RATE OF CONVERGENCE OF THE SWENDSEN-WANG DYNAMICS IN IMAGE SEGMENTATION PROBLEMS: A THEORETICAL AND EXPERIMENTAL STUDY

ISABELLE GAUDRON

ABSTRACT. We study in this paper the convergence rate of the Swendsen-Wang dynamics towards its equilibrium law, when the energy belongs to a large family of energies used in image segmentation problems. We compute the exponential equivalents of the transitions which control the process at low temperature, as well as the critical constant which gives its convergence rate. We give some theoretical tools to compare this dynamics with Metropolis, and develop an experimental study in order to calibrate both dynamics performances in image segmentation problems.

INTRODUCTION

Image processing requires several important preliminary tasks, called low level tasks, among which a crucial one is the segmentation of the image. This problem has been very efficiently solved through probabilistic methods (see Geman (1990) for a general presentation), including Markov Random Fields models (for more details one can read Wang (1994) and Wang (1997)), and from a more general point of view turns out to be a cost function minimization problem. A large class of stochastic dynamics, such as Metropolis relaxation, can then be used. However these dynamics are quite slow, and their elementary moves do not seem well adapted to image segmentation problems: it is very difficult to change the label of a whole region, which can remain a long time misclassified, with the dynamics of Metropolis.

In the late 80's, a dynamics in which the elementary objects were not the single sites of the image but regions, has been proposed by Swendsen and Wang to speed-up the convergence rate towards equilibrium for the Potts model (Swendsen and Wang (1987), Sokal (1989)) near critical temperature. In this dynamics, auxiliary variables are introduced which allow us to change in one single move the label of a whole region. Such a possibility appears very attractive in image segmentation problems, and we refer here to Besag and Green (1993) which discuss its current and future relevance in Bayesian applications. We also refer to Besag *et al.* (1995) to have a complete presentation of Bayesian Computation and Markov chains Monte Carlo techniques. But, if careful studies of the speed-up of the convergence rate towards equilibrium has been done for the Potts model and for the Ising

model with a constant external field (see for instance Martinelli (1992), Martinelli *et al.* (1991a) and Martinelli *et al.* (1991b)), an extensive comparison in image segmentation problems has to be led.

This is the main object of this paper. The image segmentation energies we will consider here are quite simple. They cover nevertheless a large class of the energies used in image segmentation problems.
Our energies U are defined by

$$U(\sigma) = \sum_{<i,j>\in B} \delta_{\sigma_i \neq \sigma_j} + \sum_{i \in S} D_{i,F}(\sigma_i),$$

where

- S is a finite set called the set of pixels or also the set of sites.
- F is the image to be segmented.
- σ is an image of labels, i.e. σ belongs to Λ with $\Lambda = \{1, \cdots, L\}^S$.
- B is a set of undirected bounds $< i,j >$ between pixels giving the neighborhood relations.
- $D_{i,F}$ is a non negative function, depending on the input image F around the site i. Let l be a label in $\{1, \cdots, L\}$, and i be a site in S. The term $D_{i,F}(l)$ will be small if the label l is locally well adapted to the image F, and large otherwise. It is built according to local characteristics of the input image, such as grey level intensities, variances, or directional contrasts, depending of the type of images or textures to be segmented. This is usually a term which has to be carefully constructed.

This kind of energies has been already used in supervised image segmentation problems, as in Graffigne (1987), or in other problems (Geman *et al.* (1986), Herlin *et al.* (1992)) where we know the maximal number of labels L in the image to be analyzed, and statistical features of each label, to be able to build the functions D_i.

In the Metropolis dynamics, a single site is allowed to change its label at each step, whereas the Swendsen-Wang dynamics allows us to change the labels of all sites in one step. As suggested before, it introduces auxiliary variables and works on a state space X, larger than Λ:

$$X = \{(\sigma, n) \in \Lambda \times \{0,1\}^B; \forall < i,j >\in B, \sigma_i \neq \sigma_j \Rightarrow n_{ij} = 1\}.$$

Each configuration is then a couple whose first component is the image of labels σ, and the second one is an edges field separating pixels with different labels. A nice characteristics of this dynamics is that the first marginal of its equilibrium law is the equilibrium law of Metropolis. The problem which has then to be faced here is to compare the dynamics of two processes which have the same static behavior. Such a problem, partly because of the high cardinality of the state spaces, is difficult. We will develop then in this article different methods to tackle it. The first one is theoretical. It is based on the approach of Freidlin and Wentzell (1984), in their study of small perturbations of dynamical systems; we compute the exponential equivalents of the transitions which control the process at low temperature, and deduce that there exist, at least at low temperature, a class of energies for which the Metropolis dynamics runs faster, and an other one where the Swendsen-Wang dynamics is preferable. The frontier between these two

classes of energies can not be practically computed, and this approach is a priori relevant only at low temperature.

We will develop then experimental methods, which can also be applied to other similar problems. These methods focus on different aspects of such dynamics. The first one is relied to the quality of the results which are obtained: the lowest energies reached in a given number of iterations, the precision of such estimates, are computed. The second one estimates the time spent under some interesting levels of energy. The last one tries to understand the proper dynamical behaviors through the computation of excursions out of low energy levels. These methods are described in our last and main part. Our experimental results calibrate more precisely in very simple cases the range of energies where we should run the Swendsen-Wang dynamics, the main fact being that we have not found any case which was relevant in image segmentation problems and where the Swendsen-Wang dynamics gave better results. This study confirms the conclusions reached in Gray (1994) and in Hurn and Jennison (1993). In the first paper, A. Gray explains in detail how the Swendsen-Wang algorithm may be used to simulate certain types of posterior Gibbs distribution, and empirically compares the behavior of the algorithm with that of Gibbs sampler, when applied to image restoration problems. The paper shows in particular the importance of the starting configuration. The paper of M. Hurn and C. Jennison focuses on the relative performances of Gibbs and Swendsen-Wang samplers in MAP (maximum a posteriori) and MPM (marginal posterior modes) estimation, when applied to Ising models, either degraded or nondegraded by additive pixel-wise independent Normal noise. Although both studies concluded to disappointing performances of the Swendsen-Wang algorithm, it was not clear that such conclusions were also relevant in the domain of image segmentation, where the fundamental objects to tackle with are not pixels but regions.

1. Description of the Swendsen-Wang dynamics

Let us first recall that the Metropolis dynamics is an homogeneous Markov chain $(\Sigma_n)_{n\in\mathbb{N}}$, with transition matrix $S_\beta = (s_\beta(\sigma,\rho))_{\sigma,\rho\in\Lambda}$,

- $s_\beta(\sigma,\rho) = \frac{1}{|S|L}\exp(-\beta(U(\rho) - U(\sigma))^+)$ if there exists i in S such that $\rho_j = \sigma_j$, for all $j \neq i$, and $\rho_i \neq \sigma_i$.
- $s_\beta(\sigma,\rho) = 0$ if there exist two sites i and j in S such that $\sigma_j \neq \rho_j$ and $\sigma_i \neq \rho_i$.
- $s_\beta(\sigma,\sigma) = 1 - \sum_{\rho\neq\sigma} s_\beta(\sigma,\rho)$.
- The parameter β is strictly non negative and β^{-1} is called a temperature.

The law of $(\Sigma_n)_{n\geq 0}$ converges towards the Gibbs law P_β such that $P_\beta(\sigma) = \frac{1}{Z_\beta}\exp(-\beta U(\sigma))$, where Z_β is the partition function $\sum_{\rho\in\Lambda}\exp(-\beta U(\sigma))$. At low temperature, this Gibbs law loads uniformly the ground states:

$$\lim_{\beta\to\infty} P_\beta = P_\infty, \text{ where } P_\infty = \frac{\sum_{\sigma;U(\sigma)=U_{min}}\delta_\sigma}{|\{\sigma;U(\sigma)=U_{min}\}|}, \text{ and } U_{min} = \inf_\sigma U(\sigma).$$

Let us state now the notations we need to define and study the Swendsen-Wang dynamics.

NOTATION 1.1.   Let β be a strictly non negative parameter; we denote p_β the real valued number $\exp(-\beta)$.

Let $x = (\sigma, n)$ be a configuration of X. We denote $\sigma(x)$ the image of labels σ, and $n(x)$ the edges field n.

Let $\sigma \in \Lambda$; we denote $n(\sigma)$ the edges field associated to σ satisfying: for all $< i, j > \in B$, $\sigma_i = \sigma_j$ if and only if $n_{ij} = 0$. This field is the minimal edges field associated with σ.

Let $n \in \{0, 1\}^B$; we call $|n|$ the number of edges of n: $|n| = \sum_{<i,j>\in B} n_{ij}$. The minimal value of the energy U is denoted U_{min}.

Let (σ, n) be a state of X, δ the Kronecker delta, and $D(\sigma) = \sum_{i \in S} D_i(F, \sigma_i)$. We define the following function u_β from $\Lambda \times \{0, 1\}^B$ to $\mathbb{R}$:

$$u_\beta(\sigma, n) = \exp(-\beta D(\sigma)) \prod_{<i,j>\in B} (\delta_{n_{ij}=0} \delta_{\sigma_i=\sigma_j} (1 - p_\beta) + p_\beta \delta_{n_{ij}=1}). \quad (1.1)$$

Let us remark here that $u_\beta(\sigma, n) \neq 0$ if and only if $(\sigma, n) \in X$, and that for all configuration $x = (\sigma, n)$ of X we have:

$$u_\beta(\sigma, n) = \exp(-\beta D(\sigma))(1 - p_\beta)^{|B|} \left(\frac{p_\beta}{1 - p_\beta} \right)^{|n|}.$$

Moreover, using (1.1), a straightforward computation gives that for any $\sigma \in \Lambda$,

$$\sum_{n \in \{0,1\}^B} u_\beta(\sigma, n) = \exp(-\beta U(\sigma));$$

so that

$$\sum_{x \in X} u_\beta(x) = \sum_{\sigma \in \Lambda} \exp(-\beta U(\sigma)) = Z_\beta.$$

Hence, considering $\frac{u_\beta}{Z_\beta}$ as a probability measure on X, we get that its marginal on Λ is exactly the Gibbs measure P_β. We have to define now the allowed moves of the Swendsen-Wang dynamics.

