A stochastic algorithm finding generalized means on compact manifolds

Marc Arnaudon[†] and Laurent Miclo[‡]

[†] Institut de Mathématiques de Bordeaux, UMR 5251
Université de Bordeaux and CNRS, France

[‡] Institut de Mathématiques de Toulouse, UMR 5219
Université de Toulouse and CNRS, France

Abstract

A stochastic algorithm is proposed, finding the set of generalized means associated to a probability measure ν on a compact Riemannian manifold and a continuous cost function κ on $M \times M$. Generalized means include p-means for $p \in (0, \infty)$, computed with any continuous distance function, not necessarily the Riemannian distance. They also include means for lengths computed from Finsler metrics, or for divergences.

The algorithm is fed sequentially with independent random variables $(Y_n)_{n\in\mathbb{N}}$ distributed according to ν and this is the only knowledge of ν required. It evolves like a Brownian motion between the times it jumps in direction of the Y_n. Its principle is based on simulated annealing and homogenization, so that temperature and approximations schemes must be tuned up. The proof relies on the investigation of the evolution of a time-inhomogeneous $\mathbb{L}^2$ functional and on the corresponding spectral gap estimates due to Holley, Kusuoka and Stroock.

Keywords: Stochastic algorithms, simulated annealing, homogenization, probability measures on compact Riemannian manifolds, intrinsic means, instantaneous invariant measures, Gibbs measures, spectral gap at small temperature.

MSC2010: first: 60D05 , secondary: 58C35, 60J75, 37A30, 47G20, 53C21, 60J65.

1 Introduction

The purpose of this paper is to present a stochastic algorithm finding the generalized means of a probability measure ν defined on a compact manifold M. A Riemannian metric is used, only for the algorithm.

Algorithms for finding means, medians or minimax centers have been the object of many investigations, see e.g. [20], [25], [13], [14], [7], [26], [9], [1], [5], [10], [6]. In these references a gradient descent algorithm is used, or a stochastic version of it avoiding to compute the gradient of the functional to minimize. Either the functional to minimize has only one local minimum which is also global, or ([9]) a local minimum is seeked. The case of Karcher means in the circle is investigated in [11] and [16]. In this special situation the global minimum of the functional can be found by more or less explicit formula.

For generalized means on compact manifolds the situation is different since the functional (1) to minimize may have many local minima, and no explicit formula for a global minimum can be expected. In [2] the case of compact symmetric spaces has been investigated and a continuous inhomogeneous diffusion process has been constructed which converges in probability to the set of p-means. In [3] the case of p-means on the circle is treated. A Markov process is constructed which has Brownian continuous part and more and more frequent jumps in the direction of independent random variables with law ν. It is proven that it converges in probability to the set of p-means of ν. Both [2] and [3] use simulated annealing techniques.

The purpose of this paper is to extend the construction in [3] to all compact manifolds, and to all generalized means.

So let be given ν a probability measure on M, a compact Riemannian manifold. Denote by $\kappa : M \times M \to \mathbb{R}$ a continuous function and consider the continuous mapping

$$U \;:\; M \ni x \mapsto \int_M \kappa(x,y)\,\nu(dy). \tag{1}$$

A global minimum of U is called a κ-mean or generalized mean of ν and let $\mathcal{M}$ be their set, which is non-empty in the above compact setting.
In practice the knowledge of ν is often given by a sequence $Y := (Y_n)_{n\in\mathbb{N}}$ of independent random variables, identically distributed according to ν. So let us present a stochastic algorithm using this data and enabling to find some elements of $\mathcal{M}$. It is based on simulated annealing and homogenization procedures. Thus we will need respectively an inverse temperature evolution $\beta : \mathbb{R}_+ \to \mathbb{R}_+$, an inverse speed up evolution $\alpha : \mathbb{R}_+ \to \mathbb{R}_+^*$ and a regularization function $\delta : \mathbb{R}_+ \to \mathbb{R}_+^*$. Typically, β_t is non-decreasing, α_t and δ_t are non-increasing and we have $\lim_{t\to+\infty}\beta_t = +\infty$, $\lim_{t\to+\infty}\alpha_t = 0$ and $\lim_{t\to+\infty}\delta_t = 0$, but we are looking for more precise conditions so that the stochastic algorithm we describe below finds $\mathcal{M}$.
For finding $\mathcal{M}$, we will use a regularization of κ with the heat kernel $p(\delta,x,z)$. So define for $\delta > 0$ $\kappa_\delta(x,y) = \displaystyle\int_M p(\delta,x,z)\kappa(z,y)\,\lambda(dz)$ where λ denotes the Lebesgue measure (namely the unnormalized Riemannian measure) and

$$U_\delta \;:\; M \ni x \mapsto \int_M \kappa_\delta(x,y)\,\nu(dy). \tag{2}$$

Notice any other regularization which would satisfy the estimates (16) and (23) below and which would be easier to compute could be used instead of the heat kernel.

