Méthodes probabilistes et 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 ?
– Deux exemples mathématiquement identiques:
... un problème de gestion du transport et
... un problème de reconstruction d'image en tomographie X
2. Problèmes inverses linéaires
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 de l'estimation
6. Problèmes ouverts, Exemple d'applications et Conclusions

1 Problèmes inverses

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

modèle	mesures	grandeurs inconnues	grandeurs inconnues intermédiaires	erreurs et bruit

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

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

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

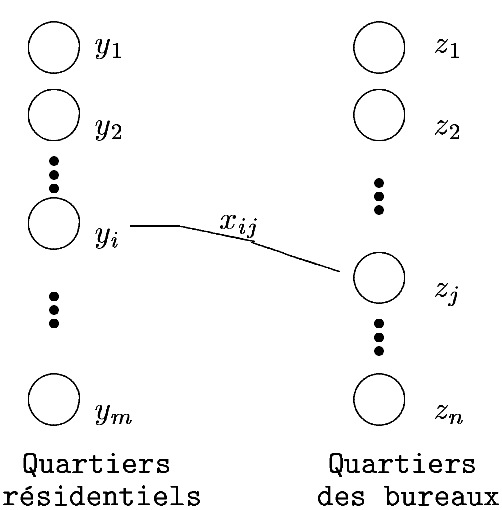
Relation entre $\boldsymbol{x}$ et $\boldsymbol{z}$: $\begin{cases} \boldsymbol{y} = \mathcal{A}_1(\boldsymbol{x}, \boldsymbol{z}) + \boldsymbol{b} \\ \mathcal{A}_2(\boldsymbol{x}, \boldsymbol{z}) = 0 \end{cases}$

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

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

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

Ex1: Problème du gestion de transport



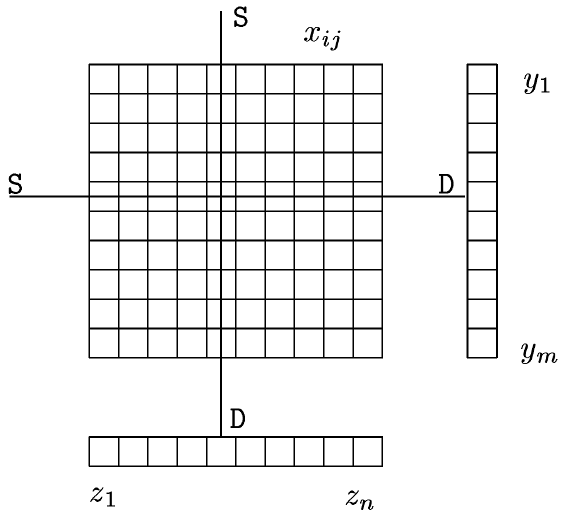
y_i nombre des résidents dans le quartier i

z_j nombre des places bureaux dans le quartier j

$$\sum_{j=1}^{n} x_{ij} = y_i, \quad i = 1, \cdots m$$

$$\sum_{i=1}^{m} x_{ij} = z_j, \quad j = 1, \cdots n$$

Ex2: Problème de la reconstruction d'image en tomographie



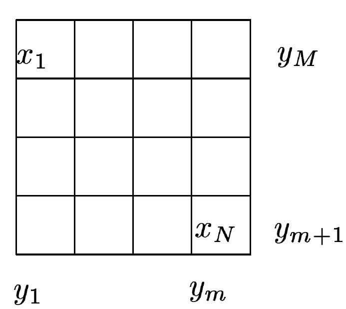
y_i projection verticale

z_j projection horizontal

$$\sum_{j=1}^{n} x_{ij} = y_i, \quad i = 1, \cdots m$$

$$\sum_{i=1}^{m} x_{ij} = z_j, \quad j = 1, \cdots n$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_8 \end{pmatrix} = \begin{pmatrix} 1000100010001000 \\ 0100010001000100 \\ 0010001000100010 \\ 0001000100010001 \\ 0000000000001111 \\ 0000000011110000 \\ 0000111100000000 \\ 1111000000000000 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ \vdots \\ x_{16} \end{pmatrix}$$



