

MODÉLISATION MARKOVIENNE
POUR LA RÉOLUTION DES
PROBLÈMES INVERSES EN IMAGERIE

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 en imagerie
2. Estimation bayésienne
3. Modélisation markovienne
 - Introduction et notations (sites, voisinages, cliques,...)
 - Tentative de classification (sens large ou sens stricte)
 - Champs de Gibbs et équivalence Gibbs-Markov
 - Exemples de modèles (Auto-modèles, Auto-logistique, Ising, ...)
 - Modèles couplés ou hiérarchiques
4. Modèles markoviens en segmentation d'image
5. Outils d'optimisation (Recuit, Échantillonneur de Gibbs)
6. Modèles markoviens en restauration ou reconstruction d'image
7. Problèmes et questions ouverts (Recuit/GNC, convexe/pas, ..)

1 Problèmes inverses en imagerie

- Modèle d'observation : $\mathbf{y} = \mathcal{A}(\mathbf{x}) + \mathbf{b}$
 - \mathbf{x} : image inconnue
 - \mathbf{y} : données (observations)
 - \mathbf{b} : bruit
 - \mathcal{A} : opérateur liant \mathbf{x} et \mathbf{y} (modèle d'observation)
- Traitement d'image : \mathbf{y} = image dégradée
 - Débruitage, Segmentation : $\mathbf{y} = \mathbf{x} + \mathbf{b}$, $\mathbf{A} = \mathbf{I}$
 - Restauration : $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$, \mathbf{A} opérateur de flou
- Reconstruction d'image : \mathbf{y} n'est pas une image
 - Tomographie, imagerie en radio-astronomie, CND, ... :
 - $\mathbf{y} = \mathcal{A}(\mathbf{x}) + \mathbf{b}$, \mathcal{A} souvent non linéaire

- Modèle linéaire + bruit additif : $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$
 - Débruitage, Segmentation : ($\mathbf{A} = \mathbf{I}$)
 - Restauration, Reconstruction d'image : \mathbf{A} mal conditionnée
- Régularisation : minimisation d'un critère

$$J(\mathbf{x}) = Q(\mathbf{y} - \mathbf{A}\mathbf{x}) + \lambda\Omega(\mathbf{x})$$

- $Q(\mathbf{y} - \mathbf{A}\mathbf{x})$: Adéquation aux données
- $U(\mathbf{x})$: Connaissances a priori
- Interprétation bayésienne :
 - estimation au sens du MAP:

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

- loi a posteriori : $p(\mathbf{x}|\mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})/p(\mathbf{y})$
- loi des observations : $p(\mathbf{y}|\mathbf{x}) \propto \exp[-Q(\mathbf{y} - \mathbf{A}\mathbf{x})]$
- loi a priori : $p(\mathbf{x}) = \exp[-\lambda U(\mathbf{x})]$

2 Inférence bayésienne

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

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

- Moyenne *a posteriori* (MP) :

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

- MAP Marginale (MAPM) :

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

où

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

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} : $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \theta\mathbf{P}_0)$:

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

- Loi a posteriori

$$p(\mathbf{x}|\mathbf{y}, \theta, \beta) \propto \exp \left[-\frac{1}{2}J(\mathbf{x}) \right] \text{ avec } J(\mathbf{x}) = Q(\mathbf{x}) + \lambda U(\mathbf{x}),$$

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

- La solution au sens du MAP :

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

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

- Hypothèse Gamma sur \mathbf{x} :

$$x_j \sim \mathcal{G}(\alpha, m_j) \rightarrow p(x_j) \propto (x_j/m_j)^\alpha \exp[-x_j/m_j]$$

$$\rightarrow U(\mathbf{x}) = \alpha \sum_j \ln \frac{x_j}{m_j} + \frac{x_j}{m_j}$$

- Hypothèse Beta sur \mathbf{x} :

$$x_j \sim \mathcal{G}(\alpha, m_j) \rightarrow p(x_j) \propto x_j^\alpha (1 - x_j)^\beta$$

$$\rightarrow U(\mathbf{x}) = \alpha \sum_j \ln x_j + \beta \sum_j \ln(1 - x_j)$$

- Modèles markoviens pour \mathbf{x} :

$$U(\mathbf{x}) = \alpha \sum_j \sum_{i \in N_j} V(x_j, x_i)$$

3 Modélisation markovienne

3.1 Introduction et notations

Site et Voisinage

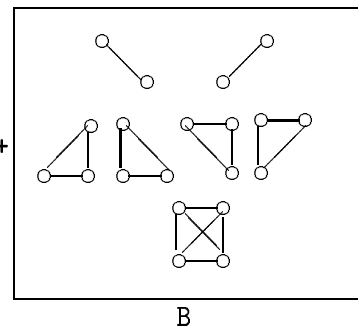
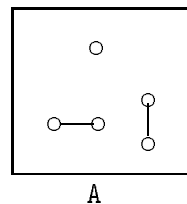
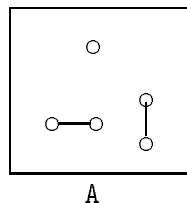
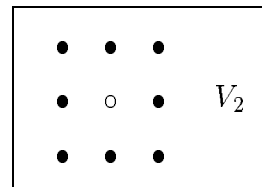
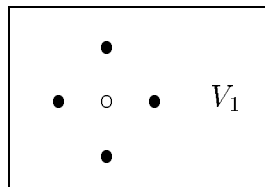
- $\mathcal{S} = \{s_1, \dots, s_n\}$ un ensemble de sites
- $\mathcal{V} = V_s, s \in \mathcal{S}$ une collection de voisinage sur \mathcal{S}
 - $s \notin V_s, r \in V_s \iff s \in V_r$
 - V_s est le voisinage du site s, V_r est le voisinage du site r
- $\{\mathcal{S}, \mathcal{V}\}$ est un graphe

