1

Gibbs fields approaches in image processing problems

X. Descombes [†], E. Zhizhina [‡]

Abstract In this paper, we address the problem of image denoising using a stochastic differential equation approach. Proposed stochastic dynamics schemes are based on the property of the diffusion dynamics to converge to a distribution on global minima of the energy function of the model, under a special cooling schedule (the annealing procedure). To derive algorithms for computer simulations we consider discrete time approximations of the stochastic differential equation. We study convergence of the corresponding Markov chains to the diffusion process. We give conditions for the ergodicity of the Euler approximation scheme. In the conclusion, we compare results of computer simulations using the diffusion dynamics algorithms and the standard Metropolis-Hasting algorithm. Results are shown on synthetic and real data.

1 Introduction. Bayesian approach in image restoration problems.

One of the main problems of image processing is to create fast algorithms which can be applied to image denoising and restoration problems. Image denoising is of fundamental importance for visualizing and interpreting images

[*]The work is partially supported by French-Russian A.M. Liapunov Institute for Applied Mathematics and Computer Science (Grant 98-02), INTAS (Grant "Stochastic Analysis" 99-00559), RFBR (Grant 02-01-00444), Scientific School Program (Grant 934.2003.1)

[†]Ariana, Joint group, CNRS/INRIA/UNSA, INRIA, 2004, route des Lucioles, BP93, 06902, Sophia-Antipolis cedex, France

[‡]Institute for Information Transmission Problems, Bolshoy Karetny per. 19, 101447 Moscow, Russia

but also as a preprocessing step to improve the performance of image processing tasks such as classification, segmentation or feature extraction. There exist numerous approaches to image denoising based on filtering. Indeed, the noise is typically a high frequency component of the image. 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.

Besides filtering, model based approaches have been traditionally used in image processing problems. The most popular method of this type is the Bayesian approach based on the analysis of a posteriori distribution on the configuration space of the model. The goal of the Bayesian image restoration is to find a configuration which maximizes a posteriori distribution. This configuration is accepted as the denoised image. There exist different methods in the framework of the Bayesian approach. They depends on the form of a posteriori distribution and on the way how to obtain the denoised image (deterministic or stochastic). Deterministic methods are known to be fast but depend on an initial configuration, and stochastic methods have been proposed to avoid this dependency with respect to an initial configuration. We consider here stochastic algorithms under a posteriori Gibbs distribution with an energy function (the Hamiltonian of the Gibbs field) depending on given noisy image. Our restoration algorithm is the further application of the Gibbs field methods in image processing. It is based on the relaxation properties of a diffusion dynamics, which is constructed as a reversible process with respect to the background Gibbs distribution. Furthermore, we give a general form of the generator of the reversible diffusion dynamics, which can be described as an adaptive diffusion process with inhomogeneous over the space coefficients depending on the current configuration.

The basic idea of our method is to construct the denoised image as the limit configuration of a stochastic iterative scheme. In this scheme, a next configuration on each iteration step is found by a distribution depending on the current configuration, the energy of the model and the current parameters of the model (the temperature and the step of the time discretization). The choice of the energy function plays a crucial role in the Gibbs fields approaches. We consider a case, when the energy function depends on given noisy image.

Let us remark that in our algorithms we propose a double annealing procedure, when the temperature and the discretization step are slowly de-

creased to zero during iterations. We formulate conditions on the speed of decrease for both parameters and prove the ergodicity of the corresponding inhomogeneous Markov chain using Dobrushin's criterion of ergodicity.

The paper is organized as follows. A spin model for image restoration is introduced in section 2. In section 3 we consider a stochastic diffusion dynamics on the configuration space with a stationary Gibbs distribution. We discuss here also a global optimization scheme called the annealing procedure. Under this cooling procedure the stationary process transforms to a non-stationary one, and it asymptotically tends to one of low energy configurations. To simulate the diffusion dynamics, we consider discrete time Markov processes, which are approximated the diffusion dynamics in time. These approximations are described in section 4. The convergence and ergodicity properties of the approximation schemes are discussed in section 5 and 6. In section 7, using computer simulations we compare the new diffusion based algorithms with the standard Metropolis – Hastings approach. First, the tests are conducted on synthetic data, then we consider the tests on real data. We compare results after a small and a large number of iterations. The results after a large number of iterations for all schemes have comparable quality. However, we observe a new more robust behavior of the diffusion 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.

2 A model for image restoration

The idea to apply Gibbs fields and corresponding stochastic dynamics to image denoising problems has been proposed in 80 years (see [1]), and since then Gibbs fields methods have been intensively developed in a framework of the Bayesian image restoration. The realization of Gibbs fields approach in image processing was made possible after the denoising procedure has been rewritten on the language of statistical physics. Under this view a digital image can be considered as a configuration of a Gibbs random field on the lattice with a single spin, running values on the real line or on a subset of the real line. We identify the values of the single spin with the intensity of a color (from black to white) at given point of the image.

Let us introduce now the model. We consider a lattice spin system (the image lattice) in a finite volume $\Lambda \subset \mathbb{Z}^2, |\Lambda| = m$ of two dimensional lattice

with a single spin space (the grey level space) $S = [0, s] \subset \mathbb{R}^1$. Then S^m is the configuration space, and any configuration

$$X \in S^m, \quad X = \{X_i,\ i \in \Lambda\}, \quad X_i \in S,$$

corresponds to some image.

