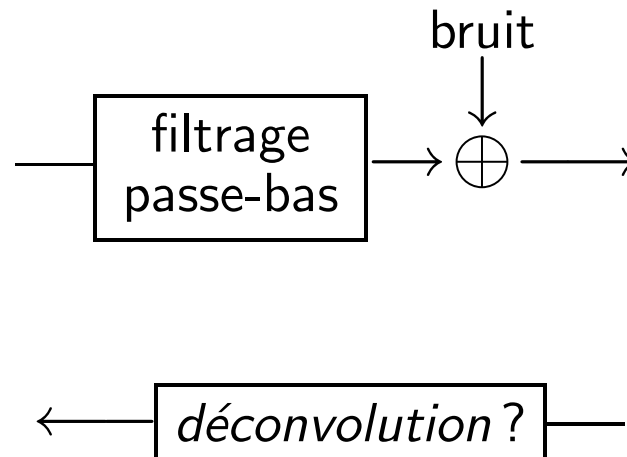


- Tomographie de diffraction (problème direct non linéaire)



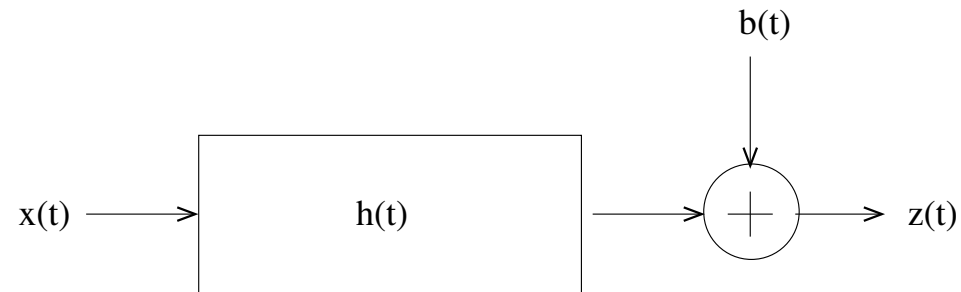
Une onde plane émise à partir d'une source S se propage dans le domaine de l'objet D_O . Le champ diffracté est mesuré par des capteurs placés dans le domaine D_M

■ Exemple 2 : déconvolution

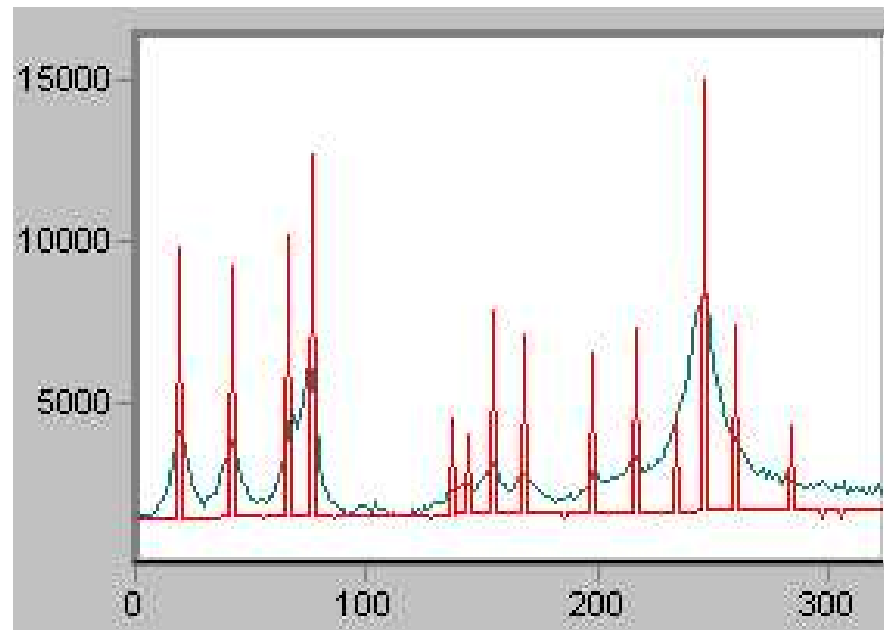


- Exemples 1D

- en temps : égalisation de canal, annulation d'écho, correction de distorsion, ...



- en fréquence : spectrométrie



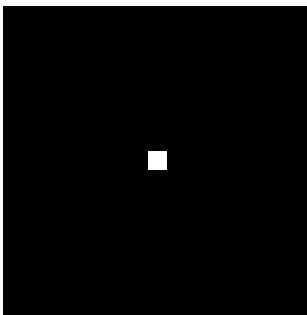
- Convolution 2D et restauration d'image

modèle :
$$z(r, s) = \iint x(u, v) h(r - u, s - v) du dv \quad (+ b(r, s))$$

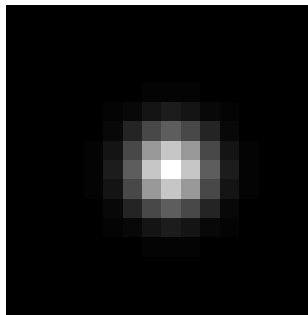
$h(r, s)$: noyau 2D ou fonction d'étalement de point

Exemples de noyaux 2D

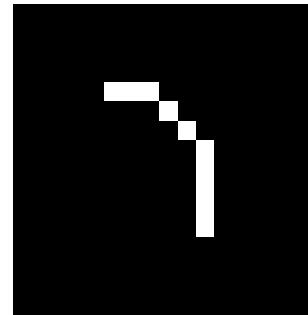
ponctuel



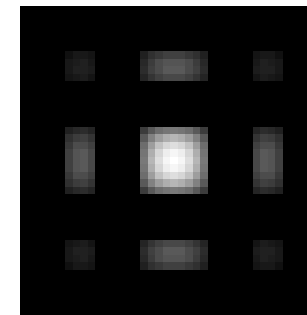
flou (gaussien)



bougé

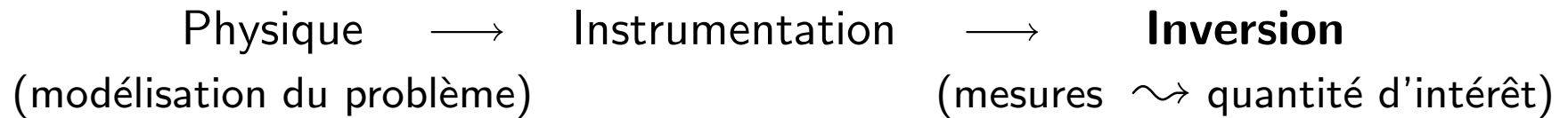


diffraction



■ Plusieurs points communs à ces exemples

- Chaîne de compétence



- Modèle direct

$$z = H(x) + \text{incertitudes}, \quad z \in \mathbb{R}^N, \quad x \in \mathbb{R}^M$$

voire

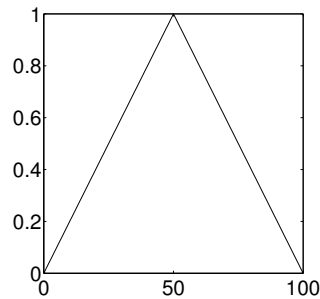
$$z = \mathbf{H} x + \text{incertitudes}, \quad \mathbf{H} \text{ matrice } N \times M$$

- Ces problèmes sont **mal posés**

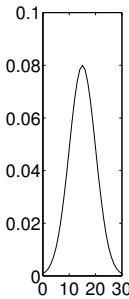
Problèmes inverses *mal posés*

Estimer x à partir de données $z = \mathbf{H}(x) + \textit{incertitudes}$

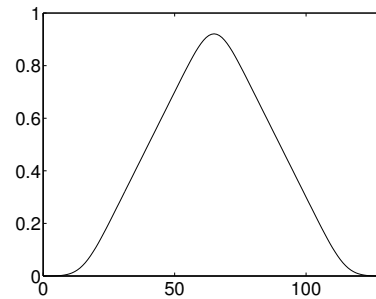
$\|\mathbf{H}(x) - \mathbf{H}(x_0)\|$ « petit » n'implique pas forcément $\|x - x_0\|$ « petit »



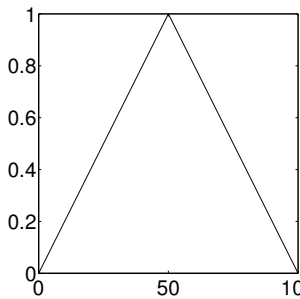
(a) signal d'entrée x



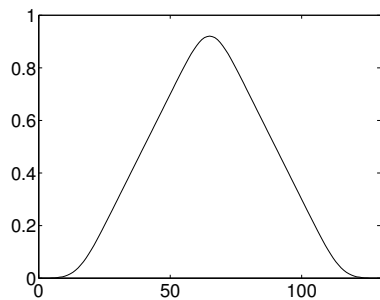
(b) RI h



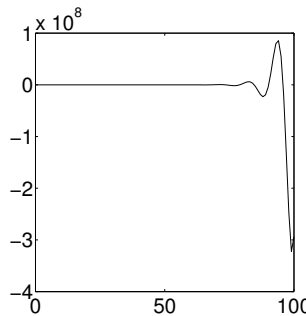
(c) $y = \text{conv}(x, h)$



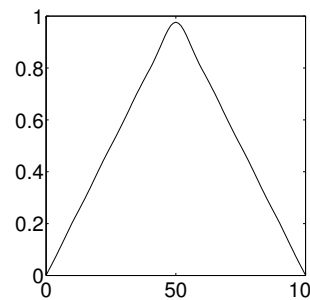
(d) $\text{deconv}(y, h)$



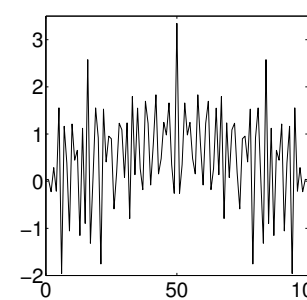
(e) sortie quantifiée z
(sur 10 bits)



(f) $\text{deconv}(z, h)$



(g) une solution
régularisée



(h) une solution
« exacte »

$$\mathbf{H}^{-1}z = \mathbf{H}^{-1}(\mathbf{H}x + b) = x + \mathbf{H}^{-1}b$$

Conditions de Hadamard [Hadamard 1902]

Soit $H : X \rightarrow Z$, X et Z espaces de Hilbert

La résolution de l'équation $z = Hx$ est dite **bien posé** si la solution $\hat{x}(z)$ vérifie :

① **existence**

② **unicité**

③ **stabilité** : $\|z - z'\| \rightarrow 0 \Rightarrow \|\hat{x}(z) - \hat{x}(z')\| \rightarrow 0$

NB : pour H opérateur linéaire, ① $\Leftrightarrow z \in \text{Im } H$; ② $\Leftrightarrow \text{Ker } H = \{0\}$

Sinon, le problème est dit **mal posé**.

Équation intégrale de Fredholm de première espèce

$$z(s) = Hx(s) = \int_a^b h(s, t) x(t) dt, \quad c \leq s \leq d$$

avec h fonction continue de dérivée partielle $\frac{\partial h}{\partial s}$ continue.

Connaissant $\tilde{z} \simeq z$, trouver une solution continue $\tilde{x} \simeq x$

... est un problème mal posé

(voir [Tikhonov et Arsénine 1976, Nashed 1981] et p. 19)

Passage en dimension finie**■ Une étape indispensable pour le calcul pratique d'une solution****● Choix d'une base de décomposition pour x**

- Famille obtenue par décalage d'un noyau $\gamma(t)$

$$x(t) = \sum_{n=1}^N x_n \gamma(t - \tau_n) + x^*(t), a \leq \tau_n \leq b$$

- Exemple : indicatrice de pixel, sinus cardinal, ondelette...

● Discrétisation de l'opérateur d'observation H

- Calcul analytique des coefficients
- *Méthode des moments* [Harrington 1987]

■ Le problème est-il bien posé en dimension finie ?

Inversion généralisée en dimension finie (cas linéaire)

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{b}, \quad \mathbf{x} \in \mathbb{R}^M, \quad \mathbf{z} \in \mathbb{R}^N$$

- ① **Existence** : si $\mathbf{z} \notin \text{Im } \mathbf{H} \rightarrow$ projection de \mathbf{z} sur $\text{Im } \mathbf{H}$
 \Leftrightarrow *Solution des moindres carrés* : $\hat{\mathbf{x}}^{\text{MC}}$ minimise $\|\mathbf{z} - \mathbf{H}\mathbf{x}\|^2$
 $\Rightarrow (\mathbf{H}^t \mathbf{H}) \hat{\mathbf{x}}^{\text{MC}} = \mathbf{H}^t \mathbf{z}$
- ② **Unicité** : si $\text{Ker } \mathbf{H} \neq \{\mathbf{0}\} \rightarrow$ contrainte de norme minimale
 \Leftrightarrow *Inverse généralisée* : $\hat{\mathbf{x}}^{\text{IG}} =$ projection de $\hat{\mathbf{x}}^{\text{MC}}$ sur $(\text{Ker } \mathbf{H})^\perp$
- ③ **Stabilité** : assurée en dimension finie

Le problème d'inversion est donc bien posé au sens de Hadamard, et pourtant...

Décomposition en valeurs singulières

■ Équation normale : $\mathbf{H}^t \mathbf{z} = (\mathbf{H}^t \mathbf{H}) \mathbf{x}$

■ Décomposition en valeurs singulières de \mathbf{H}

- $(\mathbf{u}_n)_{n=1}^N, (\mathbf{v}_m)_{m=1}^M$ bases orthonormées de $\mathbb{R}^N, \mathbb{R}^M$,
- $(\lambda_i)_{i=1}^{\max(M, N)}$, suite décroissante positive, $\lambda_i = 0$ si $i > \min(M, N)$ tels que

$$\begin{aligned} \mathbf{H} \mathbf{v}_m &= \lambda_m \mathbf{u}_m & \mathbf{H}^t \mathbf{u}_n &= \lambda_n \mathbf{v}_n \\ \Rightarrow \mathbf{H}^t \mathbf{H} \mathbf{v}_m &= \lambda_m^2 \mathbf{v}_m & \mathbf{H} \mathbf{H}^t \mathbf{u}_n &= \lambda_n^2 \mathbf{u}_n \end{aligned}$$

et $\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^t$ avec $\mathbf{U} = (\mathbf{u}_n), \mathbf{V} = (\mathbf{v}_n), \Lambda_{mn} = \begin{cases} \lambda_n & \text{si } m = n \leq \min(M, N) \\ 0 & \text{sinon} \end{cases}$

■ Inverse généralisée

$$\hat{\mathbf{x}}^{\text{IG}} = \sum_{m=1}^M \alpha_m \mathbf{v}_m \quad \text{avec} \quad \alpha_m = \begin{cases} \mathbf{u}_m^t \mathbf{z} / \lambda_m & \text{si } \lambda_m > 0 \\ 0 & \text{si } \lambda_m = 0 \end{cases}$$

Inverse généralisée et conditionnement

■ Propagation d'erreur

Soit $I = \text{rang}(\mathbf{H}) = \max_{\lambda_i > 0} i$. Si $\mathbf{z} \rightarrow \mathbf{z} + \delta\mathbf{z}$ alors $\hat{\mathbf{x}}^{\text{IG}} \rightarrow \hat{\mathbf{x}}^{\text{IG}} + \delta\hat{\mathbf{x}}^{\text{IG}}$ et

$$\frac{\|\delta\hat{\mathbf{x}}^{\text{IG}}\|^2}{\|\hat{\mathbf{x}}^{\text{IG}}\|^2} \leq \frac{\sum_1^I (\mathbf{u}_i^t \delta\mathbf{z})^2 \lambda_1^2}{\sum_1^I (\mathbf{u}_i^t \mathbf{z})^2 \lambda_I^2} \quad \left(= \frac{\|\delta\mathbf{z}\|^2}{\|\mathbf{z}\|^2} \frac{\lambda_1^2}{\lambda_N^2} \text{ si } I = N \right)$$

- La borne est atteinte pour certains couples $(\mathbf{z}, \delta\mathbf{z})$!

- $C(\mathbf{H}) = \frac{\lambda_1}{\lambda_{\min(M,N)}}$: nombre de condition de \mathbf{H} , $C(\mathbf{H}^t \mathbf{H}) = \frac{\lambda_1^2}{\lambda_{\min(M,N)}^2}$

\mathbf{H} est dite *mal conditionnée* si $C(\mathbf{H})$ est grand (par exemple $C(\mathbf{H}) > 10^5$).

Retour sur le cas continu

■ Propriétés spectrales des opérateurs compacts dans des espaces de Hilbert

- Système singulier $(\lambda_n, u_n, v_n)_{n \in \mathbb{N}}$ d'un opérateur compact H [Brézis 1983] :

$\lambda_n \geq 0$ valeurs singulières, u_n et v_n fonctions singulières telles que

- (u_n) base orthonormale de $(\text{Ker } H)^\perp$,
- (v_n) base orthonormale de $(\text{Ker } (H^*))^\perp$,
- $H u_n = \lambda_n v_n$ et $H^* v_n = \lambda_n u_n$

avec H^* opérateur adjoint de H , i.e., tel que $\langle Hx, z \rangle_Z = \langle x, H^*z \rangle_X$

- $\hat{x}^{\text{IG}} = \sum_{n \in \mathbb{N}} \lambda_n^{-1} \langle z, v_n \rangle_Z u_n$

- En général, on a $\lim_{n \rightarrow \infty} \lambda_n = 0$

Conclusion provisoire...

Problème mal posé \rightsquigarrow **le respect des données ne suffit pas !**

Toute méthode s'appuyant seulement sur ce respect sera dite **naïve...**

① **Inversion directe de $\mathbf{z} = \mathbf{H}(\mathbf{x})$**

$$\hat{\mathbf{x}}^{\text{ID}} = \mathbf{H}^{-1}(\mathbf{z}) \quad (\text{existence?})$$

② **Moindres carrés, inversion généralisée**

$$\hat{\mathbf{x}}^{\text{MC}} = \arg \min_{\mathbf{x}} \|\mathbf{z} - \mathbf{H}(\mathbf{x})\|^2 = \hat{\mathbf{x}}^{\text{ID}} \text{ si } \hat{\mathbf{x}}^{\text{ID}} \text{ existe}$$

③ **Maximum de vraisemblance**

Soit $p(\mathbf{z} | \mathbf{x}) = p_{\mathbf{B}}(\mathbf{z} - \mathbf{x})$ la densité de probabilité de \mathbf{z} sachant \mathbf{x} :

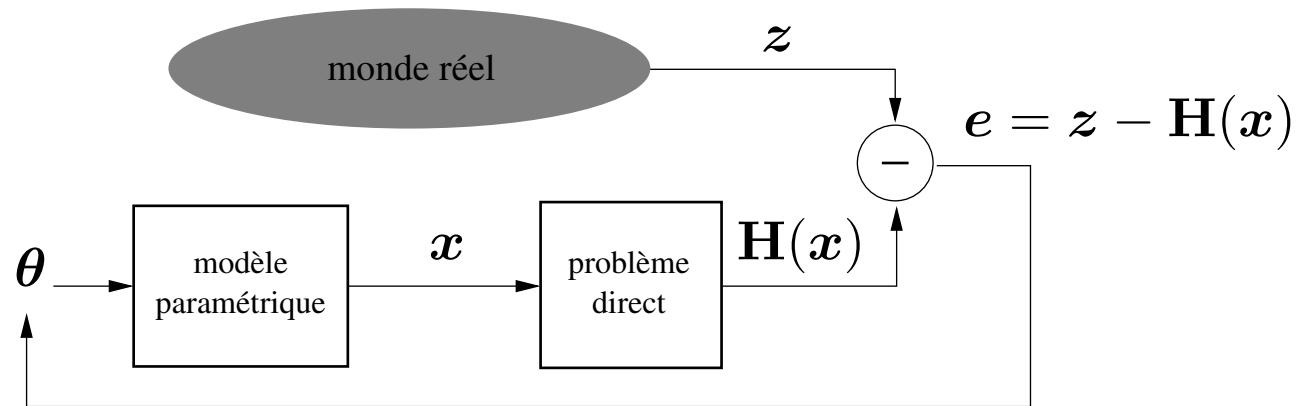
$$\hat{\mathbf{x}}^{\text{MV}} = \arg \max_{\mathbf{x}} p(\mathbf{z} | \mathbf{x}) = \hat{\mathbf{x}}^{\text{MC}} \text{ dans le cas d'un bruit gaussien}$$

④ **deconv de Matlab**

...