Cliques

$c \subset \mathcal{S}$ est une clique de $\{\mathcal{S}, \mathcal{V}\}$ si :

- $|c| = 1$ ou si $|c| > 1$ et $\forall s_1, s_2 \in c \rightarrow s_1$ et s_2 sont voisins
- $C = \{c\}$ ensemble de cliques de $\{\mathcal{S}, \mathcal{V}\}$

Exemple : $\mathcal{S} = \mathbf{Z}^2$



Champ Aléatoire

- $\mathcal{S} = \{s_1, \dots, s_n\}$ un ensemble de sites
 - $\mathcal{G} = \{g_1, \dots, g_k\}$ espace des phases (niveau de gris, couleurs, etc.)
 - $\mathcal{X} = X_s, s \in \mathcal{S}$ champ aléatoire sur \mathcal{S} à valeurs dans \mathcal{G}
 - X_s variable aléatoire à valeurs dans \mathcal{G}
- ensemble des configurations possibles :

$$\Omega = \{\mathbf{X} = (X_{s_1}, \dots, X_{s_N}) : X_{s_n} \in \mathcal{G}, n = 1, \dots, N\}$$

- une réalisation de \mathcal{X} $\{\mathbf{X} = \mathbf{x}\} = \{X_{s_1} = x_{s_1}, \dots, X_{s_N} = x_{s_N}\}$
- loi de probabilité conjointe $P(\mathbf{X} = \mathbf{x})$, cas discret
- loi de probabilité conjointe $p(\mathbf{x})$, cas continu
- loi de probabilité conditionnelle

$$P(X_s = x_s | X_r = x_r), \text{ cas discret} \quad \text{ou} \quad p(x_s | X_r = x_r), \text{ cas continu}$$

Champ Aléatoire Markovien

- \mathcal{X} champ aléatoire sur \mathcal{S} à valeurs dans \mathcal{G} est dit markovien relativement au système de voisinage \mathcal{V} si
 - $P(\mathbf{X} = \mathbf{x}) > 0, \quad \forall \mathbf{x} \in \Omega, \quad \text{et}$
 - $\forall X_s \in \mathcal{G}, \forall s \in \mathcal{S},$
$$P(X_s = x_s | X_r = x_r, r \neq s) = P(X_s = x_s | X_r = x_r, r \in V_s)$$
- Les fonctions : $\pi_s(x_s | x_{V_s}) = P(X_s = x_s | X_{V_s} = x_{V_s}), s \in \mathcal{S}$ sont les caractéristiques locales du champ markovien \mathcal{X}

Deux questions :

Etant donné un système $\{\pi_s(x_s | x_{V_s}), s \in \mathcal{S}\} = \mathcal{P}$,

1. existe-t-il un champ markovien avec \mathcal{P} comme caractéristiques locales ?
2. est-il unique ?

3.2 Tentative de Classifications

- **Champ de markov au sens strict (SSM) :**

(Définition via les probabilités conditionnelles)

Soient $\mathcal{B} \subseteq \mathcal{A} \subseteq \mathcal{S}$

$$P(X_{ij} = x_{ij} | X_{\mathcal{B}} = x_{\mathcal{B}}) = P(X_{ij} = x_{ij} | X_{\mathcal{A}} = x_{\mathcal{A}})$$

- **Champ de markov au sens large (WSM)**

(Définition via la régression linéaire)

(Estimation linéaire en moyenne quadratique ELMQ)

$$X = \{X_{ij}\}, \quad (i, j) \in \mathbf{Z}^2, \quad \mathcal{B} \subseteq \mathcal{A} \subseteq \mathcal{S}$$

$$ELMQ \{X_{ij} | X_{\mathcal{B}} = x_{\mathcal{B}}\} = ELMQ \{X_{ij} | X_{\mathcal{A}} = x_{\mathcal{A}}\}$$

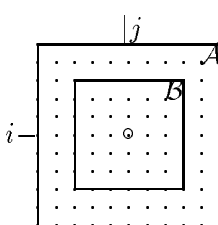
– Propriétés au sens large = Propriétés au deuxième ordre

– Représentation équivalente par un modèle AR

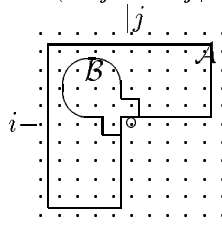
$$X_s = \sum_r a_r X_r + Z_s$$

Champ de markov au sens strict (SSM)