DEFINITION 1.2. Let n be an edges field on B. A subset C of S is called a connected component of n if and only if

- For all distinct sites i and j in C, there exists a sequence of sites in C, $i_0 = i, i_1, \cdots, i_k = j$, such that for all $l \in \{0, \cdots, k-1\}, < i_l, i_{l+1} > \in B$ and $n_{i_l i_{l+1}} = 0$.
- C is a maximal subset of S which satisfies the previous condition.

Let us remark here that for any configuration $x \in X$ all the sites inside a connected component of $n(x)$ have the same label.

NOTATION 1.3. Let n be an edges field on B. We denote $\mathcal{C}(n)$ the sets of the connected components of n. If C belongs to $\mathcal{C}(n)$, and if σ is a configuration of Λ, σ_C will be the restriction of σ to C.

DEFINITION 1.4. The transition graph G on X is a symmetric graph, with vertices set X, where two distinct states x and y of X are neighbors if $\sigma(x) = \sigma(y)$ or if $n(x) = n(y)$. The notation $x \rightarrow y$ means that the two configurations x and y are mutual neighbors.

We can notice here that the graph G is a connected graph. Steps of the Swendsen-Wang dynamics are of two types. Let us call $x = (\sigma, n)$ our current configuration. We can compute a new edges field m as following.

Let $< i, j >$ be any bound of B. If $\sigma_i \neq \sigma_j$, m_{ij} is put to 1. Otherwise, m_{ij} is put to 1 with the probability $\exp(-\beta)$, and to 0 with the probability $1 - \exp(-\beta)$.

We can also compute the connected components associated to n and adjust the labels of these regions. For each connected component C, we choose a label l_C in $\{1, \cdots, L\}$, and compute $\Delta = \sum_{i \in C} (D_i(l_C) - D_i(\sigma_i))$. The sites of C receive then the label l_C with the probability $\exp(-\beta \Delta^+)$, where x^+ is equal to $\sup(x, 0)$. This informal description leads to the following definition.

DEFINITION 1.5. The Swendsen-Wang dynamics is an homogeneous Markov chain $(Y_n)_{n \in \mathbb{N}}$ whose transition matrix is $Q_\beta = (q_\beta(x, y))_{x, y \in X}$ such that, if $x \neq y$,

$$q_\beta(x, y) = \begin{cases} \frac{1}{2} q_\beta^{(1)}(x, y) & \text{if } \sigma(x) = \sigma(y), \\ \frac{1}{2} q_\beta^{(2)}(x, y) & \text{if } n(x) = n(y), \\ 0 & \text{otherwise}, \end{cases}$$

where

$$q_\beta^{(1)}(x, y) = \prod_{<i,j> \in B; \ \sigma(x)_i = \sigma(x)_j} p_\beta^{n(y)_{ij}} (1 - p_\beta)^{1 - n(y)_{ij}},$$

$$q_\beta^{(2)}(x, y) = \prod_{C \in \mathcal{C}(n(y))} L^{-1} p_\beta^{(D(\sigma(y)_C) - D(\sigma(x)_C))^+}.$$

As announced above, we can either create new bounds, and hence new connected components, or update the label of each connected component. We easily check that Q_β is reversible for the probability law on X, $\pi_\beta = \frac{u_\beta}{Z_\beta}$, whose marginal on the set Λ is the Gibbs law P_β. The Swendsen-Wang dynamics is then a tool to simulate the Gibbs law P_β through π_β. We have to study now its convergence rate towards equilibrium and to compare it to the convergence rate of the Metropolis dynamics.

2. CONVERGENCE RATE TOWARDS EQUILIBRIUM.

The law π_β is reversible for the transitions q_β, and the Swendsen-Wang process is irreducible and aperiodic. The transition matrix Q_β is diagonalizable then in an orthogonal basis of eigenvectors in $L^2(\pi_\beta)$; its eigenvalues are real valued, and satisfy, when noted in decreasing order:

$$\lambda_1(\beta) = 1 > \lambda_2(\beta) \geq \cdots \geq \lambda_{min}(\beta) > -1.$$

Let x be a state of X, $(Y_n)_{n \geq 0}$ a Swendsen-Wang process starting from x, and μ_n^x the law of Y_n. If V is a subset of X,

$$(\mu_n^x(V) - \pi_\beta(V))^2 \leq \pi_\beta(V) \frac{1 - \pi_\beta(x)}{\pi_\beta(x)} \rho(\beta)^{2n}, \tag{2.1}$$

where $\rho(\beta) = \sup\{\lambda_2(\beta), |\lambda_{min}(\beta)|\}$. We will derive the relations between $\rho(\beta)$ and β at low temperature from the study of the transition costs and of their exponential equivalents.

LEMMA 2.1. *Let x and y be two configurations such that $q_\beta(x,y) > 0$. Then*

$$\lim_{\beta \to \infty} \frac{\log(q_\beta(x,y))}{\beta} = -V(x,y),$$

where, if $x \neq y$,

$$V(x,y) = \begin{cases} |n(y)| - |n(\sigma(x))| & \text{if } \sigma(x) = \sigma(y) \\ \sum_{C \in \mathcal{C}(n(y))} (D(\sigma(y)_C) - D(\sigma(x)_C))^+ & \text{if } n(x) = n(y). \end{cases}$$

Let x be a configuration of X. Let us denote $\overline{U}(x) = D(\sigma(x)) + |n(x)|$. Then a straightforward computation gives that

$$\pi_\beta(x) = \frac{\exp(-\beta \overline{U}(x))(1 - p_\beta)^{-|n(x)|}}{\sum_{y \in X} \exp(-\beta \overline{U}(y))(1 - p_\beta)^{-|n(y)|}}.$$

Let x and y be two states of X such that $q_\beta(x,y) > 0$. Then

$$V(x,y) + \overline{U}(x) = V(y,x) + \overline{U}(y).$$

This means that the function $\overline{U}$, defined on X, is a potential function for the transition costs V.

Since our process is reversible and aperiodic in each irreducible class at temperature zero, we have that $\lambda_{min}(\beta)$ does not tend towards -1 when β tends to infinity (Diaconis and Stroock (1991)). Applying results of Freidlin and Wentzell (1984) which concern the exponential equivalents of the eigenvalues λ_i (chapter 6, page 208), we have that $\lim_{\beta \to \infty} \lambda_2(\beta) = 1$, except in degenerated cases, where the Markov chain has only one irreducible component at temperature zero. We deduce that $\rho(\beta) = \lambda_2(\beta)$ if β is large enough.

Lemma 2.3 gives the relation between λ_2 and β.

DEFINITION 2.2. Let x and y be two states of X. We call path between x and y any sequence of vertices $x_0 = x$, $x_1, \cdots, x_k = y$, such that for all $i \in \{0, \cdots, k - 1\}$, $q_\beta(x_i, x_{i+1}) > 0$. The set of paths which join x to y will be denoted C_{xy}.

LEMMA 2.3. *There exists a strictly non negative constant K which does not depend on β such that*

$$\lambda_2(\beta) \leq 1 - \frac{\exp(-\beta \overline{m})}{K},$$

where

$$\overline{m} = \sup_{x \neq y} \inf_{\gamma \in C_{xy}} (\sup_{u \in \gamma} (\overline{U}(u) - \overline{U}(y) - \overline{U}(x) + \overline{U}_{min}). \tag{2.2}$$

Proof. The first part of the proof is based on lemmas 4.1.4, 4.1.11 and theorem 4.1.13 stated in Deuschel and Mazza (1994) (pages 1031, 1033 and 1034).

Let u and v be two states in X such that $q_\beta(u,v) > 0$. Let us define $A(u,v) = \overline{U}(u) + V(u,v)$. For each path γ we will denote $e(\gamma)$ the set of its edges, and $Evel(\gamma) = \max_{(u,v) \in e(\gamma)} A(u,v)$. Set now

$$\overline{m}^{(1)} = \max_{x \neq y} (\min_{\gamma \in C_{xy}} Evel(\gamma) + \overline{U}_{min} - \overline{U}(x) - \overline{U}(y)).$$

Since Q_β is reversible for the probability law π_β one has

$$\lambda_2(\beta) \leq 1 - \frac{\exp(-\beta \overline{m}^{(1)})}{K}.$$

We shall prove now that $\overline{m}$ is equal to this constant $\overline{m}^{(1)}$.

Let x and y be two distinct states of X and γ a path from x to y. We always have that $Evel(\gamma) \geq \sup_{u \in \gamma} \overline{U}(u)$. We claim that there exists a path g from x to y such that $Evel(g)$ is equal to $\sup_{u \in g} \overline{U}(u)$ and to $\sup_{u \in \gamma} \overline{U}(u)$, and such that $Evel(g) \leq Evel(\gamma)$.

Let (u, v) be an edge of γ such that $u = (\sigma, n)$ and $v = (\sigma, m)$. Hence $A(u, v) = \overline{U}(u) + |m| - |n(\sigma)|$ which is greater than $\overline{U}(u) \vee \overline{U}(v)$. Let g_{uv} be the path $u \to u_1 = (\sigma, n(\sigma)) \to v$. Then $Evel(g_{uv}) = \overline{U}(u) \vee \overline{U}(v) \geq \overline{U}(u_1)$. Hence the announced result is true in this case.

Let now (u, v) be an edge of γ such that $u = (\sigma, n)$ and $v = (\rho, n)$, and $C_1, C_2, \cdots, C_k$ be the connected components of n where ρ is different from σ ordered in such a way that

$$D(\rho_{C_1}) - D(\sigma_{C_1}) \leq D(\rho_{C_2}) - D(\sigma_{C_2}) \leq \cdots \leq D(\rho_{C_k}) - D(\sigma_{C_k}).$$

Let us denote $\sigma_0 = \sigma$, and $\sigma_1, \cdots, \sigma_k$ the following images of labels:

$$\forall l \in \{1, \cdots, k\}, \quad \sigma_l(i) = \begin{cases} \rho(i) & \text{if } i \in \bigcup_{1 \leq j \leq l} C_j, \\ \sigma(i) & \text{otherwise.} \end{cases}$$

Let us also denote $u_l = (\sigma_l, n)$, for all $l \in \{0, \cdots, k\}$. Note here that u_k is equal to v. We replace the edge (u, v) with the path $g_{uv} = (u, u_1, \cdots, u_k)$, and claim that $Evel(g_{uv}) \leq A(u, v)$, that $Evel(g_{uv}) = \sup_{z \in g_{uv}} \overline{U}(z)$ and $Evel(g_{uv}) = \overline{U}(u) \vee \overline{U}(v)$. We easily check that $Evel(g_{uv}) = \sup_{z \in g_{uv}} \overline{U}(z)$, and that $A(u, v) \geq \overline{U}(u) \vee \overline{U}(v)$.