Régularisation = ajout d'une information *a priori*

■ Approche paramétrée



$$\hat{\theta} = \arg \min_{\theta} \|z - \mathbf{H}(x(\theta))\|^2$$

Problème de *moindres carrés non linéaires*

↪ minimisation (locale) par l'algorithme de Levenberg-Marquadt

■ Prise en compte de contraintes

Exemple : positivité des composantes de x

↪ minimisation sous contrainte d'inégalités

■ Approche multi-objectif

Assurer $\|z - \mathbf{H}(x)\|^2$ petit, mais aussi $\Phi(x)$ petit

$$\text{Exemple : } \Phi(x) = \int (x'(t))^2 dt$$

(régularité du signal dans le cas fonctionnel [Tikhonov et Arsénine 1976])

■ Régularisation par minimisation tronquée

- E.g. : K itérations d'un algorithme de plus profonde descente pour minimiser $\|z - \mathbf{H}(x)\|^2$, initialisé par $x = \mathbf{0}$
- Comparable à une démarche multi-objectif dans certains cas, avec $\Phi(x) = \|x\|^2$ [Lagendijk et coll. 1988]

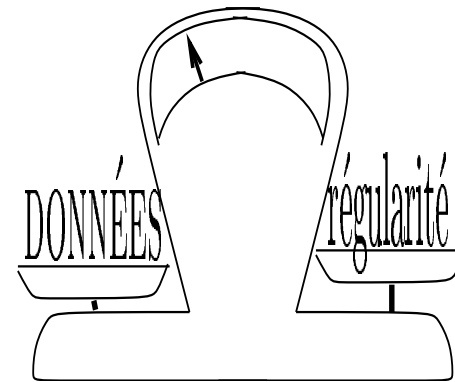
Approche multi-objectif \rightsquigarrow Approche pénalisée

■ La régularisation comme un compromis...

$$\mathcal{J}_\mu(\mathbf{x}) = \|\mathbf{z} - \mathbf{H}\mathbf{x}\|^2 + \mu\Phi(\mathbf{x})$$

μ : paramètre de régularisation

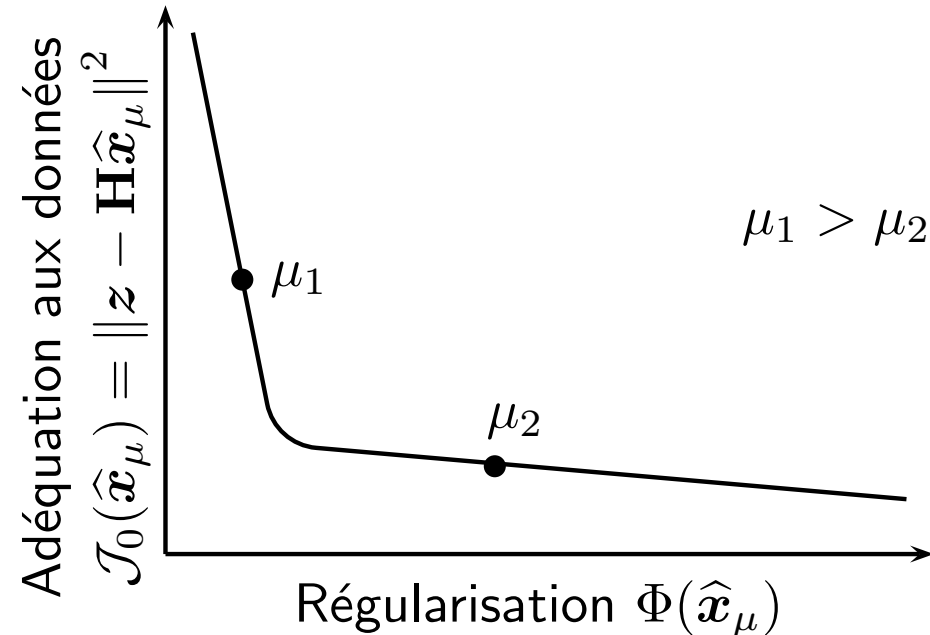
$$\hat{\mathbf{x}}_\mu(\mathbf{z}) = \arg \min_x \mathcal{J}_\mu(\mathbf{x})$$



$\mu \rightarrow 0$	$\mu \rightarrow \infty$
$\hat{\mathbf{x}}_0 = \arg \min_x \ \mathbf{z} - \mathbf{H}\mathbf{x}\ ^2$ solution non régularisée	$\hat{\mathbf{x}}_\infty = \arg \min_x \Phi(\mathbf{x})$ solution <i>a priori</i> (ne dépend pas de \mathbf{z})

■ « Courbe en L » [Hansen 1992]

(i.e., frontière de Pareto en échelle log-log)



Propriété : $\mathcal{J}_0(\hat{\mathbf{x}}_\mu)$ et $-\Phi(\hat{\mathbf{x}}_\mu)$ sont des fonctions croissantes de μ

à démontrer en combinant

$$\begin{cases} \mathcal{J}_{\mu_1}(\hat{\mathbf{x}}_{\mu_1}) \leq \mathcal{J}_{\mu_1}(\hat{\mathbf{x}}_{\mu_2}), \\ \mathcal{J}_{\mu_2}(\hat{\mathbf{x}}_{\mu_2}) \leq \mathcal{J}_{\mu_2}(\hat{\mathbf{x}}_{\mu_1}) \end{cases}$$

Estimation bayésienne

■ Règle de Bayes

$$p(\mathbf{x} | \mathbf{z}) = \frac{p(\mathbf{z} | \mathbf{x})p(\mathbf{x})}{p(\mathbf{z})} \propto p(\mathbf{z} | \mathbf{x})p(\mathbf{x})$$

$p(\mathbf{z} | \mathbf{x})$: vraisemblance, $p(\mathbf{x})$: loi *a priori* sur \mathbf{x} , $p(\mathbf{x} | \mathbf{z})$: loi *a posteriori*

■ Décision $p(\mathbf{x} | \mathbf{z}) \overset{?}{\rightsquigarrow} \hat{\mathbf{x}}$

- **Maximum *a posteriori*** : $\hat{\mathbf{x}}^{\text{MAP}} = \arg \max_{\mathbf{x}} p(\mathbf{x} | \mathbf{z})$
- **Espérance *a posteriori*** : $\hat{\mathbf{x}}^{\text{EAP}} = \mathbb{E}[\mathbf{X} | \mathbf{Z} = \mathbf{z}] = \int \mathbf{x}p(\mathbf{x} | \mathbf{z}) d\mathbf{x}$
- **MAP Marginal** : $\hat{x}_m^{\text{MAPM}} = \arg \max_{x_m} p(x_m | \mathbf{z}), \quad 1 \leq m \leq M$
- **Estimée linéaire d'erreur moyenne quadratique minimale** : $\hat{\mathbf{x}}^{\text{ELMQ}} = \mathbf{R}_{\mathbf{xz}}\mathbf{R}_{\mathbf{z}}^{-1}(\mathbf{Z} - \mathbf{m}_{\mathbf{z}}) + \mathbf{m}_{\mathbf{x}}$

avec $\mathbf{m}_{\mathbf{x}} = \mathbb{E}[\mathbf{X}]$, $\mathbf{m}_{\mathbf{z}} = \mathbb{E}[\mathbf{Z}]$, $\mathbf{R}_{\mathbf{xz}} = \mathbb{E}[\mathbf{X}\mathbf{Z}^t] - \mathbf{m}_{\mathbf{x}}\mathbf{m}_{\mathbf{z}}^t$, $\mathbf{R}_{\mathbf{z}} = \mathbb{E}[\mathbf{Z}\mathbf{Z}^t] - \mathbf{m}_{\mathbf{z}}\mathbf{m}_{\mathbf{z}}^t$

■ Théorie des coûts bayésiens

Etant donné la fonction de coût $C(\text{estimée}, \text{vraie valeur})$, $\hat{x}(\cdot)$ minimise au sens fonctionnel le coût moyen

$$E[C(\hat{x}(\mathbf{Z}), \mathbf{X}^*)] = \iint C(\hat{x}(\mathbf{z}), \mathbf{x}^*) p(\mathbf{z} | \mathbf{x}^*) p(\mathbf{x}^*) d\mathbf{z} d\mathbf{x}^*$$

Par exemple :

\hat{x}	$C(\mathbf{x}, \mathbf{x}^*)$
$\hat{x}^{\text{MAP}} = \arg \max_{\mathbf{x}} p(\mathbf{x} \mathbf{z})$	$-\delta(\mathbf{x} - \mathbf{x}^*)$
$\hat{x}^{\text{EAP}} = E[\mathbf{X} \mathbf{Z} = \mathbf{z}]$	$\ \mathbf{x} - \mathbf{x}^*\ ^2$ (risque quadratique)
$\hat{x}_m^{\text{MAPM}} = \arg \max_{x_m} p(x_m \mathbf{z}), \quad 1 \leq m \leq M$	$-\sum_m \delta(x_m - x_m^*)$
$\hat{x}^{\text{ELMQ}} = \mathbf{R}_{xz} \mathbf{R}_z^{-1} (\mathbf{Z} - \mathbf{m}_z) + \mathbf{m}_x$	$\ \mathbf{x} - \mathbf{x}^*\ ^2$ sous contrainte de linéarité

Lien entre régularisation par pénalisation et maximum *a posteriori*

$$\hat{\mathbf{x}}^{\text{MAP}} = \arg \max_{\mathbf{x}} p(\mathbf{x} | \mathbf{z}) = \arg \max_{\mathbf{x}} p(\mathbf{z} | \mathbf{x})p(\mathbf{x}) = \arg \min_{\mathbf{x}} (-\ln p(\mathbf{z} | \mathbf{x}) - \ln p(\mathbf{x}))$$

C'est donc un cas particulier d'estimateur de maximum de vraisemblance pénalisée

Cadre « énergétique »



Cadre probabiliste

énergie $\Phi(\mathbf{x})$



densité de proba $p(\mathbf{x}) = \frac{e^{-\Phi(\mathbf{x})/T}}{\int_{\mathbf{x}} e^{-\Phi(\mathbf{x})/T} d\mathbf{x}}$

$$\mathcal{J}(\mathbf{x}) = \Omega(\mathbf{z} - \mathbf{A}\mathbf{x}) + \mu\Phi(\mathbf{x})$$



règle de Bayes $p(\mathbf{x}, \mathbf{z}) = p(\mathbf{z} | \mathbf{x})p(\mathbf{x})$

solution pénalisée $\arg \min_{\mathbf{x}} \mathcal{J}(\mathbf{x}) \equiv$

$\hat{\mathbf{x}}^{\text{MAP}} = \arg \max_{\mathbf{x}} p(\mathbf{x} | \mathbf{z})$

?

marginalisation, conditionnement
(e.g., variance d'erreur d'estimation)

■ Cadre linéaire gaussien

$z = \mathbf{H}x + b$, $b \sim \mathcal{N}(m_b, \mathbf{R}_b)$, $x \sim \mathcal{N}(m_x, \mathbf{R}_x)$, (x, b) indépendant

- *Anti-log-vraisemblance a posteriori*

$$L(x | z) \propto \frac{1}{2} \|z - \mathbf{H}x - m_b\|_{\mathbf{R}_b^{-1}}^2 + \frac{1}{2} \|x - m_x\|_{\mathbf{R}_x^{-1}}^2$$

- *Forme information*

$$\hat{x}^{\text{MAP}} = (\mathbf{H}^t \mathbf{R}_b^{-1} \mathbf{H} + \mathbf{R}_x^{-1})^{-1} (\mathbf{H}^t \mathbf{R}_b^{-1} (z - m_b) + \mathbf{R}_x^{-1} m_x)$$

- *Forme covariance*

⇕ *lemme d'inversion
de matrice*

$$\hat{x}^{\text{MAP}} = m_x + \mathbf{R}_x \mathbf{H}^t (\mathbf{R}_b + \mathbf{H} \mathbf{R}_x \mathbf{H}^t)^{-1} (z - \mathbf{H} m_x - m_b)$$

- *Quatre en un...* $\hat{x}^{\text{MAP}} = \hat{x}^{\text{EAP}} = \hat{x}^{\text{MMAP}} = \hat{x}^{\text{ELMQ}}$

- *Matrice de covariance a posteriori*

$$\begin{aligned} \mathbb{E}[(\mathbf{X} - \hat{x}^{\text{EAP}})(\mathbf{X} - \hat{x}^{\text{EAP}})^t | \mathbf{Z} = z] &= (\mathbf{H}^t \mathbf{R}_b^{-1} \mathbf{H} + \mathbf{R}_x^{-1})^{-1} \\ &= \mathbf{R}_x - \mathbf{R}_x \mathbf{H}^t (\mathbf{H} \mathbf{R}_x \mathbf{H}^t + \mathbf{R}_b)^{-1} \mathbf{H} \mathbf{R}_x \end{aligned}$$

Plan

I. Généralités

II. Déconvolution de train d'impulsions

Pénalisation convexe non quadratique

Modèle de train d'impulsions : modèle Bernoulli-Gaussien

L_2+L_1 ou L_2+L_0 ?

III. Restauration d'image et optimisation

IV. Problèmes aveugles, méthodes autodidactes

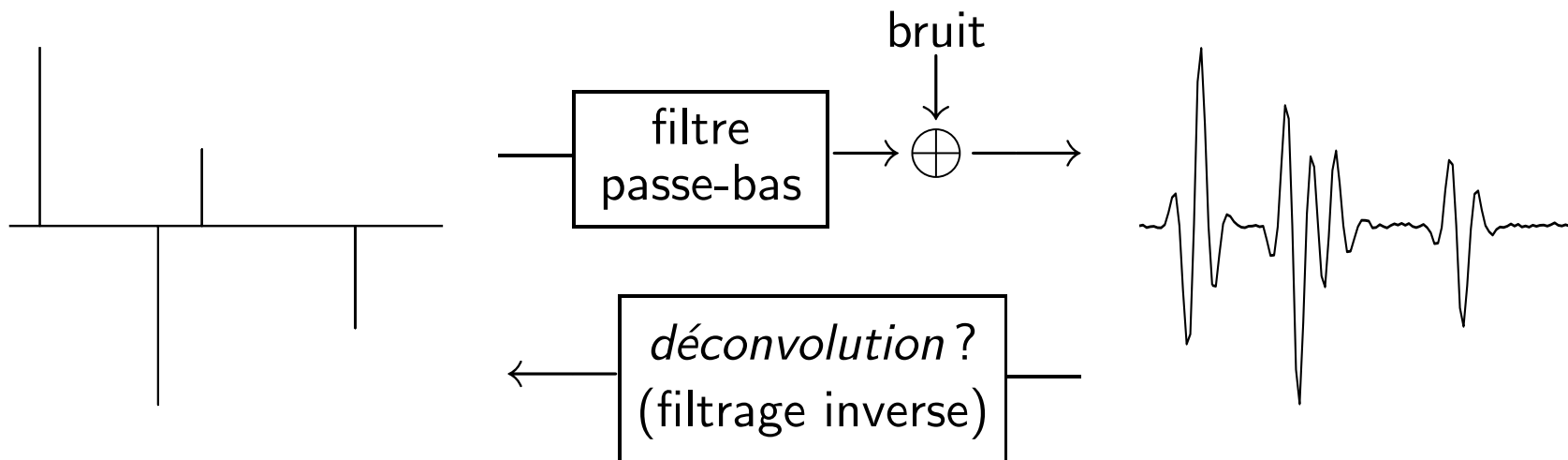
Bibliographie

Introduction

Caractérisation de défauts ponctuels, de frontières nettes, ...,
à partir de données à résolution limitée

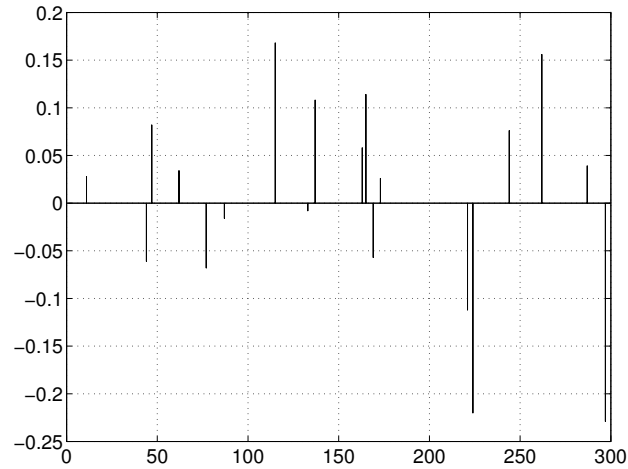
spectrométrie, astronomie, sismologie, évaluation non destructive, ...

$$\begin{aligned} \text{train d'impulsions} \star \text{ondelette} + \text{bruit} &= \text{trace} \\ x \star h + b &= z \end{aligned}$$

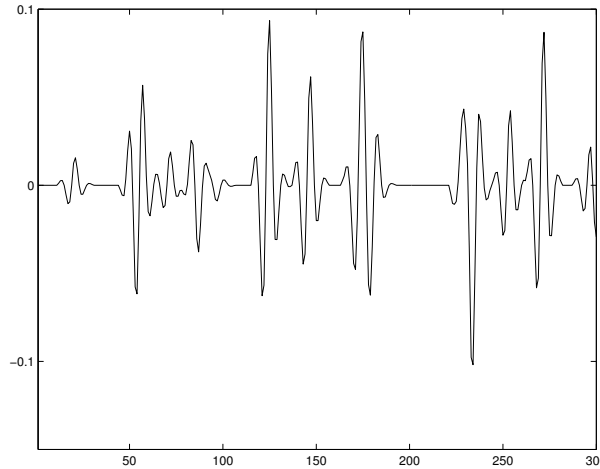


■ Exemple [Champagnat *et coll.* 2001]

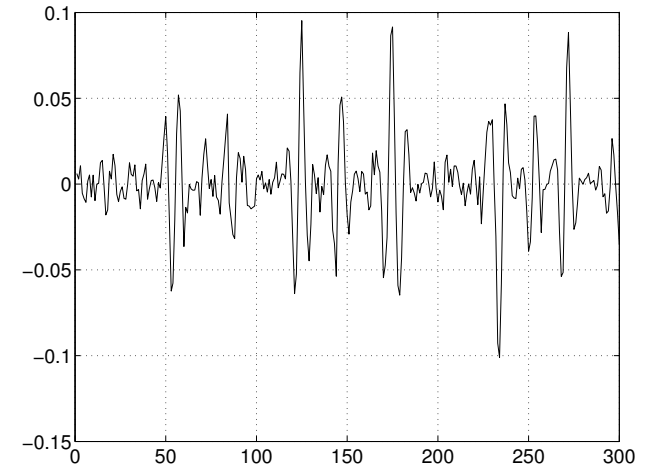
séquence de Mendel



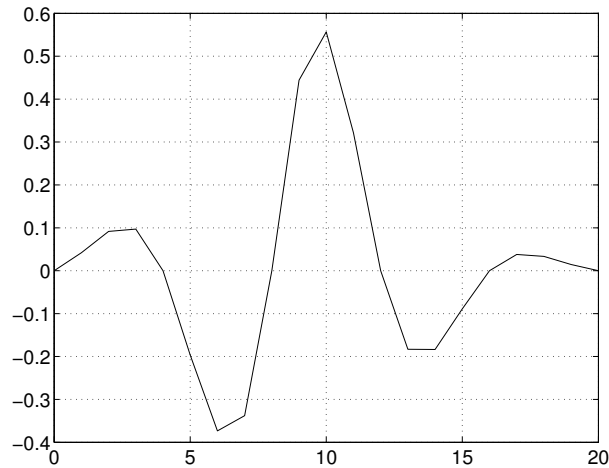
signal convolué



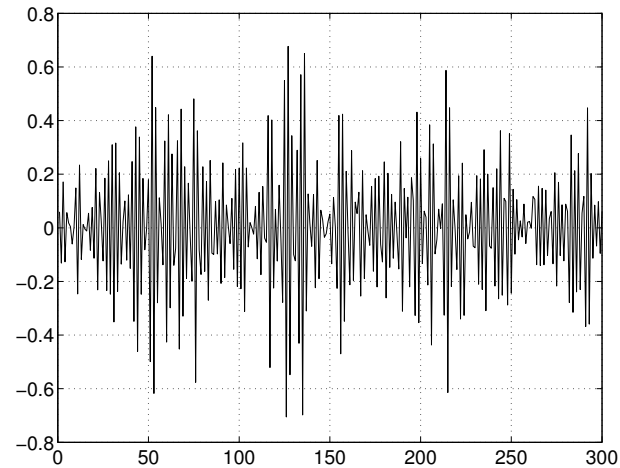
signal convolué bruité (10 dB)



ondelette



$$\hat{x} = h^{-1} \star z = x + h^{-1} \star b$$



Approches "naïves"

① Méthode des moindres carrés

trouver $\hat{\mathbf{x}}$ qui minimise $J(\mathbf{x}) = \|\mathbf{z} - \mathbf{H}\mathbf{x}\|^2$

$$\rightarrow (\mathbf{H}^t \mathbf{H}) \hat{\mathbf{x}} = \mathbf{H}^t \mathbf{z} \quad (\mathbf{H} : \text{matrice de convolution})$$

② Approche probabiliste

Maximum de vraisemblance si b est supposé blanc gaussien :

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} p(\mathbf{z} | \mathbf{x}), \quad p(\mathbf{z} | \mathbf{x}) \propto \exp \left\{ -\frac{1}{2\sigma_b^2} \|\mathbf{z} - \mathbf{H}\mathbf{x}\|^2 \right\}$$

③ Division spectrale (par transformées de Fourier)

$$\hat{X}(\nu) = \frac{Z(\nu)}{H(\nu)} \quad \left(= X(\nu) + \frac{B(\nu)}{H(\nu)} \right)$$

$$\textcircled{1} = \textcircled{2} \simeq \textcircled{3}$$

Régularisation quadratique

① Moindres carrés pénalisés (solution de Tikhonov)

trouver $\hat{\mathbf{x}}$ qui minimise $J(\mathbf{x}) = \|\mathbf{z} - \mathbf{H}\mathbf{x}\|^2 + \mu\|\mathbf{x}\|^2 \rightsquigarrow (\mathbf{H}^t\mathbf{H} + \mu\mathbf{I})\hat{\mathbf{x}} = \mathbf{H}^t\mathbf{z}$

② Approche probabiliste bayésienne

Maximum a posteriori si \mathbf{b} et \mathbf{x} sont supposés blancs gaussiens

$$\hat{\mathbf{x}} \text{ maximise } p(\mathbf{x} | \mathbf{z}) = \frac{p(\mathbf{z} | \mathbf{x}) p(\mathbf{x})}{p(\mathbf{z})},$$