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

- Problème direct: Étant donnée $\boldsymbol{x}$ calculer $\boldsymbol{y}$
- Problème inverse: Étant donnée $\boldsymbol{y}$ déterminer ou estimer $\boldsymbol{x}$

– Existence de la solution

– Unicité de la solution

– Stabilité de la solution (sensibilité vis-à-vis des perturbations)

Existence et Unicité :

				40
	?			40
				40
				40
40	40	40	40	

10	10	10	10	40
10	10	10	10	40
10	10	10	10	40
10	10	10	10	40
40	40	40	40	

40	0	0	0	40
0	40	0	0	40
0	0	40	0	40
0	0	0	40	40
40	40	40	40	

15	5	5	15	40
5	15	15	5	40
5	15	15	5	40
15	5	5	15	40
40	40	40	40	

20	0	0	20	40
0	20	20	0	40
0	20	20	0	40
20	0	0	20	40
40	40	40	40	

20	10	5	5	40
10	20	5	5	40
5	5	20	10	40
5	5	10	20	40
40	40	40	40	

Séminaire CREST-INSEE, Lab. de Statistique, Malakoff, France. Lundi 13 Mai 1996

Stabilité :

$$A = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix} \qquad A^{-1} = \begin{pmatrix} 25 & -41 & 10 & -6 \\ -41 & 68 & -17 & 10 \\ 10 & -17 & 5 & -3 \\ -6 & 10 & -3 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix} \longrightarrow \boldsymbol{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix} \begin{pmatrix} x_1 + \delta x_1 \\ x_2 + \delta x_2 \\ x_3 + \delta x_3 \\ x_4 + \delta x_4 \end{pmatrix} = \begin{pmatrix} 32.1 \\ 22.9 \\ 33.1 \\ 30.9 \end{pmatrix} \longrightarrow \boldsymbol{x} + \delta \boldsymbol{x} = \begin{pmatrix} 9.2 \\ -12.6 \\ 4.5 \\ -1.1 \end{pmatrix}$$

$$\frac{||\delta \boldsymbol{y}||}{||\boldsymbol{y}||} = \frac{1}{300} \longrightarrow \frac{||\delta \boldsymbol{x}||}{||\boldsymbol{x}||} = \frac{10}{1}$$

$$\frac{||\delta \boldsymbol{x}||}{||\boldsymbol{x}||} = \mathrm{cond}(A) \frac{||\delta \boldsymbol{y}||}{||\boldsymbol{y}||}$$

$$\boldsymbol{\lambda} = \{30.3, \quad 3.86, \quad .84, \quad 0.01\}$$

$$\longrightarrow \mathrm{cond}(A) = \frac{\max|\lambda_i|}{\min|\lambda_i} = \frac{30.3}{0.01} \sim 3000$$

Sensibilité aux erreurs sur les paramètres du modèle :

$$\begin{pmatrix} 10 & 7 & 8.1 & 7.2 \\ 7.08 & 5.04 & 6 & 5 \\ 8 & 5.98 & 9.89 & 9 \\ 6.99 & 4.99 & 99.98 & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix} \longrightarrow \boldsymbol{x} = \begin{pmatrix} -81 \\ 137 \\ -34 \\ 22 \end{pmatrix}$$

Séminaire CREST-INSEE, Lab. de Statistique, Malakoff, France. Lundi 13 Mai 1996

- Problème linéaire discret et sans bruit

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

- Problème linéaire discret avec bruit additif

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

- Problème linéaire continu

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

- Problème non linéaire continu

$$\begin{aligned} \phi_d(\boldsymbol{r}_i) &= \iint_D G_m(\boldsymbol{r}_i, \boldsymbol{r}') \phi(\boldsymbol{r}') x(\boldsymbol{r}') \,\mathrm{d}\boldsymbol{r}', \quad \boldsymbol{r}_i \in S \\ \phi(\boldsymbol{r}) &= \phi_0(\boldsymbol{r}) + \iint_D G_o(\boldsymbol{r}, \boldsymbol{r}') \phi(\boldsymbol{r}') x(\boldsymbol{r}') \,\mathrm{d}\boldsymbol{r}', \quad \boldsymbol{r} \in D \end{aligned}$$

