



Cours Doctoral EDITE
Champs de Markov et Estimation des paramètres

Marc Sigelle
Année 2008-2009

Dernière mise a jour : 24 mars 2009

Plan du cours

- rappels sur les champs de Markov
- introduction
- statistique exponentielle linéaire d'un paramètre
- estimation en données complètes
 - pseudovraisemblance - codage
 - gradient stochastique
- estimation en données incomplètes
 - gradient stochastique généralisé (ex: restauration)
 - méthode EM (ex: segmentation)
- conclusion

Rappels sur les champs de Markov

◦ champ de Markov :

- réseau de sites S + définition d'un système de voisinage \mathcal{V} sur $S \Leftrightarrow$ graphe
- ensemble de cliques \mathcal{C} et de potentiels associés
- propriété markovienne : la probabilité conditionnelle d'un site est uniquement fonction du voisinage de ce site

X est un champ de Markov relativement à un système de voisinage \mathcal{V} et $P(X = x) > 0 \Leftrightarrow X$ est un champ de Gibbs de potentiel associé à \mathcal{V}

$$P(x) = \frac{1}{Z} \exp(-U(x)) \quad U(x) = \sum_{c \in \mathcal{C}} U_c(x)$$

- accès aux probabilités conditionnelles locales

$$P(X_s = x_s / X^s = x^s) = \frac{\exp(-U_s(x_s / V_s))}{\sum_{x_s \in \Lambda} \exp(-U_s(x_s / V_s))} \quad U_s(x_s / V_s) = \sum_{c \in \mathcal{C}: s \in c} U_c(x)$$

Algorithmes de simulation et de minimisation

○ simulation d'un champ de Gibbs

tirage d'une configuration selon la loi de probabilité de Gibbs du champ

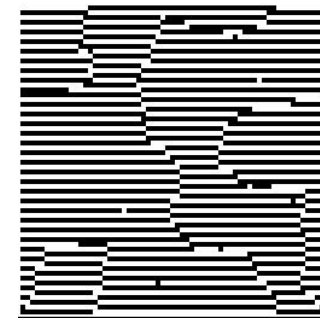
- algorithme de l'échantillonneur de Gibbs
- algorithme de Métropolis

○ recherche de la configuration la plus probable

= minimisation de l'énergie

- algorithme stochastique : recuit simulé
pour une distribution de Gibbs avec une température T , si $T \rightarrow 0$,
la distribution devient uniforme sur les minima de U et nulle ailleurs
(avec échantillonneur de Gibbs ou Métropolis)
converge *en théorie* vers un minimum global
- algorithme déterministe : ICM (Modes Conditionnels Itérés)
converge vers un minimum local qui dépend de l'initialisation

Estimation de paramètres: vous en avez déjà fait



- potentiels de cliques “les plus probables”



- réalisation d'un modèle d'Ising ou binaire

→ paramètre(s) associé(s) ? (hyperparamètre)

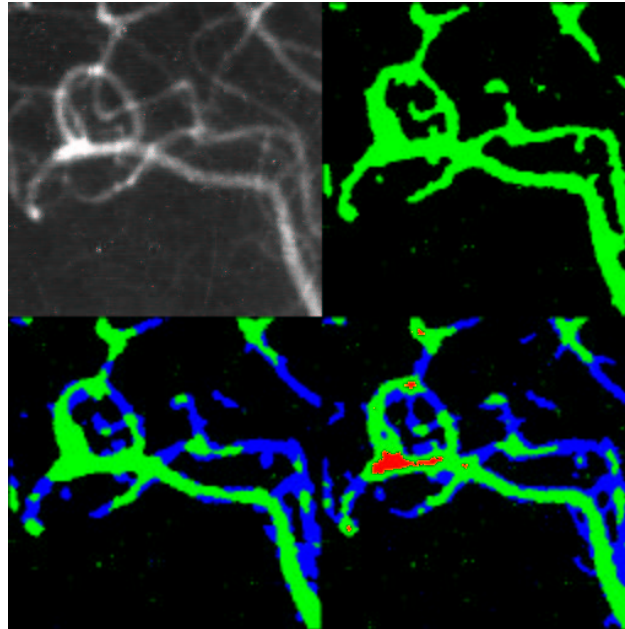
- modèle gaussien :
$$\Phi(x) = \left[\alpha \sum_{s \in S} (x_s - \mu_s)^2 + \beta \sum_{(r,s) \in \mathcal{C}} (x_r - x_s)^2 \right]$$

→ $\alpha, \beta (\mu_s)$?

but : estimation locale pour la segmentation de textures

Cas de la restauration et de la segmentation

- restauration : bruit connu \rightarrow paramètre de régularisation ?
- segmentation



paramètre de chaque classe : moyenne, variance + paramètre de régularisation

\rightarrow segmentation + estimation ?