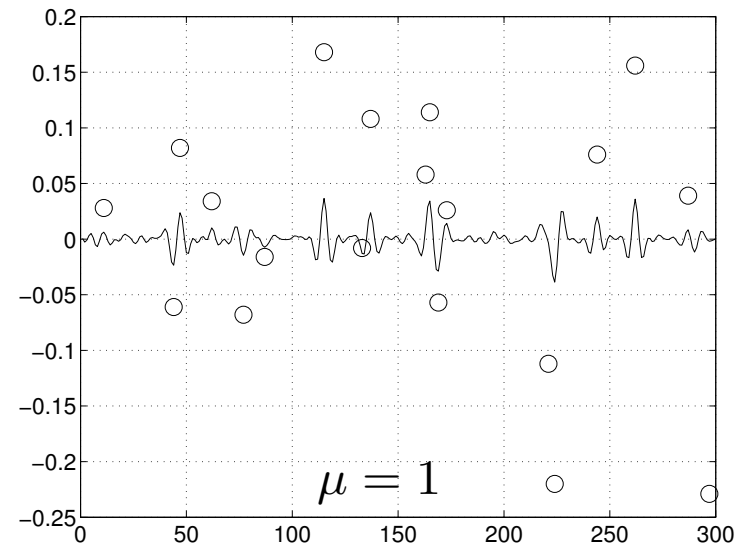
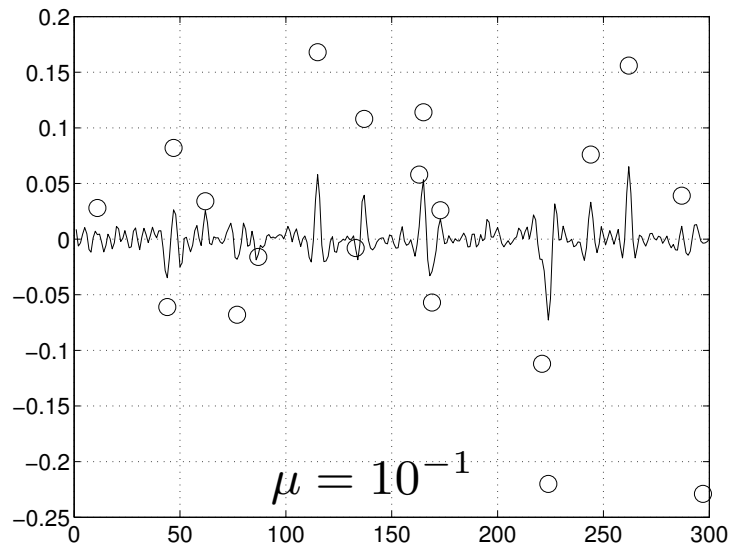
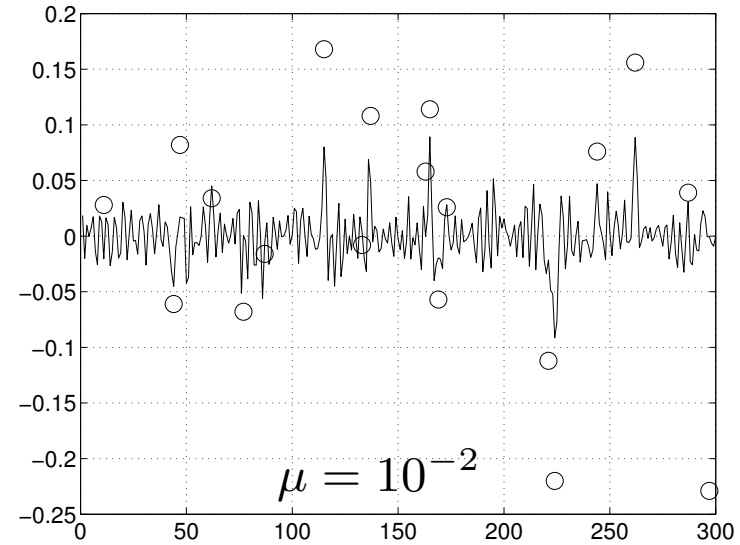
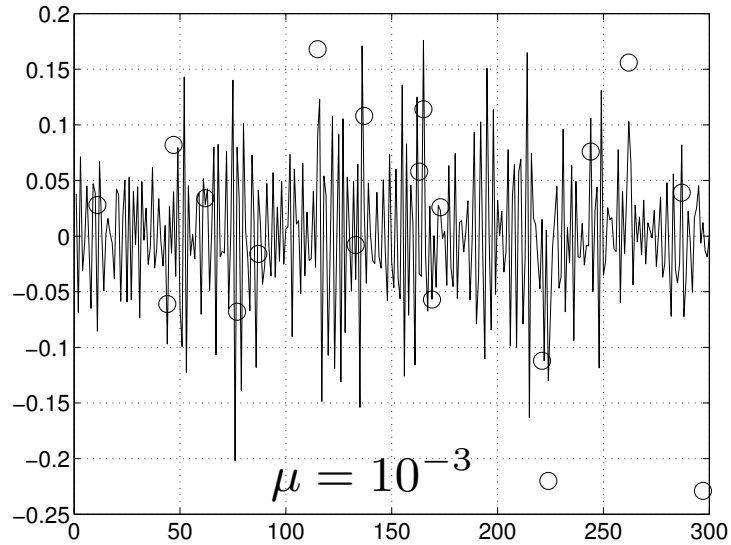
$$\iff \hat{\mathbf{x}} \text{ minimise } (-\ln p(\mathbf{z} | \mathbf{x})) + (-\ln p(\mathbf{x}))$$

③ Filtrage de Wiener :

$$\hat{X}(\nu) = \frac{H^*(\nu)}{|H(\nu)|^2 + \mu} Z(\nu) \quad \left(\mu = \frac{\sigma_b^2}{\sigma_x^2} \right)$$

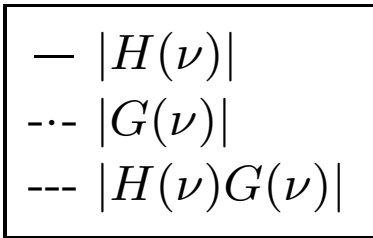
$$\textcircled{1} = \textcircled{2} \simeq \textcircled{3}$$

■ Solution linéaire simple, mais pas de compromis satisfaisant

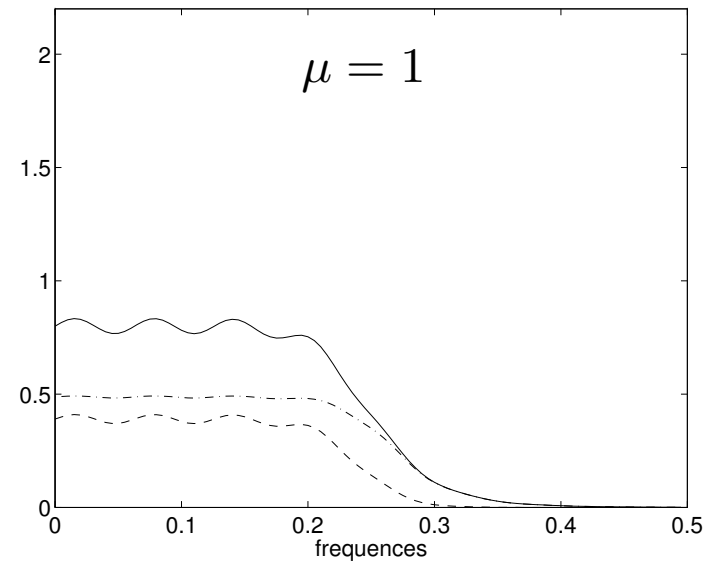
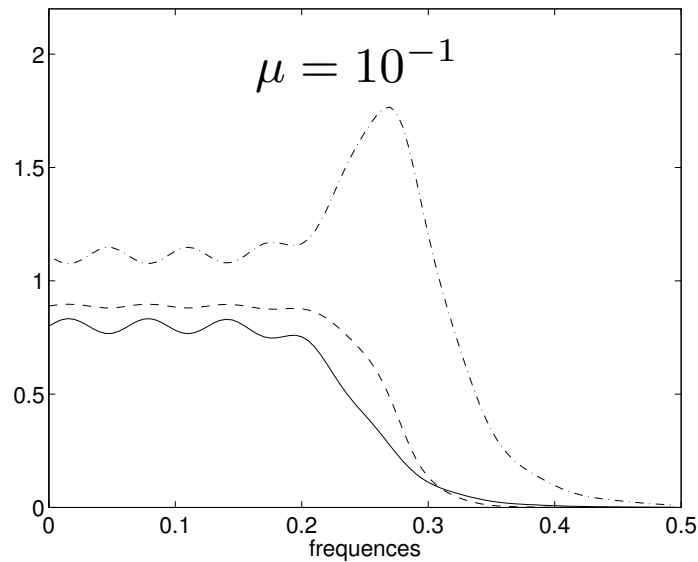
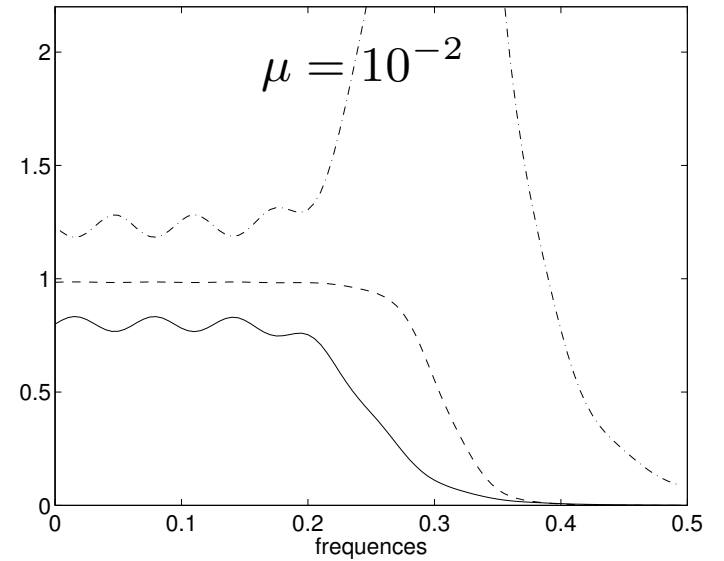
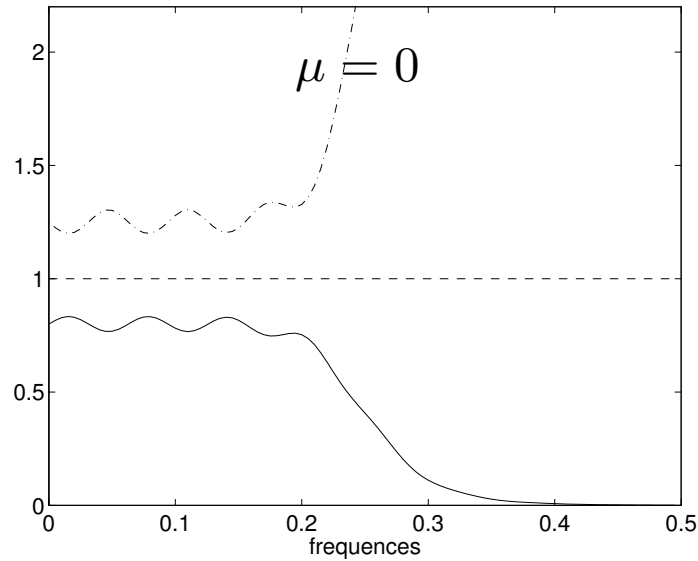


■ Égalisation spectrale dans la bande utile

$$G(\nu) = \frac{H^*(\nu)}{|H(\nu)|^2 + \mu}$$

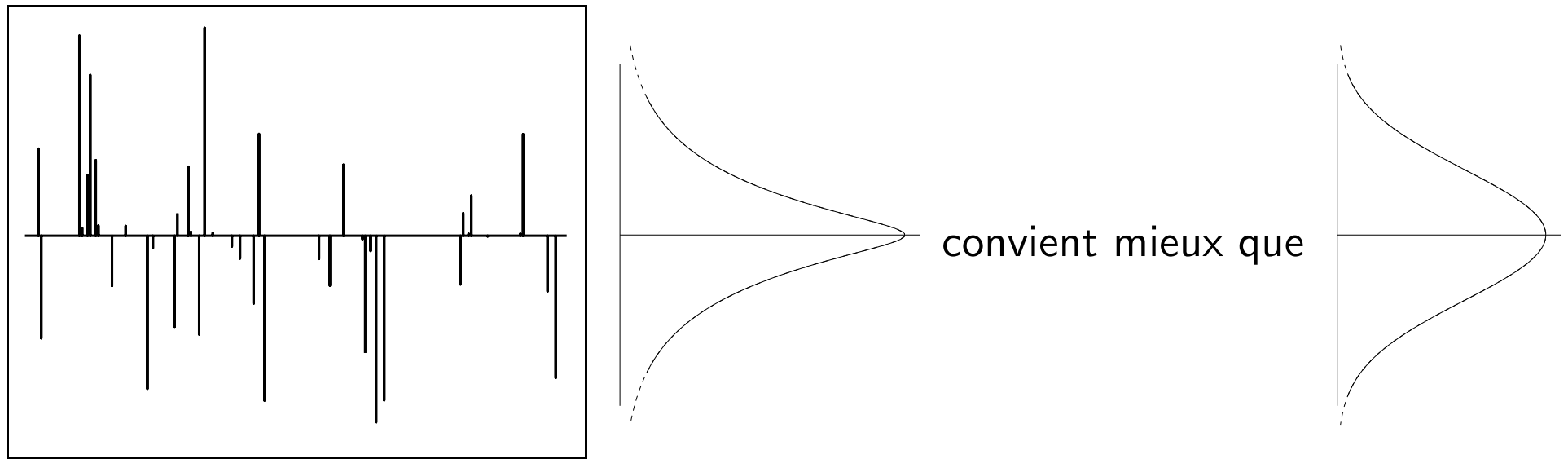


NB. $H(\nu)$ ne correspond pas à l'exemple de la p. 31



■ Vers des solutions non linéaires dans le cadre bayésien

L'histogramme d'un train d'impulsion est typiquement non gaussien !



densité de probabilité $p_{HT}(x)$ à queue plus épaisse
et plus concentrée en 0 (HT pour "heavy tail")

Pénalisation convexe non quadratique

■ Cadre bayésien

$$\text{maximiser } p(\mathbf{x} | \mathbf{z}) \propto p(\mathbf{z} | \mathbf{x}) p(\mathbf{x})$$

$$\text{où } p(\mathbf{x}) = \prod_{n=1}^N p_{\text{HT}}(x_n) \text{ est la distribution } a \text{ priori de } \mathbf{x}$$

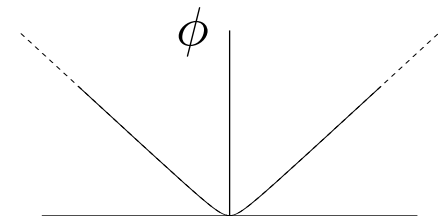
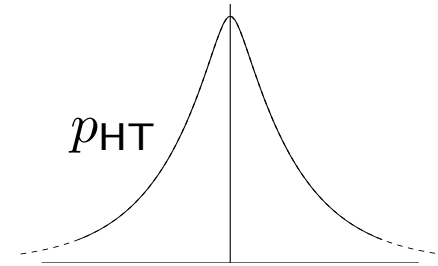


■ Moindres carrés pénalisés

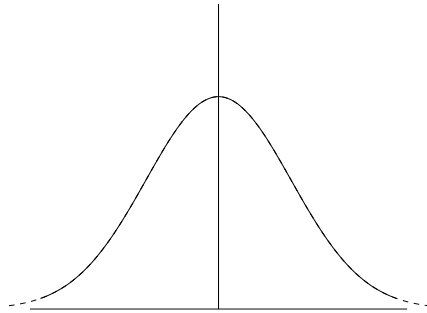
$$\text{minimiser } J(\mathbf{x}) = \|\mathbf{z} - \mathbf{h} \star \mathbf{x}\|^2 + \mu \sum_{n=1}^N \phi(x_n)$$

$$\text{où } \phi = -\ln p_{\text{HT}}$$

- $\phi(x)$ croît plus lentement que x^2
- ϕ convexe \implies pas de minima locaux

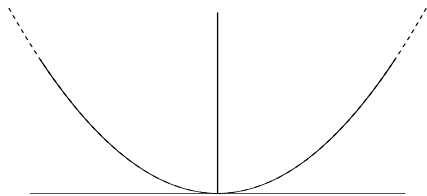


■ Exemples de lois sur \mathbb{R}

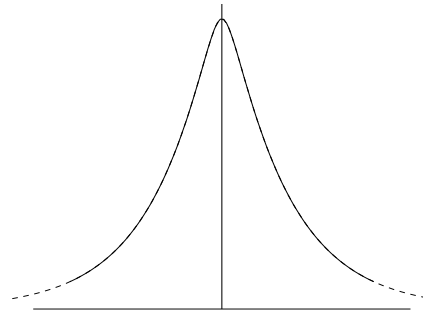


loi gaussienne

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$$

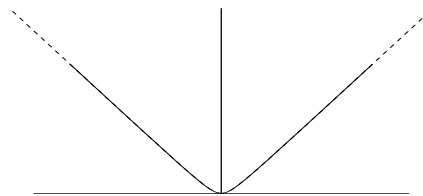


$$\phi(x) = x^2$$



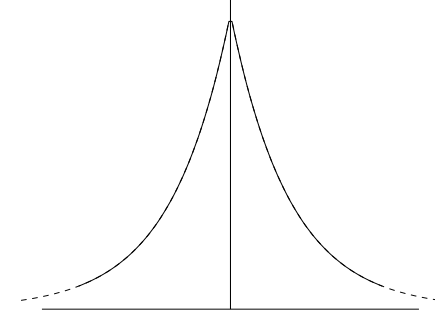
loi « hyperbolique »

$$f(x) \propto e^{-\sqrt{\delta^2+x^2}/T}$$



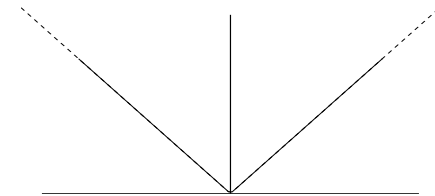
$$\phi_{\text{hyp}}(x) = \sqrt{\delta^2 + x^2} - \delta$$

$$\delta > 0$$



loi de Laplace

$$f(x) = \frac{1}{2T} e^{-|x|/T}$$

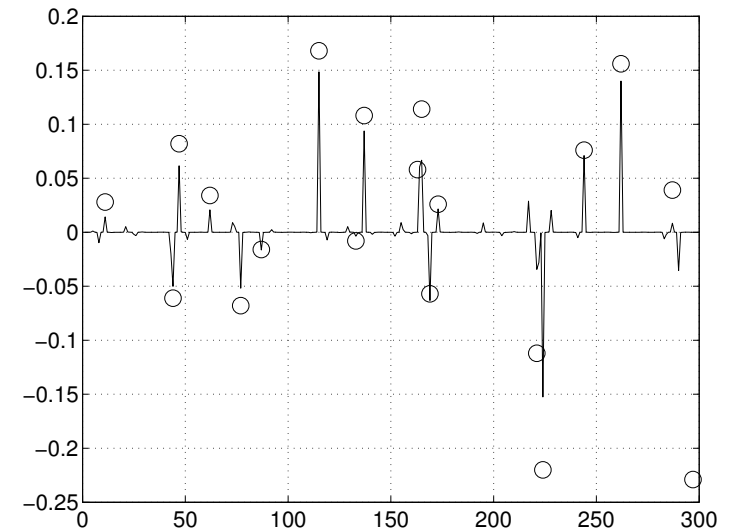


$$\phi_{L1}(x) = |x|$$

■ Calcul de \hat{x} (minimisation de J)

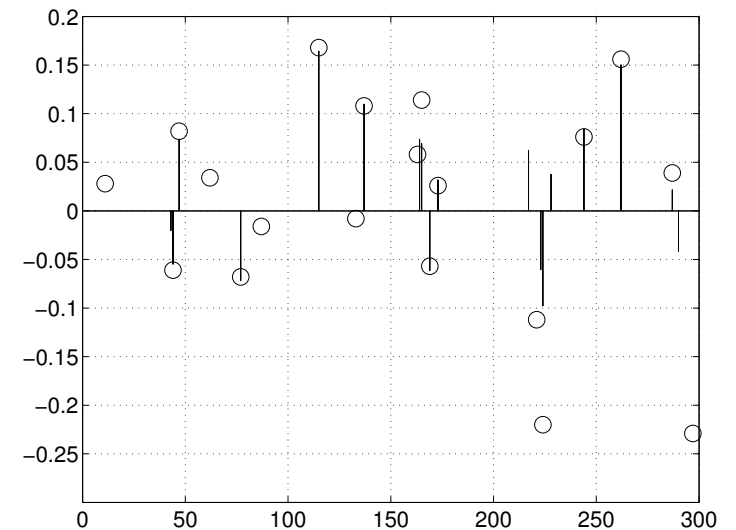
- ϕ différentiable
Optimisation locale (e.g., gradient conjugué)
[Bertsekas 1995, Nocedal et Wright 1999]

*solution
"hyperbolique"*



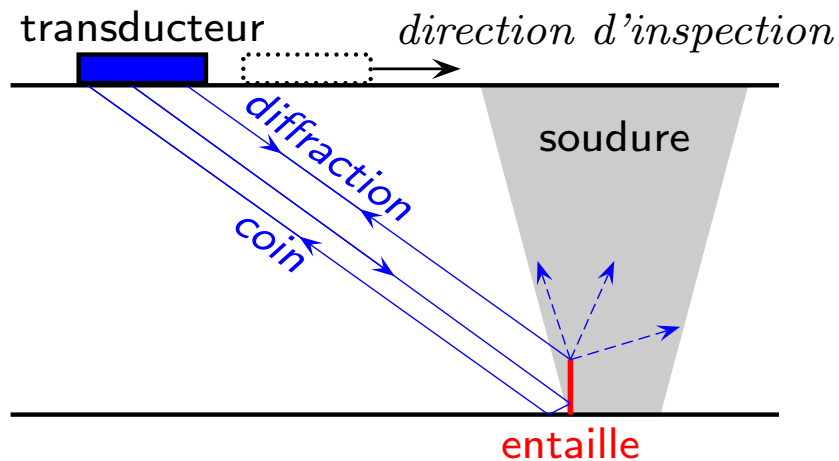
- $\phi_{L1}(x) = |x|$: Problème "L2L1"
Algorithmes plus spécifiques (homotopie, IT, ...)
voir [IEEE Selected Topics in Signal Processing,
Issue : Convex Optimization Methods, déc. 2007]

solution L1



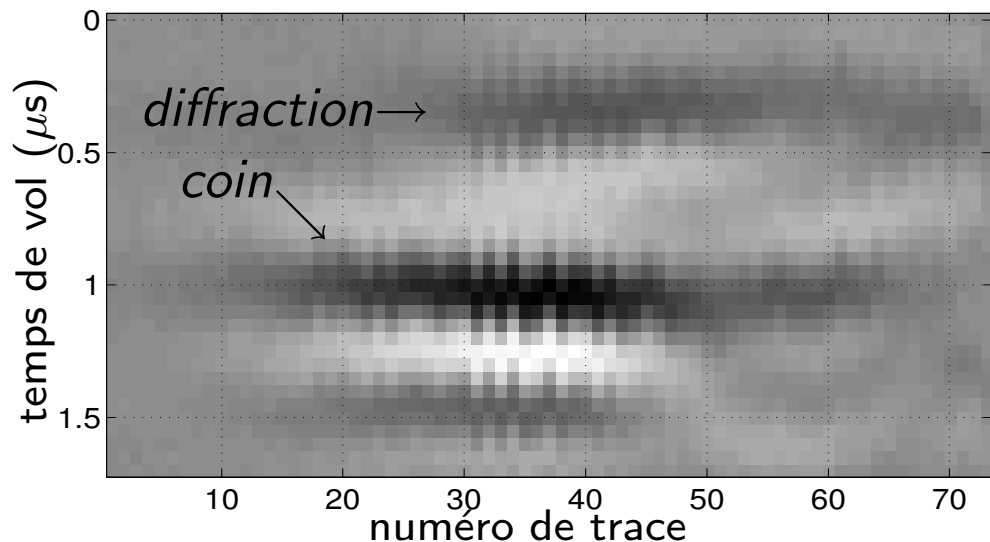
Application : contrôle ultrasonore de soudures

en collaboration avec EDF-DRD (Chatou)

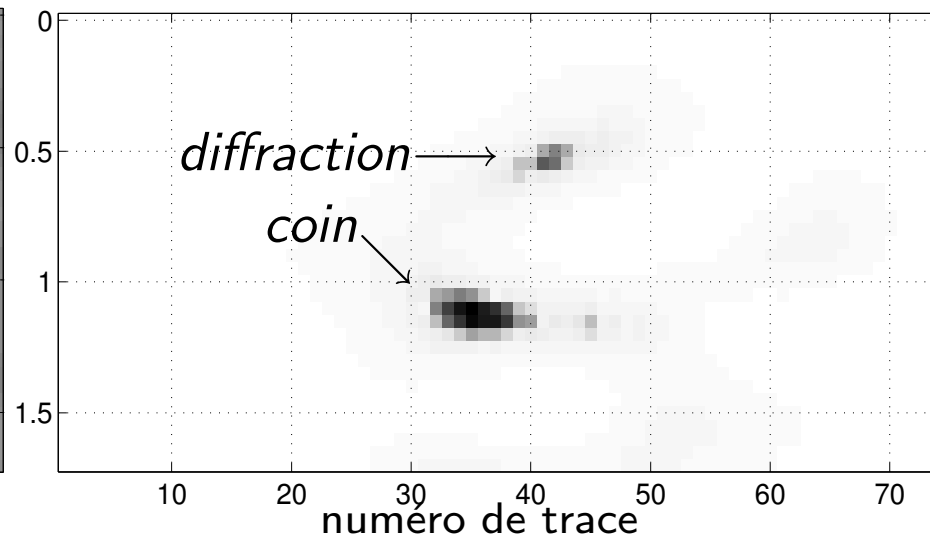


exemple tiré de [Labat et coll. 2005]

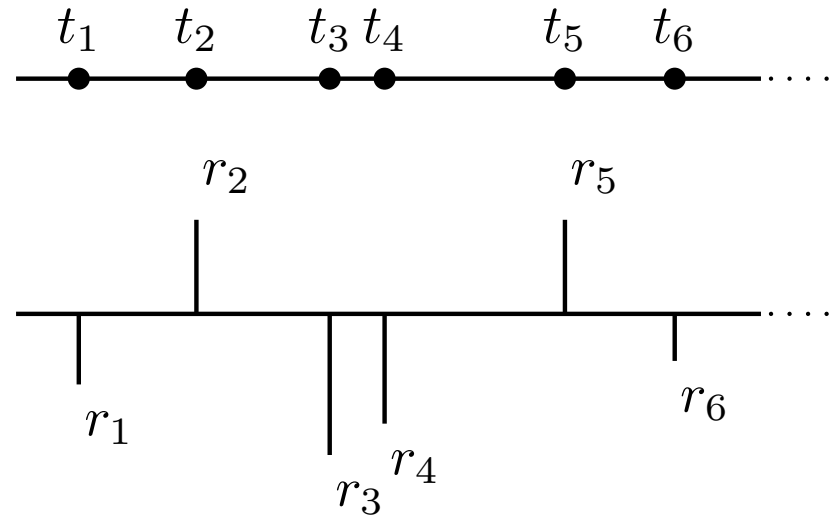
données brutes (zoom)



déconvolution 2D "hyperbolique"



Alternative : modélisation de train d'impulsions



Processus composé (t_m, r_m) : t_m est un processus ponctuel qui commande l'apparition des événements, dont l'amplitude est distribuée suivant la loi du processus r_m .

