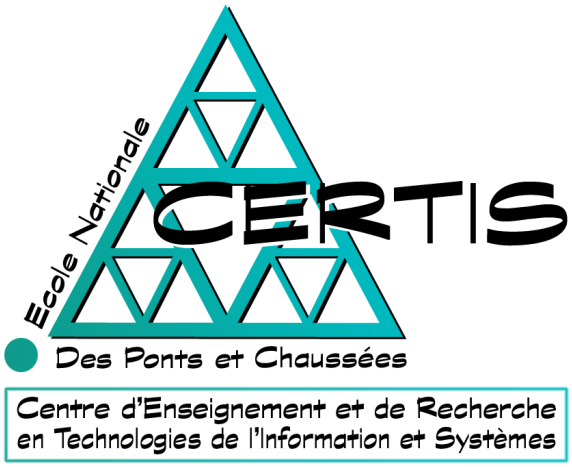
Stochastic Motion and the Level Set Method in Computer Vision: *Stochastic Active Contours*

Olivier Juan
Renaud Keriven
Gheorghe Postelnicu

Research Report 04-01
May 2004



CERTIS, ENPC,
77455 Marne la Vallee, France,
http://www.enpc.fr/certis/

Stochastic Motion and the Level Set Method in Computer Vision: *Stochastic Active Contours*

Mouvement Stochastique et la méthode des Level Sets pour la Vision par Ordinateur : *Contours Actifs Stochastiques*

Olivier Juan[1]
Renaud Keriven[1]
Gheorghe Postelnicu[1]

[1]CERTIS, ENPC, 77455 Marne la Vallee, France, http://www.enpc.fr/certis/

Abstract

Based on recent work on Stochastic Partial Differential Equations (SPDEs), this paper presents a simple and well-founded method to implement the stochastic evolution of a curve. First, we explain why great care should be taken when considering such an evolution in a Level Set framework. To guarantee the well-posedness of the evolution and to make it independent of the implicit representation of the initial curve, a *Stratonovich* differential has to be introduced. To implement this differential, a standard *Ito plus drift* approximation is proposed to turn an implicit scheme into an explicit one. Subsequently, we consider shape optimization techniques, which are a common framework to address various applications in Computer Vision, like segmentation, tracking, stereo vision etc. The objective of our approach is to improve these methods through the introduction of stochastic motion principles.

The extension we propose can deal with local minima and with complex cases where the gradient of the objective function with respect to the shape is impossible to derive exactly.

Finally, as an application, we focus on image segmentation methods, leading to what we call *Stochastic Active Contours*.

Résumé

Basé sur des travaux récents sur les équations aux dérivées partielles stochastique (SPDE), ce papier présente une approche simple et bien fondée pour l'implémentation d'évolutions stochastiques d'une courbe.

Dans un premier temps, nous expliquons pourquoi il faut faire attention en considérant de telles évolutions dans le cadre des Level Sets. Pour garantir le bien fondé de l'évolution et pour la rendre indépendante de la représentation implicite de la courbe initiale, le calcul de *Stratonovich* doit être utilisé. Pour implémenter cette différentielle stochastique, une approximation standard Ito plus drift est proposée pour rendre un schéma autrement implicite explicite.

Par la suite, nous considérons des techniques d'optimisation de formes, qui fournissent un cadre commun pour des problèmes variés de la vision par ordinateur, tels que la segmentation, le tracking, la stéréovision etc. Le but de notre approche est d'améliorer ces méthodes par l'introduction du modèle d'évolution aléatoire d'un contour.

L'extension que nous proposons peut gérer des minima locaux et des cas complexes où le gradient de la fonction objectif par rapport à la forme ne peut pas être dérivé. Finalement, en tant qu'application de ce modèle d'évolution, on s'intéresse à des méthodes de segmentation et l'on développe des *Contours Actifs Stochastiques*.

Contents

1 Introduction

1.1 Why adding noise?

Shape optimization techniques are a common framework to address various applications in Computer Vision, like segmentation, tracking, stereo vision etc. The objective of our approach is to improve these methods through the introduction of stochastic motion principles. These problems are most of the time stated as the minimization with respect to some hyper-surface Γ of $\mathbb{R}^N$ of some objective function $E(\Gamma)$. This is usually achieved using a gradient-descent method. Yet, in complex cases, E does not have any computable gradient with respect to Γ (see section (5.2)). In other cases, the minimization process gets stuck into some local minimum, while no multi-resolution approach can be invoked[2]. To deal with those two frequent problems, one can naturally turn to a stochastic optimization approach. Even a simple Simulated Annealing method might be powerful enough to escape from local minima and to cope with an approximation of the shape gradient. Indeed, adding noise to the motion of a curve is a prerequisite to developing this idea.

1.2 Context

We are interested in letting $\Gamma(t)$ evolve according to the equation

$$\frac{\partial \Gamma}{\partial t}(t,p) = \beta(t,p)\mathbf{n}(t,p) = \dot{\eta}(t,p)\mathbf{n}(t,p) \tag{1}$$

where p is some parameterization of Γ, $\mathbf{n}$ the normal to $\Gamma(t)$ at point $\Gamma(t,p)$ and where the normal velocity β depends on some stochastic perturbation $\dot{\eta}$ - here, the notation $\dot{\eta}$, standing for the "derivative" of the noise η w.r.t. time, will become clear further. The mean curvature motion $\beta = \kappa$ as well as many other problem oriented choices of β and their implementation with the Level Sets method [28] are well known. The novelty in our work is the implementation of the recently proposed stochastic flow (1) (see [37]) and its application to Computer Vision.

Stochastic dynamics of interfaces have been widely discussed in later years in the physics literature. The work in fields like front propagation or front transition is aimed at modeling and studying the properties of a moving frontier between two media that is subject to macroscopic constraints and random perturbations (which are due to the bulk). The natural translation of this dynamic in mathematical language is done through Stochastic Partial Differential Equations (SPDEs).

[2] E.g., when using some statistical region model, a change in resolution may result in smoothing out the difference between the statistics of the interior and the ones of the exterior.

These equations were introduced by Walsh in [37] and their mathematical properties were studied using mostly partial differential equations tools. Nevertheless, the problems researchers have to deal with are various and there is more than one way to add a stochastic perturbation to a PDE. An up to date survey of the existing models on stochastic motions by mean curvature can be found in [39]. It was only in recent years that the notion of viscosity solution for a SPDE was developed by Lions and Souganidis in a series of articles [19, 20, 22, 23]. Their notion of weak viscosity solution is very attractive for numerical applications, since they define the solution as a limit in a convenient space for a set of approximations. Since their pioneering work, related work has been done by Yip [38] and by Katsoulakis et al [16]. Another independent approach concerned with viscosity solutions of stochastic partial differential equations is due to Buckdahn and Ma [4]. Their approach is not well suited for Level Sets evolutions, though, since they do not allow certain functional dependencies that are common to all Level Sets evolutions.

1.3 What should not be done!

Mistake #1: Considering equation (1), one should be tempted to make the perturbation $\dot{\eta}$ depend on the parameter p, or, to make it intrinsic[3], on σ, the arc length parameter of the curve. Such an evolution, namely

$$\frac{\partial \Gamma}{\partial t}(t,p) = \dot{\eta}(t,\sigma(t,p))\mathbf{n}(t,p)$$

is actually unstable, even if η has some regularity with respect to σ and even if σ is normalized with respect to the total length of the curve. Suppose the curve develops a kink at some point. Then, its length will increase at that point and that will lead to adding more noise around the point where the kink formed itself. Consequently, that might lead to an unstable character of the evolution. That is why we will consider the stochastic perturbation as a function of space and the corresponding evolution:

$$\frac{\partial \Gamma}{\partial t}(t,p) = \dot{\eta}(t,\Gamma(t,p))\mathbf{n}(t,p)$$

Our noise will be regular in space and white in time. Indeed, there is no reason one would expect the random perturbations be correlated from one time step to the other, but considering noise that is white in space may have disastrous effects w.r.t. the regularity of the contours.

Mistake #2: A first simple choice for η is to suppose it constant in space and to consider that each of the increments $\dot{\eta}(t)$ is an independent Gaussian random

[3] That is to say, depending on the curve itself but not on the choice of the parameterization p.

variable. Therefore, $\eta(t)$ is a Brownian motion W:

$$\frac{\partial \Gamma}{\partial t} = \dot{W}(t)\mathbf{n}(t,p) \tag{2}$$

An error would be to believe that it should be implemented with the explicit scheme:

$$\Gamma(t+\Delta t,p) = \Gamma(t,p) + \Delta t\, \mathcal{N}_{(0,1)}(t)\, \mathbf{n}(t,p) \tag{3}$$

where $\mathcal{N}_{(0,1)}(t)$ denotes a standard Gaussian random variable. This would be incorrect, since the statistical properties of the curve would then depend upon the discretization of the time grid. To see this, consider independent variables $x_i \sim \mathcal{N}_{(0,1)}$ and notice that the previous evolution at time T would amount to $\sum_{i=1}^{n} x_i \Delta t = \sum_{i=1}^{n} \frac{T}{n} x_i$ where $\Delta t = T/n$ is the discretization step. Given the independence of the x_i, the previous sum is a Gaussian variable $\mathcal{N}_{(0,\frac{T}{n})}$, thus depending upon the discretization of the time interval $[0,T]$. We will see that the correct scheme involves $\sqrt{\Delta t}$ instead of Δt (see equation (13)):

$$\Gamma(t+\Delta t,p) = \Gamma(t,p) + \sqrt{\Delta t}\, \mathcal{N}_{(0,1)}(t)\, \mathbf{n}(t,p) \tag{4}$$

The previous argument is actually a trivial example that shows the difference between stochastic integration and Stieltjes integration. This is to say that our modelization will have to obey rules of Stochastic Calculus relying on intrinsic properties of Brownian motion. But we must also rely on the theory of viscosity solutions, since it is a necessary ingredient in the Level Set framework - allowing for rarefactions and shocks in evolutions and providing stable numerical schemes. The theory we need must then fill the gap between Stochastic Calculus and viscosity solutions. This is where the recent theory developed by Lions and Souganidis comes into place.

Mistake #3: Let us now try to implement (2) in a Level Set framework, where $\Gamma(t)$ is the zero level set of some implicit function $u(t,x)$ driven by the evolution $\frac{\partial u}{\partial t} = |Du|\dot{W}(t)$. Actually, a more correct way to write this SPDE is:

$$\mathrm{d}u = |Du|\mathrm{d}W(t) \tag{5}$$

As pointed out in [21], this equation is not reasonable and suffers from:

- *Non invariance:* Let $\alpha(.)$ be some smooth increasing function with $\alpha(0) = 0$. If $u(t,x)$ is solution of (5) given some initial condition $u_0(x)$, then $\alpha(u(t,x))$ is not solution of (5) with initial condition $\alpha(u_0)$. Moreover, the solution of (5) with initial condition $\alpha(u_0)$ has not the same zero level set than $u(t,x)$: the evolution depends on the implicit representation of the initial curve (see section 2.2)!

- *Ill-posedness:* Let us take $N = 1$ for the sake of simplicity. The equation $\mathrm{d}u = u_x \mathrm{d}W + \lambda u_{xx}\mathrm{d}t$, obtained by adding the curvature term λu_{xx} to (5), reveals to be an inverse heat equation for $0 < \lambda < 1/2$

It turns out that the differential used in (5) is the *Ito differential*, and that those two difficulties are overcome introducing the *Stratonovich differential*:

$$\mathrm{d}u = |Du| \circ \mathrm{d}W(t) \tag{6}$$

In the sequel, we briefly introduce the essential notions needed to understand the difference between the Ito and the Stratonovich cases. Citing [19], we will see that the notion of viscosity solution can be extended to the SPDE case. We propose a effective implementation of (6), that we extend to the case when the noise term depends on the space variable as well; then we investigate some geometrical properties of the evolution that could guide the user toward correct noise parameters. Subsequently, we explain how the stochastic motion can improve the shape optimization based methods in Computer Vision. Finally, as an application, we focus on image segmentation, leading to what we call *Stochastic Active Contours*.