Our goal is to find a denoised image, which is constructed as a solution of an optimization scheme minimizing noise on given data. Following the Bayesian approach, the required configuration should maximize a posteriori distribution on the configurations, i.e.

$$\overline{X} \text{ is a solution, if } \quad P(\overline{X}|\theta) = \max_X P(X|\theta), \tag{1}$$

where $\theta = \{\theta_i, i \in \Lambda\}$ denotes given data. A posteriori distribution has the Gibbs form:

$$P(X|\theta) = \frac{1}{Z_\sigma} e^{-\frac{2}{\sigma^2} H(X,\theta)}, \tag{2}$$

with the energy function $H(X, \theta)$ and the normalizing factor Z_σ, σ is the parameter of the model. Given data are introduced to the potential of the model by so-called self-interaction term. The energy function is defined by the sum of an interaction term, modeling some prior knowledge about the solution, and a data driven term:

$$H(X, \theta) = \Phi_1(X) + \Phi_2(X, \theta). \tag{3}$$

Here

$$\Phi_1(X) = \beta \sum_{\langle i,j \rangle \in \Lambda^2 : |i-j|=1} U_1(X_i - X_j), \tag{4}$$

$$\Phi_2(X, \theta) = \lambda \sum_{i \in \Lambda} (X_i - \theta_i)^2, \tag{5}$$

$\beta,\ \lambda > 0$ are parameters of the model. The choice of the interaction potential derives from the following properties of the required solution. On the one hand the solution should be smooth in the regions of homogeneity, and on the other hand the information about boundaries of objects on the image should be very precise. In particular, the potential of the form

$$U_1(X_i - X_j) = (X_i - X_j)^2,$$

leads to blurred image during the denoising process. The Φ-interaction term introduced in [2]

$$U_1(X_i - X_j) = -\frac{1}{1 + \frac{(X_i - X_j)^2}{d^2}},$$

is more adapted to preserve edges during the restoration process. We will use later this Φ-potential in the simulations.

Thus, under the Gibbs fields approach with a posteriori distribution (2) in the Gibbs form, the solution of the problem (1) is a configuration (or configurations) minimizing the total energy of the system:

$$E_{\min} = \arg\min H(X, \theta), \quad \overline{X} \in E_{\min}. \tag{6}$$

Either deterministic or stochastic way is usually exploited to find these configurations. Deterministic algorithms are based on the variational principle, the stochastic ones use ergodic properties of stochastic dynamics. In this paper we will study algorithms of the second type which depend on behavior of stochastic systems.

3 Stochastic dynamics based algorithms

The Metropolis – Hastings (MH) scheme has served up to now as a good tool for image restoration problems. MH algorithm is based on properties of the Glauber dynamics for the Potts models in combination with simulated annealing (see [1, 3, 4]). This small modification of the Glauber dynamics, simulating a slow cooling procedure, yields an algorithm which theoretically finds a set of global minimizers $E_{\min}$ (6), independently on starting configurations:

$$\lim_{n \to \infty} P(X(n) \in E_{\min}) = 1. \tag{7}$$

Let us remind that the MH algorithm is associated with the following single spin dynamics. If we denote by p a proposal distribution on the spin space $S = \{s_1, \ldots, s_k\}$ defined by a symmetric transition matrix $p_{x,y}$, then one can randomly picks a new configuration value $\tilde{X}_i \in S$ at the site $i \in \Lambda$ by p. We denote by Y a new configuration

$$Y = \{X_1, \ldots, \tilde{X}_i, \ldots X_m\},$$

which differs from the configuration X only at one site i. Then the new configuration Y is accepted with probability

$$Q_{X,Y} = \exp\left\{-\frac{2}{\sigma_n^2}[H(Y,\theta) - H(X,\theta)]^+\right\},$$

where $[u]^+ = u$ as $u \geq 0$ and $[u]^+ = 0$ as $u < 0$, and correspondingly, not accepted with probability $1 - Q_{X,Y}$. In such a manner, values of the new configuration $X(n+1) = \{X_i(n+1)\}$ are chosen consequently over all sites of the volume Λ. In addition the parameter σ_n is decreased to zero by a special way as $n \to \infty$ during the process (so called cooling schedule). It was proved in [1, 3] that (7) holds under

$$\lim_{n\to\infty} \sigma_n^2 \log n > R,$$

where the constant R depends on the energy function H.

If we take a model with a continuous compact single spin space, then for any σ we can consider a diffusion dynamics reversible with respect the the given Gibbs distribution, so that the Gibbs measure is the stationary measure of the process. Diffusion processes of this sort have been studied in many papers, even on the whole lattice, see for instance, [5, 6]. Then as above we modify the stationary process using the cooling procedure to obtain a non-stationary process meeting (7). It was proved in [7] - [9], that again (7) holds if the parameter $\sigma(t)$ is vanishing as $t \to \infty$ not faster than

$$\sigma^2(t) = \frac{C}{\log t}, \quad t \to \infty, \tag{8}$$

where $C > 0$ is large enough. The proof exploited on ideas and constructions of the Wentzell – Freidlin theory [10] - [12] developed for the analysis of invariant measures of diffusion processes with a small diffusion coefficient.

The corresponding diffusion processes are described by a system of stochastic interaction differential equations (see (12) below). We will not pose the problem to find a solution in the explicit form, since we consider a general class of stochastic differential equations. Our goal is to construct an approximation for the solution of any such an equation. To do this we will study discretizations (on time) for the diffusion processes, i.e. we will propose approximation schemes and prove the convergence of discrete time processes to the continuous time diffusion process.