Now three cases may appear.

In the first case, $D(\sigma(v)_{C_1}) - D(\sigma(u)_{C_1}) \geq 0$. Hence, we deduce that $D(\sigma(u)) \leq D(\sigma_1) \leq \cdots \leq D(\sigma_k)$, and that $Evel(g_{uv}) = \overline{U}(v)$.

In the second case, $D(\sigma(v)) - D(\sigma_{k-1}) \leq 0$. Hence we deduce that $D(\sigma(u)) \geq D(\sigma_1) \geq \cdots \geq D(\sigma(v))$, and that $Evel(g_{uv}) = \overline{U}(u)$.

In the third case we suppose that $D(\sigma(v)) - D(\sigma_{k-1}) > 0$ and that $D(\sigma_1) - D(\sigma(u)) < 0$. Hence $Evel(g_{uv}) = \overline{U}(u) \vee \overline{U}(v)$, and the proof is complete if we build the path g in replacing each edge (u, v) with the path g_{uv}. $\square$

Moreover, we have the exponential equivalence at low temperature:

$$\lim_{\beta \to \infty} \frac{\log(1 - \lambda_2(\beta))}{\beta} = -\overline{m}.$$

Let us recall here that we have similar results with the Metropolis dynamics, if we replace $\overline{U}$ with U, $\overline{U}_{min}$ with U_{min}, and the transition graph of the Swendsen-Wang dynamics with the transition graph of the Metropolis dynamics. The associated critical constant will be denoted m:

$$m = \sup_{x \neq y} \inf_{\gamma \in C_{xy}} (\sup_{u \in \gamma} (U(u) - U(y) - U(x) + U_{min}).$$

Let us first compare U and $\overline{U}$.

- For all x in X, $\overline{U}(x) \geq U(\sigma(x))$.
- For all x in X, $\overline{U}(x) = U(\sigma(x))$ if and only if $n(x) = n(\sigma(x))$.

- A configuration x of X is a local minimum of $\overline{U}$ if and only if $n(x) = n(\sigma(x))$ and if $\sigma(x)$ is a local minimum of the function D.
- The energies $\overline{U}$ and U have the same minimal value. Let us denote X_{min} the set of global minimums of $\overline{U}$:

$$X_{min} = \{(\sigma, n(\sigma)) \in X; U(\sigma(x)) = U_{min}\}.$$

The equilibrium law of the Swendsen-Wang process satisfies:

$$\lim_{\beta \to \infty} \pi_\beta = \pi_\infty,$$

where π_∞ is the uniform measure on X_{min}. Translating the results obtained with the Metropolis dynamics, we have the following convergence at low temperature.

Let us denote $V = \{x \in X; U(\sigma(x)) \geq E\}$, with $E > U_{min}$ and such that $\{x \in X; U(\sigma(x)) = E\} \neq \emptyset$. When the constant $\overline{m}$ is not equal to 0, the rate of convergence of the Swendsen-Wang process to the ground states is bounded by:

$$\limsup_{n \to \infty} \frac{1}{\log n}(\inf_\beta \sup_{x \in X} \log \mu_n^x(V)) \leq -\frac{E - U_{min}}{\overline{m}}.$$

3. COMPARISON OF THE SWENDSEN-WANG AND METROPOLIS DYNAMICS

This comparison is based on the comparisons of their associated critical constants m and $\overline{m}$. This leads to a comparison of the convergence rates of these dynamics which is relevant at low temperature only.

Let us recall first that the constant $\overline{m}$ is also equal to

$$\overline{m} = \sup_{x \in X_{min}, y \in X_{loc}, x \neq y} \inf_{\gamma \in C_{xy}} \sup_{u \in \gamma}(\overline{U}(u) - U(\sigma(y))),$$

where $X_{loc} = \{x \in X; \forall y \in X \text{ such that } x \to y, \overline{U}(x) \leq \overline{U}(y)\}$. An other important fact is that $\overline{m} > 0$ if and only if $\overline{U}$ has at least one local minimum which is not in the same valley of energy as one of the global minimums. We will focus in this section on the following family of energies:

$$U_h(\sigma) = hD(\sigma) + \sum_{<i,j> \in B} \delta_{\sigma_i \neq \sigma_j}, \text{ where } h > 0. \tag{3.1}$$

We recall here that the non negative function D is still equal to $\sum_{i \in S} D_i(\sigma_i)$. Such family of energies is very often used in image segmentation problems, where we usually try different smoothing rates: if we want to have a very smooth result, i.e. to have quite large regions with smooth boundaries, we run some relaxation steps with a small parameter h; if we want to have a very precise segmentation, we would prefer a large parameter h. Let us denote $\overline{m}$ the function such that

$$\overline{m}(h) = \sup_{x \neq y} \inf_{\gamma \in C_{xy}} \sup_{u \in \gamma} \left(\overline{U_h}(u) - \overline{U_h}(x) - \overline{U_h}(y) + \overline{U_h}_{min}\right),$$

and m the similar function for the Metropolis dynamics. Since these functions m and $\overline{m}$ are both continuous, we have the following proposition.

PROPOSITION 3.1. *There exists a constant $h_0 > 0$ such that for all $h < h_0$, $\overline{m}(h) < m(h)$. If D has a minimal value which is obtained with a unique and non constant image of labels $\underline{\sigma}$, there exists a constant $h_1 > 0$ such that for all $h \geq h_1$, $\overline{m}(h) > m(h)$.*

Proof. Let $h = 0$. The set X_{min} is equal to $\{(\sigma, 0_B); \forall i, j \in S \quad \sigma_i = \sigma_j\}$, where we denote 0_B the null field of edges. The energy $\overline{U}$ has then L global minimums whose first components are the constant images. The set X_{loc} is equal to $\{(\sigma, n(\sigma)) \in \Lambda \times B^{\{0,1\}}\}$. Let $x = (\sigma, 0_B)$ be a state of X_{min} and $y = (\sigma(y), n(\sigma(y)))$ be a state of X_{loc}.

- If y belongs to X_{min} too, then x and y are neighbors and $V(x, y) = 0$.
- In the other case, let us consider the following path γ from x to y: $x_0 = x \rightarrow x_1 = (\sigma, n(y)) \rightarrow x_2 = y$, and compute $\overline{U}(u) - \overline{U}(y)$ for each $u \in \gamma$. We easily check that $\sup_{u \in \gamma}(\overline{U}(u) - \overline{U}(y)) = 0$.

It follows then that $\overline{m}(0) = 0$.

The Metropolis constant $m(0)$ is strictly non negative: the energy U_0 has many strict local minimums. We end the proof of the first part of the proposition with the continuity of the two functions $\overline{m}$ and m.

Let us prove now the second part of the proposition 3.1. For h large enough, the energy U has a unique local and global minimum, for the graph of Metropolis, which is equal to $\underline{\sigma}$. The minimal value of $\overline{U}$ is unique and equal to $x = (\underline{\sigma}, n(\underline{\sigma}))$. Let us consider the following configuration of X_{loc}, $y = (\sigma(y), 0_B)$, with $\sigma(y)$ equal to a minimal constant image: for all site $i \in S$, $\sigma(y)_i = \underline{l}$, where $\sum_i D_i(\underline{l}) = \inf_{j \in \{1, \cdots, L\}} \sum_i D_i(j)$. Let γ be a path going from x to y. As $\underline{\sigma}$ is not a constant image, γ has to contain a configuration $u = (\sigma(u), n(u))$, with $\sigma(u)$ constant and $n(u)$ different from 0_B; some edges of $n(u)$ have to build a closed boundary. This fact gives us that $\overline{U}(u) - \overline{U}(y) \geq 2$. It follows then that $\overline{m}(h) \geq 2$, for h large enough. As the two functions m and $\overline{m}$ are continuous, there exists $h_1 > 0$, such that for all $h > h_1$, $m(h) < \overline{m}(h)$. $\square$

There exists then a wide domain of energies where we are sure that we have a better rate of convergence with the Metropolis dynamics than with the Swendsen-Wang dynamics. We should now estimate the domain where it is preferable to run the Swendsen-Wang dynamics than the Metropolis one. First of all we should answer the following question: is the domain $[0, h_0]$ large enough to contain some h such that U_h has a global minimum which is not a constant image? If it is not the case, it will be always worse to run the Swendsen-Wang dynamics than the Metropolis one to reach the ground states of U in practical cases of image segmentation. We have no general answer to such a question, but we have run some simulations of these two dynamics in a very simple case.

4. EXPERIMENTAL STUDY

4.1. FRAMEWORK AND PURPOSES

In all this section our goal is to compare the behaviors of Metropolis and Swendsen-Wang dynamics, when they are applied on a wide range of

energies, on quite large images, and at many temperatures. This comparison
will be led when both dynamics are run with the same fixed number of
iterations. We call iteration, or step, of the dynamics of Metropolis, one
sweep of all sites of an image. An iteration of the Swendsen-Wang dynamics
is also defined as a sweep of all sites of an image; it is cut into two steps. In
the first step, we update the edges field; in the second step, we compute the
associated connected components, and choose their new labels.

We will particularly focus our attention on three points. We will first
look at the lowest energies reached by each dynamics. We will then com-
pare their energy distributions, and compute some interesting quantiles. We
will compare at last their dynamical behaviors, when both processes are at
equilibrium (or when we can not at least discriminate their static behaviors),
and derive from this comparison indications about their energy landscapes.