↙
variante « continue »
processus de Poisson

↘
variante « discrète »
processus de Bernoulli

+ amplitudes gaussiennes

Modèle Bernoulli-Gaussien

Loi a priori

- \mathbf{q} vecteur binaire indépendant

$$P(q_m = 1) = \lambda$$

$$P(q_m = 0) = 1 - \lambda \quad \rightarrow \quad P(\mathbf{q}) = \lambda^{M_1} (1 - \lambda)^{M_0}, \quad M_0 + M_1 = M$$

λ : « densité d'événements »

- \mathbf{r} vecteur gaussien indépendant

$$r_m \sim \mathcal{N}(0, \sigma^2 q_m)$$

$$q_m = 0 \Rightarrow r_m = 0 \qquad q_m = 1 \Rightarrow r_m \sim \mathcal{N}(0, \sigma^2)$$

■ Calcul du MAP

- Problème de détection-estimation de type "L2L0" :

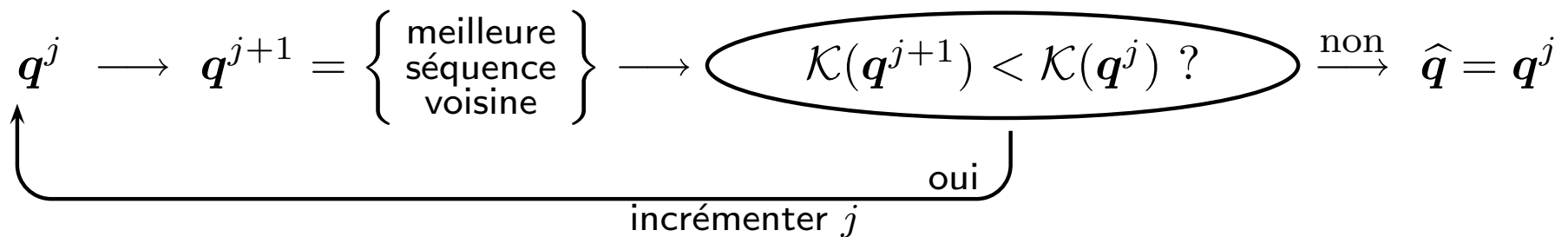
$$\mathcal{K}(\mathbf{q}) = \min_{\mathbf{r}} \|\mathbf{z} - \mathbf{h} \star \mathbf{r}\|^2 + \mu \|\mathbf{r}\|_2^2 + \beta \sum_m q_m \text{ avec } q_m = 1_{\{r_m \neq 0\}}$$

- Optimisation combinatoire

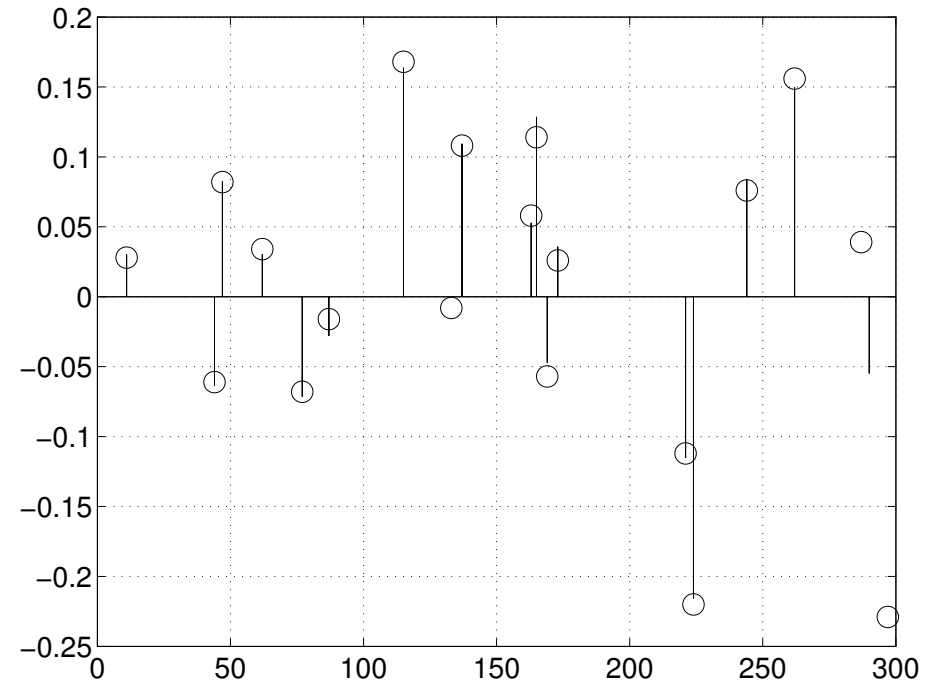
$\mathbf{q} \in \{0, 1\}^M$, soit 2^M possibilités \Rightarrow examen exhaustif impossible

\rightsquigarrow Algorithmes « sous-optimaux » itératifs : $\mathbf{q}^j \rightarrow \mathbf{q}^{j+1} \dots \rightarrow \mathbf{q}^J = \hat{\mathbf{q}}$

Méthode de parcours : *Single Most Likely Replacement* (SMLR) [Mendel 1983]



■ Résultats



L2L1 ou L2L0 ?

■ Formulation L2L1, L2Lhyp, etc

- Modèle plus grossier
- Problème convexe \rightsquigarrow solution exacte
- Solution continue en fonction des données [Bouman et Sauer 1993]
- Adaptée à l'aide au diagnostic

■ Formulation L2L0

- Modélisation plus « exacte » mais solution approchée
- Solution discontinue en fonction des données
- Adaptée à la décision automatique
(e.g., application en IRMf, cf. Ph. Ciuciu)

Plan

I. Généralités

II. Déconvolution de train d'impulsions

III. Restauration d'image et optimisation

Modèle pour les images ? Modèles markoviens, analyse/synthèse

Critères semi-quadratiques, approximations majorantes

Formes tronquées, stratégies de pas

IV. Problèmes aveugles, méthodes autodidactes

Bibliographie

Introduction

Estimer \boldsymbol{x} à partir de $\boldsymbol{z} = \mathbf{H}\boldsymbol{x} + \boldsymbol{n}$, \mathbf{H} étant connue

\mathbf{H} est une matrice de convolution, donc Toeplitz par blocs à blocs Toeplitz

■ Exemple (bateau de pêche) \boldsymbol{x}  \boldsymbol{z}

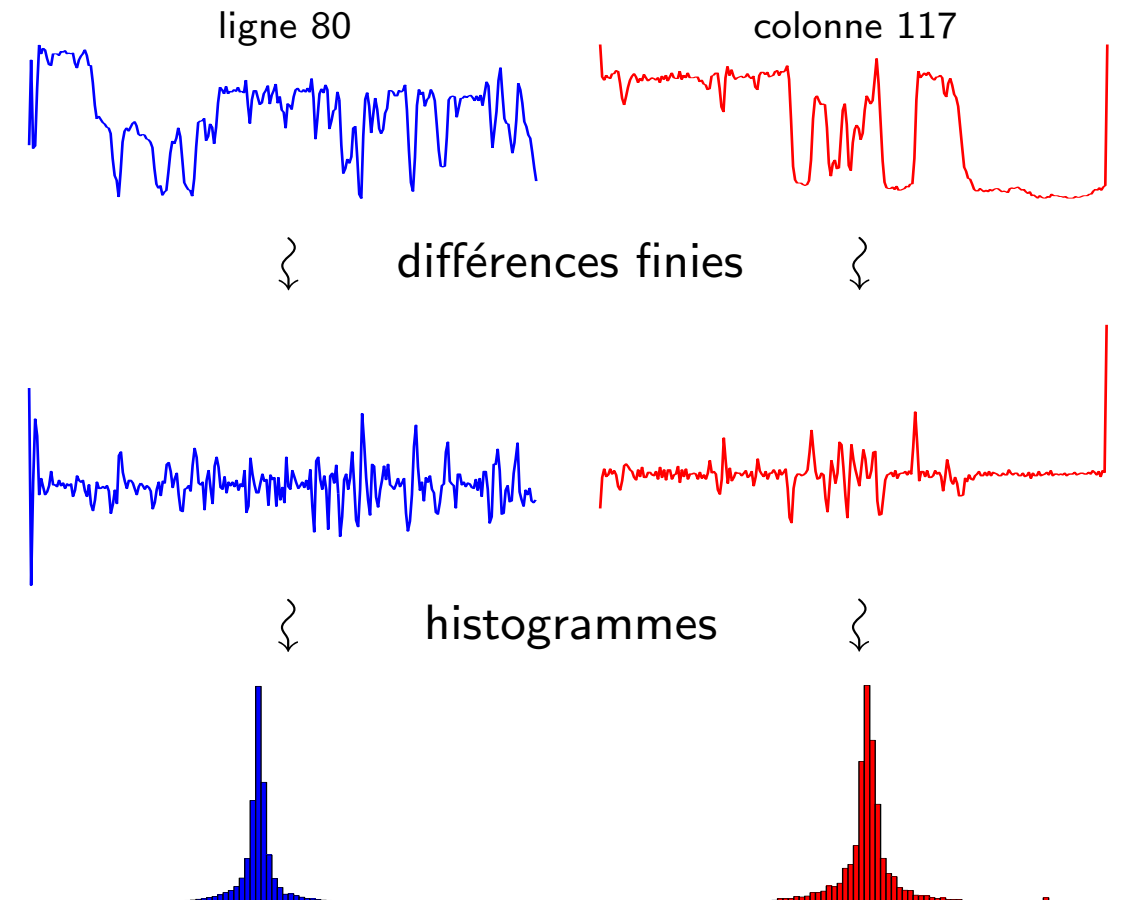
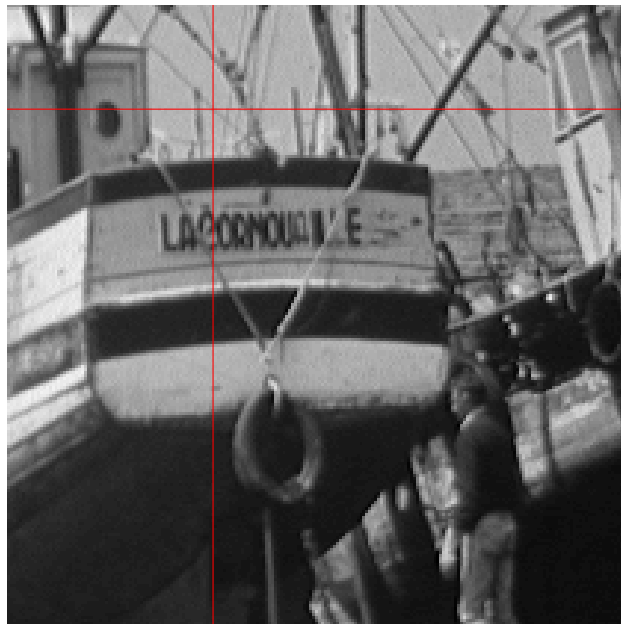
$N = 512 \times 512$
PSF gaussienne
($\sigma = 2.24$ pixels)
RSB 40dB

■ **Ingrédients nécessaires :**

- Problème mal posé \rightsquigarrow régularisation par pénalisation \rightsquigarrow modèle d'images ?
- Algorithme d'optimisation adapté aux problèmes de grande taille

Modèles *a priori* pour les images

Modèles markoviens

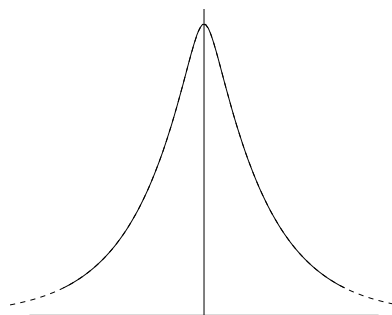


↷ Energie d'un **champ de Markov aux plus proches voisins** :

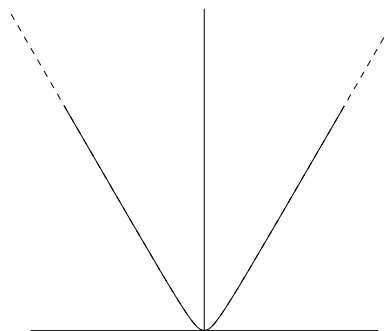
$$\Phi(\mathbf{x}) = \sum_{m,n} \phi(x_{m,n} - x_{m,n+1}) + \sum_{m,n} \phi(x_{m,n} - x_{m+1,n})$$

Préservation des discontinuités : ϕ associée à une loi à queue lourde, e.g.,

$$f(x) \propto e^{-\sqrt{\tau^2+x^2}/T}$$



$$\phi(t) = \sqrt{\delta^2 + t^2}$$



Charbonnier *et coll.* [1997]

■ Variante « **variation totale** »

$$\Phi(\mathbf{x}) = \sum_{m,n} \sqrt{(x_{m,n} - x_{m,n+1})^2 + (x_{m,n} - x_{m+1,n})^2}$$

Nondifférentiabilité ↷ parcimonie, mais « marches d'escalier » [Nikolova 2004]

■ Les champs de Markov sont-ils des bons modèles ?

✘ La loi des différence entre pixels voisins n'est pas $c \exp \{-\phi(x)\}$

✘ Aucun contrôle des propriétés au second ordre

≠ modèle gaussien, e.g., $DSP(\nu) = \frac{C}{\nu_0^p + \nu^p}$ [Kattinig et Primot 1997]



Comment imposer simultanément

— une structure de covariance spatiale,

— la loi marginale des pixels (ou des différences interpixels) ?

✔ OK malgré tout pour définir un « bon » estimateur $\hat{x}(\mu)$ à μ fixé

● Et l'estimation de μ ?

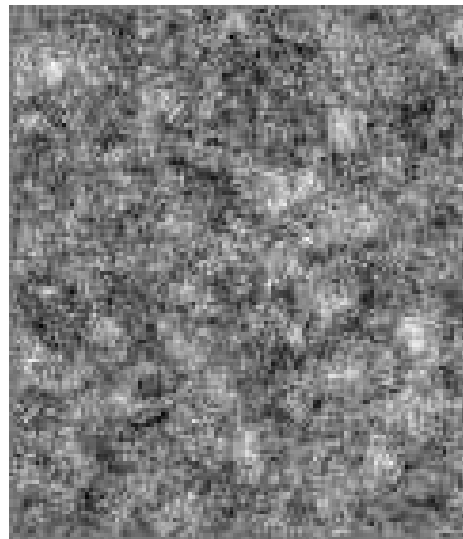
Une expérience embarrassante

- Soit $z = h \star x^* + b$ dans deux cas différents :

Cas « réel »

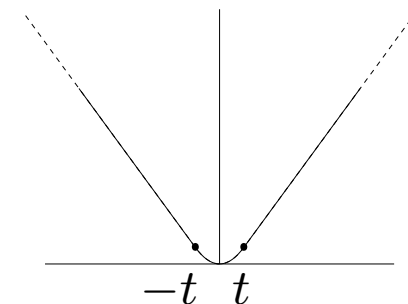


Cas « synthétique »



*réalisation d'un champ
de Markov-Huber*

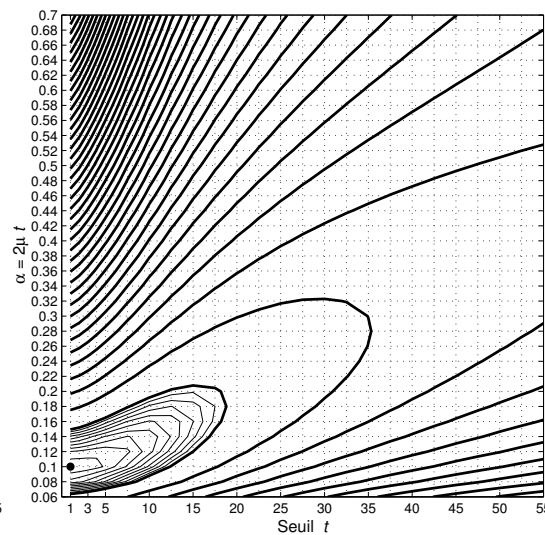
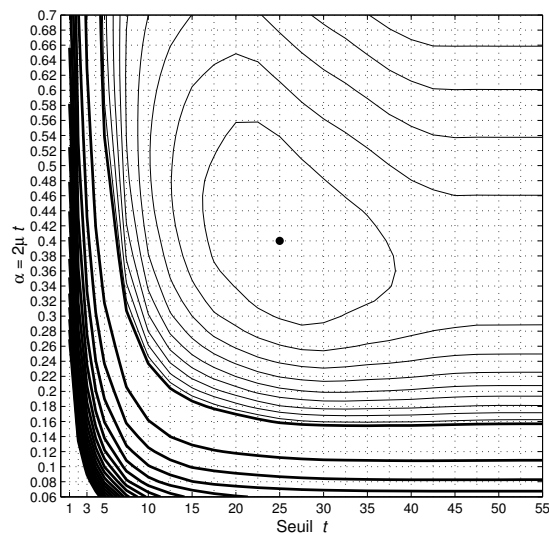
fonction de Huber ϕ_t



- Soit $\hat{x}^{\text{MAP}} = \arg \min_x \left(\frac{1}{2\sigma^2} \|z - h \star x\|^2 + \mu \sum_{r \sim s} \phi_t(x_r - x_s) \right)$

■ Cas « réel »

$$\left\| \hat{x}^{\text{MAP}} - x^* \right\|_1$$



Pseudo-vraisemblance
(fonction de x^*)

$\hat{x}^{\text{MAP}}(\mu, t)^{\text{MR1}}$
MR1 :
meilleur réglage L1



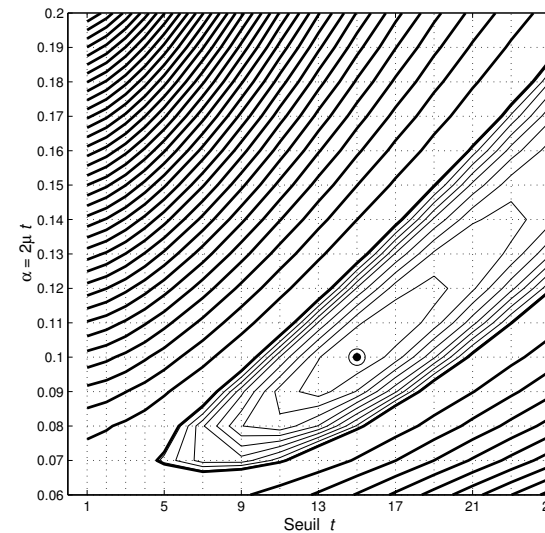
$$\hat{x}^{\text{MAP}}(\mu, t)^{\text{PMV}}$$

$$\hat{x}^{\text{MAP}}(\mu, 0)^{\text{MR1}}$$



■ Cas « synthétique »

Pseudo-vraisemblance
(vraies valeurs de (μ, t) , PMV,
MR1 et MR2 se confondent



■ Les choses se passent parfois mieux, e.g., [Trillon et coll. \[2008\]](#)

Approche « analyse » / Approche « synthèse » [Elad et coll. 2007]

■ Approche par « analyse » (champ de Markov)

$$\hat{\mathbf{x}} = \arg \min_x \left(\|\mathbf{z} - \mathbf{H}\mathbf{x}\|^2 + \mu \sum_{\ell=1}^L \phi(\alpha_\ell) \right)$$

avec $\boldsymbol{\alpha} = \mathbf{V}\mathbf{x} \in \mathbb{R}^L$ et $L \geq N$

■ Approche par « synthèse » (bases redondantes)