Let us denote by $\Omega = [-s, s]^m$, and let $\mathfrak{L}_2(\Omega, d\mu_\sigma)$ be the Hilbert space of functions on the configuration space Ω with periodical boundary conditions. The scalar product is defined by the Gibbs measure

$$d\mu_\sigma(X) = \frac{e^{-\frac{2}{\sigma^2}H(X,\theta)}}{Z_\Lambda(\sigma)} d\mu_0(X), \tag{9}$$

where $d\mu_0 = \prod_{i\in\Lambda} d\nu_i$ is the reference measure, ν is the normalized Haar measure on $[-s, s]$ and $Z_\Lambda(\sigma)$ is the normalizing factor. The generator L_σ of the reversible with respect to the Gibbs measure (9) diffusion process on Ω, i.e. such that (see [13]),

$$(L_\sigma f,\ g)_{\mu_\sigma} = (f,\ L_\sigma g)_{\mu_\sigma} \tag{10}$$

has the following general form:

$$(L_\sigma f)(X) = \frac{\sigma^2}{2} \sum_{i,j=1}^{m} b_{i,j}(X) \frac{\partial^2 f}{\partial X_i \partial X_j}(X) + \frac{\sigma^2}{2} \sum_{i,j=1}^{m} \frac{\partial b_{i,j}}{\partial X_i}(X) \frac{\partial f}{\partial X_j}(X) -$$

$$- \sum_{i,j=1}^{m} b_{i,j}(X) \frac{\partial H}{\partial X_i}(X,\theta) \frac{\partial f}{\partial X_j}(X),$$

where the matrix $B(X) = \{b_{i,j}(X)\}$ for any X is symmetric and positive definite. Condition (10) in particular implies that the process generated by L_σ has a stationary distribution (9).

In what follows we will consider the identity diffusion matrix $B(X) = E$, in this case the drift term depends only on the gradient of the energy function:

$$L_\sigma f = \frac{\sigma^2}{2} \triangle f - \nabla H \cdot \nabla f = \frac{\sigma^2}{2} \sum_{i\in\Lambda} \frac{\partial^2 f}{\partial X_i^2} - \sum_{i\in\Lambda} \frac{\partial H}{\partial X_i} \cdot \frac{\partial f}{\partial X_i}, \quad f \in \mathfrak{L}_2(\Omega, d\mu_\sigma). \tag{11}$$

The operator L_σ is the generator of a process $X^\sigma(t)$ on Ω reversible with respect to μ_σ. This process is said to be diffusion or Langevin dynamics, and it can be represented as a solution of stochastic differential equations

$$dX^\sigma(t) = a(X^\sigma(t))dt + \sigma dW(t), \quad t \geq 0. \tag{12}$$

Here $\sigma > 0$ is the parameter setting the inverse temperature of the system, the function

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

determines a drift dependent on the data θ and on the current configuration X, $dW(t)$ is the diffusion term, $W = \{W(t), t \geq 0\}$ is a m-dimensional Wiener process. We won't use later on the index σ in notations for $X^\sigma(t)$, however we will keep in mind that the solution depends on σ.

Following the above reasoning we get that for typical realizations of (12) the process $X(t)$ as $t \to \infty$ and $\sigma = \sigma(t) \to 0$ under (8) converges to one of the global minimizers $\overline{X} \in E_{\min}$ of the energy function $H(X, \theta)$. This is a crucial property for the ground of stochastic dynamics based algorithms.

4 Approximations for the diffusion dynamics

Let us rewrite the solution of (12) in the integral representation:

$$X_i(t) = X_i(s) + \int_s^t a_i(X(u), \theta)du + \sigma \int_s^t dW_i(u), \quad i = 1, \ldots, m, \tag{13}$$

with $m = |\Lambda|$, $0 < s < t$. To find the solution of (12) using computer simulations we have to approximate the continuous time process (13) by a discrete time process and to derive corresponding algorithms. Using the Ito formula for the stochastic process $X(t)$ satisfying (12) (see for instance, [14]) we have for $s < u < t$ and each $i = 1, \ldots, m$:

$$X_i(t) = X_i(s) + \int_s^t a_i(X(s), \theta)du + \sigma \int_s^t dW_i(u) +$$

$$+ \int_s^t \int_s^u \left(\frac{\partial a_i}{\partial X_i}(X(y), \theta) a_i(X(y), \theta) + \frac{\sigma^2}{2} \frac{\partial^2 a_i}{\partial X_i^2}(X(y), \theta) \right) dydu +$$

$$+ \int_s^t \int_s^u \frac{\partial a_i}{\partial X_i}(X(y), \theta) \sigma dW_i(y)du \cong$$

$$\cong X_i(s) + a_i(X(s), \theta)(t - s) + \sigma(W_i(t) - W_i(s)) +$$

$$+ \left(\frac{\partial a_i}{\partial X_i}(X(s), \theta) a_i(X(s), \theta) + \frac{\sigma^2}{2} \frac{\partial^2 a_i}{\partial X_i^2}(X(s), \theta) \right) \frac{1}{2}(t - s)^2 + \tag{14}$$

$$+\sigma \frac{\partial a_i}{\partial X_i}(X(s),\theta) \int_s^t \int_s^u dW_i(y)du.$$

Now two first approximations follow directly from (14). We will present them in the next subsections, see also [15].

Let us remark that all approximation schemes under consideration below are Markov chains with the state space $\mathbb{R}^m$. If we take the periodic continuation of Ω on $\mathbb{R}^m$, then we can identify the process on Ω with the corresponding factorization of the process on $\mathbb{R}^m$.

4.1 The Euler approximation

The Euler approximation is the simplest discretization scheme considering terms up to the order $(t-s)$ in the decomposition (14). Let

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

be a time discretization of the interval $(0,t)$ by time steps $\delta_n = \tau_{n+1} - \tau_n$. The Euler approximation

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

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

$$Y_i(n+1) = Y_i(n) + a_i(Y(n),\theta)\,\delta_n + \sigma\,(W_i(\tau_{n+1}) - W_i(\tau_n))\,. \qquad (15)$$

Here $W_i,\ i = 1, \ldots, m$ are independent random variables, and the random variable $\xi_n = W(\tau_{n+1}) - W(\tau_n)$ is distributed by a centered normal law $\mathcal{N}(0, \delta_n)$ with a mean 0 and a variance δ_n.

