Simultaneous Image Restoration and Hyperparameter Estimation for Incomplete Data by a Cumulant Analysis

Marc Sigelle

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Simultaneous Image Restoration and Hyperparameter Estimation for Incomplete Data by a Cumulant Analysis

Marc Sigelle

N° 3249

September 1997

THÈME 3

Rapport de recherche

ISSN 0249-6399

Simultaneous Image Restoration and Hyperparameter Estimation for Incomplete Data by a Cumulant Analysis

Marc Sigelle *

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet PASTIS

Rapport de recherche n˚3249 — September 1997 — 68 pages

Abstract: The purpose of this report is first to show the main properties of Gibbs distributions considered as exponential statistics on finite spaces, as well as their sampling and annealing properties. Moreover, the definition and use of their cumulant expansions enables to exhibit other important properties of such distributions. Last, we tackle the problem of hyperparameter estimation in an incomplete data frame for image restoration purposes. A detailed analysis of several joint restoration-estimation methods using generalized stochastic gradient algorithms is presented, requiring infinite, continuous configuration spaces. Using once again cumulant analysis and its relationship with Statistical Physics allows us to propose new algorithms and to extend them to an explicit boundary frame.

Key-words: exponential statistics, Gibbs distributions, hyperparameters, restoration, estimation, stochastic gradient.

(Résumé : tsvp)

* E-mail: sigelle@ima.enst.fr. This work was done while the author was on sabbatical leave from Ecole Nationale Supérieure des Télécommunications Département Images, 46 rue Barrault 75634 Paris Cedex 13 France.

Unité de recherche INRIA Sophia Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA ANTIPOLIS Cedex (France)
Téléphone : 04 93 65 77 77 – international : +33 4 93 65 77 77 – Télécopie : 04 93 65 77 65 – international : +33 4 93 65 77 65

Restauration d'images et estimation simultanée des hyperparamètres par une analyse en cumulants

Résumé : Dans ce rapport, on présente d'abord les propriétés de base des distributions de Gibbs en tant que statistiques exponentielles: échantillonnage, recuit et convergence pour un espace de configuration fini. On présente également brièvement les principales propriétés du développement en cumulants de telles distributions. On aborde ensuite le problème de l'estimation des hyperparamètres dans un contexte de restauration d'images, c'est-à-dire en données incomplètes. Cela nous amène à formuler plusieurs méthodes de restauration-estimation simultanées des hyperparamètres, sur la base d'un algorithme de gradient stochastique généralisé. Un développement en cumulants nous amène là aussi à proposer une nouvelle formulation ainsi qu'à étendre la méthode au cas d'un processus bord complémentaire, explicite, pour la gestion des discontinuités (contours).

Mots-clé : statistiques exponentielles, distributions de Gibbs, hyperparamètres, restauration, estimation, gradient stochastique.

1 Introduction

The purpose of this report is first to show as simply as possible the main properties of Gibbs distributions on finite spaces considered as exponential statistics, as well as their main sampling and annealing properties. We shall try as often as possible to give a "physical" sense to these properties and to illustrate them with figures, having clearly in mind that these arguments do not favour a scope for pure mathematicians. The pinpoint of our argumentation will be the definition, interpretation and use of cumulants [Ma-85, Malyshev-91] for statistical likelihood properties and of Dobrushin's contraction cœfficient in order to find classical sampling and annealing properties of Gibbs distributions [Winkler-95].
We shall also tackle in this report the problem of hyperparameter estimation for incomplete data in an image restoration framework. In this case, detailed analysis of some joint restoration-estimation methods using generalized stochastic gradient algorithms requires infinite, continuous configuration spaces (in the sense that the gray level space is continuous, the set of sites remaining finite). We derive first some related algorithms with classical regularization potentials, a linear dependence on hyperparameters, and pure pixel process. We extend them to joint pixel-boundary processes taking into account the preservation of discontinuities. Once again the application of cumulant analysis and its relationship with Statistical Physics proves to be fruitful and allows to exhibit a new class of restoration-estimation algorithms for the explicit boundary framework.

2 A short review of Gibbs distribution properties on finite configuration spaces - sampling and annealing

2.1 Gibbs distributions as exponential statistics - notations

Gibbs distributions can be defined on general configuration spaces, very afar from spaces related to mono- or multi-dimensional images (think for example to the traveling salesman problem on a graph.) However, we will focus on distributions related to the modeling of image distributions. Assuming from now a finite configuration space hypothesis, we shall define :

- A lattice of sites: S supposed to be finite. Its cardinal will be noted: $|S| =$ Card (S).
- A gray level space: E, also finite, and assumed to be the same at each site of lattice S for sake of simplicity.
- The configuration space is thus: $\Omega = E^S$.

2.2 The Gibbs distribution with a single parameter

We introduce first the Gibbs distribution with a single parameter for sake of simplicity. We shall then extend associated results whenever possible and when needed. We have by definition:

$$P_\theta(x) = \frac{1}{Z} \ \exp - \theta \, U(x) \text{ where } Z_\theta = \sum_{y \in \Omega} \exp \, - \theta \, U(y) \tag{1}$$

Z_θ is usually called the partition function.
U is a given function $\Omega \mapsto \mathbf{R}$. We can consider parameter θ in either two ways:
1. $\theta \geq 0$ is usually considered as the inverse temperature in statistical physics. It can be either fixed for sampling and estimation purposes or varying (annealing). Of utmost importance is the behaviour of Gibbs distribution at limit ranges, *i.e.* :

- at $\theta = 0$ (infinite temperature) : equidistribution on Ω.
- at $\theta = +\infty$ (null temperature) : equidistribution on Ω^*, the set of configurations reaching global minimal energy (noted U^*).

Both properties originate from the fact that given any configurations x and y:

$$\frac{P_\theta(y)}{P_\theta(x)} = \exp - \theta \, (U(y) - U(x))$$

The former one results immediately (finite configuration space), whereas the latter is obtained by taking $x \in \Omega^*$.
2. $\theta \in \mathbf{R}$ plays the role of an hyperparameter in disciplines such as image processing to be either fixed (sampling) or varying in sequence in order to be estimated.

The latter case is quite general, but the first one induces an asymmetry in the behaviours of likelihood which we shall tackle when necessary (especially in the simulated annealing method when temperature gœs down to zero).
In following section, we derive basic properties of the likelihood associated to a Gibbs distribution. Next section finds these results again by a general cumulant expansion which can be skipped in a first approach by the reader.

2.3 Likelihood for exponential statistics: a simple analysis

We have the following fundamental **likelihood** formula:

$$\begin{aligned}\frac{\partial}{\partial\theta}\log P_\theta(x) &= -\,U(x) - \frac{\partial}{\partial\theta}\log Z_\theta = -\,U(x) + \frac{\sum_{y\in\Omega} U(y)\exp -\theta\,U(y)}{\sum_{y\in\Omega}\exp -\theta\,U(y)} \\ &= -\,U(x) + \mathbb{E}_\theta[U] \end{aligned}\tag{2}$$

where $\mathbb{E}_\theta[X]$ means **the statistical expectation** of random variable X under the Gibbs distribution of parameter θ. In order to study the behaviour of the likelihood as a function of parameter θ in its whole variation range, we compute:

$$\begin{aligned}\frac{\partial}{\partial\theta}\,\mathbb{E}_\theta[U] &= \left(\frac{\sum_{y\in\Omega} -\,U^2(y)\exp -\theta\,U(y)}{\sum_{y\in\Omega}\exp -\theta\,U(y)}\right) + \left(\frac{\sum_{y\in\Omega} U(y)\exp -\theta\,U(y)}{\sum_{y\in\Omega}\exp -\theta\,U(y)}\right)^2 \\ &= -\mathrm{var}_\theta(U) \leq 0\end{aligned}\tag{3}$$

where $\mathrm{var}_\theta(U)$ is the statistical variance of energy U under the Gibbs distribution for θ, also known as the specific heat formula in classical thermodynamics. This formula results in fact from a more general property of exponential statistics. This states that if Ψ is any function $\Omega \mapsto \mathbf{R}$ then:

$$\frac{\partial}{\partial\theta}\,\mathbb{E}_\theta[\Psi] = -\,(\,\mathbb{E}_\theta[\Psi\,U] - \mathbb{E}_\theta[\Psi]\,\mathbb{E}_\theta[U]\,) = -\,\mathrm{cov}_\theta(\Psi,\,U)\tag{4}$$

which simply results from deriving

$$\mathbb{E}_\theta[\Psi] = \frac{\sum_{y\in\Omega}\Psi(y)\exp -\theta\,U(y)}{\sum_{y\in\Omega}\exp -\theta\,U(y)}$$

w.r.t. θ. This general property results itself from a cumulant analysis of multi-parameter Gibbs distributions, which will be detailed in further subsection 4.2.

2.4 The main properties of Gibbsian likelihood

We shall note in the following $q = |\Omega^*|$ the cardinal of the set of minimal energy configurations, which are attained at low temperature (*i.e.* for high values of θ). We shall also note, for high temperature considerations (*i.e.* at low positive values of θ), the empiric average of energy U on configuration space Ω as:

$$\bar{U} = \frac{1}{|\Omega|} \sum_{x \in \Omega} U(x)$$

Previous results (2) and (3) show that given a fixed configuration x:

• $\frac{\partial}{\partial \theta} \log P_\theta(x) \;=\; -\; U(x) \;\; + \;\; \mathbb{E}_\theta[U]$ is a monotonously decreasing function of θ (see Fig. 1 a). It varyies between extremal values:

$-\,U(x) + \bar{U} \quad (\theta = 0)$ and $-\,U(x) + U^* \quad (\theta = +\infty)$	$-\,U(x) + U^{max} \quad (\theta = -\infty)$ and $-\,U(x) + U^* \quad (\theta = +\infty)$
if θ is an inverse temperature	**if θ is a parameter.**

• then that the log-likelihood $\log P_\theta(x)$ is a **concave** function of θ (see Fig. 1 b).

• so that the likelihood itself $P_\theta(x)$ is a unimodal function, *i.e.* having at most one maximum point for parameter value θ_x (see Fig. 1 c) . More specifically, this maximum likelihood parameter θ_x verifies, due to (2) :

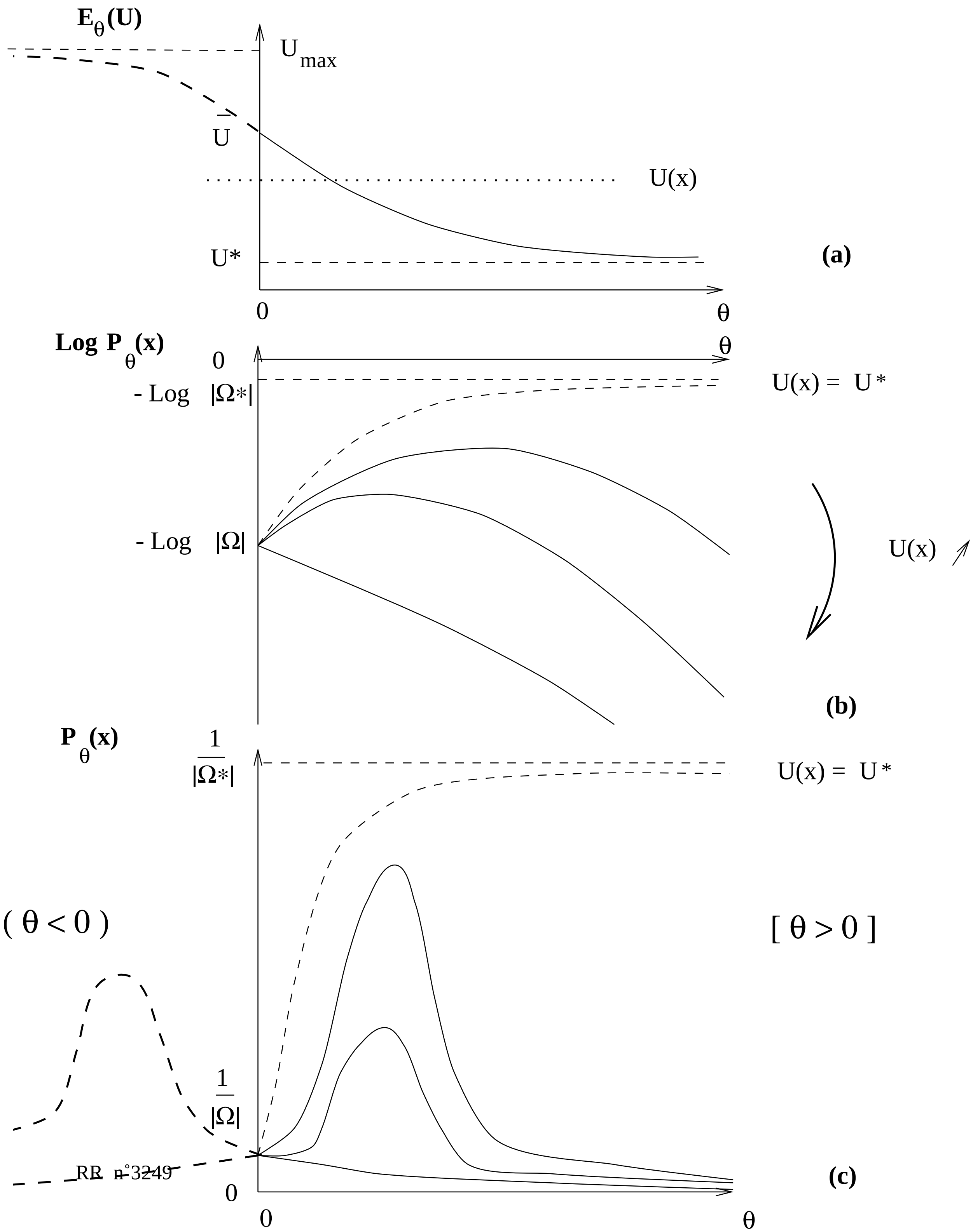


Figure 1: behaviour of the statistical average energy (a), of the log-likelihood (b) and of the likelihood itself (c), *i.e.* Gibbs distribution, as functions of parameter θ.

$(\circ)\ \theta_x = 0\ (T_x = +\infty)$	if $U(x) \geq \bar{U}$	$\theta_x = 0$	if $U(x) = \bar{U}$
$(\star)\ \theta_x > 0$	if $U^* < U(x) < \bar{U}$	$(\star)$ or $\theta_x < 0$	if $\bar{U} < U(x) < U^{max}$
$(\triangle)\ \theta_x = +\infty\ (T_x = 0)$	if $U(x) = U^*$	$(\triangle)$ or $\theta_x = -\infty$	if $U(x) = U^{max}$
if $\theta = \dfrac{1}{T}$ is an inverse temperature		**if θ is a parameter**	

and obviously satisfies the maximum-likelihood equation:

$$\mathbb{E}_{\theta_x}[U] = U(x) \tag{5}$$

Moreover, since $\mathbb{E}_\theta[U] \searrow U^*$ when $\theta \to +\infty$ we can infer the following behaviour of likelihood as parameter θ increases:

- in case $(\circ)$, $P_\theta(x) \searrow 0$ when θ increases, since $\mathbb{E}_\theta[U] < \bar{U}$ when $\theta > 0$.
- in case $(\star)$, $P_\theta(x) \searrow 0$ after $\theta > \theta_x$, which results from inspection of (2) and (5), and from the fact that $\mathbb{E}_\theta[U]$ is a monotonously decreasing function of θ .
- in case $(\triangle)$ $P_\theta(x) \nearrow \dfrac{1}{q}$, since $\mathbb{E}_\theta[U] > U^*$.

These properties will be extremely useful in the analysis of the simulated annealing algorithm convergence behaviour.

3 Homogeneous and inhomogeneous sampling of Gibbs distributions - simulated annealing

In this section we briefly adress the convergence properties of sequences of MRF Gibbs samplers of distributions associated to a sequence of parameters θ_n.

3.1 Recalls on measures, kernels and Dobrushin cœfficient

We follow in this subsection the presentation of [Winkler-95].

Distance between measures : recall that a measure μ on Ω is an application $\Omega \mapsto$ **R** such that

- $\mu(x) \geq 0 \;\; \forall x \in \Omega$
- $\sum_{x\in\Omega} \mu(x) = 1$

Now the classical (variation) distance between two measures is defined as:

$$||\mu_1 - \mu_2|| = \sum_{x\in\Omega} |\mu_1(x) - \mu_2(x)| \tag{6}$$

We have (see Fig. 2):

- $0 \leq ||\mu_1 - \mu_2|| \leq 2$
- $||\mu_1 - \mu_2|| = 0 \Leftrightarrow \mu_1 = \mu_2$
- $||\mu_1 - \mu_2|| = 2 \Leftrightarrow \mu_1$ and μ_2 have disjoint supports
- $||\mu_1 - \mu_2|| = 2 - 2\sum_{x\in\Omega} \min(\mu_1(x), \mu_2(x))$
  (hint: $a, b \geq 0 \Rightarrow |a - b| = a + b - 2\ \min(a,\ b)$)

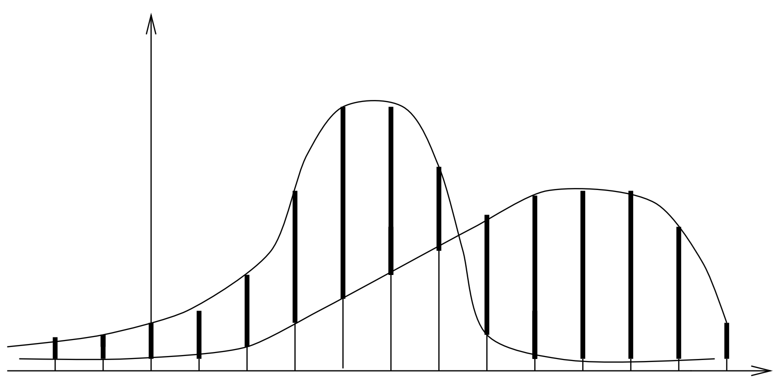


Figure 2: Variation distance between two measures

Kernels A kernel Q is an application $\Omega \times \Omega \mapsto \mathbf{R}$ such that $Q(x,.)$ is a measure $\forall x \in \Omega$
Now we can define:

- product of a measure by a kernel: it is the measure defined as:

$$(\mu\, Q)(x) = \sum_{y \in \Omega} \mu(y)\; Q(y,\; x) \;\; \forall x \in \Omega \tag{7}$$

- product of two kernels: it is the kernel defined as

$$(Q\; R)(x,y) = \sum_{z \in \Omega} Q(x,\; z)\; R(z,\; y) \;\; \forall x, y \in \Omega \tag{8}$$

The most usual interpretation of kernels is that of conditional probabilities , *i.e.* transition probabilities in Markov chain theory:

$$Q(x,\; y = \pi(Y = y \;/\; X = x) = \pi(x \to y)$$

where $\pi(.)$ is some measure on Ω . It also results from previous definitions that, as measures:

$$(Q\; R)(x,.) = Q(x,.)R \;\; \forall x \in \Omega \tag{9}$$

which will prove useful in the following.

