Introduction to Markov Random Fields

Florence Tupin
Télécom Paris

Overview of the session - first part

- Introduction
- Neighborhoods and cliques
- Definition of a Markov Random Field
- Hammersley-Clifford theorem and conditional probability
- Sampling MRF (Gibbs, Metropolis)
- Optimization

Introduction

- **Historical background**
  - Probability theory for physical phenomena (crystal structure)
  - Geman and Geman article (84)
- **Main idea of MRF**

contextual relations are necessary to model images

a local neighborhood is enough for natural images

A prior for natural images : local context

Introduction

- **Low-level applications**
  - Restoration
  - Segmentation
  - Edge detection
  - Compression
- **Higher-level applications**
  - Object recognition
  - Graph matching

Markov random field

◦ **Probabilitic model**

$S = \{s\} \subset \mathbf{Z}^d$ set of sites (discrete and finite)

$x_s \in E$ space of the “gray-levels”

$(E = \{0..255\}\ \{0..q-1\}$ (labels) $\mathbf{R})$

X_s random variable associated to s

$X = \{X_s\}_{s \in S}$ random field

$x = \{x_s\}_{s \in S} = \{x_s\} \cup x^s$ configuration (image)

$\Omega = E^{|S|}$ space of configurations

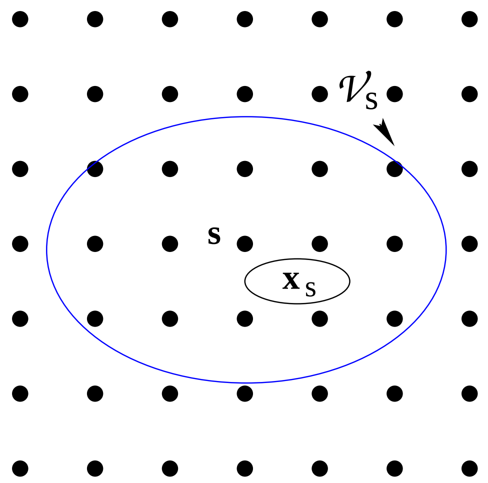
◦ **probabilities**

$P(X_s = x_s)$ local probability

$P(X = x) = P(X_1 = x_1, X_2 = x_2 \ldots X_s = x_s \ldots)$ joint probability

$\Pr(X_s = x_s \ / \ X_t = x_t, t \neq s)$conditional probability

Spatial context in natural images



s : site

$\mathcal{V}_s$: (spatial) neighborhood of s

- **homogeneous regions**

  $x_s \leftrightarrow$ radiometries of neighborhing pixels — mean

- **textured regions**

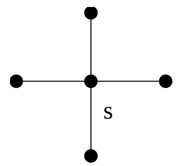
  $x_s \leftrightarrow$ radiometries of neighborhing (!!) pixels — local pattern function

global image	$\Leftrightarrow$	local neighborhood
global probability	$\Leftrightarrow$	local [conditional] probability

Topology for Markov Random Fields

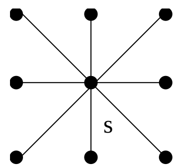
- **neighborhood system - definition (mutual relationship)**

neighborhood of site s :	$\mathcal{V}_s$
properties :	$s \notin \mathcal{V}_s \quad s \in \mathcal{V}_r \Leftrightarrow r \in \mathcal{V}_s$



$\mathcal{V} = \{\mathcal{V}_s\}_{s \in S}$ neighborhing system

$x \rightarrow V_s = \{x_r\}_{r \in \mathcal{V}_s}$ local configuration of the neighborhood



- **cliques**

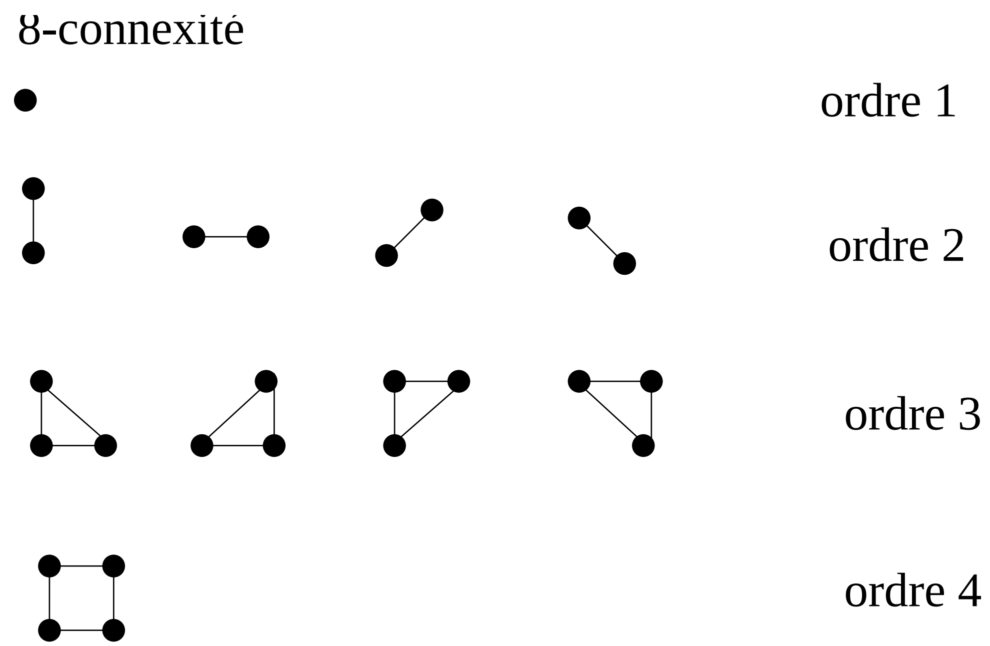
$c \subset S$ is a clique / $\mathcal{V}$ iff :
- card $(c) = 1$ (single-site)
- card $(c) \geq 2$ and $\forall\ r \neq s \in c \Rightarrow$ r, s neighbors