2 Mathematics

2.1 Some Notions of Stochastic Calculus

This subsection is meant to offer the reader an intuition of the notion of Stochastic Calculus and of the supplementary challenges it poses. Focusing on the definition of the integral itself, we suppose the reader is familiar with the Brownian motion. We shall equally use the idea of a standard probability space, martingale, quadratic variation. Rigorous and complete introductions of Stochastic Calculus can be found in [14], [12] or [18]. Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a standard probability space. We will consider that $W = (W_1(t), \ldots, W_m(t))_{t \geq 0}$ is a standard m-dimensional Brownian motion issued from 0. We are interested in finding an appropriate way of introducing the notion of stochastic differential with respect to the process W. To better understand the difficulty here, it is worth while mentioning that the paths of the Brownian motion are only $\frac{1}{2}$-Holder continuous, so they are nowhere differentiable. Hence, in order to give a meaning to what the term $\mathrm{d}W(t)$ might mean, one can first define an integral with respect to W - the stochastic integral. Once this integral is defined, the differential is obtained using the integral defined. To keep the presentation as clear as possible, we suppose that $m = 1$ and all our processes are real-valued. The considered approach for the construction of such an integral is to define it as an isometry in the appropriate functional space. Indeed, consider a square integrable process $\Phi = (\Phi(t, \omega))_{t \geq 0}$.

Trying to define the stochastic integral $\int_0^T \Phi(t,\omega)\mathrm{d}W(t)$, one would start with Riemann approximations

$$I_\Delta(\Phi)(T) = \sum_{i=1}^{n-1} \Phi(t_i,\omega)\left(W(t_{i+1}) - W(t_i)\right) \tag{7}$$

with $\Delta = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ and hope to find a suitable space where the above sum would converge when $|\Delta| \to 0$. By proving some completeness results, one can show that every square integrable continuous process is integrable w.r.t. the Brownian motion and one can obtain the **Ito stochastic integral** as the unique, square integrable martingale $I(\Phi) = (I(\Phi)(t))_{t\geq 0}$ which is the limit of $(I_\Delta(\Phi)(t))_{t\geq 0}$ when $|\Delta| \to 0$. The convergence happens in the pseudo-metric $\|X\| = \sum_{n=1}^{\infty} \frac{\min(1,\sqrt{\mathbb{E}(X_n^2)})}{2^n}$.

The price to pay for the convergence of the Riemann sums is that it happens in a process space, hence the limit, which is denoted as the stochastic integral $\int_0^T \Phi(s)\mathrm{d}W(s)$, does not hold a meaning path-wise, but only as a process. This means that the set of events $\omega \in \Omega$ where the Riemann sums mentioned above do not converge to the above introduced integral is of measure 0. In the sequel, the convergence suggested above can be extended to an arbitrary dimension. Moreover, the extension can be taken with respect to local continuous martingales. For details, the reader is referred to one of the references above. Once this integral is defined, it is imperative one is given some chain rule formula. This is where the Ito lemma comes in. Consider a process $X = (X_t)_{t\geq 0}$ and $\alpha : \mathbb{R} \to \mathbb{R}$ a function of class $\mathcal{C}^2$. Then the Ito formula states that $Y = (\alpha(X(t)))_{t\geq 0}$ verifies the dynamics

$$\mathrm{d}Y(t) = \alpha'(X(t))\mathrm{d}t + \frac{1}{2}\alpha''\mathrm{d}\langle X, X\rangle(t) \tag{8}$$

The main difference when compared to the regular chain rule is the appearance of an extra term, also called Ito term, **or drift**, which involves the second derivative of α and the quadratic form $\langle X, X\rangle$, also known as the quadratic variation of the process X. The quadratic variation is not zero when X has some dependence upon a stochastic process and can be computed in the following manner. Suppose that $X_1(t) = \int_0^t f_1(s)\mathrm{d}W(s) + A_1(t)$, where f_1 is some continuous square integrable function and $A_1 : \mathbb{R} \to \mathbb{R}$ is continuous and increasing. Suppose similarly that $X_2(t) = \int_0^t f_2(s)\mathrm{d}W(s) + A_2(t)$. Then

$$\langle X_1, X_2\rangle(t) = \int_0^t f_1(s)f_2(s)\mathrm{d}s$$

Note that $\langle X, X\rangle$, depends solely on the stochastic part of the dynamics and is independent of the bounded variation part. When $f \equiv 0$, then the classical chain rule is obtained in (8).

Now, using the Ito formula, a variation of the stochastic integral introduced above can be obtained so that the classical chain rule is satisfied. Consider a process $X(t) = x + M(t) + B(t)$ where M is a local continuous martingale[4] and B an increasing process. Consider equally another continuous process $Y(t) = y + N(t) + C(t)$ where N is a local continuous martingale and C an increasing process. The **Stratonovich integral** of Y with respect to X is then given by the formula

$$\int_0^t Y(s) \circ \mathrm{d}\, X(s) = \int_0^t Y(s)\mathrm{d}X(s) + \frac{1}{2}\langle M, N\rangle(t) \tag{9}$$

Suppose now α is of class $\mathcal{C}^3$ and apply the Ito formula (8) to $\alpha'(X)$. Then

$$\mathrm{d}\alpha'(X(t)) = \alpha''(X(t))\mathrm{d}W(t) + \frac{1}{2}\alpha^{(3)}(X(t))\mathrm{d}t$$

and consequently $\mathrm{d}\langle\alpha'(X), X\rangle(t) = \alpha''(X(t))\mathrm{d}t$. Hence,

$$\alpha(X_t) = \alpha(X(0)) + \int_0^t \alpha'(X(s)) \circ \mathrm{d}\, X(s) \tag{10}$$

by noticing that the quadratic variation term that we obtained is equal to the Ito term in equation (8).

We conclude this section by the approximation of the stochastic integral mentioned above. For the Ito integral, we saw that, if $\Delta = \{0 = t_0 < t_1 < \cdots < t_n = T\}$, then

$$\lim_{|\Delta|\to 0} \sum_{i=0}^{n-1} Y(t_i)(X(t_{i+1}) - X(t_i)) = \int_0^T Y(s)\mathrm{d}X(s)$$

For the Stratonovich case, it can be proved that we have the two equally useful limits:

$$\lim_{|\Delta|\to 0} \sum_{i=0}^{n-1} \frac{Y(t_i) + Y(t_{i+1})}{2}(X(t_{i+1}) - X(t_i)) = \int_0^T Y(s) \circ \mathrm{d}\, X(s) \tag{11}$$

$$\lim_{|\Delta|\to 0} \sum_{i=0}^{n-1} Y\left(\frac{t_i + t_{i+1}}{2}\right)(X(t_{i+1}) - X(t_i)) = \int_0^T Y(s) \circ \mathrm{d}\, X(s) \tag{12}$$

As an example that is meant to emphasize the difference between the two integrals previously introduced, consider as before a Brownian motion W and $\lambda \in [0, 1]$. Let $\Delta = \{a = t_0 < t_1 < \cdots < t_n = b\}$, $\Delta W_i = W(t_{i+1}) - W(t_i)$ and consider

[4] typically, this would be some Ito stochastic integral $M(t) = \int_0^t \Phi \mathrm{d}W$

$\phi^\lambda(t_i) = (1-\lambda)W(t_i) + \lambda W(t_{i+1})$. It can be proved that

$$\int_a^b W(s) \circ \mathrm{d}\, W(s) = \lim_{|\Delta|\to 0} \phi^{\frac{1}{2}}(t_i)\Delta W_i = \frac{1}{2}\left[W^2(b) - W^2(a)\right]$$
$$\int_a^b W(s)\, \mathrm{d}\, W(s) = \lim_{|\Delta|\to 0} \phi^{0}(t_i)\Delta W_i = \frac{1}{2}\left[W^2(b) - W^2(a)\right] - \frac{1}{2}(b-a)$$

Note also that one should be careful when simulating ΔW_i (see section 1.3):

$$\Delta W_i \sim \mathcal{N}_{(0,t_{i+1}-t_i)} \sim \sqrt{t_{i+1}-t_i}\mathcal{N}_{(0,1)} \tag{13}$$

2.2 Proposed Model for the Stochastic Curve Evolution

In applications, the stochastic term will add to a deterministic force $F(D^2u, Du, x, t)$ (one of the simplest examples, analyzed in detail in [40], is concerned with the coupled evolution $\mathrm{d}\, u = \kappa|Du|\,\mathrm{d}\, t + \dot{W}|Du|\,\mathrm{d}\, t$). Hence, a naive way to write down the coupled evolution is

$$\mathrm{d}u = F|Du|\mathrm{d}t + \dot{W}|Du|\mathrm{d}t$$

The above equation will have a meaning if written as

$$\mathrm{d}u = F|Du|\mathrm{d}t + |Du|\mathrm{d}W(t) \tag{14}$$

Concentrating on the stochastic part again, we remark that we made an implicit choice by considering the Ito integral in the above formula, but we could have decided to go for the Stratonovich integral. So what is the difference between the two integrals from a Level Sets point of view? Let us consider the following invariance property that is required when working within a Level Sets framework: consider just a random evolution of the type

$$\mathrm{d}u = |Du|\mathrm{d}W(t) \text{ with } u(0,\cdot) = u_0 \tag{15}$$

where W is a one-dimensional Brownian motion. Then this evolution codes for the corresponding contour evolution

$$\frac{\partial \Gamma}{\partial t} = \dot{W}\mathbf{n} \text{ with } \Gamma(0) = \Gamma_0 \tag{16}$$

where $\dot{W}$ is Gaussian white noise and Γ_0 is the zero-level of u_0. The idea behind the Level Sets evolution framework is to have all the level sets of the implicit function given by (15) evolve according to the same dynamics (16). A smooth change of scale of a function satisfying (15) that leaves the zero-level unchanged

should not influence the dynamics of the level sets contour - since the corresponding contour evolution (16) is not affected by this change of scale. Consider then a function $\alpha : \mathbb{R} \to \mathbb{R}$ such that $\alpha' > 0$ and $\alpha(0) = 0$ and the initial value problem

$$\mathrm{d}u = |Du|\mathrm{d}W(t) \quad \text{with} \quad u(0,\cdot) = u_0(\cdot) \tag{17}$$

If we consider u the solution of (17), then $v = \alpha(u)$ should verify the same dynamics, but with a different initial condition

$$\mathrm{d}v = |Dv|\mathrm{d}W(t) \quad \text{with} \quad v(0,\cdot) = \alpha(u_0(\cdot))$$

as is the case in the deterministic framework. Nevertheless, one can apply the Ito rule to the dynamics (15) and see that

$$\mathrm{d}v = \mathrm{d}\alpha(u) = \alpha'(u)\mathrm{d}u = |Dv|\mathrm{d}W(t) + \frac{1}{2}\alpha''(u)|Du|^2\mathrm{d}t$$

and the assertion is not verified due to the additional Ito term. Hence, the problem (15) is ill-posed from a Level Sets point of view: **for a given initial curve $\Gamma(0)$, the choice of the initial implicit function u_0 modifies the solution of the equation!**. However, as observed by Lions and Souganidis, this invariance condition is verified if one replaces the Ito integral with the Stratonovich integral, since the latter does not include any additional term anymore. Hence, the *right way* to insert stochastic evolutions in the Level Sets framework is through the Stratonovich integral. We rewrite (15) accordingly

$$\mathrm{d}u = |Du| \circ \mathrm{d}W(t) \quad \text{with} \quad u(0,\cdot) = u_0(\cdot) \tag{18}$$

Then, if we consider $v = \alpha(u)$ (hence the corresponding initial condition will be $\alpha(u_0)$) the dynamics verified by v are

$$\mathrm{d}\, v = \alpha'(u) \circ \mathrm{d}\, u = \alpha'(u)|Du| \circ \mathrm{d}\, W(t) = |Dv| \circ \mathrm{d}\, W(t)$$

and the invariance property is verified this time. Now, given the previous ingredient, the proposed random curve evolution model is given by

$$\mathrm{d}u = F\mathrm{d}t + |Du| \circ \mathrm{d}W(t) \tag{19}$$

Here, we used the Stratonovich integral, as opposed to (14).

A second example that suggests that the Stratonovich integration should be used when working with stochastic partial differential equations (rather than the Ito integral) is concerned with the 1-dimensional perturbed heat equation

$$\mathrm{d}\, u = u_x\, \mathrm{d}\, W(t) + \lambda u_{xx}\, \mathrm{d}\, t \tag{20}$$

where $\lambda \geq 0$. Using the Ito formula again, if we let $v(t,x) = u(t, x - W(t))$, then v will verify

$$\begin{aligned} \mathrm{d}\, v &= \mathrm{d}\, u - u_x \circ \mathrm{d}\, W(t) = u_x\, \mathrm{d}\, W(t) + \lambda u_{xx}\, \mathrm{d}\, t - u_x\, \mathrm{d}\, W(t) - \frac{1}{2} u_{xx}\, \mathrm{d}\, t \\ &= \left(\lambda - \frac{1}{2}\right) v_{xx}\, \mathrm{d}\, t \end{aligned} \tag{21}$$

and this reduces to an inverse heat equation for $\lambda < \frac{1}{2}$, which is ill-posed from a stability point of view. If, instead of considering the Ito integral in equation (20), we consider the Stratonovich integral,

$$\mathrm{d}\, u = u_x \circ \mathrm{d}\, W(t) + \lambda u_{xx}\, \mathrm{d}\, t \tag{22}$$

then again, by letting $v(t,x) = u(t, x - W(t))$, v will verify

$$\begin{aligned} \mathrm{d}\, v &= \mathrm{d}\, u - u_x \circ \mathrm{d}\, W(t) = u_x \circ \mathrm{d}\, W(t) + \lambda u_{xx}\, \mathrm{d}\, t - u_x \circ \mathrm{d}\, W(t) \\ &= \lambda v_{xx}\, \mathrm{d}\, t \end{aligned}$$

which is a stable equation for all $\lambda \geq 0$.