Dobrushin contraction cœffi cient This powerful conception results of the following definition

$$c(Q) = \frac{1}{2} \max_{x,y \in \Omega} ||Q(x,.) - Q(y,.)||$$

Properties resulting from its definition are:

- $0 \leq c(Q) \leq 1$
- $c(Q) = 1 - \min_{x,y \in \Omega} \sum_{z \in \Omega} \min(Q(x,z),\; Q(y,z)) \leq 1 - |\Omega| \min_{a,b \in \Omega} Q(a,b)$

In particular, suppose there exist two values x and $y \in \Omega$ such that measures $Q(x,.)$ and $Q(y,.)$ have disjoint support (let us think of them as transition probabilities $\pi(x \to .)$ and $\pi(y \to .)$). Then $c(Q) = 1$. In the opposite case, $c(Q) < 1$.
Now the fundamental properties of Dobrushin contraction cœfficient, which form the basis of sampling and annealing convergence properties, are the following:

a) $||\mu_1 \, Q - \mu_2 \, Q|| \leq c(Q) \; ||\mu_1 - \mu_2||$

b) $c(Q \; R) \leq c(Q) \; c(R)$

One can notice that latter property **b**) results almost immediately from former one **a**), since from (9) :

$$(QR)(x,.) - (QR)(y,.) = Q(x,.) \; R - Q(y,.) \; R$$

We refer the reader to an elegant demonstration of property **a**) in [Winkler-95]. Once this is done, most properties of sampling and annealing result almost easily, as we shall try to prove.

Invariant measures They follow the property

$$(\mu Q)(x) = \mu(x) \;\; \forall x \in \Omega$$

They can be seen eigenvectors of kernel Q (viewed as a stochastic matrix) associated to eigenvalue 1. Theorem of Perron-Frobenius ensures the existence of such a measure for any kernel (see [Winkler-95]).

Reversibility

$$\mu(x) \; Q(x, \; y) = \mu(y) \; Q(y, \; x) \;\; \forall x, y \in \Omega$$

Summing this formula on y and recalling previous definition of measure-kernel product yields immediately invariant measure property for μ. Let us interpret reversibility with the help of transition probabilities. Bayes theorem ensures that for any distribution π:

$$\pi(X = x) \; \pi(Y = y \; / \; X = x) = \pi(Y = y) \; \pi(X = x \; / \; Y = y) \;\; \forall x, y \in \Omega$$

i.e. $\pi(X = x) \; \pi(x \to y) = \pi(Y = y) \; \pi(y \to x)$. This shows and (conforts us !) that measure π is reversible, and therefore invariant, *w.r.t.* transition probability $\pi(. \to .)$.

Basic transition kernels of Markov Random Field Theory, such as the Gibbs and Metropolis samplers, will have the reversibility property *w.r.t.* to initial Gibbs distribution, as we shall see.

3.2 Gibbs and Metropolis samplers

Let us introduce them in a simple way and show they follow the desirable reversibility property *w.r.t.* to Gibbs distribution.

3.2.1 Gibbs sampler

Since by changing grey level value at some site s $x_s : \xi \to \eta$, we have the following energy change $\Delta U = \Delta U(x_s \,/\, \mathcal{N}_s)$ it follows

$$U(x_s = \xi) + U(x_s = \eta \,/\, \mathcal{N}_s) = U(x_s = \eta) + U(x_s = \xi \,/\, \mathcal{N}_s)$$

and thus, for all θ

$$\frac{\exp\, -\theta\, U(x_s = \xi)}{Z_\theta} \cdot \frac{\exp\, -\theta\, U(x_s = \eta \,/\, \mathcal{N}_s)}{Z_\theta^s} = \frac{\exp\, -\theta\, U(x_s = \eta)}{Z_\theta} \cdot \frac{\exp\, -\theta\, U(x_s = \xi \,/\, \mathcal{N}_s)}{Z_\theta^s}$$

This shows that following kernel, also called Gibbs sampler for parameter value θ

$$Q_\theta(x,y) = \mathbb{1}_{x_r = y_r, r \neq s} \,.\, P_\theta(y_s \,/\, \mathcal{N}_s) = \mathbb{1}_{x_r = y_r, r \neq s} \,.\, \frac{\exp\, -\theta\, U(x_s \,/\, \mathcal{N}_s)}{Z_\theta^s}$$

is reversible *w.r.t.* to initial Gibbs distribution $P_\theta(x) = \dfrac{\exp\, -\theta\, U(x)}{Z_\theta}$ and admits it as invariant measure.

Let T be an arrangement (permutation) of S. This is also called as a tour. Then given two configurations x and y, we have

$$\begin{aligned} Q_\theta(x,y) &= P(Y = y / X = x) = \sum_{z,t,u,v \in \Omega} \prod_{z,t,u,v} Q_\theta() \ldots \\ &= \prod_{s \in T} P_\theta(y_s \,/\, \mathcal{N}_s) \end{aligned}$$

Notice that transition between given x and y is certain here since only one path leads from the former to the latter given arrangement T (see Fig. 3).

Let us notice also that, since

$$\forall s \in S \;\; P_\theta(y_s \,/\, \mathcal{N}_s) = \frac{1}{\sum_{\xi \in E} \exp \theta\, [\, U(y_s \,/\, \mathcal{N}_s) - U(\xi \,/\, \mathcal{N}_s)\,]} \geq \frac{1}{|E|} \exp\, -\theta\, \delta_s \,,$$

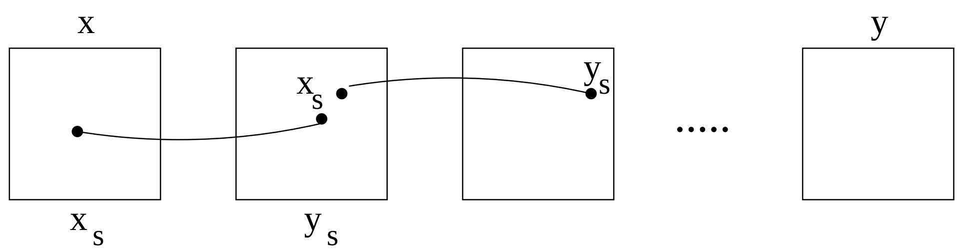


Figure 3: a tour

where $\delta_s = \max_{\mathcal{N}_s} \left(\max_{\eta \in E} U(\eta \ / \ \mathcal{N}_s) - \min_{\xi \in E} U(\xi \ / \ \mathcal{N}_s) \right)$ is the local oscillation of conditional energy $U(. \ / \ .)$, we have $Q_\theta(x, y) \geq \prod_{s \in T} \frac{1}{|E|} \exp - \theta \ \delta_s \ \ \forall x, y \in \Omega$, and thus

$$\begin{aligned} c(Q_\theta) &\leq 1 - |\Omega| \, \frac{1}{|E|^{|S|}} \, \exp(- \, \theta \, . \, \sum_{s \in S} \delta_s) \qquad \textit{i.e.} \\ c(Q_\theta) &\leq 1 - \exp(- \, \Gamma \, \theta) \quad \text{ with } \Gamma = \sum_{s \in S} \delta_s \end{aligned} \tag{10}$$

3.2.2 Metropolis sampler

Let us suppose by physical considerations that

$$Q_\theta(x, y) = 1 \text{ if } U(y) < U(x)$$

i.e. assuming sure transition when energy decreases. Is is then possible to find the nature of $Q_\theta(x, y)$ whatever x and y, in particular if $U(y) \geq U(x)$, in order to retain the desirable reversibility property:

$$P_\theta(x) \ Q_\theta(x, y) = P_\theta(y) \ Q_\theta(y, x)$$

Thus, if $U(y) > U(x)$, we have by specification $Q_\theta(y, x) = 1$ and it follows

$$Q_\theta(x, y) = \frac{P_\theta(y)}{P_\theta(x)} = \exp \, - \, \theta \ (U(y) - U(x))$$

This is known as the Metropolis sampler, which also exhibits the reversibility property *w.r.t.* Gibbs distribution $P_\theta(x)$.

Let us notice at this step that Metropolis sampler assumes from the start possible transition between any two configurations x and y, independently of their local properties. Anyway, the real transition kernel is the following: let us note $R_\theta(x, y)$ a symmetric kernel ($R_\theta(x, y) = R_\theta(y, x) \ \ \forall x, y \in \Omega$) which ensures random choice of configuration y knowing configuration x (in most case $R_\theta(x, .) = \dfrac{1}{|\Omega|}$ is the uniform law on Ω whatever configuration x). Then $S_\theta(Y = y \ / \ X = x) = R_\theta(x, y) \ . \ Q_\theta(x, y)$ is the desired real Metropolis transition kernel (see Fig. 4).

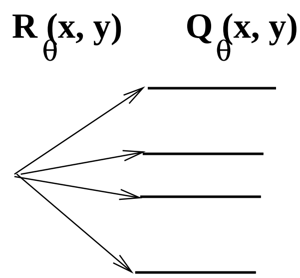


Figure 4: transition diagram for Metropolis sampler

It follows that in the uniform sampling case,

$$
\begin{aligned}
S_\theta(x, y) \ &\geq \ \frac{1}{|\Omega|} \ \exp \ - \theta \ \Delta \quad \text{with } \Delta = \max_{y \in \Omega} U(y) - \min_{x \in \Omega} U(x) \\
\Rightarrow c(S_\theta) \ &\leq \ 1 - \exp(- \ \Delta \ \theta) \qquad (11)
\end{aligned}
$$

Thus for both samplers, the following majoration applies:

$$
c(Q_\theta) \leq 1 - \exp(- \ \Gamma \ \theta)
$$

3.3 Homogeneous sampling of Gibbs distributions: convergence

3.3.1 Irreducibility

This property writes as:

$$
Q_\theta(x, y) > 0 \ \ \forall x, y \in \Omega, \quad \textit{i.e. } c(Q_\theta) < 1
$$

3.3.2 Convergence

Let us recall that with the irreducibility assumption,

$$c(Q_\theta) \leq 1 - \exp(-\,\Gamma\,\theta) \quad \text{where } \Gamma > 0$$

Let us start from initial measure μ_0 and fix some value of parameter θ. We compute the probability to obtain configuration x at step n, noted $\pi_n(x)$, by assigning kernel Q_θ to the transition probabilities between intermediate steps:

$$\begin{aligned}
\pi_n(x) = P(X^{(n)} = x) \quad &= \quad \sum_{y,t,\ldots u,v \in \Omega} P(X^{(n)} = x \;/\; X^{(n-1)} = y)\; P(X^{(n-1)} = y \;/\; X^{(n-1)} = t) \\
&\qquad \ldots\; P(X^{(1)} = u \;/\; X^{(0)} = v)\; P(X^{(0)} = v) \\
&= \quad \sum_{y,t,\ldots u,v \in \Omega} \mu_0(v)\; Q_\theta(v,u) \ldots Q_\theta(t,y)\; Q_\theta(y,x) \\
\textit{i.e. } \pi_n = \mu_0\; Q_\theta^n &
\end{aligned}$$

recall that invariant measure P_θ is such that $P_\theta = P_\theta\; Q_\theta^n \;\; \forall n \geq 0$. It implies that

$$||\pi_n - P_\theta|| = ||\mu_0\; Q_\theta^n - P_\theta\; Q_\theta^n|| \leq ||\mu_0 - P_\theta||\; c(Q_\theta)^n$$

Since $c(Q_\theta) < 1$ and $\theta < +\infty$ convergence in variation norm, whatever initial distribution μ_0, results immediately.

3.4 Inhomogeneous sampling of Gibbs distributions

Our hypotheses and notations are the following :

- a sequence $\theta_n \to \theta$
- an initial probability μ_0
- a sequence of invariant measures noted $P_n = P_{\theta_n}(.)$
- the generation of samples with transition kernels noted $Q_n = Q_{\theta_n}(.,.)$

At each step n

$$P(X^{(n)} = x) = \pi_n(x) = \mu_0\; Q_1\; Q_2 \ldots Q_n(x)$$

Since $P_n = P_n Q_n \ \forall n \geq 1$ we can compute the following difference [Winkler-95] :

$$\begin{aligned} P_n - \pi_n = P_n - \mu_0\, Q_1\, Q_2 \dots Q_n &= P_n\, Q_n - \mu_0\, Q_1\, Q_2 \dots\ Q_{n-1}\, Q_n \\ &= (P_n - P_{n-1})\, Q_n + (P_{n-1} - \pi_{n-1})\, Q_n \\ &\quad \text{and by recursion} \\ &= (P_n - P_{n-1})\, Q_{n-1} + (P_{n-1} - P_{n-2})\, Q_{n-1}\, Q_n \\ &+ (P_{n-2} - P_{n-3})\, Q_{n-2}\, Q_{n-1} \dots Q_n \ \ \dots \\ &+ (P_{n-p+1} - P_{n-p})\, Q_{n-p+1} \dots Q_n + (P_{n-p} - \pi_p)\, Q_{n-p} \dots Q_n \end{aligned} \tag{12}$$

Since each cœfficient $c(Q_{n-m}) \leq 1$, we see that a set of sufficient conditions for convergence of the inhomogeneous Markov chain so formed to the invariant measure P_θ associated to limit parameter θ is

- $\sum_n ||P_n - P_{n-1}|| < +\infty$
- $\forall p \geq 1, \ \ c(Q_{n-p} Q_{n-p+1} \dots Q_n) \to 0$ when $n \to +\infty$

First condition ensures first that $P_n \to P_\theta$ (the series $\sum_n ||P_n - P_{n-1}||$ is Cauchy, thus also the sequence P_n), so that $||P_\theta - \pi_n|| \leq ||P_\theta - P_n|| + ||P_n - \pi_n|| \to 0$ when $n \to +\infty$.

Simulated annealing case: $\theta_n \to \theta = +\infty$.
Latter set of conditions is easier to show since as seen before $\forall x \in \Omega, P_n(x)$ becomes either decreasing or increasing when θ_n increases. It remains thus to show first condition, which is more difficult, since $c(Q_n) \to 1$ when $n \to +\infty$. A sufficient condition for it is that the infinite product $\prod_{n=1}^{+\infty} (1 - \exp(-\,\Gamma\ \theta_n)) = 0$, which leads immediately to Geman conditions [Geman-84]:

$$\sum_{n=1}^{+\infty} \exp -\Gamma\ \theta_n = +\infty \Rightarrow \Gamma \theta_n \geq \frac{1}{\log n} \tag{13}$$

“Normal” case : $\theta_n \to \theta < +\infty$.
Now, first condition becomes trivial since

$$c(Q_m Q_{m+1} \dots Q_n) \leq \prod_{k=m}^{n} (1 - \exp(-\,\Gamma\,\theta_k))$$

It remains to prove thus second condition, which is not so easy.

4 Gibbsian likelihood analysis by cumulants

Previous results of section 2 for properties of Gibbs distributions viewed as linear exponential statistics can be derived from a cumulant analysis. This analysis is usual in statistical physics when studying the correlation-fluctuation properties of condensed matter [Ma-85]. We investigate first the case of single-parameter distributions.

4.1 Cumulants of single-parameter Gibbs distributions

As a matter of fact, the cumulant analysis simply states that for Gibbs distribution $P_\theta(.)$:

$$\begin{aligned} &\mathbb{E}_\theta[\exp - \lambda\, U] = \exp \sum_{n=1}^{\infty} (-1)^n\, \lambda^n \frac{C_\theta^n}{n!} \\ &\text{where } C_\theta^1 = \mathbb{E}_\theta[U],\;\; C_\theta^2 = \mathrm{var}_\theta(U) \dots \end{aligned} \tag{14}$$

$C_\theta^1, C_\theta^2 \dots C_\theta^n$ are the cumulants of order $1, 2 \dots n$ for energy U and for parameter value θ. They can be computed explicitly [1] . Let us see how it applies to our

[1] The cumulants of higher order are [Ma-85]:

$$\begin{aligned} C_\theta^3 &= \mathbb{E}_\theta[(U - \mathbb{E}_\theta[U])^3] = \mathbb{E}_\theta[U^3] - 3\,\mathbb{E}_\theta[U^2]\,.\,\mathbb{E}_\theta[U] + 2\,(\mathbb{E}_\theta[U])^3, \\ C_\theta^4 &= \mathbb{E}_\theta[U^4] - 3\,(\mathbb{E}_\theta[U^2])^2 \end{aligned}$$

C_θ^4 is also known as the kurtosis cœffi cient associated to the distribution. These cumulants will be derived later on (see also subsection 4.2).

likelihood case. We have:

$$\begin{aligned} Z_{\theta+\delta\theta} &= \sum_{y\in\Omega} \exp - \theta\, U(y) \exp\, - \delta\theta\, U(y) = Z_\theta \,.\, \mathbb{E}_\theta[\exp - \delta\theta\, U] \\ &= Z_\theta \,.\, \exp\left(-\delta\theta\; \mathbb{E}_\theta[U] + \frac{\delta\theta^2}{2}\, \mathrm{var}_\theta(U) + \ldots \right) \end{aligned}$$

Then, it follows:

$$\log Z_{\theta+\delta\theta} - \log Z_\theta = -\delta\theta\; \mathbb{E}_\theta[U] + \frac{\delta\theta^2}{2}\, \mathrm{var}_\theta(U) + \ldots \tag{15}$$

which is a Taylor expansion of $\log Z_{(.)}$ around θ, yielding thus:

$$\begin{aligned} \frac{\partial}{\partial\theta} \log Z_\theta &= -\mathbb{E}_\theta[U] \\ \frac{\partial^2 \log Z_\theta}{\partial\theta^2} &= \mathrm{var}_\theta(U) \end{aligned}$$

Since $\log P_\theta(x) = -\,\theta\, U(x) - \log Z_\theta$, we obtain previous results (2) and (3) by two successive derivations.