4.2 The Strong Taylor approximation

Let us denote by

$$\zeta(t-s) = \int_s^t \int_s^u dW(y)du$$

a random variable incoming to the last term of the decomposition (14). Then

$$1)\ \mathrm{E}\zeta(t-s) = 0,$$

$$2)\ \mathrm{E}\zeta^2(t-s) = \tfrac{1}{3}(t-s)^3,$$

$$3)\ \mathrm{E}(\zeta(t-s)(W(t)-W(s))) = \tfrac{1}{2}(t-s)^2.$$

Therefore, we can use the following representation to generate the random variable $\zeta(\delta)$:

$$\zeta(\delta) = \frac{1}{2}\left(\xi^{(1)} + \frac{1}{\sqrt{3}}\xi^{(2)}\right)\delta^{3/2}, \tag{16}$$

Here $\xi^{(1)}, \xi^{(2)}$ are independent identically distributed by $\mathcal{N}(0,1)$ random variables. Using again (14) and the above discretization of the time interval $(0,t)$ we obtain an approximation process $Z(n)$ called the Strong Taylor approximation:

$$\begin{aligned} &Z_i(0) = X_i(0), \\ &Z_i(n+1) = Z_i(n) + a_i(Z(n),\theta)\,\delta_n + \tfrac{1}{2}a_i(Z(n),\theta)\,a_i'(Z(n),\theta)\,\delta_n^2 + \\ &+\tfrac{1}{4}a_i''(Z(n),\theta)\,\sigma^2\,\delta_n^2 + a_i'(Z(n),\theta)\,\sigma\,\zeta_i(\delta_n) + \sigma\xi_i(\delta_n). \end{aligned} \tag{17}$$

Here $\xi_i(\delta_n) = \xi_i^{(1)}\sqrt{\delta_n}$, random variables $\zeta_i(\delta)$ ¨ $\xi_i^{(1)}$ are defined in (16), and $\xi_i^{(1)}, \xi_i^{(2)}$, $i = 1,\ldots,m$, are independent identically distributed by $\mathcal{N}(0,1)$ random variables.

The iteration formula (17) is not convenient for computer simulations since the computation of high order derivatives of the energy function may lead to numerical instabilities. The following modification of the formula (17) serves to avoid the computation of high order derivatives. The corresponding two-step iteration scheme is called *the Explicit Strong Taylor approximation*:

$$\begin{aligned} &\tilde{Z}_i(0) = X_i(0), \\ &\tilde{Z}_i(n+1) = \tilde{Z}_i(n) + \tfrac{1}{2}\left(a_i(\gamma(n),\theta) + a_i(\tilde{Z}(n),\theta)\right)\,\delta_n + \sigma\,\xi_i(n). \end{aligned} \tag{18}$$

Here

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

and

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

with independent identically distributed by $\mathcal{N}(0,1)$ random variables $\xi_i^{(1)}, \xi_i^{(2)}$.

4.3 α-approximation

We consider an equidistant discretization of the interval $(0,t)$ by a time step δ:

$$\tau(\delta) = \left\{ n\delta, \quad n = 0,1,2,\ldots,n_t = [\frac{t}{\delta}] \right\}.$$

The α-approximation process

$$U_n = \{U_i(n),\ i = 1,\ldots,m,\ n = 0,1,\ldots,n_t\}$$

is defined as follows:

$$\begin{aligned} &U_i(0) = X_i(0), \quad i = 1,\ldots,m, \\ &U_i(n+1) = U_i(n) + a_i(U(n),\theta)\,\delta + \eta_i^\alpha, \end{aligned} \tag{19}$$

where $X(0)$ is the initial state of $X(t)$, and $\eta_i^\alpha,\ i = 1,\ldots,m$ are independent identically distributed random variables with a mean 0 and a variance $\mathrm{D}\eta_i^\alpha = \delta\sigma^2$, defined by the probability distribution:

$$p_{\eta^\alpha}(u) = k_\alpha(\delta)e^{-b_\alpha(\delta)|u|^\alpha}, \quad u \in \mathbb{R}, \quad 0 < \alpha \le 2. \tag{20}$$

The constants $k_\alpha(\delta)$ and $b_\alpha(\delta)$ are determined by the above conditions:

$$\int p_{\eta^\alpha}(u)du = 1, \quad \int u^2 p_{\eta^\alpha}(u)du = \delta\sigma^2, \tag{21}$$

which imply

$$k_\alpha(\delta) = \frac{c_1(\alpha)}{\sigma\sqrt{\delta}}, \quad b_\alpha(\delta) = \frac{c_0(\alpha)}{\sqrt{(\delta\sigma^2)^\alpha}}.$$

The constants $c_0(\alpha)$ and $c_1(\alpha)$ depend only on α:

$$c_0(\alpha) = \left(\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right)^{\frac{\alpha}{2}}, \quad c_1(\alpha) = \frac{\alpha}{2\Gamma(1/\alpha)}\left(\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right)^{\frac{1}{2}}.$$

5 Convergence of the approximation schemes

We will state in this section results on the convergence of the approximation processes under constant σ. 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 assumptions 1) and 2) above are valid. These assumptions imply the following bounds

$$\max_i \max_{X,\theta} |a_i(X,\theta)| \le K_1, \tag{22}$$

$$\max_{\theta\in S^m} |a_i(X,\theta) - a_i(\tilde{X},\theta)| \le K_2 \left(|X_i - \tilde{X}_i| + \sum_{j:|j-i|=1} |X_j - \tilde{X}_j| \right) \tag{23}$$

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

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

Then the following theorem about the strong convergence holds.