- texture binaire, Ising et gaussien \rightarrow données complètes
- segmentation et restauration \rightarrow données incomplètes

Statistique exponentielle linéaire d'un paramètre

- **distribution exponentielle sur un espace fini Ω**

$$P_\theta(X = x) = P_\theta(x) = \frac{\exp -\theta \Phi(x)}{Z_\theta} \quad Z_\theta = \sum_{v \in \Omega} \exp -\theta \Phi(v)$$

- **espérance et variance d'une variable aléatoire T**

$$\mathbf{E}_\theta[T] = \frac{\sum_{v \in \Omega} T(v) \exp -\theta \Phi(v)}{\sum_{v \in \Omega} \exp -\theta \Phi(v)} \quad \text{var}_\theta(T) = \mathbf{E}_\theta[T^2] - (\mathbf{E}_\theta[T])^2$$

- **estimation de paramètres** \rightarrow variation de $P_\theta(X = x)$ avec θ

Statistique exponentielle linéaire (suite)

- o formules fondamentales

$$\begin{aligned}\frac{\partial \log P_\theta(x)}{\partial \theta} &= -\Phi(x) - \frac{\partial \log Z_\theta}{\partial \theta} = -\Phi(x) + \mathbf{E}_\theta[\Phi] \\ \frac{\partial \mathbf{E}_\theta[\Phi]}{\partial \theta} &= -\text{var}_\theta(\Phi) \leq 0\end{aligned}$$

- o autre forme fondamentale

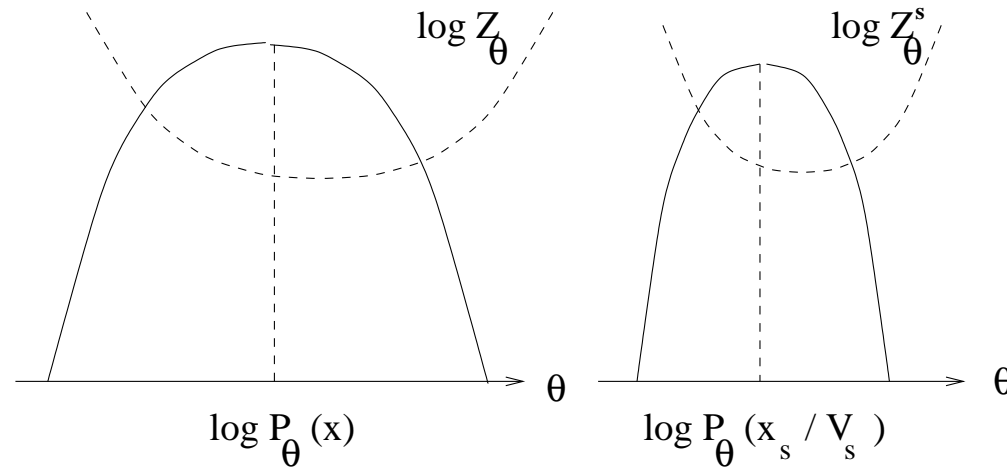
$$\frac{\partial \log Z_\theta}{\partial \theta} = -\mathbf{E}_\theta[\Phi] \quad \text{et} \quad \frac{\partial^2 \log Z_\theta}{\partial \theta^2} = \text{var}_\theta(\Phi) \geq 0$$

Démonstration

$$\begin{aligned}\log P_\theta(x) &= -\theta \Phi(x) - \log Z_\theta = -\theta \Phi(x) - \log \left(\sum_{v \in \Omega} \exp -\theta \Phi(v) \right) \\ \Rightarrow \frac{\partial \log P_\theta(x)}{\partial \theta} &= -\Phi(x) - \frac{\partial \log Z_\theta}{\partial \theta} = -\Phi(x) + \frac{\sum_{v \in \Omega} \Phi(v) \exp -\theta \Phi(v)}{\sum_{v \in \Omega} \exp -\theta \Phi(v)} \\ &= -\Phi(x) + \mathbf{E}_\theta[\Phi] \\ \Rightarrow \frac{\partial \mathbf{E}_\theta[\Phi]}{\partial \theta} &= \left(\frac{\sum_{v \in \Omega} -\Phi^2(v) \exp -\theta \Phi(v)}{\sum_{v \in \Omega} \exp -\theta \Phi(v)} \right) + \left(\frac{\sum_{v \in \Omega} \Phi(v) \exp -\theta \Phi(v)}{\sum_{v \in \Omega} \exp -\theta \Phi(v)} \right)^2 \\ &= -\text{var}_\theta(\Phi) \leq 0\end{aligned}$$

Conséquences

- $\log Z_\theta$ est une fonction convexe de θ
- $\log P_\theta(x) = -\theta \Phi(x) - \log Z_\theta$ est une fonction concave de $\theta \forall x \in \Omega$