Let $N := (N_t)_{t\geqslant 0}$ be a standard Poisson process: it starts at 0 at time 0 and has jumps of length 1 whose interarrival times are independent and distributed according to exponential random variables of parameter 1. The process N is assumed to be independent from the chain Y. We define the speeded-up process $N^{(\alpha)} := (N_t^{(\alpha)})_{t\geqslant 0}$ via

$$\forall\, t \geqslant 0, \qquad N_t^{(\alpha)} \;:=\; N_{\int_0^t \frac{1}{\alpha_s}\,ds}. \tag{3}$$

Consider the time-inhomogeneous Markov process $X := (X_t)_{t\geqslant 0}$ which evolves in M in the following way: if $T > 0$ is a jump time of $N^{(\alpha)}$, then X jumps at the same time, from X_{T-} to $X_T := \phi_\delta(\beta_T\alpha_T, X_{T_-}, Y_{N_T^{(\alpha)}})$, where

$$s \mapsto \phi_\delta(s, x, y) \in M \tag{4}$$

is the value at time s of the flow started at x of the vector field $z \mapsto -\frac{1}{2}\nabla_z \kappa_\delta(\cdot, y)$. In particular

$$\phi'_\delta(s, x, y) = -\frac{1}{2}\nabla\kappa_\delta(\cdot, y)(\phi_\delta(s, x, y)) \tag{5}$$

where ϕ'_δ denotes the derivative with respect to the first variable.

Typically we will have $\lim_{t\to+\infty} \alpha_t\beta_t\|\nabla\kappa_{\delta_t}\|_\infty = 0$, so that for sufficiently large jump-times T, X_T will be "between" X_{T-} and $Y_{N_T^{(\alpha)}}$ and quite close to X_{T-}. To proceed with the construction, we require that between consecutive jump times (and between time 0 and the first jump time), X evolves as a Brownian motion, relatively to the Riemannian structure of M (see for instance the book of Ikeda and Watanabe [19]) and independently of Y and N. Informally, the evolution of the algorithm X can be summarized by the Itô equation (in centers of exponential charts)

$$\forall\ t \geqslant 0, \qquad dX_t \quad = \quad \sigma(X_t)dB_t + \overrightarrow{X_{t-}\phi_{\delta_t}(\beta_t\alpha_t, X_{t_-}, Y_{N_t^{(\alpha)}})}\, dN_t^{(\alpha)}$$

where $(B_t)_{t\geqslant 0}$ is a Brownian motion on some $\mathbb{R}^m$, for all $x \in M$ $\sigma(x) : \mathbb{R}^m \to T_xM$ is linear satisfying $\sigma\sigma^* = \mathrm{id}$, $x \mapsto \sigma(x)$ is smooth, and where $(Y_{N_t^{(\alpha)}})_{t\geqslant 0}$ should be interpreted as a fast auxiliary process. We used the notation $\overrightarrow{xy} \in T_xM$ for the initial speed of the minimal geodesic from $x \in M$ to sufficiently close $y \in M$ in time 1 ($\overrightarrow{xy} = \exp_x^{-1}(y)$). The law of X is then entirely determined by the initial distribution $m_0 = \mathcal{L}(X_0)$. More generally at any time $t \geqslant 0$, denote by m_t the law of X_t.

We will prove that the above algorithm X finds in probability at large times the set of means $\mathcal{M}$. Let us define a constant $b \geqslant 0$, coming from the theory of simulated annealing (cf. for instance Holley, Kusuoka and Stroock [15]) in the following way. For any $x, y \in M$, let $\mathcal{C}_{x,y}$ be the set of continuous paths $p := (p(t))_{0\leqslant t\leqslant 1}$ going from $p(0) = x$ to $p(1) = y$. The elevation $U(p)$ of such a path p relatively to the function U in (1) is defined by

$$U(p) \quad := \quad \max_{t\in[0,1]} U(p(t))$$

and the minimal elevation $U(x, y)$ between x and y is given by

$$U(x, y) \quad := \quad \min_{p\in\mathcal{C}_{x,y}} U(p).$$

Then we consider

$$b(U) \quad := \quad \max_{x,y\in M} U(x, y) - U(x) - U(y) + \min_M U \tag{6}$$

This constant can also be seen as the largest depth of a well not encountering a fixed global minimum of U. Namely, if $x_0 \in \mathcal{M}$, then we have

$$b(U) \quad = \quad \max_{y\in M} U(x_0, y) - U(y) \tag{7}$$

independently of the choice of $x_0 \in \mathcal{M}$: the proof of $b(U) \leqslant \max_{y\in M} U(x_0, y) - U(y)$ is a direct consequence of the inequality $U(x, y) \leqslant \max(U(x, x_0), U(x_0, y))$ and the other direction is trivial.

