

## Chapitre 10

# Modélisation markovienne

Dans ce chapitre nous donnons un aperçu général de la modélisation markovienne des signaux et des images. Les modèles markoviens et particulièrement les champs de Markov et les champs de Gibbs sont devenus des outils indispensables pour la traduction d'une information *a priori* plus élaborée que la douceur ou la positivité en une loi de probabilité. En effet, la modélisation markovienne permet de construire des modèles qui peuvent prendre en compte les discontinuités dans un signal ou une image. Cette outil prend particulièrement son importance en traitement d'image où l'on peut construire les modèles composites (bords, contours, intensité) pour les images et les utiliser lors de la segmentation, restauration ou reconstruction d'image.

## 10.1 Introduction et notations

### 10.1.1 Site, voisinage et cliques

Considérons un ensemble de sites  $\mathcal{S} = \{s_1, \dots, s_n\}$  et un système de voisinage  $\mathcal{V} = V_s, s \in \mathcal{S}$  sur cet ensemble.

- On dit que le site  $r$  est voisin du site  $s$  si

$$r \in V_s \iff s \in V_r, \quad s \notin V_s$$

$\{\mathcal{S}, \mathcal{V}\}$  est un graphe.

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

$$|c| = 1 \quad \text{ou si} \quad |c| > 1 \quad \text{et} \quad \forall s_1, s_2 \in c \longrightarrow s_1 \text{ et } s_2 \text{ sont voisins}$$

Notons par  $\mathcal{C} = \{c\}$  l'ensemble de cliques associées au graphe  $\{\mathcal{S}, \mathcal{V}\}$ .

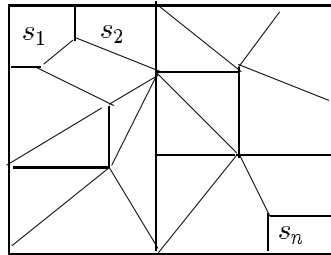
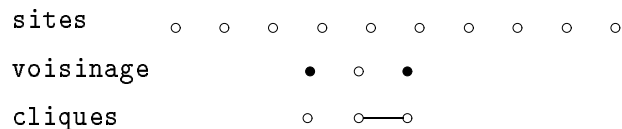


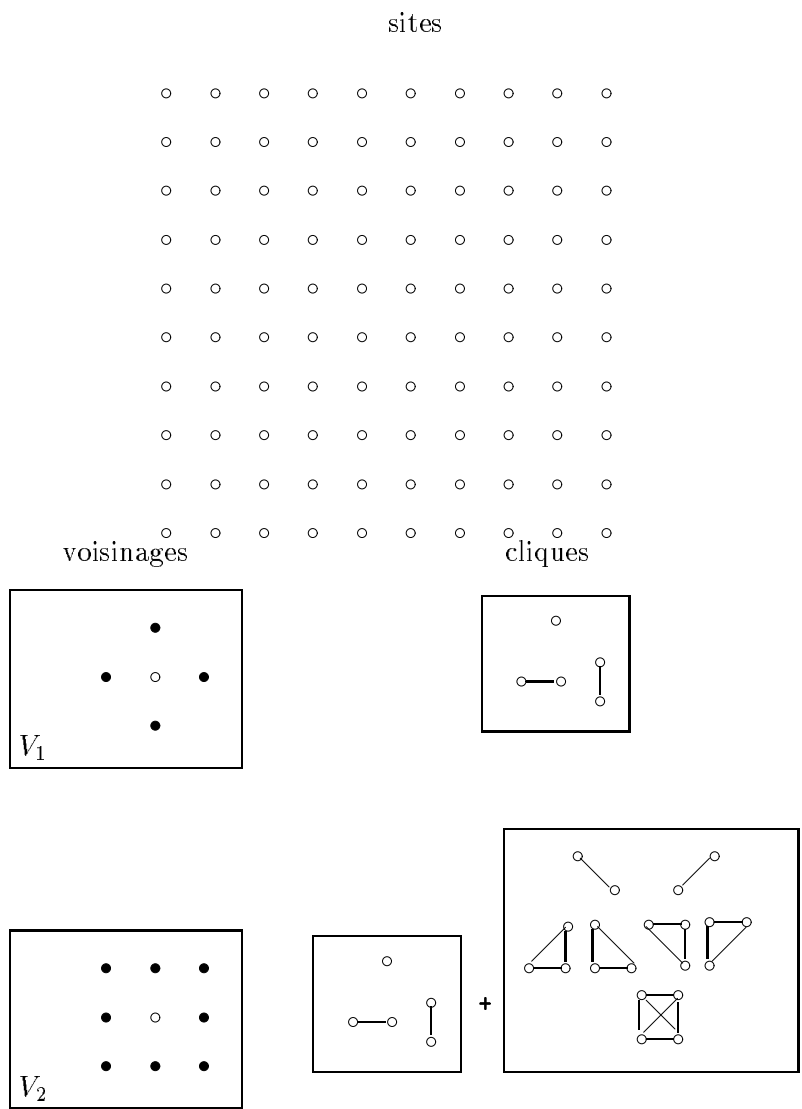
FIG. 10.1 - Sites, voisinage et champ.