- cas d'un champ de Markov-Gibbs : ceci est vrai pour

la loi globale :

$$P_\theta(x) = \frac{\exp -\theta \Phi(x)}{Z_\theta}$$

les lois conditionnelles locales (!) :

$$\Pr(X_s = x_s / V_s) = \frac{\exp -\theta \Phi(x_s / V_s)}{Z_\theta^s}$$

avec

$$\Phi(x_s / V_s) = \sum_{c \in \mathcal{C} \mid s \in c} \Phi_c(x)$$

Généralisation à plusieurs (deux) paramètres

- notations

$$P_{\theta,\lambda}(x) = \frac{\exp \{ -\theta \Phi(x) - \lambda \Psi(x) \}}{Z_{\theta,\lambda}} \quad \text{avec} \quad Z_{\theta,\lambda} = \sum_{v \in \Omega} \exp \{ -\theta \Phi(v) - \lambda \Psi(v) \}$$

- espérance et variance d'une variable aléatoire T

$$\mathbf{E}_{\theta,\lambda}[T] = \frac{\sum_{v \in \Omega} T(v) \exp \{ -\theta \Phi(v) - \lambda \Psi(v) \}}{\sum_{v \in \Omega} \exp \{ -\theta \Phi(v) - \lambda \Psi(v) \}} \quad \text{var}_{\theta,\lambda}(T) = \mathbf{E}_{\theta,\lambda}[T^2] - (\mathbf{E}_{\theta,\lambda}[T])^2$$

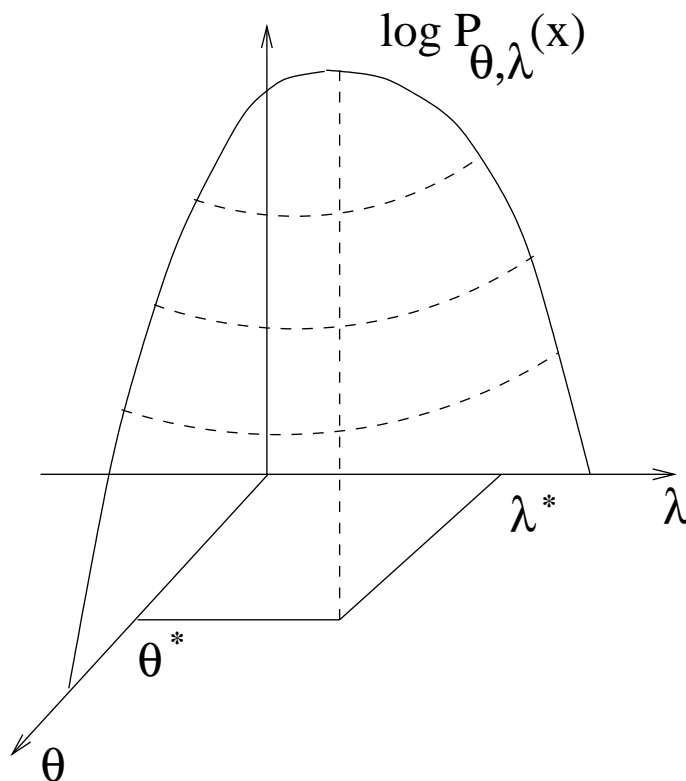
- résultats fondamentaux

$$\nabla \log Z_{\theta,\lambda} = - \begin{vmatrix} \mathbf{E}_{\theta,\lambda}[\Phi] \\ \mathbf{E}_{\theta,\lambda}[\Psi] \end{vmatrix} \quad \text{et} \quad \nabla \nabla \log Z_{\theta,\lambda} = \begin{bmatrix} \text{var}_{\theta,\lambda}(\Phi) & \text{cov}_{\theta,\lambda}(\Phi, \Psi) \\ \text{cov}_{\theta,\lambda}(\Phi, \Psi) & \text{var}_{\theta,\lambda}(\Psi) \end{bmatrix}$$

hessien de $\log Z_{\theta,\lambda}$ = matrice de variance-covariance - définie positive

Conséquences

- $\log Z_{\theta,\lambda}$ est une fonction convexe de (θ, λ)
- $\log P_{\theta,\lambda}(x)$ est une fonction concave de $(\theta, \lambda) \forall x \in \Omega$



- ceci est vrai pour les lois globale et conditionnelles locales !