What is the difference between the evolution (19) and a classical Level Sets evolution such as $\mathrm{d}u = F\mathrm{d}t$? Since the stochastic term only depends upon $|Du|$ and the time parameter, all the points of the contour will have an extra random force which will be the same on the entire contour at each time step. This type of perturbation is indeed very important from a theoretical point of view, but we would like something more flexible in our applications. Typically, we would be interested in having white noise in both the time and spatial parameters. Nevertheless, white noise in space appears to add a lot of technical difficulties to the problem and the return on investment is quite small, since most of our models will evolve on discrete grid spaces. That is why we have opted for colored spatial noise, that is typically given by

$$W(t,x) = \sum_{i=1}^{m} \phi_i(x) W_i(t)$$

where $\phi_i : \mathbb{R}^N \to \mathbb{R}$ are smooth functions with compact support. Note that other choices of colored spatial noise are possible. The final evolution model we propose is thus

$$\mathrm{d}u = F|Du|\mathrm{d}t + |Du| \sum_{i=1}^{m} \phi_i(x) \circ \mathrm{d}W_i(t) \tag{23}$$

As a simplification, in practice we choose the functions with the same profile, but centered around a number of points x_i, that we call *noise sources*. Thus, our typical choice is

$$\phi_i(x) = \phi(x - x_i)$$

where ϕ is some convenient regular function.

2.3 Stochastic Viscosity Solutions

The theory developed earlier needs some sort of convergence results. As mentioned before, the proper type of solutions need to be used, so that the previous results from the Level Sets theory apply here. The notion of stochastic viscosity solution for fully nonlinear, second-order, possibly degenerate, stochastic partial differential equations such as the ones considered previously is put forward in a series of articles: [19], [20], [22] and [23]. Their theory is meant to apply precisely to equations such as (23), with $F = F(D^2u, Du, x, t)$. So far, a limit of their theory, which stands even today as an open question, is that they do not treat equations where the noise depends upon the space parameter (they only treat the case $\phi_i \equiv 1$, with our previous notation). However, experimental data suggests that their theory applies in cases like ours as well (see section 3). Precisely, consider the equations

$$\mathrm{d}u = F(D^2u, Du, x, t)\mathrm{d}t + \epsilon|Du| \circ \mathrm{d}W(t) \quad \text{with} \quad u(\cdot, 0) = u_0(\cdot) \tag{24}$$

$$\mathrm{d}u = F(D^2u, Du, x, t)\mathrm{d}t + |Du|\dot{\xi}_\alpha(t) \quad \text{with} \quad u(\cdot, 0) = u_\alpha(\cdot) \tag{25}$$

where $\epsilon \geq 0$ and ξ_α is a family of smooth functions $\xi_\alpha : \mathbb{R}_+ \to \mathbb{R}$. Then we can cite the following theorem, summarizing their results:

Theorem 1 *The following hold a.s. in ω:*

1. *There exists a unique solution to* (24).
2. *Let $\{\xi_\alpha(t)\}_{t\geq 0}$ and $\{\eta_\beta(t)\}_{\beta>0}$ be two families of smooth functions such that as α and $\beta \to 0$, ξ_α and η_β converge to W uniformly on any compact in t and a.s. in ω. Let $\{u_\alpha\}_{\alpha>0}$ and $\{v_\beta\}_{\beta>0}$ in BUC$(\mathbb{R}_+ \times \mathbb{R}^N)$[5] be the solutions of* (25). *If $\lim_{\alpha,\beta\to 0} \|u_\alpha(\cdot, 0) - v_\beta(\cdot, 0)\|_{\mathcal{C}(\mathbb{R}^N)} = 0$, then, for all $T > 0$, $\lim_{\alpha,\beta\to 0} \|u_\alpha - v_\beta\|_{\mathcal{C}([0,T]\times\mathbb{R}^N)} = 0$. In particular, any smooth approximations of W produce solutions converging to the unique function stochastic viscosity solution of* (24).

[5] bounded uniformly continuous

3. *As $\epsilon \to 0$, the solution u^ϵ of* (24) *converges in $\mathcal{C}(\mathbb{R}_+ \times \mathbb{R}^N)$ to the solution of* (24) *with $\epsilon = 0$.*

Consequently, their result **allows us to simulate the solutions of such equations and be sure that the result of our computer simulation is what we expect it to be**. Further more, we mention that according to Lions, the convergence takes place in $C(\mathbb{R}_+ \times \mathbb{R}^N)$, which means that the numerical solutions we develop will be continuous and that they will be converge uniformly almost surely in $\omega \in \Omega$.

We end this theoretical part with an example by Souganidis on the explicit solution of the equation

$$\mathrm{d}u = |Du|\dot{\eta}\mathrm{d}t \quad \text{with} \quad u(0,x) = |x| \tag{26}$$

where $\eta : \mathbb{R}_+ \to \mathbb{R}$ is a function of class $\mathcal{C}^1$ such that $\eta(0) = 0$. The explicit solution of this equation is given by

$$u(t,x) = \max\left[(|x| + \eta(t))_+ \,,\, \max_{s\in[0,t]}(\eta(s))_+\right]$$

where $(x)_+ = \max(0,x)$; then, one can see that uniform convergence of $\eta \to W$ is sufficient to obtain the solution of the associated SPDE. Moreover, this simple case allows one to see that the random path η has a different effect on the solution that depends mainly on its sign. Indeed, as it can be observed from formula (26), there is a quantitative difference between the behavior of the solution depending on whether $\dot{\eta} > 0$ or $\dot{\eta} < 0$. This can be better understood watching a sample evolution in figure (1). Moreover, numerical artifacts will develop due to very frequent changes of sign of $\dot{\eta}$, since the use of η is only for heuristic purpose (the Brownian motion is nowhere differentiable). As a result, the regular reinitialization of the implicit function – a standard technique of the Level Set framework – is indispensable in the stochastic case.

2.4 Numerical scheme

The main problem when implementing Stratonovich evolutions is that they amount often to implementing implicit numerical schemes. Consider again the simple evolution $\mathrm{d}\,u = |Du| \circ \mathrm{d}\,W(t)$. According to the approximating scheme (11), the direct way of simulating such a process is through the following implicit scheme:

$$u_{i+1} = u_i + \frac{1}{2}\left(|Du_i| + |Du_{i+1}|\right)\Delta W_i$$

To avoid working with an implicit scheme, notice that the schema presented for the simulation of the Ito integral is an explicit one and use the fact that the Stratonovich integral is equal to the Ito integral plus an additional drift. Consider the evolution

$$\mathrm{d}u = H(Du,x) \circ \mathrm{d}W(t) \tag{27}$$

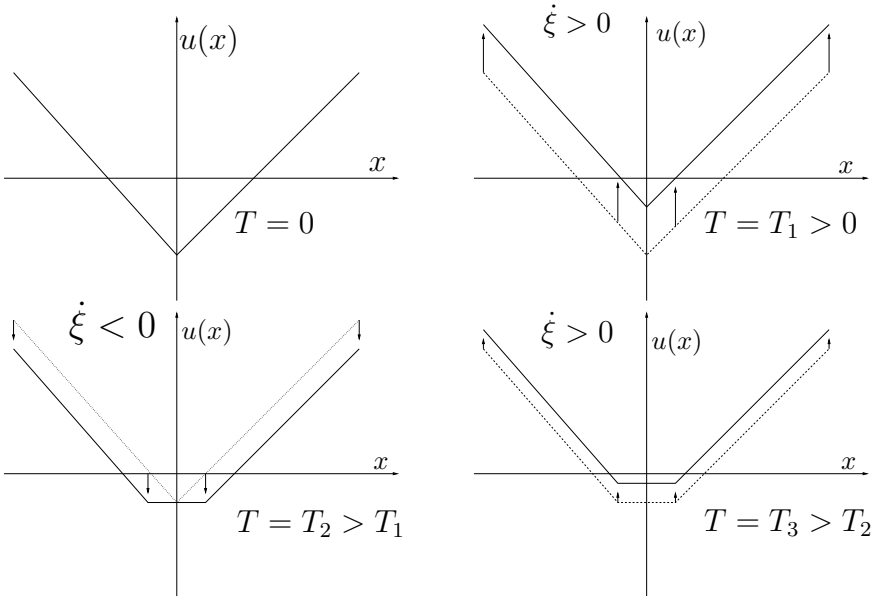


Figure 1: Examples of a typical evolution following the dynamics of equation (26). The extremely frequent changes of sign of the increments will produce singularities that might lead to numerical artifacts. Hence, from an implementation point of view, some regular *reinitialization* of u is advisable.

where we have compacted the notation used previously. Here $H(p, x)$ is a function from $\mathbb{R}^N \times \mathbb{R}$ with real values. The typical example is $H(p, x) = |p|\phi(x)$, where, ϕ is some convenient regular function which is smooth enough. Such an evolution is equivalent, according to the definition of the Stratonovich integral, with the Ito evolution given by

$$\mathrm{d}\, u = H(Du, x)\, \mathrm{d}\, W(t) + \frac{1}{2}\, \mathrm{d}\langle H(Du, x),\, W\rangle(t) \tag{28}$$

To compute the drift, we start by rewriting the above dynamics in an integral form

$$u(t, x) = u_0(x) + \int_0^t H(Du(s, x),\, x) \circ \mathrm{d}W(s)$$

We can then derive with respect to the spatial parameter x and obtain

$$Du(t, x) = Du_0(x) + \int_0^t \left[D^2u(s, x)D_pH(Du(s, x), x) + D_xH(Du(s, x), x)\right] \circ \mathrm{d}W(s)$$

where D_pH (resp. D_xH) denotes the gradients of H w.r.t. p (resp. x). Then, applying the Ito rule, we have

$$H(Du(t, x), x) = H(Du_0(x), x) + \int_0^t \left[D_pH \cdot (D^2uD_pH) + D_pH \cdot D_xH\right] \circ \mathrm{d}W(s)$$

Finally, if we consider the simplifying notation $A[u] = A(u,u)$, when A is some quadratic form, then the drift from equation (28) can be written as

$$\frac{1}{2}\langle H(Du,x),\, W\rangle(t) = \frac{1}{2}\int_0^t (D^2u(s,x)\,[D_pH(Du(s,x),x)] \\ + D_pH(Du(s,x),x)\cdot D_xH(Du(s,x),x))\mathrm{d}s$$

When $H = |p|\phi(x)$, the previous formula becomes

$$\langle H(Du,x),\, W\rangle(t) = \int_0^t \left[\phi^2(x)D^2u(s,x)\left[\frac{Du(s,x)}{|Du(s,x)|}\right] + \phi(x)\,D\phi(x)\cdot Du(s,x)\right]\mathrm{d}s$$

We can remark that the second order term in the above formula is a smoothing term. It can also be written

$$D^2u\left[\frac{Du}{|Du|}\right] = \Delta u - |Du|\,\mathrm{div}\left(\frac{Du}{|Du|}\right) = \Delta u - |Du|\kappa$$

where κ denotes the mean curvature of the level set at point x. One can be alarmed by the presence of $-|Du|\kappa$. Nevertheless, the overall term is positive, since D^2u is a semi-positive definite matrix.