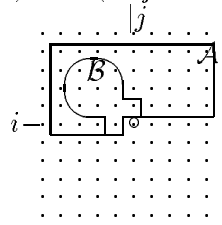
Soient $\mathcal{B} \subseteq \mathcal{A} \subseteq \mathcal{S}$ $P(X_{ij} = x_{ij} | X_{\mathcal{B}} = x_{\mathcal{B}}) = P(X_{ij} = x_{ij} | X_{\mathcal{A}} = x_{\mathcal{A}})$



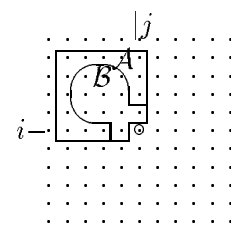
Bilatéral



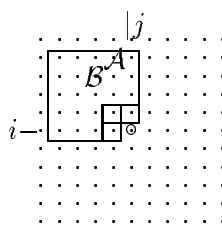
MMRF



Unilateral



Causal



Pickard

- Pickard \subseteq Causal**
- Causal \subseteq Unilatéral**
- Unilatéral \subseteq MMRF**
- MMRF \subseteq Bilatéral**

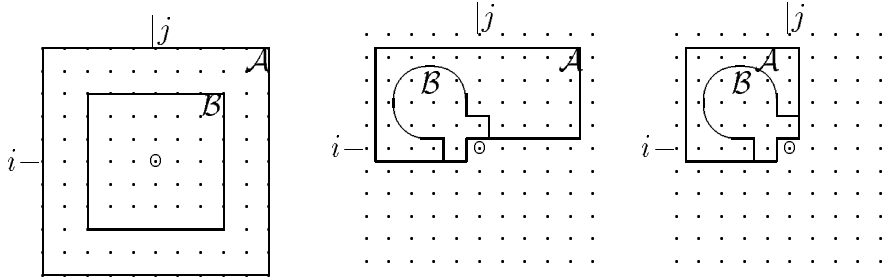
Champ de markov au sens large (WSM)

(Définition via la régression linéaire)

(Estimation linéaire en moyenne quadratique ELMQ)

$$X = \{X_{ij}\}, \quad (i, j) \in \mathbf{Z}^2, \quad \mathcal{B} \subseteq \mathcal{A} \subseteq \mathcal{S}$$

$$ELMQ \{X_{ij} | X_{\mathcal{B}} = x_{\mathcal{B}}\} = ELMQ \{X_{ij} | X_{\mathcal{A}} = x_{\mathcal{A}}\}$$



Bilatéral

Unilateral

Causal

Champ de markov gaussien : sens large = sens stricte

Un exemple classique utilisé en traitement d'image :

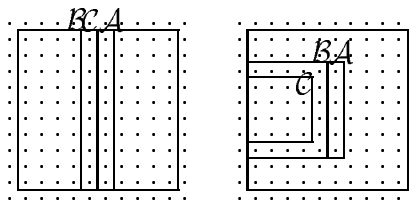
$$\mathcal{S} = \mathbf{Z}_{mn}^2 = \{(i, j) : 1 \leq i \leq m, 1 \leq j \leq n\}, \quad \mathcal{X} = \{X_{i,j}, (i, j) \in \mathcal{S}\}$$

$$\mathcal{V}_1 = \begin{pmatrix} \bullet & & \\ \bullet & \circ & \bullet \\ \bullet & & \bullet \end{pmatrix} \quad \mathcal{V}_2 = \begin{pmatrix} \bullet & \bullet & \bullet \\ \bullet & \circ & \bullet \\ \bullet & \bullet & \bullet \end{pmatrix}$$

$$P(X_{ij} = x_{ij} | X_{kl} = x_{kl}, (k, l) \in \mathcal{S} - \{i, j\}) = P(X_{ij} = x_{ij} | X_{kl} = x_{kl}, (k, l) \in \mathcal{V}_{ij})$$

Quelques propriétés :

$$P(x_a, a \in \mathcal{A} | x_b, b \in \mathcal{B}) = P(x_a, a \in \mathcal{A} | x_b, b \in \mathcal{C})$$



3.3 Champs de Gibbs

Energie et Potentiel

- $\{V_a, a \subseteq \mathcal{S}\} : \Omega \rightarrow \mathbb{R}^+$ telle que $V_a(\mathbf{x})$ ne dépend que de \mathbf{x}_a
- $V_c, c \in \mathcal{C}$ est un potentiel de voisinage si $V_c = 0$ lorsque c n'est pas une clique
- $U(\mathbf{x}) = \sum_{a \subseteq \mathcal{S}} V_a(\mathbf{x})$ est l'énergie associée au potentiels V_a

Lorsque U est l'énergie associée au potentiel de voisinage $V_c, c \in \mathcal{C}$ on a

$$U(\mathbf{x}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{x})$$

Mesure de Gibbs

- la mesure de Gibbs associée à l'énergie $U(\mathbf{x})$ et à la température T est la mesure de probabilité sur Ω :

$$\pi_T(\mathbf{x}) = \frac{1}{Z_T} \exp \left[-\frac{U(\mathbf{x})}{T} \right]$$

– la constante

$$Z_T = \sum_{\mathbf{x} \in \Omega} \exp \left[-\frac{U(\mathbf{x})}{T} \right]$$

est la fonction de partition.