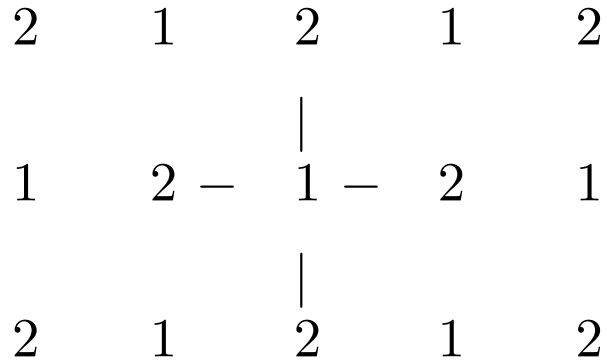
Données complètes - simplification : un seul paramètre θ

(exemple : Ising)



- donnée (configuration) observée x
- distribution de Gibbs de paramètre θ : $L(\theta) = P_\theta(x) = \frac{\exp -\theta \Phi(x)}{Z_\theta}$
- maximum de vraisemblance (MV) \nearrow
- deux stratégies :
 - A) approximer par les probabilités conditionnelles locales
→ codage, pseudo-vraisemblance
 - B) algorithmes itératifs à partir de la vraisemblance exacte
→ gradient stochastique

A) Codage (4-connexité)



$$\begin{aligned}
 P_\theta(x) \approx P_\theta(\text{Cod}_1 / \text{Cod}_2) &= \prod_{s \in \text{Cod}_1} P_\theta(X_s = x_s / V_s) \\
 &= \prod_{s \in \text{Cod}_1} \frac{\exp -\theta \Phi(x_s / V_s)}{Z_s(\theta)}
 \end{aligned}$$

$$Z_s(\theta) = \sum_{\xi \in E} \exp -\theta \Phi(\xi / V_s) \quad - \text{ local - calculable}$$

Pseudovraisemblance

$$P_{\theta}(x) \approx PL_{\theta}(x) = \prod_{s \in S} P_{\theta}(X_s = x_s / V_s)$$

- produit de probabilités conditionnelles locales paramétrées par θ
- théorème :

$\log P_{\theta}(\mathbf{Cod}_1 / \mathbf{Cod}_2)$ et $\log PL_{\theta}(x)$ sont des fonctions concaves de θ

- sommes de fonctions concaves : $\log R_{\theta}(x) = \sum_{s \in A} \log P_{\theta}(X_s = x_s / V_s)$

$\Rightarrow \exists$ une valeur de paramètre optimale $\hat{\theta}$

- méthodes de recherche d'un maximum : gradient (conjugué) etc ..
 - méthode consistante et convergente
 - biais expérimentaux dans les valeurs obtenues

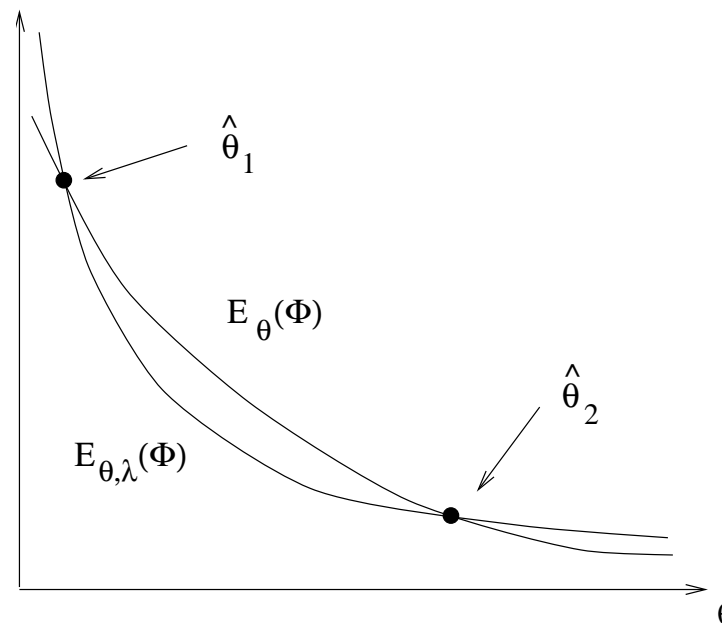
B) Gradient stochastique : principe

- **distribution de Gibbs** : $P_\theta(x) = \frac{\exp -\theta \Phi(x)}{Z_\theta}$

- **maximum de vraisemblance** : $\frac{\partial \log P_\theta(x)}{\partial \theta} = -\Phi(x) - \frac{\partial \log Z_\theta}{\partial \theta} = 0$

