

Méthodes probabilistes pour les problèmes inverses

Ali MOHAMMAD-DJAFARI

Laboratoire des Signaux et Systèmes

Supélec, Plateau de Moulon

91192 Gif-sur-Yvette Cedex, FRANCE.

e-mail : djafari@lss.supelec.fr

Plan de l'exposé :

1. Problèmes inverses ?
2. Exemples de problèmes inverses
3. Classification des méthodes d'inversion
 - Analytiques, Algébriques déterministes et Probabilistes
4. Méthodes probabilistes
 - Maximum de vraisemblance (MV)
 - Maximum d'entropie classique (ME)
 - Maximum d'entropie sur la moyenne (MEM)
 - Approche bayésienne
5. Approche bayésienne et ses difficultés
6. Invariance d'échelle, GNC, Estimation des hyperparamètres
7. Exemples d'application
8. Problèmes ouverts

1 Problèmes inverses

$$\mathcal{A}(\mathbf{y}, \mathbf{x}, \mathbf{z}, \mathbf{b}) = 0$$

modèle mesures grandeurs inconnues grandeurs inconnues intermédiaires erreurs et bruit

Relation explicite : $\mathbf{y} = \mathcal{A}(\mathbf{x}, \mathbf{z}, \mathbf{b})$

Erreur en sortie : $\mathbf{y} = \mathcal{A}(\mathbf{x}, \mathbf{z}) \diamond \mathbf{b}$

Erreur additive : $\mathbf{y} = \mathcal{A}(\mathbf{x}, \mathbf{z}) + \mathbf{b}$

Relation entre \mathbf{x} et \mathbf{z} :

$$\begin{cases} \mathbf{y} = \mathcal{A}_1(\mathbf{x}, \mathbf{z}) + \mathbf{b} \\ \mathcal{A}_2(\mathbf{x}, \mathbf{z}) = 0 \end{cases}$$

Modèle non linéaire simple : $\mathbf{y} = \mathcal{A}(\mathbf{x}) + \mathbf{b}$

Modèle linéaire : $\mathbf{y} = \mathcal{A}\mathbf{x} \diamond \mathbf{b}$

Modèle linéaire + bruit additif : $\mathbf{y} = \mathcal{A}\mathbf{x} + \mathbf{b}$

3

2 Exemples de problèmes inverses

2.1 Suivre de front de fusion liquide-solide

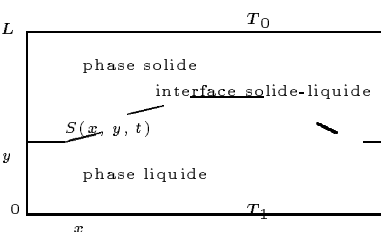
Solide : $\frac{\partial T_s}{\partial t} = \alpha_s \left(\frac{\partial^2 T_s}{\partial x^2} + \frac{\partial^2 T_s}{\partial y^2} \right)$

Liquide : $\frac{\partial T_l}{\partial t} = \alpha_l \left(\frac{\partial^2 T_l}{\partial x^2} + \frac{\partial^2 T_l}{\partial y^2} \right)$

Bilan d'énergie à l'interface : $k_s \frac{\partial T_s}{\partial n} - k_l \frac{\partial T_l}{\partial n} = \rho L_f \vec{v} \cdot \vec{n}$

\vec{v} : vitesse de l'interface solide-liquide

\vec{n} : vecteur normal à l'interface



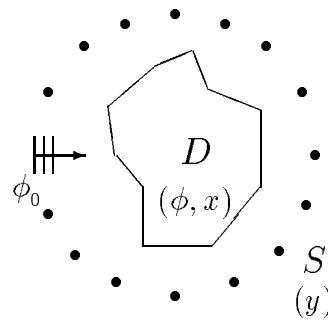
Mesures : flux de chaleur sur la paroi chaude $\frac{\partial T_s(x, 0, t)}{\partial t}$

Inconnue : position de l'interface $S(x, y, t)$

Grandeurs intermédiaire : champs de température $T_s(x, y, t)$ et $T_l(x, y, t)$

4

2.2 Imagerie à ondes diffractées



Mesures : champ diffracté par objet $\phi_d(\mathbf{r}_i)$

Grandeur inconnue : $x(\mathbf{r}) = k_0^2(n^2(\mathbf{r}) - 1)$

Grandeurs intermédiaires : $\phi(\mathbf{r})$

$$\phi_d(\mathbf{r}_i) = \iint_D G_m(\mathbf{r}_i, \mathbf{r}') \phi(\mathbf{r}') x(\mathbf{r}') d\mathbf{r}', \quad \mathbf{r}_i \in S$$

$$\phi(\mathbf{r}) = \phi_0(\mathbf{r}) + \iint_D G_o(\mathbf{r}, \mathbf{r}') \phi(\mathbf{r}') x(\mathbf{r}') d\mathbf{r}', \quad \mathbf{r} \in D$$