Theorem 1. *Let*

$$\delta = \max_{k=1,\dots,n_t} \delta_k,$$

then under assumptions (22), (23) *approximation processes* (15), (17), (18) *converge strongly to* $X(t)$ *as* $\delta \to 0$ *with order* $\frac{1}{2}$, $\frac{3}{2}$ *and* 1 *respectively: for any* t

$$\max_i \mathrm{E}\left(|X_i(t) - Y_i(n_t)|\right) \le C\delta^{1/2}, \tag{24}$$

C *being a positive constant which does not depend on* δ *(but depends on* t*). The analogous estimates are true for processes* (17), (18).

The proof of theorem 1 see in [16]. It follows the standard reasoning from [15] modified for the case of a finite system of interacting stochastic

processes and in so doing makes use of inequality (22), (23) to estimate the difference between the diffusion process and its approximations on each interval of the time discretization.

Since processes (13) and (19) are defined on different probability spaces we state a result on a uniform weak convergence of the processes U_n and $X(t)$.

Theorem 2. *For each continuous function $f \in C(S^m)$ on S^m and any $t > 0$*

$$\max_{x \in S^m} \max_{0 \le s \le t} |\mathrm{E}_x f(U(n_s)) - \mathrm{E}_x f(X_s)| \to 0 \quad \bar{a}\ddot{} \ \delta \to 0, \tag{25}$$

where E_x is a conditional mean under the condition that x is the initial state.

The proof of theorem 2 see in [17]. The main idea of the proof is to establish convergence of the generators of the processes U_n and $X(t)$ as $\delta \to 0$. The generator L of the diffusion process (13) is given by (11), and the generator of the process (19) – (20) has the following form

$$L^\delta f(y) = \frac{1}{\delta} \left(\int_{S^m} k_\alpha^m(\delta) \prod_{i=1}^m \left(\sum_{n \in Z} \exp\left\{ -b_\alpha(\delta) |x_i + 2ns - \tilde{y}_i|^\alpha \right\} \right) f(x) dx - f(y) \right), \tag{26}$$

with

$$\tilde{y}_i = y_i - a_i(y)\delta, \quad i = 1, \ldots, m. \tag{27}$$

Using the Taylor expansion of the function $f(y)$ in a neighborhood of $\tilde{y}_i$ one can see that for each $f \in C^4(S^m)$ $f(y)$

$$L^\delta f(y) = Lf(y) + g^\delta(y),$$

with $\max\limits_y |g^\delta(y)| \to 0$ as $\delta \to 0$. This formula together with the approximation technique (see [18]) imply weak convergence (25) of the corresponding processes in the uniform norm of the Banach space of continuous functions on S^m.

6 Ergodic properties of approximations

In this section, we will study ergodic properties of the Euler approximation scheme $Y(n)$ (15) as the parameters σ and δ are decreased to zero. This discrete time scheme is corresponding to the Euler approximation of a non-stationary diffusion process (12) under a certain cooling schedule, when $\sigma = \sigma(t) \to 0$.

Consider an inhomogeneous Markov chain which is analogous to (15):

$$Y_i(n+1) = Y_i(n) + a_i(Y(n), \theta)\,\delta_n + \sigma_n \xi_n. \tag{28}$$

Here $\sigma_n \to 0$, $\delta_n \to 0$ are sequences of vanishing positive numbers, and the random variables ξ_n are distributed on $[-s, s]$ by

$$p(x) = \frac{1}{\sqrt{2\pi\delta}\sigma} \sum_{n\in\mathbb{Z}} \exp^{-\frac{1}{2\delta\sigma^2}(x+2sn)^2}, \quad x \in [-s, s]. \tag{29}$$

Denote by $P_n^{(i)}(x, A),\ n = 1, 2, \ldots,$ transition functions of the Markov chain (28), and let $P_{k,n}^{(i)}(x, A)$ be a probability to jump from the position $x \in S$ at time k to the set A at time n. The Markov chain (28) is said to be ergodic, if for any $k,\ i,\ x_1, x_2 \in S$

$$\sup_A \lim_{n\to\infty} \left| P_{k,n}^{(i)}(x_1, A) - P_{k,n}^{(i)}(x_2, A) \right| = 0.$$

Theorem 3. *If $\sigma_n\sqrt{\delta_n} \to 0$ in such a way that*

$$\sum_n \frac{1}{\sigma_n\sqrt{\delta_n}} e^{-\frac{s^2}{2\sigma_n^2\delta_n}} = \infty, \tag{30}$$

then the Markov chain (28) *is ergodic.*

The proof of theorem 3 is based on Dobrushin's ergodic criterion, see [19]. The results of [19] imply that the following condition

$$\sum_{n=1}^{\infty} \alpha_n = \infty,$$

is necessary and sufficient for the ergodicity of the inhomogeneous Markov chain with transition functions $P_n(x, A)$. Here $\alpha_n = \alpha(P_n)$ is the ergodic coefficient for the transition function P_n, which is given by the following way (see [19, 9]):

$$\alpha(P) = 1 - \frac{1}{2} \sup_{x,y} ||P(x, \cdot) - P(y, \cdot)||_{\mathrm{Var}},$$

with the norm of total variation $||\cdot||_{\mathrm{Var}}$.