⇒ $\mathbf{E}_{\hat{\theta}}[\Phi] = \Phi(x)$ équation stochastique

- **une seule solution**



Gradient : schéma itératif exact

$$\theta_{n+1} = \theta_n - \frac{\mathbf{E}_{\theta_n}[\Phi] - \Phi(x)}{\left(\frac{\partial \mathbf{E}_{\theta}[\Phi]}{\partial \theta}\right)_{\theta_n}} = \theta_n + \frac{\mathbf{E}_{\theta_n}[\Phi] - \Phi(x)}{\text{var}_{\theta_n}(\Phi)} \rightarrow \mathbf{E}_{\theta_n}[\Phi], \text{var}_{\theta_n}(\Phi) ?$$

o le gradient stochastique [Younes(1988)]

$$\left\{ \begin{array}{l} \theta_0, x^{(0)} \text{ arbitraires} \\ \theta_{n+1} = \theta_n + \frac{\Phi(x^{(n)}) - \Phi(x)}{(n+1)V} \text{ pour } n \geq 1 \end{array} \right. \quad V \simeq \text{var}_{\theta_n}(\Phi)$$

$$\begin{array}{ccccccc} x^{(0)} & \rightarrow & x^{(1)} \dots & \rightarrow & x^{(n-1)} & \rightarrow & x^{(n)} \dots \\ & & \text{Gibbs}(\theta_0) & & & & \text{Gibbs}(\theta_n) \end{array}$$

Données incomplètes : cas de la restauration

- donnée manquante x → observation y cas gaussien (λ)

$$\begin{aligned}
 X &\rightarrow Y \\
 \Pr(Y = y / X = x, \Lambda = \lambda) &= \frac{\exp - \lambda U(y / x)}{Z_\lambda} = \frac{\exp - \lambda \|y - x\|^2}{Z_\lambda} \\
 \text{avec } Z_\lambda &= \left(\sqrt{\frac{\pi}{2\lambda}} \right)^{|S|}
 \end{aligned}$$

- a priori sur l'image : markovien $\Pr(X = x / \Theta = \theta) = \frac{\exp - \theta \Phi(x)}{Z_\theta}$

- probabilité a posteriori de x

$$\begin{aligned}
 \Pr(X = x / Y = y) &= \frac{\Pr(Y = y / X = x, \Lambda = \lambda) \cdot \Pr(X = x / \Theta = \theta)}{\Pr(Y = y)} \\
 &= \frac{\exp \{ -\lambda U(y / x) - \theta \Phi(x) \}}{Z_{\theta, \lambda}} \\
 \text{avec } Z_{\theta, \lambda} &= \sum_{x \in \Omega} \exp \{ -\lambda U(y / x) - \theta \Phi(x) \}
 \end{aligned}$$

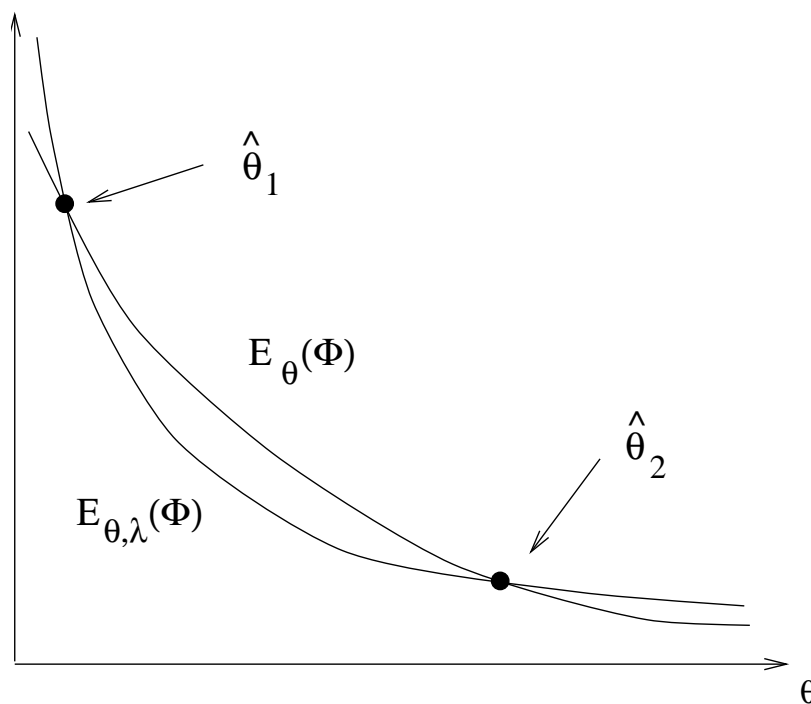
Vraisemblance du paramètre de régularisation θ

$$\begin{aligned} L(\theta) &= \Pr(Y = y / \Lambda = \lambda, \Theta = \theta) \\ &= \frac{\Pr(X = x, Y = y / \Lambda = \lambda, \Theta = \theta)}{\Pr(X = x / Y = y, \Lambda = \lambda, \Theta = \theta)} \\ &= \frac{\Pr(Y = y / X = x, \Lambda = \lambda) \Pr(X = x, \Theta = \theta)}{\Pr(X = x / Y = y, \Lambda = \lambda, \Theta = \theta)} \\ &= \frac{\exp - \lambda U(y / x) - \theta \Phi(x)}{Z_\lambda \cdot Z_\theta} / \frac{\exp - \lambda U(y / x) - \theta \Phi(x)}{Z_{\theta, \lambda}} \\ &= \frac{Z_{\theta, \lambda}}{Z_\lambda \cdot Z_\theta} \end{aligned}$$