– T est la température

- Un champ aléatoire \mathcal{X} défini sur \mathcal{S} est un champ de Gibbs si :

$$P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \exp \left[-\frac{U(\mathbf{x})}{T} \right] \quad \text{avec} \quad U(\mathbf{x}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{x})$$

Équivalence Gibbs-Markov

Théorème : Soit \mathcal{V} un système de voisinage sur \mathcal{S} . Alors, \mathcal{X} est un champ markovien sur \mathcal{S} à valeurs dans \mathcal{G} relativement au système de voisinage \mathcal{V} ssi $p(\mathbf{x}) = P(\mathbf{X} = \mathbf{x})$ est une distribution de Gibbs relativement à \mathcal{V} .

(Hammersley & Clifford)
 Gibbs \iff Markov
 Potentiels \iff caractéristiques locales

Gibbs \longrightarrow Markov :

$$P(X_s = x_s | X_{\mathcal{S}-s} = x_{\mathcal{S}-s}) = \frac{\exp \left[-1/T \sum_{c \in \mathcal{C}} V_c(x_s) \right]}{\sum_{y_s \in \mathcal{G}} \exp \left[-1/T \sum_{c \in \mathcal{C}} V_c(y_s, x_{\mathcal{S}-s}) \right]}$$

Markov \longrightarrow Gibbs : plus délicat. Potentiels canoniques

Quelques exemples de modèles classiques

- **Auto-modèles :**

$$U(\mathbf{x}) = \sum_r x_r g_r(x_r) + \sum_r \sum_s \beta_{rs} x_r x_s, \quad r, s \in \mathcal{S}$$

$g_r(\cdot)$ fonctions arbitraires

β_{rs} coefficients d'interaction entre les sites r et s

$\beta_{rs} \neq 0$ ssi les sites r et s sont voisins : $|r - s|^2 \leq d$

- **Auto-logistique :**

$\mathcal{G} = \{0, 1\}$ ou $\{-1, +1\}$

$$U(\mathbf{x}) = \sum_r \alpha_r x_r + \sum_r \sum_s \beta_{rs} x_r x_s, \quad r, s \in \mathcal{C}$$

Si $\alpha = 0$, et $\beta_{rs} = \begin{cases} \beta & \text{si } |r - s|^2 \leq d \\ 0 & \text{sinon} \end{cases}$, U mesure la longueur d'un

contour.

- **Ising** : Auto-logistique avec voisinage d'ordre 1

Deux exemples spécifiques :

- Chaîne de Markov binaire : $\mathcal{G} = \{0, 1\}$, $\mathcal{S} = \mathbf{Z}$,

$$\text{Matrice de transitions} \begin{pmatrix} 0-0 & 0-1 \\ 1-0 & 1-1 \end{pmatrix} \quad P = \begin{pmatrix} s & 1-s \\ 1-t & t \end{pmatrix}$$

$$\pi(\mathbf{x}) = \frac{1}{Z} \exp \left[\bar{V}_0 \sum_{i=0,n} x_i + V_0 \sum_{0 < i < n} x_i + V_1 \sum_{|i-j|=1} x_i x_j \right]$$

$$\bar{V}_0 = \log \frac{1-s}{s}, \quad V_0 = \log \frac{(1-s)(1-t)}{s^2}, \quad V_1 = \log \frac{st}{(1-s)(1-t)}$$

- Images binaires et voisinage d'ordre 1 :

$$\mathcal{G} = \{0, 1\}, \quad \mathcal{S} = \mathbf{Z}^2, \quad V_1 = \left\{ \begin{array}{ccc} & \bullet & \\ \bullet & \circ & \bullet \\ & \bullet & \end{array} \right\}$$

$$U(\mathbf{x}) = \alpha \sum x_{ij} + \beta \sum x_{i,j} v_{i,j}$$

avec

$$v_{i,j} = x_{i,j-1} + x_{i-1,j} + x_{i,j+1} + x_{i+1,j}$$

• **Multi-Level Logistic Model (MLL) ou Ising généralisé**
ou encore **Champ de labels**

$\mathcal{G} = g_1, \dots, g_M$ niveaux des gris dans chaque région

$$U(\mathbf{x}) = \sum_k V_k(x) + \sum_c V_c(\mathbf{x})$$

- pour les cliques simples

$$V_c(x) = \begin{cases} \alpha_k & \text{si } x_{ij} = g_k \\ 0 & \text{sinon} \end{cases}$$

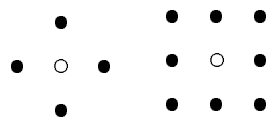
- pour les cliques multiples

$$V_c(\mathbf{x}) = \begin{cases} +\beta & \text{si } x_{ij} \text{ dans } C \text{ ont les mêmes valeurs} \\ -\beta & \text{sinon} \end{cases}$$

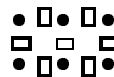
3.4 Modèles couplés ou hiérarchique

– \mathbf{Z}_m	réseau des entiers		○ ○ ○ ○ ○ ○ ○
– $\mathcal{S}_1 = \mathbf{Z}_m$	sites pixels	○	○ ○ ○ ○ ○ ○ ○
– $\mathcal{S}_2 = \mathbf{Z}_m$	sites contours	–	○ ○ ○ ○ ○ ○ ○
– $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$	sites image		○ ○ ○ ○ ○ ○ ○