Frame d'ondelettes, représentation temps-fréquence,
décomposition sur un dictionnaire :

$$\hat{\mathbf{x}} = \mathbf{W}\hat{\boldsymbol{\alpha}} \text{ avec } \boldsymbol{\alpha} \in \mathbb{R}^L \text{ ou } \mathbb{C}^{L/2}, L \gg N,$$

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \left(\|\mathbf{z} - \mathbf{H}\mathbf{W}\boldsymbol{\alpha}\|^2 + \mu \sum_{\ell=1}^L \phi(\alpha_\ell) \right)$$

et ϕ est préférentiellement **non différentiable** pour créer de la parsimonie

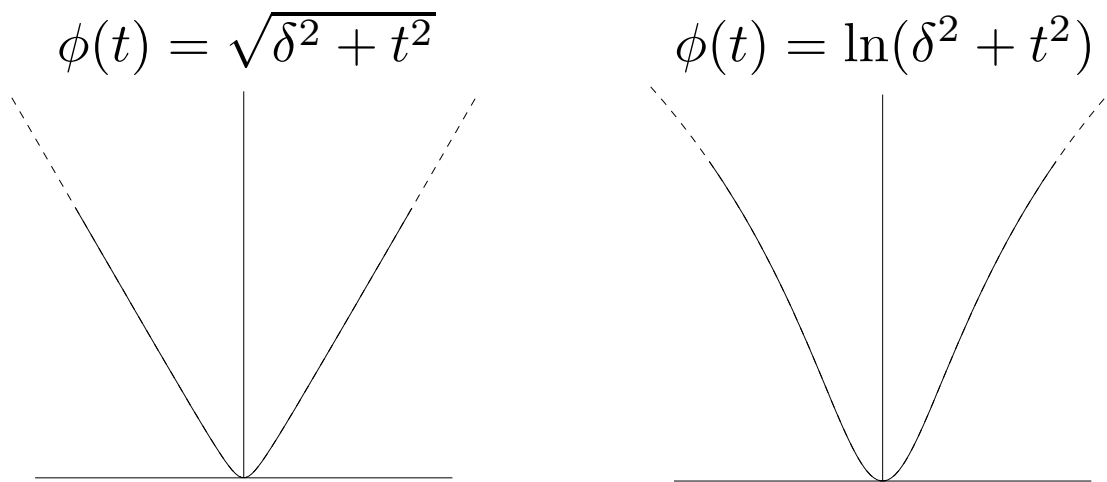
Optimisation de critères pénalisés :
Approximations majorantes pour la restauration d'image

Contexte markovien

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \mathcal{J}(\mathbf{x}) = \|\mathbf{z} - \mathbf{H}\mathbf{x}\|^2 + \mu \sum_{c \in \mathcal{C}} \phi(\mathbf{v}_c^t \mathbf{x})$$

e.g., $\mathbf{v}_c^t \mathbf{x} = x_r - x_s$, $c = \{r, s\}$: paires de pixels voisins horizontaux ou verticaux

- Préservation des discontinuités $\rightsquigarrow \phi$ non quadratique



Charbonnier et coll. [1997] Hebert et Leahy [1989]

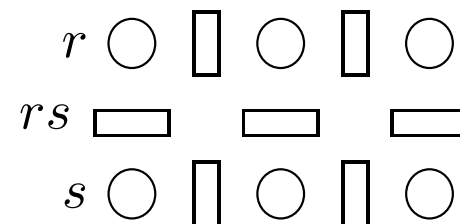
- Pas de marches d'escalier $\rightsquigarrow \phi$ différentiable

Algorithmes semi-quadratiques

■ Origine : modèles composites et variables binaires de contours

[Blake et Zisserman 1987, Jeng et Woods 1991]

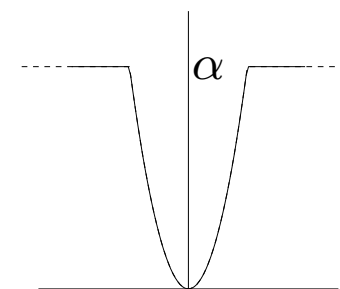
e.g., $\Psi(\mathbf{x}, \ell) = \sum_{r \sim s} (1 - \ell_{rs})(x_r - x_s)^2 + \alpha \sum_{r \sim s} \ell_{rs}$



- $\mathcal{K}(\mathbf{x}, \ell) = \|\mathbf{z} - \mathbf{H}\mathbf{x}\|^2 + \mu\Psi(\mathbf{x}, \ell)$ est *semi-quadratique* (SQ)

- $\min_{\mathbf{x}, \ell} \mathcal{K}(\mathbf{x}, \ell) = \min_{\mathbf{x}} \left(\|\mathbf{z} - \mathbf{H}\mathbf{x}\|^2 + \mu\Phi(\mathbf{x}) \right)$

avec $\Phi(\mathbf{x}) = \sum_{r \sim s} \phi(x_r - x_s)$, $\phi(x) = \min \{ \alpha, x^2 \}$
 quadratique tronquée



Algorithme SQ = Minimisation alternée d'un critère composite \mathcal{K}

Extension à d'autres fonctions ϕ ? Geman et Reynolds [1992], Geman et Yang [1995]

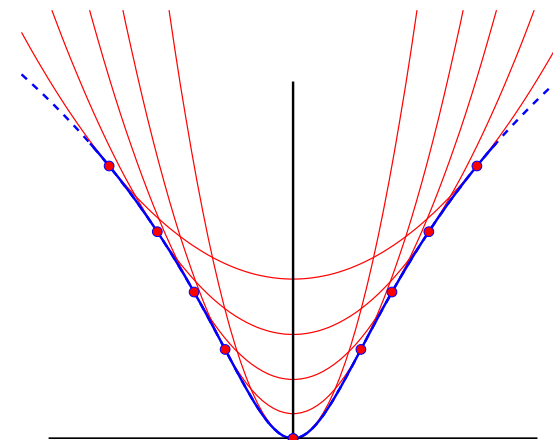
Deux variantes SQ

■ **ARTUR/GR** [Geman et Reynolds 1992]

$\phi \in \mathcal{C}^1$, paire, $\phi(\sqrt{\cdot})$ concave sur \mathbb{R}^+

⇓

$$\exists \psi \text{ tel que } \phi(u) = \inf_{\ell \in \mathbb{R}} (\ell u^2 + \psi(\ell))$$

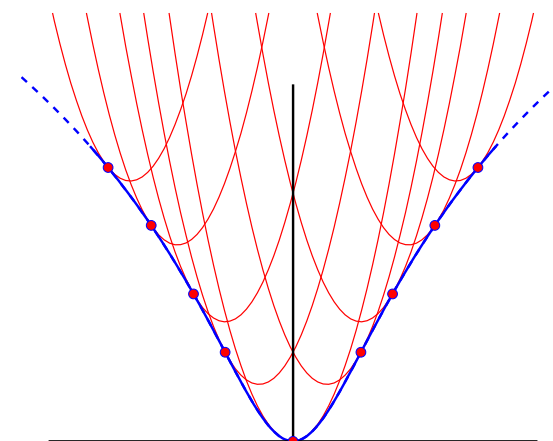


■ **LEGEND/GY** [Geman et Yang 1995]

$\frac{1}{2}u^2 - \phi$ convexe

⇓

$$\exists \zeta \text{ tel que } \phi(u) = \inf_{\ell \in \mathbb{R}} \left(\frac{1}{2}(u - \ell)^2 + \zeta(\ell) \right)$$



■ **Remarques**

$\ell(\cdot)^2 + \psi(\ell)$ et $(\cdot - \ell)^2 + \zeta(\ell)$ sont quadratiques, majorantes et tangentes à ϕ

De même, $\Psi(\cdot, \ell)$ est quadratique, majorante et tangente à Φ

Approximations quadratiques majorantes (AMQ)

Ortega et Rheinboldt [1970], Voss et Eckhardt [1980]

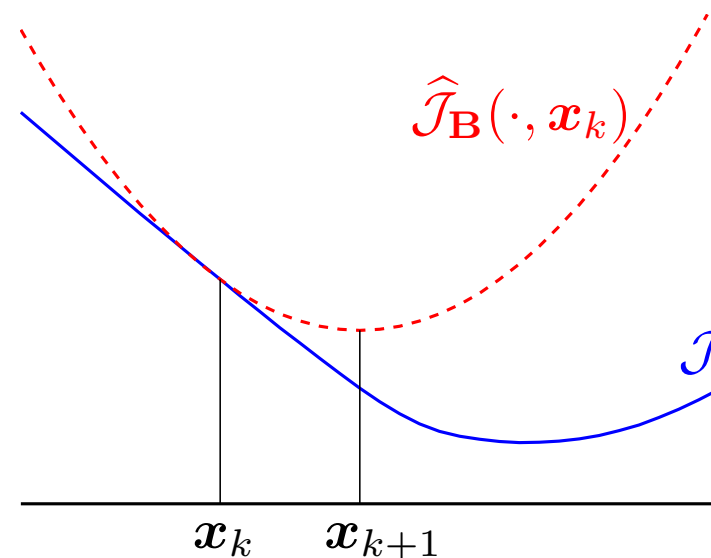
Soit $\hat{\mathcal{J}}_{\mathbf{B}}(\mathbf{x}', \mathbf{x}) = \mathcal{J}(\mathbf{x}) + (\mathbf{x}' - \mathbf{x})^t \nabla \mathcal{J}(\mathbf{x}) + (\mathbf{x}' - \mathbf{x})^t \mathbf{B}(\mathbf{x}) (\mathbf{x}' - \mathbf{x})/2$

Supposons qu'il existe $\mathbf{B}(\cdot) > 0$ tel que $\hat{\mathcal{J}}_{\mathbf{B}}(\mathbf{x}', \mathbf{x}) \geq \mathcal{J}(\mathbf{x}')$, $\forall \mathbf{x}, \mathbf{x}' \in \mathbb{R}^N$

Alors :

- $\hat{\mathcal{J}}_{\mathbf{B}}(\mathbf{x}, \mathbf{x}) = \mathcal{J}(\mathbf{x})$,
- $\min_{\mathbf{x}'} \hat{\mathcal{J}}_{\mathbf{B}}(\mathbf{x}', \mathbf{x}) \leq \mathcal{J}(\mathbf{x})$,
- $\arg \min_{\mathbf{x}'} \hat{\mathcal{J}}_{\mathbf{B}}(\mathbf{x}', \mathbf{x}) = \mathbf{x} - \mathbf{B}^{-1}(\mathbf{x}) \nabla \mathcal{J}(\mathbf{x})$

algorithmme AMQ :
 $\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{B}^{-1}(\mathbf{x}_k) \nabla \mathcal{J}(\mathbf{x}_k)$



- $\mathbf{B}_{\text{GY}} = 2\mathbf{H}^t \mathbf{H} + \mu \mathbf{V}^t \mathbf{V}$, $\mathbf{V} = [\mathbf{v}_1 | \dots | \mathbf{v}_C]^t$
- $\mathbf{B}_{\text{GR}}(\mathbf{x}) = 2\mathbf{H}^t \mathbf{H} + \mu \mathbf{V}^t \mathbf{L}(\mathbf{x}) \mathbf{V}$, $\mathbf{L}(\mathbf{x}) = \text{Diag} \{ \phi'(\mathbf{v}_c^t \mathbf{x}) / \mathbf{v}_c^t \mathbf{x} \}$

■ Bref historique

- **Approche markovienne en restauration d'image**
[Geman et Reynolds 1992, Geman et Yang 1995, Charbonnier *et coll.* 1997]

Algorithmes SQ : ARTUR / GR, LEGEND / GY

- **Régression robuste** [Beaton et Tukey 1974, Byrd et Payne 1979, Huber 1981]

$$\min_{\mathbf{x}} \sum_{n=1}^N \phi(y_n - \mathbf{h}_n^t \mathbf{x}) \rightsquigarrow \text{moindres carrés repondérés : IRLS, RSD}$$

- **Cas non différentiables**

— Problème de Fermat-Weber : $\min_{\mathbf{x}} \sum_{n=1}^N \|\mathbf{y}_n - \mathbf{x}\|_2$ [Weiszfeld 1937]

— Dictionnaires et parcimonie : $\min_{\mathbf{x}} \left(\|\mathbf{z} - \mathbf{HW}\mathbf{x}\|^2 + \mu \|\mathbf{x}\|_p^p \right), \quad p \leq 1$

FOCUSS [Gorodnitsky et Rao 1997, Rao *et coll.* 2003]; voir aussi Fuchs [2007]

IRLS = ARTUR / GR = FOCUSS = Weiszfeld
RSD = LEGEND /GY

■ Applicabilité en restauration d'image

$$\mathbf{B}(\mathbf{x}) = \mathbf{B}_{\text{GY}} = 2\mathbf{H}^t\mathbf{H} + \mu\mathbf{V}^t\mathbf{V} \quad \text{ou} \quad \mathbf{B}(\mathbf{x}) = \mathbf{B}_{\text{GR}}(\mathbf{x}) = 2\mathbf{H}^t\mathbf{H} + \mu\mathbf{V}^t\mathbf{L}(\mathbf{x})\mathbf{V}$$

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

✘ Sauf cas particulier, le coût par itération explose quand N croît !

■ Variante 1 [Charbonnier *et coll.* 1997, Nikolova et Ng 2005]

Calcul approché de $\mathbf{B}^{-1}(\mathbf{x}_k)\nabla\mathcal{J}(\mathbf{x}_k)$ par GC (**préconditionné**) : **SQ+GC**

■ Variante 2 [Fessler et Booth 1999]

Algorithme GC non linéaire (**préconditionné**) + SQ scalaire : **GCNL+SQ1D**

... mais SQ+GC et GCNL+SQ1D ne sont pas des algorithmes AMQ

Variante 1 : SQ+GC**■ Principe**

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{x}_k + \mathbf{d}_k \\ \mathbf{d}_k &= \mathbf{u}_{I_k}(\mathbf{x}_k)\end{aligned}$$

où $\mathbf{u}_I(\mathbf{x})$ est la solution après I itérations de GC(P) appliquées à

$$\mathbf{B}(\mathbf{x}) \mathbf{u} = -\nabla \mathcal{J}(\mathbf{x})$$

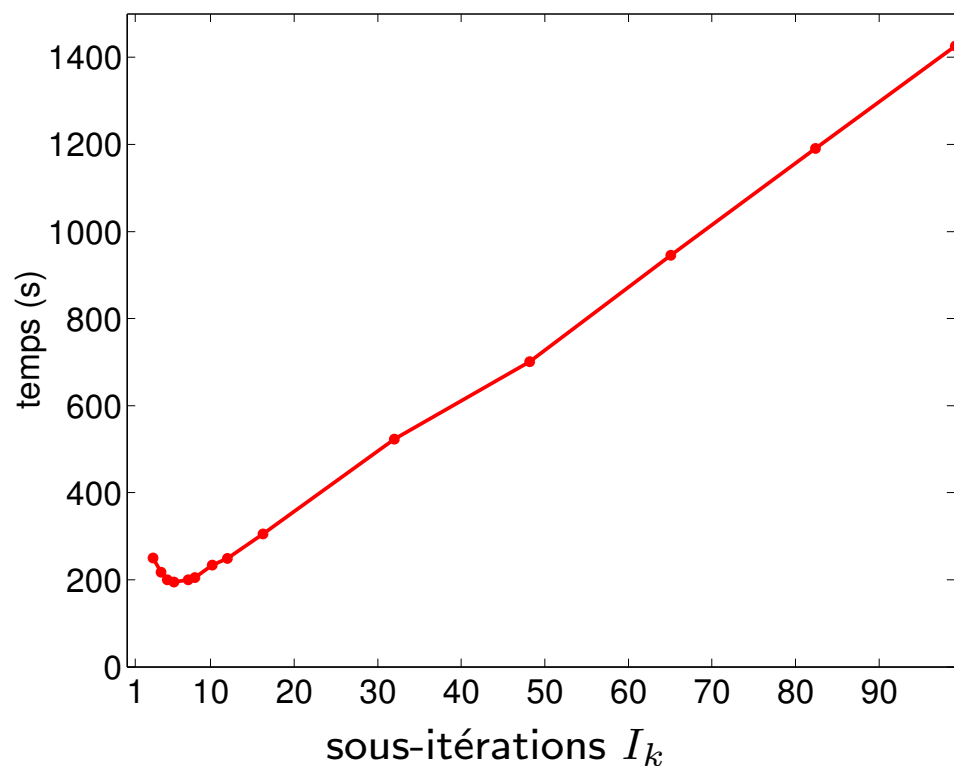
■ Questions ouvertes

- Comment choisir I_k ? Tel que $\frac{\|\text{résidu}_{I_k}\|}{\|\text{résidu}_0\|} \leq \varepsilon$, e.g., $\varepsilon = 10^{-6}$?
- **Convergence ?**

$$\forall I_k \geq 1, \quad \lim_{k \rightarrow \infty} \nabla \mathcal{J}(\mathbf{x}^k) = \mathbf{0} \quad [\text{Labat et Idier 2007}]$$

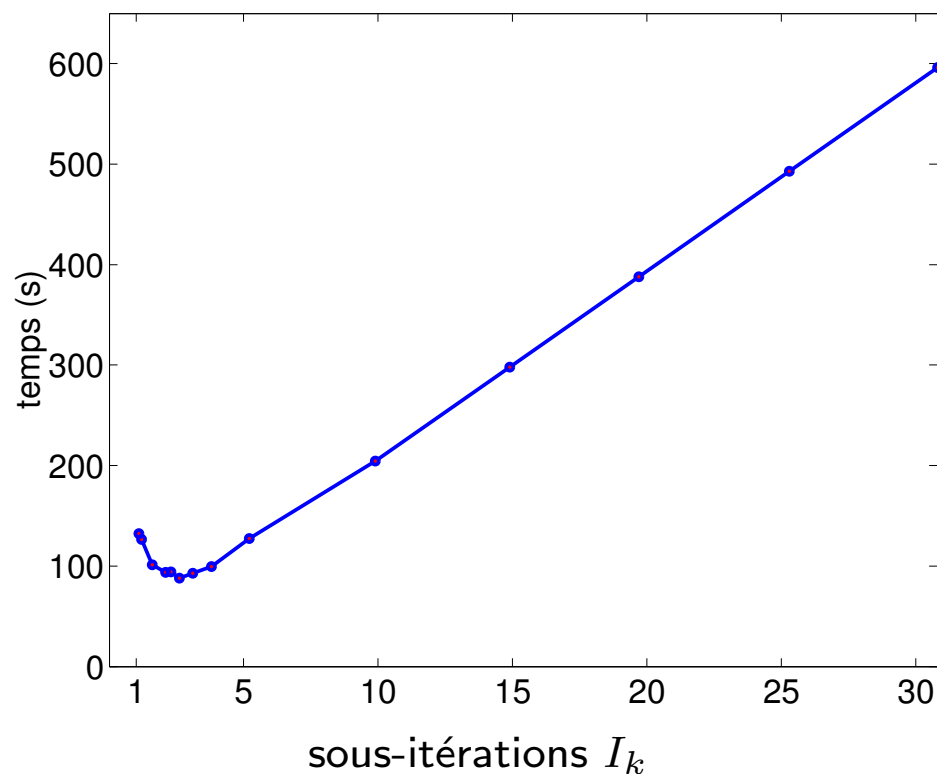
■ **Problème simulé (bateau de pêche)**

GR+GC



$$\varepsilon = 10^{-6} \implies I_k \in [60, 100]$$

GR+GCP (transformée en cosinus)



$$\varepsilon = 10^{-6} \implies I_k \in [20, 30]$$

Variante 2 : GCNL+SQ1D

■ Principe

$$\mathbf{d}_k = -\mathbf{M}^{-1} \nabla \mathcal{J}(\mathbf{x}_k) + \beta_k \mathbf{d}_{k-1}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$$

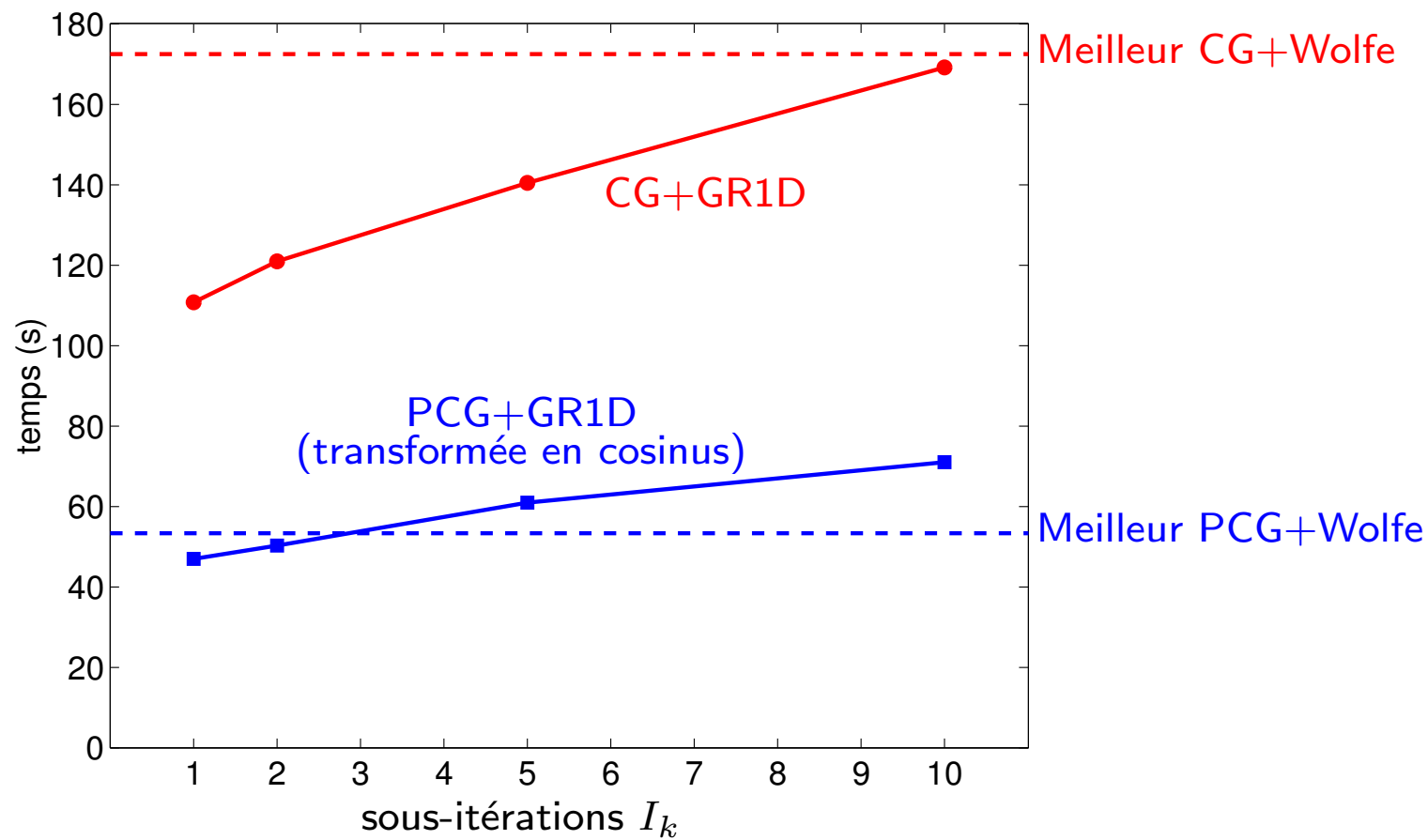
$\mathbf{M} > 0$: matrice de préconditionnement