Données incomplètes : restauration (suite)

- **maximum de la vraisemblance** $L(\theta) = \frac{Z_{\theta,\lambda}}{Z_\lambda \cdot Z_\theta} \Rightarrow \frac{\partial \log Z_{\theta,\lambda}}{\partial \theta} = \frac{\partial \log Z_\theta}{\partial \theta}$

$$\Rightarrow \begin{array}{|c|} \hline \mathbf{E}_{\hat{\theta},\lambda}[\Phi] = \mathbf{E}_{\hat{\theta}}[\Phi] \\ \hline \text{a posteriori} \quad \quad \quad \text{a priori} \\ \hline \end{array}$$



Données incomplètes : cas de la segmentation

○ loi de l'observation : attache aux données

- m classes - cas gaussien: $\lambda = \{\mu_i, \sigma_i^2\}_{i=1,m}$ moyennes , écarts-type

$$\begin{aligned} \Pr(Y = y / X = x, \Lambda = \lambda) &= \prod_{s \in S} \frac{1}{\sqrt{2\pi} \sigma_{x_s}} \exp - \frac{1}{2 \sigma_{x_s}^2} (y_s - \mu_{x_s})^2 \\ &= \left(\frac{1}{\sqrt{2\pi}} \right)^{|S|} \exp - \left\{ \underbrace{\sum_{s \in S} \frac{1}{2 \sigma_{x_s}^2} (y_s - \mu_{x_s})^2 + \log \sigma_{x_s}}_{U_\lambda(y / x) \text{ déjà normalisé} / \lambda \Rightarrow Z_\lambda = \text{Cte} !!} \right\} \end{aligned}$$

○ vraisemblance des paramètres

$$U_\lambda(y / x) \text{ déjà normalisé} / \lambda \Rightarrow L(\theta, \lambda) = \frac{Z_{\theta, \lambda}}{Z_\lambda \cdot Z_\theta} \propto \frac{Z_{\theta, \lambda}}{Z_\theta} !!$$

Cas de la segmentation (suite)

- estimation du paramètre de régularisation θ par MV

$$\mathbf{E}_{\hat{\theta}, \lambda}[\Phi] = \mathbf{E}_{\hat{\theta}}[\Phi]$$

- estimation des hyperparamètres λ (par MV)

$$\forall i \in [1..m] , \left\{ \begin{array}{l} \mu_i = \frac{\sum_{s \in S} y_s P_{\theta, \lambda}(X_s = i)}{\sum_{s \in S} P_{\theta, \lambda}(X_s = i)} \\ \sigma_i^2 = \frac{\sum_{s \in S} (y_s - \mu_i)^2 P_{\theta, \lambda}(X_s = i)}{\sum_{s \in S} P_{\theta, \lambda}(X_s = i)} \end{array} \right.$$

probabilités a posteriori

Estimation des hyperparamètres λ : démonstration

$$\begin{aligned}
 Z_{\theta,\lambda} &= \sum_{x \in \Omega} \exp - \left[\sum_{s \in S} \left(\frac{1}{2 \sigma_{x_s}^2} (y_s - \mu_{x_s})^2 + \log \sigma_{x_s} \right) + \theta \Phi(x) \right] \\
 &= \sum_{x \in \Omega} \exp - \left[\sum_{i=1}^m \sum_{s \in S} \left(\frac{1}{2 \sigma_i^2} (y_s - \mu_i)^2 + \log \sigma_i \right) \mathbf{I}_{x_s=i} + \theta \Phi(x) \right]
 \end{aligned}$$

o moyennes

$$\frac{\partial \log Z_{\theta,\lambda}}{\partial \mu_i} = \frac{1}{2 \sigma_i^2} \mathbf{E}_{\theta,\lambda} \left[\sum_{s \in S} (\mu_i - y_s) \mathbf{I}_{x_s=i} \right] \Rightarrow \sum_{s \in S} (\mu_i - y_s) P_{\theta,\lambda}(X_s = i) = 0$$

o variances

$$\begin{aligned}
 \frac{\partial \log Z_{\theta,\lambda}}{\partial \sigma_i} &= \frac{1}{\sigma_i} \mathbf{E}_{\theta,\lambda} \left[\sum_{s \in S} \left(\frac{-(y_s - \mu_i)^2}{\sigma_i^2} + 1 \right) \mathbf{I}_{x_s=i} \right] \\
 &\Rightarrow \sum_{s \in S} \left(\frac{-(y_s - \mu_i)^2}{\sigma_i^2} + 1 \right) P_{\theta,\lambda}(X_s = i) = 0 !
 \end{aligned}$$