- **notations** $c = (r, s, t, \ldots);$ $\mathcal{C} = \{c\}$

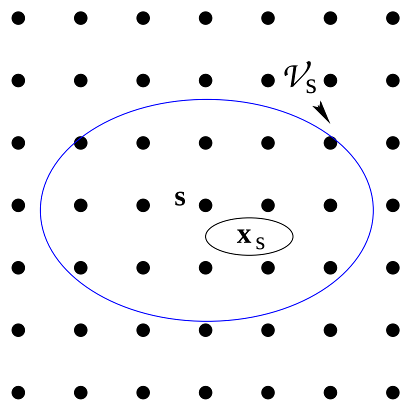
Topology for Markov Random Fields (2)

- **4-connexity**

- **8-connexity**



MRF : definition



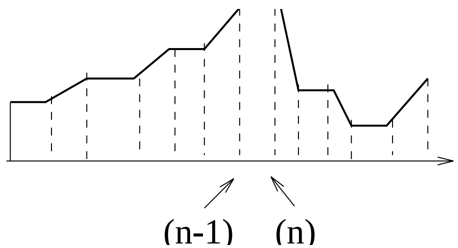
s : site

$\mathcal{V}_s$: voisinage (spatial) de s

$$\begin{aligned} \Pr(X_s = x_s \ / \ \{X_r = x_r\} \ , \ r \neq s) &= \Pr(X_s = x_s \ / \ \{X_r = x_r\} \ , \ r \in \mathcal{V}_s) \\ &= \Pr(X_s = x_s \ / \ V_s) \end{aligned}$$

Global		Local
Probability	$\leftrightarrow$	Probability

- **MRF = 2D extension of Markov chain**



Hammersley-Clifford theorem :
$P(X = x) > 0 \;\; \forall x \in \Omega$ is a MRF iff

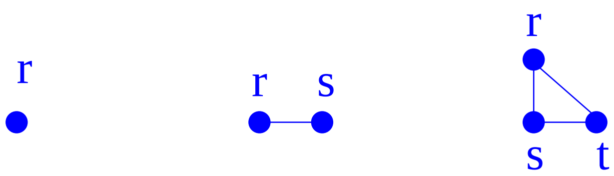
$$P(X = x) = \frac{\exp - U(x)}{Z} \qquad \text{Gibbs distribution}$$

$$U(x) = \sum_{c \in \mathcal{C}} V_c(x) \qquad \text{global energy}$$

$$V_c(x) = V_c(x_s, \;\; s \in c) \qquad \text{clique potential}$$

$$Z = \sum_{y \in \Omega} \exp - U(y) \qquad \text{partition function}$$

- **Example : cliques**



$$U(x) = A \sum_{s \in S} f(x_s) + B \sum_{(r,s)} g(x_r, x_s) + C \sum_{(r,s,t)} h(x_r, x_s, x_t)$$

possible non-stationarity : $A \rightarrow A_s \;,\;\; B \rightarrow B_{rs} \;\; \ldots$

- **important : low energy $U(x)$ $\Leftrightarrow$ high probability $P(X = x)$**

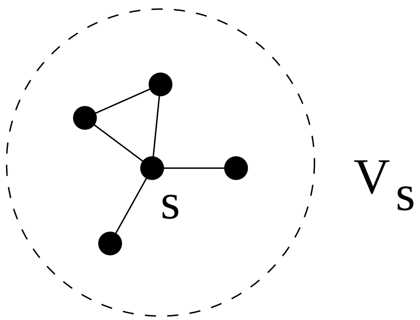
Conditional probability

$$\Pr(X_s = x_s \;/\; V_s) = \frac{1}{Z^s}\ \exp -U_s(x_s, V_s)$$

with :

$$U_s(x_s, V_s) = \sum_{c \subset C, s \in c} V_c(x_s, V_s) \quad \text{local conditional energy}$$

$$Z^s = \sum_{\xi \in E} \exp\ - U_s(\xi, V_s) \qquad \text{local partition function}$$



- ⇒ **local form of the Gibbs distribution**

Local conditional probability (2)

- **demonstration**

$$\Pr(X_s = x_s \ / \ X^s = x^s) = \frac{\Pr(X_s = x_s, \ X^s = x^s)}{\Pr(X^s = x^s)}$$

$$= \frac{\Pr(X_s = x_s, \ X^s = x^s)}{\sum_{\xi \in E} \Pr(X_s = \xi, \ X^s = x^s)} = \frac{\Pr(X = x)}{\sum_{\xi \in E} \Pr(X = x')}$$