It follows from (28), (29) that for any $i = 1, \ldots, m$ the density $p_n^{(i)}(x, z)$ of the transition function $P_n^{(i)}(x, A)$ is bounded away from zero uniformly over i and x:

$$m_n = m_n^{(i)} = \inf_z p_n^{(i)}(x, z) \geq \frac{2s}{\sigma_n\sqrt{2\pi\delta_n}} e^{-\frac{s^2}{2\sigma_n^2\sigma_n}}.$$

Consequently, for any i and any x, y we have

$$\left| |P_n^{(i)}(x,\cdot) - P_n^{(i)}(y,\cdot)| \right|_{\mathrm{Var}} = \int\limits_S |p_n^{(i)}(x,z) - p_n^{(i)}(y,z)| d\nu(z) \leq$$

$$\leq \int\limits_S (|p_n^{(i)}(x,z) - m_n| + |p_n^{(i)}(y,z) - m_n|) d\nu(z) = 2 - 2m_n,$$

that results in the following bound on $\alpha(P_n^{(i)})$:

$$\alpha(P_n^{(i)}) \geq m_n \geq \frac{2s}{\sigma_n \sqrt{2\pi\delta_n}} e^{-\frac{s^2}{2\sigma_n^2 \sigma_n}}. \tag{31}$$

Thus under assumption (30) estimate (31) implies that for any i

$$\sum_{n=1}^{\infty} \alpha_n^{(i)} = \infty,$$

which guarantees the ergodicity of the Markov chain (28).

Remark. Obviously, condition (30) should be met, if

$$\sigma_n \to 0, \quad \delta_n \to 0$$

in such a way that

$$\sigma_n \sqrt{\delta_n} \geq \frac{c}{\sqrt{\ln n}}, \quad \text{with} \quad c \geq \frac{s}{\sqrt{2}}. \tag{32}$$

Thus, the Markov chain (28) is ergodic if a cooling schedule and an approximation step decay are chosen by estimate (32).

7 Results of computer simulations

In this section, we present results of computer simulations for image denoising performed by X. Descombes. Let us describe algorithms being used for the simulations. We consider two discrete approximations of the continuous process, namely the Euler and the Taylor schemes. Each approximation leads to two different iterative algorithms based on two optimization criteria: simulated annealing or the expectation criterion. In the annealing algorithms

the required configuration is constructed as a result of the corresponding approximation schemes (Euler or Taylor) where the parameters of the model are decreased during iterations: $\sigma_n,\ \delta_n \to 0$. The expectation algorithms on the first stage of iterations run in the same way as the corresponding annealing algorithms (Euler or Taylor). The parameters σ and δ are fixed, attaining some small values, and the required configuration is constructed as the average of the subsequent iterated configurations under given values σ and δ.

Computer experiments have been conducted both on synthetic and real images. The results of simulations are presented in figures 1-7. Figures 2-4 and 6-7 demonstrate images provided by four proposed above denoising algorithms. Besides, we compare two approximation schemes (Euler and Taylor) with the Metropolis – Hastings algorithm widely used in the Markov Random Field approaches. Each figure contains an information about a number of iterations, a range of the parameters σ_n and δ_n in the annealing algorithms (no δ_n in the Metropolis – Hastings algorithm), fixed values of σ and δ in the expectation algorithms. Although theoretical results require a logarithmic decay of the parameters of the model during the annealing procedure but in practice, we use an exponential decreasing scheme for both $\sigma_n,\ \delta_n$: $\sigma_n = \alpha^n \sigma_0$ with a constant $\alpha < 1$ close to 1, and the same for δ_n. Furthermore, to estimate results not only visually we calculate the value of the energy H_f of the final configuration for each algorithm under consideration. The parameters of the energy function $H(X, \theta)$ here are the same for all schemes, namely, $\beta = 2,0,\ \lambda = 0,0001,\ d = 10$. All images have 256×256 pixels, and the interval $[0, 256]$ is the grey level space.

In figures 2-4, we compare two proposed criteria using two different approximations (Euler and Taylor). On can see that the result derived by the Taylor schemes has a better quality, but it requires more computational time. Comparing tests on the synthetic image using simulations by different stochastic schemes – the MH algorithm and the diffusion dynamics algorithms (Euler and Taylor), one can see that the convergence of the latter schemes is faster than for the MH scheme, especially when using a small number of iterations. Moreover, the final images obtained by the diffusion schemes (Euler and Taylor) and by the MH schemes have different properties. The result of the realization of the diffusion schemes appears to be globally similar to the synthetic image without noise, except that the image still has pixelwise noise. This phenomenon is less appreciable with the expectation algorithms. To remove the residual noise, we propose to apply the median

filtering on the final stage of the simulations, see figures 4,6,7. Using the MH scheme we do not get this pixelwise noise but numerous areas which are not as smooth as the optimum, see figures 3,4. In this case, the median filtering doesn't improve the quality of the result.

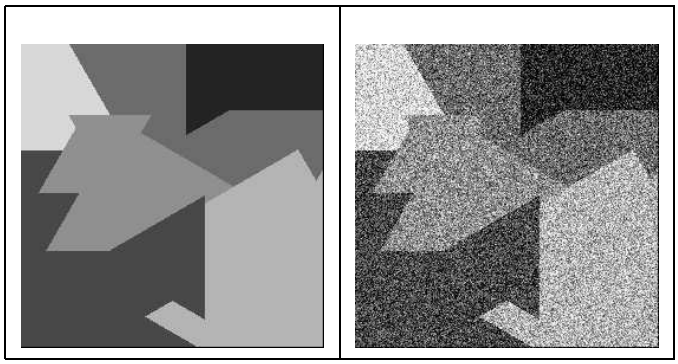


Table 1: test image (left); noisy image $\sigma = 50$ (right)

References

[1] *Geman S., Geman D.* Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images // IEEE Trans. Pattern Anal. Machine Intelligence. 1984. V. 6. ü 6. P. 721-741.

