INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Image Denoising using Stochastic Differential Equations

Xavier Descombes — Elena Zhizhina

N° 4814

Mai 2003

THÈME 3

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4814--FR+ENG

Image Denoising using Stochastic Differential Equations

Xavier Descombes , Elena Zhizhina *

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

Rapport de recherche n° 4814 — Mai 2003 — 30 pages

Abstract: We address the problem of image denoising using a Stochastic Differential Equation approach. We consider a diffusion process which converges to a Gibbs measure. The Hamiltonian of the Gibbs measure embeds an interaction term, providing smoothing properties, and a data term. We study two discrete approximations of the Langevin dynamics associated with this diffusion process: the Euler and the Explicit Strong Taylor approximations. We compare the convergence speed of the associated algorithms and the Metropolis-Hasting algorithm. Results are shown on synthetic and real data. They show that the proposed approach provides better results when considering a small number of iterations.

Key-words: Stochastic Differential Equations, Image denoising, Euler approximation, Explicit Strong Taylor approximation

This work was partially supported by the Lyapunov Institute

* Institute for Information Transmission Problems, Bolshoy Karetny per. 19, 101447 Moscow, Russia

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Restauration d'image par Équations Différentielles Stochastiques

Résumé : Ce rapport concerne le problème de la restauration d'image avec une approche par Équation Différentielle Stochastique. Nous considérons un processus de diffusion convergeant vers une mesure de Gibbs. L'hamiltonien de la mesure de Gibbs contient un terme d'interactions, apportant des contraintes de lissage sur la solution, et un terme d'attache aux données. Nous étudions deux schémas d'approximation discrète de la dynamique de Langevin associée à ce processus de diffusion : les approximation d'Euler et explicite forte de Taylor. La vitesse de convergence des algorithmes correspondants est comparée à celle de l'algorithme de Metropolis-Hasting. Des résultats sont montrés sur des images de synthèse et réelles. Il montrent la supériorité de l'approche proposée lorsque l'on considère un faible nombre d'itérations.

Mots-clés : Équations Différentielles Stochastiques, Restauration d'image, Approximation d'Euler, Approximation explicite forte de Taylor.

1 Introduction

Image denoising is of fundamental importance for visualizing and interpreting images but also as a preprocessing step to improve the performance of image processing tasks such as classification, segmentation or feature extraction [1]. There exist numerous approaches to image denoising based on filtering. Indeed, using the assumption that the noise is a high frequency component of the image, early works concerning this subject exploited low pass filtering. However, high frequencies contain also some important features of an image, such as edge information. Simple low pass filtering results in blurring the image during the denoising process. Therefore, the trade-off consists in preserving edges and informative features while denoising.

There are two main approaches which address this problem: adaptive filtering and model based approaches. In the first method, the image is locally filtered using adaptive windows which avoid filtering over the edges [2]. In the second type of methods, a global model is defined by an energy function. The solution is then the configuration which minimizes in some sense the energy function. Deterministic approaches, consisting of solving a partial differential equation, converge to a local minimum of the energy [3]. These methods are known to be fast but depend on an initial configuration: if $N_1 \neq N_2$ are two different initial configurations, then in the general case we get two different images $I_1 \neq I_2$ under a deterministic procedure F_D

$$N_1 \stackrel{F_D}{\Longrightarrow} I_1, \quad N_2 \stackrel{F_D}{\Longrightarrow} I_2 \quad \text{with } I_1 \neq I_2$$

Stochastic methods have been proposed to avoid this dependency with respect to an initial configuration. In these approaches, the energy is interpreted as the Hamiltonian of a Gibbs field, which is minimized by a Metropolis-Hastings algorithm embedded in a simulated annealing scheme [4]. The counterpart is the speed of convergence which leads to costly iterative algorithms.

In this paper, we propose a new restoration algorithm, based on stochastic relaxation and annealing. We consider an interaction diffusion process and the associated Langevin dynamics as a stochastic relaxation process, described in section 2. It is a stationary Markov process reversible with respect to a background Gibbs distribution. Such diffusion processes have been proposed for matching images [5] or coupled with a jump process for features extraction [6, 7]. To simulate the Langevin dynamics, we consider discrete time Markov processes, which are called the Euler approximation and the Explicit Strong Taylor approximation of the Langevin dynamics. These approximations are described in section 3. The main idea of our approach is to combine the discrete approximations of the Langevin dynamics with some optimization criteria leading to the expectation with respect to the corresponding Gibbs measure or to the global minimum of the Hamiltonian (the denoised image) under given data (the degraded image). In section 4, we compare these two discretization processes with a standard Metropolis-Hasting approach. Firstly, the tests are conducted on synthetic data. We compare results after a small and a large number of iterations, and we observe a new more robust behavior of the Langevin dynamics algorithms with respect to the Metropolis-Hastings one when using a small number of iterations. That allows us to propose a new fast optimization scheme, which can be applied to image denoising problems. Finally, we consider tests on real data.

2 The Stochastic Differential Equation approach

2.1 A model for image restoration

Following the Bayesian approach based on Gibbs fields in image processing, we consider a Hamiltonian defined by the sum of an interaction term, modeling some prior knowledge about the solution, and a data driven term. We compare the Gaussian interaction term with the $\Phi-$model interaction term, which has been proved to preserve edges during the restoration process [8, 9, 10].

We consider a lattice spin system (the image lattice) in a volume $\Lambda \subset \mathbb{Z}^2, |\Lambda| = m$ with a continuous spin space (the grey level space) $S \subset \mathbb{R}^1$. Let S be a compact subset of $\mathbb{R}^1$, then $\Omega = S^m$ is the configuration space. The Hamiltonian of the model has the following form :

$$H(X,\theta) = \Phi_1(X) + \Phi_2(X,\theta), \tag{1}$$

where $X, \theta \in \Omega$, $X = \{X_i, i \in \Lambda\}$, and $\theta = \{\theta_i, i \in \Lambda\}$ is a fixed configuration (the "data"),

$$\Phi_1(X) = \beta \sum_{(i,j)\in\Lambda^2:|i-j|=1} U_1(X_i - X_j), \tag{2}$$

$$\Phi_2(X,\theta) = \lambda \sum_{i\in\Lambda} (X_i - \theta_i)^2, \tag{3}$$

$\beta > 0$ and $\lambda > 0$ are parameters of the model.

We first consider two test models:

$$\begin{aligned} TM1 : U_1^{(1)}(X_i - X_j) &= (X_i - X_j)^2; & (4)\\ TM2 : U_1^{(2)}(X_i - X_j) &= -\frac{1}{1 + \frac{(X_i - X_j)^2}{d^2}}. & (5) \end{aligned}$$

2.2 The method

The basic idea of our method is to construct the image as the limit configuration $\hat{X} \in \Omega$ of some iterative scheme $\{Y_n\}$. In this scheme the n-th configuration Y_n is found by a distribution depending on the previous configuration Y_{n-1} and the Hamiltonian of the model. To construct the process $\{Y_n\}$ we will use the following three steps.
1. First we construct a background continuous in time interaction diffusion process $X(t)$ on the configuration space Ω in such a way that the corresponding process on the space of measures converges to the Gibbs measure, defined by the Hamiltonian $H(X,\theta)$ of equation (1).
2. Then we consider an approximation of the process by a discrete time Markov process.

This discretization is necessary to derive algorithms which allow us to perform some computer simulations.
3. Finally, we define an estimator which optimize in some sense the above Hamiltonian. We apply a special procedure to the approximation process (simulated annealing or an expectation scheme) to find a configuration, giving either the global minimum of the Hamiltonian, so-called ground state for the corresponding spin system (MAP criterion), or the average over the Gibbs distribution (EXP criterion).