- $\rightarrow$ **Let** $U(x) = U(x_s \ / \ V_s) + \sum_{c \in C, s \notin c} V_c(x)$

$\nwarrow \mathcal{W}$

$$\Pr(X_s = x_s \ / \ X^s = x^s) = \Pr(X_s = x_s \ / \ V_s) = \frac{\exp - U(x_s \ / \ V_s)}{\sum_{\xi \in E} \exp - U(\xi \ / \ V_s)} \qquad \swarrow \quad Z^s$$

Local conditional probability : example

binary field ($E = \{0, 1\}$)

- **Neighborhood** 4-connexity
- **Clique potentials**

cliques of order 2

$V_{c=(s,t)}(x_s, x_t) = \beta 1_{x_s \neq x_t}$ (0 if $x_s = x_t$, β if not)

- **Local conditional probabilities**

$$U(x_s = 0, V_s = (0, 0, 0, 1)) = \beta$$
$$U(x_s = 1, V_s = ((0, 0, 0, 1)) = 3\beta$$

$$P(X_s = 0|V_s) = \frac{\exp(-\beta)}{\exp(-\beta) + \exp(-3\beta)}$$

$$P(X_s = 1|V_s) = \frac{\exp(-3\beta)}{\exp(-\beta) + \exp(-3\beta)}$$

Sampling of MRF

◦ **problem**

X (P(X=x)) being defined (neighborhood system, clique potentials)

how sampling a configuration following P(X) ?

◦ **solutions**

two possible algorithms :

— Gibbs sampler
— *Metropolis algorithm*

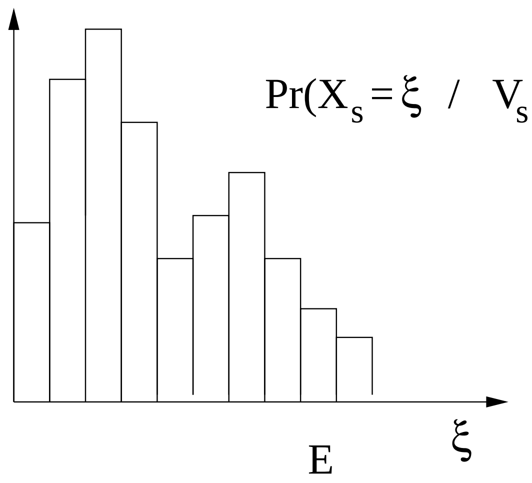
Gibbs sampler

- **principle** building a sequence of configurations $x(n)$ by visiting each site (randomly sampled)

- **local change :** $x = \{x_s,\ x^s\} \ \rightarrow \ x' = \{\xi,\ x^s\}$

$$\frac{\Pr(X = x')}{\Pr(X = x)} = \frac{\Pr(X_s = \xi \ / \ X^s = x^s)\ \Pr(X^s = x^s)}{\Pr(X_s = x_s \ / \ X^s = x^s)\ \Pr(X^s = x^s)} = \frac{\Pr(X_s = \xi \ / \ V_s)}{\Pr(X_s = x_s \ / \ V_s)}$$

- **$\Rightarrow$ sampling of the new state according to the conditional probability**



$$\Pr(X_s = \xi \ / \ V_s) = \frac{\exp - \frac{U(\xi \ / \ V_s)}{T}}{Z_T^s}$$

Metropolis sampling

◦ **principle**

sampling of a site s and a new state x'_s

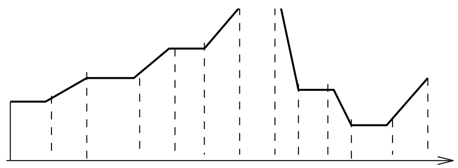
computation of the energy variation between the two states $x = (x_s, x_t, t \neq s)$ and $x' = (x'_s, x_t, t \neq s)$

— if $\Delta U = U(x') - U(x) < 0$ accept the new state

— else, accept (reject) the new state with the probability $p = \exp(-\Delta U)$ $(1 - p)$