4.2 A cumulant analysis of multi-parameter Gibbs distributions

We shall consider in the following a more general class of distributions:

$$\begin{aligned} P_{\beta_1,\beta_2\ldots\beta_n}(x) &= \frac{1}{Z_{\beta_1,\beta_2\ldots\beta_n}}\ \exp - \left(\sum_{i=1}^{n} \beta_i\, U_i(x) \right) \\ \text{where } Z_{\beta_1,\beta_2\ldots\beta_n} &= \sum_{y\in\Omega} \exp\, - \left(\sum_{i=1}^{n} \beta_i\, U_i \right)(y) \end{aligned}$$

We shall remark first that there is a general theorem which states that for any distribution P, any set of random variables [2] Ψ_i and cœfficients λ_i and under suitable asymptotic behaviour conditions [Ma-85, Malyshev-91] :

$$\mathbb{E}[\exp\ \sum_{i=1}^{n} \lambda_i\ \Psi_i] = \exp \sum_{k_1+k_2\ldots+k_n\geq 1}^{\infty} \left(\prod_{i=1}^{n} {\lambda_i}^{k_i} \right) \frac{C^{k_1,k_2\ldots k_n}}{(k_1+k_2+\ldots k_n)!}$$

[2] Some of them could be identical..

where cumulant coefficients $C^{k_1,k_2\ldots k_n}\ldots$ can be recursively computed in the following way [Ma-85, Malyshev-91] . For sets of n integers of the form $\mathcal{L}=\{l_1,l_2,\ldots l_n\}$, we note: $\mathcal{L}\oplus\mathcal{M}=\{l_1+m_1,\ l_2+m_2,\ldots,\ l_n+m_n\}$. Then we have for any given set $\mathcal{K}=\{k_1,k_2,\ldots k_n\}$:

$$\mathbb{E}[\prod_{i=1}^{n}\Psi_i^{k_i}] = \sum_{\mathcal{L}_1,\mathcal{L}_2\ldots\ :\ \mathcal{L}_1\oplus\mathcal{L}_2\oplus\ldots\mathcal{L}_j\oplus\ldots=\mathcal{K}} \left(\prod_j C^{\mathcal{L}_j}\right)$$

This results from a Taylor expansion of $\exp\left(\log\left\{\mathbb{E}[\exp\sum_{i=1}^{n}\lambda_i\ \Psi_i]\right\}\right)$. Let us consider for example the case of a two-parameter exponential Gibbs statistics:

$$P_{\theta,\lambda}(x) = \frac{1}{Z_{\theta,\lambda}}\ \exp - \theta\, U(x) - \ \lambda\,\Psi(x)$$
$$\text{where } Z_{\theta,\lambda} = \sum_{y\in\Omega}\ \exp\, -\theta\, U(y) - \ \lambda\,\Psi(y) \tag{16}$$

First and second-order cumulants associated to random variables $\Psi_1=U$ and $\Psi_2=\Psi$ are then given by:

$$\begin{array}{ll}
\mathbb{E}_{\theta,\lambda}[U] = C^{1,0} & \mathbb{E}_{\theta,\lambda}[\Psi] = C^{0,1} \\
\mathbb{E}_{\theta,\lambda}[U^2] = C^{2,0} + C^{1,0}\ .\ C^{1,0} & \textit{i.e. } C^{2,0} = \mathrm{var}_{\theta,\lambda}(U) \\
\mathbb{E}_{\theta,\lambda}[\Psi^2] = C^{0,2} + C^{0,1}\ .\ C^{0,1} & \textit{i.e. } C^{0,2} = \mathrm{var}_{\theta,\lambda}(\Psi) \\
\mathbb{E}_{\theta,\lambda}[\Psi U] = C^{1,1} + C^{1,0}\ .\ C^{0,1} & \textit{i.e. } C^{1,1} = \mathrm{cov}_{\theta,\lambda}(U,\Psi)
\end{array}$$

Going back as before to the expansion of the partition function associated to this two-parameter distribution, we compute:

$$\begin{aligned}
Z_{\theta+\delta\theta,\lambda+\delta\lambda} &= \sum_{y\in\Omega}\exp - \theta\, U(y) - \ \lambda\,\Psi(y)\exp\, -\delta\theta\, U(y) - \delta\lambda\,\Psi(y) \\
&= Z_{\theta,\lambda}\ .\ \mathbb{E}_{\theta,\lambda}[\exp -\delta\theta\, U - \delta\lambda\,\Psi] \\
&= Z_{\theta,\lambda}\ .\ \exp\,(\ -\delta\theta\,\mathbb{E}_{\theta,\lambda}[U] - \delta\lambda\,\mathbb{E}_{\theta,\lambda}[\Psi] \\
&\qquad + \frac{1}{2}\ \{\ \delta\theta^2\ \mathrm{var}_{\theta,\lambda}(U) + \delta\lambda^2\ \mathrm{var}_{\theta,\lambda}(\Psi) + 2\ \delta\theta\ .\ \delta\lambda\ \mathrm{cov}_{\theta,\lambda}(U,\Psi)\ \} + \ldots\ )
\end{aligned}$$

which gives as before a Taylor expansion of $\log Z_{(.,.)}$ around (θ,λ), yielding

$$
\begin{aligned}
\frac{\partial}{\partial\theta} \log Z_{\theta,\lambda} &= -\ \mathbb{E}_{\theta,\lambda}[U] && (17)\\
\frac{\partial}{\partial\lambda} \log Z_{\theta,\lambda} &= -\ \mathbb{E}_{\theta,\lambda}[\Psi] && (18)\\
\frac{\partial^2 \log Z_{\theta,\lambda}}{\partial\theta^2} &= \mathrm{var}_{\theta,\lambda}(U) && (19)\\
\frac{\partial^2 \log Z_{\theta,\lambda}}{\partial\lambda^2} &= \mathrm{var}_{\theta,\lambda}(\Psi) && (20)\\
\frac{\partial^2 \log Z_{\theta,\lambda}}{\partial\theta\partial\lambda} &= \mathrm{cov}_{\theta,\lambda}(U,\Psi) && (21)
\end{aligned}
$$

Let us note that when second parameter λ is fixed, the associated likelihood (16) remains a unimodal function of θ, due to relations (17) and (19) . This will be useful in the following. Another important consequence is that combination of equations (18) and (21) yields

$$\frac{\partial}{\partial\theta}\ \mathbb{E}_{\theta,\lambda}[\Psi] = -\ \mathrm{cov}_{\theta,\lambda}(U,\Psi) = -\left(\ \mathbb{E}_{\theta,\lambda}[U\ \Psi] - \mathbb{E}_{\theta,\lambda}[U]\ \mathbb{E}_{\theta,\lambda}[\Psi]\ \right)$$

a relation valid for all values of parameters λ and θ. At the limit: $\lambda \to 0$ we obtain the scalar Gibbs distribution with parameter θ (assuming everything is continuous ..), and thus related formula:

$$\frac{\partial}{\partial\theta}\ \mathbb{E}_{\theta}[\Psi] = -\ \mathrm{cov}_{\theta}(U,\ \Psi) = -\left(\ \mathbb{E}_{\theta}[U\ \Psi] - \mathbb{E}_{\theta}[U]\ \mathbb{E}_{\theta}[\Psi]\ \right) \qquad (22)$$

i.e. relation (4) . This is similar to an evolution equation in quantum mechanics. This relation is also extremely useful in classical thermodynamics, implying that the linear response of a statistical system such as a diluted gas to an external pertubation (for example a magnetic field) is linked to the internal correlation properties of the system itself.

4.2.1 Application: how to compute higher order cumulants for single-parameter distributions

Let's indicate how previous formula (22) can be simply used to iteratively find cumulants of order higher than two. We shall derive as an example the order three

cumulant of a single-parameter distribution by noticing that
$\frac{\partial^3 Z_\theta}{\partial \theta^3} = -C_\theta^3 = \frac{\partial}{\partial \theta}\,\mathrm{var}_\theta(U)$ in reason of Taylor expansion (15) . Since
$\frac{\partial}{\partial \theta}\,\mathrm{var}_\theta(U) = \frac{\partial}{\partial \theta}\,(\mathbb{E}_\theta[U^2] - (\mathbb{E}_\theta[U])^2)$ and by assigning then successively Ψ as U^2 and U in formula (22) we obtain $\frac{\partial}{\partial \theta}\,\mathrm{var}_\theta(U) = -\,\mathrm{cov}_\theta(U^2, U) + 2\mathbb{E}_\theta[U]\,\mathrm{var}_\theta(U)$ and thus

$$C_\theta^3 = \mathrm{cov}_\theta(U^2, U) - 2\mathbb{E}_\theta[U]\,\mathrm{var}_\theta(U) = \mathbb{E}_\theta[U^3] - 3\mathbb{E}_\theta[U]\,\mathbb{E}_\theta[U^2] + 2\,(\mathbb{E}_\theta[U])^3$$

which was previously given in footnote 1. A nicer way to procede is as follows:

$$\begin{aligned}
C_\theta^3 &= \frac{\partial}{\partial \theta}\,\mathbb{E}_\theta[(U - \mathbb{E}_\theta[U])^2] = \mathrm{cov}_\theta((U - \mathbb{E}_\theta[U])^2, U) \\
&= \mathrm{cov}_\theta((U - \mathbb{E}_\theta[U])^2, U - \mathbb{E}_\theta[U] + \mathbb{E}_\theta[U]) \\
&= \mathrm{cov}_\theta((U - \mathbb{E}_\theta[U])^2, U - \mathbb{E}_\theta[U]) = \mathbb{E}_\theta[(U - \mathbb{E}_\theta[U])^3]
\end{aligned}$$

since for any random variable X and constant a, $\mathrm{cov}_\theta(X, a) = 0$ and since obviously $\mathbb{E}_\theta[U - \mathbb{E}_\theta[U]] = 0$. This will prove to be useful in the following sections.

5 Hyperparameter Estimation

5.1 Hyperparameter estimation for complete data

We recall briefly in this section the main properties which result from previous sections and are known from literature [Younes-88]. Younes has shown that in order to find the (unique) hyperparameter value $\hat{\theta}$ satisfying $\mathbb{E}_{\hat{\theta}}[\Phi(.)] = \Phi(d)$, the following algorithm:

$$\theta_{n+1} = \theta_n + \frac{\Phi(X_{n+1}) - \Phi(d)}{(n+1)\,V} \tag{23}$$

converges almost surely, where X_{n+1} is a sample generated for hyperparameter value θ_n and where V is some positive large constant. It can be interpreted as a gradient descent scheme, where $\mathbb{E}_{\theta_n}[\Phi]$ is replaced by the empiric estimate $\Phi(X_{n+1})$.

5.2 Hyperparameter estimation for imperfectly observed fields

We shall try to show in this section that hyperparameter estimation for a given prior regularization model cannot be dissociated from the observation probability density function (pdf) model *i.e.* the attachment to data. Herein, we place ourselves in a continuous framework, meaning that the configuration space is now: $\Omega = \mathbf{R}^S$ For example, we assume a Gaussian attachment to data for sake of simplicity:

$$\Pr(D = d \;/\; F = f) \;=\; \frac{\exp - \lambda \,||\, d - Rf \,||^2}{Z_\lambda}$$

More generally, which we shall note:

$$\Pr(D = d \;/\; F = f) \;= \Pr_{\lambda,d}(f) = \frac{\exp - \lambda \, U_d(f)}{Z_\lambda}$$

where λ a positive cœfficient. Then the *posterior* probability (usually employed in stochastic restoration algorithms such as simulated annealing or maximum of the posterior marginal) of original scene f knowing single observation d and hyperparameter value θ writes:

$$\begin{aligned}
\Pr(F = f \;/\; D = d,\; \Theta = \theta) \;&=\; \frac{\Pr(D = d \;/\; F = f) \;.\; \Pr(F = f \;/\; \Theta = \theta)}{\Pr(D = d)} \\
&\propto\; \exp - \lambda \, U_d(f) - \;\theta \, \Phi(f)
\end{aligned}$$

where $\Phi(f)$ is the regularization prior component of posterior energy,
We shall also assume in the following that every admissibility condition is fulfilled:

$a)$ $\Phi(f) > 0 \;, \; \forall f \in \Omega$

$b)$ $Z_\theta = \int_\Omega \exp - \theta \, \Phi(f) \;\; \mathrm{df} \;<+\infty \;, \; \forall \theta > 0$

$c)$ hence $Z_{\theta,\lambda} = \int_\Omega \exp - \lambda \, U_d(f) \;\exp - \theta \, \Phi(f) \;\; \mathrm{df} \;<+\infty \;, \; \forall \theta > 0$

Then, the posterior probability of hyperparameter θ knowing incomplete observation d writes as:

$$\begin{aligned}
\Pr(\Theta = \theta \;/\; D = d) \;&=\; \int_\Omega \Pr(\Theta = \theta, F = f \;/\; D = d) \;\; \mathrm{df} \\
&=\; \int_\Omega \frac{\Pr(D = d \;/\; \Theta = \theta, F = f) \;.\; \Pr(\Theta = \theta, F = f)}{\Pr(D = d)} \;\; \mathrm{df}
\end{aligned}$$

where $\Pr(D = d)$ is totally independent of any value of f or θ (cf. Bayesian analysis).
Since we assume the observation pdf to be independent of hyperparameter θ, and θ to follow some *a priori* law, it follows :

$$\Pr(\Theta = \theta \,/\, D = d) = \frac{1}{\Pr(D = d)} \int_\Omega \Pr(D = d \,/\, F = f)\ \Pr(F = f \,/\, \Theta = \theta) \Pr(\Theta = \theta)\ \mathrm{df}$$

When no *a priori* information on θ is known (*i.e.* θ follows the uniform distribution on $\mathbf{R}^+$), we have more explicitly:

$$\Pr(\Theta = \theta \,/\, D = d) = \frac{1}{\Pr(D = d)} \int_\Omega \Pr(D = d \,/\, F = f)\ .\ \frac{\exp - \theta\ \Phi(f)}{Z_\theta}\ \mathrm{df} \tag{24}$$

yielding thus:

$$\Pr(\Theta = \theta \,/\, D = d) = \frac{1}{\Pr(D = d)} \frac{\int_\Omega \Pr_{\lambda,d}(f)\ \exp - \theta\ \Phi(f)\ \mathrm{df}}{\int_\Omega \exp - \theta\ \Phi(f)\ \mathrm{df}} \propto \mathbb{E}_\theta[\Pr_{\lambda,d}] \tag{25}$$

where proportionality term $\dfrac{1}{\Pr(D = d)}$ only depends on observation d, and where, as in previous sections, $\mathbb{E}_\theta[.] = \mathbb{E}_{\theta,\lambda=0}[.]$ denotes the statistical expectation *w.r.t.* to the prior Markov random field of energy $\theta\ \Phi(f)$ (notice also that $\Pr(\Theta = \theta \,/\, D = d) \propto \mathbb{E}_\theta[\exp - \lambda \,||\, d - R\ .\ ||^2]$ in the Gaussian case). It follows then that parameter likelihood maximization for incomplete data is equivalent to find the regularization MRF yielding the highest statistical attachment to observed data :

$$\hat{\theta} \ = \ \arg\max \mathbb{E}_\theta[\Pr_{\lambda,d}] = \arg\max \mathbb{E}_\theta[\exp - \lambda \,||\, d - R\ .\ ||^2] \text{ in the Gaussian case.}$$

This seems physically sound in the sense that neighborhood of observed d in configuration space Ω should be "consistent" at most with the regularization properties induced by the Markov random field associated to prior energy $\theta\ \Phi$. Now, previous cumulant analysis, which we shall assume to remain valid in this continuous case (for example when moments of every order of Gibbs distributions exist and are finite), yields that optimal hyperparameter value $\hat{\theta}$ should verify:

$$\left(\frac{\partial \mathbb{E}_\theta[\Pr_{\lambda,d}]}{\partial \theta}\right)_{\hat{\theta}} = -\mathrm{cov}_{\hat{\theta}}(\Pr_{\lambda,d},\ \Phi) = \mathbb{E}_{\hat{\theta}}[\Phi]\ \mathbb{E}_{\hat{\theta}}[\Pr_{\lambda,d}] - \mathbb{E}_{\hat{\theta}}[\Phi\ \Pr_{\lambda,d}] = 0 \tag{26}$$

In the Gaussian observation case this gives:

$$\mathbb{E}_{\hat{\theta}}[\Phi] \, . \, \mathbb{E}_{\hat{\theta}}[\exp \, -\lambda \, || \, d - R \, . \, ||^2] - \mathbb{E}_{\hat{\theta}}[\Phi \, \exp \, -\lambda \, || \, d - R \, . \, ||^2] = 0$$

It could be "roughly" interpreted that optimal hyperparameter value $\hat{\theta}$ enables random variables $\Pr_{\lambda,d}(f)$ and $\Phi(f)$ to be "quasi-independent" *w.r.t.* to the *prior* MRF associated to energy $\hat{\theta}\ \Phi$.
Another way to look at this optimal hyperparameter estimation problem is (thanks to G. Gimel'farb) the following (see also [Younes-89, Zhou-97]) . Let us note

$$L(\theta) = \Pr(\Theta = \theta \ / \ D = d)$$

We start directly from admissibility condition c) and from Eq.(24) in order to notice that:

$$L(\theta) \propto \frac{Z_{\theta,\lambda}}{Z_\lambda \, . \, Z_\theta} \quad (\text{ proportionality term : } \frac{1}{\Pr(D=d)} \,)$$

Thus,

$$\left(\frac{\partial \log L}{\partial \theta} \right)_{\hat{\theta}} = \mathbb{E}_{\hat{\theta}}[\Phi] - \mathbb{E}_{\hat{\theta},\lambda}[\Phi] = 0 \Rightarrow \mathbb{E}_{\hat{\theta}}[\Phi] = \mathbb{E}_{\hat{\theta},\lambda}[\Phi] \qquad (27)$$

where it is recalled that $\mathbb{E}_\theta[.]$ is the expectation under the *prior* MRF associated to energy function $\theta\ \Phi(f)$ and that $\mathbb{E}_{\theta,\lambda}[.]$ means statistical expectation under the posterior Markov random field of energy $\mathcal{U}(f) = \lambda \, || \, d - Rf \, ||^2 + \theta \, \Phi(f)$.
This is very similar to an approach used in unsupervised Bayesian-Markovian segmentation [Gimel'farb-97]. Notice also that due to a preceding remark in section 4.2 , both expectations $\mathbb{E}_{\theta,\lambda}[\Phi]$ and $\mathbb{E}_\theta[\Phi]$ are monotonously decreasing functions of θ, implying thus that *several* optimal hyperparameter values can be found (See Fig. 5 for example). To our opinion a study of higher order derivatives of these expectations following cumulant analysis is needed. It will be carried out afterwards.