The above calculation remains valid if the dynamics depends on more than one Brownian motion. In conclusion, to simulate an evolution of the type

$$\mathrm{d}u = F\mathrm{d}t + |Du|\sum_{i=1}^m \phi_i(x)\circ\mathrm{d}W_i(t) \tag{29}$$

we use

$$\begin{aligned}\mathrm{d}u =& F\mathrm{d}t + |Du|\sum_{i=1}^m \phi_i(x)\mathrm{d}W_i(t) \\ &+ \frac{1}{2}\left((\sum_{i=1}^m \phi_i^2(x))D^2u\left[\frac{Du}{|Du|}\right] + (\sum_{i=1}^m \phi_i(x)D\phi_i(x))\cdot Du\right)\mathrm{d}t\end{aligned} \tag{30}$$

or, in the general case when the stochastic Hamiltonian is given by $H(p,x)$:

$$\mathrm{d}u = F\mathrm{d}t + H(Du,x)\mathrm{d}W(t) + \frac{1}{2}\left(D^2u\,[D_pH] + D_pH\cdot D_xH\right)\mathrm{d}t$$

3 Validation

In this section, we test our scheme and investigate some simple geometrical properties the evolution that could guide the user toward a correct choice of noise.

3.1 One Gaussian noise

Let us begin with the simple case of a Gaussian noise constant in space. We thus consider $\mathrm{d}u = |Du| \circ \mathrm{d}W(t)$ and implement:

$$\mathrm{d}u = |Du|\mathrm{d}W(t) + \frac{1}{2}D^2u(t,x)\left[\frac{Du(t,x)}{|Du(t,x)|}\right]$$

We use a standard WENO3 scheme [13] in space with step Δx and a first order explicit scheme in time with step Δt and verify the convergence of the approximation when the space step and/or the time step tend to zero. Again, please remind the use of $\sqrt{\Delta t}$:

$$u(t+\Delta t,x) = u(t,x) + |Du(t,x)|\sqrt{\Delta t}\mathcal{N}_{(0,1)}(t) + \frac{1}{2}D^2u(t,x)\left[\frac{Du(t,x)}{|Du(t,x)|}\right]$$

Because of the stochastic character of the evolution, one can only compare the different approximations through some statistical quantity[6]. For a given initial condition and a given final time T, *the variance of the area of the interior of the final curve* provides a simple and meaningful way to compare two approximations. We approximate the area of the interior region by an approximating Heavyside integral over the image:

$$A_\epsilon = \int_D H_\epsilon(u(x))\,\mathrm{d}\,x$$

with H_ϵ an approximation of the Heaviside function, given by

$$H_\epsilon(x) = \begin{cases} 1 & \text{si } x > \epsilon \\ 0 & \text{si } x < -\epsilon \\ \frac{1}{2}\left[1 + \frac{x}{\epsilon} + \frac{1}{\pi}\sin\left(\frac{\pi x}{\epsilon}\right)\right] & \text{pour } |x| \leq \epsilon \end{cases}$$

We preferred this approximation of the Heavyside function for its local character, that gives a low error level for the implicit representations we consider. The left part of figure 2 shows, for different values of Δx, the convergence of this variance when $1/\Delta t$ increases. As a reminder to avoid a naive mistake, we also implemented the evolution with Δt instead of $\sqrt{\Delta t}$ and verify that the variance of the area tends to zero!

[6] Actually, for a given time step Δt, we might fix the event ω and compare the approximations for different Δx but with the same Brownian. We also successfully used such a path-wise comparison when testing the invariance property of our scheme.

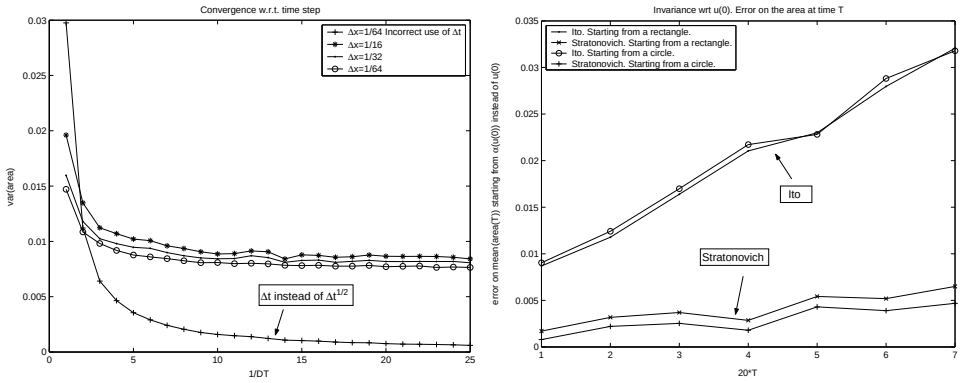


Figure 2: One Gaussian source. Left: convergence of the variance of the area at a given time T when Δt tends to 0 (plus the erroneous case when using (3)). Right: invariance of Stratonovich w.r.t. the choice of the initial implicit function.

As a test of the invariance of the Stratonovich differential, we compare, for a given initial curve $\Gamma(0)$, the mean of the area of the curve at a given final time T for different choices of the initial implicit function $u(0)$ (namely the signed distance function $d_{\Gamma(0)}$ to $\Gamma(0)$ and $\alpha(d_{\Gamma(0)})$ with $\alpha(x) = e^x - 1$). The right part of figure 2 shows, for different values of T and different initial curves, the relative difference between the means of the final area for the initial conditions $d_{\Gamma(0)}$ and for $\alpha(d_{\Gamma(0)})$ in both the Ito and the Stratonovich cases . Note how the Stratonovich scheme is much more invariant with respect to the choice of $u(0)$.

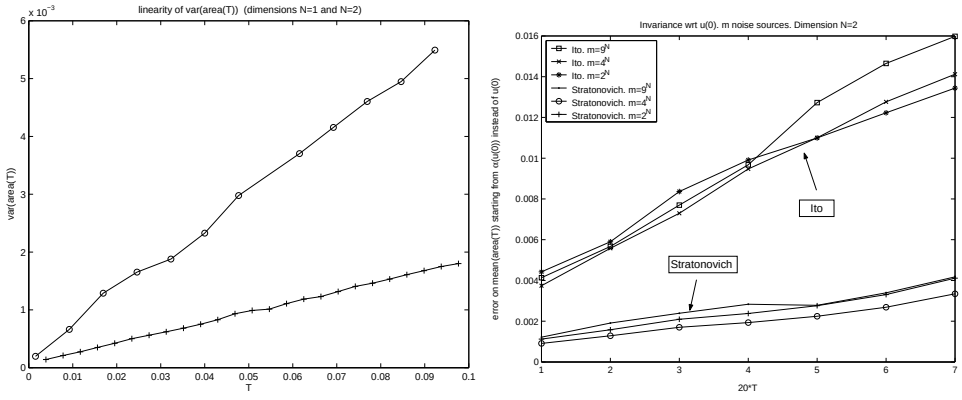


Figure 3: Left: Linear dependency between the final time and the variance of the area (one Gaussian). Right: invariance of Stratonovich wrt to the choice of the implicit function (several Gaussian sources).

Let us consider now the following simple example: the initial contour is a circle and we choose the signed distance to the contour to be our implicit evolution initialization. Then, the stochastic term that will influence the evolution will result in a random force (uniform on the entire contour) that modifies the radius of the circle. Of course, the drift will tend to come into play as well, but let us neglect it for a few moments. It is then easy to see the area of region of interest will be of the form

$$\pi \times (r + \eta)^2$$

where r is the initial radius and η is the stochastic perturbation. If we consider this evolution at a time t (with t small), then $\eta \sim \mathcal{N}(0, t)$. Hence, the area difference between times t and 0 is given by

$$2\pi r\eta + \pi\eta^2$$

One cand thus see the first term is a first order term that contains the length of the initial contour. It is then easy to see the law of the area difference (neglecting the influence of the drift term) is given by a gaussian random variable with mean 0 and variance

$$t\,(2\pi r)^2 + t^2\,\pi^2$$

which is a second-order polynomial in the time parameter. When adding back the drift term (or when considering more complicated evolutions) the mean will modify, but the first-order term should follow the same type of dynamics.

3.2 Several Gaussian noise sources

Having the whole curve shrink or grow at the same time is not very useful. We will use a spatially dependent noise although the viscosity solution result is still an open question in this case. For a given number m of random sources, we implement the evolution (29) with $F = 0$ using the scheme (30). The m sources are equally distributed on a grid $\{x_i\}$ and $\phi_i(x) = \phi(x - x_i)$ where ϕ is such that $\phi_i(x_j) = \delta_{ij}$ and ϕ_i decreases smoothly from x_i to its neighbors. In practice, although not derivable in x_i, the classical multi-linear interpolation functions are sufficient. Note also that $\sum_{i=1}^{m} \phi_i(x)\mathrm{d}W_i(t)$ is no more of variance 1 for all x, so that the stochastic motion would be weaker between two sources. Using $\phi_i(x)/\left(\sum_{j=1}^{m} \phi_j^2(x)\right)^{\frac{1}{2}}$ instead of $\phi_i(x)$ recovers a constant variance 1.

The drift will have a spatial derivative term (see (30)). Like figure 2 for one noise, the right part of figure 3 shows, for different values of m, how the Stratonovich scheme makes the evolution invariant with respect to the choice of $u(0)$.

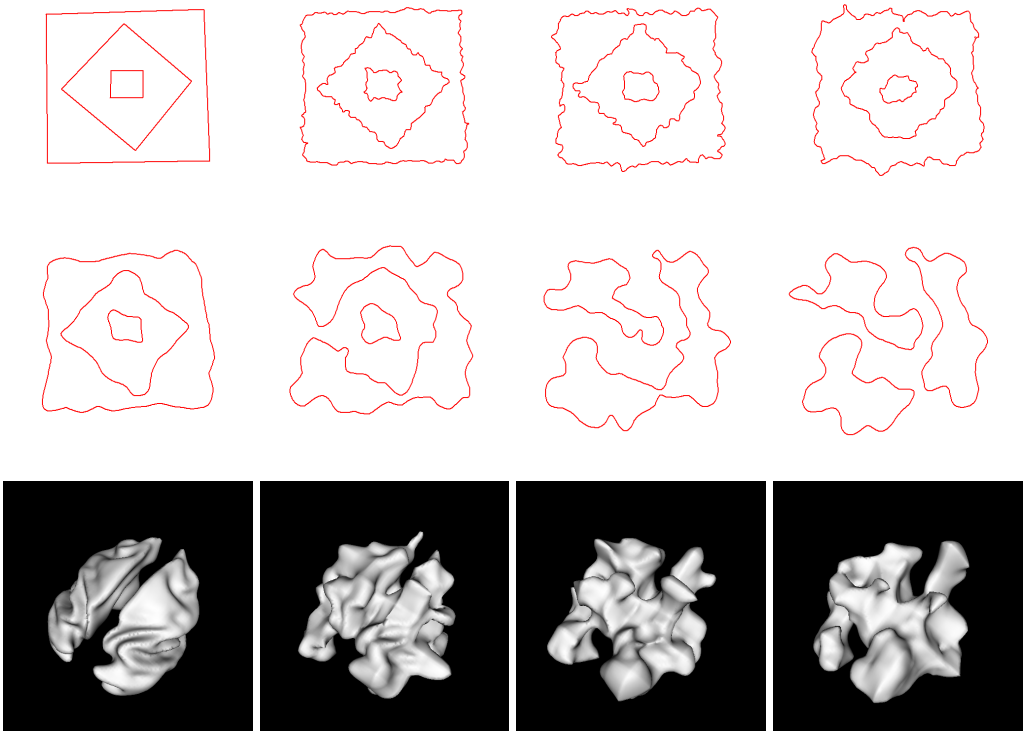


Figure 4: Different number of Gaussian noise sources. Top row: starting from the initial curve (top left), three time steps of the evolution with a large number of Gaussian sources. Middle row: from the same initial curve, four time steps of the evolution with a spatially smoother noise (small number of sources). Bottom row: a 3D example starting from the cortex of a monkey.

With more than one source of noise, the points of the curve do not move at the same speed anymore, leading to the desired stochastic global deformation. As one should expect, with a large number of sources, the deformation is very noisy but the contributions of the sources tend to annihilate one each other. Thus, the curve does not move very far from its initial position. On the contrary, with a medium number of sources, the deformation is smoother but with ampler motions (see figure 4). Depending on his/her own application, the user might want to choose the optimal number of sources. As a first attempt to quantify the phenomenon, we measure how long it takes to the curve to move away from its initial position. For a given distance δ, we call the *expected exit time* the quantity $T(\delta) = \mathbb{E}(\inf\{t : \exists x \in \Gamma(t), d(x, \Gamma(0)) \geq \delta\})$ where $\mathbb{E}$ denotes the expectation. For a Brownian motion, the expected exit time from a ball is a quadratic function of the radius of the ball. In our case, such a result would be certainly hard to prove. Yet, our experiments show a similar relationship $T(\delta) \approx \alpha(m)\delta^2$: (see the left part of figure 3.2). This useful relationship indicates clearly how long the user have to wait to see his/her curve getting away from its initial position. The right part of figure 3.2 plots α as a function of the number of sources. As expected, a large number of sources m induces a larger exit time, thus a larger α. Surprisingly, the smallest values of m give also a large α. We do not have any satisfactory explanation for this phenomenon... Anyway, these are only some very first step toward the understanding of the geometric properties of this kind of stochastic motion and many other quantities would be of great interest: the variations of the curvature, the time to get the curve split, etc.