With these notations, the main result of this paper is:

Theorem 1 *For any scheme of the form*

$$\forall\ t \geqslant 0, \qquad \begin{cases} \alpha_t &:= (1+t)^{-1} \\ \beta_t &:= c^{-1}\ln(1+t) \\ \delta_t &:= \ln(2+t)^{-1} \end{cases} \tag{8}$$

where $c > b(U)$, we have for any neighborhood $\mathcal{N}$ of $\mathcal{M}$ and for any m_0,

$$\lim_{t\to+\infty} m_t[\mathcal{N}] \quad = \quad 1 \tag{9}$$

Thus to find a element of $\mathcal{M}$ with a high probability, one should pick up the value of X_t for sufficiently large times t.

A crucial ingredient of the proof of this convergence are the Gibbs measures associated to the potentials U_δ. They are defined as the probability measures $\mu_{\beta,\delta}$ given for any $\beta \geqslant 0$ by

$$\mu_{\beta,\delta}(dx) \quad := \quad \frac{\exp(-\beta U_\delta(x))}{Z_{\beta,\delta}}\,\lambda(dx) \tag{10}$$

where $Z_{\beta,\delta} := \int \exp(-\beta U_\delta(x))\,\lambda(dx)$ is the normalizing factor.
Indeed we will show that $\mathcal{L}(X_t)$ and μ_{β_t,δ_t} become closer and closer as $t \geqslant 0$ goes to infinity in the sense of $\mathbb{L}^2$ variation.

By uniform continuity of κ we can easily prove that U_δ converges uniformly to U as $\delta \to 0$. As a consequence, for any neighbourhood $\mathcal{N}$ of $\mathcal{M}$, $\mu_{\beta,\delta}(\mathcal{N})$ converges to 1 as $\beta \to \infty$, uniformly in $\delta \leqslant \delta_0$ for some $\delta_0 > 0$ depending on $\mathcal{N}$. All this will prove the theorem.

The main difference in the method between the present work and [3] concerns the definition of the jumps. Instead of following the geodesic from the current position to a realization of ν, the process jumps to $\phi_{\delta_t}(\beta_t\alpha_t, X_{t-}, Y_{N_t^{(\alpha)}})$. The calculations are much easier, and this allows to consider more general cost functions κ. The drawback is that the implementation may be more complicated.

The cost functions κ are only assumed to be continuous. So this includes distances at the power p, ρ^p, which lead to p-means, for all $p \in (0,\infty)$. Notice the case $p \in (0,1)$ has never been considered in previous works. This also includes lengths for Finsler metrics, all divergences for parametric statistical models (Kullback-Leibler, Jeffrey, Chernoff, Hellinger...).

The paper is constructed on the following plan. In Section 2 we obtain an estimate of $L^*_{\alpha,\beta,\delta}\mathbb{1}$ where $L^*_{\alpha,\beta,\delta}$ is the adjoint of $L_{\alpha,\beta,\delta}$ in $L^2(\mu_{\beta,\delta})$, $L_{\alpha,\beta,\delta}$ is the generator of the process X_t described above but with constant α, β, δ. The proof of Theorem 1 is given in Section 3. For this proof, the estimate of Section 2 is crucial to see how close is the instantaneous invariant measure associated to the algorithm at large times $t \geqslant 0$ to the Gibbs measures associated to the potential U_{δ_t} and to the inverse temperature β_t^{-1}.

2 Regularity issues

Let us consider a general probability measure ν on M. For any $\alpha > 0$, $\beta \geqslant 0$ and $\delta > 0$, we are interested into the generator $L_{\alpha,\beta,\delta}$ defined for f from $\mathcal{C}^2(M)$ via

$$\forall\ x \in M, \qquad L_{\alpha,\beta,\delta}[f](x) \quad := \quad \frac{1}{2}\triangle f(x) + \frac{1}{\alpha}\int \left(f(\phi_\delta(\beta\alpha, x, y)) - f(x)\right)\,\nu(dy) \tag{11}$$

We will prove in Section 3 that $\mathcal{L}(X_t)$ gets closer and closer to the Gibbs distribution μ_{β_t,δ_t} as $t \to \infty$. Since for large $\beta \geqslant 0$, $\mu_{\beta,\delta}$ concentrates around $\mathcal{M}$ uniformly in δ sufficiently small, this will be sufficient to establish Theorem 1.