2.3 Problèmes inverses linéaires

$$y(\mathbf{s}_i) = \int x(\mathbf{r}) h(\mathbf{r}, \mathbf{s}_i) d\mathbf{r} + b(\mathbf{s}_i), \quad i = 1, \dots, M$$

Débruitage : $y(\mathbf{r}_i) = x(\mathbf{r}_i) + b(\mathbf{r}_i)$

Interpolation : $y(\mathbf{r}_i) = x(\mathbf{r}) + b(\mathbf{r}_i)$

Déconvolution : $y(\mathbf{r}_i) = \int x(\mathbf{r}') h(\mathbf{r}_i - \mathbf{r}') d\mathbf{r}' + b(\mathbf{r}_i)$

Restauration d'image :

$$y(\zeta_i, \eta_j) = \iint x(\zeta', \eta') h(\zeta_i - \zeta', \eta_j - \eta') d\zeta' d\eta' + b(\zeta_i, \eta_j)$$

Reconstruction d'image :

$$y(r_i, \phi_j) = \iint x(\zeta, \eta) \delta(r_i - \zeta \cos \phi - \eta \sin \phi) d\zeta d\eta + b(r_i, \phi_j)$$

Synthèse de Fourier :

$$y(\Omega_i, \phi_j) = \iint x(\zeta, \eta) \exp [j(\zeta \Omega_i \cos \phi_j + \eta \Omega_i \sin \phi_j)] d\zeta d\eta + b(\Omega_i, \phi_j)$$

3 Classification des méthodes d'inversion

3.1 Méthodes analytiques

On se place dans le cadre de mesures continues et en utilisant la notion d'opérateur adjoint on cherche une solution analytique qui est ensuite approximée lors du calcul numérique

$$y(\mathbf{r}_i) = \int h(\mathbf{r}_i, \mathbf{r}') x(\mathbf{r}') d\mathbf{r}' + b(\mathbf{r}_i)$$

$$x(\mathbf{r}) = \int g(\mathbf{r}', \mathbf{r}) y(\mathbf{r}') d\mathbf{r}' \simeq \sum_i w_i y(\mathbf{r}_i) g(\mathbf{r}', \mathbf{r}_i)$$

- Inversion directe
- Inversion dans l'espace dual
- Inversion par décomposition tronquée sur une base

Exemple : Reconstruction d'image et la Transformée de Radon

$$p(r, \phi) = \iint_D f(x, y) \delta(r - x \cos \phi - y \sin \phi) dx dy$$

Inversion directe :

$$f(x, y) = \left(-\frac{1}{2\pi^2} \right) \int_0^\pi \int_{-\infty}^{+\infty} \frac{\frac{\partial}{\partial r} p(r, \phi)}{(r - x \cos \phi - y \sin \phi)} dr d\phi$$

Inversion dans le domaine de Fourier :

$$F(\omega_x, \omega_y) = P(\Omega, \phi), \quad \omega_x = \Omega \cos \phi, \omega_y = \Omega \sin \phi$$

Limitations :

- Méthodes limitées aux cas de modèles simples
- Bruit non pris en compte
- Nécessite un ensemble complet de données non bruitées

3.2 Méthodes algébriques déterministes

On discrétise le problème $\rightarrow \mathbf{y} = \mathbf{A}(\mathbf{x}) + \mathbf{b}$

et on définit la solution comme le minimum d'un critère

$$\mathbf{x} = \arg \min_{\mathbf{x}} \{J(\mathbf{x})\}$$

- Moindres carrés (MC) : $J(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}(\mathbf{x})\|^2$
- MC de norme minimal (MCNM) : $J(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}(\mathbf{x})\|^2 + \lambda \|\mathbf{x}\|^2$
- Régularisation classique : $J(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}(\mathbf{x})\|^2 + \lambda \|\mathbf{D}\mathbf{x}\|^2$
- Régularisation plus générale :

ou encore $J(\mathbf{x}) = \mathcal{Q}(\mathbf{y} - \mathbf{A}(\mathbf{x})) + \lambda \Phi(\mathbf{D}\mathbf{x})$

$$J(\mathbf{x}) = \Delta_1(\mathbf{y}, \mathbf{A}(\mathbf{x})) + \lambda \Delta_2(\mathbf{x}, \mathbf{x}_\infty)$$

Limitations :

- Bruit implicitement considéré blanc et gaussien
- Prise en compte d'information *a priori* sur la solution limitée
- Manque d'outil pour la détermination des hyperparamètres