■ Stratégies de pas

1. Pas constant $\alpha_k = \theta$: pas de garantie de convergence
2. Recherche de pas classique : dichotomie, interpolation, ... + conditions de Wolfe
3. Minimisation de $f(\alpha) = \mathcal{J}(\mathbf{x}_k + \alpha \mathbf{d}_k)$ (I_k sous-itérations) [Fessler et Booth 1999]
 - Comment choisir I_k ? $I_k = 5$?
 - Convergence ?

$$\forall I_k \geq 1, \liminf_{k \rightarrow \infty} \nabla \mathcal{J}(\mathbf{x}^k) = \mathbf{0} \text{ (Polak-Ribière, ...)} \quad [\text{Labat et Idier 2008}]$$

■ Problème simulé (bateau de pêche)



■ Bilan comparatif

	préconditionnement	
	non	oui
GR+GC	194,3 s	87,9 s
CG+GR1D	110,7 s	46,9 s
CG+Wolfe	175,5 s	53,3 s
SQ exact	>2000 s	

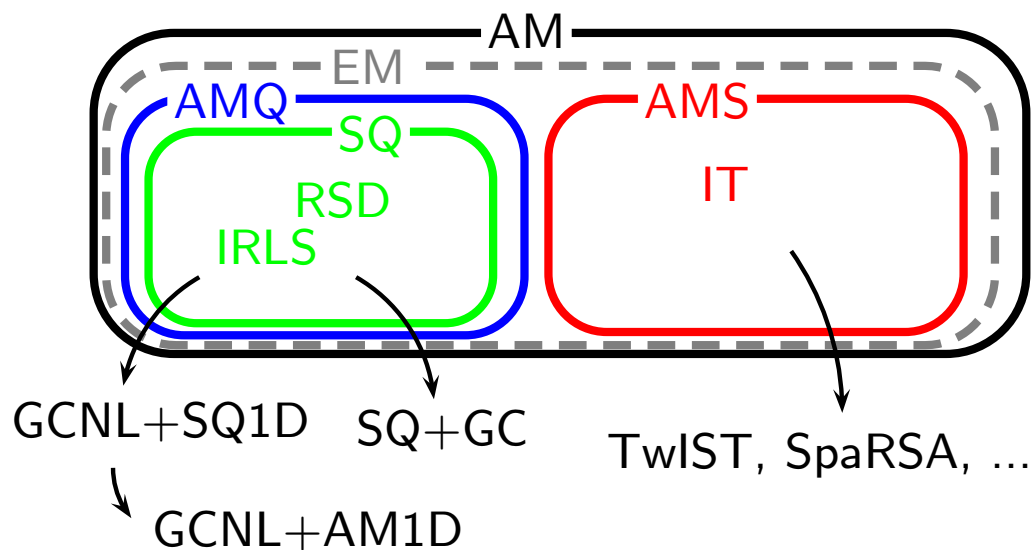
Conclusions

■ Adaptation des algorithmes SQ aux problèmes de grande taille

- Algorithmes de Newton inexact tronqué
- Algorithmes de gradient conjugué sans recherche de pas

Pas d'algorithme emboîté : simplicité + faible coût par itération + convergence

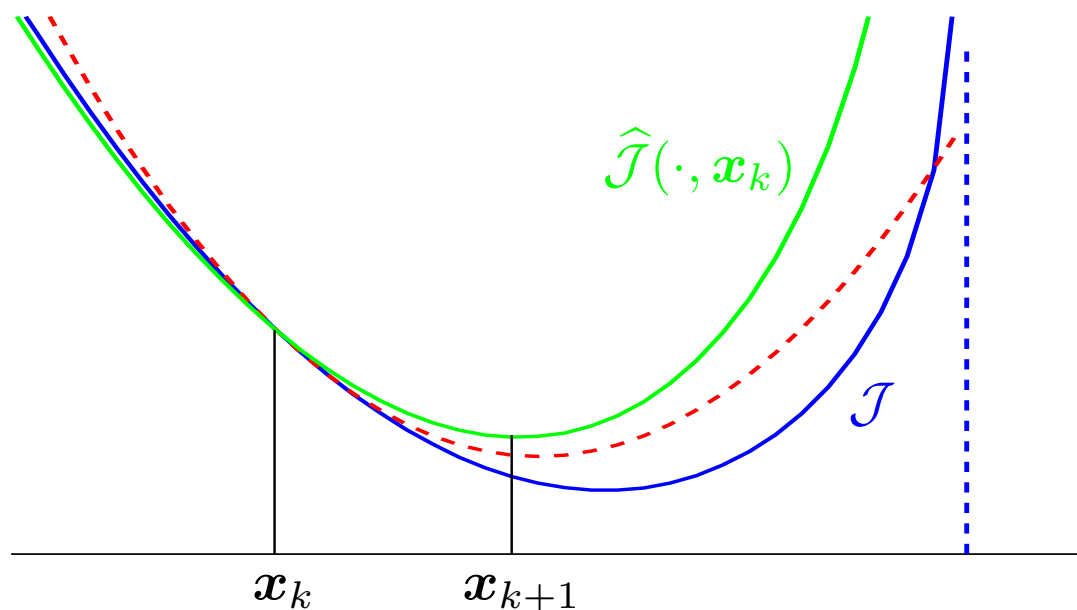
■ Typologie d'algorithmes de minimisation par approximation majorante



- AM : approximation majorante
[Lange et coll. 2000, Vaida 2005]
= minimize-maximize
= iterative majorization
= optimization transfer
- EM : expectation maximization
- AMQ : AM quadratique
- AMS : AM séparable
- IT : iterative thresholding

■ GCNL+AM1D : Critères à barrière logarithmique

- Tomographie par émission de positons
- Pénalisation par maximum d'entropie
- Algorithmes de points intérieurs (contraintes inégalités, non différentiabilité)



↪ GCNL + recherche de pas par AM non quadratique de type

$$h(\alpha, \alpha_0) = p_0 + p_1\alpha + p_2\alpha^2 - p_3 \log(\bar{\alpha} - \alpha)$$

[Chouzenoux *et coll.* 2009a;b]

Plan

- I. Généralités
- II. Déconvolution de train d'impulsions
- III. Restauration d'image et optimisation

IV. Problèmes aveugles, méthodes autodidactes

Obstacle informationnel, obstacle technique

Méthodes de Monte-Carlo par chaîne de Markov (MCMC)

Application : déconvolution aveugle et autodidacte de train d'impulsions

Echantillonnage inspiré par l'optimisation

Bibliographie

Généralités

Introduction

■ Problèmes aveugles

i.e., l'opérateur \mathbf{H} du problème direct contient des inconnues
e.g., la déconvolution aveugle : h est inconnu

■ Problèmes autodidactes

On souhaite un réglage automatique des *hyperparamètres*
e.g., du paramètre de régularisation μ

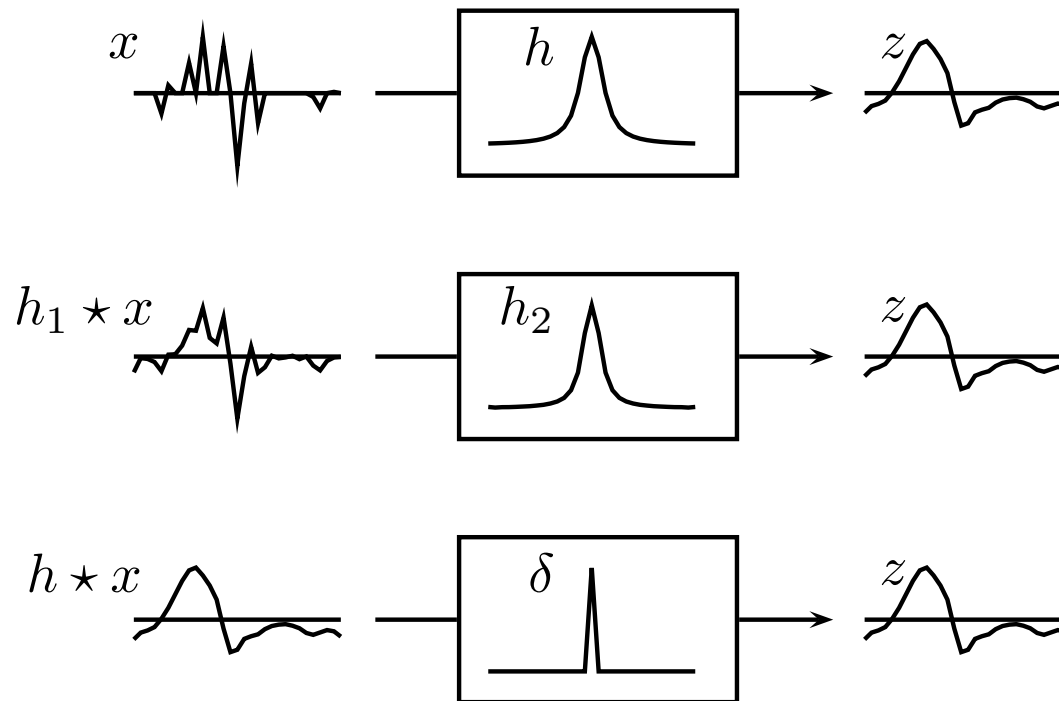
■ Deux obstacles

- Obstacle « informationnel » : Y a-t-il un espoir de solution exploitable ?
- Obstacle « technique » : Comment calculer cette solution ?
simulation bayésienne (méthodes MCMC)

Obstacle informationnel : exemple en déconvolution aveugle

Il existe une infinité de solutions équivalentes !

- Retards, changement d'échelle
- Déphasages
- Contenu fréquentiel



Obstacle technique : quel estimateur calculer ?**■ Diverses approches non bayésiennes**

i.e., exploitent seulement la statistique de $p(\mathbf{z} | \mathbf{x}, \mathbf{h}, \boldsymbol{\theta})$

- Adéquation aux données : trouver μ tel que $\|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}_{\mu}\|^2 = N\sigma^2$
(tendance sous-régularisante)
- courbe en L
- validation croisée
- minimum description length (MDL), Akaike information criterion (AIC), ...
- estimateur de Stein

■ MAP joint

Maximiser $p(\mathbf{x}, \mathbf{h}, \boldsymbol{\theta} \mid \mathbf{z})$ par rapport à $(\mathbf{x}, \mathbf{h}, \boldsymbol{\theta})$

- Extension simple de la minimisation de critère pénalisé, e.g.,

$$\arg \min_{\mathbf{x}, \mathbf{h}} \|\mathbf{z} - \mathbf{h} \star \mathbf{x}\|^2 + \mu_x \Phi_x(\mathbf{x}) + \mu_h \Phi_h(\mathbf{h})$$

- Biais asymptotique, voire dégénéscence
[Little et Rubin 1983, Gassiat *et coll.* 1992]
- Parfois fonctionnel...

Moindres carrés conjoints en diversité de phase [Gonsalves et Chidlaw 1979]

$$\mathbf{z}_k = \mathbf{h}_k \star \mathbf{x} + \mathbf{b}_k, \quad k = 1, \dots, K$$

avec $\mathbf{h}_k = \mathbf{h}(\boldsymbol{\theta}_k)$ et $\boldsymbol{\theta}_k = \boldsymbol{\theta} + \boldsymbol{\delta}_k$, $\boldsymbol{\delta}_k$ connus, $\boldsymbol{\theta}$ inconnu

- **Moindres carrés conjoints**

$$(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = \min_{\mathbf{x}, \boldsymbol{\theta}} \sum_k \|\mathbf{z}_k - \mathbf{h}_k \star \mathbf{x}\|^2 + \mu \|\mathbf{F}\mathbf{x}\|$$

- **Propriété :** Si $K > 1$, $\text{taille}(\mathbf{x}) \rightarrow \infty$, $\mu \rightarrow 0$ et des hypothèses « minimales », alors $\hat{\boldsymbol{\theta}}$ est **asymptotiquement sans biais** !

($\hat{\boldsymbol{\theta}}$ est un estimateur à minimum de contraste [Idier et coll. 2005])

$\mathbf{h} = |\text{TF}(\mathbf{A} \exp(i\boldsymbol{\theta}))|^2$, **fonction non linéaire de $\boldsymbol{\theta}$!**

\mathbf{A} : fonction d'ouverture de la pupille

$\boldsymbol{\theta}$: paramètres de la phase aberrante

$\boldsymbol{\delta}_k$: paramètres de défocalisation

■ Estimateurs marginaux

☞ Construire $\hat{\mathbf{h}}$ et $\hat{\boldsymbol{\theta}}$ à partir de $p(\mathbf{h}, \boldsymbol{\theta} | \mathbf{z}) = \int p(\mathbf{x}, \mathbf{h}, \boldsymbol{\theta} | \mathbf{z}) d\mathbf{x}$

↪ EM, SEM, variational Bayes EM

(rapport *nb de données/nb d'inconnues* favorable),

puis estimer \mathbf{x} sachant \mathbf{z} , $\hat{\mathbf{h}}$ et $\hat{\boldsymbol{\theta}}$

☞ Construire $\hat{\mathbf{x}}$ à partir de $p(\mathbf{x} | \mathbf{z}) = \iint p(\mathbf{x}, \mathbf{h}, \boldsymbol{\theta} | \mathbf{z}) d\mathbf{h} d\boldsymbol{\theta}$

(i.e., intégrer les paramètres de nuisance hors du problème)

↪ Méthodes de Monte-Carlo par chaîne de Markov (MCMC)

Méthodes MCMC

Méthodes de Monte-Carlo par chaîne de Markov (MCMC) [Liu 2001]

Elles permettent le calcul d'estimateurs marginaux tels que l'espérance *a posteriori*

$$\mathbb{E}[\mathbf{x} | \mathbf{z}] = \int \mathbf{x} p(\mathbf{x} | \mathbf{z}) d\mathbf{x} = \int \mathbf{x} p(\mathbf{x}, \mathbf{h}, \boldsymbol{\theta} | \mathbf{z}) d\mathbf{x} d\mathbf{h} d\boldsymbol{\theta}$$

1) Choisir $\mathbf{x}_0, \mathbf{h}_0, \boldsymbol{\theta}_0$ arbitrairement ;

2) pour $k = 1, \dots, K$,

– tirer au hasard $\mathbf{x}^{(k)}$ selon $p(\mathbf{x} | \mathbf{h}^{(k-1)}, \boldsymbol{\theta}^{(k-1)}, \mathbf{z}) \propto p(\mathbf{x}, \mathbf{h}^{(k-1)}, \boldsymbol{\theta}^{(k-1)} | \mathbf{z})$

– tirer au hasard $\mathbf{h}^{(k)}$ selon $p(\mathbf{h} | \mathbf{x}^{(k)}, \boldsymbol{\theta}^{(k-1)}, \mathbf{z}) \propto p(\mathbf{x}^{(k)}, \mathbf{h}, \boldsymbol{\theta}^{(k-1)} | \mathbf{z})$

– tirer au hasard $\boldsymbol{\theta}^{(k)}$ selon $p(\boldsymbol{\theta} | \mathbf{x}^{(k)}, \mathbf{h}^{(k)}, \mathbf{z}) \propto p(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}, \boldsymbol{\theta} | \mathbf{z})$

3) Calculer $\hat{\mathbf{x}} = \frac{1}{K} \sum_{k=1}^K \mathbf{x}^{(k)}$

• La suite aléatoire $(\mathbf{x}_0, \mathbf{h}_0, \boldsymbol{\theta}_0), (\mathbf{x}_1, \mathbf{h}_1, \boldsymbol{\theta}_1), \dots$ est une **chaîne de Markov**

• **Aspect « Monte-Carlo »** : on approche $\mathbb{E}[f(\mathbf{x}, \mathbf{h}, \boldsymbol{\theta} | \mathbf{z})]$ par $\frac{1}{K} \sum_{k=1}^K f(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}, \boldsymbol{\theta}^{(k)})$

Propriétés

■ Convergence en loi

Si $\{\mathbf{X}^{(k)}\}_{k \in \mathbb{N}}$ une chaîne de Markov homogène « ergodique » (irréductible récurrente positive de période 1) dont le noyau de transition φ vérifie la *condition d'équilibre*

$$\varphi(\mathbf{x}' | \mathbf{x}) p(\mathbf{x}) = \varphi(\mathbf{x} | \mathbf{x}') p(\mathbf{x}') \quad (\text{CE})$$

alors la densité de probabilité de $\mathbf{X}^{(\infty)}$ est p .

■ Loi des grands nombres

$$\text{si } E_p[f^2] < \infty, \quad \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K f(\mathbf{x}^{(k)}) = E_p[f] \text{ p.s.}$$

■ Principe de l'augmentation de données

$$\{\mathbf{X}^{(k)}, \Theta^{(k)}\} \sim p(\mathbf{x}, \theta | \mathbf{z}) \implies \begin{cases} \Theta^{(k)} \sim p(\theta | \mathbf{z}) = \int p(\mathbf{x}, \theta | \mathbf{z}) d\mathbf{x} \\ \mathbf{X}^{(k)} \sim p(\mathbf{x} | \mathbf{z}) = \int p(\mathbf{x}, \theta | \mathbf{z}) d\theta. \end{cases}$$

Echantillonneurs de Gibbs et de Metropolis-Hastings

On souhaite générer des échantillons de $p(\mathbf{x})$, $\mathbf{x} = (x_1, \dots, x_N)$ en enchaînant des « mouvements » satisfaisant la condition d'équilibre (CE)

■ Echantillonneur de Gibbs [Geman et Geman 1984]

Pour $n = 1, \dots, N$, tirer x_n au hasard suivant $p(x_n \mid x_1, \dots, x_{n-1}, x_{n+1} \dots)$

■ Echantillonneur de Metropolis-Hastings [Hastings 1970]

① Configuration courante : \mathbf{x}

② Proposer \mathbf{x}' par échantillonnage d'un *noyau de proposition* $q(\mathbf{x}' \mid \mathbf{x})$

③ $P(\mathbf{x}' \text{ remplace } \mathbf{x}) = \min \left\{ 1, \frac{p(\mathbf{x}')}{p(\mathbf{x})} \frac{q(\mathbf{x} \mid \mathbf{x}')}{q(\mathbf{x}' \mid \mathbf{x})} \right\}$;

retour en ① pour l'itération suivante

Cas particuliers classiques :

— $\mathbf{x}' = \mathbf{x} + \text{perturbation}$ (marche aléatoire)

— Echantillonneur de Gibbs

Quantités calculables par MCMC

■ Estimateurs à coût bayésien séparable

Si $C(\mathbf{x}, \mathbf{x}^*) = \sum_n c_n(x_n, x_n^*)$, alors

$$\min_{\hat{\mathbf{x}}} \mathbb{E}[C(\hat{\mathbf{x}}(\mathbf{z}), \mathbf{x}^*)] = \sum_n \min_{\hat{x}_n} \mathbb{E}[c_n(\hat{x}_n(\mathbf{z}), x_n^*)] \approx \sum_n \min_{\hat{x}_n} \frac{1}{K} \sum_{k=1}^K c_n(\hat{x}_n(\mathbf{z}), x_n^{(k)})$$

à minimiser composante par composante analytiquement (EAP)
ou numériquement (MAPM)

estimateur	$C(\mathbf{x}, \mathbf{x}^*)$	séparabilité
MAP	$-\delta(\mathbf{x} - \mathbf{x}^*)$	non
EAP	$\ \mathbf{x} - \mathbf{x}^*\ ^2$ (risque quadratique)	oui
MAP marginal	$-\sum_m \delta(x_m - x_m^*)$	oui
ELMQ	$\ \mathbf{x} - \mathbf{x}^*\ ^2$ sous contrainte de linéarité	oui

■ Covariance a posteriori (barres d'erreur, etc)

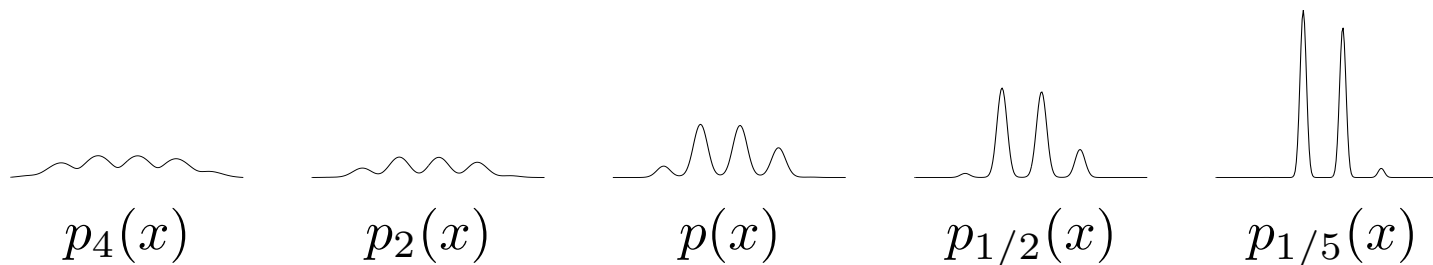
$$\lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K (\mathbf{x}^{(k)} - \hat{\mathbf{x}})(\mathbf{x}^{(k)} - \hat{\mathbf{x}})^t = \text{Cov}(\mathbf{x} | \mathbf{z})$$

■ Calcul d'évidences (choix de modèles)

$$p(\mathbf{z} | \mathcal{M}) = \int p(\mathbf{z}, \mathbf{x}, \mathbf{h}, \boldsymbol{\theta}) d\mathbf{x} d\mathbf{h} d\boldsymbol{\theta} \quad [\text{Chib 1995}]$$

Estimation MAP : principe du recuit simulé [Geman et Geman 1984]

Si la loi instantanée de $\{\mathbf{X}^{(k)}\}$ est $p_{T(k)}(\mathbf{x} | \mathbf{z})$, avec $T(k) \searrow 0$ et $p_T(\mathbf{x} | \mathbf{z}) \propto (p(\mathbf{x} | \mathbf{z}))^{1/T}$, alors : $\lim_{T \rightarrow 0} p_T(\mathbf{x} | \mathbf{z}) = \delta_{\hat{\mathbf{x}}^{\text{MAP}}}(\mathbf{x})$