Sampling of a MRF

- **Markov chain**

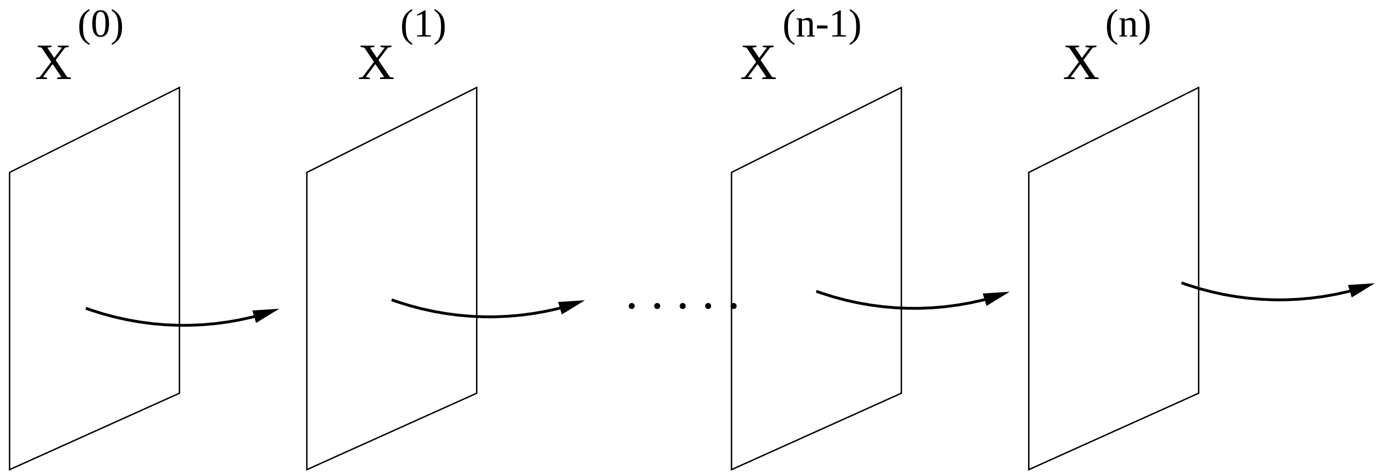


$$\begin{aligned}\Pr(X^{(n)} = x^{(n)} \;/\; X^{(0)} &= x^{(0)}, X^{(1)} = x^{(1)} \ldots X^{(n-1)} = x^{(n-1)}) \\ &= \Pr(X^{(n)} = x^{(n)} \;/\; X^{(n-1)} = x^{(n-1)})\end{aligned}$$

- **transition kernel**

$$Q_n(x,\; y) = \Pr(X^{(n)} = y \;/\; X^{(n-1)} = x) \qquad x \to y$$

- **homogeneous Markov chain :** $Q_n(x,\ y)$ **independant of** n
- **Sampling : Markov chain of images !**



$$X^{(0)},\ X^{(1)}\ \ldots X^{(n-1)},\ X^{(n)},\ X^{(n+1)}\ldots$$

- **homogeneous sampling : finding** $Q(x,\ y)$ **such that**

$$\lim_{n\to+\infty} \Pr(X^{(n)} = x) = P(X = x)$$

Examples of Markov models

- **Ising model**

$$U(x) = -\beta \sum_{c=(s,t)} x_s \; x_t - B \sum_{s \in S} x_s \qquad E = \{-1, +1\}$$