[2] *Geman S., Reynolds G.* Constrained Restoration and Recovery of Discontinuities // IEEE Trans. Pattern Anal. Machine Intelligence. 1992. V. 14. ü 3. P. 367-383.

[3] *Hajek B.* Cooling schedules for optimal annealing: Preprints Math. Op. Research, 1987.

[4] *Kirkpatrick S., Gelatt C.D., Vecchi M.P.* Optimization by Simulated Annealing // Science. 1983. V. 220. P. 671-680.

[5] *Holley R., Strook D.* Diffusions on an Infinite Dimensional Torus // Journal Funct. Anal. 1981. V. 42. P. 29-63.

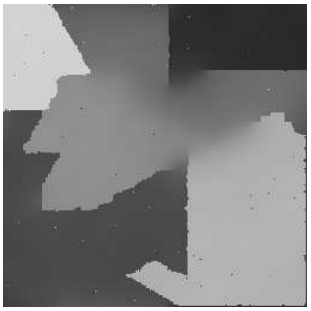
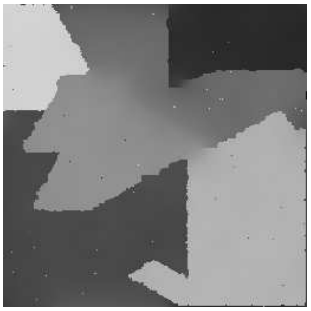
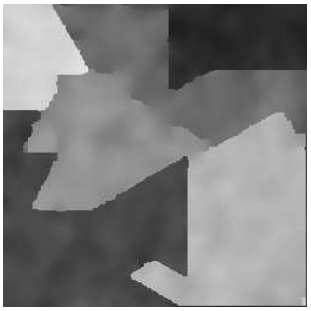
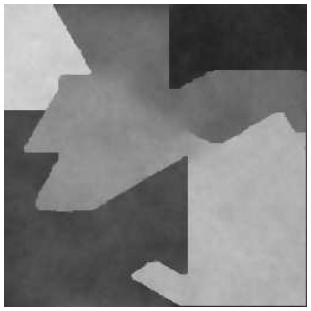
 The annealing algorithm: Euler $\sigma : 10 \rightarrow 0,001;\ \delta : 1000 \rightarrow 0,1$ $H_f = -245404$	 The expectation algorithm: Euler $\sigma = 1,0; \delta = 10$ $H_f = -244928$
 The annealing algorithm: Taylor $\sigma : 10 \rightarrow 0,001;\ \delta : 1000 \rightarrow 0,1$ $H_f = -245385$	 The expectation algorithm: Taylor $\sigma = 1,0; \delta = 10$ $H_f = -244954$
 The annealing algorithm: Metropolis $\sigma : 10 \rightarrow 0,001$ $H_f = -245018$	 The expectation algorithm: Metropolis $\sigma = 0,5$ $H_f = -241139$

Table 2: Results for the different algorithms and criteria using a high number of iterations (10000)

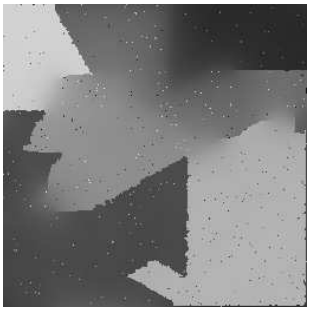
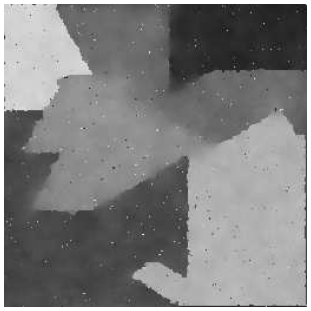
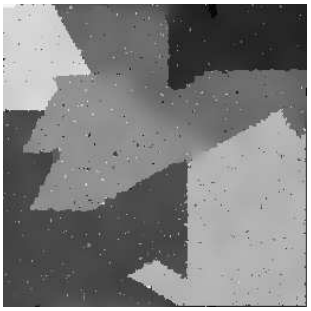
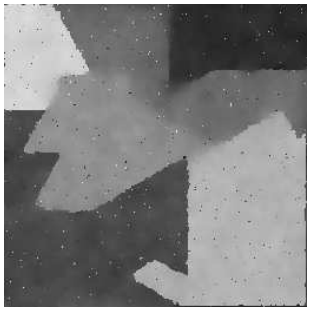
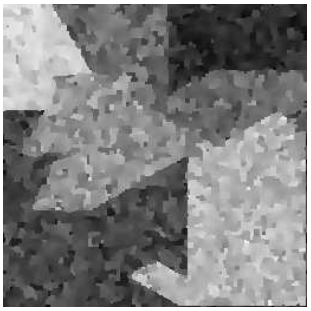
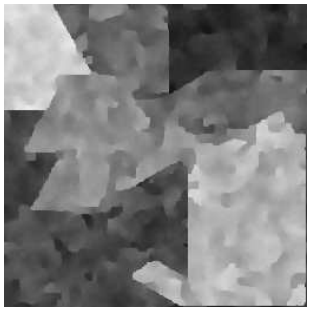
 The annealing algorithm: Euler $\sigma : 10 \rightarrow 0,01;\ \delta : 1000 \rightarrow 5$ $H_f = -242820$	 The expectation algorithm: Euler $\sigma = 1,0;\ \delta = 20$ $H_f = -237813$
 The annealing algorithm: Taylor $\sigma : 10 \rightarrow 0,01;\ \delta : 1000 \rightarrow 5$ $H_f = -241719$	 The expectation algorithm: Taylor $\sigma = 1,0;\ \delta = 20$ $H_f = -238169$
 The annealing algorithm: Metropolis $\sigma : 10 \rightarrow 0,01$ $H_f = -221939$	 The expectation algorithm: Metropolis $\sigma = 0,5$ $H_f = -229612$