L'algorithme EM (Expectation Maximization)

$$(\theta_{n+1}, \lambda_{n+1}) = \arg \max_{\theta, \lambda} \underbrace{\mathbf{E}_{\theta_n, \lambda_n} [\log \Pr(X = x, Y = y / \Lambda = \lambda, \Theta = \theta)]}_{\text{espérance a posteriori courante}}$$

- **théorème (Baum *et al.* 1970) :** $L(\theta_{n+1}, \lambda_{n+1}) \geq L(\theta_n, \lambda_n)$
- **loi jointe** → “séparabilité” des optimisations sur chaque paramètre

$$\begin{cases} \theta_{n+1} &= \arg \max_{\theta} \mathbf{E}_{\theta_n, \lambda_n} [\log \Pr(X = x / \Theta = \theta)] \\ \lambda_{n+1} &= \arg \max_{\lambda} \mathbf{E}_{\theta_n, \lambda_n} [\log \Pr(Y = y / X = x, \Lambda = \lambda)] \end{cases}$$

- **paramètre de régularisation**

$$v(\theta) = \mathbf{E}_{\theta_n, \lambda_n} [\log \Pr(X = x / \Theta = \theta)] = \mathbf{E}_{\theta_n, \lambda_n} [-\theta \Phi - \log Z_{\theta}]$$

$$\frac{\partial v(\theta)}{\partial \theta} = -\mathbf{E}_{\theta_n, \lambda_n} [\Phi] - \frac{\partial \log Z_{\theta}}{\partial \theta} = 0 !$$

$$\mathbf{E}_{\theta_{n+1}} [\Phi] = \mathbf{E}_{\theta_n, \lambda_n} [\Phi]$$

Méthode EM : paramètres de régions

$$\forall i \in [1..m] \left\{ \begin{array}{l} \mu_i^{(n+1)} = \frac{\sum_{s \in S} y_s P_{\theta_n, \lambda_n}(X_s = i)}{\sum_{s \in S} P_{\theta_n, \lambda_n}(X_s = i)} \\ (\sigma_i^{(n+1)})^2 = \frac{\sum_{s \in S} (y_s - \mu_i^{(n+1)})^2 P_{\theta_n, \lambda_n}(X_s = i)}{\sum_{s \in S} P_{\theta_n, \lambda_n}(X_s = i)} \end{array} \right.$$

somme sur tous les pixels (pondérés)

à convergence on retrouve les équations issues du MV !

$$\begin{aligned} w(\lambda) &= \mathbf{E}_{\theta_n, \lambda_n} [\log \Pr(Y = y / X = x, \Lambda = \lambda)] \\ &= \mathbf{E}_{\theta_n, \lambda_n} \left[- \sum_{i=1}^m \sum_{s \in S} \left(\frac{1}{2 \sigma_i^2} (y_s - \mu_i)^2 + \log \sigma_i \right) \mathbf{I}_{x_s=i} \right] \end{aligned}$$

$$\frac{\partial w(\lambda)}{\partial \mu_i} \propto \mathbf{E}_{\theta_n, \lambda_n} \left[\sum_{s \in S} (y_s - \mu_i) \mathbf{I}_{x_s=i} \right] = \sum_{s \in S} (y_s - \mu_i) P_{\theta_n, \lambda_n}(X_s = i) = 0 !$$

Méthode EM : implémentation pratique

- distribution a posteriori courante $\text{Gibbs}(\theta_n, \lambda_n)$
- plusieurs échantillons nécessaires : X_p , $p = 1..N$
- → paramètre de régularisation

- Estimation : $\mathbf{E}_{\theta_n, \lambda_n} [\Phi] \approx \frac{1}{N} \sum_{p=1}^N \Phi(X_p)$

- Maximisation : trouver θ_{n+1} tel que $\mathbf{E}_{\theta_{n+1}} [\Phi] = \mathbf{E}_{\theta_n, \lambda_n} [\Phi]$

estimateur MV pour la donnée complète : $\mathbf{E}_{\theta_n, \lambda_n} [\Phi]$!!

- $\Rightarrow \left\{ \begin{array}{l} \text{pseudo-vraisemblance, codage [Chalmond(1989)]} \\ \text{gradient stochastique} \end{array} \right.$

Méthode EM : implémentation pratique (suite)

○ → paramètres de régions [Chalmond(1989)]

- **Estimation** : labels obtenus en chaque site par Gibbs(θ_n, λ_n)

$$P_{\theta_n, \lambda_n}(X_s = i) \approx \frac{N_s(i)}{N} \leftarrow X_p, p = 1..N$$