3.3 Méthodes probabilistes

On se place dans le cadre de l'estimation :

- Maximum de vraisemblance (MV)
- Maximum d'entropie (ME)
- Maximum d'entropie sur la moyenne (MEM)
- Estimation bayésienne (EB)

Avantages :

- Prise en compte de la nature du bruit
- Prise en compte d'information *a priori* sur la solution
- Cadre cohérent permettant aussi la détermination des hyperparamètres

Limitations : Mise en œuvre plus difficile en pratique

4 Méthodes probabilistes

4.1 Maximum de vraisemblance (MV)

$$\left\{ \begin{array}{l} \text{Modèle d'observation : } \mathbf{y} = \mathbf{A}(\mathbf{x}) + \mathbf{b} \\ \text{Caractéristiques du bruit : } p_b(\mathbf{b}) \end{array} \right. \longrightarrow p(\mathbf{y}|\mathbf{x})$$

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \{p(\mathbf{y}|\mathbf{x})\} = \arg \min_{\mathbf{x}} \{-\ln p(\mathbf{y}|\mathbf{x})\}$$

- Cas linéaire gaussien : \longrightarrow Moindres Carrés
- Méthode rarement satisfaisante pour les problèmes inverses
- Maximum de vraisemblance pénalisée

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \{-\ln p(\mathbf{y}|\mathbf{x}) + \Phi(\mathbf{x})\}$$

4.2 Maximum d'entropie classique (ME)

$$\text{maximiser } H(\mathbf{x}) = -\sum_j \left[x_j \ln \frac{x_j}{m_j} + (x_j - m_j) \right] \quad \text{s.c. } \mathbf{y} - \mathbf{A}\mathbf{x} = 0$$

$$\text{Lagrangien : } \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = H(\mathbf{x}) - \sum_i \lambda_i (y_i - [\mathbf{A}\mathbf{x}]_i)$$

Solution :

$$\frac{\partial \mathcal{L}}{\partial x_j} = 0 \longrightarrow x_j = m_j \exp [-\lambda_0 - [\mathbf{A}^t \boldsymbol{\lambda}]_j]$$

Les variables duales (paramètres de Lagrange) sont solution de :

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = 0 \longrightarrow y_i - \sum_j a_{ij} m_j \exp [-\lambda_0 - [\mathbf{A}^t \boldsymbol{\lambda}]_j] = 0$$

Prise en compte du bruit :

$$\max H(\mathbf{x}) \text{ s.c. } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \leq c \longrightarrow \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \lambda H(\mathbf{x})\}$$

Difficultée : choix de λ

4.3 Maximum d'entropie sur la moyenne (MEM)

$\mathbf{x} \in \mathcal{C}$, $\mu(\mathbf{x})$: mesure de référence

Modèle de mesure : $\mathbf{y} = \mathbf{A}(\langle \mathbf{x} \rangle)$ où $\langle \mathbf{x} \rangle = \int_{\mathcal{C}} \mathbf{x} p(\mathbf{x}) d\mu(\mathbf{x})$

maximiser $-\int_{\mathcal{C}} p(\mathbf{x}) \ln p(\mathbf{x}) d\mu(\mathbf{x})$ s.c. $\mathbf{y} = \mathbf{A}(\langle \mathbf{x} \rangle)$

ou d'une manière équivalente

minimiser $\int_{\mathcal{C}} p(\mathbf{x}) \ln \frac{p(\mathbf{x})}{p_0(\mathbf{x})} d\mathbf{x}$ s.c. $\mathbf{y} = \mathbf{A}(\langle \mathbf{x} \rangle)$

avec $d\mu(\mathbf{x}) = \frac{1}{p_0(\mathbf{x})} d\mathbf{x}$.

Solution :

$$p(\mathbf{x}) = \frac{1}{Z(\boldsymbol{\lambda})} \mu(\mathbf{x}) \exp [\boldsymbol{\lambda}^t [\mathbf{A}\mathbf{x}]] = \mu(\mathbf{x}) \exp [\boldsymbol{\lambda}^t [\mathbf{A}\mathbf{x}] - \ln Z(\boldsymbol{\lambda})]$$

$$Z(\boldsymbol{\lambda}) = \int_{\mathcal{C}} \mu(\mathbf{x}) \exp [\boldsymbol{\lambda}^t [\mathbf{A}\mathbf{x}]] d\mathbf{x}$$

$$\frac{\partial \ln Z(\boldsymbol{\lambda})}{\partial \lambda_i} = y_i, \quad i = 1, \dots, M \quad \longrightarrow \boldsymbol{\lambda} \quad \longrightarrow \hat{\mathbf{x}}(\boldsymbol{\lambda}) = \int \mathbf{x} dP(\mathbf{x}, \boldsymbol{\lambda})$$