- $\mathcal{F} = \{F_{i,j}, (i,j) \in \mathcal{S}_1\}$ champ intensité
- $\mathcal{L} = \{L_{\alpha,\beta}, (\alpha,\beta) \in \mathcal{S}_2\}$ champ des lignes
- $\mathcal{X} = (\mathcal{F}, \mathcal{L})$ image



Voisinage pixels



voisinage contours

Modèle hiérarchique :

$$P(\mathbf{f}, \mathbf{l}) = P(\mathbf{f}|\mathbf{l})P(\mathbf{l})$$

$$P(\mathbf{l}) = \frac{1}{Z_L} \exp[-U_L(\mathbf{l})]$$

$$P(\mathbf{f}|\mathbf{l}) = \frac{1}{Z_{F|L}} \exp[-U_{F|L}(\mathbf{f}|\mathbf{l})]$$

$$\rightarrow P(\mathbf{f}, \mathbf{l}) = \frac{1}{Z_{F,L}} \exp[-U_{F,L}(\mathbf{f}, \mathbf{l})],$$

avec

$$U_{F,L}(\mathbf{f}, \mathbf{l}) = U_{F|L}(\mathbf{f}|\mathbf{l}) + U_L(\mathbf{l})$$

4 Application en segmentation d'image

- $\mathcal{S} = \{(i, j) : 1 \leq i \leq m, 1 \leq j \leq n\}$ sites pixels ◦
- $\mathcal{X} = \{X_{i,j}, (i, j) \in \mathcal{S}\}$ image
- $\mathcal{Y} = \{Y_{i,j}, (i, j) \in \mathcal{S}\}$ image dégradée
- $\mathcal{N} = \{N_{i,j}, (i, j) \in \mathcal{S}\}$ bruit

Modèle de dégradation : $\mathcal{Y} = \mathcal{X} + \mathcal{N}$

$X_{i,j} = g_m$, si $X_{i,j} = m$ avec

$\mathcal{G} = \{g_1, \dots, g_M\}$ niveaux des gris dans chaque région

Hypothèses statistiques :

- \mathcal{X} un champ de Markov à valeur dans $\mathcal{G} = \{g_1, \dots, g_M\}$
- \mathcal{N} un champ gaussien blanc
- $N_{i,j}$ v.a. indépendantes et de lois $\mathcal{N}(0, \sigma^2)$
- \mathcal{N} et \mathcal{X} indépendants

Modèle a posteriori et choix de l'estimateur

- \mathcal{X} est un champ de Gibbs d'énergie $U(\mathbf{x})$:

$$P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \exp[-U(\mathbf{x})]$$

$$U(\mathbf{x}) = \sum_j \alpha_j x_j$$

$$+ \beta_1 \sum_{(i,j)0} x_i x_j + \beta_2 \sum_{(i,j)90} x_i x_j + \beta_3 \sum_{(i,j)-45} x_i x_j + \beta_4 \sum_{(i,j)45} x_i x_j$$

$$+ \gamma_1 \sum_{(i,j,k)} x_i x_j x_k + \gamma_2 \sum_{(i,j)90} x_i x_j x_k + \dots$$

$$+ \xi_1 \sum_{(i,j,k,l)} x_i x_j x_k x_l$$

avec M paramètres $\alpha_m, m = 1 = \dots, M$,

4 paramètres β_1, \dots, β_4 , 4 paramètres $\gamma_1, \dots, \gamma_4$ et

un paramètre ξ_1 .

- \mathcal{N} est un champ gaussien, blanc et indépendant de \mathcal{X} ,

$$P(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) = K \exp \left[-\frac{1}{2\sigma_b^2} \|\mathbf{y} - \mathbf{x}\|^2 \right]$$

- Règle de Bayes :

$$P(\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}) = \frac{P(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) P(\mathbf{X} = \mathbf{x})}{P(\mathbf{Y} = \mathbf{y})}$$

- Énergie a posteriori : $U^p(\mathbf{x}) = \frac{1}{2\sigma_b^2} \|\mathbf{y} - \mathbf{x}\|^2 + U(\mathbf{x})$

Deux problèmes :

- **Classification ou segmentation supervisée** : On connaît $\boldsymbol{\theta} = \{\{g_1, \dots, g_M\}, \{\beta_1, \dots, \beta_4\}, \{\gamma_1, \dots, \gamma_4\}, \xi_1\}$ et $\mathbf{y} \rightarrow \hat{\mathbf{x}}$.
- **Segmentation non supervisée** : $\mathbf{y} \rightarrow \hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}$
 - Apprentissage supervisée : $\mathbf{x}, \mathbf{y} \rightarrow \hat{\boldsymbol{\theta}}$
 - Apprentissage non supervisée : $\mathbf{y} \rightarrow \hat{\boldsymbol{\theta}}$