The remaining part of this section is devoted to a quantification of what separates $\mu_{\beta,\delta}$ from being an invariant probability of $L_{\alpha,\beta,\delta}$, for $\alpha > 0$, $\delta > 0$ and $\beta \geqslant 0$. As it will become clearer in the next section, a practical way to measure this discrepancy is through the evaluation of $\mu_{\beta,\delta}[(L^*_{\alpha,\beta,\delta}[\mathbb{1}])^2]$, where $L^*_{\alpha,\beta,\delta}$ is the dual operator of $L_{\alpha,\beta,\delta}$ in $\mathbb{L}^2(\mu_{\beta,\delta})$ and where $\mathbb{1}$ is the constant function taking the value 1. Indeed, it can be seen that $L^*_{\alpha,\beta,\delta}[\mathbb{1}] = 0$ in $\mathbb{L}^2(\mu_{\beta,\delta})$ if and only if $\mu_{\beta,\delta}$ is invariant for $L_{\alpha,\beta,\delta}$. Before being more precise about the definition of $L^*_{\alpha,\beta,\delta}$, we need an elementary result, where we will use the following notations: for any $s \in \mathbb{R}$, $T_{\delta,y,s}$ is the operator acting on measurable functions f defined on M via

$$\forall\, x \in M, \qquad T_{\delta,y,s}f(x) \;\; := \;\; f(\phi_\delta(s,x,y)) \tag{12}$$

Lemma 2 *For any $y \in M$, any $s \in [0,1)$ and any measurable and bounded functions f,g, we have*

$$\int_M g(x)T_{\delta,y,s}f(x)\,\lambda(dx) \;\; = \;\; \int_M f(z)T_{\delta,y,-s}g(z)|J\phi_\delta(-s,\cdot,y)|(z)\,\lambda(dz)$$

where dx and dz denote Lebesgue measure on M and $J\phi_\delta(-s,\cdot,y)(z)$ is the determinant at z of the Jacobian matrix of $\phi_\delta(-s,\cdot,y)$.

Proof

Just make the change of variable $z = \phi_\delta(s,x,y)$ in the first integral, which yields $x = \phi_\delta(-s,z,y)$. ■

A consequence of this lemma is the next result, where $\mathcal{D}$ is the subspace of $\mathbb{L}^2(\lambda)$ consisting of functions whose second derivatives in the distributional sense belong to $\mathbb{L}^2(\lambda)$ (or equivalently to $\mathbb{L}^2(\mu_{\beta,\delta})$ for any $\beta \geqslant 0$ and $\delta > 0$).

Lemma 3 *For $\alpha > 0$, $\beta \geqslant 0$ and $\delta > 0$, the domain of the maximal extension of $L_{\alpha,\beta,\delta}$ on $\mathbb{L}^2(\mu_{\beta,\delta})$ is $\mathcal{D}$. Furthermore the domain of its dual operator $L^*_{\alpha,\beta,\delta}$ in $\mathbb{L}^2(\mu_{\beta,\delta})$ is also $\mathcal{D}$ and we have for any $f \in \mathcal{D}$,*

$$\begin{aligned} & L^*_{\alpha,\beta,\delta}f \\ = \;\; & \frac{1}{2}\exp(\beta U_\delta)\Delta[\exp(-\beta U_\delta)f] + \frac{\exp(\beta U_\delta)}{\alpha}\int T_{\delta,y,-\alpha\beta}[\exp(-\beta U_\delta)f]|J\phi_\delta(-\alpha\beta,\cdot,y)|\,\nu(dy) - \frac{f}{\alpha} \end{aligned}$$

Proof

With the previous definitions, we can write for any $\alpha > 0$, $\beta \geqslant 0$ and $\delta > 0$,

$$L_{\alpha,\beta,\delta} \;\; = \;\; \frac{1}{2}\Delta + \frac{1}{\alpha}\int T_{\delta,y,\alpha\beta}\,\nu(dy) - \frac{I}{\alpha}$$

where I is the identity operator. Note furthermore that the identity operator is bounded from $\mathbb{L}^2(\lambda)$ to $\mathbb{L}^2(\mu_{\beta,\delta})$ and conversely. Thus to get the first assertion, it is sufficient to show that $\int T_{\delta,y,\alpha\beta}\,\nu(dy)$ is bounded from $\mathbb{L}^2(\lambda)$ to itself, or even only that $\|T_{\delta,y,\alpha\beta}\|_{\mathbb{L}^2(\lambda)\circlearrowleft}$ is uniformly bounded in $y \in M$. To see that this is true, consider a bounded and measurable function f and assume that $\alpha\beta \geqslant 0$. Since $(T_{\alpha\beta}f)^2 = T_{\alpha\beta}f^2$, we can apply Lemma 2 with $s = \alpha\beta$, $g = \mathbb{1}$ and f replaced by f^2 to get that