Notons aussi que les sites ne sont pas obligés être tous identiques, mais en traitement du signal ces sites peuvent correspondre aux positions des échantillons. En traitement d'image on a l'habitude de parler de *pixels*. Ci-dessous deux exemples

**Exemple 1:**  $\mathcal{S} = \mathbf{Z}$



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



### 10.1.2 Champ aléatoire

Ensembles des variables aléatoires  $\mathcal{X} = \{X_s, s \in \mathcal{S}\}$  est appelé un champ aléatoire sur les sites  $\mathcal{S}$ . La variable aléatoire  $X_s$  prend sa valeur dans l'ensemble  $\mathcal{G} = \{g_1, \dots, g_k\}$  qui appelé l'espace des phases (niveau de gris, couleurs, etc.).

Une réalisation du champ aléatoire  $\mathcal{X}$  est notée :

$$\{\mathbf{X} = \mathbf{x}\} = \{X_{s_1} = x_{s_1}, \dots, X_{s_N} = x_{s_N}\}$$

et l'ensemble des réalisations (ou configurations) possibles est noté :

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

Lorsque l'espace des phases est à dimension finie, (cas discret) le champ est caractérisé par la loi de probabilité conjointe  $P(\mathbf{X} = \mathbf{x})$ , et lorsque l'espace des phases est à dimension infinie, (cas continu) le champ est caractérisé par sa densité de probabilité conjointe  $p(\mathbf{x})$ .

On peut aussi définir les lois de probabilités conditionnelles :

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

### 10.1.3 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
  - $P(\mathbf{X} = \mathbf{x}) > 0, \quad \forall \mathbf{x} \in \Omega, \quad \text{et}$
  - $\forall X_s \in \mathcal{G}, \forall s \in \mathcal{S}, \quad P(X_s = x_s | X_r = x_r, r \neq s) = P(X_s = x_s | X_r = x_r, r \in V_s)$
- 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?

## 10.2 Champs de Gibbs et équivalence Gibbs-Markov

### 10.2.1 Énergie 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

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

### 10.2.2 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  $\Omega$  :

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

- 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 :

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

### 10.2.3 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  $\mathcal{V}$ .

$$\begin{array}{ccc} \text{(Hammersley} & \& \text{Clifford)} \\ \text{Gibbs} & \iff & \text{Markov} \\ \text{Potentiels} & \iff & \text{caracteristiques locales} \end{array}$$

Gibbs  $\rightarrow$  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  $\rightarrow$  Gibbs : plus délicat. Potentiels canoniques

### 10.3 Exemples de modèles classiques

#### 10.3.1 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

$\beta_{rs}$  coefficients d'interaction entre les sites  $r$  et  $s$

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

#### 10.3.2 Auto-logistique

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

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

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.