Différents estimateurs possibles pour le problème 1

- MAP (Maximum A Posteriori) : $\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \{P(\mathbf{x}|\mathbf{y})\}$
- MPM (Maximum of the Posterior Marginal) :
 $\hat{\mathbf{x}} = \{\hat{x}_s \in S\}$ avec $\hat{x}_s = \arg \max_{x_s} \{p_s(X_s = x_s|\mathbf{y})\}$
- PM (Posterior Mean) Moyenne a posteriori :

$$\hat{\mathbf{x}} = \{\hat{x}_s \in S\} \quad \text{avec} \quad \hat{x}_s = \sum_{x_i \in \mathcal{G}} x_s p_s(X_s = x_s|\mathbf{y})$$

Deux algorithmes stochastiques :

- **Le recuit simulé ou relaxation stochastique**
qui permet d'estimer le MAP
- **L'échantillonneur de Gibbs**
qui permet de simuler des réalisations de $p(x_s|\mathbf{y}, x_{j \neq s})$, et de $p(\mathbf{x}|\mathbf{y})$ et donc d'estimer le MPM ou la moyenne *a posteriori*.

Différents estimateurs possibles pour les problèmes 2

- Restauration : $\mathbf{y}, \theta \longrightarrow \hat{\mathbf{x}}$ MAP, PM, MPM
- Apprentissage non supervisée : $\mathbf{y} \longrightarrow \hat{\theta}$ MV (EM)
- Apprentissage supervisée : $\mathbf{x}, \mathbf{y} \longrightarrow \hat{\theta}$

Lorsque \mathbf{x} est modélisé par un champ unilatéral ou par une chaîne de Markov :

- $\mathbf{y}, \theta \longrightarrow \hat{\mathbf{x}}_{\text{MAP}}$ Viterbi (Programmation dynamique)
- $\mathbf{y} \longrightarrow \hat{\theta}_{\text{MV}}$ EM, (Forward-Backward)
- $\mathbf{y} \longrightarrow \hat{\mathbf{x}}_{\text{MAP}}, \hat{\theta}_{\text{MV}}$: Algorithme itératif, par ex. MVG:

$$\hat{\theta}^{(0)}, \mathbf{y} \longrightarrow \hat{\mathbf{x}}_{\text{MAP}}^{(1)}, \mathbf{y} \longrightarrow \hat{\theta}_{\text{MV}}^{(1)}, \mathbf{y} \longrightarrow \hat{\mathbf{x}}_{\text{MAP}}^{(1)}, \mathbf{y} \longrightarrow \dots$$

5 outils de simulation et d'optimisation

Echantillonneur de Gibbs

- Soit $P(\mathbf{X} = \mathbf{x})$ une distribution de probabilité.
- Soit $(s_k), k \in \mathcal{N}$ une suite de visite des sites telle que, pour tout $s \in \mathcal{S}$, s est visité un nombre infini de fois.

$$X_s^{(k+1)} = \begin{cases} X_s^{(k+1)} & \text{si } s \neq s_k \\ \xi & \text{si } s = s_k \end{cases}$$

ξ est une v.a. tirée suivant la loi conditionnelle

$$p_{s_k}(\xi = \lambda) = p_{s_k}(\xi = \lambda | x_j = x_j^{(k)}, j \neq s_k)$$

Si l'ensemble des états est fini ou si X est gaussien on a :

Théorème : Quelle que soit la configuration initiale x_0

$$\lim_{k \rightarrow \infty} P(\mathbf{X}(k) = \mathbf{x} | \mathbf{X}(0) = \mathbf{x}(0)) = P(\mathbf{X} = \mathbf{x})$$

Remarques :

- Si X est un champ de Markov, la loi conditionnelle s'exprime simplement en fonction de la configuration sur un voisinage de s_k , ce qui rend cette simulation raisonnable en temps de calcul.
- Pour la segmentation $X|Y$ a la même taille de voisinage que X , mais
- Pour la restauration la taille de voisinage a posteriori dépend à la fois de la taille de voisinage a priori et du support de l'opérateur de flou.

Recuit simulé

$$P(\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}) = \frac{1}{Z} \exp[-U(\mathbf{x})] = \pi(\mathbf{x})$$

- **Physique statistique** : niveau d'intensité d'un pixel=état d'une particule élémentaire
- distribution de Gibbs : $\pi_T(\mathbf{x}) = \frac{1}{Z_T} \exp\left[-\frac{1}{T}U(\mathbf{x})\right]$
- chaque réalisation d'image représente l'état de l'ensemble des particules auquel on associe une énergie $U(\mathbf{x})$

$$T \rightarrow \infty \quad \pi_T(\mathbf{x}) \rightarrow \text{loi uniforme}$$

$$T = 1 \quad \pi_T(\mathbf{x}) \rightarrow \pi(\mathbf{x})$$

$$T \rightarrow 0 \quad \pi_T(\mathbf{x}) \rightarrow \text{loi concentrée sur l'estimée MAP}$$

Algorithme de recuit simulé :

(avec l'échantillonneur de Gibbs)

$$X_s(k+1) = \begin{cases} X_s^{(k+1)} & \text{si } s \neq s_k \\ \xi & \text{si } s = s_k \end{cases}$$