- Problème non linéaire discret

$$\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{x}) + \boldsymbol{b}$$

2 Problèmes inverses linéaires

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

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

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

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

Restauration d'image :

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

Reconstruction d'image :

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

Synthèse de Fourier :

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

3 Classification des méthodes

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

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

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

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

Limitations :

- Méthodes limitées aux cas de modèles simples
- Bruit et perturbations ne sont pas pris en compte
- Nécessite un ensemble complet de données non bruitées

3.2 Méthodes algébriques déterministes

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

- Rétrojection ou filtre adapté : $\widehat{\boldsymbol{x}} = \boldsymbol{A}^t\boldsymbol{y}$
- Inversion direct lorsque possible: $\widehat{\boldsymbol{x}} = \boldsymbol{A}^{-1}\boldsymbol{y}$
- Inverse gńéralisée:

– $M > N$ et $\text{rang}\{A\} = N \longrightarrow \widehat{\boldsymbol{x}} = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t\boldsymbol{y}$

– $M < N$ et $\text{rang}\{A\} = M \longrightarrow \widehat{\boldsymbol{x}} = \boldsymbol{A}^t(\boldsymbol{A}\boldsymbol{A}^t)^{-1}\boldsymbol{y}$

– Décomposition en valeurs singulières : $\widehat{\boldsymbol{x}} = \sum_{k=1}^{r} \frac{< \boldsymbol{u}_k, \boldsymbol{y} >}{\lambda_k} \boldsymbol{v}_k$

- Restriction de l'espace des solutions :

$$\min f(\boldsymbol{x}) \text{ s.c. } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}, \quad \text{avec, par exemple}$$

$$f(\boldsymbol{x}) = \sum_{j=1}^{n} x_j^2, \quad f(\boldsymbol{x}) = -\sum_{j=1}^{n} x_j \log x_j, \quad f(\boldsymbol{x}) = \sum_{j=1}^{n} \log x_j, \quad ...$$

- Définition de la solution comme le minimum d'un critère :

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

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

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

ou encore

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

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)

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

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

- Cas linéaire gaussien : $\longrightarrow$ Moindres Carrés

$$-\ln p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{\sigma^2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 \longrightarrow \widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\{\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2\right\}$$

- Méthode rarement satisfaisante pour les problèmes inverses
- Maximum de vraisemblance pénalisée

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

4.2 Maximum d'entropie classique (ME)

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

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

Solution :

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

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

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

Prise en compte du bruit :

max $H(\boldsymbol{x})$ s.c. $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 \leq c \longrightarrow \widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\{\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 + \lambda H(\boldsymbol{x})\right\}$

Difficultée : choix de λ

4.3 Maximum d'entropie sur la moyenne (MEM)

$\boldsymbol{x} = \mathrm{E}\{\boldsymbol{Z}\} \in \mathcal{C}$, $\boldsymbol{Z}$ un vecteur aléatoire de loi de probabilité $p(\boldsymbol{z})$

Modèle de mesure : $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\mathrm{E}\{\boldsymbol{Z}\}$ où $\mathrm{E}\{\boldsymbol{Z}\} = \int \boldsymbol{z} p(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}$

$$\text{minimiser} \quad \int_{\mathcal{C}} p(\boldsymbol{z}) \ln \frac{p(\boldsymbol{z})}{p_0(\boldsymbol{z})}\,\mathrm{d}\boldsymbol{z} \quad \text{s.c.} \quad \boldsymbol{y} = \boldsymbol{A}\mathrm{E}\{\boldsymbol{Z}\}$$