#### 10.3.3 Modèle d'Ising

Auto-logistique avec voisinage d'ordre 1

Deux exemples spécifiques :

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

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

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

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

- 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

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

**10.3.4 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

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

## 10.4 Tentative de classification

- **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)

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

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

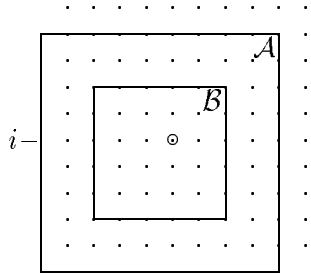
- Propriétés au sens large = Propriétés au deuxième ordre

- Représentation équivalente par un modèle AR :  $X_s = \sum_r a_r X_r + Z_s$

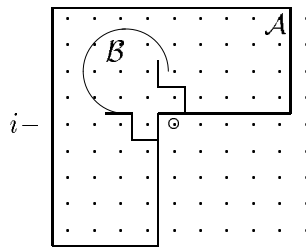


**Champ de markov au sens strict (SSM)**

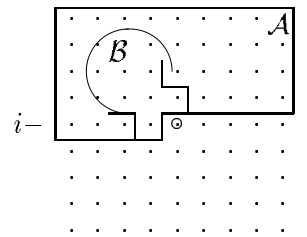
Soient  $\mathcal{B} \subseteq \mathcal{A} \subseteq \mathcal{S}$   $P(X_{ij} = x_{ij} | X_{\mathcal{B}} = x_{\mathcal{B}}) = P(X_{ij} = x_{ij} | X_{\mathcal{A}} = x_{\mathcal{A}})$



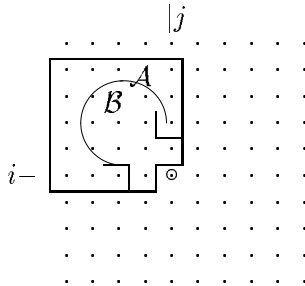
Bilatéral



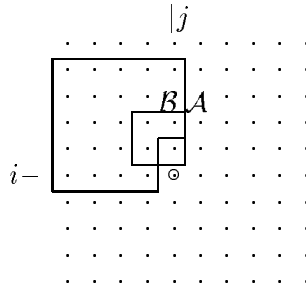
MMRF



Unilateral



Causal



Pickard

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

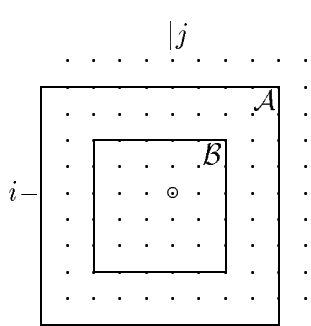
**Champ de markov au sens large (WSM)**

(Définition via la régression linéaire)

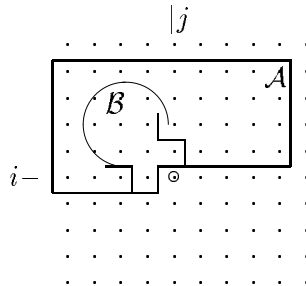
(Estimation linéaire en moyenne quadratique ELMQ)

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

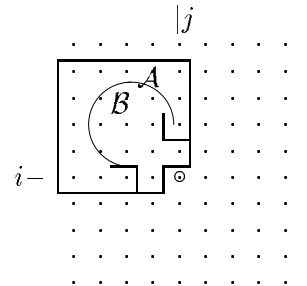
$$ELMQ \{X_{ij} | X_{\mathcal{B}} = x_{\mathcal{B}}\} = ELMQ \{X_{ij} | X_{\mathcal{A}} = x_{\mathcal{A}}\}$$



Bilatéral



Unilateral



Causal

**Champ de markov gaussien : sens large = sens stricte**

Un exemple classique utilisé en traitement d'image :

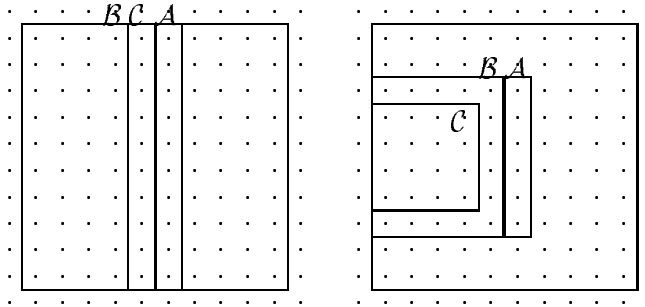
$$\mathcal{S} = \mathbf{Z}_{mn}^2 = \{(i, j) : 1 \leq i \leq m, 1 \leq j \leq n\}, \mathcal{X} = \{X_{i,j}, (i, j) \in \mathcal{S}\}$$