$$\begin{aligned} \int (T_{\delta,y,\alpha\beta}f)^2(x)\,\lambda(dx) \;\; &= \;\; \int f^2(z)T_{\delta,y,-\alpha\beta}\mathbb{1}|J\phi_\delta(-\alpha\beta,\cdot,y)|(z)\,\lambda(dz) \\ &\leqslant \;\; J_{\delta,\infty}\int f^2\,d\lambda \end{aligned}$$

with $J_{\delta,\infty} = \sup_{z,y \in M} |J\phi_\delta(-\alpha\beta,\cdot,y)|(z)$. This quantity is finite, since $\kappa_\delta(\cdot,\cdot)$ belongs to the class $\mathcal{C}^{\infty,0}$ of functions which are smooth in the first variables and whose all derivatives with respect to the first variable are continuous in the two variables, due to its definition by convolution with a smooth kernel. Next to see that for any $f,g\in\mathcal{C}^2(M)$,

$$\int gL_{\alpha,\beta,\delta}f\,d\mu_{\beta,\delta} \quad=\quad \int fL^*_{\alpha,\beta,\delta}g\,d\mu_{\beta,\delta} \tag{13}$$

where $L^*_{\alpha,\beta,\delta}$ is the operator defined in the statement of the lemma, we note that, on one hand,

$$\begin{aligned}\int g\Delta f\,d\mu_{\beta,\delta} &= Z^{-1}_{\beta,\delta}\int \exp(-\beta U_\delta)g\Delta f\,d\lambda\\ &= \int f\exp(\beta U_\delta)\Delta[\exp(-\beta U_\delta)g]\,d\mu_{\beta,\delta}\end{aligned}$$

and on the other hand, for any $y\in M$,

$$\begin{aligned}\int gT_{\delta,y,\alpha\beta}f\,d\mu_{\beta,\delta} &= Z^{-1}_{\beta,\delta}\int \exp(-\beta U_\delta)gT_{\delta,y,\alpha\beta}f\,d\lambda\\ &= Z^{-1}_{\beta,\delta}\int_M fT_{\delta,y,-\alpha\beta}\left(\exp(-\beta U_\delta)g\right)|J\phi_\delta(-\alpha\beta,\cdot,y)|(x)\,\lambda(dx)\end{aligned}$$

by Lemma 2. After an additional integration with respect to $\nu(dy)$, (13) follows without difficulty. To conclude, it is sufficient to see that for any $f\in\mathbb{L}^2(\mu_{\beta,\delta})$, $L^*_{\alpha,\beta,\delta}f\in\mathbb{L}^2(\mu_{\beta,\delta})$ (where $L^*_{\alpha,\beta,\delta}f$ is first interpreted as a distribution) if and only if $f\in\mathcal{D}$. This is done by adapting the arguments given in the first part of the proof, in particular we get that

$$\left\|\frac{\exp(\beta U_\delta)}{\alpha}\int T_{\delta,y,-\alpha\beta}[\exp(-\beta U_\delta)\cdot]|J\phi_\delta(-\alpha\beta,\cdot,y)|\,\nu(dy)\right\|^2_{\mathbb{L}^2(\lambda)\circlearrowleft} \quad\leqslant\quad \frac{J^2_{\delta,\infty}\exp(2\beta\mathrm{osc}(U_\delta))}{\alpha^2}$$

■

with $\mathrm{osc}(U_\delta):=\sup_{x,y\in M}(U_\delta(y)-U_\delta(x))$. For any $\alpha>0$ and $\beta\geqslant 0$, denote $\eta=\alpha\beta$. As a consequence of the previous lemma, we get that for any $x\in M$,

$$\begin{aligned}L^*_{\alpha,\beta,\delta}\mathbb{1}(x) &= \frac{1}{2}\exp(\beta U_\delta(x))\Delta\exp(-\beta U_\delta(x))-\frac{1}{\alpha}\\ &\quad+\frac{\exp(\beta U_\delta(x))}{\alpha}\int T_{\delta,y,-\eta}[\exp(-\beta U_\delta)](x)|J\phi_\delta(-\eta,\cdot,y)|(x)\,\nu(dy)\\ &= \frac{\beta^2}{2}(|\nabla U_\delta|(x))^2-\frac{\beta}{2}\Delta U_\delta(x)-\frac{1}{\alpha}\\ &\quad+\frac{1}{\alpha}\int_M\exp(\beta[U_\delta(x)-U_\delta(\phi_\delta(-\eta,x,y))])|J\phi_\delta(-\eta,\cdot,y)|(x)\,\nu(dy)\end{aligned} \tag{14}$$

It appears that $L^*_{\alpha,\beta,\delta}\mathbb{1}$ is continuous. The next result evaluates the uniform norm of this function.