$p_0(\boldsymbol{z})$ une loi a priori pour $\boldsymbol{Z}$. **Solution :**

$$p(\boldsymbol{z};\boldsymbol{\lambda}) = \frac{1}{Z(\boldsymbol{\lambda})} p_0(\boldsymbol{z}) \exp\left[\boldsymbol{\lambda}^t[\boldsymbol{A}\boldsymbol{z}]\right] = p_0(\boldsymbol{z}) \exp\left[\boldsymbol{\lambda}^t[\boldsymbol{A}\boldsymbol{z}] - \ln Z(\boldsymbol{\lambda})\right]$$

$$Z(\boldsymbol{\lambda}) = \int p_0(\boldsymbol{z}) \exp\left[\boldsymbol{\lambda}^t[\boldsymbol{A}\boldsymbol{z}]\right]\,\mathrm{d}\boldsymbol{z}$$

$$\frac{\partial \ln Z(\boldsymbol{\lambda})}{\partial \lambda_i} = y_i, \quad i = 1, \cdots, M \longrightarrow \boldsymbol{\lambda} \longrightarrow \widehat{\boldsymbol{x}}(\boldsymbol{\lambda}) = \mathrm{E}\{\boldsymbol{Z}\} = \int \boldsymbol{z} p(\boldsymbol{z};\boldsymbol{\lambda})\,\mathrm{d}\boldsymbol{z}$$

Notant :

$$\boldsymbol{s} = \boldsymbol{A}^t\boldsymbol{\lambda}, \quad G(\boldsymbol{s}) = \ln Z(\boldsymbol{s}) = \ln \int p_0(\boldsymbol{z}) \exp\left[\boldsymbol{s}^t\boldsymbol{z}\right]\,\mathrm{d}\boldsymbol{z},$$

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

on a

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

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

$$\widehat{\boldsymbol{x}}(\boldsymbol{s}) = \frac{\mathrm{d}G(\boldsymbol{s})}{\mathrm{d}\boldsymbol{s}}, \qquad \text{relation explicite}$$

- G et H dépendent de la loi a priori $p_0(\boldsymbol{z})$.
- Si on note $\boldsymbol{m} = \int \boldsymbol{z} p_0(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}$ alors on montre que $H(\boldsymbol{x}) = H(\boldsymbol{x}, \boldsymbol{m})$ est une mesure de distance entre $\boldsymbol{x}$ et $\boldsymbol{m}$:
  - $H(\boldsymbol{x}, \boldsymbol{m}) \geq 0$, et $H(\boldsymbol{x}, \boldsymbol{m}) = 0$ ssi $\boldsymbol{x} = \boldsymbol{m}$
  - $H(\boldsymbol{x}, \boldsymbol{m})$ est dérivable et strictement convexe sur $\mathcal{C}$
  - $H(\boldsymbol{x}, \boldsymbol{m}) = \infty$ si $\boldsymbol{x} \notin \mathcal{C}$.

Si

$$p_0(\boldsymbol{z}) = \prod_{j=1}^{N} \mu_j(z_j) \quad \text{alors} \quad p(\boldsymbol{z}, \boldsymbol{\lambda}) = \prod_{j=1}^{N} p_j(z_j, \boldsymbol{\lambda})$$

et

$$G(\boldsymbol{s}) = \sum_j g_j\left(s_j\right), \quad H(\boldsymbol{x}, \boldsymbol{m}) = \sum_j h_j(x_j, m_j), \quad \widehat{x}_j = g_j'(s_j)$$

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

$$G(\boldsymbol{\lambda}) = \sum_j g_j\left([\boldsymbol{A}^t\boldsymbol{\lambda}]_j\right), \; H(\boldsymbol{x}, \boldsymbol{m}) = \sum_j h_j(x_j, m_j), \; \widehat{x}_j = g_j'\left([\boldsymbol{A}^t\widehat{\boldsymbol{\lambda}}]_j\right)$$

- $h_j(x, m)$ et $g_j(x)$ dépendent de $\mu_j(x)$.
- $g_j(x)$ est le convexe conjugué de $h_j(x, m)$

- $g_j(s)$ est la transformée de Cramer (log transformée de Laplace) de $\mu_j(x)$

$$g_j(s) = \ln \int \mu_j(x) \exp\left[sx\right] \mathrm{d}x,$$

Exemples :

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

• Lorsque $p_0(\boldsymbol{x})$ n'est pas séparable il devient très difficile de mener les calcul (excepté le cas gaussien).