$$\mathcal{V}_1 = \left\{ \begin{array}{ccc} & \bullet & \\ \bullet & \circ & \bullet \\ & \bullet & \end{array} \right\} \quad \mathcal{V}_2 = \left\{ \begin{array}{ccc} \bullet & \bullet & \bullet \\ \bullet & \circ & \bullet \\ \bullet & \bullet & \bullet \end{array} \right\}$$

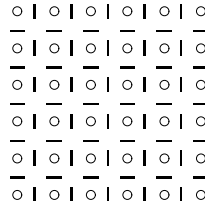
$$P(X_{ij} = x_{ij} | X_{kl} = x_{kl}, (k, l) \in \mathcal{S} - \{i, j\}) = P(X_{ij} = x_{ij} | X_{kl} = x_{kl}, (k, l) \in \mathcal{V}_{ij})$$

Quelques propriétés :

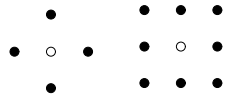
$$P(x_a, a \in \mathcal{A} | x_b, b \in \mathcal{B}) = P(x_a, a \in \mathcal{A} | x_b, b \in \mathcal{C})$$



### 10.5 Modèles couplés ou hiérarchiques

|   |  |                    |  |
|---|--|--------------------|--|
|   |  |                    |  |
| - | $\mathbf{Z}_m$                                   | 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 :

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

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

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

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

avec

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

## 10.6 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  $U(\mathbf{x})$  :

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

$$\begin{aligned} U(\mathbf{x}) = & \sum_j \alpha_j x_j \\ & + \beta_1 \sum_{(i,j)0} \sum_j x_i x_j + \beta_2 \sum_{(i,j)90} \sum_j x_i x_j + \beta_3 \sum_{(i,j)-45} \sum_j x_i x_j + \beta_4 \sum_{(i,j)45} \sum_j x_i x_j \\ & + \gamma_1 \sum_{(i,j,k)} x_i x_j x_k + \gamma_2 \sum_{(i,j)90} \sum_j x_i x_j x_k + \dots \\ & + \xi_1 \sum_{(i,j,k,l)} x_i x_j x_k x_l \end{aligned}$$

avec  $M$  paramètres  $\alpha_m, m = 1 = \dots, M$ , 4 paramètres  $\beta_1, \dots, \beta_4$ , 4 paramètres  $\gamma_1, \dots, \gamma_4$  et un paramètre  $\xi_1$ .

- $\mathcal{N}$  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 types de problèmes :****1. Classification ou segmentation supervisée :**

On connaît  $\theta = \{\{g_1, \dots, g_M\}, \{\beta_1, \dots, \beta_4\}, \{\gamma_1, \dots, \gamma_4\}, \xi_1\}$  et  $\mathbf{y} \rightarrow \hat{\mathbf{x}}$ .

**2. Segmentation non supervisée :**  $\mathbf{y} \rightarrow \hat{\mathbf{x}}, \hat{\theta}$ 

– Apprentissage supervisée :  $\mathbf{x}, \mathbf{y} \rightarrow \hat{\theta}$

– Apprentissage non supervisée :  $\mathbf{y} \rightarrow \hat{\theta}$

**Quelques 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) :

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

– PM (Posterior Mean) Moyenne *a posteriori* :

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

**Deux algorithmes stochastiques :**

– **Le recuit simulé ou relaxation stochastique**

permet d'estimer le MAP

– **L'échantillonneur de Gibbs**

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*.