$$U(x) = \beta \sum_{c=(s,t)} 1_{x_s \neq x_t} \qquad E = \{0, 1\}$$

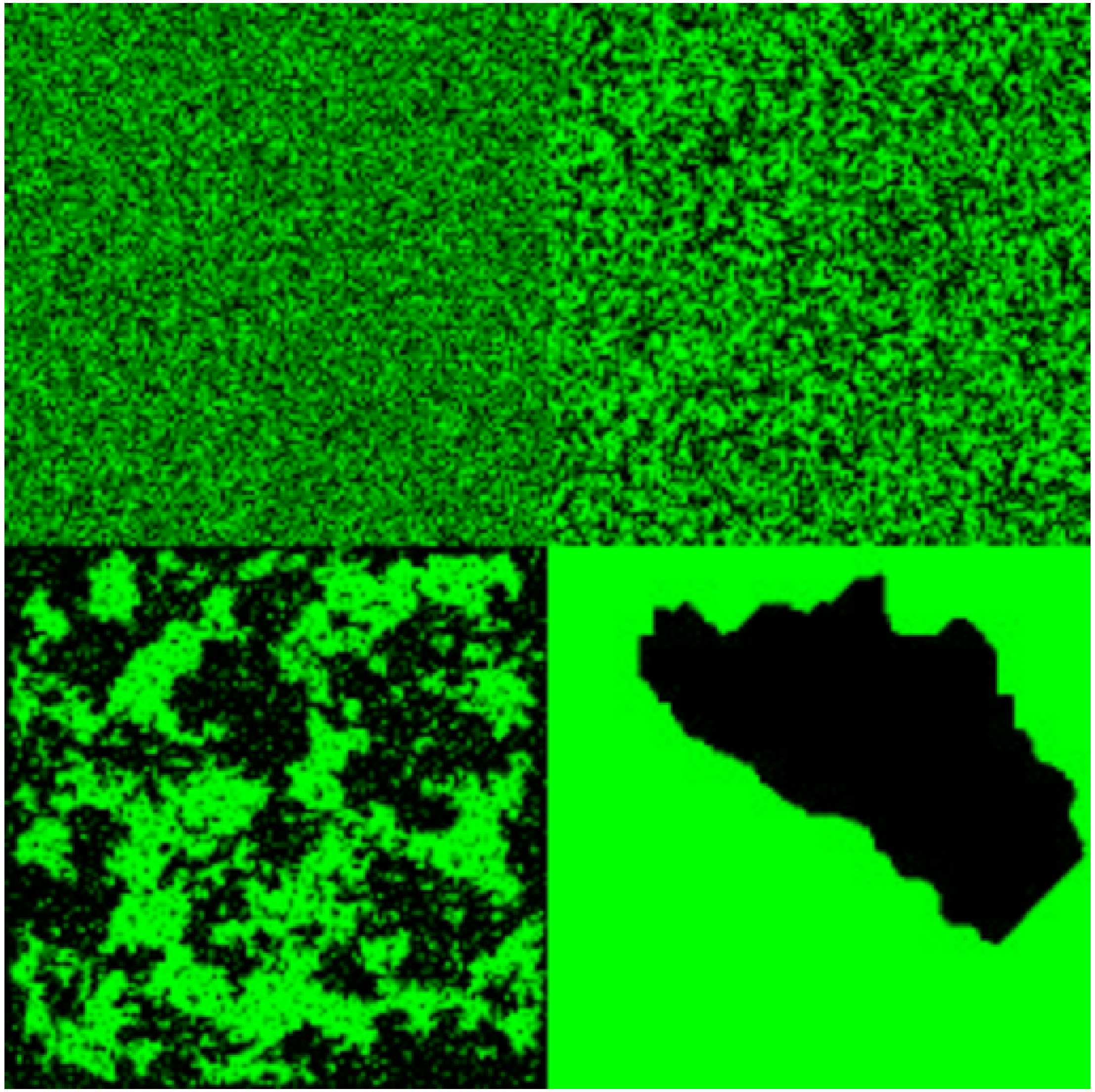
- **Potts model**

$$U(x) = \beta \sum_{c=(s,t)} 1_{x_s \neq x_t} \qquad E = \{0..q-1\}$$

- **Gaussian Markov model**

$$U(x) = \beta \sum_{c=(s,t)} (x_s - x_t)^2 + \alpha \sum_{s \in S} (x_s - \mu_s)^2 \qquad E = \mathbf{R}$$

Ising model in 4-connexity and $B = 0$

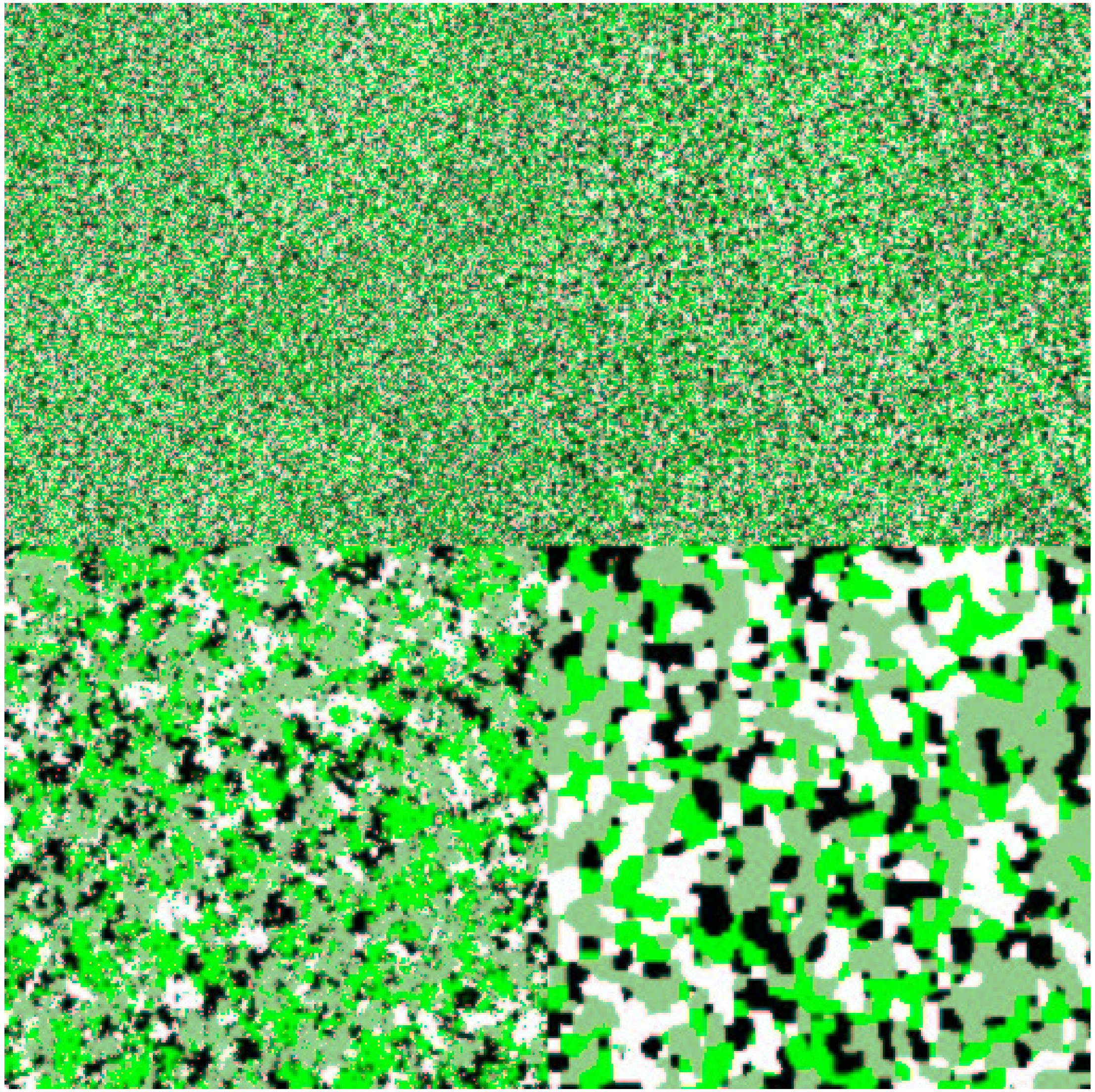


A	B
C	D

- A : *random image* : $\beta = 0$ - B : *weak regularization* : $\beta = 0.2$

- C : *"critical" regularization* : $\beta \approx 0.44$ - D : *strong regularization* : $\beta = 4.0$