The energies we will work with belong to the family described in the
equation (3.1), with the four nearest neighbors system. The second term of
the energy $D(\sigma)$ is very simple and easy to build

$$D(\sigma) = D_R(\sigma) = \sum_{i \in S} \delta_{\sigma_i \neq R_i},$$

where the image $R = (R_i)_{i \in S}$ can be interpreted as the result of a prior esti-
mate of the segmentation, based on a local classification of the pixels which
does not take into account the spatial relationships between neighbors. We
think that even if this second term is too rough to give good segmentations,
it gives relevant conclusions about the comparison of the Metropolis and
Swendsen-Wang dynamics.

The energy U_h will then provide us with smooth segmentations close to
R. The value of the parameter h will be adjusted according to the expected
smoothness of the result: a small value of h will provide us with a high
regularization (note that if h is close to zero, the energy ground states are
constant images, and are of no interest for image segmentation problems);
if h is large, our segmented image will be very close to R.

4.2. Energies based on synthetic images

We have built several binary images R, of size 100 by 100, and present
here some significant and generic results obtained with three of them. The
first image R_1 is a perfect segmented image, and provides us with a family
of energies where we know what are the ground states, and some important
local minimums. We will see that even in such an easy case, starting from
a random image, the Swendsen-Wang dynamics has some problems to find
the right segmentation. The second one, denoted R_2, is the realization of
a 2D Ising model with no external field at low temperature, to study the
behaviors of both dynamics when we want to segment images with wide
regions which have no linear boundaries, and where holes of misclassified
pixels appear. The third one, called R_3, is a noisy half white and black
image; we also know in this case what is the expected segmentation.

For each second term D_R, we have run both dynamics with five different
temperatures (β belongs to $\{0.7, 1, 2, 3, 5\}$), and with several parameters

h which belong to a domain where the energy ground states are smooth versions of the prior classification R.

For each experiment, we have first run 1000 steps of both dynamics, starting from random noise. We have kept the final configuration and discarded the 999 first images. These 1000 steps seem to be sufficient to reach equilibrium for the Metropolis dynamics at all temperatures, and for the Swendsen-Wang dynamics only at median or high temperatures ($\beta \leq 2$).

Let $X_1^{R,h,1,\beta}$ be the final configuration obtained with Metropolis and $X_1^{R,h,2,\beta}$ the final configuration obtained with Swendsen-Wang. We have then run 4000 steps of each dynamics and computed the energy U of each configuration. Let us denote $(X_i^{R,h,1,\beta})_{1 \leq i \leq 4000}$ the 4000 configurations of Metropolis and $(X_i^{R,h,2,\beta})_{1 \leq i \leq 4000}$ the configurations of Swendsen-Wang and let us compare their energies.

4.2.1. LOWEST ENERGIES REACHED BY BOTH DYNAMICS.

Let h, R and the type j of the dynamics be fixed and let us call $u = (R, h, j)$. For each sequence $(X_i^{u,\beta})_{1 \leq i \leq 4000}$ we have computed the minimal energy reached $U_m^{u,\beta} = \min_{1 \leq i \leq 4000} U(X_i^{u,\beta})$, the mean energy $\overline{U}^{u,\beta} = (4000)^{-1} \sum_{i=1}^{4000} U(X_i^{u,\beta})$, and the standard deviation. We have also computed the energies of the two constant images U_{white} and U_{black} and of the image R, the two first images being at least local minima of both graphs in the domain where lives h. In each following table we have reported U_{white}, U_{black}, U_R and the lowest energy reached among all different temperatures $\widehat{U}_{min}^u = \min_\beta U_m^{u,\beta}$. We have also reported the set $\mathcal{B}$ of the inverse temperatures β for which $\widehat{U}_{min}^u$ is reached. The bold typed parameter β corresponds to values for which $\overline{U}_{low} = \min_{\beta \in \mathcal{B}} \overline{U}^{u,\beta}$ is reached. The reported standard deviation is associated to this bold typed parameter.
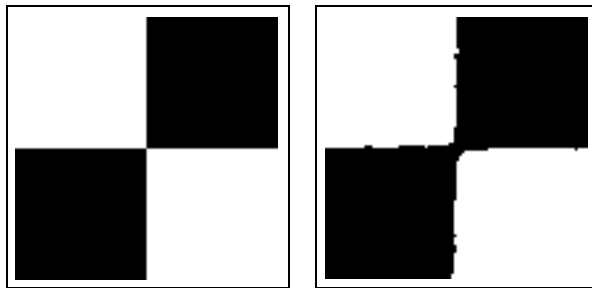

FIGURE 1. Left: first image R_1; right: best result obtained with Swendsen-Wang for $h = 0.1$ and $\beta = 2$.

In the first case (see table I and figure 1), $U_{white} = U_{black} = 5000h$, and $U(R_1) = 200$. The global minimum is a non constant image for $h > 0.04$. When h is equal or larger than 4, the Metropolis energy landscape has no strict local minimum (in this case, $m(4)$ is equal to 0).

The second image (see figure 2), denoted R_2, is the realization of a Markov random field of energy $\beta \sum_{<i,j>} \delta_{\sigma_i \neq \sigma_j}$, with β large.

The third image (see figure 3), denoted R_3, is a half white and black image I^0 corrupted with an independent noise W as follows: $R_3(i) = I_i^0 \delta_{W_i=0} +$
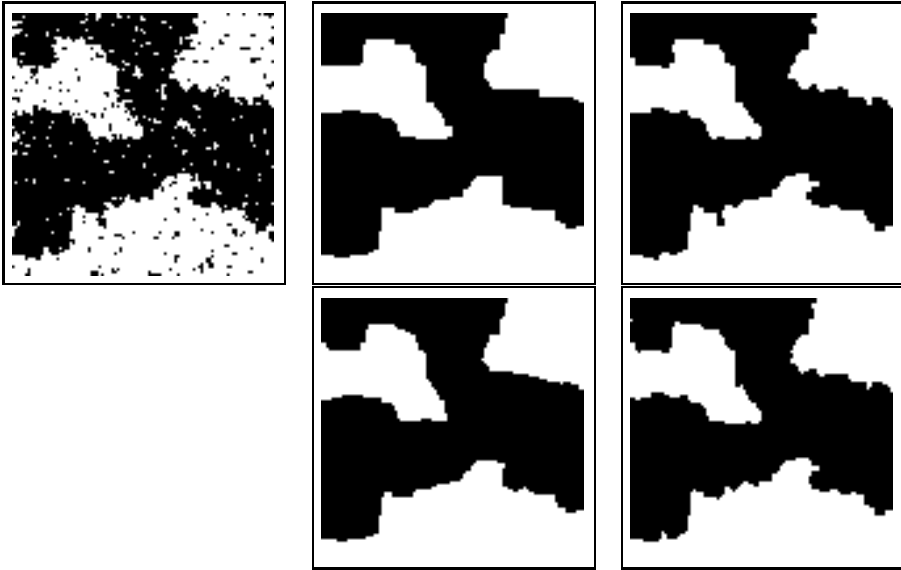
FIGURE 2. Examples of segmented images of lowest energy, associated with the second image R_2. Up-left: Second image R_2; up-middle: Metropolis with $h = 0.5$ and $\beta = 5$; up-right: Metropolis with $h = 1$ and $\beta = 5$. Down-middle: Swendsen-Wang with $h = 0.5$ and $\beta = 2$; down-right: Swendsen-Wang with $h = 1$ and $\beta = 2$. Results obtained with the two dynamics are very close from a qualitative point of view, even if $\widehat{U}_{min}(Met)$ is slightly smaller than $\widehat{U}_{min}(S.W.)$.
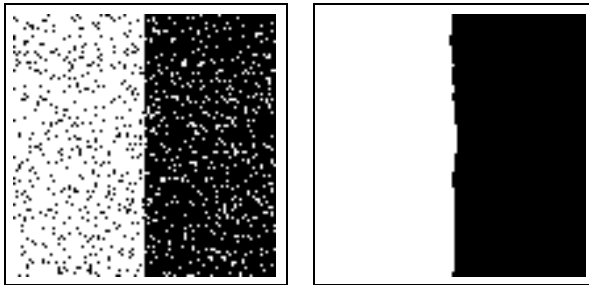

FIGURE 3. Left: image R_3; Right: best result obtained by Swendsen-Wang on R_3 with $h = 0.1$ and $\beta = 2$.

$(1 - I_i^0)\delta_{W_i=1}$, for all site i in S, where the W_i are i.i.d. random variables such that $P(W_i = 1) = 0.1$, and $P(W_i = 0) = 0.9$. In table III we have reported the energy of the perfect corresponding segmented image I^0.

In all our experiments, the energies observed with both dynamics at high and sometimes median temperatures ($\beta \leq 2$) are very similar (we have similar mean energies, standard deviations, and lowest energies). When temperature becomes too low, these dynamics have different quantitative and qualitative behaviors. The energies of Metropolis configurations are getting smaller and smaller, whereas the energies of Swendsen-Wang configurations

are much larger than they were at $\beta = 2$. The best results of the Swendsen-Wang dynamics are quite always obtained with $\beta = 2$. At low temperature, Swendsen-Wang stays in local minima, and can not create easily boundaries to change the structure of the connected components of the image: it has to pay a high price whereas it does not know if it would be able to decrease the energy. For instance it is very difficult for this dynamics to move the boundary of a region at low temperature, or to break a region into several pieces.

h	0.05	0.1	0.5	4
$U_{white} = U_{black}$	250	500	2500	20 000
$U(R_1)$	200	200	200	200
Metropolis				
$\widehat{U}_{min}$	200.05	200	200	200
β	5	5	2, 3, **5**	1, 2, **3, 5**
$\overline{U}_{low}$	202.85	201.28	200.19	200
Std	1.29	0.87	0.34	0
Swendsen-Wang				
$\widehat{U}_{min}$	233.3	226.60	200	200
β	2	2	2	1, **2**, 3
$\overline{U}_{low}$	272.76	256.25	218.98	200.02
Std	12.49	10.25	7.85	0.35

TABLE I. Results obtained with the first image R_1. For $h = 4$, the energy U has no local minimum in the graph of Metropolis, this explains why Metropolis reaches R_1 after a few steps and stays in this state for $\beta \geq 3$.