We will now introduce the diffusion process $X(t) = \{X_i(t), i \in \Lambda\}$, corresponding to the first step of our construction. Let us consider for any given $\sigma > 0$ a stationary process $\{X^\sigma(t), t \geq 0\}$ on Ω with the invariant measure

$$d\mu_\sigma(X) = \frac{e^{-\frac{2}{\sigma^2}H(X,\theta)}}{Z_\Lambda(\sigma)} d\mu_0(X), \tag{6}$$

where $\mu_0 = \times_{i\in\Lambda}\nu_0^{(i)}$ is the product of uniform distributions ν_0 on S. To describe this process we introduce its generator L_σ as an operator in the Hilbert space $\mathcal{L}_2(\Omega, d\mu_\sigma)$ of the functions on Ω:

$$L_\sigma f = \frac{1}{2}\sigma^2 \triangle f - \nabla H \cdot \nabla f = \frac{\sigma^2}{2}\sum_{i\in\Lambda}\frac{\partial^2 f}{\partial x_i^2} - \sum_{i\in\Lambda}\frac{\partial H}{\partial x_i}\cdot\frac{\partial f}{\partial x_i}, \quad f \in \mathcal{L}_2(\Omega, d\mu_\sigma).$$

The operator L_σ meets the detailed balance condition, which is equivalent to the self-adjointness of L_σ in the space $\mathcal{L}_2(\Omega, d\mu_\sigma)$ (see for instance [12]). Consequently, the operator L_σ generates a reversible stochastic process $X^\sigma(t)$ on Ω with respect to μ_σ. This process is called the Langevin dynamics and can be found as the solution of the following stochastic differential equation :

$$dX^\sigma(t) = a(X^\sigma(t))dt + \sigma dW(t), \quad t \geq 0, \tag{7}$$

where $\sigma > 0$ is the parameter and:

$$a(X) = a(X,\theta) = -\nabla_X H(X,\theta)$$

is a deterministic drift term, depending on the data θ, $dW(t)$ is a diffusive term, $W = \{W(t), t \geq 0\}$ is the m-dimensional Wiener process. To simplify the notations we will omit σ, so that in what follows $X \equiv X^\sigma$. The solution of (7) can be rewritten as :

$$X_i(t) = X_i(s) + \int_s^t a_i(X(u),\theta)du + \sigma\int_s^t dW(u), \quad i = 1,\ldots,m \tag{8}$$

with $m = |\Lambda|$, $0 < s < t$.

3 Algorithms

To simulate the process, we have to discretize the time. In this section, we present two approximation schemes to perform this discretization.

3.1 The Euler approximation

We consider a time discretization of the interval $(0, t)$:

$$\tau(\delta) = \{\tau_n, \quad n = 0, 1, 2, \dots, n_t\}$$

by time steps $\delta_n = \tau_{n+1} - \tau_n$. The approximation process

$$Y(n) = \{Y_i(n)\},\ i = 1, \dots, m; \quad n = 0, 1, \dots, n_t$$

has the same initial state $X(0)$ as the process $X(t)$, and can be constructed by the following iterative scheme:

$$\begin{aligned} &Y_i(0) = X_i(0), \\ &Y_i(n+1) = Y_i(n) + a_i(Y(n), \theta)\, \delta_n + \sigma\, (W(\tau_{n+1}) - W(\tau_n))\,. \end{aligned} \tag{9}$$

$W(\tau_{n+1}) - W(\tau_n)$ is simulated by sampling a centered normal law $\mathcal{N}(0, \delta_n)$ with a variance equal to δ_n.

Remark. *Let us remark that the process $Y(n)$ can be determined by the following transition distribution:*

$$\begin{aligned} &P(Y(n) \to Y(n+1)) = \\ &\frac{k^m}{(2\,\pi\,\delta_n\,\sigma^2)^{m/2}} \exp\left\{ -\frac{1}{2\,\delta_n\,\sigma^2} \sum_{i=1}^{m} (Y_i(n+1) - Y_i(n) - a_i(Y(n), \theta)\,\delta)^2 \right\}, \end{aligned}$$

where k^m is a normalizing constant, and σ is a given parameter which can be interpreted as the temperature of the Gibbs state. This formula refers to parallel sampling. All the results hold also for the sequential sampling.

3.2 The Strong Taylor approximation

Using the stochastic Taylor formula [11], we obtain from (8) an approximation process called the Strong Taylor scheme:

$$\begin{aligned} &Z_i(0) = X_i(0), \\ &Z_i(n+1) = Z_i(n) + a_i(Z(n), \theta)\, \delta_n + \tfrac{1}{2} a_i(Z(n), \theta)\, a_i'(Z(n), \theta)\, \delta_n^2 + \\ &\tfrac{1}{4} a_i''(Z(n), \theta)\, \sigma^2\, \delta_n^2 + a_i'(Z(n), \theta)\, \sigma\, \Delta V_i(n) + \sigma\, \Delta W_i(n), \end{aligned} \tag{10}$$

Here

$$\Delta V_i(n) = \frac{1}{2} \left(\xi_i^{(1)} + \frac{1}{\sqrt{3}} \xi_i^{(2)} \right) \delta_n^{3/2}, \quad \Delta W_i(n) = \xi_i^{(1)} \sqrt{\delta_n},$$

$\xi_i^{(1)}, \xi_i^{(2)},\ i = 1, \dots, m$ are independent identically distributed $\mathcal{N}(0, 1)$ random variables.

We also consider a modification of the Strong Taylor approximation, which is called *the Explicit Strong Taylor approximation*:

$$\begin{aligned} \tilde{Z}_i(0) &= X_i(0), \\ \tilde{Z}_i(n+1) &= \tilde{Z}_i(n) + \tfrac{1}{2}\left(a_i(\gamma(n),\theta) + a_i(\tilde{Z}(n),\theta)\right)\,\delta_n + \sigma\,\Delta W_i(n), \end{aligned} \tag{11}$$

with

$$\gamma(n) = \tilde{Z}(n) + a(\tilde{Z}(n),\theta)\,\delta_n + \sigma\,\Delta\tilde{W}(n),$$

and

$$\Delta W_i(n) = \xi_i^{(1)}\sqrt{\delta_n}, \quad \Delta\tilde{W}_i(n) = \xi_i^{(2)}\sqrt{\delta_n},$$

$\xi_i^{(1)}, \xi_i^{(2)}$ are i.i.d. $\mathcal{N}(0,1)$ random variables.

To perform the different tests, we considered the last approximation as well as the first order scheme (9). This explicit strong Taylor approximation avoids the computation of high order derivatives of the Hamiltonian which may lead to numerical instabilities.

3.3 Convergence and ergodicity properties

Using a modification of the reasoning from [11] we will show in this section that the approximation processes (9)-(11) converge in the strong sense. We assume that
1) $a_i(X,\theta) \in C(S^{2m})$ is a continuous function on X and θ for any $i = 1,\dots,m$;
2) $a_i(X,\theta)$ depends only on the configurations X and θ in a neighborhood of the site i:

$$a_i(X,\theta) = a_i(X_j,\theta_j), \quad j : |i-j| \le c_0$$

with some c_0.

We note, that under our definitions of $a_i(X,\theta)$ and $H(X,\theta)$, where

$$a_i(X,\theta) = -\frac{\partial}{\partial X_i} H(X,\theta),$$

both the assumptions 1) and 2) above are valid. These assumptions imply the following bounds

$$\max_i \max_{X,\theta} |a_i(X,\theta)| \le K_1, \tag{12}$$

$$\max_{\theta\in S^m} |a_i(X,\theta) - a_i(\tilde{X},\theta)| \le K_2 \left(|X_i - \tilde{X}_i| + \sum_{j:|j-i|=1} |X_j - \tilde{X}_j| \right) \tag{13}$$

for some constants K_1, K_2. We denote by τ_{n_s} the nearest discretization point to s:

$$|\tau_{n_s} - s| < \delta_{n_s}.$$

In order to prove convergence we require the following lemma.
Lemma 1 (Gronwall inequality) [11]. *Let* $\alpha, \beta : [t_0, T] \to R$ *be integrable with*

$$0 \leq \alpha(t) \leq \beta(t) + C \int_{t_0}^{t} \alpha(s) ds$$

for $t \in [t_0, T]$ *with* $C > 0$*. Then*

$$\alpha(t) \leq \beta(t) + C \int_{t_0}^{t} e^{C(t-s)} \beta(s) ds$$

for $t \in [t_0, T]$*.*

Theorem 1 (Strong convergence).