- **Maximisation** : on obtient

$$\forall i \in [1..m] \left\{ \begin{array}{l} \mu_i^{(n+1)} = \frac{\sum_{s \in S} y_s N_s(i)}{\sum_{s \in S} N_s(i)} \\ (\sigma_i^{(n+1)})^2 = \frac{\sum_{s \in S} (y_s - \mu_i^{(n+1)})^2 N_s(i)}{\sum_{s \in S} N_s(i)} \end{array} \right.$$

○ à convergence : échantillons $\{\tilde{x}^{(n)}\}_{n > N_0} \Rightarrow$ estimateur MPM

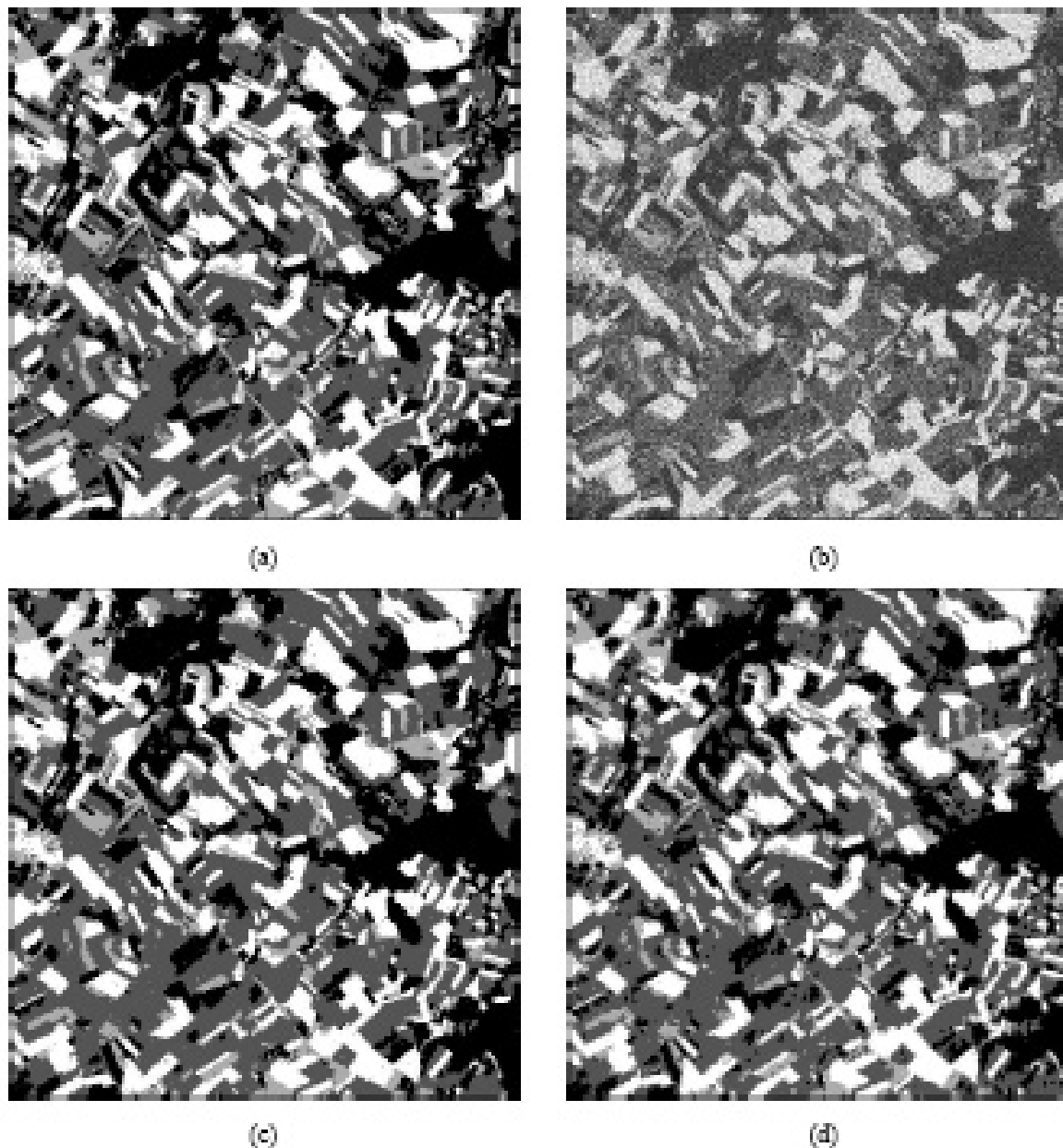


Fig. 5. Classification of a 512×512 simulated 3-look SAR image into four classes: (a) Ideal class image, (b) Speckled amplitude image. Results obtained (c) with the Markov random field method and (d) with the Markov chain method. The fraction of correctly classified pixels is (c) 87.0 % and (d) 84.2 %.

Conclusion sur l'estimation des paramètres

- données complètes : algorithmes rapides
- données incomplètes : algorithmes lourds mais avec restauration ou segmentation “simultanée”