**Quelques estimateurs possibles pour les problèmes 2**

– Apprentissage supervisée :  $\mathbf{x}, \mathbf{y} \rightarrow \hat{\theta}$  MAP, PM, MPM

– Apprentissage non supervisée :  $\mathbf{y} \rightarrow \hat{\theta}$  MV (EM)

– Segmentation non supervisée :  $\mathbf{y} \rightarrow \hat{\mathbf{x}}, \hat{\theta}$  MVG

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

–  $\mathbf{y}, \theta \rightarrow \hat{\mathbf{x}}_{\text{MAP}}$  Viterbi (Programmation dynamique)

–  $\mathbf{y} \rightarrow \hat{\theta}_{\text{MV}}$  EM, (Forward-Backward)

–  $\mathbf{y} \rightarrow \hat{\mathbf{x}}_{\text{MAP}}, \hat{\theta}_{\text{MV}}$  : Algorithme itératif, par ex. MVG:

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

## 10.7 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)**

$$\begin{aligned} \mathcal{X} &= (\mathcal{F}, \mathcal{L}), & \mathbf{x} &= (\mathbf{f}, \mathbf{l}) \\ \mathcal{Y} &= (\mathcal{G}, \mathcal{L}), & \mathbf{y} &= (\mathbf{g}, \mathbf{l}) \end{aligned}$$

–  $\mathcal{X}$  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.

## 10.8 Outils de simulation et d'optimisation

### 10.8.1 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.

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

$\xi$  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.

### 10.8.2 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  $U(\mathbf{x})$ 
  - $T \rightarrow \infty$      $\pi_T(\mathbf{x}) \rightarrow$  loi uniforme
  - $T = 1$          $\pi_T(\mathbf{x}) \rightarrow \pi(\mathbf{x})$
  - $T \rightarrow 0$          $\pi_T(\mathbf{x}) \rightarrow$  loi concentrée sur l'estimée MAP

#### Algorithme de recuit simulé :

(avec l'échantillonneur de Gibbs)

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

$\xi$  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  $U$

### 10.8.3 Algorithme d'ICM (Iterated Conditional Modes)

(BESAG)

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

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

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



## 10.9 Algorithmes de relaxations déterministes

Lors de calcul de la solution au sens du MAP, deux situations peuvent se présenter :

1. La loi *a posteriori* est unimodale.

Alors il n'y a qu'une seule solution et, en général, l'optimisation du critère MAP (l'énergie *a posteriori*) ne pose pas de grandes difficultés et on peut utiliser les algorithmes de descentes du type gradient ou autre pour effectuer cette optimisation.

2. La loi *a posteriori* est multimodale.

Dans ce cas le problème devient plus difficile et il faut aller à la recherche de techniques d'optimisation globale. L'algorithme du recuit simulé (RS) peut alors être utilisé, mais son coût de calcul devient très vite prohibitif, surtout lorsque la taille de voisinage de la loi *a posteriori* devient plus grande. En effet, la taille de voisinage de la loi *a posteriori* est directement liée à la taille de voisinage de la loi *a priori*, mais aussi et surtout aux caractéristiques de la l'opérateur liant la grandeur observable à la grandeur inconnue. Par exemple, lorsqu'il s'agit de la ségmentation ou la restauration d'image avec un noyau de taille très réduite, alors on peut utiliser la RS. Mais dans d'autres applications du genre la reconstruction d'image en tomographie X ou par ondes diffractées où le support de l'opérateur devient important, on ne peut plus envisager l'utilisation de tel algorithme. On peut alors faire recours aux techniques de relaxations déterministes qui cherchent à faire un compromis entre l'optimalité au sens de la recherche de l'optimum global et le coût de calcul. Parmi ces techniques on peut citer la technique de non convexité graduelle (Graduated Non Convexity GNC) que nous allons la décrire ici dans un cas particulier.

### 10.9.1 Principe de base du GNC

Soit  $J(\mathbf{f})$  un critère multimodal dont nous cherchons à trouver l'argument  $\mathbf{f}$  correspondant à son minimum global. La principale idée du GNC est de construire une suite de critères  $J_{c_k}(\mathbf{f})$  telles que :

- $J_{c_0}(\mathbf{f})$  soit convexe et
- $\forall \mathbf{f} \lim_{c \rightarrow c_\infty} J_c(\mathbf{f}) = J(\mathbf{f})$ .