5.2.1 At the limit $\lambda \to +\infty$: revisiting the complete data case

Let us also give a hint helping to justify the coherence of Eq. (26) and (27) . Assume that matrix $R = \mathbf{1}$ for sake of simplicity and that at the limit $\lambda \to +\infty$, one has: $\Pr_{\lambda,d}(f) \to \delta(d-f)$ in the sense of distributions (this of course should be developed and checked). This means that observation d becomes now *complete* since it is the only configuration to occur with nonzero probability, which is indeed the case for

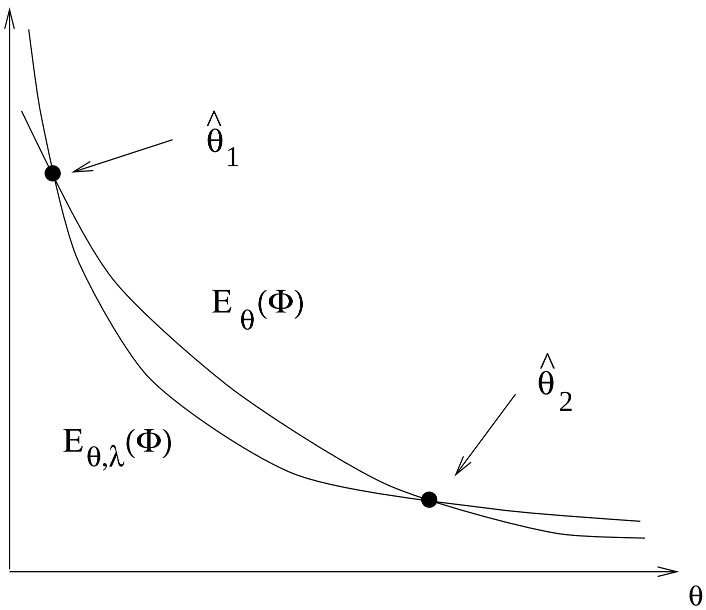


Figure 5: Estimation of optimal regularization parameter value(s) $\hat{\theta}$ for incomplete data.

previous Gaussian distribution when standard deviation comes down to 0. Let us investigate first Eq. (27) which is simpler. It comes then:

$$\mathbb{E}_{\hat{\theta},\lambda}[\Phi] \rightarrow \frac{\int_\Omega \delta(d-f)\ \Phi(f) \exp -\hat{\theta}\,\Phi(f)\ \ \mathrm{df}}{\int_\Omega \delta(d-f)\ \exp -\hat{\theta}\,\Phi(f)\ \ \mathrm{df}} = \Phi(d)$$

leading thus to

$$\mathbb{E}_{\hat{\theta}}[\Phi] = \Phi(d) \tag{28}$$

i.e. the classical Maximum Likelihood principle (5) for parameter θ and energy Φ.

In the same line, Eq.(26) yields:

$$\begin{aligned} \mathbb{E}_{\hat\theta}[\mathrm{Pr}_{\lambda,d}\,.\,\Phi] &\rightarrow \frac{\int_\Omega \delta(d-f)\;\Phi(f)\;\exp - \hat\theta\,\Phi(f)\;\;\mathrm{df}}{\int_\Omega \exp - \hat\theta\,\Phi(f)\;\;\mathrm{df}} \\ &= \frac{\Phi(d)\;\exp - \hat\theta\;\Phi(d)}{Z_{\hat\theta}} \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}_{\hat\theta}[\mathrm{Pr}_{\lambda,d}]\,.\,\mathbb{E}_{\hat\theta}[\Phi] &\rightarrow \frac{\int_\Omega \delta(d-f)\;\exp - \hat\theta\,\Phi(f)\;\;\mathrm{df}}{\int_\Omega \exp - \hat\theta\,\Phi(f)\;\;\mathrm{df}}\,.\,\mathbb{E}_{\hat\theta}[\Phi] \\ &= \frac{\exp - \hat\theta\;\Phi(d)}{Z_{\hat\theta}}\,.\,\mathbb{E}_{\hat\theta}[\Phi] \end{aligned}$$

Equating two members of the covariance equations yields thus also (28) , due to admissibility conditions $\exp\, - \theta\;\Phi(d) > 0$ and $Z_{\hat\theta} = \int_\Omega \exp\, - \hat\theta\,\Phi(f)\;\;\mathrm{df}\; < +\infty$. We observe that second method (27) yields more easily Maximum Likelihood principle than first one (26) , although more complicate to simulate since the first one only needs to estimate expectations under the regularizing field.

5.3 Towards a stochastic gradient-like algorithm for imperfectly observed fi elds

We should devise a Newton-Raphson-like method for estimating the optimal hyperparameter value $\hat\theta$ following both previous methods. Recall that pure Newton-Raphson iterative scheme for optimizing some criterion $F(\theta)$ could be written as:

$$\theta_{n+1} = \theta_n - \frac{F'(\theta_n)}{F''(\theta_n)} \tag{29}$$

which requires to compute (or at least to estimate) the second-order derivative of the criterion to be maximized *w.r.t.* hyperparameter.

5.3.1 The Gimel'farb approach od Eq. (27)

Newton-Raphson scheme leads straightforwardly to:

$$\theta_{n+1} = \theta_n - \frac{\mathbb{E}_{\theta_n}[\Phi] - \mathbb{E}_{\theta_n,\lambda}[\Phi]}{\dfrac{\partial}{\partial \theta_n}\left(\mathbb{E}_{\theta_n}[\Phi] - \mathbb{E}_{\theta_n,\lambda}[\Phi]\right)} = \theta_n + \frac{\mathbb{E}_{\theta_n}[\Phi] - \mathbb{E}_{\theta_n,\lambda}[\Phi]}{\mathrm{var}_{\theta_n}(\Phi) - \mathrm{var}_{\theta_n,\lambda}(\Phi)} \tag{30}$$

It is of course the most simply formulated method, but needs in general to estimate statistical expectations and covariances under two MRF: the pure regularization (prior) one and the posterior one with attachment to data. However, we shall see further on (see subsection 5.4) that on many usual circumstances the prior expectation and variance of potential Φ can be exactly computed as functions of regularization hyperparameter θ. Noting therefore $< X >_{\theta_n,\lambda}$, the posterior empirical statistical average of quantity X under hyperparameters $\{\theta_n, \lambda\}$, previous equations writes as:

$$\theta_{n+1} = \theta_n + \frac{\mathbb{E}_{\theta_n}[\Phi] - < \Phi >_{\theta_n,\lambda}}{\mathrm{var}_{\theta_n}(\Phi) - < \mathrm{var}\ \Phi >_{\theta_n,\lambda}} \tag{31}$$

Notice the following sign condition at optimal value:

$$-\left(\frac{\partial^2 L}{\partial \theta^2}\right)_{\hat{\theta}} \simeq \mathrm{var}_{\hat{\theta}}(\Phi) - < \mathrm{var}\ \Phi >_{\hat{\theta},\lambda} > 0$$

to ensure at least that a (local) maximum is reached.

We infer that under suitable conditions (log-likelihood monomodality for the data attachment pdf) this condition is always valid. Let us also remark for this purpose that at the limit $\lambda \to +\infty$, $\mathrm{var}_{\hat{\theta},\lambda}(\Phi) \to 0$ since f become a certain variable (so that previous sign condition is automatically fulfilled) and that at the limit $\lambda \to +0$, $\mathrm{var}_{\hat{\theta},\lambda}(\Phi) \to \mathrm{var}_{\theta}(\Phi)$.

Previous equation (30) should be compared with the stochastic gradient approach of Younes *et al.* [Younes-88] by noticing once again that at the limit $\lambda \to +\infty$, it becomes

$$\theta_{n+1} = \theta_n + \frac{\mathbb{E}_{\theta_n}[\Phi]\ - \Phi(d)}{\mathrm{var}_{\theta_n}(\Phi)}$$

to be compared with (23) (assuming $\mathbb{E}_{\theta_n}[\Phi] \simeq \Phi(X_{n+1})$ as previously).

5.3.2 The covariance approach Eq. (26)

We shall in this case compute the second derivative $\dfrac{\partial^2 L}{\partial \theta^2}$ with the help of previous cumulant analysis results up to third order. As a matter of fact:

$$\begin{aligned}
\frac{\partial^2 L}{\partial \theta^2} &= -\frac{\partial}{\partial \theta} \operatorname{cov}_\theta(\mathrm{Pr}_{\lambda,d},\ \Phi) = \frac{\partial}{\partial \theta} \left(\mathbb{E}_\theta[\mathrm{Pr}_{\lambda,d}] \,.\, \mathbb{E}_\theta[\Phi] - \mathbb{E}_\theta[\mathrm{Pr}_{\lambda,d} \,.\, \Phi] \right) \\
&= \operatorname{cov}_\theta(\mathrm{Pr}_{\lambda,d} \,.\, \Phi,\ \Phi) - \mathbb{E}_\theta[\Phi] \,.\, \frac{\partial}{\partial \theta} \mathbb{E}_\theta[\mathrm{Pr}_{\lambda,d}] - \mathbb{E}_\theta[\mathrm{Pr}_{\lambda,d}] \,.\, \operatorname{var}_\theta(\Phi)
\end{aligned}$$

So here we have more covariance terms to estimate, but under the regularizing MRF only. It could result in our opinion into a lower computational cost.
Let us also notice that for optimal hyperparameter value, the second part of the right member of last equation should vanish. This should yield a sign condition:

$$\left(\frac{\partial^2 L}{\partial \theta^2} \right)_{\hat{\theta}} = \operatorname{cov}_{\hat{\theta}}(\mathrm{Pr}_{\lambda,d} \,.\, \Phi,\ \Phi) - \mathbb{E}_{\hat{\theta}}[\mathrm{Pr}_{\lambda,d}] \,.\, \operatorname{var}_{\hat{\theta}}(\Phi) < 0$$

to ensure at least that a (local) maximum is reached.

5.3.3 An alternative: the Monte Carlo Markov Chain (MCMC) approach

An alternative to previous generalized stochastic gradient presentation was developed in [Descombes-96] for hyperparameter estimation purpose. It consists in evaluating the dependence of partition function *w.r.t.* to slight hyperparameter changes from different samples of unperturbed hamiltonian. We also refer to the excellent book by C. Robert [Robert-96] which makes the most recent point on this technique.

5.4 The case of homogeneous potentials

An important case (also coming from Statistical Physics) occurs when the regularization potential is an homogeneous function, *i.e.* :

$$\Phi(\nu\ g) = \nu^\alpha\ \Phi(g)\ ,\ \forall g \in \Omega \text{ and } \nu \geq 0 \tag{32}$$

where α is the homogeneity degree of Φ. The most classical example by far is the Gaussian model for which $\Phi(f) = \sum_{(r,s) \in \mathcal{C}} (f_r - f_s)^2$ and $\alpha = 2$.

Other potentials with homogeneity degree $\alpha \leq 1$ can be used in order to preserve discontinuities between boundaries in image processing. We can now compute the partition function and its derivatives for such potentials using the auxiliar variable change $f = \nu\, g$:

$$\begin{aligned} Z_\theta &= \int_\Omega \exp - \theta\, \Phi(f)\ \mathrm{df} = \int_\Omega \exp - \theta\, \Phi(\nu\, g)\ .\ \nu^{|S|}\ \mathrm{dg} \\ &= \nu^{|S|}\ .\ \int_\Omega \exp - \nu^\alpha\, \theta\, \Phi(g)\ \mathrm{dg} = \nu^{|S|}\, Z_{\theta \nu^\alpha} \end{aligned}$$

The partition function considered as a function of hyperparameter θ is thus also homogeneous and applying Euler formula (which is equivalent to expand previous equation around $\nu = 1$) we find:

$$\theta \left(\frac{\partial Z_\theta}{\partial \theta} \right)_\theta = - \frac{|S|}{\alpha}\, Z_\theta \ \ \forall\, \theta > 0$$

i.e. :

$$- \frac{\partial}{\partial \theta} \log Z_\theta = \mathbb{E}_\theta[\Phi] = \frac{|S|}{\alpha}\, \frac{1}{\theta} \tag{33}$$

In a similar way we find:

$$- \frac{\partial^2 \log Z_\theta}{\partial \theta^2} = \mathrm{var}_\theta(\Phi) = \frac{|S|}{\alpha}\, \frac{1}{\theta^2} \tag{34}$$

This is very interesting since we need no more to estimate previous expectation and variance of Φ *w.r.t.* prior model in previous stochastic optimization approaches.

5.5 A synthetic comparison between different estimation variants

5.5.1 Joint probability of observation and result conditionally to hyperparameter [Lakshmanan-89, Descombes-96]

Suppose that (at some step) the optimal restoration image f^* is known. Thus we can write: $\hat{\theta} = \arg\max_\theta \Pr(F = f^*\ ,\ D = d\ /\ \Theta = \theta)$. Since

$$\begin{aligned} \Pr(F = f^*\ ,\ D = d\ /\ \Theta = \theta) &= \Pr(D = d\ /\ F = f^*\ ,\ \Theta = \theta)\ \Pr(F = f^*\ /\ \Theta = \theta) \\ &= \Pr(D = d\ /\ F = f^*)\ \Pr(F = f^*\ /\ \Theta = \theta) \end{aligned}$$

due to the independence of observation model *w.r.t.* hyperparameter θ , one has:

$$\hat{\theta} = \arg\max_{\theta} \Pr(F = f^* \ / \ \Theta = \theta) = \frac{\exp \ -\theta \ \Phi(f^*)}{[\ Z_\theta = \int_\Omega \exp \ -\theta \ \Phi(f) \ \ \mathrm{df} \ \ ]}$$

i.e. the Maximum Likelihood estimate of prior MRF with energy $\theta \ \Phi(f)$ and for the complete data f^*, given by:

$$\mathbb{E}_{\hat{\theta}}[\Phi] = \Phi(f^*) \tag{35}$$

5.5.2 Conditional probability of result *w.r.t.* to observation and hyperparameter

In this case we have, using Bayes formula:

$$\begin{aligned}
\hat{\theta} &= \arg\max_{\theta} \Pr(F = f^* \ / \ D = d \ , \ \Theta = \theta) \\
&= \arg\max_{\theta} \frac{\Pr(D = d \ / \ F = f^* \ , \ \Theta = \theta) \ \Pr(F = f^* \ / \ \Theta = \theta)}{\int_\Omega \Pr(D = d \ / \ F = f \ , \ \Theta = \theta) \ \Pr(F = f \ / \ \Theta = \theta) \ \ \mathrm{df}} \\
&= \arg\max_{\theta} \frac{\exp \ -\lambda \ || \ d - Rf^* \ ||^2 - \ \theta \ \Phi(f^*)}{[\ Z_{\theta,\lambda} = \int_\Omega \exp \ -\lambda \ || \ d - Rf \ ||^2 - \ \theta \ \Phi(f) \ \ \mathrm{df} \ \ ]}
\end{aligned}$$

i.e. the Maximum Likelihood estimate of posterior MRF with energy $\theta \ \Phi(f) + \lambda \ || \ d - Rf \ ||^2$ and for the complete data f^*, given by:

$$\mathbb{E}_{\hat{\theta},\lambda}[\Phi] = \Phi(f^*) \tag{36}$$

Due to previous unimodality results of partial and compound log-likelihood, both methods can be implemented with a stochastic gradient [Younes-88] or MCMC algorithm [Descombes-96] .

5.5.3 Probability of hyperparameter conditionally to observation only

This is the real incomplete data case we developed before:

$$\mathbb{E}_{\hat{\theta},\lambda}[\Phi] = \mathbb{E}_{\hat{\theta}}[\Phi] \tag{37}$$

See also [Younes-89] .