Potts model in 4-connexity and $B = 0$ $(q = 4)$

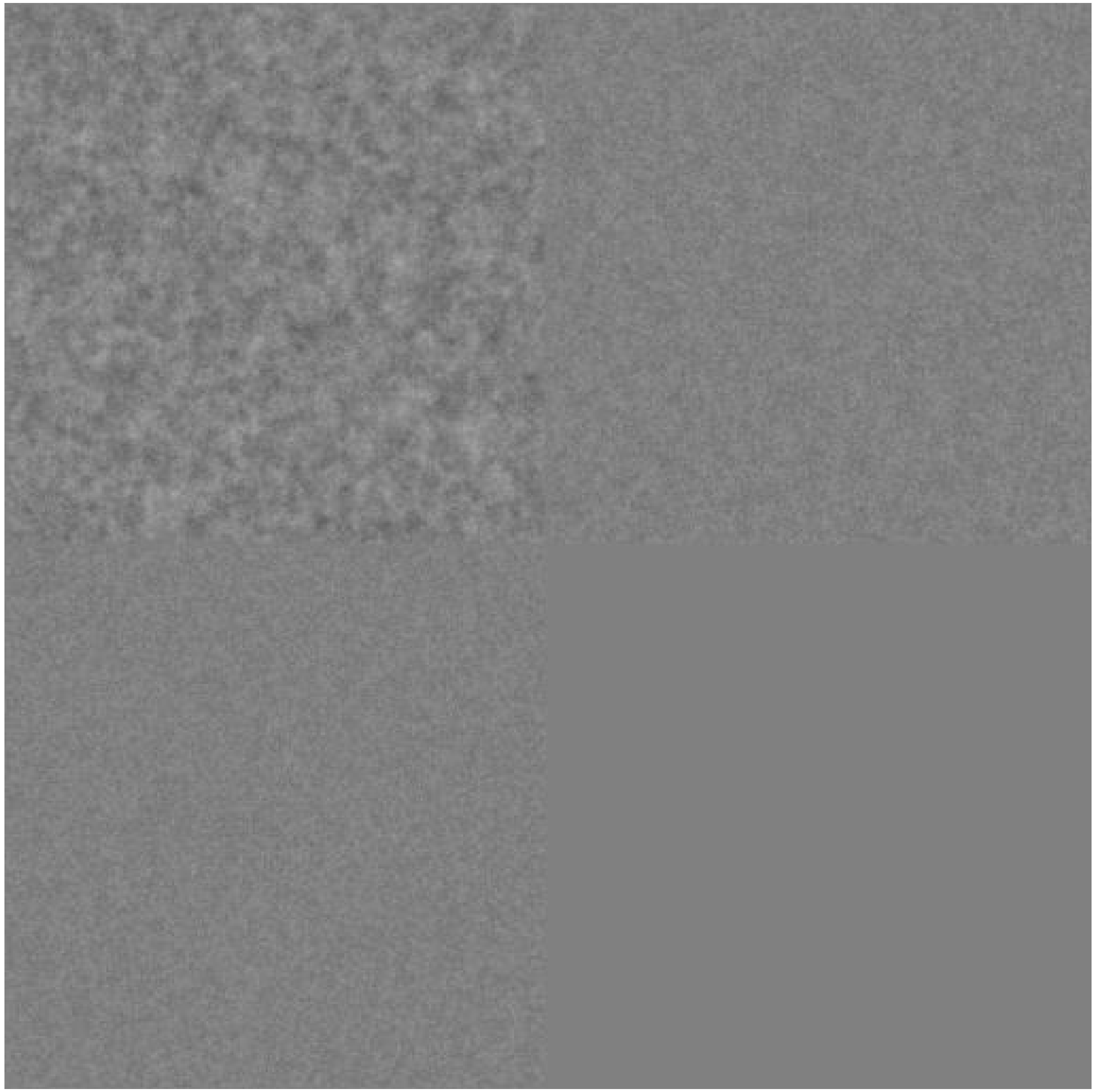


A	B
C	D

- A : *random image :* $\beta = 0$ - B : *weak regularization :* $\beta = 0.2$

- C : *"critical" regularization :* $\beta \approx 1,099$ - D : *strong regularization :* $\beta = 4.0$

Gaussian Markov model in 4-connexity



$$U(x) = \beta \sum_{c=(s,t)} (x_s - x_t)^2$$
$$+ \alpha \sum_{s \in S} (x_s - \mu_s)^2$$

A	B
C	D

- A : $\alpha = 5.10^{-4}$ - B : $\alpha = 5.10^{-3}$

- C : $\alpha = 2.10^{-3}$ - D : $\alpha = \infty$ *($\mu = 127$ for all simulations)*