Application :
déconvolution aveugle et autodidacte de train d'impulsions

Application « directe » d'une méthode MCMC [Cheng et coll. 1996]

● *Modèle a priori*

- $\mathbf{x} = (\mathbf{q}, \mathbf{r})$ modèle Bernoulli-gaussien, paramètres : λ, σ^2
- \mathbf{h} *a priori* gaussien centré « non informatif », paramètre : σ_h^2
- $\boldsymbol{\theta} = (\lambda, \sigma^2, \sigma_h^2, \sigma_b^2)$: *a priori* conjugués non informatifs

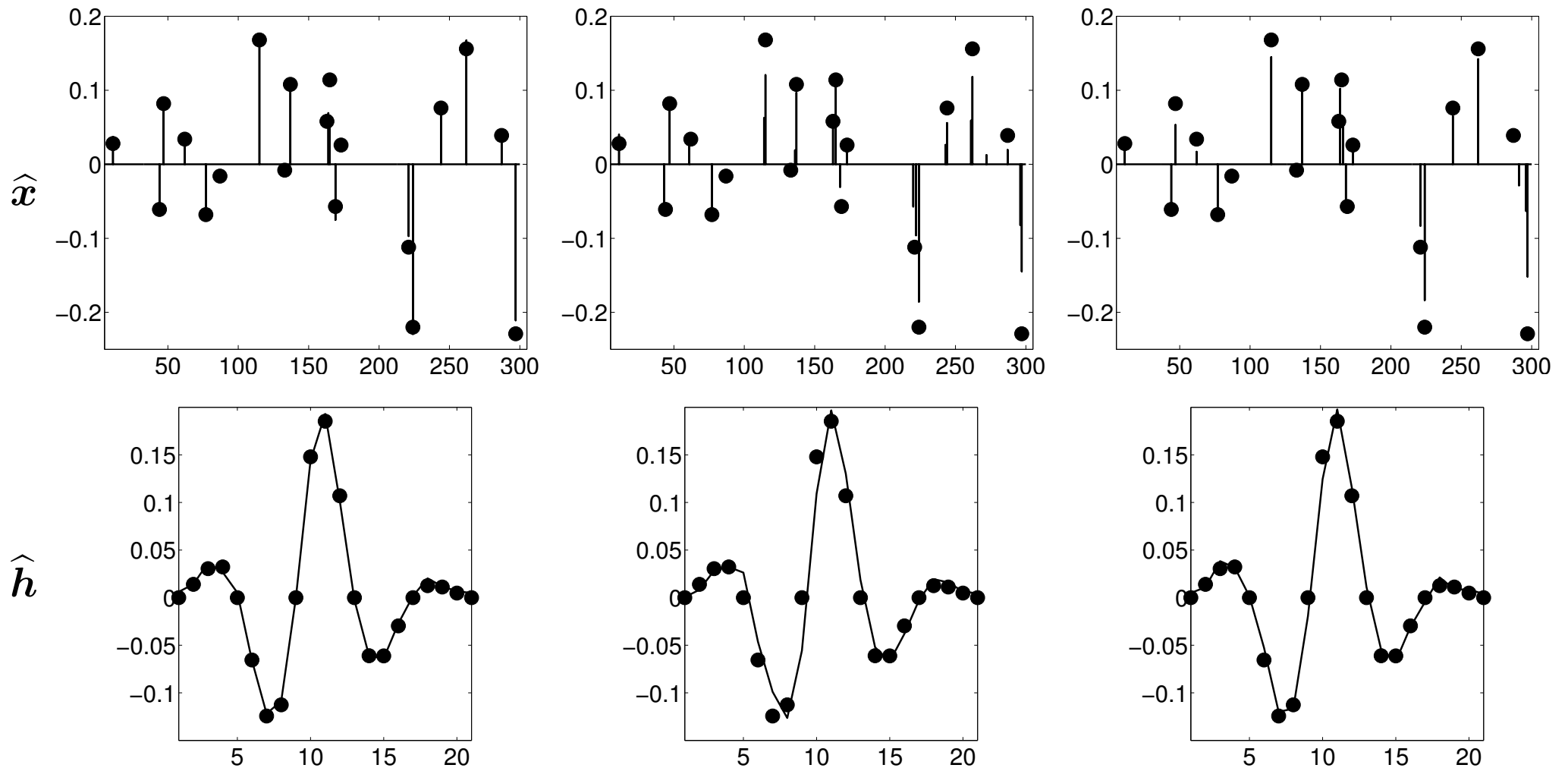
● *Echantillonneur de Gibbs*

$$(x_1, \dots, x_M, \mathbf{h}, \lambda, \sigma^2, \sigma_h^2, \sigma_b^2)$$

● *Estimation*

- $\hat{\mathbf{q}}^{\text{MMAP}}$ (vote majoritaire composante par composante)
- $\hat{r}_m = 0$ si $\hat{q}_m = 0$, $\hat{r}_m = \sum r_m^{(k)} / \sum q_m^{(k)}$ sinon
- \mathbf{h}^{EAP}
- $\boldsymbol{\theta}^{\text{EAP}}$

● Résultats typiques ($K = 1000$)

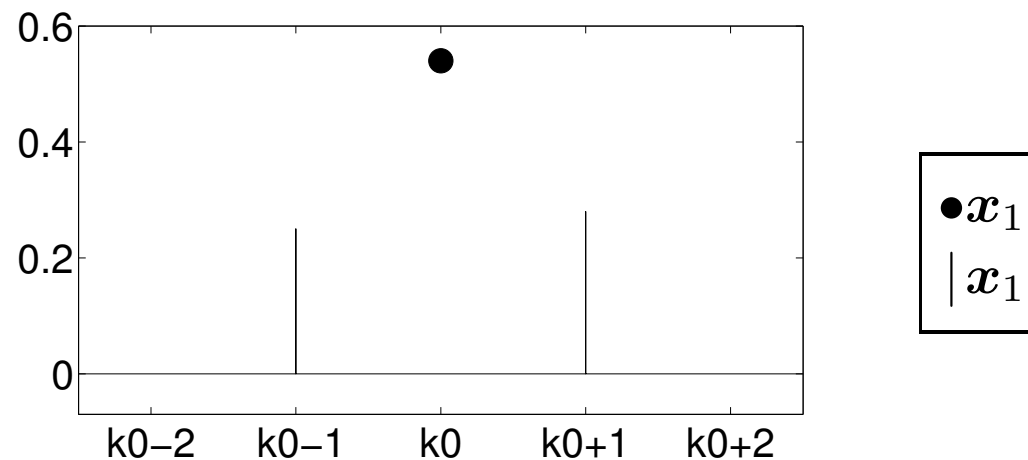


Reproductibilité imparfaite...

Trois types de problèmes

■ **Problème 1 : « Piégage »** [Labat 2006, Ge et coll. 2008]

- Rareté des transitions de type $(h, x) \leftrightarrow (retard \star h, retard^{-1} \star x)$
- Rareté des transitions de type $x_1 \leftrightarrow x_2$



Analogie avec le problème des solutions locales en minimisation

~> Labat [2006], Ge et coll. [2008] : ajout de mouvements spécifiques

■ Problème 2 : Définition des estimateurs

$$— p(\mathbf{x}, \mathbf{r}, \mathbf{h}, \boldsymbol{\theta} | \mathbf{z}) = p(\mathbf{q}, -\mathbf{r}, -\mathbf{h}, \boldsymbol{\theta} | \mathbf{z}) \implies \mathbb{E}[\mathbf{h} | \mathbf{z}] = \mathbf{0} !$$

$$— p(\text{retard} \star \mathbf{x}, \text{retard}^{-1} \star \mathbf{h}, \boldsymbol{\theta} | \mathbf{z}) \approx p(\mathbf{x}, \mathbf{h}, \boldsymbol{\theta} | \mathbf{z})$$

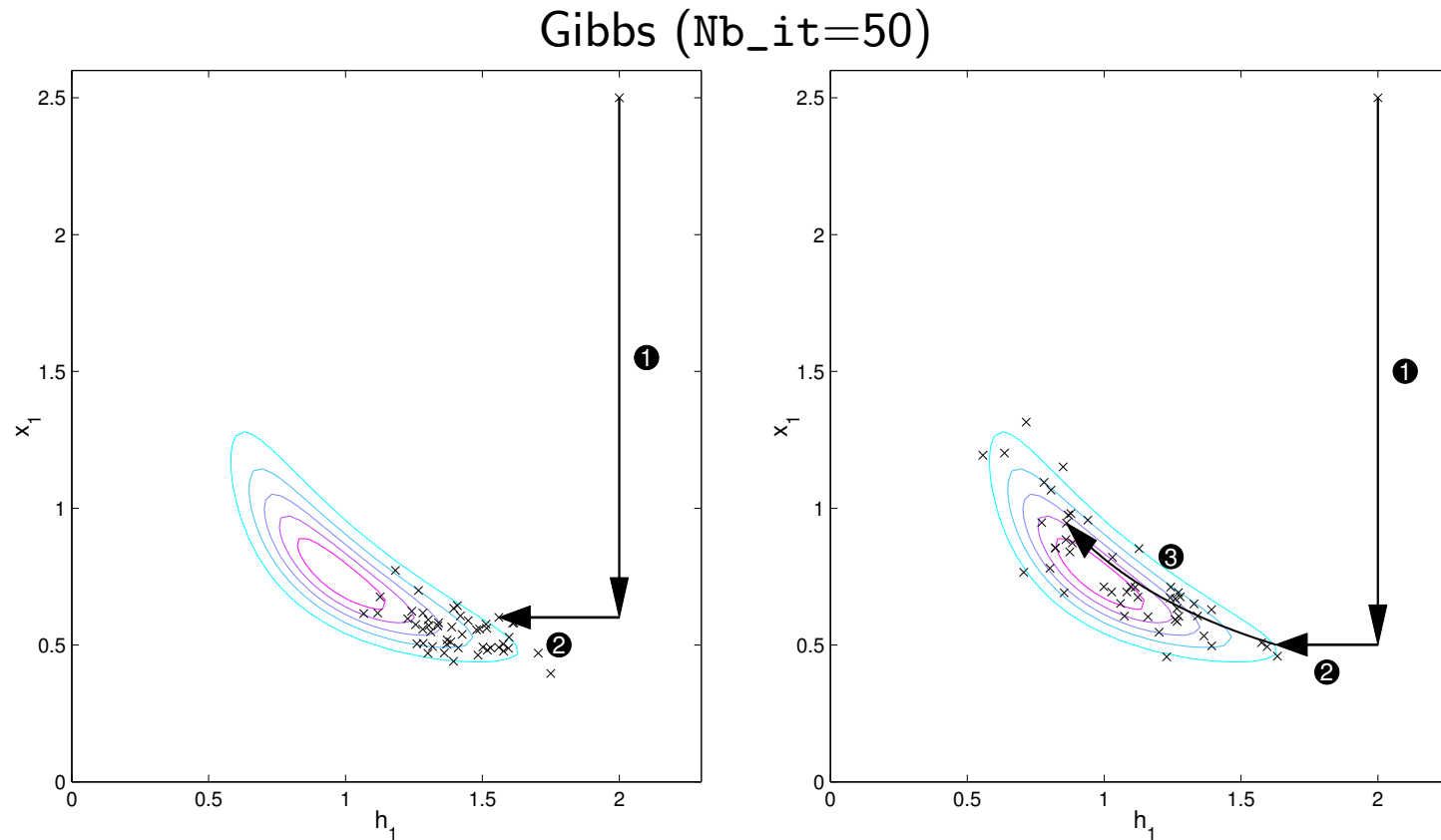
Problèmes voisins de l'« échange d'étiquettes » (*label switching*) dans le cas des mélanges de population [Stephens 2000])

A résoudre par tri des échantillons suivant un critère de similarité [Labat 2006]

■ Problème 3 : Lenteur d'évolution

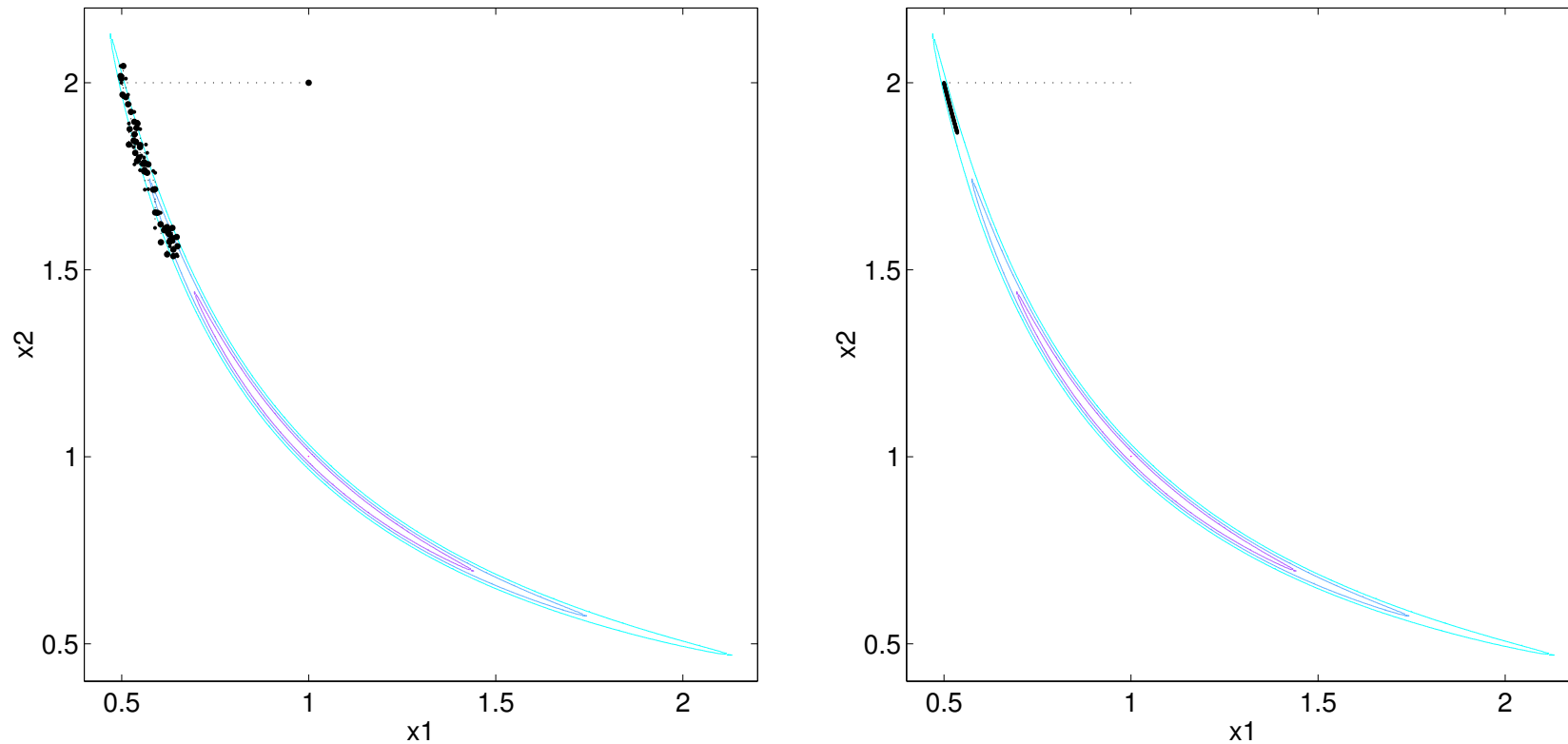
Evolution de l'échelle, *i.e.*, exploration des configurations $(sh, x/s)$ en fonction de s

↪ Veit *et coll.* [2008] : ajout de mouvements spécifiques



Problème 3 : Lenteur d'évolution
↪ Echantillonnage inspiré par l'optimisation

Gibbs / Gauss-Seidel (Nb_it=50)



↪ Echantillonneur exploitant la structure locale de l'énergie (gradient, Hessien) ?

■ Echantillonneur de Langevin-Hastings [Neal 1993, Besag 2001]

- Diffusion de Langevin

$$d\mathbf{x}(t) = -\nabla \mathcal{J}(\mathbf{x}(t)) dt + \sqrt{2} d\mathbf{w}(t)$$

avec $p(\mathbf{x}) \propto \exp(-\mathcal{J}(\mathbf{x}))$, $\mathbf{w}(t)$ processus brownien vectoriel

Propriété : \mathbf{x} converge en loi vers p

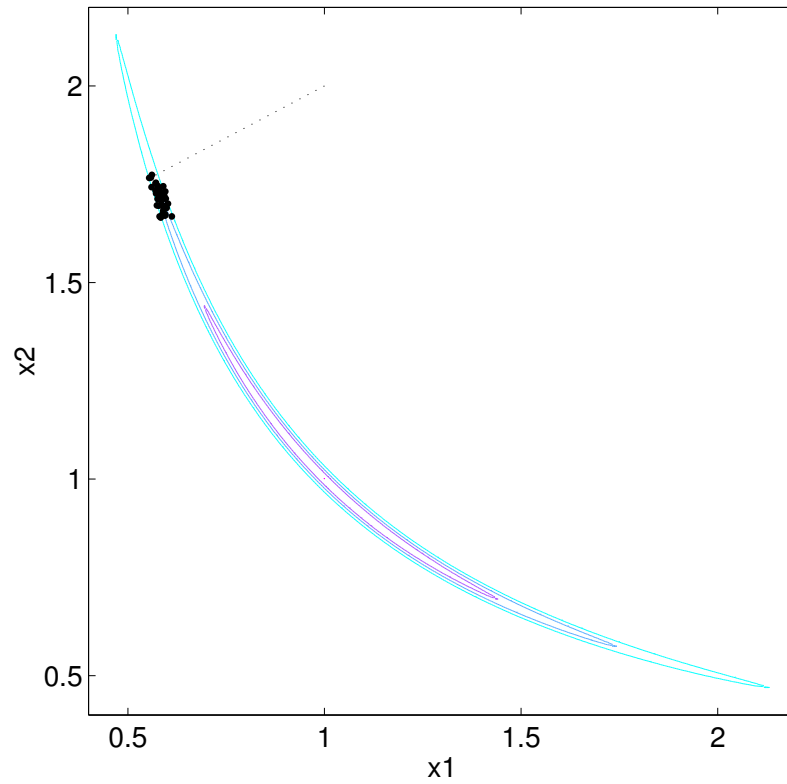
- Echantillonneur de Langevin-Hastings

= Echantillonneur de Metropolis-Hastings avec loi de proposition obtenue par diffusion de Langevin discrétisée :

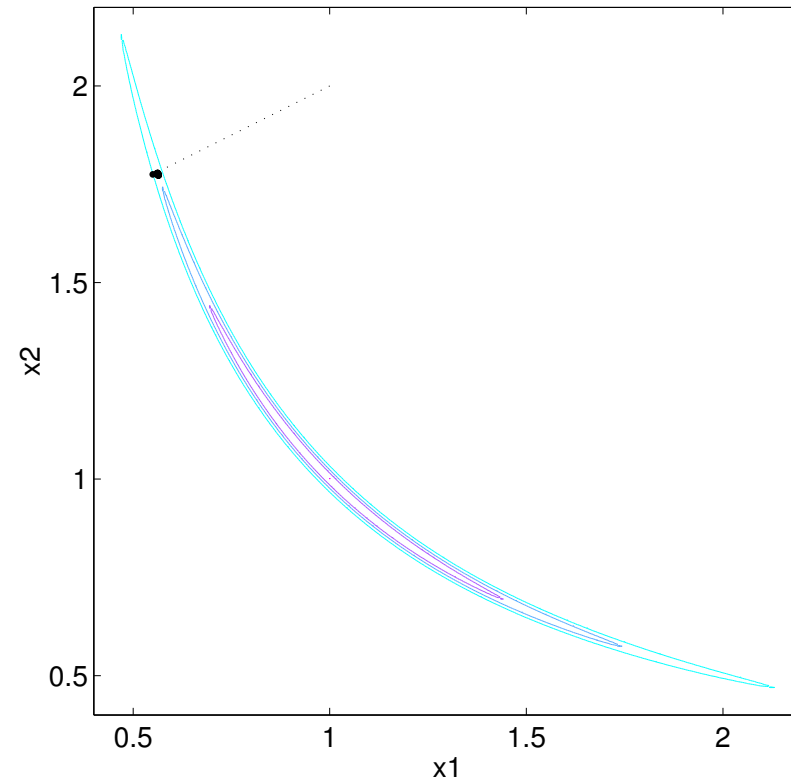
$$\mathbf{x}' = \mathbf{x} - \tau \nabla \mathcal{J}(\mathbf{x}) + \mathbf{z} \sqrt{2\tau}, \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

= Descente de gradient à pas fixe « perturbée »

Langevin-Hastings
(Nb_it=50)



gradient
(pas fixe, Nb_it=50)



● Version préconditionnée ?