• Cas gaussien : $p_0(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{m}, \boldsymbol{R}_x)$ où on a

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

Autre cas ?

Prise en compte du bruit :

Deux approches possibles :

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

1- Remplacer $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ par $|y_i - [\boldsymbol{A}\boldsymbol{x}]_i| < \epsilon$

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

2- Remplacer $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ par $\boldsymbol{y} = [\boldsymbol{A}|\boldsymbol{I}]\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{b}\end{bmatrix} \longrightarrow \boldsymbol{y} = \tilde{\boldsymbol{A}}\,\tilde{\boldsymbol{x}}$

Si $p_0(\tilde{\boldsymbol{x}}) = \mu_x(\boldsymbol{x})\mu_b(\boldsymbol{b})$ alors $\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\mathcal{C}} \{\mathcal{Q}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) + \alpha H(\boldsymbol{x})\}$

$$\text{avec} \qquad H(\boldsymbol{x}) = \sum_{j=1}^{N} h_j(x_j), \qquad \text{et} \qquad \mathcal{Q}(\boldsymbol{z}) = \sum_{i=1}^{M} q_i(z_j),$$

$h_j(x_j)$ et $q_i(z_i)$ dépendent des mesures de référence $\mu_x(\boldsymbol{x})$ et $\mu_b(\boldsymbol{x})$.

Cependant la détermination du α reste empirique.

5 Approche bayésienne de l'estimation

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

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

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

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

Estimateurs ponctuels :

- Maximum *a posteriori* (MAP) :

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

- Moyenne *a posteriori* (MP) :

$$C(\boldsymbol{x},\widehat{\boldsymbol{x}}) = [\boldsymbol{x}-\widehat{\boldsymbol{x}}]^t \boldsymbol{Q} [\boldsymbol{x}-\widehat{\boldsymbol{x}}]^t \longrightarrow \widehat{\boldsymbol{x}} = \mathrm{E}_{x|y}\{\boldsymbol{x}\} = \int \boldsymbol{x}\, p(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\beta})\,\mathrm{d}\boldsymbol{x}$$

- MAP Marginale (MAPM) :

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

où

$$p(x_i|\boldsymbol{y};\boldsymbol{\theta}) = \int p_{x|y}(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta})\,\mathrm{d}x_1 \cdots \mathrm{d}x_{i-1} \cdots \mathrm{d}x_{i+1} \cdots \mathrm{d}x_n$$

Cas des problèmes inverses linéaires

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

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

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

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

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

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

$$\widehat{\boldsymbol{x}} = \boldsymbol{P}\boldsymbol{A}^t(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_0), \quad \boldsymbol{P} = \left(\boldsymbol{A}^t\boldsymbol{A} + \lambda \boldsymbol{P}_0^{-1}\right)^{-1}$$

- La solution au sens du MAP :

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

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

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

• Hypothèse Gamma sur $\boldsymbol{x}$:

$$p(x_j) \propto (x_j/m_j)^\alpha \exp\left[-x_j/m_j\right] \longrightarrow \quad \Phi(\boldsymbol{x}) = \alpha \sum_j \ln \frac{x_j}{m_j} + \frac{x_j}{m_j},$$

• Hypothèse Béta sur $\boldsymbol{x}$:

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

• Hypothèse gaussienne généralisée sur $\boldsymbol{x}$:

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

• Modèles markoviens pour $\boldsymbol{x}$:

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

Cadre général : Estimé MAP :

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\{J(\boldsymbol{x})\right\}, \text{avec } J(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 + \Phi_\alpha(\boldsymbol{x})$$