Ensuite, on minimise  $J_{c_0}(\mathbf{f})$  pour obtenir  $\mathbf{f}_0$ , et finalement, pour une suite de  $\{c_1, \dots, c_\infty\}$ , on minimise localement  $J_{c_k}(\mathbf{f})$  au voisinage de la solution obtenue à l'itération précédente.

On espère ainsi que la suite des solutions  $\hat{\mathbf{f}}_k$  ainsi obtenues converge vers le minimum global du critère. Notons que, la convexité de  $J_{c_0}$  assure l'unicité de la solution, mais, il n'y a aucune garantie que l'algorithme converge vers le minimum global. Parmi les points importants concernant cet algorithme, il faut noter que son coût, contrairement aux algorithmes à base du recuits simulé, ne dépend pas de la taille de voisinage de la loi *a posteriori* dans le cas d'un estimateur au sens du MAP.

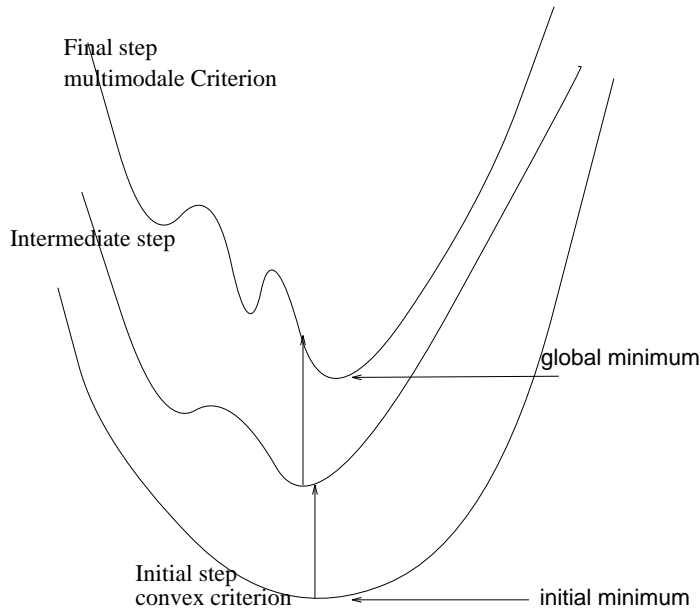


FIG. 10.2 - Principe de base de la technique du GNC.

Pour illustrer la mise en œuvre de cette technique nous allons considérer deux cas :

- le cas du calcul de la solution au sens du MAP d'un problème linéaire avec une loi *a priori* markovienne dont le terme d'énergie est nonconvexe ;
- le cas d'un problème non linéaire où la multimodalité du critère MAP vient plutôt de la non linéarité du problème que du choix de la loi *a priori*.

### Cas des problèmes linéaires :

Considérons le cas d'un problème inverse linéaire  $\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{b}$  en une dimension et supposons que le signal d'entrée est un signal continu par morceau et notons par  $\mathbf{f}$  le vecteur contenant les valeurs du signal et par  $\mathbf{l}$  le vecteur contenant les positions des discontinuités ( $l_j = 0$ , pas de discontinuité,  $l_j = 1$ , il y a une discontinuité). L'objectif de la reconstruction est alors d'estimer à la fois  $\mathbf{f}$  et  $\mathbf{l}$ .

- Modèles avec un processus de ligne binaire

$$p(\mathbf{f}, \mathbf{l} | \mathbf{g}) = p(\mathbf{g} | \mathbf{f}) p(\mathbf{f} | \mathbf{l}) p(\mathbf{l}) \longrightarrow$$

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

- Si le processus de ligne est non-interactif, ce qui signifie  $p(\mathbf{l}) = \prod_j p(l_j)$  ou encore  $U(\mathbf{l}) = \sum_j \phi_1(l_j)$ , alors on peut montrer que la solution

$$(\hat{\mathbf{l}}, \hat{\mathbf{f}}) = \arg \min_{(\mathbf{f}, \mathbf{l})} \{U^p(\mathbf{f}, \mathbf{l} | \mathbf{g})\}$$

peut être calculés en deux étapes :

- Calculer  $\hat{\mathbf{f}}$  en minimisant un critère ne dépendant que de  $\mathbf{f}$  :

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

avec  $D\mathbf{f}$  un opérateur de différences finies et  $\phi$  une fonction non-convexe et ;

- Dédire  $\hat{\mathbf{l}}$  par une opération directe à partir de  $\hat{\mathbf{f}}$ .