$$\textit{Let} \;\; \delta = \max_{k=1,\ldots,n_t} \delta_k.$$

Then under assumptions 1)-2) on the functions $a_i(X, \theta)$*,* $i = 1, \ldots, m$*, the approximation process (9) converges strongly to the process* $X(t)$ *as* $\delta \to 0$ *with the order* $1/2$ *which means that, for any* t*, we have the following:*

$$\max_i E\left(|X_i(t) - Y_i(n_t)|\right) \leq C\delta^{1/2} \tag{14}$$

C *being a positive constant which does not depend on* δ *(but depends on* t*).*
Proof. We give a modification of the proof in [11]. Denote:

$$Z(t) = \max_i \sup_{0 \leq s \leq t} E\left(|X_i(s) - Y_i(n_s)|\right). \tag{15}$$

Then the representations (8) and (9) imply

$$
\begin{aligned}
Z(t) &= \max_i \sup_{0\le s\le t} E\left(\left|\sum_{n=0}^{n_s-1}(Y_i(n+1)-Y_i(n))\right.\right. \\
&\qquad\qquad \left.\left. -\int_0^s a_i(X(u),\theta)du - \sigma\int_0^s dW(u)\right|\right) \\
&= \max_i \sup_{0\le s\le t} E\left(\left|\sum_{n=0}^{n_s-1} a_i(Y(n),\theta)\cdot\delta_n - \int_0^s a_i(X(u),\theta)du\right.\right. \\
&\qquad\qquad \left.\left. -\sigma\int_0^s dW(u) + \sigma\int_0^{\tau_{n_s}} dW(u)\right|\right) \\
&= \max_i \sup_{0\le s\le t} E\left(\left|\sum_{n=0}^{n_s-1}\left(a_i(Y(n),\theta)\cdot\delta_n - \int_{\tau_n}^{\tau_{n+1}} a_i(X(u),\theta)du\right)\right.\right. \\
&\qquad\qquad \left.\left. -\int_{\tau_{n_s}}^s a_i(X(u),\theta)du - \sigma\int_{\tau_{n_s}}^s dW(u)\right|\right) \\
&\le \max_i \sup_{0\le s\le t}\left\{E\left(\left|\sum_{n=0}^{n_s-1}\int_{\tau_n}^{\tau_{n+1}}(a_i(Y(n),\theta)-a_i(X(u),\theta))du\right|\right)\right. \\
&\qquad \left. +E\left(\int_{\tau_{n_s}}^s |a_i(X(u),\theta)|du\right) + E\left(\left|\sigma\int_{\tau_{n_s}}^s dW(u)\right|\right)\right\}
\end{aligned}
\tag{16}
$$

Since $|\tau_{n_s}-s|<\delta_{n_s}$ we have for any s and i:

$$
E\left(\int_{\tau_{n_s}}^s |a_i(X(u),\theta)|du\right) \le K_1\delta_{n_s} \le K_1\delta, \tag{17}
$$

$$
E\left(\left|\sigma\int_{\tau_{n_s}}^s dW(u)\right|\right) \le \sigma\sqrt{\delta_{n_s}} \le \sigma\sqrt{\delta}. \tag{18}
$$

We now estimate the first term in (16) using (13):

$$
\begin{aligned}
&\left|\sum_{n=0}^{n_s-1}\int_{\tau_n}^{\tau_{n+1}}(a_i(Y(n),\theta)-a_i(X(u),\theta))du\right| \\
&\qquad = \left|\int_0^{\tau_{n_s}}(a_i(Y(n_u),\theta)-a_i(X(u),\theta))du\right| \\
&\qquad \le \int_0^{\tau_{n_s}}|(a_i(Y(n_u),\theta)-a_i(X(u),\theta))|\,du \\
&\qquad \le K_2\int_0^{\tau_{n_s}}\left(|Y_i(n_u)-X_i(u)| + \sum_{j:|j-i|=1}|Y_j(n_u)-X_j(u)|\right)du
\end{aligned}
$$

Hence,

$$\max_i \sup_{0\le s\le t} E\left(\left|\sum_{n=0}^{n_s-1}\int_{\tau_n}^{\tau_{n+1}}(a_i(Y(n),\theta)-a_i(X(u),\theta))du\right|\right)$$
$$\le 5K_2\int_0^{\tau_{n_t}} Z(u)du \le 5K_2\int_0^t Z(u)du. \tag{19}$$

Using (16)-(19) we have

$$Z(t)\le 5K_2\int_0^t Z(u)du + K_1\delta+\sigma\sqrt{\delta}.$$

Finally, by the Gronwall inequality (see Lemma 1) we obtain for small enough $\delta>0$ the following bound

$$Z(t)\le C\sqrt{\delta}\cdot\sigma,$$

where $C=C(t)$ does not depend on δ, which proves theorem 1.

By similar considerations using the Taylor-Ito expansion and more restrictions on the smoothness of the functions $a_i(X,\theta)$ one can prove the strong convergence for the processes (10)-(11) with the order 1.5 and 1.0 respectively.

Corollary (Uniform weak convergence). *For each $f\in C(S^m)$ and any $t>0$, we have the following:*

$$\max_{X\in S^m}\sup_{0\le s\le t}|E_X f(Y([s/\delta]))-E_X f(X(s))|\to 0 \quad \textit{as} \quad \delta\to 0, \tag{20}$$

where E_X is a conditional mean under the condition that X is the initial state.

By our constructions the process $X(t)$ is ergodic:

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T f(X_s^{X_0})ds=\int f(y)d\mu_\sigma(y) \tag{21}$$

for any μ_σ-integrable function f and for a.s. initial conditions X_0. We denote:

$$F^\delta=\lim_{T\to\infty}\frac{1}{[\frac{T}{\delta}]}\sum_{n=0}^{[\frac{T}{\delta}]-1} f(Y(n)), \tag{22}$$

and:

$$F=\int f(y)d\mu_\sigma=\lim_{T\to\infty}\frac{1}{T}\int_{n=0}^T f(X(s))ds. \tag{23}$$

Then the following ergodic convergence criterion holds for the Euler scheme [13]: for small enough δ

$$|F^\delta-F|\le C_f\delta, \tag{24}$$

where C_f is a positive constant which does not depend on δ. Due to the interpretation of F as $E(f(X_\infty))$ and of F^δ as $E(f(Y_\infty))$ the ergodic convergence criterion can be considered as an extension of the weak convergence criterion to the infinite time $T = \infty$.

3.4 Convergence diagnosis

To control the path-wise convergence we consider a statistical test proposed by Brooks and Gelman in [14]. We run n simulations independently. For each simulation we consider m iterations. To test the convergence, we compare the statistics of the m_0 next samples obtained with the different simulations independently with the statistics of the set defined by the union of the n simulations. We compute these statistics on single pixels.

Consider a given pixel. Denote x_k^i the pixel value at the k^{th} iteration in the i^{th} simulation. Consider the set $X_m^i = \{x_k^i, k = m, ..., m + m_0\}$ for some m_0. Discard the $\alpha\%$ lowest and the $\alpha\%$ highest values in X_m^i. Denote the minimal and the maximal values of the resulting set by A_m^i and B_m^i, respectively. Consider now the set $X_m = \cup_{i=1,..,n} X_m^i$. Discard the $\alpha\%$ lowest and the $\alpha\%$ highest values in X_m. Denote the minimal and the maximal values of the resulting set by A_m and B_m, respectively . We consider the following ratio:

$$R_m = \frac{B_m - A_m}{\sum_{i=1,..n} A_m^i/n - \sum_{i=1,..n} B_m^i/n}$$

If the algorithm converge, we have:

$$\lim_{m\to\infty} R_m = 1$$

For the different experiments we plot R_m as a function of m and consider that the convergence is reached when R_m is close to 1.

We compare the Langevin dynamics with the Metropolis-Hasting algorithm [15, 16]. Figure 1 shows an example on one pixel trajectory of the evolution of R_m with respect to m ($m_0 = 100$) for the model TM2 (5). The Euler approximation exhibits the fastest convergence speed and the Metropolis-Hasting algorithm the slowest one (note that an iteration requires approximately the same amount of CPU time for the Euler approximation as for the Metropolis-Hasting algorithms and about twice as much for the explicit strong Taylor approximation). Therefore, the Langevin dynamics approximations appear to converge faster than the Metropolis-Hasting algorithm for simulating a Gibbs distribution.