h	0.1	0.5	1	1.5	2	4
U_{white}	551.7	2578.5	5517	8275.5	11 034	22068
U_{black}	448.3	2241.5	4483	6724.5	8966	17932
$U(R_2)$	1908	1908	1980	1908	1908	1908
Metropolis						
$\widehat{U}_{min}$	377.3	721	1023	1278.5	1504	1908
β	3	5	5	5	3, **5**	2, 3, **5**
$\overline{U}_{low}$	394.65	726.63	1025.36	1281.98	1504.36	1908.13
Std	15.21	2.02	1.43	1.42	0.64	0.39
Swendsen-Wang						
$\widehat{U}_{min}$	478.1	757	1046	1305	1523	1924
β	2	2	2	2	2	2
$\overline{U}_{low}$	510.17	783.25	1074.61	2535.57	1544.76	1961.8
Std	11.64	8.77	9.76	2079.64	8.03	29.7

TABLE II. Results obtained with the second image R_2.

h	0.1	0.5	1	4
U_{white}	495.8	2479	4958	19832
U_{black}	504.2	2521	5042	20168
$U(R_3)$	3605	3605	3605	3605
$U(I^0)$	199.2	596	1092	4068
Metropolis				
$\widehat{U}_{min}$	199.2	596	1092	3605
β	3, **5**	2, 3, **5**	2, 3, **5**	3, **5**
$\overline{U}_{low}$	199.31	596.04	1092.11	3605.2
Std	0.46	0.24	0.35	0.53
Swendsen-Wang				
$\widehat{U}_{min}$	215.2	596	1092	3637
β	2	2	2	2
$\overline{U}_{low}$	239.5	614.675	1108.1	3710.3
Std	9.2	8	7	45.2

TABLE III. Results obtained with the third image R_3.

We also see in tables I, II and III that the lowest energies reached by the Swendsen-Wang dynamics are higher than the lowest energies reached by Metropolis. Even when the Swendsen-Wang dynamics reaches the minimal energy, the standard deviation is worse than with the Metropolis dynamics (we have to recall here that the minimal energy is reached by at a lower temperature for Metropolis than for Swensen-Wang).

4.2.2. COMPARISON OF THE ENERGY DISTRIBUTIONS.

We have shown yet that the Metropolis dynamics reached ground states at many temperatures, whereas the Swendsen-Wang dynamics was very often trapped in local minimums at low temperature. Since the Swendsen-Wang dynamics always obtains its best results with $\beta = 2$, we focus from now on both dynamics at this inverse temperature. We limit our study to the cases where global minimums were reached or nearly reached. Let us look at the energy distributions of both sequences, and denote $U^1 = (U(X_i^{R,h,1,\beta}))_{1 \leq i \leq 4000}$ and $U^2 = (U(X_i^{R,h,2,\beta}))_{1 \leq i \leq 4000}$.

Each following figure shows both distributions and cumulative distribution functions of U^1 (in solid line) and U^2 (in dashed line), as well as interesting quantiles. The main fact is that both distributions are similar, and can not be discriminated when $h = 0.5$ or 1, confirming the fact that both dynamics seem to be at equilibrium.

This shows that if we have to work at a median temperature we will obtain similar results with both dynamics. If we want to reach ground states with a very precise value, which means that we have to work at low temperature, we should run the Metropolis dynamics, which appears much more robust.
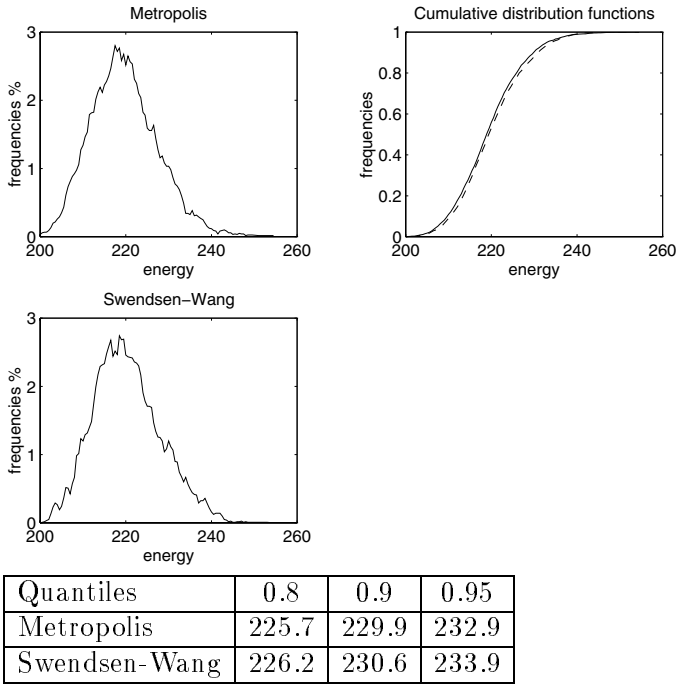
Quantiles	0.8	0.9	0.95
Metropolis	225.7	229.9	232.9
Swendsen-Wang	226.2	230.6	233.9

FIGURE 4. Distributions of U^1 (in solid line) and U^2 (in dashed line), with $h = 0.5$, $\beta = 2$, and $R = R_1$.

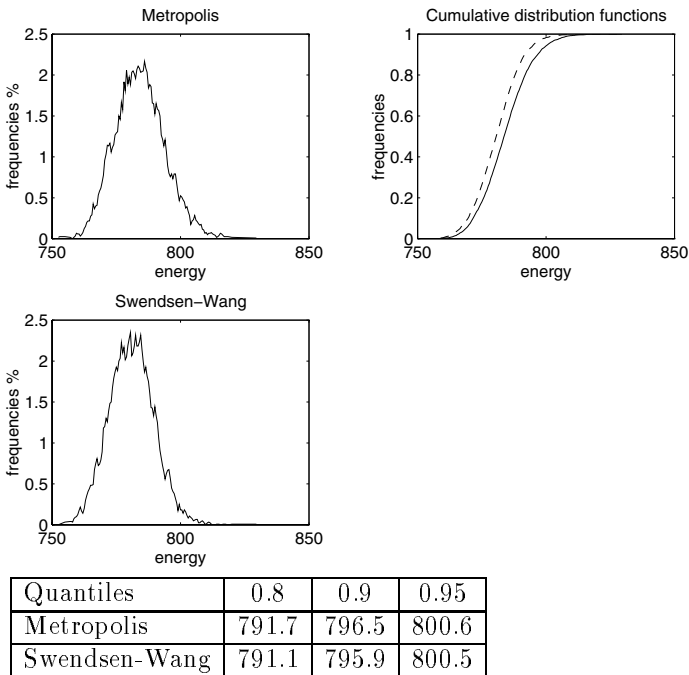| Quantiles	0.8	0.9	0.95
Metropolis	791.7	796.5	800.6
Swendsen-Wang	791.1	795.9	800.5

FIGURE 5. Distributions of U^1 and U^2, with $h = 0.5$, $\beta = 2$, and $R = R_2$.

Metropolis

Cumulative distribution functions

Swendsen–Wang

Quantiles	0.8	0.9	0.95
Metropolis	621.4	625.6	629.0
Swendsen-Wang	621.2	625.0	629.1
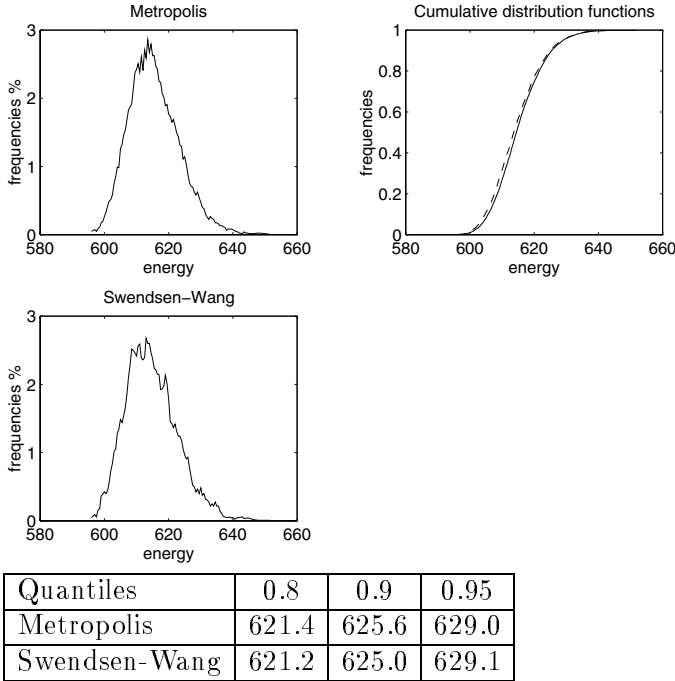
FIGURE 6. Distributions of U^1 and U^2 when $h = 0.5$, $\beta = 2$ and $R = R_3$.

4.2.3. COMPARISON OF THE ENERGY LANDSCAPES.

What happens at this median temperature $\frac{1}{2}$ gives us significant informations about the proper dynamical behaviors of these Markov chains. More precisely, we focus here on the energy landscape nearby interesting subsets of Λ, i.e. sets of low energy states, and work with the complete processes U^1 and U^2 (and not only with their distributions). A simple way to proceed is to look at the excursions out of these low energy level sets. Long excursions will denote the existence of a barrier of high energy in the neighborhood of these sets.