Proposition 4 *There exists a constant $C>0$, depending on M and $\|\kappa\|_\infty$, such that for any $\beta\geqslant 1$, $\delta\in(0,1]$ and $\alpha\in(0,\delta^2/(2\beta^2))$ we have*

$$\|L^*_{\alpha,\beta,\delta}\mathbb{1}\|_\infty \quad\leqslant\quad C\alpha\beta^4\delta^{-4}$$

Proof

In view of the expression of $L^*_{\alpha,\beta,\delta}\mathbb{1}(x)$ given before the statement of the proposition, we want to estimate for any fixed $x\in M$, the quantity

$$\int_M\exp(\beta[U_\delta(x)-U_\delta(\phi_\delta(-\eta,x,y))])|J\phi_\delta(-\eta,\cdot,y)|(x)\,\nu(dy).$$

Consider the function

$$\psi(s) = U_\delta(x) - U_\delta(\phi_\delta(s,x,y)).$$

It has derivative

$$\psi'(s) = \frac{1}{2}\langle \nabla U_\delta, \nabla \kappa_\delta(\cdot,y)\rangle(\phi_\delta(s,x,y))$$

and second derivative

$$\psi''(s) = -\frac{1}{4}\text{Hess}\, U_\delta\left(\nabla\kappa_\delta(\cdot,y)(\phi_\delta(s,x,y)), \nabla\kappa_\delta(\cdot,y)(\phi_\delta(s,x,y))\right) - \frac{1}{4}\left\langle \nabla U_\delta, \nabla_{\nabla\kappa_\delta(\cdot,y)(\phi_\delta(s,x,y))}\nabla\kappa_\delta(\cdot,y)\right\rangle.$$

For any $\eta = \alpha\beta$, there exists $s \in [0,\eta]$ such that

$$\psi(-\eta) = \psi(0) - \eta\psi'(0) + \frac{\eta^2}{2}\psi''(-s).$$

This yields

$$\begin{aligned}\beta\left(U_\delta(x) - U_\delta(\phi(-\eta,x,y))\right) &= \frac{-\beta\eta}{2}\langle \nabla U_\delta, \nabla\kappa_\delta(\cdot,y)\rangle(x)\\ &-\frac{\beta\eta^2}{8}\left(\text{Hess}\, U_\delta\left(\nabla\kappa_\delta(\cdot,y)(\phi_\delta(-s,x,y)), \nabla\kappa(\cdot,y)(\phi_\delta(-s,x,y))\right) + \left\langle \nabla U_\delta, \nabla_{\nabla\kappa(\cdot,y)(\phi_\delta(-s,x,y))}\nabla\kappa(\cdot,y)\right\rangle\right)\end{aligned}$$

Observe that for any $a, b \in \mathbb{R}$, we can find $u, v \in (0,1)$ such that

$$\exp(a+b) = (1 + a + a^2\exp(ua)/2)(1 + b\exp(vb)). \tag{15}$$

Apply this equality with

$$a = \frac{-\beta\eta}{2}\langle \nabla U_\delta, \nabla\kappa_\delta(\cdot,y)\rangle(x)$$

and

$$b = -\frac{\beta\eta^2}{8}\left(\text{Hess}\, U_\delta\left(\nabla\kappa_\delta(\cdot,y)(\phi_\delta(s,x,y)), \nabla\kappa_\delta(\cdot,y)(\phi_\delta(-s,x,y))\right) + \left\langle \nabla U_\delta, \nabla_{\nabla\kappa_\delta(\cdot,y)(\phi_\delta(-s,x,y))}\nabla\kappa_\delta(\cdot,y)\right\rangle\right).$$