6 A method for simultaneous restoration and hyperparameter estimation

In this section, we present results in restoration for the homogeneous potential case. We simultaneously perform hyperparameter estimation and image restoration, since a Large Numbers law relative to inhomogeneous sampling and allows us to perform the Maximum of Posterior Marginal estimate for the pixel image to be restored. In our case we chose to implement the Posterior Mean estimate [3] , *i.e.* averaging on the samples obtained by the generalized stochastic gradient method since sufficient convergence criterium for the hyperparameter is obtained.

6.1 Principle

generalized stochastic gradient implementation We assume in the following the homogeneous potential case. At each iteration, the prior MRF characteristics $\mathbb{E}_{\theta_n}[\Phi]$ and $\mathrm{var}_{\theta_n}(\Phi)$ are estimated from (33) and (34) , so just the posterior MRF is simulated and sampled. From Eq. (31), the increment of hyperparameter at iteration n is given by:

$$\theta_{n+1} - \theta_n = \frac{\mathbb{E}_{\theta_n}[\Phi] - < \Phi >_{\theta_n,\lambda}}{V * (\mathrm{var}_{\theta_n}(\Phi) - < \mathrm{var}\ \Phi >_{\theta_n,\lambda})}$$

[3]on the convergence of estimates:

- if hyperparameter $\theta \neq 0$ then previous Dobrushin's theorem ensures that inhomogenous Markov sampling will converge to limit invariant law P_θ

$$c(\prod(Q_n)) \leq \prod(1 - \exp(-\Delta\theta_n)) \to 0 \ \ (\text{ since } T = 1)$$

- then naively, if X_n are samples of P_{θ_n}, then

$$\mathbb{E}[\frac{\sum_{n=1}^{N} \Phi(X_n)}{N}] \to \frac{\sum_{n=1}^{N} \mathbb{E}_{\theta_n}[\Phi(X)]}{N} \to \mathbb{E}_\theta[\Phi(X)] \text{ (Cesaro)}$$

This justifies the posterior mean estimate for pixel image f. Notice however that samples are not independent due to the Markov Chain principle itself.

which we put for statistical purpose in the following form:

$$\theta_{n+1}-\theta_n = \frac{|S| * (\mathbb{E}_{\theta_n}[\phi] - \mathbb{E}_{\theta_n,\lambda}[\phi])}{V * |S|^2 * (\text{var}_{\theta_n}(\phi) - < \text{var}\ \phi >_{\theta_n,\lambda})} = \frac{\mathbb{E}_{\theta_n}[\phi] - \mathbb{E}_{\theta_n,\lambda}[\phi]}{V * |S| * (\text{var}_{\theta_n}(\phi) - < \text{var}\ \phi >_{\theta_n,\lambda})}$$

where $\phi = \dfrac{\Phi}{|S|}$ is the reduced regularization potential by site. This writes more explicitly as:

$$\theta_{n+1} - \theta_n = \frac{\dfrac{1}{\alpha\theta_n} - < \phi >_{\theta_n,\lambda}}{V * (\dfrac{1}{\alpha\theta_n^2} - |S|\ .\ < \text{var}\ \phi >_{\theta_n,\lambda})}$$

since obviously: $|S|\ \text{var}_{\theta_n}(\phi) = \dfrac{\text{var}_{\theta_n}(\Phi)}{|S|} = \dfrac{1}{\alpha\theta_n^2}$. As a matter of fact, the statistical quantities which can be estimated during any posterior Gibbs simulation are precisely related to reduced potential ϕ, since this quantity can be empirically estimated on any convenient window (sub-image), provided the window should be statistically representative of the whole image. Reduced empiric average and variance are then easily computed at each statistical step by letting several Metropolis sweeps (noted N, with $N \geq 2$) run with the same hyperparameter value. For example the reduced empirical mean at iteration n is computed as: $< \phi >_{\theta_n,\lambda}= \dfrac{\sum_{p=1}^{N} \phi(X_n^{(p)})}{N}$ (we could chose $\phi(X_n^{(N)})$ as well, *i.e.* the reduced potential of the last sample so generated, but experimental results proved better convergence of the estimates in the former case). Moreover, we experimentally found that precise convergence restoration results were obtained when performing statistical averages on the <u>whole</u> image. The results presented hereafter were thus all obtained with this condition.

Generally speaking, the Younes factor V can be taken as $V = A{*}(n{+}1)^\gamma$. Exponent γ may be either eqal to 0 (constant step, deterministic-like, gradient algorithm), to 1 (the Younes case), or strictly speaking such that the series $u_n = \dfrac{1}{(n+1)^\gamma}$ diverges whereas the series $v_n = u_n^2$ converges (see general conditions in [Metivier-87]) . We choose for example $\gamma = \dfrac{2}{3}$ which leads to good results (the excursion of hyperparameter variation may be large while preserving final convergence) The normalization constant A may be chosen by user. However we found that assigning A to value 1.0 with previous exponent $\gamma \in [\dfrac{1}{2}, 1]$ leads to convenient results in all cases.

initial guess of hyperparameter θ^0 Assuming an homogeneous regularization potential with exponent α and substituting (crudely !) $\mathbb{E}_{\theta,\lambda}[\Phi]$ by its empiric realization $\Phi(d)$ in (27) yields $\dfrac{|S|}{\alpha}\,\dfrac{1}{\theta^0} = \Phi(d)$, *i.e.* $\theta^0 = \dfrac{1}{\alpha\,\phi(d)}$. We assume eight-neighbor connectivity in all examples.

posterior mean estimate averaging We decide here to include Metropolis sample X_n for posterior mean estimate when the relative statistical difference $\left|\dfrac{\mathbb{E}_{\theta_n}[\phi] - <\phi>_{\theta_n,\lambda}}{\mathbb{E}_{\theta_n}[\phi]}\right|$ is below some threshold (lying typically in the range $[0.10,\ 0.15]$).

6.2 Results

quantitative result analysis To compare the results on several synthetic images generated with different signal-to-noise ratios we perform statistical tests (namely empirical mean μ and standard deviation σ) in two selected regions:

Region	X	Y	width	height
R_1:	100	0	40	20
R_2:	160	30	40	20

Results Nos. A, B and C (shown on respective Figs. 6, 7 and 8) correspond to initial and noisy image shown on top of Fig. 6. Result D corresponds to the noisy image shown on top of Fig. 9.

They remain of course to be compare with results from other methods [Charbonnier-96, Khoumri-97]).

Result A

Initial guess of hyperparameter $\theta^0 = 0.097$.

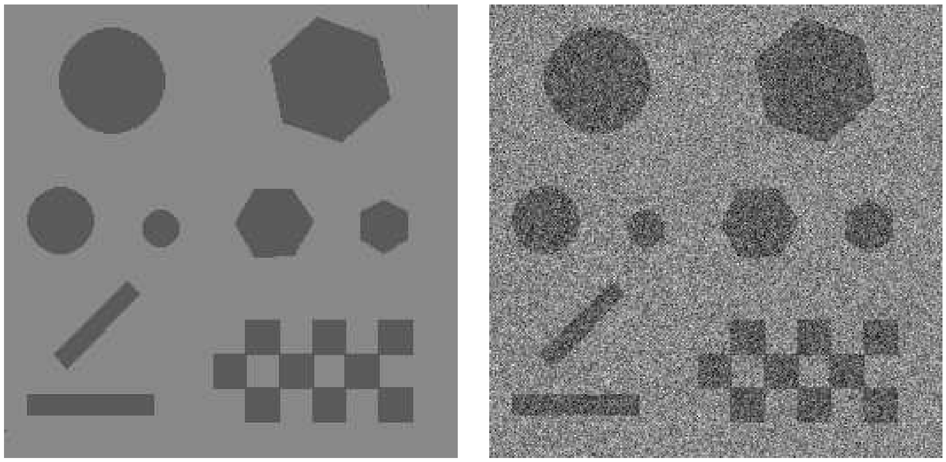
Conditions

Homogenenous potential exponent	$\alpha = \frac{1}{2}$
Younes factor at step n:	$V = A * (n+1)^\gamma, \;\; A = 1.0, \;\; \gamma = \frac{2}{3}$
Estimating empiric potential mean and	variances every 5 Metropolis sweeps
Posterior mean threshold	$\left\| \frac{\mathbb{E}_{\theta_n}[\phi] - < \phi >_{\theta_n,\lambda}}{\mathbb{E}_{\theta_n}[\phi]} \right\| < 0.10$
Total number of such averaging steps:	40

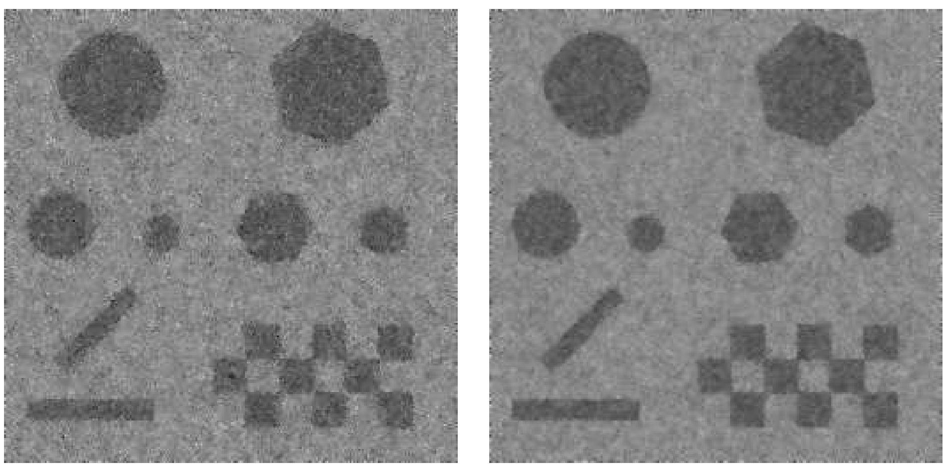
Results 225 Metropolis iterations - Final hyperparameter estimate $\theta^* = 0.159$

	Initial image	Last Metropolis sample	Posterior Mean estimate
Region R_1:	μ = 135.744 σ = 27.799	μ = 137.080 σ = 15.139	μ = 136.613 σ = 10.886
Region R_2:	μ = 90.969 σ = 27.118	μ = 88.929 σ = 12.355	μ = 89.865 σ = 7.857

Last Metropolis sweep:
prior energy/site = 12.626 - empiric posterior energy/site = 12.220
prior normalized variance = 79.7 - empiric normalized posterior variance = 22.4



Original image | Noisy image



Last Metropolis sample | Posterior mean image

Figure 6: Result A

Result B

Initial guess of hyperparameter $\theta^0 = 0.097$.

Conditions

Homogenenous potential exponent	$\alpha = \frac{1}{2}$
Younes factor at step n:	$V = A * (n+1)^\gamma, \ \ A = 1.0, \ \ \gamma = \frac{2}{3}$
Estimating empiric potential mean and	variances every 5 Metropolis sweeps
Posterior mean threshold	$\left\lvert \frac{\mathbb{E}_{\theta_n}[\phi] - <\phi>_{\theta_n,\lambda}}{\mathbb{E}_{\theta_n}[\phi]} \right\rvert < 0.10$
Total number of such averaging steps:	100

Results 535 Metropolis iterations - Final hyperparameter estimate $\theta^* = 0.177$

	Initial image	Last Metropolis sample	Posterior Mean estimate
Region R_1:	μ = 135.744 σ = 27.799	μ = 136.916 σ = 11.947	μ = 136.164 σ = 9.347
Region R_2:	μ = 90.969 σ = 27.118	μ = 88.675 σ =8.978	μ = 88.942 σ = 5.420

Last Metropolis sweep:
prior energy/site = 11.288 - empiric posterior energy/site = 11.042.
prior normalized variance = 63.7 - empiric normalized posterior variance = 2.42

Result C

Initial guess of hyperparameter $\theta^0 = 0.097.$

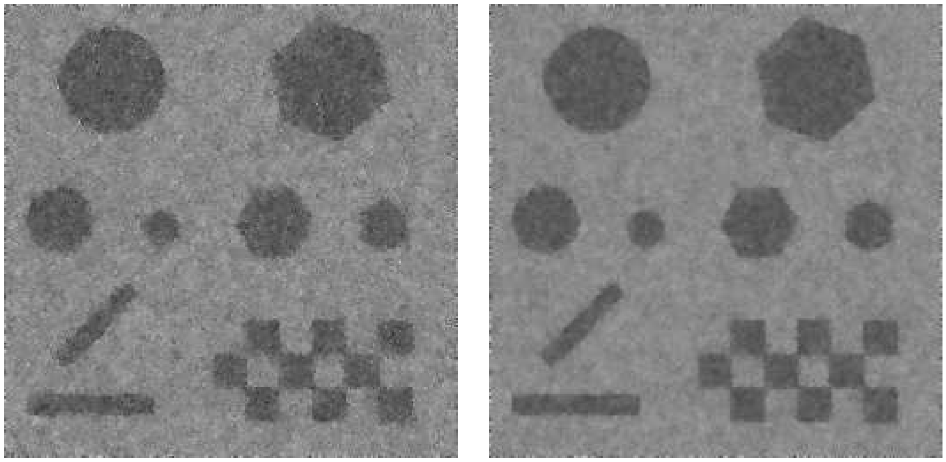
Conditions

Homogenenous potential exponent	$\alpha = \frac{1}{2}$
Younes factor at step n:	$V = A * (n+1)^\gamma, \;\; A = 1.0, \;\; \gamma = \frac{2}{3}$
Estimating empiric potential mean and	variances every 2 Metropolis sweeps
Posterior mean threshold	$\left\lvert \frac{\mathbb{E}_{\theta_n}[\phi] - < \phi >_{\theta_n,\lambda}}{\mathbb{E}_{\theta_n}[\phi]} \right\rvert < 0.10$
Total number of such averaging steps:	400

Results 842 Metropolis iterations - Final hyperparameter estimate $\theta^* = 0.193$

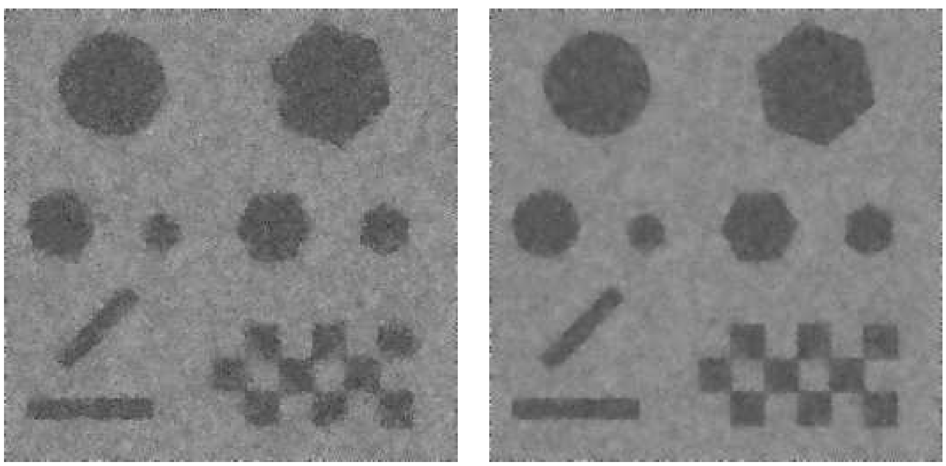
	Initial image	Last Metropolis sample	Posterior Mean estimate
Region R_1:	μ = 135.744 σ = 27.799	μ = 136.671 σ = 11.949	μ = 136.254 σ = 8.744
Region R_2:	μ = 90.969 σ = 27.118	μ = 88.160 σ = 8.709	μ = 89.654 σ = 4.901

Last Metropolis sweep:
prior energy/site = 10.378 - empiric posterior energy/site = 10.177
prior normalized variance = 53.9 - empiric normalized posterior variance = 0.0942