Let us denote $V_\epsilon^1 = \{\sigma \in \Lambda; U(\sigma) \leq (1 + \epsilon)U_{min}\}$, and $V_\epsilon^2 = \{x \in X; U(\sigma(x)) \leq (1+\epsilon)U_{min}\}$. Let us consider the Metropolis (resp. Swendsen-Wang) Markov chain starting from V_ϵ^1 (resp. V_ϵ^2) with its equilibrium law. Let $\tau_\epsilon^1 = \inf\{i > 0; X_{i+1}^1 \in V_\epsilon^1\}$ be the return time associated with Metropolis and $\tau_\epsilon^2 = \inf\{i > 0; X_{i+1}^2 \in V_\epsilon^2\}$ be the return time associated with Swendsen-Wang. We have estimated their distributions for small values of ϵ. They have similar means (we are at equilibrium)[1], but look very different: τ_ϵ^1 has median values whereas the Swendsen-Wang process has quite a lot of short excursions and always a few long ones. This implies that the variance of τ_ϵ^2 is always much larger than the variance of τ_ϵ^1. The existence of long excursions also means that at low temperature, the Swendsen-Wang process will take a long time to reach a ground state if it does not start from

---

[1] Let us recall here that $E^V(\tau_V) = \frac{1}{\pi(V)}$ for an irreducible Markov chain, starting from V with its equilibrium law π, and where τ_V is the return time in V.
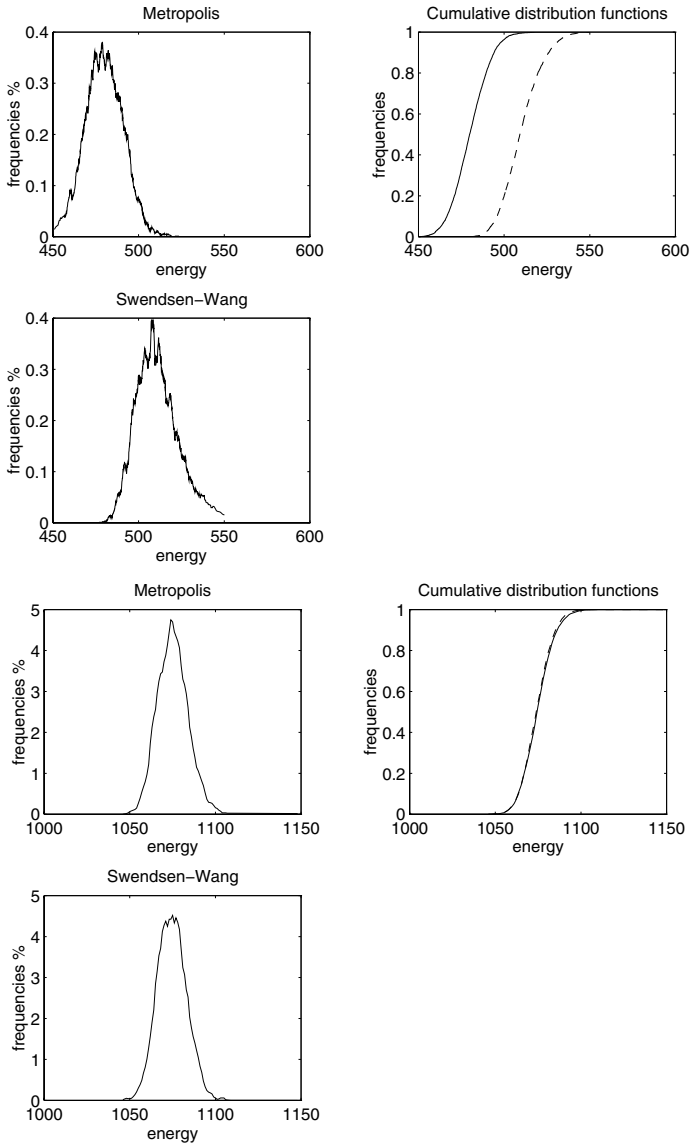
FIGURE 7. Distributions of U^1 and U^2 when $R = R_2$ and $\beta = 2$; top: h=0.1; bottom: h= 1. Let us notice here that these distributions are quite different when $h = 0.1$, i.e. with a high regularization rate and non constant global minimums; this confirms the fact that in this case Metropolis performs better (see table II). This difference disappears when h increases.

the neighborhood of a global minimum (we see such a phenomenon in our experiments).

We show the different results we have obtained with $\epsilon = 5\%$, with values of h for which static results were similar for both dynamics, in the figures 8, 9 and 10. The number of the return time lies on the x-axis, and its value

on the y-axis. After each figure, we express the mean value, denoted $\overline{\tau}$, and the empirical variance, denoted $var(\tau)$.
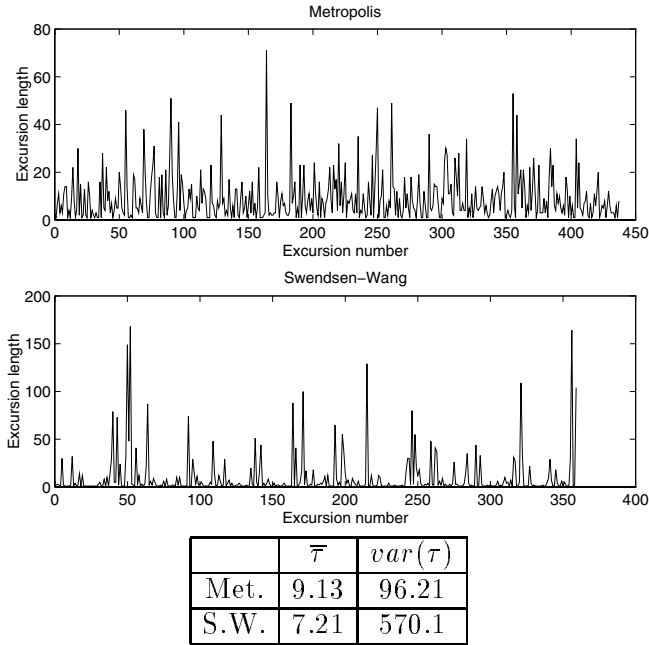

	$\overline{\tau}$	$var(\tau)$
Met.	9.13	96.21
S.W.	7.21	570.1

FIGURE 8. Return times of Metropolis and Swendsen-Wang for $h = 0.5$, R_1, and $\epsilon = 5\%$.

	$\overline{\tau}$	$var(\tau)$
Met.	1.34	0.62
S.W.	1.30	3.455
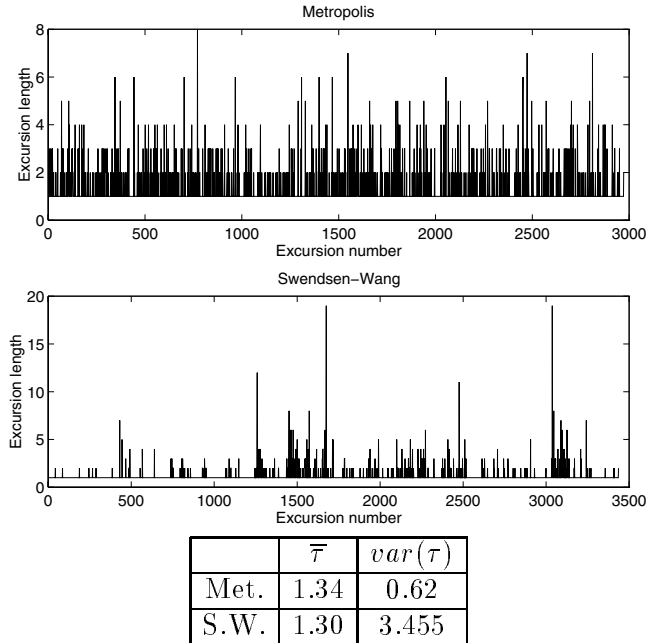
FIGURE 9. Return times of Metropolis and Swendsen-Wang for $h = 0.5$, R_2, and $\epsilon = 5\%$.

We have also estimated, for the inverse temperature $\beta = 2$ and for $h \in \{0.5, 1\}$, the autocorrelation functions of the two processes U^1 and U^2, i.e. $\rho_U^1(t) = \frac{cov(U(X_1^1), U(X_{1+t}^1))}{var(U(X_1^1))}$, and $\rho_U^2(t) = \frac{cov(U(X_1^2), U(X_{1+t}^2))}{var(U(X_1^2))}$. The autocorrelation function of the Swendsen-Wang process has a much slower decrease than Metropolis (see figures 11 and 12). This also confirms the fact that Swendsen-Wang has a slower rate of convergence to equilibrium than Metropolis. One can read Sokal (1989) to have a quite complete idea on the relations between the autocorrelation function and the convergence rate of a Markov process to equilibrium in the reversible case. We also refer to Gaudron and Trouvé (1996) in which we study the integrated autocorrelation times of a Markov chain at low temperature with no reversibility assumption.
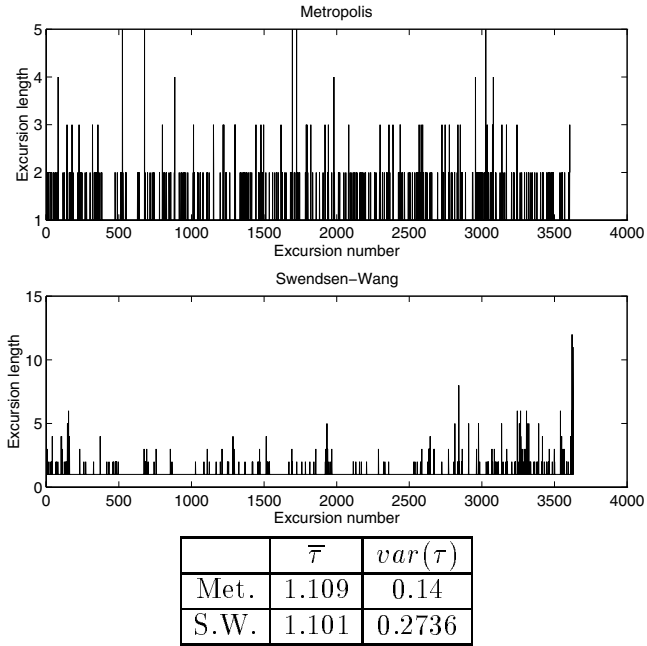