Notant :

$$\mathbf{s} = \mathbf{A}^t \boldsymbol{\lambda}, \quad G^*(\mathbf{s}) = \ln Z(\mathbf{s}) = \ln \int_{\mathcal{C}} \exp [\mathbf{s}^t \mathbf{x}] d\mu(\mathbf{x}),$$

$$H(\mathbf{x}) = \mathbf{s}^t \mathbf{x} - G^*(\mathbf{s}), \quad D(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^t \mathbf{y} - G^*(\mathbf{A}^t \boldsymbol{\lambda})$$

on a

$$\hat{\boldsymbol{\lambda}} = \arg \min_{\boldsymbol{\lambda}} \{D(\boldsymbol{\lambda})\}, \quad \text{Critère dual}$$

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{C}} \{H(\mathbf{x})\} \text{ s.c. } \mathbf{y} = \mathbf{A}\mathbf{x}, \quad \text{Critère primal}$$

$$\hat{\mathbf{x}}(\mathbf{s}) = \frac{dG^*(\mathbf{s})}{d\mathbf{s}}, \quad \text{relation explicite}$$

- G et H dépendent de la mesure de référence $d\mu(\mathbf{x})$.
- Si on note $\mathbf{m} = \int \mathbf{x} d\mu(\mathbf{x})$ alors on montre que $H(\mathbf{x}) = H(\mathbf{x}, \mathbf{m})$ est une mesure de distance entre \mathbf{x} et \mathbf{m} :
 - $H(\mathbf{x}, \mathbf{m}) \geq 0$, et $H(\mathbf{x}, \mathbf{m}) = 0$ ssi $\mathbf{x} = \mathbf{m}$
 - $H(\mathbf{x}, \mathbf{m})$ est dérivable et strictement convexe sur \mathcal{C}
 - $H(\mathbf{x}, \mathbf{m}) = \infty$ si $\mathbf{x} \notin \mathcal{C}$.

Si $d\mu(\mathbf{x}) = \prod_{j=1}^N d\mu_j(x_j)$ alors $dP(\mathbf{x}, \boldsymbol{\lambda}) = \prod_{j=1}^N dP_j(x_j, \lambda_j)$

et

$$G(\mathbf{s}) = \sum_j g_j(s_j), \quad H(\mathbf{x}, \mathbf{m}) = \sum_j h_j(x_j, m_j), \quad \hat{x}_j = g'_j(s_j)$$

Remplaçant $\mathbf{s} = \mathbf{A}^t \boldsymbol{\lambda}$ on a

$$G(\boldsymbol{\lambda}) = \sum_j g_j([\mathbf{A}^t \boldsymbol{\lambda}]_j), \quad H(\mathbf{x}, \mathbf{m}) = \sum_j h_j(x_j, m_j), \quad \hat{x}_j = g'_j([\mathbf{A}^t \hat{\boldsymbol{\lambda}}]_j)$$

- $h_j(x, m)$ et $g_j(x)$ dépendent des mesures de référence $\mu_j(x)$.
- $g_j(x)$ est le convexe conjugué de $h_j(x, m)$
- $g_j(x)$ est la transformée de Laplace de $\mu_j(x)$.

Exemples :

Cas	$\mu_j(x)$	$g_j(s)$	$h_j(x, m)$
Gaussienne	$\exp[-\frac{1}{2}(x-m)^2]$	$\frac{1}{2}(s-m)^2$	$\frac{1}{2}(x-m)^2$
Poissonnienne	$\frac{m^x}{x!} \exp[-m]$	$\exp[m-s]$	$-\frac{x}{m} \ln \frac{x}{m} + m - x$
Gamma	$x^{\alpha-1} \exp[-x/m]$	$\ln(s-m)$	$-\ln \frac{x}{m} + \frac{x}{m} - 1$

- Lorsque $\mu(\mathbf{x})$ n'est pas séparable il devient très difficile de mener les calcul (excepté le cas gaussien).
- Cas gaussien : $\mu(\mathbf{x}) = \mathcal{N}(\mathbf{m}, \mathbf{R}_x)$ où on a

$$H(\mathbf{x}, \mathbf{m}) = -\frac{1}{2}(\mathbf{x} - \mathbf{m})^t \mathbf{R}_x^{-1} (\mathbf{x} - \mathbf{m}) \quad \text{et} \quad G(\boldsymbol{\lambda}) = -\frac{1}{2} \boldsymbol{\lambda}^t \boldsymbol{\lambda}$$

Autre cas ?