Gaussian Markov models

- **model with independent pixels**

$$\Pr(X = x) = \left[\sqrt{\frac{\alpha}{\pi}}\right]^{|S|} \prod_{s \in S} e^{-\alpha\ (x_s - \mu)^2} \Leftrightarrow \frac{\exp\ -U(x)}{Z}$$

$$\text{with } U(x) = \alpha \sum_{s \in S} (x_s - \mu)^2$$

- **General case - auto-normal model**

$$U(x) = \alpha \sum_{s \in S} (x_s - \mu_s)^2 + \beta \sum_{c=(s,t)} (x_s - x_t)^2$$

$\downarrow$ local mean

$\downarrow$ coupling

- variable illumination μ_s

- constant illumination $\mu_s = \mu = 128$

Conditional probability for γ connexity : gaussian

$$\Pr(X_s = x_s \,/\, V_s) = \frac{1}{z}\ \exp - \left[\alpha (x_s - \mu_s)^2 + \beta \sum_{c=(s,t),\ t \in \mathcal{V}_s} (x_s - x_t)^2 \right]$$

$$= \sqrt{\frac{2\ (\alpha + \beta\ \gamma)}{\pi}}\ .\ \exp - \ (\alpha + \beta\ \gamma) \left[x_s - \left(\frac{\alpha\ \mu_s + \beta \displaystyle\sum_{t \in \mathcal{V}_s} x_t}{\alpha + \beta\ \gamma} \right) \right]^2$$

- **conditional expectation**

$$\mathbf{E}[X_s \,/\, V_s] = \frac{\alpha\ \mu_s + \beta \displaystyle\sum_{t \in \mathcal{V}_s} x_t}{\alpha + \beta\ \gamma} = \frac{\alpha\ \mu + \beta \displaystyle\sum_{t \in \mathcal{V}_s} x_t}{\alpha + \beta\ \gamma} \ \rightarrow\ \text{gravity center } (\mu_s = \mu)$$

- **conditional variance**

$$\mathrm{var}(X_s \,/\, V_s) = \frac{1}{2\ (\alpha + \beta\ \gamma)} \ \rightarrow\ \text{independent of } \mu_s \text{ and } x_t, t \in \mathcal{V}_s$$

- $\Rightarrow$ **statistics computation for fixed** $X_{\mathbf{V}} = \sum_{t \in \mathcal{V}_s} x_t$

Sampling versus optimization

- **Finding the configuration with highest probabibility**

equivalent to searching for the global minimum of the energy

- **Problem categorization**

  — global / local inima
  — continuous / discrete labels
  — convex / non convex energy

ICM - Iterated Conditional Modes

◦ Discrete labels and local minimum

Sequential update of each site :

- choice of a site
- computation of the conditional probabilities (conditional local energies) for the fixed local neighborhood
- choice of the state maximizing (minimizing) the conditional probability (conditional local energy)

◦ ICM algorithm

- converges to a local minimum
- depends very much of the initial configuration
- very fast
- similar to a “gradient descent” with continuous labels

Gibbs distribution with temperature parameter

$$P_T(X = x) = \frac{1}{Z_T} \; \exp - \frac{U(x)}{T}$$

$$U(x) = \sum_{c \in \mathcal{C}} U_c(x) \qquad \text{global energy}$$

$$Z_T = \sum_{y \in \Omega} \exp - \frac{U(y)}{T} \qquad \text{partition function}$$

Behaviour for extreme temperatures

- **intuition**

$$\frac{P_T(X=y)}{P_T(X=x)} = \exp -\frac{[U(y)-U(x)]}{T} \quad \forall x,\ y \in \Omega$$

$$T \to \infty \quad \exp -\frac{[U(y)-U(x)]}{T} \to 1 \quad \forall x, y \in \Omega \ \underline{\text{fini}}$$

$$T \to 0 \quad \exp -\frac{[U(y)-U(x)]}{T} \to 0 \quad \text{si } U(y) > U(x)$$

- **demonstration for $T \to \infty$**

$$P_T(X=x) = \frac{\exp -\frac{U(x)}{T}}{\sum_{y\in\Omega} \exp -\frac{U(y)}{T}} = \frac{1}{\sum_{y\in\Omega} \exp -\frac{[U(y)-U(x)]}{T}}$$

$$\to \frac{1}{\text{Card } \Omega} \quad \forall x \in \Omega \quad \textcolor{blue}{\text{equidistribution on } \Omega}$$

Gibbs distribution with temperature parameter (2)