Using the bounds

$$\forall\ \delta > 0, \qquad \|\nabla \ln p(\delta,\cdot,y)(x)\| \leqslant \frac{C'}{\delta} \quad \text{and} \quad \|\nabla d \ln p(\delta,\cdot,y)(x)\| \leqslant \frac{C'}{\delta^2} \tag{16}$$

for some $C' > 0$ where ∇d denotes the Hessian map (see e.g. [17]), writing

$$\nabla\kappa_\delta(\cdot,y)(x) = \int_M \nabla \ln p(\delta,\cdot,z)p(\delta,x,z)\kappa(z,y)\,\lambda(dz),$$

we get

$$\forall\ \delta > 0, \qquad \|\nabla\kappa_\delta(\cdot,y)(x)\| \leqslant \frac{C}{\delta} \quad \text{and} \quad \|\nabla d\kappa_\delta(\cdot,y)(x)\| \leqslant \frac{C}{\delta^2} \tag{17}$$

with $C = 2C'\|\kappa\|_\infty$, together with

$$\|\nabla U_\delta(x)\| \leqslant \frac{C}{\delta} \quad \text{and} \quad \|\nabla d U_\delta(x)\| \leqslant \frac{C}{\delta^2}. \tag{18}$$

It follows that $|a| = \mathcal{O}(\alpha\beta^2\delta^{-2})$ and $|b| = \mathcal{O}(\alpha^2\beta^3\delta^{-4})$, so in conjunction with the assumption $\alpha\beta^2\delta^{-2} \leqslant 1/2$, we can write with (15) that

$$\exp(\beta[U_\delta(x) - U_\delta(\phi_\delta(-\eta,x,y))]) \;=\; 1 - \frac{\beta\eta}{2}\langle \nabla U_\delta, \nabla\kappa_\delta(\cdot,y)\rangle(x) + \mathcal{O}(\alpha^2\beta^4\delta^{-4}). \tag{19}$$

Integrating this expression, we get that

$$\begin{aligned}
&\int_M \exp(\beta[U_\delta(x) - U_\delta(\phi(-\eta, x, y))])|J\phi_\delta(-\eta, \cdot, y)|(x)\,\nu(dy) \\
&\quad = \int_M |J\phi_\delta(-\eta, \cdot, y)|(x)\,\nu(dy) \\
&\quad - \frac{\beta\eta}{2}\int_M \langle \nabla U_\delta, \nabla\kappa_\delta(\cdot, y)\rangle(x)|J\phi_\delta(-\eta, \cdot, y)|(x)\,\nu(dy) + \mathcal{O}(\alpha^2\beta^4\delta^{-4})
\end{aligned}$$

where we used the fact that $|J\phi_\delta(-\eta, \cdot, y)|(x)$ is uniformly bounded (see (20) below). We can now return to (14) and we obtain that for any $x \in M$,

$$\begin{aligned}
L^*_{\alpha,\beta,\delta}\mathbb{1}(x) &= \frac{\beta^2}{2}|\nabla U_\delta(x)|^2 - \frac{\beta}{2}\Delta U_\delta(x) + \frac{1}{\alpha}\int_M (|J\phi_\delta(-\eta, \cdot, y)|(x) - 1)\ \nu(dy) \\
&\quad - \frac{\beta^2}{2}\int_M \langle \nabla U_\delta, \nabla\kappa(\cdot, y)\rangle(x)|J\phi_\delta(-\eta, \cdot, y)|(x)\,\nu(dy) + \mathcal{O}(\alpha\beta^4\delta^{-4}).
\end{aligned}$$

Note that for $\eta/\delta \geqslant 0$ small enough (up to a universal factor, less than the injection radius of M), we have by Taylor formula with remainder in integral form applied to $\log_x \circ\phi_\delta$:

$$\phi_\delta(-\eta, x, y) = \exp_x\left(\frac{\eta}{2}\nabla\kappa_\delta(\cdot, y)(x) + \eta^2\int_0^1 (\log_x \circ\phi_\delta)''\,(-s\eta, x, y)(1-s)\,ds)\right)$$

where $\log_x$ is the inverse function of $\exp_x$ and $(\log_x \circ\phi_\delta)''$ is the second derivative in the first variable. From this equality, in conjunction with (5) and (17), we get

$$|J\phi_\delta(-\eta, \cdot, y)|(x) = 1 + \frac{\eta}{2}\Delta\kappa_\delta(\cdot, y)(x) + \mathcal{O}(\eta^2\delta^{-4}) \tag{20}$$

first for $\alpha\beta/\delta^2$ small enough and next by a compactness argument for all α, β, δ in the range described in the statement of Proposition 4. It also appears that $|J\phi_\delta(-\eta, \cdot, y)|(x)$ is uniformly bounded when $\alpha\beta^2\delta^{-2} \leqslant \frac{1}{2}$. This yields

$$\begin{aligned}
\frac{1}{\alpha}\int_M (|J\phi_\delta(-\eta, \cdot, y)|(x) - 1)\ \nu(dy) &= \frac{\beta}{2}\int_M \Delta\kappa_\delta(\cdot, y)(x)\,\nu(dy) + \mathcal{O}(\alpha\beta^2\delta^{-4}) \\
&= \frac{\beta}{2}\Delta U_\delta(x) + \mathcal{O}(\alpha\beta^2\delta^{-4}).
\end{aligned}$$