ξ est une v.a. tirée suivant la loi conditionnelle

$$p_{s_k, T_k}(X = \xi) = \frac{1}{Z_k} \exp\left[-\frac{1}{T_k}U(X_{s_k} = x | x_j, j \neq s_k)\right]$$

- **Théorème** : Si $\lim_{k \rightarrow \infty} T_k \ln k \geq K$ où K est une constante assez grande, alors quelle que soit la configuration initiale $X(0) = x_0$, $P(X(k) = x | X(0) = x_0)$ converge en loi vers une distribution uniforme sur les minima globaux de U

Algorithme d'ICM (Iterated Conditional Modes)

(Besag)

Il s'agit en effet d'une adaptation déterministe de l'algorithme précédent.

$$X_s^{(k+1)} = \begin{cases} X_s^{(k)} & \text{si } s \neq s_k \\ \arg \max_X \{p_i(X_{s_k} = x | x_j, j \neq s_k)\} & \text{si } s = s_k \end{cases}$$

Cet algorithme est rapide mais le résultat dépend du point initial et fournit un minimum local de U .

6 Application en restauration d'image

Modèle de Geman-Geman : modèle couplé pixels-contours

- $\mathcal{F} = \{F_{i,j}, (i,j) \in \mathbf{Z}_m\}$ champ intensité $U(\mathbf{f}|\mathbf{l})$
- $\mathcal{L} = \{L_{\alpha,\beta}, (\alpha,\beta) \in \mathcal{D}_m\}$ champ des lignes $U(\mathbf{l})$
- $\mathcal{X} = (\mathcal{F}, \mathcal{L})$ image $U(\mathbf{f}, \mathbf{l}) = U(\mathbf{f}|\mathbf{l}) + U(\mathbf{l})$
- $\mathcal{G} = \{G_{i,j}, (i,j) \in \mathbf{Z}_m\}$ image dégradée

$$G_{i,j} = \sum_{(k,l)} H_{i-k,j-l} F_{k,l} + N_{i,j} \quad \text{ou} \quad \mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}$$

Remarque : le champ des ligne n'est pas dégradée

- $\mathcal{N} = \{N_{i,j}, (i,j) \in \mathbf{Z}_m\}$ champ bruit d'intensité (gaussien, blanc)
- $\mathbf{H} = \{H_{i,j}, (i,j) \in \mathbf{Z}_m\}$ matrice de flou

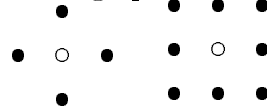
Estimateur du maximum a posteriori (MAP)

$$\mathcal{X} = (\mathcal{F}, \mathcal{L}), \quad \mathbf{x} = (\mathbf{f}, \mathbf{l})$$

$$\mathcal{Y} = (\mathcal{G}, \mathcal{L}), \quad \mathbf{y} = (\mathbf{g}, \mathbf{l})$$

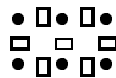
- \mathcal{X} est un champ de Markov d'énergie $U(\mathbf{f}, \mathbf{l})$:

Voisinage pixels



et

voisinage contours



- \mathcal{N} est gaussien, blanc et indépendant de \mathcal{F} ,
- \mathbf{H} est invariant par translation :

$$P(\mathbf{G} = \mathbf{g} | \mathbf{X} = \mathbf{x}) = K \exp \left[-\frac{1}{2\sigma_b^2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 \right]$$

- Règle de Bayes : $P(\mathbf{X} = \mathbf{x} | \mathbf{G} = \mathbf{g}) = A \exp[-U^p(\mathbf{f}, \mathbf{l})]$ avec

$$U^p(\mathbf{f}, \mathbf{l}) = \frac{1}{2\sigma_b^2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + U(\mathbf{f}, \mathbf{l})$$

- Loi a posteriori est aussi une mesure de Gibbs.
Il reste à déterminer le système de voisinage de cette loi.
- Le champ des lignes \mathcal{L} n'est pas affecté par la dégradation
- Le voisinage du champ d'intensité \mathcal{X} est élargi par la réponse impulsionnelle de la dégradation.

7 Problèmes et questions ouverts

Problème essentiel :

Coût de calcul très élevé du MAP

- potentiels convexes/non convexes \rightarrow optimisation locale/globale
- Optimisation stochastique ou déterministe

Solution proposée :

Relaxation déterministe : GNC

Non-convexité graduelle (GNC): Technique d'optimisation non-locale

Blake & Zisserman (87)

- Modèles avec un processus de ligne

$$p(\mathbf{f}, \mathbf{l} | \mathbf{g}) = p(\mathbf{g} | \mathbf{f}) p(\mathbf{f}, \mathbf{l}) \rightarrow U^p(\mathbf{f}, \mathbf{l}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + U(\mathbf{f}, \mathbf{l})$$

- Si le processus de ligne est non-interactif $\rightarrow \mathbf{l}$ peut être éliminé

$$U^p(\mathbf{f} | \mathbf{g} \mathbf{b}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda \sum_{s \in \mathcal{S}} \phi([\mathbf{D}\mathbf{f}]_s)$$

avec $\mathbf{D}\mathbf{f}$ un opérateur de différence et ϕ une fonction non-convexe

- $U^p(\mathbf{f} | \mathbf{g})$ est alors multimodale et la recherche de la solution MAP ne peut se faire par une méthode d'optimisation locale.
- Recuit simulé trop coûteux lorsque l'opérateur \mathbf{H} a un support large
- GNC est une technique d'optimisation non-locale dont le coût de calcul n'est pas dépendant du support de l'opérateur \mathbf{H} .