	$\overline{\tau}$	$var(\tau)$
Met.	1.109	0.14
S.W.	1.101	0.2736

FIGURE 10. Return times of Metropolis and Swendsen-Wang for $h = 0.5$, R_3, and $\epsilon = 5\%$.
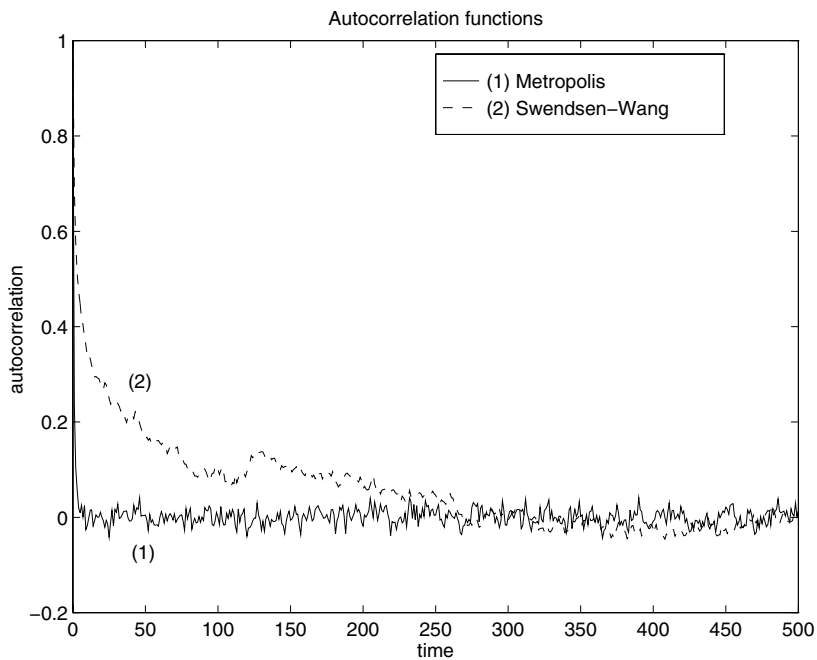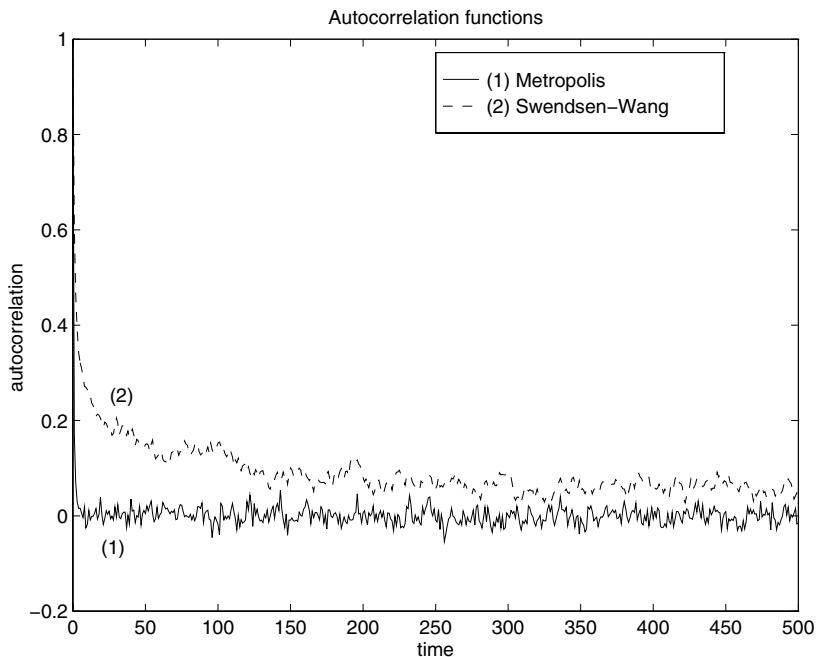
FIGURE 11. Autocorrelation functions of the energy of the Metropolis and Swendsen-Wang processes for $h = 0.5$ and $\beta = 2$. Top: R_1; bottom: R_2.
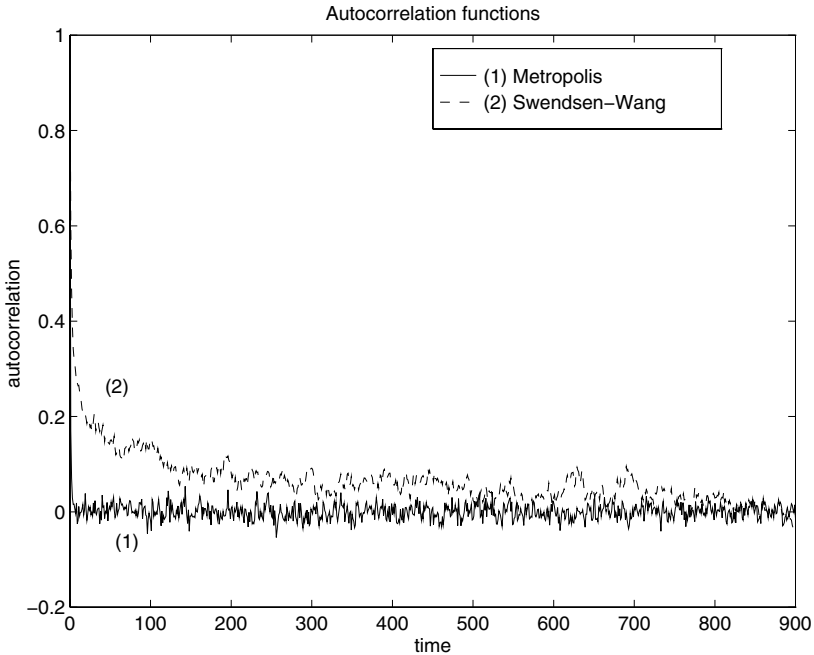
FIGURE 12. Autocorrelation functions of the energy of the Metropolis and Swendsen-Wang processes for $h = 0.5$, $\beta = 2$ and R_3.

4.3. COMPARISON WHEN THE GROUND STATES ARE CONSTANT IMAGES

We focus in this part on the energies U_h with a small value of h, i.e. with a high regularization term. Although such family of energies is of no interest in image segmentation problems (ground states are constant images in such a case), this study will help us to answer some questions about the value of the threshold h_0 which satisfies: $\forall h < h_0$, $\overline{m}(h) < m(h)$. In all our experiments, the fact that the energy ground states were constant images did not appear to be sufficient to assure that the Swendsen-Wang dynamics had a higher rate of convergence than Metropolis. In fact, the Swendsen-Wang process behavior seems better than the Metropolis behavior for very small values of h ($h \leq 0.001$), and worse as soon as $h \geq 0.01$ (see tables IV and V).

We have also estimated return times of the two dynamics for $h = 0.001$, and we have noticed that both return times had very similar distributions. Let us recall here that Swendsen-Wang seemed more efficient than Metropolis for such a small value of h. Hence we can deduce that the study of some return times distributions is a very useful tool to compare different Markovian dynamics. In particular, the existence of long excursions out of sets of low energy states at a median temperature will give a very slow rate of convergence to equilibrium at low temperature.

h	0.001	0.01	0.02
$U_{white} = U_{black}$	5	50	100
$U(R_1)$	200	200	200
Metropolis			
$\widehat{U}_{min}$	5	50	100
β	2, 3, **5**	2, **3**, 5	**2**, 3, 5
$\overline{U}_{low}$	5.002	50.002	104.16
Std	0.08	12.98	8.5
Swendsen-Wang			
$\widehat{U}_{min}$	5	243.88	247.88
β	2, 3, **5**	2	2
$\overline{U}_{low}$	5	280.05	282.06
Std	0	10.75	10.67

TABLE IV. Results obtained with the first image R_1.

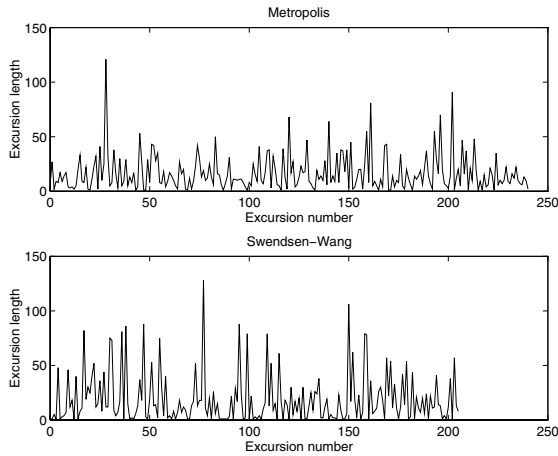| h	0.001	0.01	0.02	0.05
U_{white}	5.517	55.17	110.34	275.85
U_{black}	4.483	44.83	89.66	224.15
$U(R_2)$	1908	1908	1908	1908
Metropolis				
$\widehat{U}_{min}$	4.483	44.83	89.66	272.1
β	2, **3**	2, 3, **5**	2, **3**	5
$\overline{U}_{low}$	4.90	44.83	90.09	275.43
Std	1.22	0.08	1.24	1.31
Swendsen-Wang				
$\widehat{U}_{min}$	4.483	238.11	333.9	394.05
β	2, 3, **5**	2	2	2
$\overline{U}_{low}$	4.49	274.92	386.97	443.56
Std	0.07	13.92	26.7	25.2

TABLE V. Results obtained with the second image R_2.

FIGURE 13. Return times of Metropolis and Swendsen-Wang for $h = 0.001$, $\beta = 2$, R_2, and $\epsilon = 100\%$ (the minimal energy is very small). Let us notice that both return times have here similar distributions: in this case (which has no interest for image segmentation problems), the Swendsen-Wang dynamics is very efficient.

4.4. SEGMENTATION OF A REAL IMAGE

We will show here some results obtained with a real image of size 256×256, which is composed of four different textures $(^2)$. The energy we work with is $U_h(\sigma) = \sum_{<i,j> \in B} \delta_{\sigma_i \neq \sigma_j} + h \sum_{i \in S} \delta_{\sigma_i \neq R_i}$, where $< i,j > \in B$ if and only if $|i - j| = 1$, and where R is our prior unsupervised segmentation we have computed according to local statistical characteristics of the image. This energy is too simple to solve such a problem: a realistic second term would involve a kind of distance between statistical characteristics of the different textures. Anyway, the quantitative and qualitative comparisons of the convergence rates of Metropolis and Swendsen-Wang dynamics would lead to similar conclusions.