Prise en compte du bruit :

(Thèses de G. Le Besnerais (93) et J.F. Bercher(95))

Deux approches possibles :

1- Remplacer $\mathbf{y} = \mathbf{Ax}$ par

$$\begin{cases} |y_i - [\mathbf{Ax}]_i| < \epsilon \\ \|\mathbf{y} - \mathbf{Ax}\|^2 < \sigma^2 \end{cases}$$

$$\begin{cases} \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \{H(\mathbf{x})\} \text{ s.c. } |y_i - [\mathbf{Ax}]_i| < \epsilon, \text{ avec } H(\mathbf{x}) = \sum_j h_j(x_j) \\ \hat{\boldsymbol{\lambda}} = \arg \min_{\boldsymbol{\lambda}} \{\tilde{D}(\boldsymbol{\lambda})\} \text{ avec } \tilde{D}(\boldsymbol{\lambda}) = D(\boldsymbol{\lambda}) + \alpha \|\boldsymbol{\lambda}\|^2 \end{cases}$$

 α dépend de σ^2 .2- Remplacer $\mathbf{y} = \mathbf{Ax}$ par $\mathbf{y} = \mathbf{Ax} + \mathbf{b}$ et la réécrire

$$\mathbf{y} = [\mathbf{A}|\mathbf{I}] \begin{bmatrix} \mathbf{x} \\ \mathbf{b} \end{bmatrix} \longrightarrow \mathbf{y} = \tilde{\mathbf{A}} \tilde{\mathbf{x}}$$

Si $\mu(\tilde{\mathbf{x}}) = \mu_x(\mathbf{x})\mu_b(\mathbf{b})$ alors

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{C}} \{Q(\mathbf{y} - \mathbf{Ax}) + \alpha H(\mathbf{x})\}$$

$$\text{avec } H(\mathbf{x}) = \sum_{j=1}^N h_j(x_j), \quad \text{et } Q(\mathbf{z}) = \sum_{i=1}^M q_i(z_i),$$

 $h_j(x_j)$ et $q_i(z_i)$ dépendent des mesures de référence $\mu_x(\mathbf{x})$ et $\mu_b(\mathbf{b})$.Pendant la détermination du α reste empirique.

5 Approche bayésienne de l'estimation

- Modèle d'observation et caractéristiques du bruit : $\longrightarrow p(\mathbf{y}|\mathbf{x};\boldsymbol{\beta})$
- Information *a priori* sur \mathbf{x} : $\longrightarrow p(\mathbf{x}|\boldsymbol{\theta})$
- Bayes : $\longrightarrow p(\mathbf{x}|\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta}) = p(\mathbf{y}|\mathbf{x};\boldsymbol{\beta}) p(\mathbf{x}|\boldsymbol{\theta})/p(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})$
- Règle d'estimation : fonction coût $c(\mathbf{x},\hat{\mathbf{x}})$

$$\hat{\mathbf{x}}(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta}) = \arg \min_{\mathbf{z}} \left\{ \int c(\mathbf{x},\mathbf{z}) p(\mathbf{x}|\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta}) d\mathbf{x} \right\}$$

Exemple :

Estimation au sens du *Maximum A Posteriori* (MAP) :

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \{-\ln p(\mathbf{x}|\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})\} = \arg \min_{\mathbf{x}} \{-\ln p(\mathbf{y}|\mathbf{x};\boldsymbol{\beta}) - \ln p(\mathbf{x}|\boldsymbol{\theta})\}$$

Estimateurs ponctuels :

- Maximum *a posteriori* (MAP) :

$$C(\mathbf{x},\hat{\mathbf{x}}) = 1 - \delta(\mathbf{x} - \hat{\mathbf{x}}) \longrightarrow \hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \{p(\mathbf{x}|\mathbf{y};\boldsymbol{\theta},\boldsymbol{\beta})\}$$

- Moyenne *a posteriori* (MP) :

$$C(\mathbf{x},\hat{\mathbf{x}}) = [\mathbf{x} - \hat{\mathbf{x}}]^t \mathbf{Q} [\mathbf{x} - \hat{\mathbf{x}}]^t \longrightarrow \hat{\mathbf{x}} = E_{x|y} \{\mathbf{x}\} = \int \mathbf{x} p(\mathbf{x}|\mathbf{y};\boldsymbol{\theta},\boldsymbol{\beta}) d\mathbf{x}$$

- MAP Marginale (MAPM) :

$$C(\mathbf{x},\hat{\mathbf{x}}) = \prod_i 1 - \delta(x_i - \hat{x}_i) \longrightarrow \hat{\mathbf{x}} = \arg \max_{x_i} \{p(x_i|\mathbf{y};\boldsymbol{\theta})\},$$

où

$$p(x_i|\mathbf{y};\boldsymbol{\theta}) = \int p_{x|y}(\mathbf{x}|\mathbf{y};\boldsymbol{\theta}) dx_1 \cdots dx_{i-1} \cdots dx_{i+1} \cdots dx_n$$