- **deemonstration for $T \to 0$**

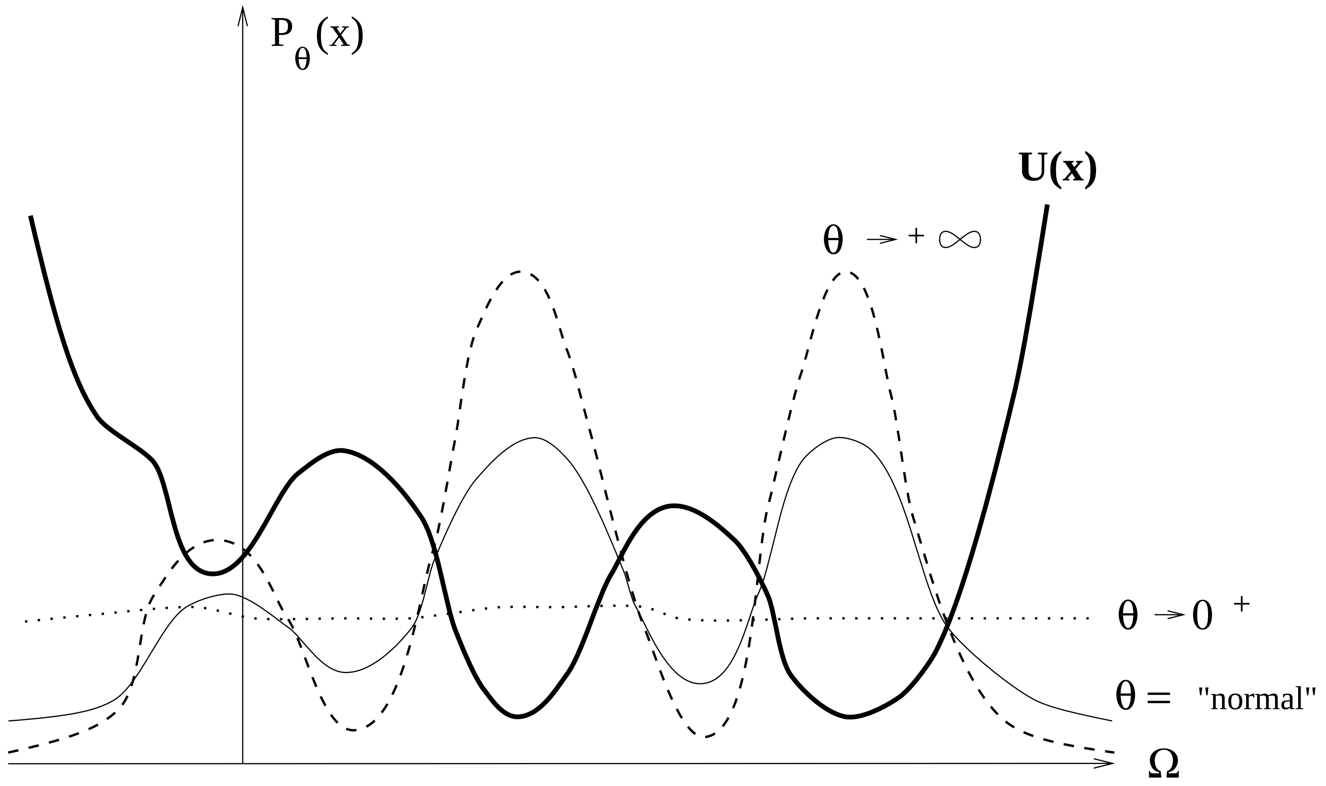
$$U^* = \min_{x \in \Omega} U(x) \quad \Omega^* = \{x \in \Omega \mid U(x) = U^*\}$$

$$P_T(X = x) = \frac{\exp -\frac{[U(x) - U^*]}{T}}{\sum_{y \in \Omega} \exp -\frac{[U(y) - U^*]}{T}} = \frac{\exp -\frac{[U(x) - U^*]}{T}}{\text{Card } \Omega^* + \sum_{y \in \Omega, y \notin \Omega^*} \exp -\frac{[U(y) - U^*]}{T}}$$

$$\to \begin{cases} \frac{1}{\text{Card } \Omega^*} & \text{si } x \in \Omega^* \\ 0 & \text{si } x \notin \Omega^* \end{cases} \qquad \text{equidistribution on } \Omega^*$$

(Recall : $\exp -\frac{[U(y) - U(x)]}{T} \to 0$ si $U(y) > U(x)$)

Gibbs distribution with temperature parameter(3)



Simulated Annealing

- **theorem (Geman and Geman 1984)**
  - — building a sequence of images with sampling for $P_{T_n}(X)$ with T_n decreasing slowly and initializing the sampler with the current configuration
  - — the configuration obtained when the temperature is close to 0 is a global minimum of the energy
  - — Conditions : temperature decrease should be very low (cooling schedule with logarithmic rate) and initial temperature should be high enough

Simulated Annealing

- **theorem (Geman and Geman 1984)**

if $\quad Q_n(x,y)$ with $T_n \searrow 0 \quad , T_n \geq \dfrac{T_0}{\log(1+n)}$
then $\lim_{n \to +\infty} \Pr(X^{(n)} = x) = \dfrac{1}{\Omega^*} \; \delta(x \in \Omega^*) \leftarrow$ energy global minimum

building a sequence of images with samplers for $P_{T_n}(X)$

and T_n following a logarithmic decreasing

- **theoretical condition**

$T_0 = \Delta\, U_{max}$ Metropolis — $T_0 = \sum_{s \in S} \delta\, U(.\ /\ V_s)_{max}$ Gibbs

- **in practice : $T_n = T_0\ \alpha^n$ with :**

$T_0 \approx \delta\, U(.\ /\ V_s)_{max} \quad , \alpha \approx 0.98$

Simulated Annealing