The experiment is the following: we start from the first segmented image R, and run n_1 simulations of the Metropolis dynamics and n_2 simulations of the Swendsen-Wang dynamics. Let us remark that R is here a local minimum for the Swendsen-Wang transition graph: this dynamics has to increase its energy quite a long time, before being able to decrease it. Hence its stabilization time is much longer than the stabilization time of Metropolis, and this explains why we have taken n_2 much larger than n_1. On figure 14, we show the initial image to be segmented, as well as the first segmentation based on statistical features of the textures. The image of smallest energy obtained with the Swendsen-Wang dynamics is quite close from the image obtained with the Metropolis dynamics, but after much more steps.

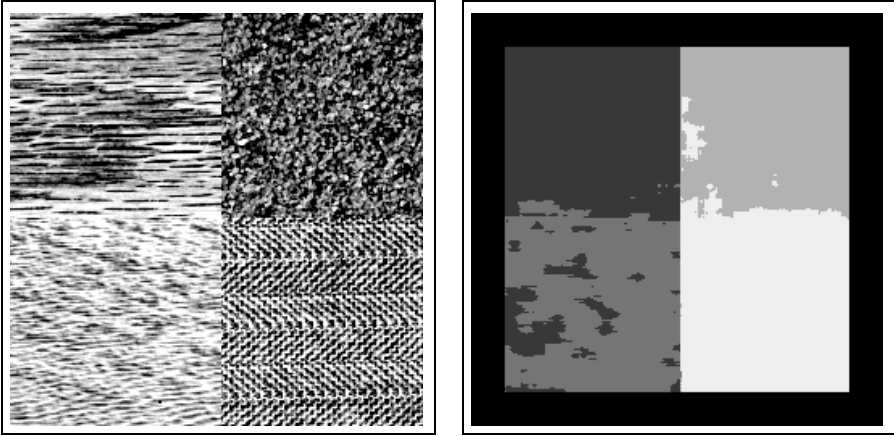[2]We thank very much Professor C. Graffigne who has kindly provided us with this image.

FIGURE 14. Left: original image; right: prior segmentation R.
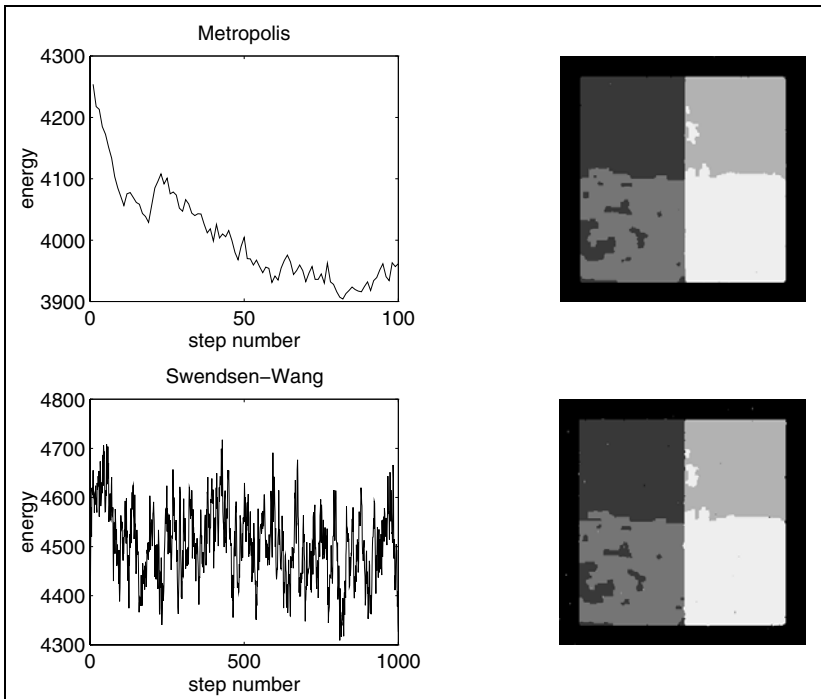

FIGURE 15. Experiments with $h = 0.5$, and $\beta = 2$. Up-left: Metropolis energy values; up-right: segmented image of lowest energy obtained with Metropolis dynamics. Down-left: Swendsen-Wang energy values; down-right: segmented image of lowest energy obtained with Swendsen-Wang dynamics. Note that the Swendsen-Wang dynamics has not reached equilibrium at the last step of Metropolis.

5. Conclusion

We have theoretically stated in this article that we should better run Swendsen-Wang than Metropolis for energies with a large regularization term, and that we should run Metropolis when the regularization term is very small. In our experiments we have tried to calibrate what was the meaning of small and large.

The first point which has appeared in all our experiments is that we shall not run the Swendsen-Wang dynamics at too low temperatures: we have always obtained the best results with $\beta \simeq 2$, i.e. at a median temperature. For $\beta = 3$ or $\beta = 5$ the Swendsen-Wang dynamics gets trapped in the neighborhood of local minimums of the energy. We can also notice here that results obtained with Swendsen-Wang are very sensitive to this choice of temperature.

In a second step, we have studied more precisely the Swendsen-Wang and Metropolis dynamics at this median temperature $\beta = \frac{1}{2}$, in significant cases for image segmentation problems (i.e. $h = 0.5$ or 1), and when stationarity seemed to be reached. We have computed and compared the distributions of the excursions out of sets of low energy. In all cases, these distributions were very different: the Swendsen-Wang dynamics has always very short excursions and a few very long ones, whereas the excursions of Metropolis have a median length.

This behavior at a median temperature explains what happens at low temperature: the Swendsen-Wang state space seems to have deeper cycles, standing not too far from sets of low energy, than the Metropolis state space. At a median temperature, it stays quite a long time in these cycles, but can still get out of them. When the temperature decreases, it gets trapped in them. We also think that the qualitative properties of such excursions is a very efficient and helpful way to compare Markovian dynamics.

The main disappointing conclusion is that we did not find any interesting case for the image segmentation problem, which means that the global minimums of the energy are not constant images, where the Swendsen-Wang dynamics provided us with better results than the Metropolis dynamics.

Hence we do not believe that the Swendsen-Wang dynamics is interesting to be run by alone for image segmentation problems. Nevertheless it may be interesting to follow a mixed strategy, and we refer here to the current thesis work of O. Cherif (CMLA) who builds an effective image segmentation algorithm, partly based on the Swendsen-Wang dynamics, and to Hurn (1995) who considers adaptative strategies in a classification task in confocal fluorescence microscopy.

References

Besag, J. and Green, P. J. (1993). Spatial statistics and bayesian computation. *J. R. Statis. Soc. B* **55** 25–37.

Besag, J., Green, P. J., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science* **10** 3–66.

DEUSCHEL, J.-D. and MAZZA, C. (1994). L^2 convergence of time nonhomogeneous Markov processes: I. Spectral estimates. *Ann. Appl. Prob.* **4** 1012–1056.

DIACONIS, P. and STROOCK, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Prob.* **1** 36–61.

FREIDLIN, M. I. and WENTZELL, A. D. (1984). *Random perturbations of dynamical systems*, **260**, Springer-Verlag.

GAUDRON, I. and TROUVÉ, A. (1996). Fluctuations of empirical means at low temperature for finite Markov chains with rare transitions in the general case. Preprint CMLA, Cachan, France.

GEMAN, D. (1990). Ch. Random fields and inverse problems in imaging. *Lectures on Probability Theory and Statistics. XVIIIème École d'Été de Probabilités de Saint-Flour*, Lecture Notes in Mathematics, Springer-Verlag.

GEMAN, D., GEMAN, S., and GRAFFIGNE, C. (1986). Locating texture and object boundaries, in *Pattern Recognition Theory and Applications*, Devijver ed., NATO ASI, Springer-Verlag, Heidelberg.

GRAFFIGNE, C. (1987). *Experiments in texture analysis and segmentation*. PhD thesis, Brown University.

GRAY, A. (1994). Simulating posterior Gibbs distributions: A comparison of the Swendsen-Wang and Gibbs sampler methods. *Statistics and Computing* **4** 189–201.

HERLIN, I., NGUYEN, C., and GRAFFIGNE, C. (1992). Stochastic Segmentation of ultrasound images, in *11th IAPR International Conference on Pattern Recognition*, IEEE Computer Society Press, **1** 289–292.

HURN, M. (1995). On the use of auxiliary variables in Markov chain Monte-Carlo methods, Tech. Rep. #95-07, Statistics Group at the University of Bath, School of Mathematical Sciences, University of Bath, Bath, BA2 7AY.

HURN, M. and JENNISON, C. (1993). Multiple-site updates in maximum a posteriori and marginal posterior modes image estimation, in *Advances in Applied Statistics: Statistics and Images*, Mardia and Kanji eds., Oxford - Carfax, 155–186.

MARTINELLI, F. (1992). Dynamical analysis of low-temperature Monte-Carlo cluster algorithms. *J. Stat. Physics* **66** 1245–1276.

MARTINELLI, F., OLIVIERI, E., and SCOPPOLA, E. (1991). On the Swendsen-Wang dynamics. I. Exponential convergence to equilibrium. *J. Stat. Physics* **62** 117–133.

MARTINELLI, F., OLIVIERI, E., and SCOPPOLA, E. (1991). On the Swendsen-Wang dynamics. II. Critical droplets and homogeneous nucleation at low temperature for the two-dimensional Ising models. *J. Stat. Physics* **62** 117–133.

SOKAL, A. D. (1989). *Monte-Carlo methods in statistical mechanics: Foundations and new algorithms*. Cours de troisième cycle de la physique en Suisse Romande, Lausanne.

SWENDSEN, R. H. and WANG, J. S. (1987). Nonuniversal critical dynamics in Monte-Carlo simulation. *Physical Review Letters* **58** 86–88.

WANG, J. (1994). *Multiscale Markov fields: applications to the segmentation of textured images and film fusion*, PhD thesis, Orsay University.

WANG, J. (1997). Stochastic relaxation on partitions with connected components and its application to image segmentation. Preprint CMLA, Cachan, France.

DIAM, CMLA, URA 1611, ÉCOLE NORMALE SUPÉRIEURE DE CACHAN, 61, AV. DU PRÉSIDENT WILSON, 94235 CACHAN AND UNIVERSITÉ PARIS-NORD, URA 742, AV. JEAN-BAPTISTE CLÉMENT, VILLETANEUSE. E-MAIL: gaudron@cmla.ens-cachan.fr