4 Application to image denoising

In this section, we are not only interested in making choice of the model for image denoising, but we need to optimize the model in some sense. We define herein two criteria for which we derive estimators from the different dynamics under study.

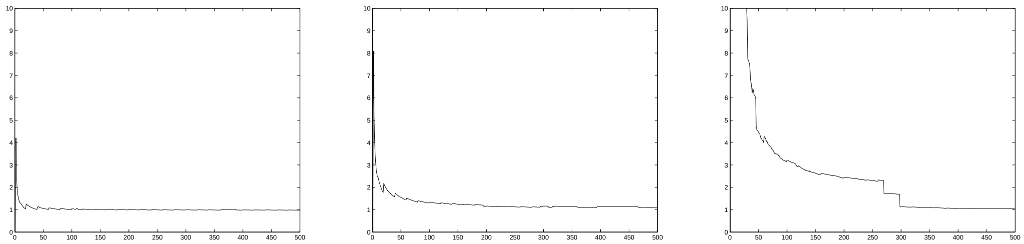


Figure 1: R_m as a function of m ($m_0 = 100$) for the Euler approximation, the Explicit Strong Approximation and the Metropolis-Hasting Algorithm, respectively

4.1 Two criteria

The first criterion, refered as the Maximum A Posteriori (MAP) criterion in the Bayesian framework, consists in minimizing the Hamiltonian defined in equation (1):

$$\hat{X}_1 = \arg\min_X H(X, \theta) \tag{25}$$

To estimate $\hat{X}_1$ we apply a simulated annealing scheme where the parameter σ decreases during iterations [17, 4]. For the Langevin dynamics the results of [11] implies that the solution of equation (7) under $\sigma = \sigma(t) = \frac{1}{\sqrt{\ln(t+2)}}$ converges to $\hat{X}_1$ as $t \to \infty$. The optimal value (in the sense of a fast convergence) of the time discretization step δ depends on the value of σ. We therefore have also to consider a decreasing scheme for the parameter δ. In practice, we considered an exponential decreasing scheme for both σ and δ.

As follows from the above ergodicity arguments, we can consider a criterion (EXP) based on the estimation of the expectation:

$$\hat{X}_2 = E_\mu(X) \tag{26}$$

To obtain the estimate $\hat{X}_2$, we run the Langevin dynamics for a given fixed σ until convergence. The estimate $\hat{X}_2$ is then given as follows:

$$\hat{X}_2 = \frac{1}{N} \sum_{n=M}^{M+N} Y(n)$$

with some large enough M and N, so that by ergodicity property (22) we have: $\hat{X}_2 \sim E(Y_\infty)$.

4.2 Results on a test image

We first consider a synthetic image consisting of several uniform areas (see table 1.a) on which we have added a centered Gaussian noise with standard deviation equal to 50 (see table 1.b).

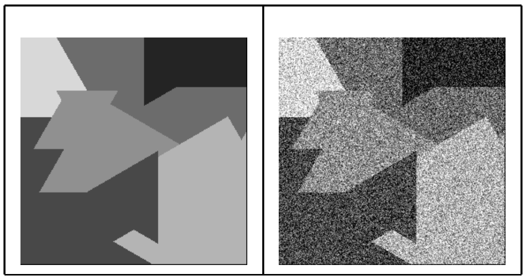
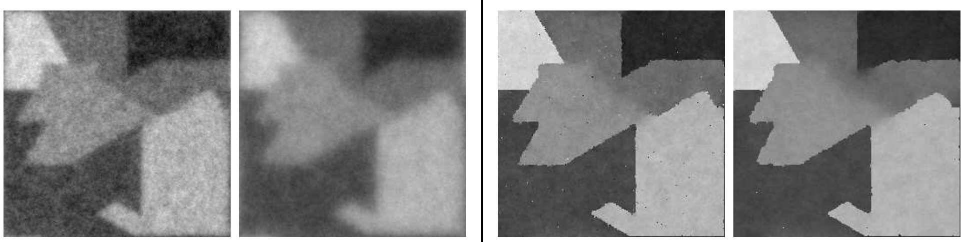


Table 1: a (left): First test image, b (right): Noisy image $\sigma = 50$

We first compare, in table 2, the two different models (4)-(5), using the Euler approximation and the MAP criterion. The Gaussian prior (model TM1) leads to a blurred image. For $\beta = 0.001$, the result is under-regularized as the difference image shows that there is still some noise. When increasing the weight of the interactions ($\beta = 0.005$), edges are blurred. The model behaves as a low pass filter and as such cuts the high frequencies representing edges. Model TM2 seems to be the most adequate for image denoising as it allows us to remove noise while preserving edges.



Model 1: $\lambda = 0.0001$ $\beta = 0.001$ and $\beta = 0.005$	Model 2: $d = 10$, $\lambda = 0.0001$ $\beta = 1.5$ and $\beta = 2.0$

Table 2: Result for the 2 models studied using different parameters

In table 3, we compare two proposed criteria (MAP and EXP) using two different approximations (the Euler and Taylor schemes). Considering the value of the final energy, the MAP criterion is better than the EXP criterion, which is not surprising as the MAP criterion consists in minimizing the energy. However, a visual inspection shows that results derived by two above criteria are very similar. Small differences between the energy is not translated into differences in the image quality. The Taylor and Euler gives similar results at this stage, i.e. with a high number of iterations.

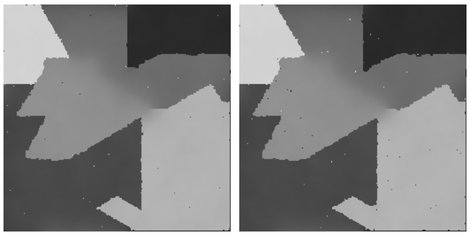
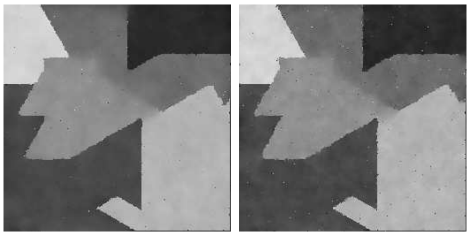
	
MAP : Euler and Taylor app. $H_f = -181145$ and $H_f = -181034$	EXP : Euler and Taylor app. $H_f = -178560$ and $H_f = -175080$

Table 3: Result using model 2 ($\beta = 1.5$, $d = 10$, $\lambda = 0.0001$) for the Euler and Taylor approximations using MAP and EXP criteria (20000 iterations)

It is not surprising to obtain similar results with a high number of iterations. The main point of this study is to consider a small number of iterations to overcome the main drawback of stochastic methods, i.e. long computational time. We thus study the robustness of the different dynamics with respect to the number of iterations. Is the optimality, guarantied by the convergence theorems, still valid when considering a small number of iterations ? We compare the results for the test image using simulations by the MAP and EXP criteria with the Metropolis-Hastings algorithm (MHA) and with the Langevin dynamics algorithms (LDA) after different numbers of iterations (see table 4 for 10000 iterations and table 5 for 1000 iterations).

One can see that the convergence by the LDA scheme is faster than by the MHA scheme on the first stage of the simulations for both criteria, as denoted by the final energy value. Moreover, the obtained local minima have different properties with the LDA schemes and with the MHA scheme. Using the LDA schemes, with 1000 iterations, the result appears to be globally similar to the optimum obtained with 10000 iterations, except that the image still has pixelwise noise. This phenomenon is less appreciable with the EXP criterion. Using the MHA scheme we do not get this pixelwise noise but the different areas are not as smooth as the optimum. To remove the pixelwise noise, we propose to apply the median filtering on the final stage of the simulations. Table 6 presents the result of simulations by the MAP and the EXP criterion using MHA and LDA by 1000 iterations after filtering.

The filtering evidently improves the result of LDA, but is almost helpless in the MHA case. Finally, as for 10000 iterations, the Taylor approximation is not essentially better than the Euler approximation.

To show the final energy of the model with respect to the iteration number for the different schemes (without or with filtering), we collect our results on the energy behaviour in tables 7 and 8.