We end this section with another heuristic for the variance of the area in the case when several noise sources come into play. For a given initial curve, the variance of the area of the curve at time T can be approximated the following way. When working with a process that follows $\mathrm{d}\, u = |Du| \circ \mathrm{d}\, W(t)$ with the initial condition u_0, if we suppose that $|Du_0| \equiv 1$, then one can consider the first order approximation $u(t, x) - u_0(x) = \sum_{i=1}^{m} \phi_i(x) W_i(t)$ Hence, a very simplistic approximation of the area difference for a short time delta is given by

$$\begin{aligned} A_\epsilon(t) - A_\epsilon(0) &= \int_D \left(H_\epsilon(u(t,x)) - H_\epsilon(u(0,x))\right) \mathrm{d}\, x \\ &\sim \int_D \delta_\epsilon(u(0,x)) \left[\sum_{i=1}^{m} \phi_i(x) W_i(t)\right] \mathrm{d}\, x \\ &= \sum_{i=1}^{m} W_i(t) \int_D \delta_\epsilon(u(0,x)) \phi_i(x) \, \mathrm{d}\, x \end{aligned}$$

Considering this first order approximation amounts to

$$Var(A_\epsilon(t) - A_\epsilon(0)) = t \left[\sum_{i=1}^{m} \left(\int_D \delta_\epsilon(u(0,x)) \phi_i(x) \, \mathrm{d}\, x \right)^2 \right]$$

One can see that when $m = 1$ and $\phi \equiv 1$, this formula becomes

$$Var(A_\epsilon(t) - A_\epsilon(0)) = t \left(\int_D \delta_\epsilon(u(0,x)) \, \mathrm{d}\, x \right)^2$$

and the last term on the right-hand side is exactly the length of the initial contour. We recover thus the result we noticed in the last section when working with one noise source.

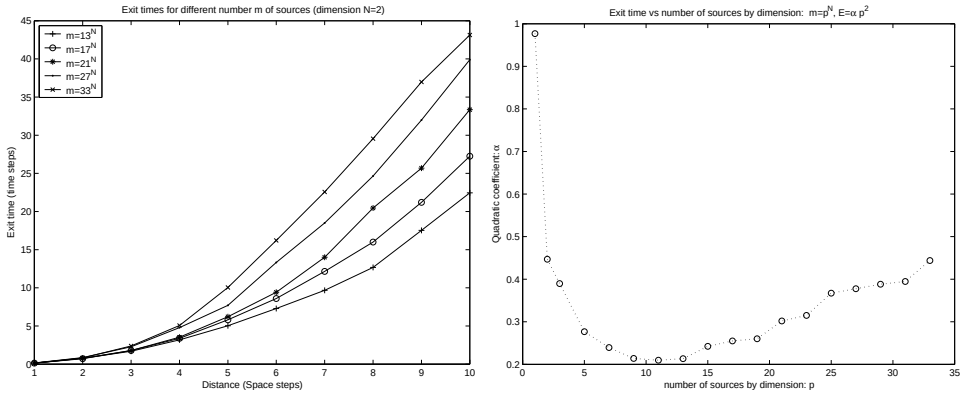


Figure 5: Several Gaussian sources. Left: Quadratic relation between the distance δ and the expected exit time from the band of thickness δ. Right: variation of the exit time w.r.t. the number of sources.

4 Applications to Computer Vision

Many Computer Vision problems consist in recovering a certain surface or region through a shape optimization framework [6, 11, 29]. These methods suffer from being sometimes stuck in local minima. The dynamics presented earlier, coupled with a decision mechanism, can be used to overcome this problem. As a first step toward more sophisticated genetic methods, we turned our attention to the Simulated Annealing algorithm. Based on the work of Metropolis et al. [27], Simulated Annealing was first mentioned by Kirkpatrick in [17] as a nice application of statistical physics to optimization problems. Its purpose is to introduce a

probabilistic decision mechanism for finding global minima in higher dimension. First, we would like to briefly comment upon the previous works oriented toward the use of genetic programming in Computer Vision.

4.1 Comparison with previous work Computer Vision

In a lot of cases, the stochastic theory is used to help researchers develop an intuition of the macroscopic dynamics at a microscopic level. This is the case in [2, 35], where an algorithm for stochastic approximations to a curve shortening flow is built. Another example is given by [34], where the authors develop a model of anisotropic diffusion using the information gained by analyzing the stochastic differential equation associated to a linearized version of the geometric heat equation. Note that one should not confuse these stochastic simulations of a deterministic motion with our use of stochastic motion. In other cases, stochastics are actively used in selection algorithms meant to overcome some classical dynamics difficulties. In [32] Storvik used Simulated Annealing combined with a Bayesian dynamics and developed applications in medical imagery. He used a node-oriented representation technique for the contour representation. Thus, his algorithm can only detect simply connected domains in an image. Moreover, self-intersections are not allowed, due to the complications they would involve. More recently, Ballerini et al developed in [1] an interesting application to medical image segmentation using a genetic algorithm, *genetic snakes*. They used a model that they fit using a number of control points. Their application cannot, therefore, be extended to a more general framework.

Please note that the main ingredient of our work is not the Simulated Annealing part, but rather the implementation of the stochastic motion and its use in shape optimization problems. It is obvious that the stochastic approach adds to the power and flexibility of the Level Sets technique into a very powerful tool. We can thus use this mechanism through skillfully applied controls, while continuing to allow for topological changes and weak regularity assumptions. Simulated Annealing is used in our experiments. In the future, more evolved genetic programming selection techniques might be considered, but it is encouraging that such simple ingredients added to the Level Sets framework provide good practical results. *Sketchily, one can see the same difference between our method and the previous ones, than between geodesic active contours and the pioneering snakes [15].*

4.2 Principle

Given some Computer Vision problem in a variational framework where we have to find the region Γ that minimizes an energy $E(\Gamma) = E(u)$, we use the following

simple Simulated Annealing decision scheme:

1. Start from some initial guess u_0
2. compute u_{n+1} from u_n using some dynamics, e.g. $\mathrm{d}u = |Du| \sum_{i=1}^{m} \phi_i(x) \circ \mathrm{d}W_i(t)$
3. compute the energy $E(u_{n+1})$
4. accept u_{n+1}:
   - if $E(u_{n+1}) < E(u_n)$
   - otherwise, accept u_{n+1} with probability $\exp\left(-\frac{E(u_{n+1})-E(u_n)}{T(n)}\right)$
5. loop back to step 2, until some stopping condition is fulfilled

as a decreasing temperature. Its choice is not obvious. If the temperature decreases too fast, the process may get stuck in a local minimum; on the contrary, decreasing too slowly may be postpone convergence. A classical choice is $T(n) = T_0/\sqrt{n}$. The classical way to solve the previous minimization problem is to use a gradient descent method. The Euler-Lagrange equation is computed, leading to some evolution $\partial\Gamma/\partial t = \beta_c \mathbf{n}$, or equivalently, in the Level Set framework, to $\partial u/\partial t = \beta_c |Du|$. We will actually use the classical motion as heuristics that drive the evolution faster toward a minimum, and replace the dynamics of step 2, by

$$\mathrm{d}u = \beta_c |Du| \mathrm{d}t + |Du| \sum_{i=1}^{m} \phi_i(x) \circ \mathrm{d}W_i(t) \tag{31}$$

As often with genetic algorithms, the proof of the convergence of this algorithm toward a global minimum is still an open problem. However, practical simulations indicate that the above algorithm is more likely to overcome local minima than the classical approach. This is our main motivation, since local minima are the major problem of classical approaches. Note also, as already mentioned, that our framework can be used in cases when the shape gradient is too complex from a mathematical or computational point of view, or even impossible to compute.

5 Stochastic Active Contours

Our scheme could be used in the Geodesic Active Contours framework [6] where segmentation is based upon gradient intensity variations. Yet, a multiscale approach is often used successfully in that context to overcome the local minimum problem. Other segmentation schemes [29] use a region model (eg. texture, statistics) that is less adapted to multiscale. We will first focus on one such case, namely the single Gaussian statistics model in [31].

5.1 Single Gaussian model

In their unsupervised segmentation framework [31], the authors model each region of a gray-valued or color image I by a single Gaussian distribution of unknown mean μ_i and variance Σ_i. The case of two regions segmentation turns into minimizing the following energy:

$$E(\Gamma, \mu_1, \Sigma_1, \mu_2, \Sigma_2) = \int_{\Omega_1} e_1(x) + \int_{\Omega_2} e_2(x) + \nu \text{length}(\Gamma)$$

where Ω_1 is the region inside Γ, Γ_2 the outside, and $e_i(x) = -\log p_{\mu_i \Sigma_i}(I(x))$ with $p_{\mu_i \Sigma_i}(I(x)) = C|\Sigma_i|^{-1/2} e^{-(I(x)-\mu_i)^T \Sigma_i^{-1}(I(x)-\mu_i)/2}$ being the conditional probability density function of a given value $I(x)$ with respect to the hypothesis (μ_i, Σ_i). The parameters (μ_i, Σ_i), estimated from the pixel actually inside and outside Γ, are functions of Γ. Thus, the energy is a function of Γ only: $E(\Gamma, \mu_1, \Sigma_1, \mu_2, \Sigma_2) = E(\Gamma)$. Its Euler-Lagrange equation is not obvious, but finally simplifies into the minimization dynamics

$$\beta_c = e_2(x) - e_1(x) + \nu \, \text{div}\left(\frac{\text{D}\, u}{|\text{D}\, u|}\right)$$

The authors successfully segment two regions, even when they have the same mean but only different variances. However, the evolution could easily be stuck into some local minimum while a multiscale approach might modify the statistics so that no segmentation would be possible anymore. As demonstrated figure 6, a simple Simulated Annealing scheme with dynamics (31) overcomes this problem. Figure 7 shows the same phenomenon on a real image. Note that this image was successfully segmented by the authors of [29]. Yet, they used an adapted model of texture. Here, the Stochastic Active Contours framework succeeds in making a simple unsupervised single Gaussian model recover the correct regions.

5.2 Gaussian mixtures

As an illustration of the case when the Euler-Lagrange equation cannot be computed, we extend the previous method to region statistics modeled by a mixture of Gaussian distributions of parameters $\Theta_i = (\pi_i^1, \mu_i^1, \Sigma_i^1, ..., \pi_i^{n_i}, \mu_i^{n_i}, \Sigma_i^{n_i})$. with $\sum_j \pi_i^j = 1$. The conditional probability density function of a given value $I(x)$ becomes:

$$p_{\Theta_i}(I(x)) = \sum_{j=1}^{n_i} \pi_j p_{\mu_i^j \Sigma_i^j}(I(x))$$

The number of Gaussian distributions can be given, estimated at the initial time step, or dynamically evaluated using a Minimum Description Length criterion[30]

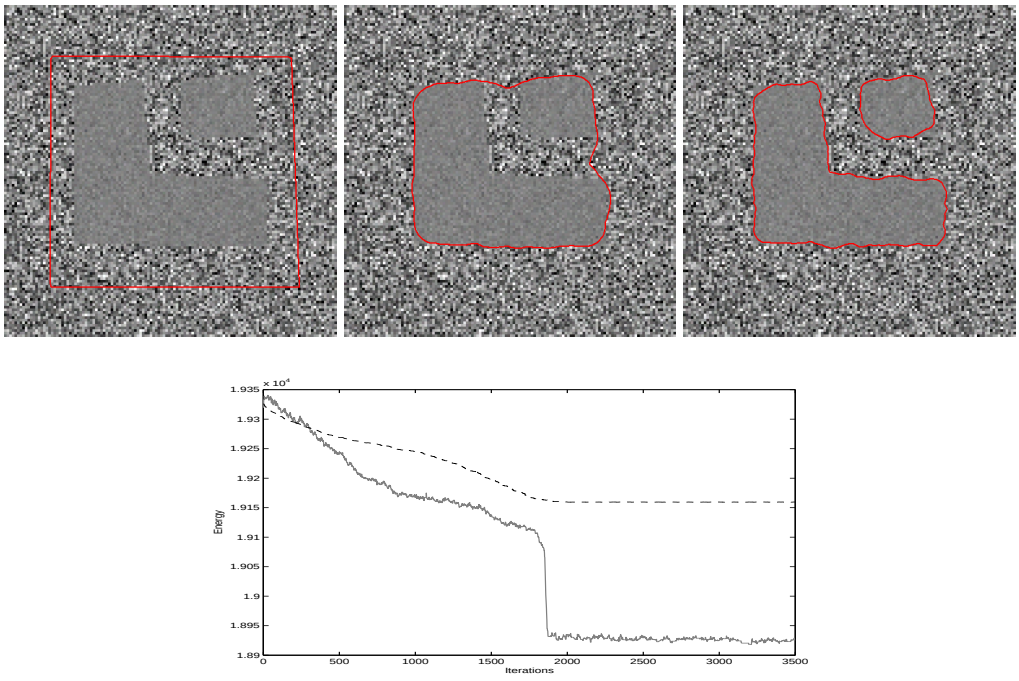


Figure 6: Segmentation of two regions modeled by two unknown Gaussian distributions (same mean, different variances). Top row: the initial curve, the final state of the classical approach stuck in a local minimum, and the final state of our method. Bottom row: evolution of the energy (dashed: deterministic method, solid: our method)