• Lois gaussiennes : $\Phi_\alpha(\boldsymbol{x})$ quadratique
$\longrightarrow J(\boldsymbol{x})$ quadratique $\longrightarrow$ Estimateur linéaire

• Lois entropiques : $\Phi_\alpha(\boldsymbol{x})$ convexe et séparable : $\Phi(\boldsymbol{x}) = \sum_j \phi_\alpha(x_j)$
avec $\quad \phi_\alpha(x_j) = \left\{\alpha x_j^p, \quad \alpha(x_j \ln x_j - x_j), \quad \alpha(\ln x_j - x_j)\right\}$
$\longrightarrow J(\boldsymbol{x})$ convexe $\longrightarrow$ Estimateur nonlinéaire mais facile à calculer

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

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

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

6 Problèmes ouverts

- Comment choisir $p(\boldsymbol{x}|\boldsymbol{\beta})$
  Dans le cas des modèles markoviens comment choisir les fonctions potentiels
  $\longrightarrow$ Invariance d'echelle (Djafari93, Brette *et al.* 93, 94)
- Comment déterminer les hyperparamètres $(\boldsymbol{\theta}, \boldsymbol{\beta})$
  $\longrightarrow$ MAP généralisée, MAPM, EM, SEM, ...
  (Djafari 95, Idier 94)
- Comment choisir un estimateur : MAP, MP ou MAPM
  $\longrightarrow$ Plutôt considération calculatoires
- Comment calculer efficacement la solution $\widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\beta})$
  $\longrightarrow$ Algorithmes d'optimisation :
  – Recuit simulé,

– Relaxation déterministe (GNC) (Thèse Nikolova 95)

6.1 Estimation des hyperparamètres

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

Cas où les hyperparamètres sont inconnus :

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

• JMAP (ou MV généralisé):

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

• Un schéma d'optimisation possible :

$$\left\{\begin{array}{lcl} \widehat{\boldsymbol{x}}^{(k+1)} &=& \arg\max_{\boldsymbol{x}} \left\{p(\boldsymbol{x}|\boldsymbol{y},\widehat{\boldsymbol{\theta}}_1^{(k)},\widehat{\boldsymbol{\theta}}_2^{(k)})\right\} \\ \widehat{\boldsymbol{\theta}}_1^{(k+1)} &=& \arg\max_{\boldsymbol{\theta}_1} \left\{p(\boldsymbol{\theta}_1|\boldsymbol{y},\widehat{\boldsymbol{x}}^{(k)},\widehat{\boldsymbol{\theta}}_2^{(k)})\right\} \\ \widehat{\boldsymbol{\theta}}_2^{(k+1)} &=& \arg\max_{\boldsymbol{\theta}_2} \left\{p(\boldsymbol{\theta}_2|\boldsymbol{y},\widehat{\boldsymbol{x}}^{(k)},\widehat{\boldsymbol{\theta}}_1^{(k)})\right\} \end{array}\right.$$

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

• Marginalisation (paramètres de nuisance):

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

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

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

$$\left\{\begin{array}{lcl} (\widehat{\boldsymbol{\theta}}_1,\widehat{\boldsymbol{\theta}}_2) &=& \arg\max_{(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)} \{p(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2|\boldsymbol{y})\}, \\ \widehat{\boldsymbol{x}} &=& \arg\max_{\boldsymbol{x}} \left\{p(\boldsymbol{x}|\boldsymbol{y},\widehat{\boldsymbol{\theta}}_1,\widehat{\boldsymbol{\theta}}_2)\right\} \end{array}\right.$$

Mais en général le calcul de $p(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2|\boldsymbol{y})$ est difficile $\longrightarrow$
– Approximation gaussienne,
– Algorithme EM, SEM, etc.

7 Exemples d'application

- Restauration des signaux mesurés par des capteurs imparfaits
- 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 Foucaut en CND
- Tomographie d'impédance
- Fusion de données gammagraphie et ultrasonore en CND

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