The shape of these plots implies again, that the convergence is much faster for the LDA scheme. Although we do not get the optimal value, the energy is, after a small number

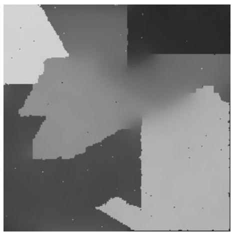
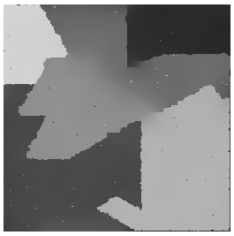
MAP criterion : $T, \sigma : 1000 \rightarrow 0.001$	 Euler $\delta : 1000 \rightarrow 0.1$ $H_f = -245404$	 Taylor $\delta : 1000 \rightarrow 0.1$ $H_f = -245385$	 Metropolis $H_f = -245018$
EXP criterion :	 Euler $T = 1.0, \delta = 10$ $H_f = -244928$	 Taylor $T = 1.0, \delta = 10$ $H_f = -244954$	 Metropolis $T = 0.5$ $H_f = -241139$

Table 4: Result for the different algorithms and criteria using a high number of iterations (10000), Model 2, $\beta = 2.0$, $d = 10$, $\lambda = 0.0001$

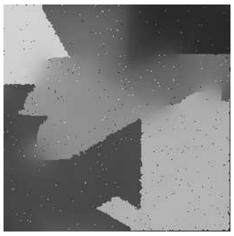
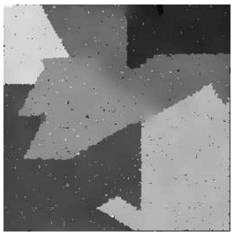
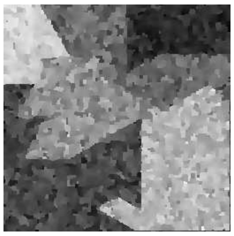
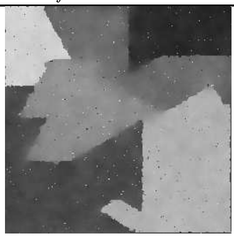
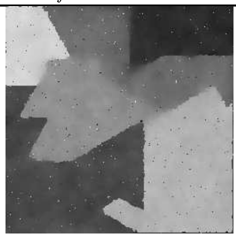
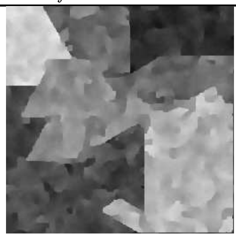
MAP criterion :	 Euler $\sigma : 10 \rightarrow 0.01$ $\delta : 1000 \rightarrow 5$ $H_f = -242820$	 Taylor $\sigma : 10 \rightarrow 0.01$ $\delta : 1000 \rightarrow 5$ $H_f = -241719$	 Metropolis $T : 500 \rightarrow 0.001$ $H_f = -221939$
EXP criterion :	 Euler $T = 1.0, \delta = 20$ $H_f = -237813$	 Taylor $T = 1.0, \delta = 20$ $H_f = -238169$	 Metropolis $T = 0.5$ $H_f = -229612$

Table 5: Result for the different algorithms and criteria using a small number of iterations (1000), Model 2, $\beta = 2.0$, $d = 10$, $\lambda = 0.0001$

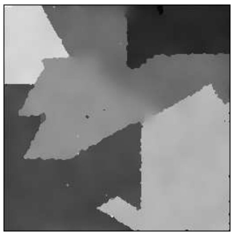
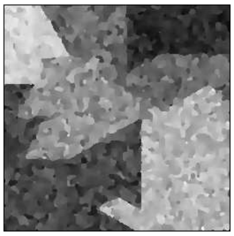
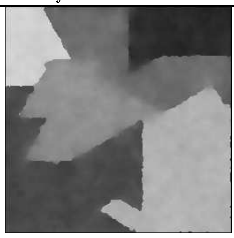
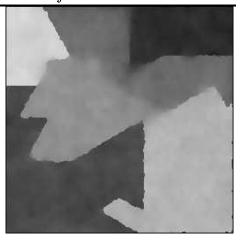
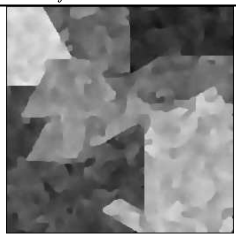
MAP criterion :	 Euler $\sigma : 10 \rightarrow 0.01$ $\delta : 1000 \rightarrow 5$ $H_f = -243499$	 Taylor $\sigma : 10 \rightarrow 0.01$ $\delta : 1000 \rightarrow 5$ $H_f = -243147$	 Metropolis $T : 500 \rightarrow 0.001$ $H_f = -219963$
EXP criterion :	 Euler $T = 1.0, \delta = 20$ $H_f = -241021$	 Taylor $T = 1.0, \delta = 20$ $H_f = -241233$	 Metropolis $T = 0.5$ $H_f = -230552$

Table 6: Result for the different algorithms and criteria using a small number of iterations (1000) followed by a 3×3 median filter, Model 2, $\beta = 2.0$, $d = 10$, $\lambda = 0.0001$

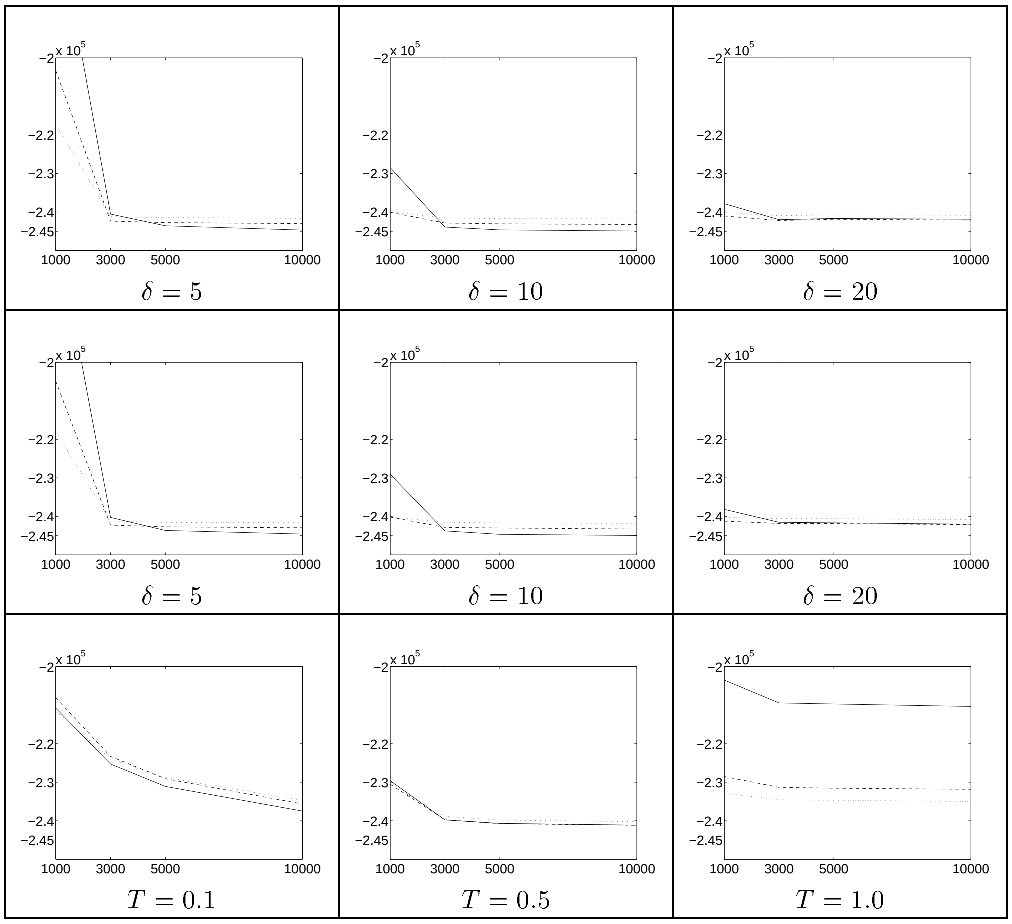


Table 7: Energy w.r.t. the number of iterations for the EXP criterion for Euler (top row), Taylor (middle row) and Metropolis (bottom row) : without filtering (plain line), with a 3×3 median filter (dashed line) and with a 5×5 median filter (dotted line)