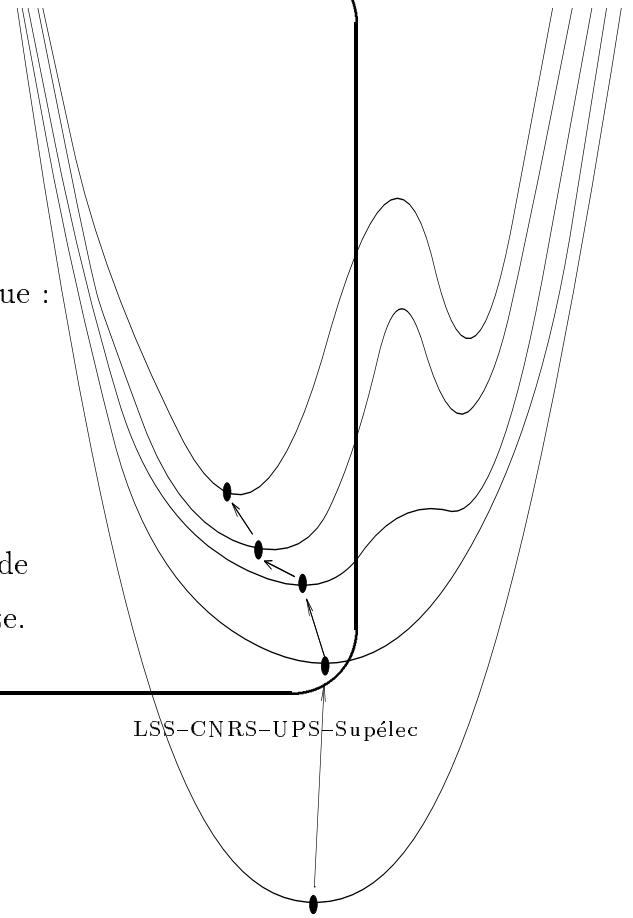
Principe du base du GNC:

Solution MAP :

$$\hat{\mathbf{f}} = \arg \max_{\mathbf{f}} \{p(\mathbf{f}|\mathbf{g})\} = \arg \min_{\mathbf{f}} \{J(\mathbf{f})\}$$

$J(\mathbf{f})$ multimodal

- Définir une suite de critères $J_r(\mathbf{f})$ telles que :
 - $J_{r_0}(\mathbf{f})$ soit convexe
 - $\forall \mathbf{f} \lim_{r \rightarrow r_\infty} J_r(\mathbf{f}) = J(\mathbf{f})$
- Minimiser $J_{r_0}(\mathbf{f})$ pour obtenir \mathbf{f}_0
- Pour une suite de $r = \{r_1, \dots, r_\infty\}$
 - minimiser localement $J_r(\mathbf{f})$ au voisinage de la solution obtenue à l'itération précédente.



Cas des problèmes linéaires

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

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

- [Blake et Zisserman] :
 - Segmentation d'image $\mathbf{A} = \mathbf{I} \rightarrow$ Problème bien posé
 - $J_{r_k}(\mathbf{f}) = \|\mathbf{g} - \mathbf{f}\|^2 + \Phi_{r_k}(\mathbf{f}), \quad \exists r_0$ 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

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

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

References

- [1] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 721–741, November 1984.
- [2] D. Geman, "Random fields and inverse problems in imaging," in *Lecture Notes in Mathematics* (P. L. Hennequin, ed.), vol. 1427, pp. 117–193, Springer-Verlag, 1990.
- [3] S. Geman and G. Reynolds, "Constrained restoration and recovery of discontinuities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-14, pp. 367–383, March 1992.
- [4] J. E. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of Royal Statistical Society B*, vol. 36, no. 2, pp. 192–236, 1974.
- [5] J. E. Besag, "On the statistical analysis of dirty pictures," *Journal of Royal Statistical Society B*, vol. 48, no. 3, pp. 259–302, 1986.

- [6] J. Besag, "Digital image processing : Towards Bayesian image analysis," *Journal of Applied Statistics*, vol. 16, no. 3, pp. 395–407, 1989.
- [7] J. Besag and P. Green, "Spatial statistics and Bayesian computation," *J. R. Statist. Soc. B*, vol. 55, pp. 25–37, 1993.
- [8] A. Blake and A. Zisserman, *Visual reconstruction*. Cambridge: The MIT Press, 1987.
- [9] M. Nikolova and A. Mohammad-Djafari, "Eddy current tomography using a binary Markov model," *To appear in Signal Processing*, vol. 49, pp. 000–000, May 1996.
- [10] J. Idier, A. Mohammad-Djafari and G. Demoment "Regularization Methods and Inverse Problems: An Information Theory Standpoint," *Submitted to 2nd Intern. Conf. on Inverse Problems in Eng.*, Le Croisic, June 1996.