Cas des problèmes inverses linéaires

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

- Hypothèse sur le bruit : $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, 1/\beta\mathbf{I}) \rightarrow \mathbf{y}|\mathbf{x} \sim \mathcal{N}(\mathbf{A}\mathbf{x}, 1/\beta\mathbf{I})$

$$p(\mathbf{y}|\mathbf{x}, \beta) \propto \exp \left[-\frac{1}{2}\beta \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \right]$$

- Hypothèse gaussienne sur \mathbf{x} : $\mathbf{x} \sim \mathcal{N}(\mathbf{x}_0, \theta\mathbf{P}_0)$:

$$p(\mathbf{x}|\theta) \propto \exp \left[-\frac{1}{2}\theta [\mathbf{x} - \mathbf{x}_0]^t \mathbf{P}_0^{-1} [\mathbf{x} - \mathbf{x}_0] \right]$$

- $\mathbf{x}|\mathbf{y} \sim \mathcal{N}(\hat{\mathbf{x}}, \mathbf{P})$ avec

$$\hat{\mathbf{x}} = \mathbf{P}\mathbf{A}^t(\mathbf{y} - \mathbf{A}\mathbf{x}_0), \quad \mathbf{P} = (\mathbf{A}^t\mathbf{A} + \lambda\mathbf{P}_0^{-1})^{-1}$$

- La solution au sens du MAP :

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \{p(\mathbf{x}|\mathbf{y})\} = \arg \min_{\mathbf{x}} \{J(\mathbf{x})\}, \quad \text{avec } J(\mathbf{x}) = Q(\mathbf{x}) + \lambda\Phi(\mathbf{x})$$

$$Q(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2, \quad \Phi(\mathbf{x}) = \mathbf{x}^t \mathbf{P}_0^{-1} \mathbf{x} = \|\mathbf{D}\mathbf{x}\|, \quad \lambda = \frac{\theta}{\beta}$$

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \lambda\Phi(\mathbf{x})$$

- Hypothèse Gamma sur \mathbf{x} :

$$p(x_j) \propto (x_j/m_j)^\alpha \exp[-x_j/m_j] \rightarrow \Phi(\mathbf{x}) = \alpha \sum_j \ln \frac{x_j}{m_j} + \frac{x_j}{m_j},$$

- Hypothèse Béta sur \mathbf{x} :

$$p(x_j) \propto x_j^\alpha (1-x_j)^\beta \rightarrow \Phi(\mathbf{x}) = \alpha \sum_j \ln x_j + \beta \sum_j \ln(1-x_j),$$

- Hypothèse gaussienne généralisée sur \mathbf{x} :

$$p(x_j) \propto \exp[-\alpha|x_j|^p], \quad 1 < p < 2 \rightarrow \Phi(\mathbf{x}) = \alpha \sum_j |x_j|^p,$$

- Modèles markoviens pour \mathbf{x} :

$$p(x_j|\mathbf{x}) \propto \exp \left[-\alpha \sum_{i \in N_j} \phi(x_j, x_i) \right] \rightarrow \Phi(\mathbf{x}) = \alpha \sum_j \sum_{i \in N_j} \phi(x_j, x_i),$$

Cadre général : Estimé MAP :

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{arg\,min}} \{J(\mathbf{x})\}, \text{ avec } J(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \Phi_\alpha(\mathbf{x})$$

- Lois gaussiennes : $\Phi_\alpha(\mathbf{x})$ quadratique

→ $J(\mathbf{x})$ quadratique → Estimateur linéaire

- Lois entropiques : $\Phi_\alpha(\mathbf{x})$ convexe et séparable : $\Phi(\mathbf{x}) = \sum_j \phi_\alpha(x_j)$

avec $\phi_\alpha(x_j) = \{\alpha x_j^p, \alpha(x_j \ln x_j - x_j), \alpha(\ln x_j - x_j)\}$

→ $J(\mathbf{x})$ convexe → Estimateur nonlinéaire mais facile à calculer

- Modèles markoviens : $\Phi_\alpha(\mathbf{x}) = \sum_j \sum_{i \in N_j} \phi_\alpha(x_j - x_i)$ avec $\phi_\alpha(t) =$

$$\left\{ \begin{array}{ll} |t|^2 & \text{si } |t| < \alpha, \\ \alpha^2 & \text{sinon,} \end{array} \right\}, \left\{ \begin{array}{ll} t^2 & \text{si } |t| < \alpha, \\ 2\alpha t - \alpha^2 & \text{sinon,} \end{array} \right\}, \frac{\alpha^2 t^2}{1+t^2}, \log \cosh(t/\alpha)$$