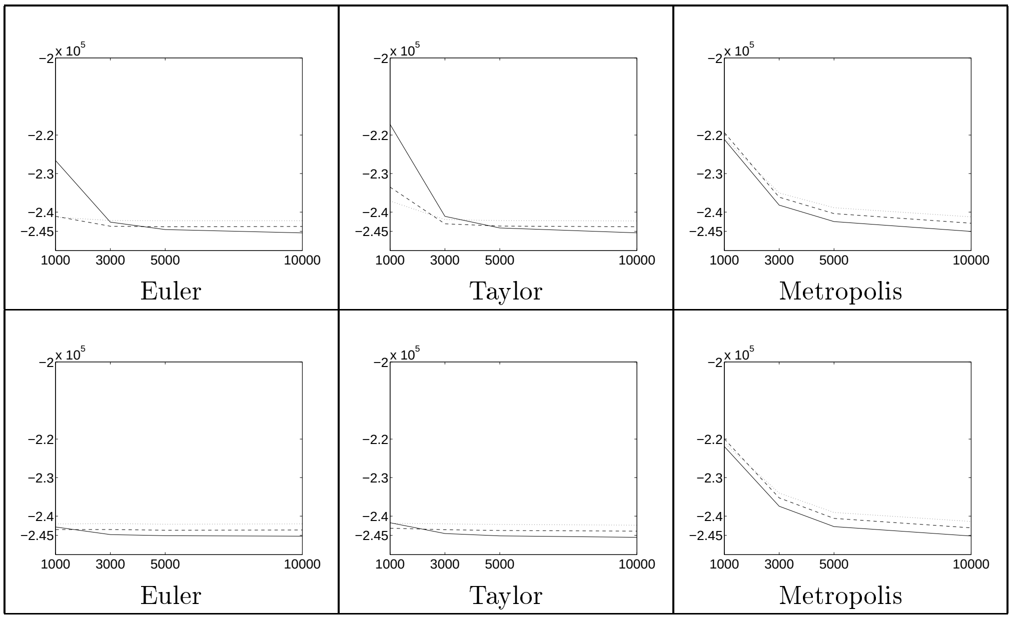


Table 8: Energy w.r.t. the number of iterations for the MAP criterion : top row, for $T = 1000 \rightarrow 0.001$ (Metropolis) and $T = 10 \rightarrow 0.01, \delta = 1000 \rightarrow 0.01$ (Euler and Taylor), and bottom row, for $T = 500 \rightarrow 0.001$ (Metropolis) $T = 10 \rightarrow 0.01, \delta = 1000 \rightarrow 5.0$ (Euler and Taylor) : without filtering (plain line), with a 3×3 median filter (dashed line) and with a 5×5 median filter (dotted line)

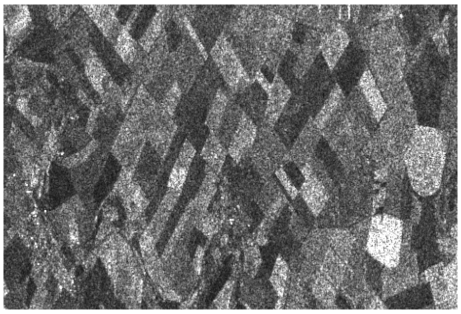


Figure 2: 4-look ERS1 image

of iterations, lower for the LDA schemes than for the MHA scheme. Moreover, applying a median filter reduces the energy to a value close to the optimal one in the case of the LDA schemes.

4.3 Tests on real data

We first consider a radar image (a 4-look ERS1 image shown in figure 2). Radar images contain a speckle noise which is a multiplicative, correlated noise. In this paper, we do not consider a speckle model but keep the quadratic term for the data term. However, such a simple model has proved to be efficient. We compare the Euler and Metropolis schemes on this real severely noised image for a small number of iterations (3000, 1000 and 300 on tables 9, 10 and 11, respectively). Considering both the image quality and the energy value, the MHA provides slightly better results for 3000 iterations. However, when applying a median filter on the result, both schemes are equivalent. The main point is that when we reduce the number of iterations, down to 300, the LDA scheme appears to be more robust, since there is almost no degradation of the result, that is not the same in the MHA scheme case. Besides, in the LDA algorithm, the median filter improves the result of the LDA schemes and we get an energy close to the optimal one.

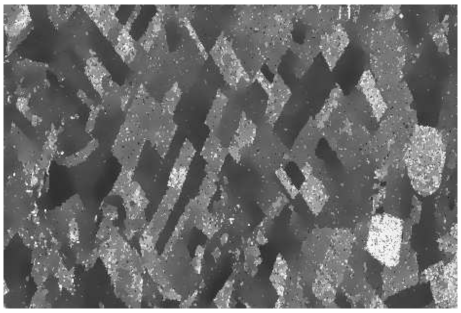
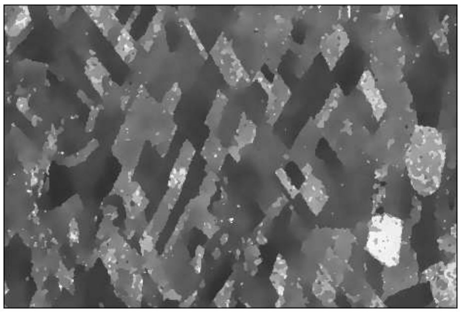
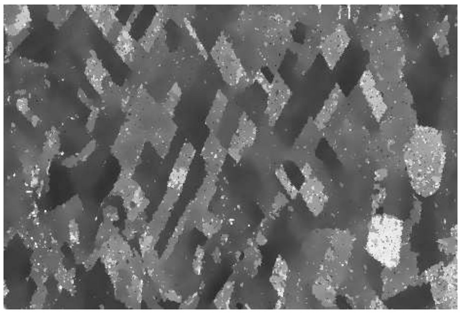
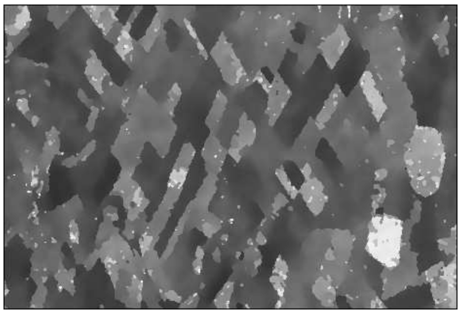
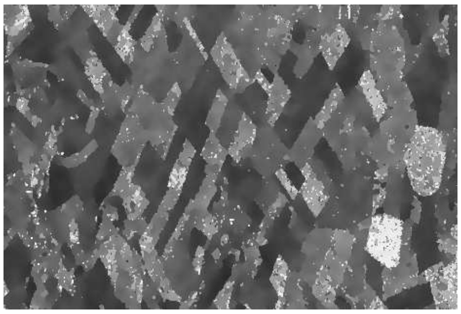
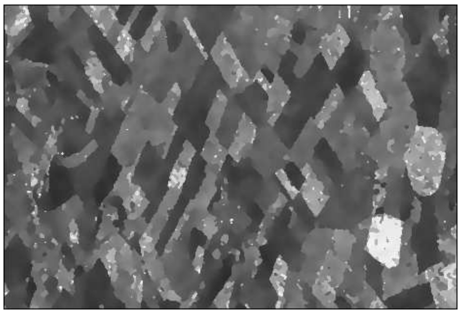
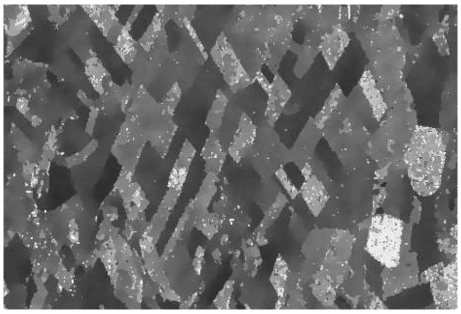
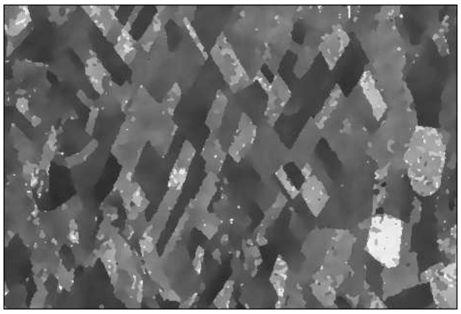
 Euler with MAP criterion $E = -213845$	 plus 3×3 median filter $E = -201382$
 Euler with EXP criterion $E = -216915$	 plus 3×3 median filter $E = -203780$
 Metropolis with MAP criterion $E = -220075$	 plus 3×3 median filter $E = -202661$
 Metropolis with EXP criterion $E = -218675$	 plus 3×3 median filter $E = -203780$