Last Metropolis sample Posterior mean image

Figure 7: Result B



Last Metropolis sample Posterior mean image

Figure 8: Result C

Result D

Initial guess of hyperparameter $\theta^0 = 0.0968$

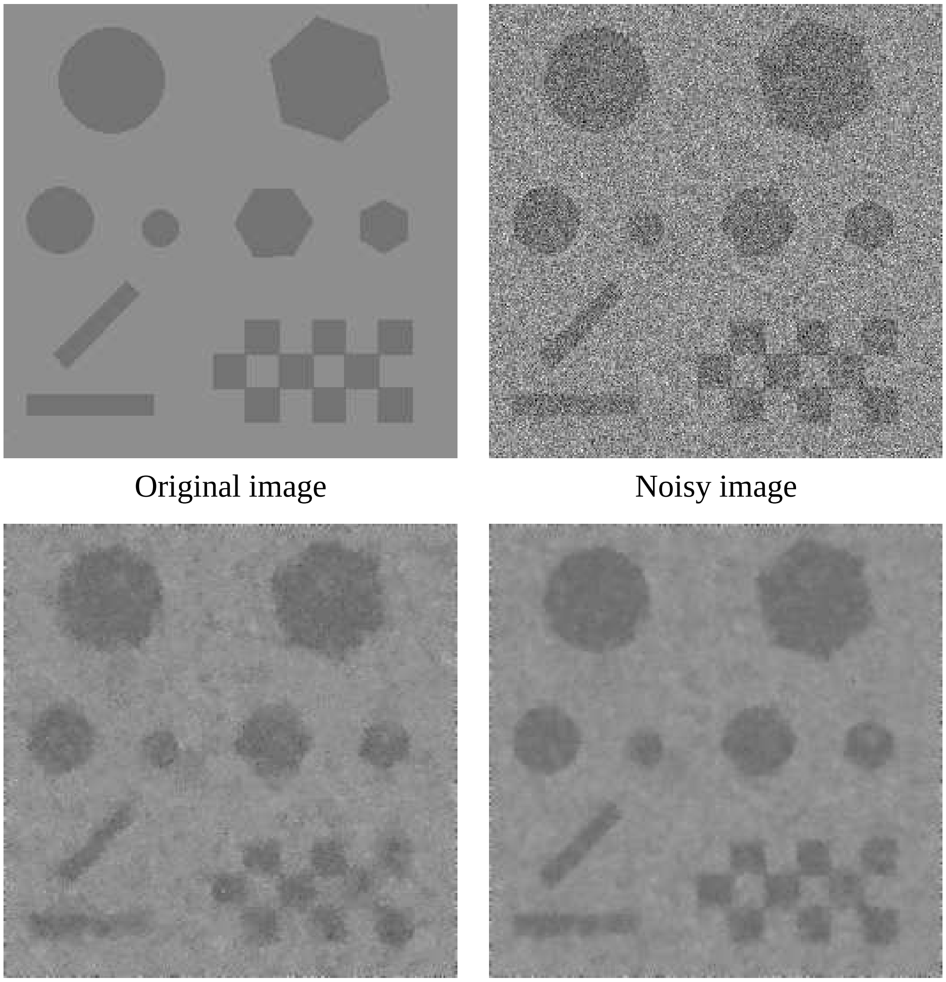
Conditions

Homogenenous potential exponent	$\alpha = \frac{1}{2}$
Younes factor at step n:	$V = A * (n+1)^\gamma, \;\; A = 1.0, \;\; \gamma = \frac{2}{3}$
Estimating empiric potential mean and	variances every 2 Metropolis sweeps
Posterior mean threshold	$\left\| \frac{\mathbb{E}_{\theta_n}[\phi] - < \phi >_{\theta_n,\lambda}}{\mathbb{E}_{\theta_n}[\phi]} \right\| < 0.10$
Total number of such averaging steps:	400

Results 842 Metropolis iterations - Final hyperparameter estimate $\theta^* = 0.238$

	Initial image	Last Metropolis sample	Posterior Mean estimate
Region R_1:	μ = 142.645 σ = 28.154	μ = 144.056 σ = 8.562	μ = 142.946 σ = 7.793
Region R_2:	μ = 114.873 σ = 29.638	μ = 112.811 σ = 6.138	μ = 113.535 σ = 4.197

Last Metropolis sweep:
prior energy/site = 8.426 - empiric posterior energy/site = 7.975
prior normalized variance = 35.5 - empiric normalized posterior variance = 0.224



Original image | Noisy image

Last Metropolis sample | Posterior mean image

Figure 9: Result D

7 Extending to restoration with boundary processes

7.1 Presentation of the model

Our purpose is to extend previous approach to simultanoues hyperparameter estimation and image restoration while preserving discontinuities. Two main variants are possible at this point:
• extending hyperparameter estimation to regularization potentials such as Φ-functions, which are known to preserve boundaries, but which contain an intrinsic non-linear dependence *w.r.t.* to hyperparameters [Geman-85, Geman-92]. Such a typical form is

$$\Phi_{rs}(f_r,\ f_r) = \frac{\mu}{1 + \left(\dfrac{f_r - f_s}{\delta}\right)^2} \quad ,$$

where μ and δ are the hyperparameters of the model. This is called the implicit boundary approach, since boundary values can be iteratively addressed and estimated during restoration algorithms (see for example [Charbonnier-96]). Classical stochastic gradient algorithms for hyperparameter estimation must then be adapted to this non-linear framework. This problem was first raised in [Younes-89] and is thoroughly investigated in [Khoumri-97].
• preserving the linear dependence framework by a complementary set of random variables, namely the boundary process, to the initial pixel process, and a related regularization potential. This is called the explicit boundary framework [Geman-84, Geman-90]. We shall develop here this latter approach, namely the explicit boundary framework. We thus add to the set of intensities to be restored a set of complementary (boundary) variables noted $b = \{b_{rs}\}_{(r,s)\in\mathcal{C}}$. They can be either binary or continuous (in $\mathbf{R}^+$). We shall assume in the following the binary case for boundary for sake of simplicity, and we shall note $\mathcal{B} = \{0,1\}^{\mathcal{C}}$ the set of all boundary configurations. Even so, the marginal energy associated to pixels is sufficiently similar to Geman-McClure Φ-function to be interesting enough, as we shall see further on. Moreover, nonlinear estimation of the Φ-function hyperparameters (associated to implicit boundary values) will now become a set of linear hyperparameters estimations within this framework.
One must as in previous section write first the posterior probability density function of parameters (intensities, boundaries) knowing various hyperparameters, in order

to compute the explicit likelihood of hyperparameters. We present thus the different contributing parts to the joint posterior pdf of boundaries-intensities.

7.1.1 The observation probability knowing initial scene and boundary process

It is assumed that observation model neither depends on boundary process value b, nor on other hyperparameters such as regularization cœfficient Θ etc.. :

$$\Pr(D = d \ / \ F = f, B = b, \text{ hyperparameters }) = \frac{\exp - \lambda \mid\mid d - Rf \mid\mid^2}{Z_\lambda}$$

7.1.2 The prior probability of intensity knowing boundary value and regularization hyperparameter

It is assumed to be written as:

$$\Pr(F = f \ / \ B = b, \ \Theta = \theta) = \frac{\exp - \sum_{(r,s)\in\mathcal{C}} b_{rs}\Phi_{rs}(f)}{Z_{\theta/b}}$$

$$\text{where } Z_{\theta/b} = \int_\Omega \exp - \sum_{(r,s)\in\mathcal{C}} b_{rs}\Phi_{rs}(f) \ \text{ df}$$

i.e. value $b_{rs} = 0$ corresponds to a boundary between sites r and s, since no regularization occurs between them, whereas $b_{rs} = 1$ correspond to no boundary.
In the following, we shall note:

$$b.\Phi(f) = \sum_{(r,s)\in\mathcal{C}} b_{rs}\Phi_{rs}(f) \text{ so that } Z_{\theta/b} = \int_\Omega \exp - \theta \ b.\Phi(f) \ \text{ df}$$

for sake of simplicity.

7.1.3 The prior probability of boundary process

It is assumed to be of the following linear exponential form:

$$\Pr(B = b \ / \ M = \mu) = \frac{\exp - \mu \ \Psi(b)}{Z_\mu}$$

$$\text{where } Z_\mu = \sum_{b\in\mathcal{B}} \exp - \mu \ \Psi(b)$$

Typically, we have $\Psi(b) = \sum_{(r,s)\in\mathcal{C}} (1 - b_{rs})$ for an independent clique boundary model, an hypothesis we shall keep here for sake of simplicity. Notice also that:
$Z_\mu = (1 + \exp -\mu)^{|\mathcal{C}|}$ in this case. In the following, we shall also use the notation:

$$\text{param} = (\theta,\ \mu) \text{ and } \text{Param} = (\Theta,\ M)$$

related to the set of hyperparameters to be estimated. With all these assumptions, the joint intensity-boundary process law writes finally as:

$$\begin{aligned}
&\Pr(F = f, B = b \;/\; D = d,\ \text{Param} = \text{param}) \\
&= \frac{\Pr(D = d \;/\; F = f, B = b,\ \text{Param} = \text{param}) \;.\; \Pr(F = f, B = b \;/\; \text{Param} = \text{param})}{\Pr(D = d)} \\
&= \frac{\Pr(D = d \;/\; F = f) \;.\; \Pr(F = f \;/\; B = b, \Theta = \theta) \;.\; \Pr(B = b \;/\; M = \mu)}{\Pr(D = d)} \\
&= \frac{\exp - \lambda \;||\; d - Rf \;||^2}{Z_\lambda} \cdot \frac{\exp - \theta\; b.\Phi(f)}{Z_{\theta/b}} \cdot \frac{\exp - \mu\; \Psi(b)}{Z_\mu} \cdot \frac{1}{\Pr(D = d)}
\end{aligned} \tag{38}$$

In order to go further it is necessary to investigate the behaviour of conditional partition function $Z_{\theta/b}$, in particular its dependence *w.r.t.* hyperparameter θ and boundary process b, in order to derive a Gibbs-like expression for the previously written joint intensity-boundary process law, which will form the basis of our analysis within this frame.

7.2 Computing the conditional boundary partition function

The conditional boundary partition function $Z_{\theta/b}$ can be expanded thanks once again to cumulant analysis. Two points of view can be described:

7.2.1 Global evaluation

First, we shall tackle a global point of view. Assuming the <u>rate of contours</u> in most natural images to remain <u>low</u>, we compute directly:

$$Z_{\theta/b} \;=\; \int_\Omega \exp - \theta \sum_{(r,s)\in\mathcal{C}} b_{rs}\; \Phi_{rs}(f)\;\; \text{df}$$

$$
\begin{aligned}
&= \int_\Omega \exp - \theta\, \Phi(f)\ \exp - \theta \sum_{(r,s)\in\mathcal{C}} (1-b_{rs})\ \Phi_{rs}(f)\ \ \mathrm{df} \\
&= Z_\theta\ .\ \mathbb{E}_\theta[\exp - \theta \sum_{(r,s)\in\mathcal{C}} (1-b_{rs})\ \Phi_{rs}] \qquad (39)
\end{aligned}
$$

which gives by the usual cumulant line:

$$
\log Z_{\theta/b} = \log Z_\theta + \theta\, \mathbb{E}_\theta[\sum_{(r,s)\in\mathcal{C}} (1-b_{rs})\ \Phi_{rs}] + \frac{1}{2}\,\theta^2\ \mathrm{var}_\theta(\sum_{(r,s)\in\mathcal{C}} (1-b_{rs})\ \Phi_{rs}) + \ldots
$$

Assuming now a stationary case, let us introduce local expectations and various range order correlations (see Fig. 10) :

$$
\begin{aligned}
\theta\, \mathbb{E}_\theta[\Phi_{rs}] &= \theta\, \frac{\mathbb{E}_\theta[\Phi]}{|\mathcal{C}|}\ , \\
\theta^2\ \mathrm{cov}_\theta(\Phi_{rs}, \Phi_{tu}) &= \kappa_\theta(\vec{\delta} = \vec{tu} - \vec{rs})\ \ \forall (r,s),(t,u) \in \mathcal{C} \quad \text{etc ..} \qquad (40)
\end{aligned}
$$

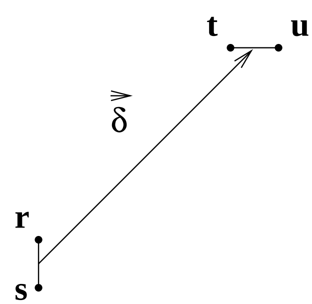


Figure 10: correlation between cliques

so that one obtains

$$
\log Z_{\theta/b} = \log Z_\theta + w(b)
$$

where

$$
w(b) = \theta\, \frac{\mathbb{E}_\theta[\Phi]}{|\mathcal{C}|} \sum_{(r,s)\in\mathcal{C}} (1-b_{rs}) + \sum_{\vec{\delta}} \kappa_\theta(\vec{\delta})\ [\sum_{(r,s)\in\mathcal{C}} (1-b_{rs})(1-b_{\vec{rs}+\vec{\delta}})\ ] + \ldots
$$

Joint posterior probability density function of intensity and boundaries (38) writes now:

$$\Pr(F=f, B=b \;/\; D=d, \text{ Param} = \text{param}) = \frac{1}{\Pr(D=d)} \; .$$
$$\frac{\exp\, -\lambda \,||\, d - Rf \,||^2}{Z_\lambda} \; . \; \frac{\exp\, -\theta\, b.\Phi(f)}{Z_\theta} \; . \; \frac{\exp\, -\mu\, \Psi(b) - w(b)}{Z_\mu} \qquad (41)$$

Thus, expansion term $w(b)$ can be interpreted as a correction to the boundary prior. It is worth noticing that its first-order term is exactly similar to the usual-form prior, whereas its second-order one, Ising-like, acts as a pairwise interaction terms between boundaries. We can say in this sense that conditional partition function contains in itself intrinsic prior knowledge on the boundary process.
We also see that in the important case of (stationary) homogeneous potentials, all expansion terms in $w(b)$ are independent of θ, since:

$$\theta \, \frac{\mathbb{E}_\theta[\Phi]}{|\mathcal{C}|} = \frac{|S|}{|\mathcal{C}|}\frac{1}{\alpha} = K \;\;,\;\; \theta^2 \text{ var}_\theta(\Phi) = \sum_{\vec{\delta}} \theta^2 \text{ cov}_\theta(\Phi_{rs}, \Phi_{\vec{rs}+\vec{\delta}}) = K \quad \text{etc } \ldots$$

so that all correlation functions $\kappa_\theta(\vec{\delta})$, noted now $\kappa(\vec{\delta})$, are independent on θ [4]. Thus:

$$w(b) = K\; \Psi(b) + \sum_{\vec{\delta}} \kappa(\vec{\delta}) \; [\sum_{(r,s)\in\mathcal{C}} (1-b_{rs})(1-b_{\vec{rs}+\vec{\delta}}) \;] + \ldots$$

is now independent on θ in expression (41). This will be extremely useful in the following developments, and also gives an insight on the power of cumulant analysis. Similar results can be found in Statistical Physics [Ma-85].

7.2.2 Local evaluation

One can need to estimate the variation of conditional partition function when the boundary process varies during some sampling algorithm. We have

$$Z_{\theta/b+\delta b} = \int_\Omega \exp\, -\theta\, b.\Phi(f) \;.\; \exp\, -\theta\, \delta b_{rs}\; \Phi_{rs}(f) \;\; \text{df} \; = Z_{\theta/b} \,.\, \mathbb{E}_{\theta/b}[\exp\, -\theta\, \delta b_{rs}\; \Phi_{rs}(f)]$$

[4] typically they have an exponential decreasing behaviour as functions of range $||\vec{\delta}||$: they can be written as $\kappa(\vec{\delta}) = \exp\, -\dfrac{||\vec{\delta}||}{\rho} \,.\, \kappa(\vec{0})$.

Thus:

$$\log Z_{\theta/b+\delta b} - \log Z_{\theta/b} = \log \mathbb{E}_{\theta/b}[\exp - \theta\, \delta b_{rs}\; \Phi_{rs}(f)] \approx -\theta\, \delta b_{rs}\; \mathbb{E}_{\theta/b}[\Phi_{rs}(f)] + \ldots$$

Since $b.\Phi(f)$ is an homogeneous potential as well, one has:

$$\mathbb{E}_{\theta/b}[b.\Phi(f)] = \frac{|S|}{\alpha\,\theta}$$

Assuming a stationarity hypothesis yields:

$$\mathbb{E}_{\theta/b}[\Phi_{rs}(f)] \approx \frac{|S|}{|\mathcal{C}(b)|} \cdot \frac{1}{\alpha\theta}$$

where $\mathcal{C}(b) = \sum_{(r,s)\in\mathcal{C}} b_{rs}$ is the number of remaining cliques when b is present. Thus:

$$\log Z_{\theta/b+\delta b} - \log Z_{\theta/b} \approx -\frac{|S|}{|\mathcal{C}(b)|}\frac{1}{\alpha} \,.\, \delta b_{rs}$$

which is fairly close to the energy variation obtained by the previous global approach at first order:

$$\log Z_{\theta/b+\delta b} - \log Z_{\theta/b} = \delta w(b) = -K\; \delta b_{rs} = -\frac{|S|}{|\mathcal{C}|}\frac{1}{\alpha} \,.\, \delta b_{rs}$$

when the rate of contours is low (*i.e.* $|\mathcal{C}(b)| \approx |\mathcal{C}|$). The stationarity assumption invoked before is of course generally untrue, especially when the clique being investigated is close to an already existing contour, (see Fig. 11), but we shall assume it fulfilled.

7.2.3 Application : the marginalization method

We shall use previous global expansion of partition function $Z_{\theta/b}$ and retain only expansion term up to first order. We shall see that even so, marginalization on boundaries random variables yields fairly close effective regularization potential

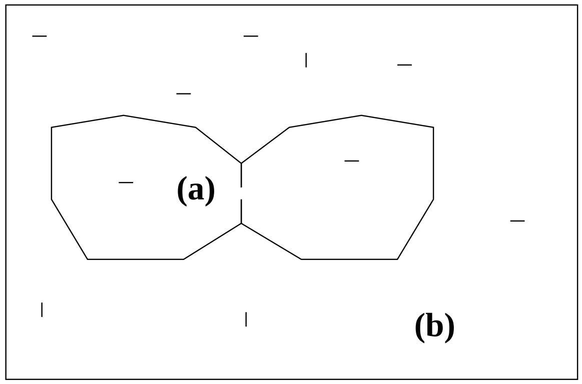


Figure 11: Adding a boundary value to current boundary process.
(a) : an area of dense existing contours, stationarity hypothesis untrue.
(b) : an area of low contour rate, stationarity hypothesis true.

intensities to Φ-functions with specified hyperparameters θ and μ. As a matter of fact summing previous expression (38) on all possible values of the process b yields:

$$\begin{aligned}
\Pr(F = f \;/\; D = d, \text{ Param} = \text{param}) &= \sum_{b \in \mathcal{B}} \Pr(F = f, B = b \;/\; D = d, \text{ Param} = \text{param}) \\
&= \frac{1}{\Pr(D = d)} \cdot \sum_{b \in \mathcal{B}} \frac{\exp -\lambda \mid\mid d - Rf \mid\mid^2}{Z_\lambda} \cdot \frac{\exp -\theta\, b.\Phi(f)}{Z_\theta} \cdot \frac{\exp -(\mu + K)\; \Psi(b)}{Z_\mu} \\
&= \frac{\exp -\lambda \mid\mid d - Rf \mid\mid^2}{\Pr(D = d)\; Z_\lambda\; Z_\theta\; Z_\mu} \cdot \sum_{b \in \mathcal{B}} \exp -\theta\, b.\Phi(f) - (\mu + K)\; \Psi(b) \qquad (42)
\end{aligned}$$

Now, due to the separability hypothesis of Ψ, this writes:

$$\Pr(F = f \;/\; D = d, \text{ Param} = \text{param}) \propto \frac{\exp -\lambda \mid\mid d - Rf \mid\mid^2}{Z_\lambda} \cdot \prod_{(r,s) \in \mathcal{C}} \left(\exp -\theta\; \Phi_{rs}(f) + \exp -(\mu + K) \right)$$

yielding thus the announced effective "marginalized" regularization potential:

$$\tilde{\Phi}_{rs}(f) = -\log\; \left(\exp -\theta\; \Phi_{rs}(f) + \exp -(\mu + K) \right)$$

An example is shown for gaussian quadratic potential and reasonable values of hyperparameters θ and μ (See Fig. 12). This is indeed close to the classical Φ-function shape.