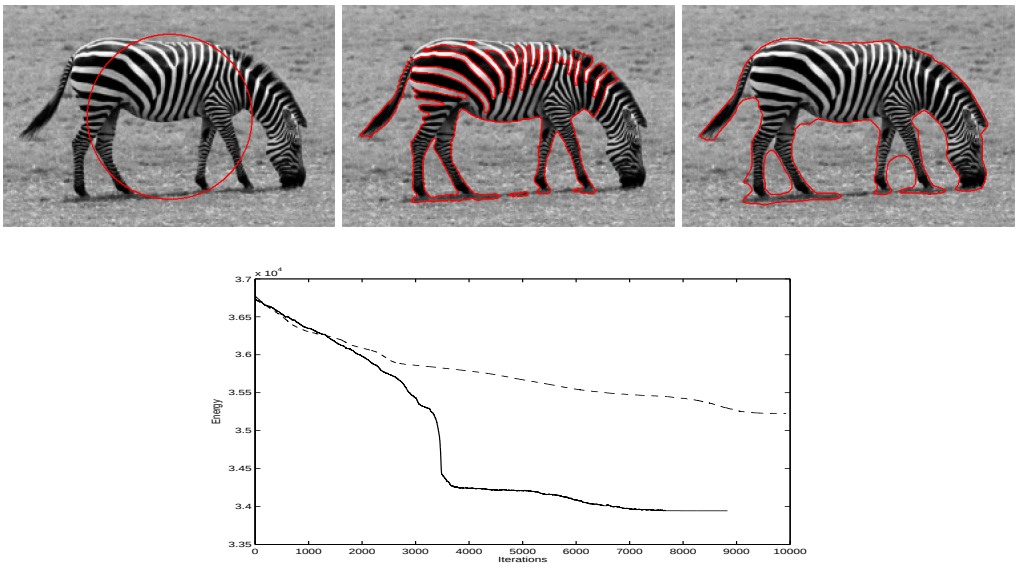


Figure 7: Segmentation of two regions modeled by two unknown Gaussian distributions. Top row: the initial curve, the final time step of the classical method, again stuck in a local minimum and the final step of our method. Bottom row: evolution of the energy (dashed: deterministic method, solid: our method)

or the Minimal Message Length method [36]. A large literature is dedicated to the problem of estimating Θ_i from input samples. We have used the original K-Means algorithm pioneered by MacQueen [24], although we have tested extensions like the Fuzzy-K-Means [3, 8], the K-Harmonic-Means [41], and the Expectation-Maximization algorithm (EM), first proposed in [9]. The latter solves iteratively

$$\widehat{\Theta}_i = \underset{\Theta_i}{\operatorname{argmax}} \int_{x \in \Omega_i} \log p_{\Theta_i}\left(I\left(x\right)\right) \, dx$$

(Please refer to appendix for details and references).

Our segmentation problem still consists in minimizing the same energy, with now $e_i(x) = -\log p_{\Theta_i}(I(x))$. Unfortunately, we now have to deal with a complex dependency of Θ_i with respect to Γ. In fact, the learning algorithm acts as a "black box" implementing $\Gamma \rightarrow \Theta_i(\Gamma)$. As a consequence, the Euler-Lagrange equation of the energy $E(\Gamma, \Theta_1(\Gamma), \Theta_2(\Gamma)) = E(\Gamma)$ cannot be computed. A deterministic contour evolution driven by $\beta_c = e_2 - e_1 + \nu\kappa$ may get stuck just because $\beta_c \mathbf{n}$ is not the exact gradient. Yet, the Stochastic Active Contours can still be used, with β_c as heuristics. As a simple illustration of this, let us consider the synthetic example of figure 8. The region to segment is a square. The square and the background are each modeled by a mixture of two equally weighted Gaussian distributions: $\Theta_i =$

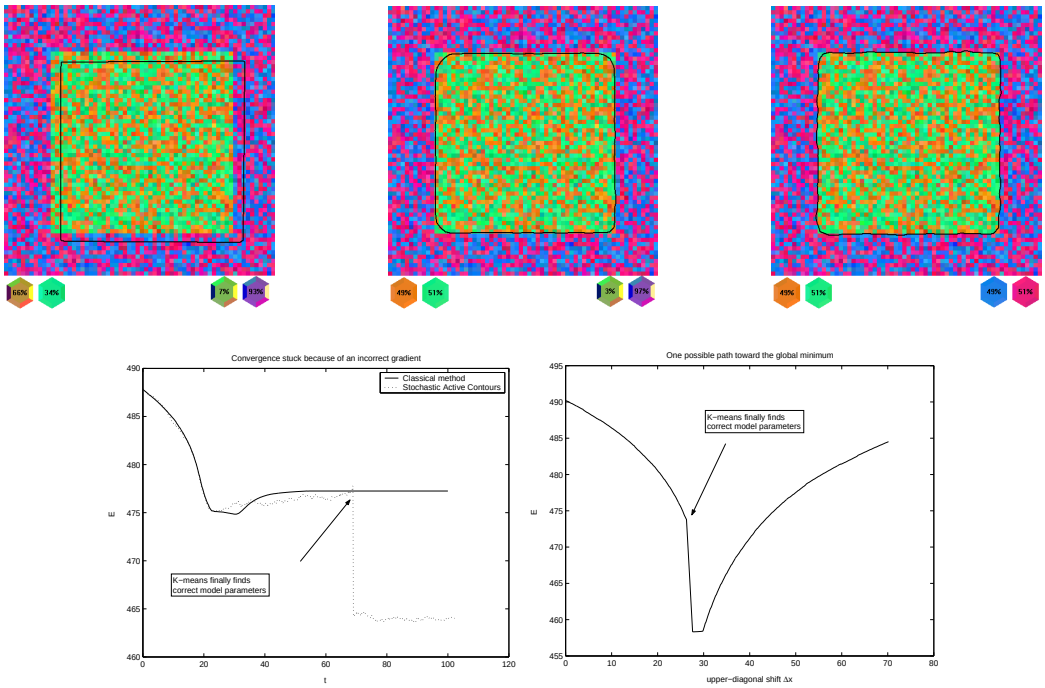


Figure 8: A case where the gradient is not correct (see text). Top row, from left to right: initial position, final position with the classical method (the model is not correctly recovered - see percentages in the hexagons), leading to rounded corners), final position with our method (the model is correctly recovered). Bottom left: evolution of the energy in both cases. Bottom right: energy for a translation of the curve that goes through the correct segmentation.

$(\frac{1}{2}, \mu_i^1, \Sigma, \frac{1}{2}, \mu_i^2, \Sigma)$. As an initial guess, we shift the square toward the bottom-right corner. Although a bit of the background is in Ω_1, Θ_1 is correctly estimated. Yet, for some reason, the K-Means algorithm estimates Θ_2 approximatively by $(1-\epsilon, \frac{\mu_2^1+\mu_2^2}{2}, \Sigma', \epsilon, \frac{\mu_1^1+\mu_1^2}{2}, \Sigma'')$. During its convergence, the deterministic method keeps such an incorrect Θ_2 and finally gets stuck at a roughly correct place but with an incorrect model, leaving some interior pixels outside (especially in the corners, because of the smoothing term of the energy). The colored hexagons below the images indicate the means and variances of the mixtures components and their respective weights. See also how the energy increases in the end! On the contrary, our method does not rely completely on the incorrect gradient only and finally *"discovers"* the correct model, leading to a somehow better fit. Notice the energy level drop-down when the K-Means algorithm ejects the interior pixels as negligible and shifts to the correct model. The last graph of figure 8 is a plot of the energy when the initial square is *manually* translated from the bottom-right corner to the upper-left one, going through the correct position. It clearly shows that the heuristic gradient by itself gets stuck in a local minimum , whereas our method comes much closer to the desired minimum.

6 Results

Even when the deterministic scheme converge more or less, our method shows a better ability to overcome local minima: figure 9 shows how Γ can be stuck leading to a dramatic evolution toward completely false regions. Finally, figure 10 shows some more examples on other real images. Animations corresponding to all the presented examples can be downloaded at

7 Conclusion

Based on recent work on Stochastic Partial Differential Equations by Lions and Souganidis, we have presented a simple and well-founded method to implement the stochastic motion of a surface in a Level Set framework. This method is used as the key point of a stochastic extension to standard shape optimization methods in Computer Vision. In the particular case of segmentation, we introduced the *Stochastic Active Contours*, a natural extension of the well-known active contours. Our method overcomes the local minima problem and can also be used when the Euler-Lagrange equation of the energy is out of reach. This extension is not time consuming: the only computational effort is computing the energy . Convincing results are presented with the segmentation of regions modeled by unknown statistics, namely single Gaussian distributions or mixtures of Gaussian

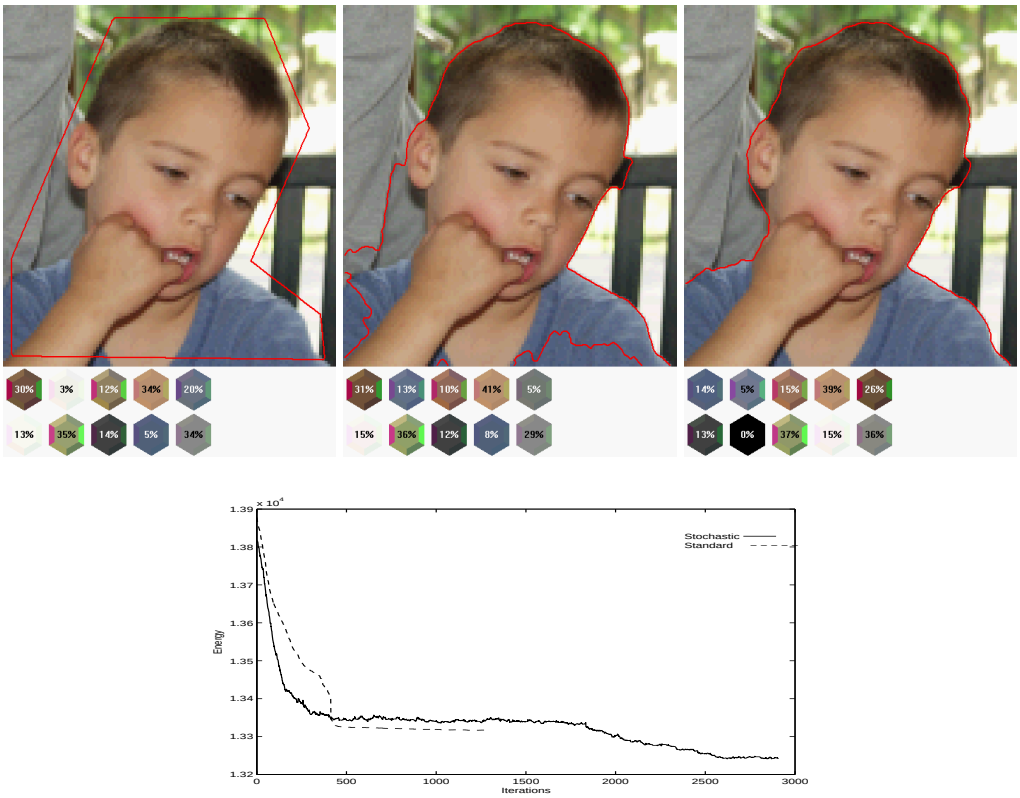


Figure 9: Segmentation of two regions modeled by two unknown Gaussian mixtures. Top row: the initial curve, the final state of the deterministic method, stuck in a local minimum and the final state of our method. Bottom row: evolution of the energy.