Table 9: Result on image shown in figure 2 for 3000 iterations

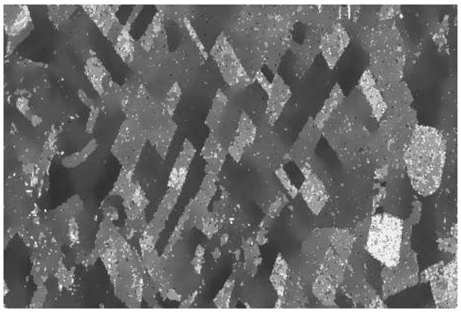
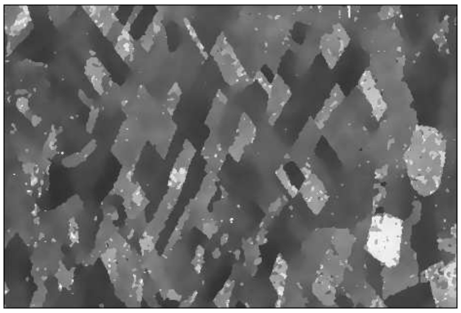
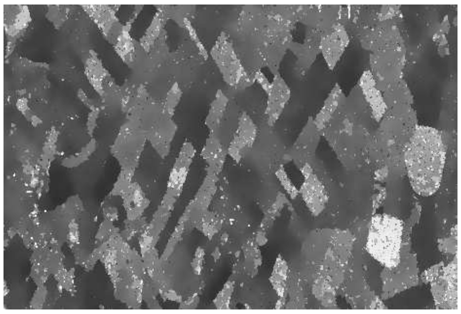
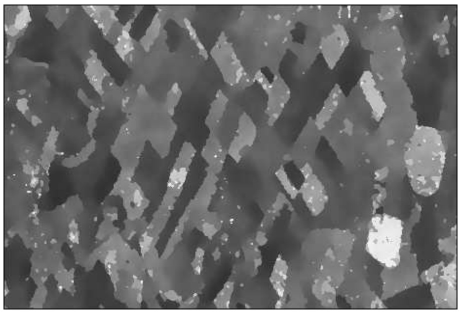
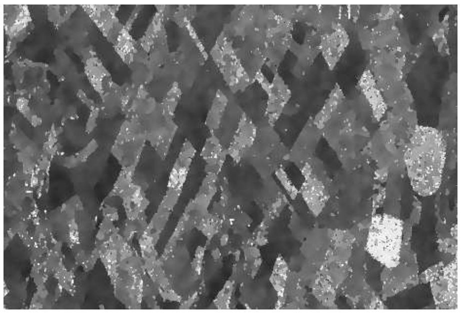
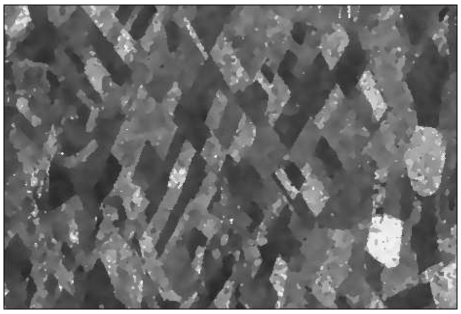
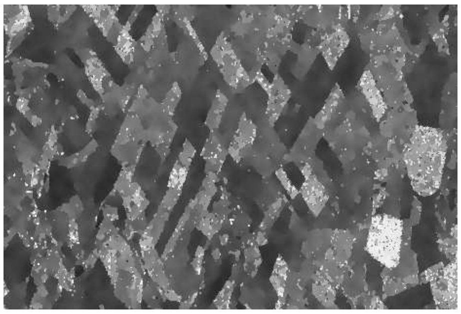
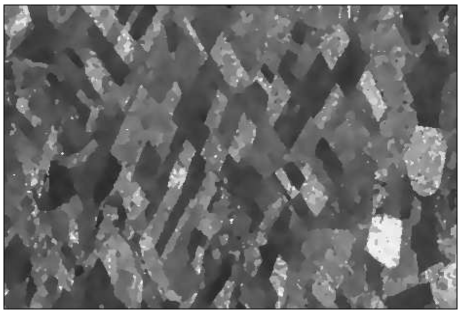
 Euler with MAP criterion $E = -218142$	 plus 3×3 median filter $E = -203800$
 Euler with EXP criterion $E = -216022$	 plus 3×3 median filter $E = -203456$
 Metropolis with MAP criterion $E = -214695$	 plus 3×3 median filter $E = -197734$
 Metropolis with EXP criterion $E = -213948$	 plus 3×3 median filter $E = -197767$

Table 10: Result on image shown in figure 2 for 1000 iterations

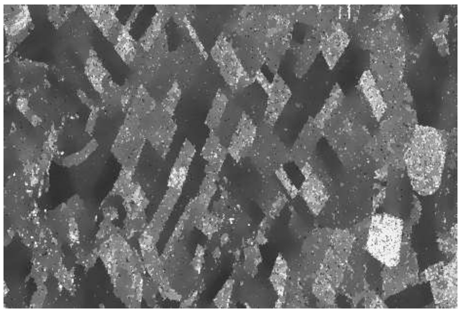
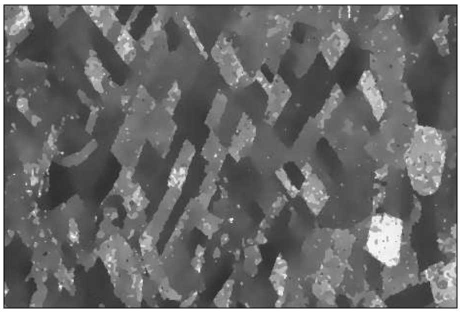
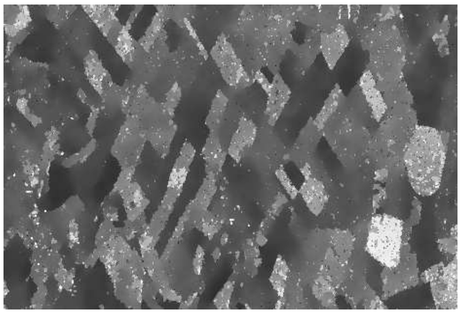
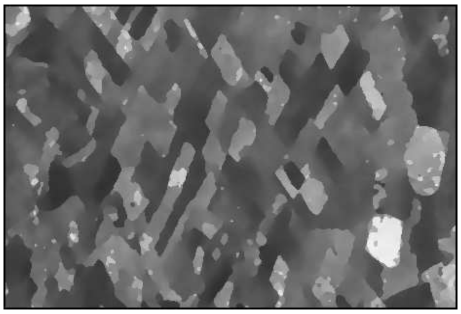
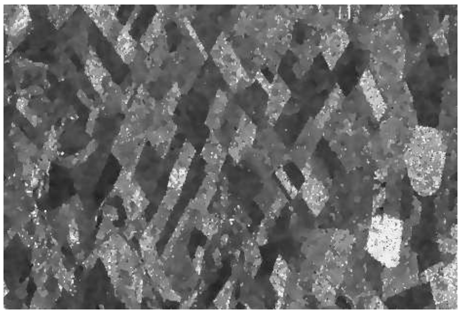
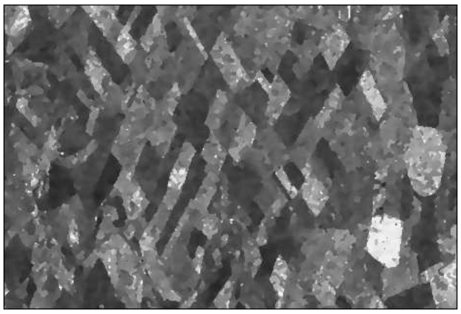
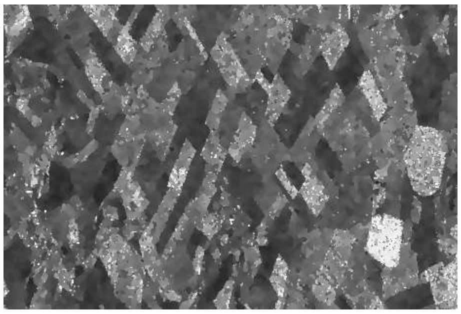
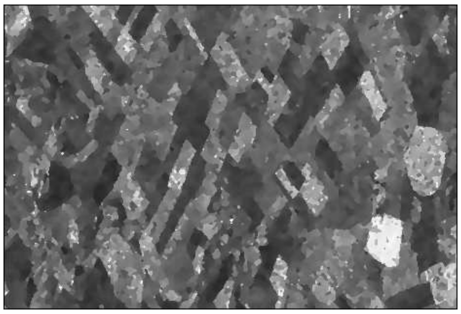
 Euler with MAP criterion $E = -209869$	 plus 3×3 median filter $E = -199785$
 Euler with EXP criterion $E = -215052$	 plus 3×3 median filter $E = -203251$
 Metropolis with MAP criterion $E = -204654$	 plus 3×3 median filter $E = -190264$
 Metropolis with EXP criterion $E = -203618$	 plus 3×3 median filter $E = -190332$