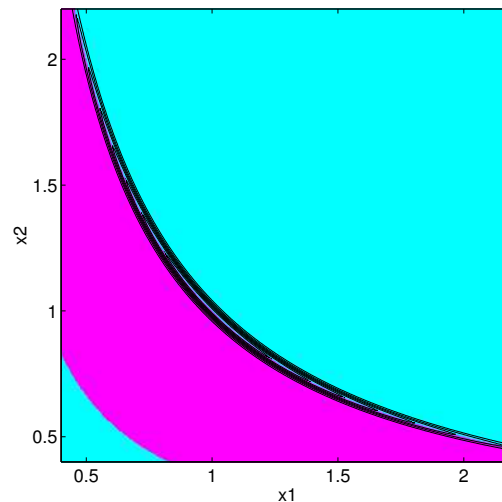
= Version de base après un changement de variable $\mathbf{y} = \mathbf{P}^t \mathbf{x}$
 (typiquement, $\mathbf{P}\mathbf{P}^t \approx (\nabla_{\mathbf{x}}^2 \mathcal{J})^{-1}$)



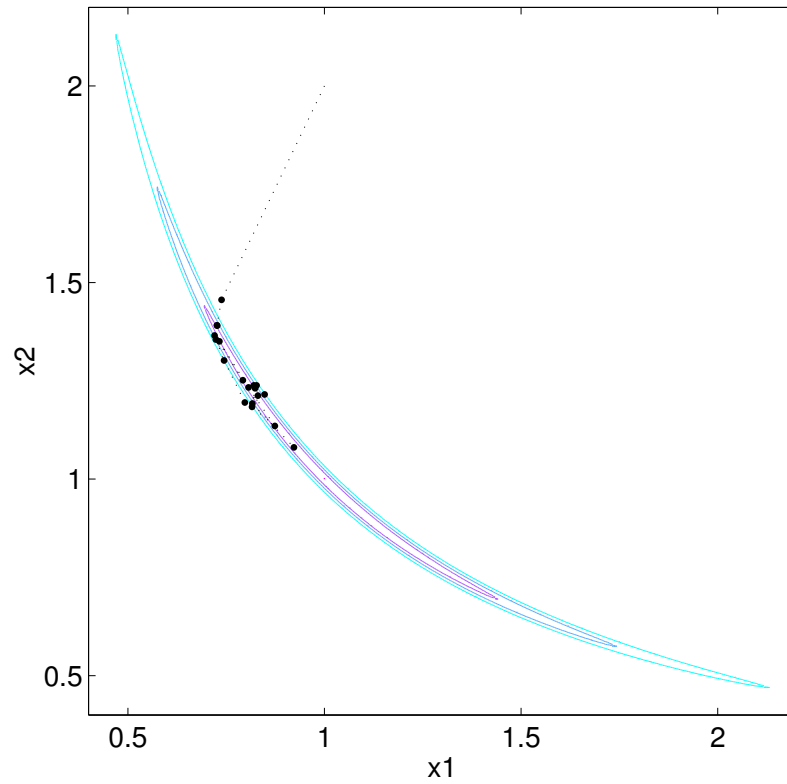
$$\mathbf{x}' = \mathbf{x} - \tau \mathbf{P}^{-1} \mathbf{P}^{-t} \nabla \mathcal{J}(\mathbf{x}) + \mathbf{P}^{-1} \mathbf{z} \sqrt{2\tau}, \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

Par exemple : $\mathbf{P} = \begin{cases} \text{Cholesky}((\nabla_{\mathbf{x}}^2 \mathcal{J})^{-1}) & \text{si } \nabla_{\mathbf{x}}^2 \mathcal{J} > 0 \\ \text{diag}(\text{diag}((\nabla_{\mathbf{x}}^2 \mathcal{J})^{-1})^{1/2}) & \text{sinon} \end{cases}$

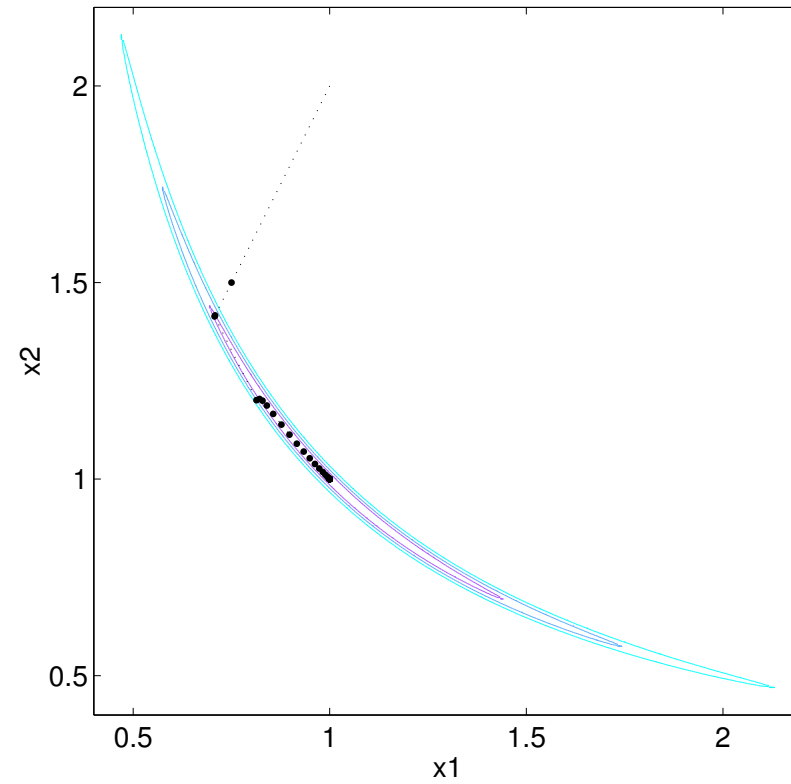
Zone à hessien positif (en magenta)

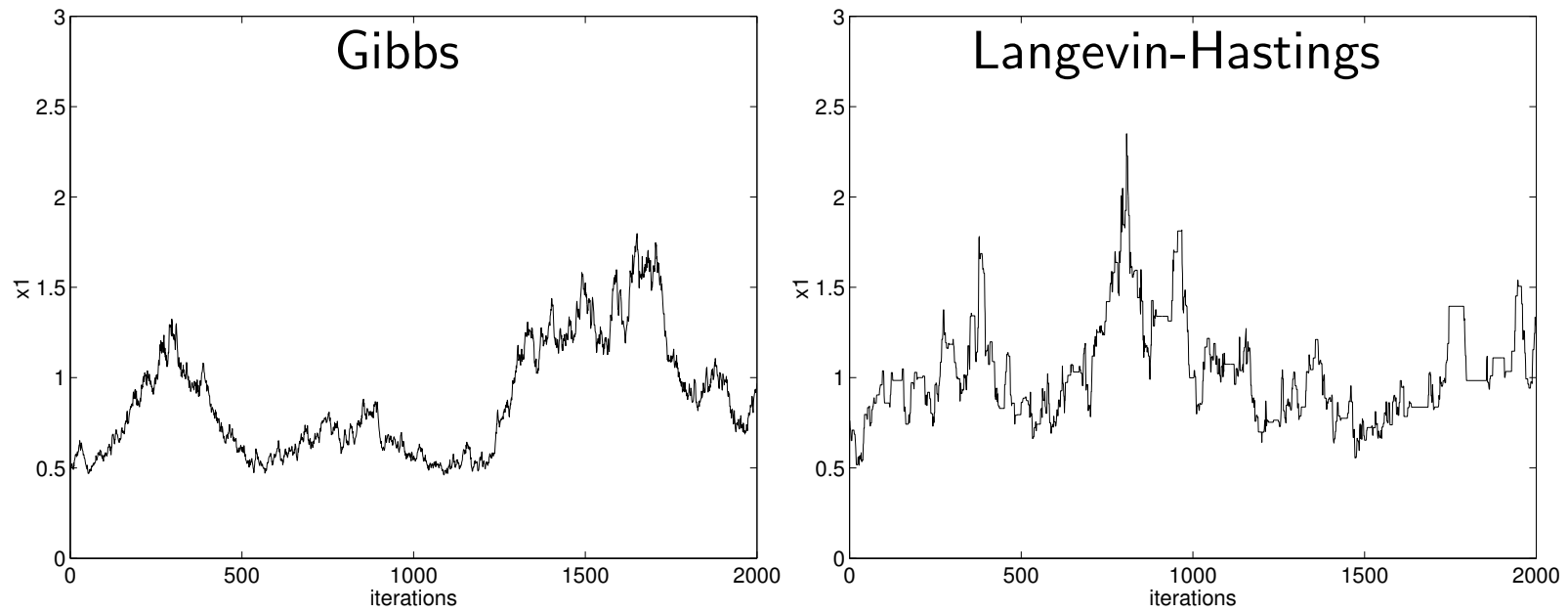


Langevin-Hastings préconditionné
(Nb_it=50)



gradient préconditionné
(pas fixe, Nb_it=50)





Trajectoire de x_1

- Perspective : échantillonnage de vecteurs gaussiens

Alternative classique :

- ☞ Echantillonnage exact par factorisation LU de la covariance
- ☞ Echantillonneur de Gibbs

Perspective :

- ☞ Echantillonneur de Langevin-Hastings préconditionné (par LU tronqué?)

- [Beaton et Tukey 1974] A. E. Beaton et J. W. Tukey. The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16 : 147–185, 1974.
- [Bertsekas 1995] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, USA, 1995.
- [Besag 2001] J. Besag. Markov chain monte carlo for statistical inference. Working Paper 9, Center for Statistics and the Social Sciences, Univ. of Washington, USA, 2001.
- [Blake et Zisserman 1987] A. Blake et A. Zisserman. *Visual reconstruction*. The MIT Press, Cambridge, MA, USA, 1987.
- [Bouman et Sauer 1993] C. A. Bouman et K. D. Sauer. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Transactions on Image Processing*, 2 (3) : 296–310, juillet 1993.
- [Brézis 1983] H. Brézis. *Analyse fonctionnelle : théorie et applications*. Masson, Paris, 1983.
- [Byrd et Payne 1979] R. H. Byrd et D. A. Payne. Convergence of the iteratively reweighted least squares algorithm for robust regression. Rapport Interne 313, The Johns Hopkins Univ., Baltimore, MD, USA, juin 1979.
- [Champagnat *et coll.* 2001] F. Champagnat, Y. Goussard, S. Gautier et J. Idier. Déconvolution impulsionnelle. In J. Idier, éditeur, *Approche bayésienne pour les problèmes inverses*, pages 115–138, Paris, novembre 2001. Traité IC2, Série traitement du signal et de l'image, Hermès.
- [Charbonnier *et coll.* 1997] P. Charbonnier, L. Blanc-Féraud, G. Aubert et M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6 (2) : 298–311, février 1997.
- [Cheng *et coll.* 1996] Q. Cheng, R. Chen et T.-H. Li. Simultaneous wavelet estimation and deconvolution

- of reflection seismic signals. *IEEE Transactions on Geoscience and Remote Sensing*, 34 : 377–384, mars 1996.
- [Chib 1995] S. Chib. Marginal likelihood from the Gibbs output. *Journal of Acoustical Society America*, 90 (432) : 1313–1321, 1995.
- [Chouzenoux et coll. 2009a] E. Chouzenoux, S. Moussaoui et J. Idier. A majorize-minimize line search algorithm for barrier function optimization. In *EUSIPCO*, Glasgow, UK, août 2009.
- [Chouzenoux et coll. 2009b] E. Chouzenoux, S. Moussaoui et J. Idier. A new line search method for barrier functions with strong convergence properties. Rapp. tech., IRCCyN, janvier 2009. <http://hal.archives-ouvertes.fr/docs/00/39/18/18/PDF/Main.pdf>.
- [Demoment 1989] G. Demoment. Image reconstruction and restoration : Overview of common estimation structure and problems. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-37 (12) : 2024–2036, décembre 1989.
- [Elad et coll. 2007] M. Elad, P. Milanfar et R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23 (3) : 947–968, juin 2007.
- [Fessler et Booth 1999] J. A. Fessler et S. D. Booth. Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction. *IEEE Transactions on Image Processing*, 8 (5) : 668–699, mai 1999.
- [Fuchs 2007] J.-J. Fuchs. Convergence of a sparse representations algorithm applicable to real or complex data. *IEEE Journal of Selected Topics in Signal Processing*, 1 (4) : 598–605, décembre 2007. Issue : Convex Optimization Methods for Signal Processing.
- [Gassiat et coll. 1992] E. Gassiat, F. Monfront et Y. Goussard. On simultaneous signal estimation and parameter identification using a generalized likelihood approach. *IEEE Transactions on Information Theory*, 38 : 157–162, janvier 1992.

- [Ge *et coll.* 2008] D. Ge, J. Idier et E. Le Carpentier. A new MCMC algorithm for blind Bernoulli-Gaussian deconvolution. In *EUSIPCO*, Lausanne, Suisse, août 2008.
- [Geman et Reynolds 1992] D. Geman et G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14 (3) : 367–383, mars 1992.
- [Geman et Yang 1995] D. Geman et C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing*, 4 (7) : 932–946, juillet 1995.
- [Geman et Geman 1984] S. Geman et D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6 (6) : 721–741, novembre 1984.
- [Gonsalves et Chidlaw 1979] R. A. Gonsalves et R. Chidlaw. Wavefront sensing by phase retrieval. In A. Tescher, éditeur, *Applications of digital image processing III*, pages 32–39. SPIE 207, 1979.
- [Gorodnitsky et Rao 1997] I. F. Gorodnitsky et B. D. Rao. Sparse signal reconstruction from limited data using focuss : a re-weighted minimum norm algorithm. *IEEE Transactions on Signal Processing*, 45 (3) : 600–616, mars 1997.
- [Hadamard 1902] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton Univ. Bull.*, 13 : 49–52, 1902.
- [Hansen 1992] P. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34 : 561–580, 1992.
- [Harrington 1987] R. F. Harrington. The method of moments in electromagnetics. *Journal of Electromagnetic Waves and Applications*, 1 (3) : 181–200, 1987.
- [Hastings 1970] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57 : 97, janvier 1970.

- [Hebert et Leahy 1989] T. Hebert et R. Leahy. A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Transactions on Medical Imaging*, 8 (2) : 194–202, juin 1989.
- [Huber 1981] P. J. Huber. *Robust Statistics*. John Wiley, New York, NY, USA, 1981.
- [Idier 2001] J. Idier, éditeur. *Approche bayésienne pour les problèmes inverses*. Traité IC2, Série traitement du signal et de l'image, Hermès, Paris, novembre 2001.
- [Idier et coll. 2005] J. Idier, L. Mugnier et A. Blanc. Statistical behavior of joint least square estimation in the phase diversity context. *IEEE Transactions on Image Processing*, 14 (12) : 2107–2116, décembre 2005.
- [Jeng et Woods 1991] F. C. Jeng et J. W. Woods. Compound Gauss-Markov random fields for image estimation. *IEEE Transactions on Signal Processing*, 39 (3) : 683–697, mars 1991.
- [Kattnig et Primot 1997] A. P. Kattnig et J. Primot. Model of the second-order statistic of the radiance field of natural scenes, adapted to system conceiving. In *Aerosense'97 - Visual information processing VI*, volume 3074, Washington DC, USA, 1997. SPIE.
- [Labat 2006] C. Labat. *Algorithmes d'optimisation de critères pénalisés pour la restauration d'images. Application à la déconvolution de trains d'impulsions en imagerie ultrasonore*. Thèse de doctorat, École Centrale de Nantes, décembre 2006.
- [Labat et Idier 2007] C. Labat et J. Idier. Convergence of truncated half-quadratic and Newton algorithms, with application to image restoration. Rapp. tech., IRCCyN, juin 2007.
- [Labat et Idier 2008] C. Labat et J. Idier. Convergence of conjugate gradient methods with a closed-form stepsize formula. *Journal of Optimization Theory and Applications*, 136 (1) : 43–60, janvier 2008.
- [Labat et coll. 2005] C. Labat, J. Idier, B. Richard et L. Chatellier. Ultrasonic nondestructive testing

- based on 2D deconvolution. In *PSIP'2005 : Physics in signal and Image processing*, Toulouse, janvier 2005.
- [Lagendijk *et coll.* 1988] R. Lagendijk, J. Biemond et D. E. Boekee. Regularized iterative image restoration with ringing reduction. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-36 : 1874–1888, 1988.
- [Lange *et coll.* 2000] K. Lange, D. R. Hunter et I. Yang. Optimization transfer using surrogate objective functions (with discussion). *Journal of Computational and Graphical Statistics*, 9 (1) : 1–20, mars 2000.
- [Little et Rubin 1983] R. J. A. Little et D. B. Rubin. On jointly estimating parameters and missing data by maximizing the complete-data likelihood. *The American Statistician*, 37 : 218–220, août 1983.
- [Liu 2001] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer Series in Statistics. Springer Verlag, New York, NY, USA, 2001.
- [Mendel 1983] J. M. Mendel. *Optimal Seismic Deconvolution*. Academic Press, New York, NY, USA, 1983.
- [Nashed 1981] M. Z. Nashed. Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory. *IEEE Transactions on Antennas and Propagation*, 29 : 220–231, 1981.
- [Neal 1993] R. M. Neal. Probabilistic inference using Markov chain Monte Carlo methods. Rapport Interne CRG-TR-93-1, Department of Computer Science, University of Toronto, <http://www.cs.toronto.edu/~radford>, septembre 1993.
- [Nikolova 2004] M. Nikolova. Weakly constrained minimization. Application to the estimation of images and signals involving constant regions. *Journal of Mathematical Imaging and Vision*, 21 (2) : 155–175, 2004.

- [Nikolova et Ng 2005] M. Nikolova et M. Ng. Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM Journal on Scientific Computing*, 27 : 937–966, 2005.
- [Nocedal et Wright 1999] J. Nocedal et S. J. Wright. *Numerical optimization*. Springer Texts in Operations Research. Springer-Verlag, New York, NY, USA, 1999.
- [Ortega et Rheinboldt 1970] J. Ortega et W. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, NY, USA, 1970.
- [Rao et coll. 2003] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer et K. Kreutz-Delgado. Subset selection in noise based on diversity measure minimization. *IEEE Transactions on Signal Processing*, 51 (3) : 760–770, mars 2003.
- [Stephens 2000] M. Stephens. Dealing with label-switching in mixture models. *Journal of the Royal Statistical Society B*, 62 : 795–809, 2000.
- [Tarantola et Valette 1982] A. Tarantola et B. Valette. Inverse problems = quest for information. *Journal of Geophysics*, 50 : 159–170, 1982.
- [Tikhonov et Arsénine 1976] A. Tikhonov et V. Arsénine. *Méthodes de résolution de problèmes mal posés*. Éditions MIR, Moscou, Russie, 1976.
- [Trillon et coll. 2008] A. Trillon, J. Idier et P. Peureux. Unsupervised Bayesian 3D reconstruction for non-destructive evaluation using gammagraphy. In *EUSIPCO*, Lausanne, Suisse, août 2008.
- [Vaida 2005] F. Vaida. Parameter convergence for EM and MM algorithms. *Statistica Sinica*, 15 : 831–840, 2005.
- [Veit et coll. 2008] T. Veit, J. Idier et S. Moussaoui. Rééchantillonnage de l'échelle dans les algorithmes MCMC pour les problèmes inverses bilinéaires. *Traitement du Signal*, 25 (4) : 329–343, 2008.
- [Voss et Eckhardt 1980] H. Voss et U. Eckhardt. Linear Convergence of Generalized Weiszfeld's Method.

Computing, 25 : 243–251, 1980.

[Weiszfeld 1937] E. Weiszfeld. Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tôhoku Mathematical Journal*, 43 : 355–386, 1937.

Lemme d'inversion de matrice

Soit \mathbf{A} ($N \times N$) ; \mathbf{B} ($N \times M$) ; \mathbf{C} ($M \times M$) ; \mathbf{D} ($M \times N$).

- Sous réserve d'inversibilité,

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$$

- Corrolaires

$$\mathbf{D}(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{C}^{-1}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$$

$$\mathbf{CD}(\mathbf{A} + \mathbf{BCD})^{-1} = (\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$$



Matrices de convolution 2D

■ Convolution 2D et ordre lexicographique

Soit $\mathbf{Y} = \mathbf{h} \circledast \mathbf{X}$ pour la convolution 2D : $Y_{\ell,c} = \sum_p \sum_q h_{p,q} X_{\ell-p,c-q}$

Remarque 1 : plusieurs hypothèses de bord possibles

Remarque 2 : convention « matricielle » (origine en haut à gauche)

■ Ordre lexicographique

Soit $\text{lex}(\mathbf{X}) = \begin{bmatrix} \text{ligne 1 en colonne} \\ \text{ligne 2 en colonne} \\ \vdots \\ \text{ligne } L \text{ en colonne} \end{bmatrix} \in \mathbb{R}^{LC \times 1}$ si $\mathbf{X} \in \mathbb{R}^{L \times C}$

Alors $\mathbf{Y} = \mathbf{h} \circledast \mathbf{X}$ s'écrit aussi $\mathbf{y} = \mathbf{H}\mathbf{x}$ avec $\mathbf{y} = \text{lex}(\mathbf{Y})$ et $\mathbf{x} = \text{lex}(\mathbf{X})$

- Taille de \mathbf{H} : $N \times M$ avec $M = LC$ et N de l'ordre de LC !
- Structure ?

Matrices Toeplitz-bloc-Toeplitz

$$\mathbf{H} = \begin{bmatrix} \ddots & \ddots & \ddots & & & & \\ \dots & \mathbf{H}_{-1} & \mathbf{H}_0 & \mathbf{H}_1 & \dots & & \\ & \dots & \mathbf{H}_{-1} & \mathbf{H}_0 & \mathbf{H}_1 & \dots & \\ & & & \ddots & \ddots & \ddots & \\ & & & & & & \end{bmatrix} \quad \text{avec } \mathbf{H}_p = \begin{bmatrix} \ddots & \ddots & \ddots & & & & \\ \dots & \mathbf{h}_{p,-1} & \mathbf{h}_{p,0} & \mathbf{h}_{p,1} & \dots & & \\ & \dots & \mathbf{h}_{p,-1} & \mathbf{h}_{p,0} & \mathbf{h}_{p,1} & \dots & \\ & & & \ddots & \ddots & \ddots & \\ & & & & & & \end{bmatrix}$$

- \mathbf{H}_p : **matrice de Toeplitz** des coefficients des contributions de la ligne $\ell + p$ de l'image \mathbf{X} à la ligne ℓ de l'image $\mathbf{h} \circledast \mathbf{X}$
- \mathbf{H} : **matrice Toeplitz-bloc-Toeplitz** (Toeplitz par bloc, à blocs Toeplitz)

Cas particuliers :

- convolution circulaire : **\mathbf{H} matrice circulante-bloc-circulante**, diagonalisable dans la base de Fourier 2D
- \mathbf{h} symétrique et convolution miroir + circulaire : **\mathbf{H} diagonalisable par transformée en cosinus 2D**



Algorithmes préconditionnés

- Correspond à l'algorithme de base après un changement de variable inversible $\mathbf{y} = \mathbf{P}^t \mathbf{x}$
- Exemple : gradient conjugué, forme de Polak-Ribière préconditionnée

$$\mathbf{p}_k = -(\mathbf{M}_k)^{-1} \nabla \mathcal{J}(\mathbf{x}^k) \quad (\text{préconditionnement}) \quad (1)$$

$$\beta_k = \begin{cases} 0 & \text{si } k = 0 \\ \frac{(\nabla \mathcal{J}(\mathbf{x}^k) - \nabla \mathcal{J}(\mathbf{x}^{k-1}))^t \mathbf{p}_k}{(\nabla \mathcal{J}(\mathbf{x}^{k-1}))^t \mathbf{p}_{k-1}} & \text{si } k > 0 \end{cases} \quad (2)$$

$$\mathbf{d}_k = \mathbf{p}_k + \beta_k \mathbf{d}_{k-1} \quad (3)$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k \quad (\text{mise à jour}) \quad (4)$$

$$(\mathbf{M}_k = \mathbf{P}_k^{-1} \mathbf{P}_k^{-t})$$