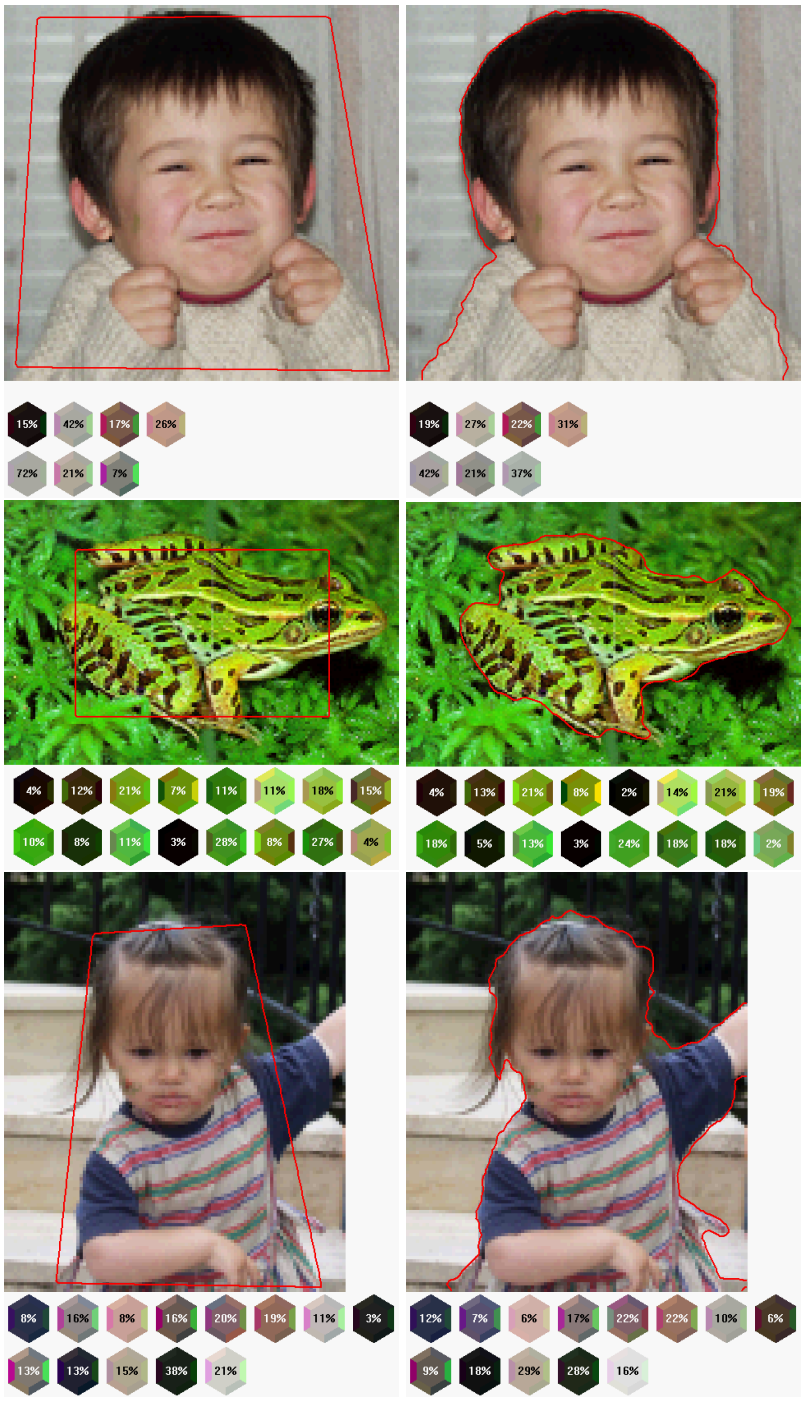


Figure 10: Segmentation of two regions modeled by two unknown Gaussian mixtures. Left column: the initial states. Right column: the corresponding final states of our method.

distributions. The way is now open for applying our principle to other Computer Vision problems but also in different fields where shape optimization problems arise, like in theoretical chemistry [5].

A Appendix A: Data Mining and Unsupervised Learning Survey

This part presents a survey of well known data mining algorithms (K-Means and EM algorithms) in case of fitting Gaussian Mixture Model (GMM) on a given data set. In our case, the data set's samples are pixels of a segmented regions.

Firstly, we will discuss of the K-Means algorithm and some extensions and, in a second part, of the EM algorithm.

K-Means algorithm is much more fast than EM and is often used in critical running time applications and in initialization step of EM algorithm.

However our segmentation scheme is based on minimizing:

$$E\left(\Gamma,\Theta\left(\Gamma\right)\right)=\int_{\Omega}-logp_{\Theta(\Gamma)}\left(I\left(x\right)\right)+\nu\text{length}\left(\Gamma\right)$$

So the gradient is:

$$\frac{\partial E\left(\Gamma,\Theta\left(\Gamma\right)\right)}{\partial\Gamma}=\frac{\partial E}{\partial\Gamma}\left(\Gamma,\Theta\left(\Gamma\right)\right)+\frac{\partial E}{\partial\Theta}\left(\Gamma,\Theta\left(\Gamma\right)\right)\frac{\partial\Theta}{\partial\Gamma}\left(\Gamma\right) \tag{32}$$

Since the second term of the gradient is too complex, we completely neglect it and use only the first term as a heuristic in a simulate annealing framework. Then using EM algorithm is more adapted than K-Means since it attempts at each time step to find the parameter Θ by a log-likelihood optimization such that the GMM fits best the region data set. If the EM algorithm manages to achieve the log-likelihood optimization, then $\frac{\partial\Theta}{\partial\Gamma}\left(\Gamma\right)=0$, so the second term vanishes in equation (32). And the heuristic gradient we use in our simulate annealing, is now the correct gradient.

A.1 K-Means and co.

K-Means is historically the first efficient and fast algorithm for data clustering problem [24], due to its computational low cost. Since, some extensions, improving robustness and classification, have been provided like Fuzzy-K-Means [3] and [8].

In the next sections, we will first elaborate on the K-Means algorithm and then we will elaborate on some of its extensions (concentrating on the Fuzzy-K-Means algorithm).

A.1.1 K-Means

This method provides K partitions of a data set X. Each sample of the data set, here pixel, is assigned with a class in order to minimize the within-class sum square errors functional:

$$KM(X) = \sum_{i=1}^{N} \min_{j=1..K} d(x_i, c_j)^2$$

where $x_i \in X$, c_j is the center of jth class and d is the distance between x_i and c_j.

Different distances can be used like Manhattan, Euclidian or Mahalanobis distance. Each distance has its own advantages and disadvantages. Manhattan distance (L1 distance), is more robust to outliers although its unit sphere is a cube unlike Euclidian distance (L2 distance) which has a more natural unit sphere. Mahalanobis distance - the distance associated to $\langle\ ,\ \rangle_\Sigma$ where Σ is the covariance matrix of the considered class - takes into account much more information from data set (covariance matrix) but using Mahalanobis distance makes K-Means algorithm mush more numerically instable.

Unlike Mahalanobis distance, when working with L1 or L2 distances, one has to deal with another problem: clusters (or classes) will have approximately the same size. However Mahalanobis is computationally costly and L2 distance is a good compromise between speed and clustering quality.

A detailed description of K-Means algorithms can be found in [10] pages 526-527. There are many ways to choose initial centers, like choosing them randomly or uniformly in the data set X or in the data space. It does not exist a unique way to initialize K-Means, many ideas have been proposed in that way to enhance robustness of K-Means algorithm. Although K-Means is very sensitive to initialization step and in most case, random initialization is the best choice.

Sensitivity wrt. the initial data is an unwanted feature of the K-Means algorithm. A modified version of the algorithm, named *Fuzzy K-Means*, was proposed by deGruijter and McBratney in order to improve robustness.

A.1.2 Fuzzy-K-Means and other extensions

The basic idea was simple: if K-Means makes hard partitioning of data space, why don't try a soft partitioning ? In fact, in K-Means, a data's sample assigned to a class can't influence another class. Fuzzy-K-Means lets each sample of the data set interfere with *all* the classes. (see figure 12)

The energy to minimize is now:

$$FKM(X) = \sum_{i=1}^{N} \sum_{j=1}^{K} m_{ij} d(x_i, c_j)^2$$

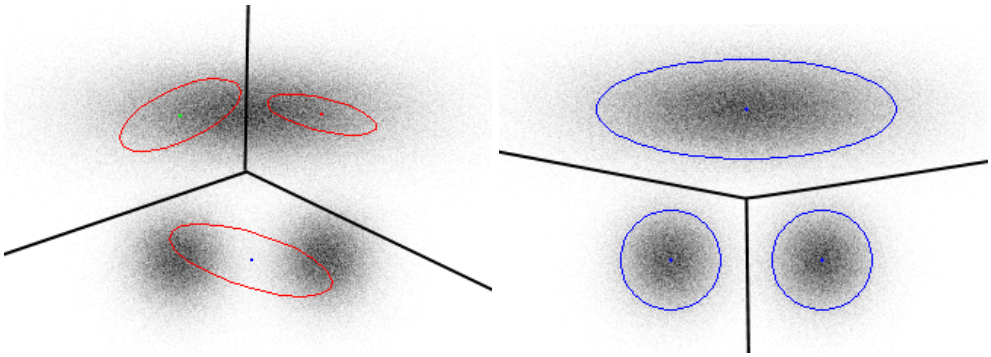


Figure 11: On the left, in red, result of the classification by K-Means and, in black, the borders of the partition. On the right, the true solution.

where:

$$\sum_{j=1}^{K} m_{ij} = 1$$

$$\sum_{i=1}^{N} m_{ij} > 0$$

m_{ij} is called *membership coefficient* and it influences the impact of x_i on c_j. It is computed through a new parameter, ϕ, which represents the fuzziness of the partition. When $\phi = 1$, Fuzzy K-Means does the same thing as K-Means. The greater ϕ is, the more groups overlap.

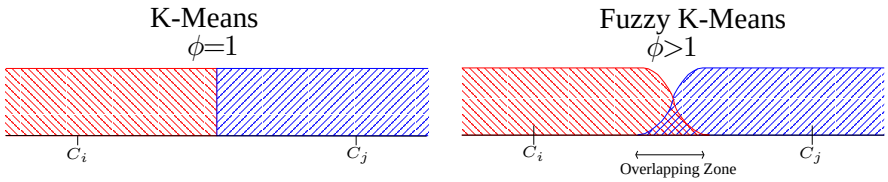


Figure 12: On the left, the representation of the *membership function* in case of Hard partitionning (K-Means). On the right, the *membership function* in case of Soft partitionning (Fuzzy K-Means with ϕ>1).

$$m_{ij} = \frac{d\left(x_i, c_j\right)^{2/(\phi-1)}}{\sum_{m=1}^{K} d\left(x_i, c_m\right)^{2/(\phi-1)}} \tag{33}$$

and c_j is updated using:

$$c_j = \frac{\sum_{i=1}^{N} m_{ij} x_i}{\sum_{i=1}^{N} m_{ij}} \tag{34}$$

The Fuzzy-K-Means algorithm could be found in [10] pages 528-529.

A.1.3 Conclusion

K-Means algorithm is quite sensitive to initialization and does not often classify correctly a synthetic data set built from a gaussian mixture distribution. Fuzzy-K-Means is more robust and less sensitive to initial conditions than K-Means algorithm. But K-Means and Fuzzy-K-Means are Centers-based classifier algorithm and they are very mush alike. Both of them suffer from common disadvantages, in particular the fact that clusters have approximately the same size.

A.2 Expectation Maximization: EM algorithm

The Expectation-Maximization algorithm (EM) is a parameter estimation algorithm based on likelihood maximization. Here we note Θ all the parameters of the model and denote by $p(x|\Theta)$ the density function. If χ denotes the data set, likelihood is defined by:

$$L(\Theta|\chi) = p(\chi|\Theta) = \prod_{x \in \chi} p(x|\Theta) \tag{35}$$

EM deals with incomplete data observation. We denote by X the observed data, Y the missing data and $Z = (X, Y)$ the complete data.

EM is a two step algorithm. During the first step, called E-Step, it finds the expected value of $\log L(\Theta|Z) = \log L(\Theta|X, Y) = \log p(X, Y|\Theta)$ wrt. Y, i.e. calculate:

$$Q(\Theta, \Theta') = E\left[\log p(X, Y|\Theta) \,|X, \Theta'\right] \tag{36}$$

The second step, the Maximization Step (M-Step) maximizes $Q(\Theta, \Theta')$ wrt. Θ and solves the equation:

$$\widehat{\Theta} = \underset{\Theta}{\operatorname{argmin}}\, Q(\Theta, \Theta') \tag{37}$$

EM consists in applying iteratively E-Step and M-Step until convergence (i.e. until the log-likelihood delta between two iterations drops below a given threshold).