$\Phi_\alpha(\mathbf{x})$ non convexe → Estimateur nonlinéaire et difficile à calculer

6 Problèmes ouverts

- Comment choisir $p(\mathbf{x}|\boldsymbol{\beta})$

Dans le cas des modèles markoviens comment choisir les fonctions potentiels

→ Invariance d'échelle (Djafari93, Brette *et al.* 93, 94)

- Comment déterminer les hyperparamètres $(\boldsymbol{\theta}, \boldsymbol{\beta})$

→ MAP généralisée, MAPM, EM, SEM, ...

(Djafari 95, Idier 94)

- Comment choisir un estimateur : MAP, MP ou MAPM

→ Plutôt considération calculatoires

- Comment calculer efficacement la solution $\hat{\mathbf{x}}(\mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\beta})$

→ Algorithmes d'optimisation :

– Recuit simulé,

– Relaxation déterministe (GNC) (Thèse Nikolova 95)

6.1 Invariance d'échelle

Estimation linéaire :

$$\begin{cases} \mathbf{y}_1 \mapsto \hat{\mathbf{x}}_1 \\ \mathbf{y}_2 \mapsto \hat{\mathbf{x}}_2 \end{cases} \implies \alpha_1 \mathbf{y}_1 + \alpha_2 \mathbf{y}_2 \mapsto \alpha_1 \hat{\mathbf{x}}_1 + \alpha_2 \hat{\mathbf{x}}_2,$$

Estimation invariant par changement d'échelle :

$$\mathbf{y} \mapsto \hat{\mathbf{x}} \implies k\mathbf{y} \mapsto k\hat{\mathbf{x}}$$

ou encore

$$\text{échelle 1: } \mathbf{y} \mapsto \hat{\mathbf{x}}(\mathbf{y}, \boldsymbol{\theta}_1)$$

$$\text{échelle } k: k\mathbf{y} \mapsto k\hat{\mathbf{x}}(k\mathbf{y}, \boldsymbol{\theta}_k), \text{ avec } \boldsymbol{\theta}_k = f(\boldsymbol{\theta}_1, k)$$

Cette propriété permet de réduire l'espace des lois *a priori* et par conséquent l'ensemble des fonctions potentiels.

voir (Djafari93, Brette et al. 94)

25

6.2 Non convexité graduelle (GNC)

Principe :

Solution MAP :

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \{p(\mathbf{x}|\mathbf{y})\} = \arg \min_{\mathbf{x}} \{J(\mathbf{x})\}$$

$J(\mathbf{x})$ multimodal

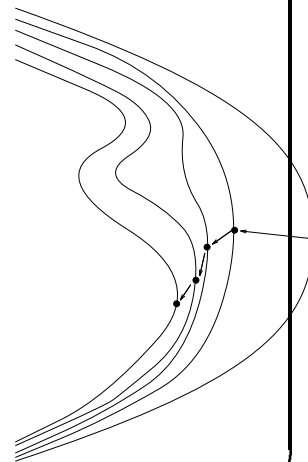
- Définir une suite de critères $J_r(\mathbf{x})$ telles que :

- $J_{r_0}(\mathbf{x})$ soit convexe
- $\forall \mathbf{x} \lim_{r \rightarrow r_\infty} J_r(\mathbf{x}) = J(\mathbf{x})$

- Minimiser $J_{r_0}(\mathbf{x})$ pour obtenir \mathbf{x}_0

- Pour une suite de $r = \{r_1, \dots, r_\infty\}$

minimiser localement $J_r(\mathbf{x})$ au voisinage de la solution obtenue à l'itération précédente.



26

Cas des problèmes linéaires

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \Phi(\mathbf{x})$$

Non convexité vient du choix de la loi *a priori* $\Phi(\mathbf{x})$

- [Blake et Zisserman] :
 - Segmentation d'image $\mathbf{A} = \mathbf{I} \rightarrow$ Problème bien posé
 - $J_{r_0}(\mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|^2 + \Phi_{r_0}(\mathbf{x}), \quad \exists r_0$ tel que J_{r_0} soit convexe
- [Nikolova, Djafari, Idier] :
 - Extension pour les problème mal posés \mathbf{A}
 - $\nexists r_0$ tel que J_{r_0} soit convexe \rightarrow

$$J_{r_k}(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \Phi_{r_k}(\mathbf{x}) + \Psi_{a_k}(\mathbf{x})$$

Double relaxations: $a_k \mapsto 0$ et $r_0 \mapsto r_\infty$

– Extension pour d'autres modèles que celui de la chaîne ou membrane faible

Cas des problèmes nonlinéaires

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}(\mathbf{x})\|^2 + \lambda\Phi(\mathbf{x})$$

Non convexité vient surtout du premier terme.