Table 3: Results for the different algorithms and criteria using a small number of iterations (1000)

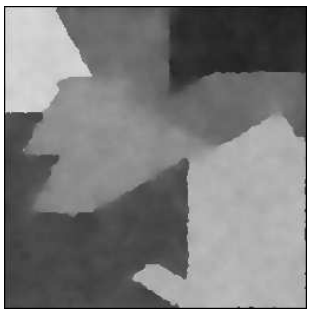
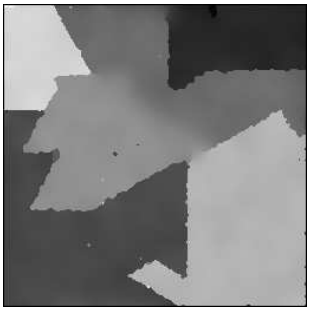
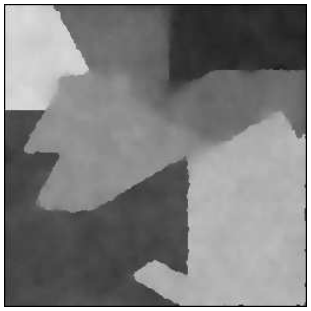
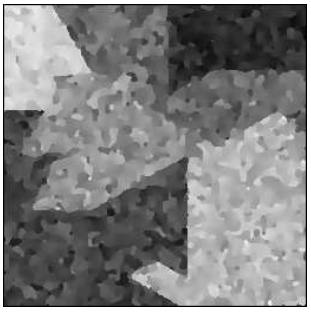
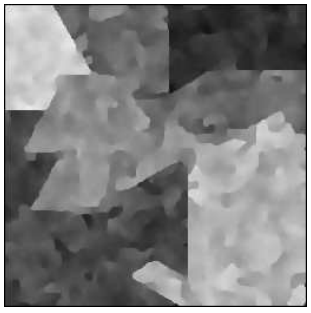
 The annealing algorithm: Euler $\sigma : 10 \rightarrow 0,01;\ \delta : 1000 \rightarrow 5$ $H_f = -243499$	 The expectation algorithm: Euler $\sigma = 1,0;\ \delta = 20$ $H_f = -241021$
 The annealing algorithm: Taylor $\sigma : 10 \rightarrow 0,01;\ \delta : 1000 \rightarrow 5$ $H_f = -243147$	 The expectation algorithm: Taylor $\sigma = 1,0;\ \delta = 20$ $H_f = -241233$
 The annealing algorithm: Metropolis $\sigma : 10 \rightarrow 0,01$ $H_f = -219963$	 The expectation algorithm: Metropolis $\sigma = 0,5$ $H_f = -230552$

Table 4: Results for the different algorithms and criteria using a small number of iterations (1000) followed by a 3×3 median filter

Table 5: Lenna picture (left); noisy Lenna picture $\sigma = 50$ (right)

[6] *Albeverio S., Kondratiev Yu. G., Roeckner M.* Uniqueness of the Stochastic Dynamics for Continuous Spin Systems on a Lattice // J. Funct. Anal. 1995. V. 133. P. 10-20.

[7] *Geman S., Hwang C.* Diffusions for Global Optimization // SIAM J. Control and Optimization. 1986. V. 24. ü 5. P. 1031-1043.

[8] *Chiang T.-S., Hwang C.-R., Sheu S.-J.* Diffusion for Global Optimization in R^n// SIAM J. Control and Optimization. 1987. V. 25. ü 3. P. 737-753.

[9] *Winkler G.* Image Analysis, Random Fields and Markov Chain Monte Carlo Methods, A Mathematical Introduction. NY.: Springer, 2003.

[10] *Freidlin M.I., Wentzell A.D.* Random perturbations of dynamical systems. NY: Springer-Verlag, 1984.

[11] *Wentzell A.D.* On the asymptotic behaviour of the largest eigenvalue of an elliptic second-order differential operator with a small parameter by the higher derivatives // Dokl. cad. Nauk USSR. 1972. V. 202. ü 1. P. 19-21.

Euler
$H_f = -94703$

plus a 3×3 median filter
$H_f = -92759$

Metropolis
$H_f = -88504$

plus a 3×3 median filter
$H_f = -90226$

Table 6: Results on image for 50 iterations with the annealing algorithms

Euler
$H_f = -98194$

plus a 3×3 median filter
$H_f = -93968$

Metropolis
$H_f = -92220$

plus a 3×3 median filter
$H_f = -90795$

Table 7: Results on image for 50 iterations with the expectation algorithms

[12] *Wentzell A.D.* Formulas for eigen-functions and eigen-measures associated with Markov process // Theory Prob. and Appl. 1973. V. 18. ü 1. P. 3-29.

[13] *Liggett Th.* Interacting particle systems. NY: Springer-Verlag, 1985.

[14] *Watanabe S., Ikeda N.* Stochastic differential equations and diffusion processes. NY.: NH Publ. Company, 1981.

[15] *Kloeden P., Platen E.* Numerical solution of stochastic differential equations. NY.: Springer-Verlag, 1992

[16] *Descombes X., Zhizhina E.* Image Denoising using Stochastic Differential Equations, Research Report 4814, INRIA, mai 2003.

[17] *Krim H., Zhizhina E.* Approximation approach in stochastic algorithms, in preparation.

[18] *Ethier S. N., Kurtz T. G.* Markov processes. Characterization and convergence. NY: , 1986.

[19] *Dobrushin R.L.* Central limit theorem for non-stationary Markov chains I. Theory Prob. Appl. 1, 65-80, 1956.