A.2.1 EM and Gaussian Mixture Model

Here we use a Gaussian Mixture Model:

$$p\left(x|\Theta\right) = \sum_{i=1}^{K} \pi_i G_{\mu_i,\Sigma_i}\left(x\right) \tag{38}$$

So the log-likelihood for the incomplete data is:

$$\log\left(L\left(\Theta|X\right)\right) = \log\prod_{i=1}^{N} p\left(x_i|\Theta\right) = \sum_{i=1}^{N}\log\left(\sum_{j=1}^{K}\pi_j G_{\mu_j,\Sigma_j}\left(x_i\right)\right)$$

And the log-likelihood for the complete data is:

$$\begin{aligned}
\log\left(L\left(\Theta|X,Y\right)\right) &= \log\prod_{i=1}^{N} p\left(x_i,y_i|\Theta\right) &(39)\\
&= \sum_{i=1}^{N}\log\left(\sum_{j=1}^{K}\delta_{y_i}\left(j\right)\pi_j G_{\mu_j,\Sigma_j}\left(x_i\right)\right) &(40)\\
&= \sum_{i=1}^{N}\log\left(\pi_{y_i} G_{\mu_{y_i},\Sigma_{y_i}}\left(x_i\right)\right) &(41)
\end{aligned}$$

where y_i, the missing data, is such that $G_{\mu_{y_i},\Sigma_{y_i}}$ is the gaussian from which x_i is a realization.

By using Bayes' rule, we find that E-Step consists in:

$$p\left(y_i|x_i,\Theta'\right) = \frac{\pi'_{y_i} G_{\mu'_{yi},\Sigma'_{yi}}\left(x_i\right)}{\sum_{j=1}^{K}\pi'_j G_{\mu'_j,\Sigma'_j}\left(x_i\right)} \tag{42}$$

and by derivating M-Step consists in:

$$\pi_m^{new} = \frac{1}{N}\sum_{i=1}^{N} p\left(m|x_i,\Theta'\right) \tag{43}$$

$$\mu_m^{new} = \frac{\sum_{i=1}^{N} p\left(m|x_i,\Theta'\right) x_i}{\sum_{i=1}^{N} p\left(m|x_i,\Theta'\right)} \tag{44}$$

$$\Sigma_m^{new} = \frac{\sum_{i=1}^{N} p\left(m|x_i,\Theta'\right)\left(x_i-\mu_i^{new}\right)\left(x_i-\mu_i^{new}\right)^T}{\sum_{i=1}^{N} p\left(m|x_i,\Theta'\right)} \tag{45}$$

$$\tag{46}$$

As it was said in previous section, EM algorithm consists in iterating E-Step followed by M-Step until convergence. It can be prove that one iteration of EM always increase the likelihood $L\left(\Theta|\chi\right)$ in [26] and [25].

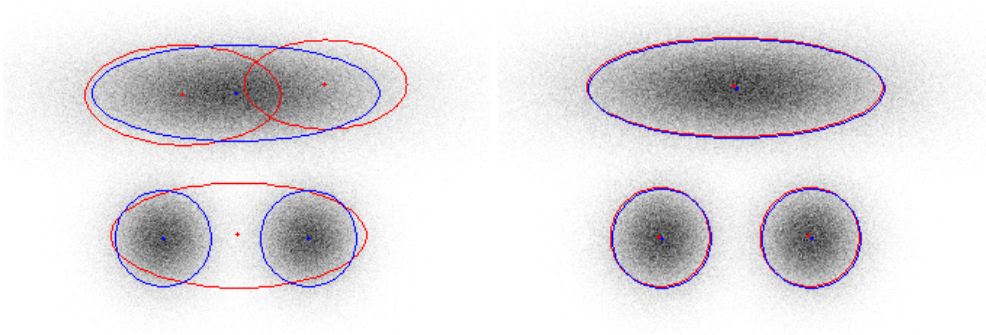


Figure 13: On the left, in red, best result of K-Means over 20 shots and, in blue, the the true solution. On the right, in red, best result of EM over 20 shots and, in blue, the the true solution.

EM algorithm provides better results and is more robust to initialization than K-Means or Fuzzy-K-Means (see figure 13).

A.2.2 EM and Extensions

Like K-Means, extensions of EM algorithm have been proposed to improved convergence speed or robustness. Stochastic EM (SEM) is one of the most used ex-

tensions of EM [7]. However to avoid Minima loci, Split and Merge EM (SMEM) improves greatly the robustness of the algorithm but it is computationally costly and works well on large data sets [33]. Color images of 256 pixels by side is not sufficient since pixels live in 3D color space.

A.3 Unsupervised Learning ?

The classic model of GMM isn't totally unsupervised. In fact the number of gaussians has to be set by the user. To provide a totally user-friendly algorithm, we have tested a MDL based criterion [30]. Minimum Description Length criterion allows us to compare the log-likelihood of two GMM who don't have the same number of underlying meaningful distributions. The MDL criterion is defined by:

$$MDL\left(K\right) = \frac{K}{2}\left[1 + d + \frac{d\left(d+1\right)}{2}\right]\log N$$

where K is the number of gaussians, d the space dimension and N the number of data samples. We introduce a criterion to compare two GMM:

$$J\left(\Theta, K\right) = -\sum_{i=1}^{N}\log p\left(x_i|\Theta\right) + MDL\left(K\right)$$

where p is a GMM of K gaussians defined as (38), Θ represents all the parameters (π_i, μ_i, Σ_i) of the GMM. Nevertheless when using MDL, one will have an extra parameter to calibrate K_{max} which represents the maximum number of gaussians in GMM.

As $MDL(K)$ does not depend on Θ, minimizing $J(\Theta, K)$ consists in K_{max} runs of EM algorithm to minimize the first part of J wrt. Θ with K fixed and then choosing the couple (θ_K, K) that minimize J over all $K \in \{1..K_{max}\}$.

However if this method provides a better fitted model, it is K_{max} times more long that a simple EM algorithm and needs large data set.

References

[1] L. Ballerini. Genetic snakes for image segmentation. In *Evolutionary Image Analysis, Signal Processing and Telecommunication*, volume 1596 of *Lecture Notes in Computer Science*, pages 59–73. Springer, 1999.

[2] G. Ben-Arous, A. Tannenbaum, and O. Zeitouni. Stochastic approximations to curve shortening flows via particle systems. Technical report, Technion Institute, 2002.

[3] James C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, 1981.

[4] R. Buckdahn and J. Ma. Stochastic viscosity solutions for nonlinear stochastic partial differential equations. *Stochastic Processes and their Applications*, 93:181–228, 2001.

[5] E. Cances, R. Keriven, F Lodier, and A. Savin. How electrons guard the space: shape optimization with proba bility distribution criteria. *Theoret. Chem. Acc.*, 2004.

[6] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *The International Journal of Computer Vision*, 22(1):61–79, 1997.

[7] Gilles Celeux, Didier Chauveau, and Jean Diebolt. On stochastic versions of the EM algorithm. Technical Report RR-2514.

[8] J.J. deGruijter and A.B. McBratney. A modified fuzzy k means for predictive classification. In *Bock,H.H.(ed) Classification and Related Methods of Data Analysis*, pages 97–104. Elsevier Science, Amsterdam., 1988.

[9] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39:1–38, 1977.

[10] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley, 2001. DUD r3 01:1 1.Ex.

[11] Olivier Faugeras and Renaud Keriven. Variational principles, surface evolution, PDE's, level set methods and the stereo problem. *IEEE Transactions on Image Processing*, 7(3):336–344, March 1998.

[12] T.C. Gard. *Introduction to Stochastic Differential Equations*. Marcel Dekker, New York and Basel, 1988.

[13] G.-S. Jiang and D. Peng. Weighted ENO schemes for Hamilton-Jacobi equations. *SIAM Journal of Scientific Computing*, 21(6):2126–2143, 2000.

[14] I. Karatzas and S.E. Shreve. *Brownian motion and stochastic calculus*. Springer-Verlag, New York, 1991.

[15] M. Kass, A. Witkin, and D. Terzopoulos. SNAKES: Active contour models. In *Proceedings of the 2nd International Conference on Computer Vision*, volume 1, pages 321–332, Tampa, FL, January 1988. IEEE Computer Society Press.

[16] M.A. Katsoulakis and A.T. Kho. Stochastic curvature flows: Asymptotic derivation, level set formulation and numerical experiments. *Journal of Interfaces and Free Boundaries*, 3:265–290, 2001.

[17] S. Kirkpatrick, C.D. Jr. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598), 1983.

[18] H. Kunita. *Stochastic Flows and Stochastic Differential Equations*. Cambridge University Press, 1990.

[19] P.L. Lions and P.E. Souganidis. Fully nonlinear stochastic partial differential equations. *C.R. Acad. Sci. Paris Ser. I Math*, 326:1085–1092, 1998.

[20] P.L. Lions and P.E. Souganidis. Fully nonlinear stochastic partial differential equations: nonsmooth equations and applications. *C.R. Acad. Sci. Paris Ser. I Math*, 327:735–741, 1998.

[21] P.L. Lions and P.E. Souganidis. Equations aux derivees partielles stochastiques non lineaires et solutions de viscosite. Seminaire X-EDP (Unpublished Lecture), 1999.

[22] P.L. Lions and P.E. Souganidis. Fully nonlinear stochastic partial differential equations with semilinear stochastic dependence. *C.R. Acad. Sci. Paris Ser. I Math*, 331:617–624, 2000.

[23] P.L. Lions and P.E. Souganidis. Uniqueness of weak solutions of fully nonlinear stochastic partial differential equations. *C.R. Acad. Sci. Paris Ser. I Math*, 331:783–790, 2000.

[24] J. MacQueen. Some methods for classification and analysis of multivariate. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.

[25] G.J. McLachlan and P. David. *Finite Mixture Models*. Wiley, New York, 2000.

[26] G.J. McLachlan and T. Krishnan. *The EM algorithm and extensions*. Wiley, New York, 1997.

[27] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and Teller; E. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.

[28] S. Osher and J. Sethian. Fronts propagating with curvature dependent speed: algorithms based on the Hamilton–Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.

[29] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *The International Journal of Computer Vision*, 46(3):223, 2002.

[30] J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scientific Publishing Company, 1989.

[31] M. Rousson and R. Deriche. A variational framework for active and adaptative segmentation of vector valued images. In *Proc. IEEE Workshop on Motion and Video Computing*, pages 56–62, Orlando, Florida, December 2002.

[32] G. Storvik. A bayesian approach to dynamic contours through stochastic sampling and simulated annealing. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(10):976–986, october 1994.

[33] Naonori Ueda, Ryohei Nakano, Zoubin Ghahramani, and Geoffrey E. Hinton. Smem algorithm for mixture models. *Neural Computation*, 12(9):2109–2128, 2000.

[34] G. Unal, H. Krim, and A. Yezzi. Stochastic differential equations and geometric flows. *IEEE Transactions in Image Processing*, 11(12):1405–1416, 2002.

[35] G. Unal, D. Nain, G. Ben-Arous, N. Shimkin, A. Tannenbaum, and O. Zeitouni. Algorithms for stochastic approximations of curvature flows. In *IEEE International Conference on Image Processing*, 2003.

[36] C.S. Wallace and D.M. Boulton. An information measure for classification. *Computer Journal*, 11(2):185–194, 1968.

[37] J.B. Walsh. An introduction to stochastic partial differential equations. In *Ecole d'Été de Probabilités de Saint-Flour*, volume XIV-1180 of *Lecture Notes in Math.* Springer, 1994.

[38] N.K. Yip. Stochastic motion by mean curvature. *Arch. Rational Mech. Anal.*, 144:331–355, 1998.

[39] N.K. Yip. Stochastic curvature driven flows. In G. Da Prato and L. Tubaro, editors, *Stochastic Partial Differential Equations and Applications*, volume

227 of *Lecture Notes in Pure and Applied Mathematics*, pages 443–460. Springer, 2002.

[40] N.K Yip and P.E. Souganidis. Uniqueness of motion by mean curvature perturbed by stochastic noise. *Annales de l'Institut Henri Poincare - analyse non lineaire*, (21):1–23, 2004.

[41] B. Zhang and M. Hsu. K-harmonic means – a data clustering algorithm, 1999.