– **Exemple :**

$$U(\mathbf{f}|\mathbf{l}) = \lambda^2 \sum_j (f_j - f_{j-1})^2 (1 - l_j), \quad U(\mathbf{l}) = \mu \sum_j l_j,$$

Si

$$(\hat{\mathbf{l}}, \hat{\mathbf{f}}) = \arg \min_{(\mathbf{f}, \mathbf{l})} \{U^p(\mathbf{f}, \mathbf{l}|\mathbf{g})\}$$

alors

$$\hat{\mathbf{f}} = \arg \min_{(\mathbf{f})} \{U^p(\mathbf{f}|\mathbf{g})\} \quad \text{avec} \quad U^p(\mathbf{f}|\mathbf{g}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \sum_j \phi(x_j - x_{j-1})$$

avec

$$\phi(t) = \begin{cases} (\lambda t)^2 & \text{si } |t| \leq T \\ \alpha = (\lambda T)^2 & \text{si } |t| > T \end{cases}$$

et

$$\hat{l}_j = \begin{cases} 1 & \text{si } |x_j - x_{j-1}| \leq T \\ 0 & \text{si } |x_j - x_{j-1}| > T \end{cases}$$

- $U^p(\mathbf{f}|\mathbf{g})$  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. En effet le coût de calcul de cet algorithme croît d'une manière exponentielle avec la taille de son support.
- GNC est une technique d'optimisation non-locale dont le coût de calcul n'est pas dépendant du support de l'opérateur  $\mathbf{H}$ .

Non convexité vient du choix du terme  $U^p(\mathbf{f}|\mathbf{g})$  qui est une conséquence du choix de la loi *a priori*. C'est pourquoi pour mettre en œuvre l'algorithme du GNC dans ce cas, on peut envisager de définir une suite de critères  $J_c(\mathbf{f})$ :

$$J_c(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \sum_j \phi_c(x_j - x_{j-1})$$

avec

$$\phi_c(t) = \begin{cases} (\lambda t)^2 & \text{si } |t| < q_c \\ \alpha - \frac{1}{2}c(|t| - r_c)^2 & \text{si } q_c > |t| > r_c \\ \alpha = (\lambda T)^2 & \text{si } |t| > r_c \end{cases}$$

où

$$\begin{cases} q_c = T(1 + 2\lambda^2/c)^{-1/2} \\ r_c = T(1 + 2\lambda^2/c)^{1/2} \end{cases}$$

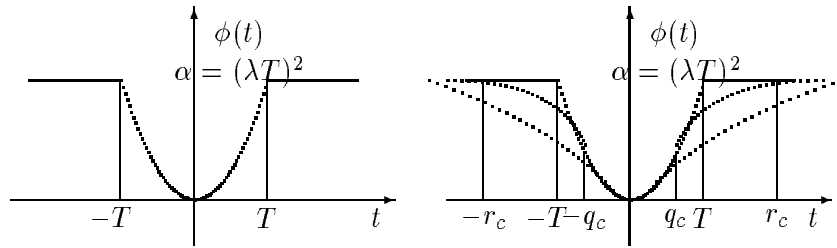
La principale difficulté est ensuite de trouver un  $c_0$  telle que  $J_{c_k}(\mathbf{f})$  soit convexe et ensuite choisir la suite  $\{c_0, c_1, \dots\}$ . La solution est alors calculée par

$$\hat{\mathbf{f}}_{c_k} = \arg \min_{\mathbf{f} \in V(\hat{\mathbf{f}}_{c_{k-1}})} \{J_{c_k}(\mathbf{f})\}$$

– [Blake et Zisserman]:

– Segmentation d'image  $\mathbf{A} = \mathbf{I} \rightarrow$  Problème bien posé

$J_{c_k}(\mathbf{f}) = \|\mathbf{g} - \mathbf{f}\|^2 + \Phi_{c_k}(\mathbf{f}), \quad \exists c_0$  tel que  $J_{c_0}$  soit convexe