Table 11: Result on image shown in figure 2 for 300 iterations

We finally consider the lena picture on which we have added a Gaussian noise with $\sigma = 50$ (see table 12). We test our restoration schemes on this example where edges are not as clearly defined as in the previous tests (one can see some textures and non uniform illumination on the different areas). We only consider 50 iterations in order to have a very fast restoration process (8 seconds on a 1 GHZ processor for a 256×256 image). Here again, the Euler scheme, followed by a 3×3 median filter, provides the best results.The EXP criterion here is slightly better than the MAP criterion (see tables 13 and 14).

Table 12: Lena picture 256×256 and noisy Lena ($\sigma = 50$)

 Euler $E = -94703$	 plus 3×3 median filter $E = -92759$
 Metropolis $E = -88504$	 plus 3×3 median filter $E = -90226$

Table 13: Result on image shown in figure 12 for 50 iterations with the MAP criterion

 Euler $E = -98194$	 plus 3×3 median filter $E = -93968$
 Metropolis $E = -92220$	 plus 3×3 median filter $E = -90795$

Table 14: Result on image shown in figure 12 for 50 iterations with the EXP criterion

5 Conclusion

In this paper, we have proposed the use of stochastic differential equations (SDE) for image denoising. The Langevin dynamics associated with SDE allows us to sample a Gibbs distribution defined by a given Hamiltonian modeling the searched solution. Therefore a link is obvious with the Markov Random Field approach commonly used in image processing. Besides, the considered SDE can be interpreted as a stochastic version of the partial derivative equation approach employed in image processing. The corresponding diffusion process associated with SDE is constructed by the drift term, which is the same as in the deterministic version, and in addition, by the stochastic term, which is the Wiener process. That provides fluctuations allowing us to consider some algorithms converging to the global minimum of the Hamiltonian. We have studied two discrete approximations of the continuous process, the Euler and the Taylor schemes. These approximations lead to two iterative algorithms giving some estimators of the solution (expectation and maximum a posteriori ones). We compared two approximation schemes with the Metropolis-Hastings dynamics being widely used in the Markov Random Field approaches. Some experiments have been conducted on both synthetic and real images.

All the schemes under consideration present very similar final results after a large enough number of iterations (around 10 000), so that all algorithms (Euler LDA, Taylor LDA, MHA) are comparable. However, our goal is to compare these algorithms when using a small number of iterations to reduce the computation time which is the main drawback of stochastic methods. We have tested the robustness of these approaches with respect to the number of iterations. The result of the simulations by MAP or EXP criterion using LDA after 50-1000 iterations is better than the analogous result for MHA. One can see the main improvement on the visual quality of the results, whereas the energies of the corresponding local minima have minor difference. Indeed, the local minima obtained with the LDA approaches are similar to the results obtained with a greater number of iterations except in some isolated points containing a residual noise. The residual noise can be easily removed with a simple median filtering. Applying a median filter we obtain a result, which is close w.r.t. the visual and energy inspection to the final image (after 20 000 iterations by either the LDA or the MHA scheme). That means that we can reduce the number of iterations without loss of quality of the final result using the LDA scheme followed by filtering.

Comparing results, obtained by the Euler and Taylor approximations, we can conclude that there is no main difference in the convergence properties for the models under consideration, and we prefer to use the Euler approximation as the faster one. SDE appears to be a good tool to sample a Gibbs distribution and therefore to construct algorithms solving image processing problems. They are a nice compromise between the deterministic variational approaches, which are fast but sub-optimal, and the stochastic MCMC approach, which is optimal but slow.

We are going to continue the study of the SDE approach in image processing problems, such as parameter estimation or image segmentation.

References

[1] Andrews H., Hunt B., Digital image restoration, Prentice Hall, Englewood Cliffs, NJ, 1977.

[2] Lee J.S., Digital image smoothing and the sigma filter, Computer Vision, Graphics, Image Processing, vol. 24, pp. 255-269, 1983.

[3] Rudin L.I., Osher S., Fatemi E., Nonlinear total variation based noise removal algorithms, Physica D, vol. 60, pp. 259-268, 1992.

[4] Geman S., Geman D., Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images, IEEE Trans. Pattern Anal. Machine Intelligence, vol. 6, No. 6, pp. 721-741, 1984.

[5] Salomon M., Perrin G.R., Heitz F., Parallel sampling with stochastic differential equations for 3D deformable matching of medical images, 2002 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2002), Las Vegas, USA, June 24-27, 2002.

[6] Grenander U., Miller M.I., Representations of knowledge in complex systems, Journal of the royal Stat. Soc. Series B, vol. 56, no, 4, 1994.

[7] Lanterman A.D. , Miller M.I. , D.L. Snyder D.L., Implementation of jump-diffusion algorithms for understanding FLIR scenes, in Proc. SPIE, vol. 2485, 309-320, 1995.

[8] Geman S., Reynolds G., Constrained restoration and recovery of discontinuities, IEEE Trans. Pattern Anal. Machine Intelligence, vol. 14, n0. 3, pp. 367-383, 1992.

[9] Nikolova M., Regularisation functionss and estimators, In proc. IEEE Int. Conf. Ima. Proc., ICIP-96, pp. 457-460, 1996.

[10] Descombes X., Kruggel F., von Cramon, D.Y., fMRI signal restoration using a spatio-temporal Markov Random Field preserving transitions, Neuroimage, vol. 8, pp. 340-349, 1998.

[11] Kloeden P., Platen E., Numerical solution of stochastic differential equations, Springer-Verlag, 1992

[12] Liggett Th., Interacting particle systems, Springer-Verlag, 1985.

[13] Talay D., Second order discretization schemes of stochastic differential systems for the computation of the invariant law. Stochastics and Stochastic Reports, vol. 29, no. 1, pp. 13-36, 1990.

[14] Brooks S.P., Gelman A., General Methods for Monitoring Convergence of Iterative Simulations, J. of Computational and Graphical Statistics, 7:4, pp. 434-455, 1998.

[15] Hastings W.K., Monte Carlo sampling methods using Markov chains and their applications, Biometrika, vol. 57, pp. 97-109, 1970.

[16] Winkler G., Image analysis, random fields and dynamic Monte Carlo methods: a mathematical introduction, Springer-Verlag, 1995.

[17] Kirkpatrick S., Gelatt C.D., Vecchi M.P., Optimization by simulated annealing, Science, vol. 220, pp. 671-680, 1983.

Contents

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399