Exemple : Imagerie tomographique par ondes diffractées

$$\mathbf{A}(\mathbf{x}) = \mathbf{G}_M \mathbf{X} (\mathbf{I} - \mathbf{G}_O \mathbf{X})^{-1} \phi_0, \quad \mathbf{X} = \text{diag}(\mathbf{x})$$

Il fallait alors trouver un schéma de relaxation

(Thèse de H. Carfantan)

$$J_n(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}_n(\mathbf{x})\|^2 + \lambda\Phi(\mathbf{x})$$

$$\text{avec} \quad \mathbf{A}_n(\mathbf{x}) = \mathbf{G}_M \mathbf{X} (\mathbf{I} - r_n \mathbf{G}_O \mathbf{X})^{-1} \phi_0$$

$$r_0 = 0 \quad \rightarrow \mathbf{A}_0(\mathbf{x}) = \mathbf{G}_M \mathbf{X} \phi_0, \quad \text{Approximation de Born}$$

$$\lim_{n \rightarrow 1} r_n = 1 \quad \rightarrow \mathbf{A}_\infty(\mathbf{x}) = \mathbf{A}(\mathbf{x})$$

6.3 Estimation des hyperparamètres

$$p(\mathbf{x}|\boldsymbol{\theta}_1), \quad p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_2) \longrightarrow p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$$

Cas où les hyperparamètres sont inconnus :

$$p(\boldsymbol{\theta}_1), p(\boldsymbol{\theta}_2) \longrightarrow p(\mathbf{x}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y}) = p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)p(\boldsymbol{\theta}_1)p(\boldsymbol{\theta}_2)$$

- JMAP (ou MV généralisé):

$$(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) = \arg \max_{(\mathbf{x}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)} \{p(\mathbf{x}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y})\}$$

- Un schéma d'optimisation possible :

$$\begin{cases} \hat{\mathbf{x}}^{(k+1)} &= \arg \max_{\mathbf{x}} \left\{ p(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2^{(k)}) \right\} \\ \hat{\boldsymbol{\theta}}_1^{(k+1)} &= \arg \max_{\boldsymbol{\theta}_1} \left\{ p(\boldsymbol{\theta}_1|\mathbf{y}, \hat{\mathbf{x}}^{(k)}, \hat{\boldsymbol{\theta}}_2^{(k)}) \right\} \\ \hat{\boldsymbol{\theta}}_2^{(k+1)} &= \arg \max_{\boldsymbol{\theta}_2} \left\{ p(\boldsymbol{\theta}_2|\mathbf{y}, \hat{\mathbf{x}}^{(k)}, \hat{\boldsymbol{\theta}}_1^{(k)}) \right\} \end{cases}$$

- Existence du maximum ? - Convergence de l'algorithme ?
- Caractéristiques de la solution ?

29

- Marginalisation (paramètres de nuisance):

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \{p(\mathbf{x}|\mathbf{y})\} \text{ avec } p(\mathbf{x}|\mathbf{y}) = \iint p(\mathbf{x}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y}) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2$$

- MAP marginal (ou MV marginal, Empirical Bayes ou encore Evidence framework):

$$p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y}) = \int p(\mathbf{x}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y}) d\mathbf{x}$$

$$\begin{cases} (\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) &= \arg \max_{(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)} \{p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y})\}, \\ \hat{\mathbf{x}} &= \arg \max_{\mathbf{x}} \{p(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)\} \end{cases}$$

Mais en général le calcul de $p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y})$ est difficile \longrightarrow

- Approximation gaussienne,
- Algorithme EM, SEM, etc.

30

7 Exemples d'application

- Restauration d'image
- Reconstruction d'image en tomographie X
- Synthèse d'ouverture et reconstruction d'image en radioastronomie
- Synthèse de Fourier et reconstruction d'image par ondes diffractées
- Imagerie par courants de Foucault en CND
- Tomographie d'impédance

8 Problèmes ouverts

- Modélisation du problème direct
linéaire, bilinéaire, non linéaire, ...
- Modélisation des objets (signaux, images, bruit) étudiés
Paramétrique / Non paramétrique
Gaussien / Non Gaussien, Entropique / Markovien
- Choix des critères ou des fonctions coût
MAP, MP, MMP, ...
- Algorithmes d'optimisation et techniques de mise en œuvre
Locale / Globale, Déterministe / Stochastique
techniques multigrilles
- Choix des hyperparamètres et sélection du modèle
- Mesure de qualité de l'estimation