FIG. 10.3 - *Fonction quadratique tronquée et sa relaxation*

- [Nikolova, Djafari, Idier]:
  - Extension pour les problème mal posés  $\mathbf{A}$
  - $\exists c_0$  tel que  $J_{c_0}$  soit convexe  $\rightarrow$

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

Double relaxations:  $a_k \mapsto 0$  et  $c_0 \mapsto c_\infty$

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

[7, 8, 9, 10, 11, 12, 13, 14, 15]

A COMPLETER

**Cas d'un problème inverse non linéaire :** Dans le cas d'un problème inverse non linéaire  $\mathbf{g} = \mathbf{H}(\mathbf{f}) + \mathbf{b}$  si on choisit de définir la solution comme le minimum d'un critère du type :

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \left\{ J(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}(\mathbf{f})\|^2 + \lambda \Omega(\mathbf{f}) \right\},$$

même si on choisit  $\Omega(\mathbf{f})$  convexe,  $J(\mathbf{f})$  est en général non convexe et multimodale à cause du terme de l'adéquation aux données  $\|\mathbf{g} - \mathbf{H}(\mathbf{f})\|^2$ .

Ici, pour illustrer l'usage de l'algorithme du GNC, il faut bien entendu préciser la nature de l'opérateur non linéaire  $\mathbf{H}(\mathbf{f})$ . Pour ceci, nous allons considérer le cas d'un problème de l'imagerie à ondes diffractées :

$$\begin{aligned} g(\mathbf{r}_i) &= \iint_D G_m(\mathbf{r}_i, \mathbf{r}') \phi(\mathbf{r}') f(\mathbf{r}') d\mathbf{r}', \quad \mathbf{r}_i \in S \\ \phi(\mathbf{r}) &= \phi_0(\mathbf{r}) + \iint_D G_o(\mathbf{r}, \mathbf{r}') \phi(\mathbf{r}') f(\mathbf{r}') d\mathbf{r}', \quad \mathbf{r} \in D \end{aligned}$$

La version discrétisée de ces deux équations :

$$\begin{aligned} \mathbf{g} &= \mathbf{G}_m \mathbf{F} \phi + \mathbf{b} \\ \phi &= \phi_0 + \mathbf{G}_o \mathbf{F} \phi \end{aligned}$$

Remplaçant la deuxième équation dans la première on a :

$$\mathbf{g} = \mathbf{H}(\mathbf{f}) + \mathbf{b} \quad \text{avec} \quad \mathbf{H}(\mathbf{f}) = \mathbf{G}_m \mathbf{F} (\mathbf{I} - \mathbf{G}_o \mathbf{F})^{-1} \phi_0$$

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}(\mathbf{x})\|^2 + \Omega(\mathbf{x}) \quad (10.1)$$

$$\mathbf{A}(\mathbf{x}) = \mathbf{G}_m \mathbf{X} (\mathbf{I} - \mathbf{G}_o \mathbf{X})^{-1} \phi_0 \quad (10.2)$$

Un choix possible pour créer une suite de critères relaxés est :

$$\mathbf{H}_{c_k}(\mathbf{f}) = \mathbf{G}_m \mathbf{F} (\mathbf{I} - c_k \mathbf{G}_o \mathbf{F})^{-1} \phi_0 \quad (10.3)$$

avec  $c_0 = 0$ , and  $\lim_{k \rightarrow \infty} c_k = 1$ .

Notons que pour  $c_0 = 0$  on a :

$$\mathbf{H}_0(\mathbf{f}) = \mathbf{G}_m \mathbf{F} \phi_0 \longrightarrow J_0(\mathbf{f}) = \|\mathbf{g} - \mathbf{G}_m \mathbf{F} \phi_0\|^2 + \Omega(\mathbf{f}).$$

et on retrouve l'approximation linéaire de BORN.

Pour  $c_k = 1$ , effectivement on retrouve :

$$\mathbf{H}_1(\mathbf{f}) = \mathbf{H}(\mathbf{f}) \longrightarrow J_1(\mathbf{f}) = J(\mathbf{f})$$

A COMPLETER

## 10.10 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

A COMPLETER