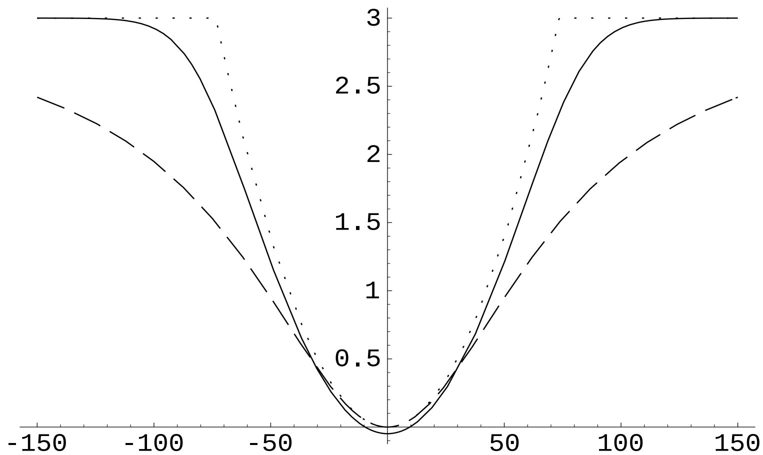


Figure 12: Comparison of boundary discrete marginal with $\Phi_{rs}(f) = \theta\ (f_r - f_s)^2$ (full), quadratic truncated (semi-dotted) and Φ-function (dotted) for the same values of parameters: $\mu + K = 3.0,\ \sigma = 30.0$ and $\theta = \dfrac{1}{2\sigma^2} = 5.6\ .\ 10^{-4}$

7.3 Estimation of hyperparameters within the explicit boundary framework

Our purpose here is to develop a methodology for the joint estimation of hyperparameters param $= \{\theta, \mu\}$. We follow the way previously adopted for estimation of hyperparameter θ alone by computing the posterior probability density function of param knowing observation d:

$$
\begin{aligned}
&\mathrm{Pr}(\text{Param} = \text{param} \;/\; d) \\
&= \sum_{b \in \mathcal{B}} \int_{\Omega} \mathrm{Pr}(\text{Param} = \text{param},\ F = f,\ B = b \;/\; D = d)\ \text{df} \\
&= \sum_{b \in \mathcal{B}} \int_{\Omega} \mathrm{Pr}(F = f, B = b \;/\; \text{Param} = \text{param},\ D = d)\ .\ \mathrm{Pr}(\text{Param} = \text{param})\ \text{df} \\
&= \frac{\mathrm{Pr}(\text{Param} = \text{param})}{\mathrm{Pr}(D = d)}\ . \\
&\quad \sum_{b \in \mathcal{B}} \int_{\Omega} \frac{\exp\ -\lambda \,||\, d - Rf \,||^2}{Z_\lambda} \cdot \frac{\exp\ -\theta\ b.\Phi(f)}{Z_\theta} \cdot \frac{\exp\ -\mu\ \Psi(b) - w(b)}{Z_\mu}\ \text{df} \quad (43)
\end{aligned}
$$

Assuming as before uniform prior distribution for $\mathrm{Pr}(\text{Param} = \text{param})$ and adopting the Gimel'farb approach, we write previous expression as:

$$
\mathrm{Pr}(\text{Param} = \text{param} \;/\; d) = \frac{1}{\mathrm{Pr}(D = d)}\ \frac{Z_{\lambda,\ \theta,\ \mu}}{Z_\theta\ .\ Z_\mu} \tag{44}
$$

where

$$
Z_{\lambda,\ \theta,\ \mu} = \sum_{b \in \mathcal{B}} \int_{\Omega} \exp\ -\lambda \,||\, d - Rf \,||^2 - \theta\ b.\Phi(f) - \mu\ \Psi(b) - w(b)\ \text{df} \tag{45}
$$

is the partition function associated to joint intensity-boundary effective posterior energy:

$$
\mathcal{U}(f,\ b) = \lambda \,||\, d - Rf \,||^2 + \theta\ b.\Phi(f) + \mu\ \Psi(b) + w(b) \tag{46}
$$

Many parameter-hyperparameter estimation variants are possible at this point. However, from what precedes, all of them will share the following common hyperparameter estimation scheme, valid for the stationary homogeneous regularization potential case (*i.e.* when $w(b)$ is independent on hyperparameters) :

$$
\mathbb{E}_{\lambda,\ \hat{\theta},\ \hat{\mu}}[b.\Phi(f)] = \left(\frac{\partial \log Z_\theta}{\partial \theta} \right)_{\hat{\theta}} \tag{47}
$$

$$
\mathbb{E}_{\lambda,\ \hat{\theta},\ \hat{\mu}}[\Psi(b)] = \left(\frac{\partial \log Z_\mu}{\partial \mu} \right)_{\hat{\mu}} \tag{48}
$$

Calling as before α the homogeneity degree of regularizatiom potential Φ and supposing that $\Psi(b) = \sum_{(r,s)\in\mathcal{C}} (1-b_{rs})$, this yields:

$$\mathbb{E}_{\lambda,\ \hat{\theta},\ \hat{\mu}}[b.\Phi(f)] = \frac{|S|}{\alpha\,\hat{\theta}} \tag{49}$$

$$\mathbb{E}_{\lambda,\ \hat{\theta},\ \hat{\mu}}[\sum_{(r,s)\in\mathcal{C}} (1-b_{rs})] = |\mathcal{C}|\ \frac{\exp -\hat{\mu}}{1+\exp -\hat{\mu}}$$

$$\Rightarrow\quad \mathbb{E}_{\lambda,\ \hat{\theta},\ \hat{\mu}}[\sum_{(r,s)\in\mathcal{C}} b_{rs}] = \frac{|\mathcal{C}|}{1+\exp -\hat{\mu}} \tag{50}$$

We try in the following to present some of these variants. Notice that in both cases, respective updatings of intensities and pixels associated to joint energy (46) are very simple:

- One updates intensity f at fixed boundary value b with a Metropolis dynamics associated to energy

$$\mathcal{U}(f) = \lambda\ ||\ d - Rf\ ||^2 + \theta\ b.\Phi(f) \tag{51}$$

  Its local variation at site s is given by:

$$\Delta\mathcal{U} = \lambda\ \Delta(\ ||\ d - Rf\ ||^2) + \theta \sum_{r\in\mathcal{N}_s} b_{rs}\ \Delta\Phi_{rs}(f)$$

- One updates boundary process b at fixed intensity f with a Metropolis dynamics associated to energy

$$\mathcal{W}(b) = \theta\ b.\Phi(f) + (\mu + K)\ \Psi(b) \tag{52}$$

  whose updating scheme for clique $c = (r, s)$ is given by

$$\Delta\mathcal{W} = \delta b_{rs}\ [\ \theta\ \Phi_{rs}(f) - (\mu + K)\ ]\ \text{where}\ \delta b_{rs} = \pm 1$$

First proposed iterative method

Initialization: initial boundary image $b^0_{rs} = 1 \; \forall (r,s) \in \mathcal{C}$ (*i.e.* no initial contours)
- initialize θ in a similar way as before: $\dfrac{|S|}{\alpha\theta} = b^0.\Phi(d) = \Phi(d)$.
- initialize μ with an admissible a priori value (*i.e.* no initial boundary detector).

(1) Estimating intensity and boundaries for given hyperparameters Θ and μ

- Updating f with a Metropolis dynamics associated to energy

$$\mathcal{U}(f) = \lambda \, || \, d - Rf \, ||^2 + \theta \, b.\Phi(f)$$

- Updating b with a Metropolis dynamics associated to energy

$$\mathcal{W}(b) = \theta \, b.\Phi(f) + (\mu + K) \; \Psi(b)$$

(2) Estimation of hyperparameters knowing f and b

- Estimating hyperparameter θ with following equation:

$$\mathbb{E}_{\lambda,\, \hat{\theta},\, \hat{\mu}}[b.\Phi(f)] = \frac{|S|}{\alpha\, \hat{\theta}} \quad , \quad \textit{i.e. } (49) \; .$$

- Estimating hyperparameter μ with following equation:

$$\mathbb{E}_{\lambda,\, \hat{\theta},\, \hat{\mu}}[\sum_{(r,s)\in\mathcal{C}} b_{rs}] = \frac{|\mathcal{C}|}{1 + \exp -\hat{\mu}} \quad , \quad \textit{i.e. } (50) \; .$$

Go to (1) if necessary

(3) Retain final intensity and boundary images

- Apply Posterior Mean estimate to the intensity and boundary samples obtained near convergence of hyperparameters.

Second proposed iterative method

Initialization: find a contour detector b^0 (e.g. from the value of λ), helping to
- initialize θ in a similar way as before: $\dfrac{|S|}{\alpha\theta} = b^0.\Phi(d)$
- initialize μ by the complete data estimation formula:

$$\Psi(b^0) = \sum_{(r,s)\in\mathcal{C}} (1 - b^0_{rs}) = |\mathcal{C}| \, \frac{\partial \log Z_\mu}{\partial\mu} \Leftrightarrow \frac{\sum_{(r,s)\in\mathcal{C}} b^0_{rs}}{|\mathcal{C}|} = \frac{1}{1+\exp -\mu}$$

(1) Estimating intensity f and hyperparameter θ knowing boundary value b

- Updating f with a Metropolis dynamics associated to energy

$$\mathcal{U}(f) = \lambda \mid\mid d - Rf \mid\mid^2 + \theta \, b.\Phi(f)$$

- Estimating hyperparameter θ with following equation

$$\mathbb{E}_{\hat\theta}[\Phi] = \mathbb{E}_{\lambda,\hat\theta/b}[b.\Phi]$$

  where right member expectation is taken along the Gibbs line (51). Here we see the importance of initial boundary process value b^0 in order to speed up the whole algorithm.

- Retaining image f^* by the associated Posterior Mean estimate for example.

(2) Estimating boundary value b and hyperparameter μ knowing intensity f^*

- Using a Metropolis dynamics associated to energy

$$\mathcal{W}(b) = \theta \, b.\Phi(f^*) + \mu \, \Psi(b) + \log Z_{\theta/b}$$

- Estimating hyperparameter μ with extended stochastic equation:

$$\mathbb{E}_{\lambda,\,\hat\theta,\,\hat\mu}[\sum_{(r,s)\in\mathcal{C}} b_{rs}] = \frac{|\mathcal{C}|}{1+\exp -\hat\mu} \quad , \quad \textit{i.e.}\ (50)\ .$$

Go to (1) if necessary

8 Results

In this section we present results for the first iterative method. Of course, these must be once again checked and compared with results from other existing methods in the same field, namely resulting form the use of Φ-functions and combination with the ARTUR algorithm [Charbonnier-96, Khoumri-97].

8.1 Principle

We assume as before an homogeneous regularization potential. We implement first the Metropolis updating schemes associated to respective posterior energies (51) and (52). Then we implement the generalized stochastic gradient algorithms associated to stochastic equations (49) and (50). For each of the related stochastic gradient-like algorithms, we can attribute a different normalization constant (the Younes factor). However we choose the same one in all our experiments. We also decide that the convergence criterium should bring on hyperparameter θ only, since our experiments show in general faster convergence for the boundary-related hyperparameter μ. Variances are computed separately in each stochastic equations (which means the diagonal hessian assumption in the gradient-like iterative scheme).

Result E corresponds to initial and noisy image shown on Fig. 13, whereas result F corresponds to initial and noisy image shown on Fig. 14.

Unfortunately, at this moment we cannot restore a proper posterior boundary image (in fact we find no more contour in such a posterior image). This remains to be explained.

Result E

Initial guess θ^0 hyperparameter values $\theta^0 = 0.097$ - $\mu^0 = 3.0$.

Conditions

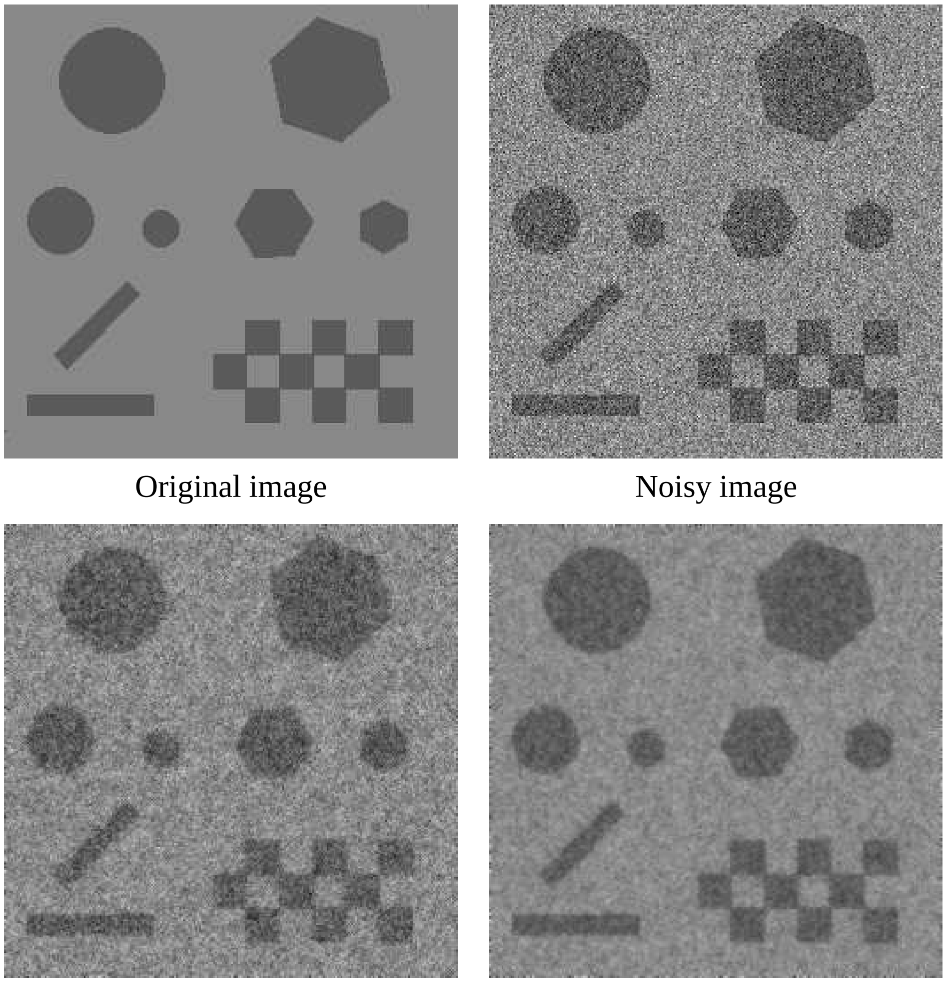
Homogenenous potential exponent	$\alpha = 2$ (boundary-gaussian potential)
Younes factor at step n:	$V = A * (n+1)^\gamma, \;\; A = 1.0, \;\; \gamma = \frac{2}{3}$ (boundaries+pixels)
Estimating empiric potential mean and	variances every 2 Metropolis sweeps
Posterior mean threshold	$\left\| \frac{\mathbb{E}_{\theta_n}[\phi] - < \phi >_{\theta_n,\lambda}}{\mathbb{E}_{\theta_n}[\phi]} \right\| < 0.10$
Total number of such averaging steps:	40

Results 434 Metropolis iterations -
Final hyperparameter estimates $\theta^* = 0.259$ - $\mu^* = 2.86$

	Initial image	Last Metropolis sample	Posterior Mean estimate
Region R_1:	μ = 135.744 σ = 27.799	μ = 137.206 σ = 18.787	μ = 136.471 σ = 11.851
Region R_2:	μ = 90.969 σ = 27.118	μ = 89.162 σ = 17.377	μ = 90.707 σ = 9.132

Last Metropolis sweep:

<u>pixels:</u>
prior energy/site = 1931.15 - empiric posterior energy/site = 1931.089
prior normalized variance = 7.46e+06 - empiric normalized posterior variance = 7.77e+05
<u>boundaries:</u>
prior energy/clique = 0.054 - empiric posterior energy/clique = 0.065
prior normalized variance = 0.0507 - empiric normalized posterior variance = 0.00988



Original image Noisy image

Last Metropolis sample Posterior mean image

Figure 13: Result E

Result F

Initial guess θ^0 and μ^0 hyperparameter value $\theta^0 = 0.0968$ $\mu^0 = 3.0$

Conditions

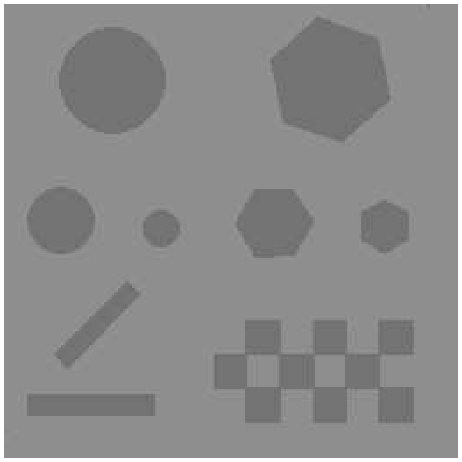
Homogenenous potential exponent	$\alpha = \frac{1}{2}$
Younes factor at step n:	$V = A * (n+1)^\gamma, \;\; A = 1.0, \;\; \gamma = \frac{2}{3}$ (boundaries+pixels)
Estimating empiric potential mean and	variances every 2 Metropolis sweeps
Posterior mean threshold	$\left\lvert \frac{\mathbb{E}_{\theta_n}[\phi] - <\phi>_{\theta_n,\lambda}}{\mathbb{E}_{\theta_n}[\phi]} \right\rvert < 0.10$
Total number of such averaging steps:	400

Results 818 Metropolis iterations -
Final hyperparameter estimates $\theta^* = 0.231$ - $\mu^* = 5.04$

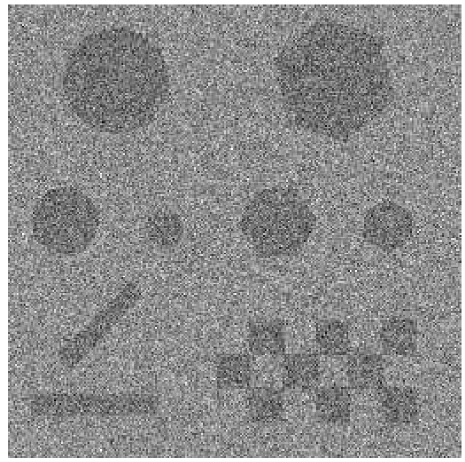
	Initial image	Last Metropolis sample	Posterior Mean estimate
Region R_1:	μ = 142.645 σ = 28.154	μ = 144.574 σ = 8.959	μ = 143.727 σ = 7.837
Region R_2:	μ = 114.873 σ = 29.638	μ = 113.316 σ = 6.343	μ = 113.984 σ = 4.251

Last Metropolis sweep:

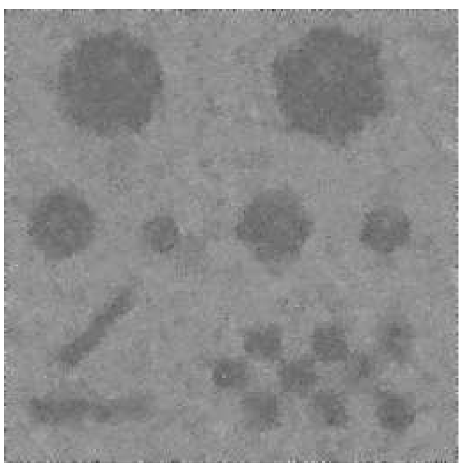
pixels:
prior energy/site = 8.651 - empiric posterior energy/site = 8.331
prior normalized variance = 37.4 - empiric normalized posterior variance = 2.25
boundaries:
prior energy/clique = 0.006 - empiric posterior energy/clique = 0.054
prior normalized variance = 0.00641 - empiric normalized posterior variance = 0.00205



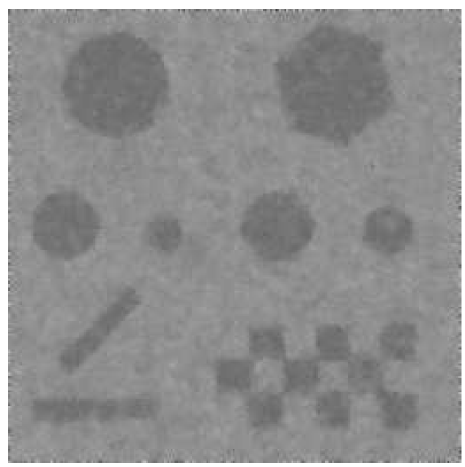
Original image



Noisy image



Last Metropolis sample



Posterior mean image

Figure 14: Result F

9 Last minute results

In this section we present a last minute idea : it consists in computing the current statistical empiric variance using current Metropolis sample and also previous one generated with previous hyperparameter values, assumed these values do not vary too much between consecutive iterations. This can indeed reduce the number of required iterations by a factor two at least ! It also can help provide better evaluation of the generalized stochastic gradient algorithm presented so forth in the long-term iteration range. We perform thus two initial Metropolis sweeps aimed to both first guess of hyperparameter θ and to the statistical empiric mean and variance estimates used in next iterations.

We give here two results relative to Synthetic Aperture Radar imagery. The former is related to a synthetic mire image (whose size is 160 x 160 pixels): whereas the latter refers to a sub-image of SAR ERS-1 image of Netherlands (whose size is 340 x 380 pixels).

To evaluate the performances of this method, we perform statistical averages in four significant sub-regions of the synthetic mire image:

Region	X	Y	width	height
R_1	0	0	80	20
R_2	20	20	80	20
R_3	40	40	80	20
R_4	60	60	40	40

We compute in particular the equivalent number of looks $\tilde{l} = \frac{\mu}{\sigma}$ in each region and for each image of interest. This is known to provide a good performance measure of the algorithm. We also present in Fig. 15 comparative slices profiles of initial and restored images.

Result G

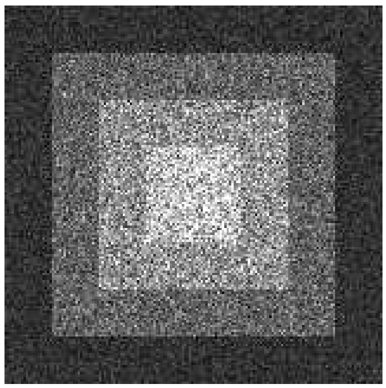
Initial guess θ^0 hyperparameter values $\theta^0 = 0.101$ - $\mu^0 = 3.0$.

Conditions

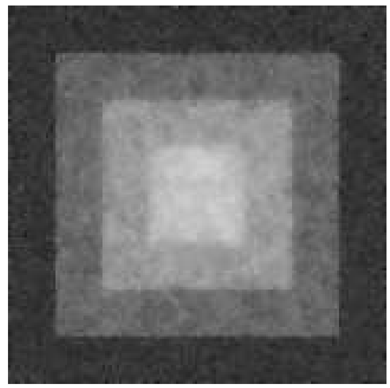
Homogeneous potential exponent	$\alpha = \frac{1}{2}$
Younes factor at step n:	$V = A * (n+1)^\gamma, \;\; A = 1.0, \;\; \gamma = 1.0$ (boundaries+pixels)
Estimating empiric potential mean and	variance EVERY Metropolis sweep
Posterior mean threshold	$\left\| \frac{\mathbb{E}_{\theta_n}[\phi] - < \phi >_{\theta_n,\lambda}}{\mathbb{E}_{\theta_n}[\phi]} \right\| < 0.10$
Total number of such averaging steps:	800

Results 2 (initial) + 803 Metropolis iterations -
Final hyperparameter estimates $\theta^* = 0.163$ - $\mu^* = 3.76$

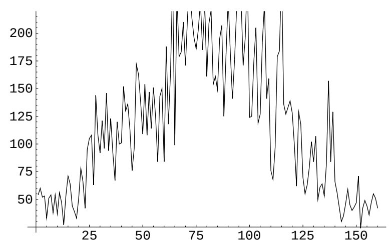
	Region	μ	σ	equivalent number of looks
Initial noisy image				
	R_1	48.527	14.784	3.282
	R_2	98.957	28.668	3.452
	R_3	145.578	42.653	3.413
	R_4	190.732	47.645	4.003
Last Metropolis sweep				
	R_1	51.611	10.661	4.841
	R_2	103.446	16.124	6.416
	R_3	150.903	15.409	9.793
	R_4	188.779	14.560	12.966
Posterior mean image				
	R_1	52.103	7.137	7.300
	R_2	102.697	9.579	10.721
	R_3	151.074	9.787	15.437
	R_4	189.952	8.476	22.409



Noisy image



Posterior mean image



Noisy image

Posterior mean result

Figure 15: Result G

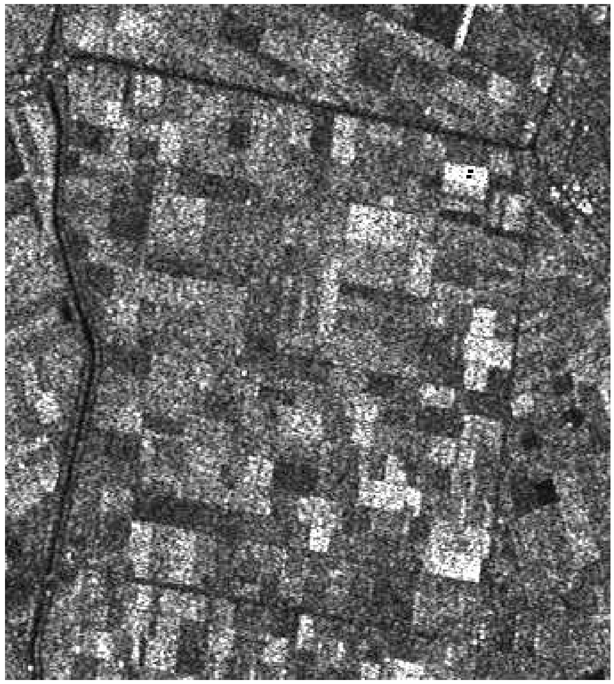
Result H

Initial guess θ^0 hyperparameter values $\theta^0 = 0.098$ - $\mu^0 = 3.0$.

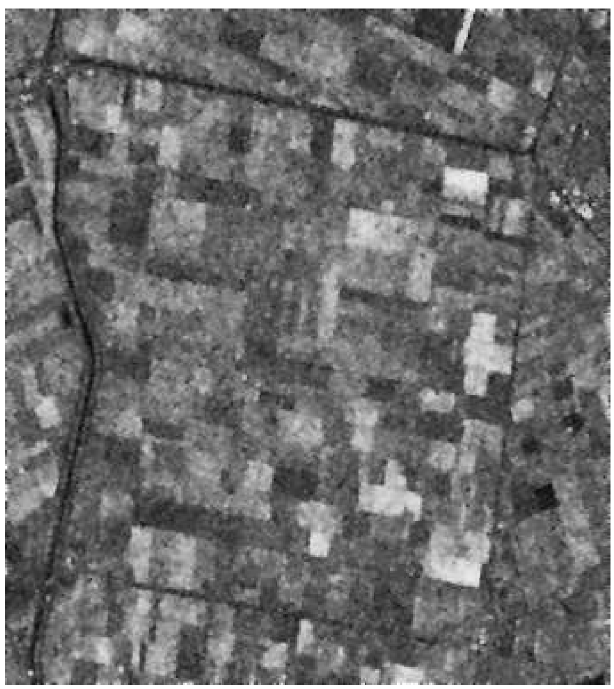
Conditions

Homogenenous potential exponent	$\alpha = \frac{1}{2}$
Younes factor at step n:	$V = A * (n+1)^\gamma, \ \ A = 1.0, \ \ \gamma = 1.0$ (boundaries+pixels)
Estimating empiric potential mean and	variance EVERY Metropolis sweep
Posterior mean threshold	$\left\lvert \frac{\mathbb{E}_{\theta_n}[\phi] - < \phi >_{\theta_n,\lambda}}{\mathbb{E}_{\theta_n}[\phi]} \right\rvert < 0.10$
Total number of such averaging steps:	800

Results 2 (initial) + 825 Metropolis iterations -
Final hyperparameter estimates $\theta^* = 0.1283$ - $\mu^* = 3.66$



Original SAR ERS-1 image © ESA



Posterior mean image

Figure 16: Result H

10 Conclusion

In this report, we have tried to present and to compare briefly current hyperparameter estimation methods with perhaps some new ones, the main challenge being the ability to simultaneously perform image restoration together with hyperparameter estimation. To our point of view, many tasks and investigations remain to be achieved, among them:

1. Test and implement the covariance method in the pure regularisation case (*i.e.* with no boundary process). Examine in particular the case where the prior potential expectation cannot be theoretically computed (so that only one MFR remains to be sampled). In the homogeneous potential case, compare with the generalized stochastic gradient method presented so forth.

2. Extend the covariance method in presence of a boundary process. This should not be too difficult, since in view of Eq.(43) one can write the posterior probability density function of hyperparameters θ and μ as:

$$\Pr(\text{Param} = \text{param} \;/\; d) \propto \mathbb{E}_{\theta,\lambda,\mu}[\exp - \; || \, d - Rf \, ||^2 - w(b)]$$

so that related covariance equations for θ and μ result easily.

3. Investigate the behaviour of $\mathbb{E}_{\theta,\lambda}[\Phi]$ at limit conditions, in particular when $\theta \searrow 0$ (thanks to a discussion with F. Baccelli). Here also a cluster development should be particularly adapted to this task. This should give precise indication of the relative behaviour of respective prior and posterior expectations $\mathbb{E}_{\theta}[\Phi]$ and $\mathbb{E}_{\theta,\lambda}[\Phi]$.

4. Implement and compare the second boundary iterative method.

5. Extend the presented methods in the non-stationary case, *i.e.* when hyperparameter values can locally depend of the image context. This should result into a third level (pixels, hyperparameters and "superhyperparameters") to be estimated as well.

6. Last (and in relationship with the previous point) examine the soundness of the regularization model chosen with respect to the image(s) to be restored, in terms of ground states (*i.e.* the configurations of lowest energy) and phase transition possibilities. To our opinion, this only could help validating the choice of such models in restoration for a given set of images, and thus the related task of hyperparameter(s) estimation.

Acknowledgements

I would like to thank Dr. Josiane Zerubia for her welcome at INRIA (PASTIS project) and motivating research orientation, Profs. François Baccelli and Georgy Gimel'farb, Dr. Laure Blanc-Féraud and Mr. Mustapha Khoumri for impressively stimulating and enlighting discussions.

References

[Chalmond-89] B. Chalmond, An Iterative Gibbsian Technique for Reconstruction of *m*-ary Images. *Pattern Recognition*, 22(6):747-761, 1989.

[Charbonnier-96] P. Charbonnier, L. Blanc-Féraud, G. Aubert and M. Barlaud. Deterministic Edge-Preserving Regularization in Computed Imaging. *IEEE Transactions on Image Processing*, 5(12):1-13, 1996.

[Descombes-96] X. Descombes, R. Morris, J. Zerubia and M. Berthod, Estimation of Markov Random Field Prior Parameters Using Markov Chain Monte Carlo Likelihood. INRIA Research Report N. 3015, 1996. Accepted for publication in *IEEE Transactions on Image Processing*.

[Djafari-96] A. Mohammad-Djafari, Joint Estimation of Parameters and Hyperparameters in a Bayesian Approach of Solving Inverse Problems. *ICIP'96*:473-476, Lausanne, Switzerland, 1996.

[Figueiredo-97] M.A. Figueiredo and J.M.N. Leitão, Unsupervised Image restoration and Edge Location Using Compound Gauss-Markov Random Fields and the MDL Principle. *IEEE Transactions on Image Processing*, 6(8):1089-1102, 1997.

[Geman-84] S. Geman and D. Geman, Stochastic relaxation, Gibbs Distribution, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721-741, 1984.

[Geman-85] S. Geman and D.E. McClure, Bayesian image analysis: An application to single photon emission tomography. *Proc. Stat. Comput. Sect. Amer. Stat. Assoc.*, Washington D.C. :13-18, 1985.

[Geman-90] D. Geman, S. Geman, C. Graffigne and P. Dong, Boundary detection by constrained optimization, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:609-628, 1990.

[Geman-92] D. Geman and G. Reynolds, Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3):367-383, March 1992.

[Gimel'farb-96] G.L. Gimel'farb, Texture modeling by multiple pairwise pixel interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(11):1110-1114, 1996.

[Gimel'farb-97] G.L. Gimel'farb, Gibbs Fields with Multiple Pairwise Pixel Interactions for texture Simulation and Segmentation. INRIA Research Report N. 3202, 1997.

[Khoumri-97] M. Khoumri, Estimation d'hyperparamètres pour la déconvolution d'images satellitaires. Rapport de stage DEA, INRIA Sophia, september 1997.

[Lakshmanan-89] S. Lakshmanan and H. Derin, Simultaneous Parameter Estimation and Segmentation of Gibbs Random Fields Using Simulated Annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(8):799-813, 1989.

[Ma-85] S.K. Ma, Statistical Mechanics. World Scientific Philadelphia Singapore 1985.

[Malyshev-91] V.A. Malyshev and R.A. Minlos, Gibbs Random Fields. Cluster Expansions. *Kluwer Academic Publishers* Dordrecht Boston London 1991.

[Metivier-87] M. Métivier and P. Priouret, Théorèmes de convergence presque sûre pour une classe d'algorithmes stochastiques à pas décroissant, *Probability Theory and Related Fields*, 74:403-428, 1987.

[Robert-96] C. Robert, Méthodes de Monte Carlo par Chaînes de Markov. Statistique Mathématique et Probabilités *Economica* Paris 1996.

[Winkler-95] G. Winkler, Image Analysis, Random Fields and Dynamic Monte Carlo Methods. A Mathematical Introduction. *Applications of mathematics*, Springer-Verlag Berlin Heidelberg New York 1995.

[Younes-88] L. Younes, Estimation and annealing for Gibbsian Fields. *A. Inst Henri Poincaré*, 24(2):269-294, 1988.

[Younes-89] L. Younes, Parametric Inference for Imperfectly Observed Gibbsian Fields. *Probability Theory and Related Fields*, 82:625-645, 1989.

[Zhou-97] Z. Zhou, R. Leahy and J. Qi, Approximate maximum Likelihood Hyperparameter Estimation for Gibbs Priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):799-813, 1997.

Contents

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifi que,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399