Notice the first term in the right cancels with the second in the right of (14). We also have

$$\begin{aligned}
&-\frac{\beta^2}{2}\int_M \langle \nabla U_\delta, \nabla\kappa_\delta(\cdot, y)\rangle(x)|J\phi_\delta(-\eta, \cdot, y)|(x)\,\nu(dy) \\
&= -\frac{\beta^2}{2}\left\langle \nabla U_\delta, \int_M \nabla\kappa_\delta(\cdot, y)(x)\,\nu(dy)\right\rangle + \mathcal{O}(\alpha\beta^3\delta^{-4}) \\
&= \frac{-\beta^2}{2}|\nabla U_\delta(x)|^2 + \mathcal{O}(\alpha\beta^3\delta^{-4}).
\end{aligned}$$

Here the first term in the right cancels with the first term in the right of (14). The bound announced in the lemma follows at once.

■

In particular, under the hypotheses of the previous proposition we get

$$\sqrt{\mu_{\beta,\delta}[(L^*_{\alpha,\beta,\delta}\mathbb{1})^2]} \leqslant C\alpha\beta^4\delta^{-4} \tag{21}$$

The l.h.s. will be used in the next section, when $\alpha\beta^4\delta^{-4}$ is small, as a discrepancy for the fact that $\mu_{\beta,\delta}$ is not necessarily an invariant measure for $L_{\alpha,\beta,\delta}$.

3 Proof of Theorem 1

This is the main part of the paper: we are going to prove Theorem 1 by the investigation of the evolution of a $\mathbb{L}^2$ type functional.

On M consider the algorithm $X \coloneqq (X_t)_{t\geqslant 0}$ described in the introduction. For the time being, the schemes $\alpha : \mathbb{R}_+ \to \mathbb{R}_+^*$, $\beta : \mathbb{R}_+ \to \mathbb{R}_+$ and $\delta : \mathbb{R}_+ \to \mathbb{R}_+^*$ are assumed to be continuously differentiable. Only later on, in Proposition 9, we will present the conditions insuring the wanted convergence (9). On the initial distribution m_0, the last ingredient necessary to specify the law of X, no hypothesis is made. We also denote the law of X_t by m_t, for any $t > 0$. We have that m_t admits a $\mathcal{C}^1$ density with respect to λ, which is equally written m_t (for a proof we refer to the appendix of [3]). As it was mentioned in the previous section, we want to compare these temporal marginal laws with the corresponding instantaneous Gibbs measures, which were defined in (10) with respect to the potentials U_δ given in (2). A convenient way to quantify this discrepancy is to consider the variance of the density of m_t with respect to μ_{β_t,δ_t} under the probability measure μ_{β_t,δ_t}:

$$\forall\ t > 0, \qquad I_t \ \coloneqq\ \int \left(\frac{m_t}{\mu_{\beta_t,\delta_t}} - 1\right)^2 d\mu_{\beta_t,\delta_t} \tag{22}$$

(here and in the sequel we use the same notation for a probability measure with density, and for its density with respect to Lebesgue measure). Our goal here is to derive a differential inequality satisfied by this quantity, which implies its convergence to zero under appropriate conditions on the schemes α and β. More precisely, our purpose is to obtain:

Proposition 5 *There exists two constants c_1, $c_2 > 0$ such that for any $t > 0$ with $\beta_t \geqslant 1$ and $\alpha_t\beta_t^2\delta_t^{-2} \leqslant 1/2$, we have*

$$\begin{aligned} I'_t \ &\leqslant\ -c_1\left[(\beta_t\delta_t^{-1})^{2-5m}\exp(-b(U)\beta_t) - \alpha_t\beta_t^3\delta_t^{-3} - \left|\beta'_t\right| - \beta_t\delta_t^{-2}\left|\delta'_t\right|\right] I_t \\ &+\ c_2\left[\alpha_t\beta_t^4\delta_t^{-4} + \left|\beta'_t\right| + \beta_t\delta_t^{-2}\left|\delta'_t\right|\right]\sqrt{I_t} \end{aligned}$$

where $b(U)$ was defined in (6).

Let us stress the fact that even if the inequality above is true under the stated assumption, it will be usable only in a situation where all terms in the right are neglectable compared to the first one.

Proof

At least formally, there is no difficulty to differentiate the quantity I_t with respect to the time $t > 0$. For a rigorous justification of the following computations, we refer to the appendix of [3], where the regularity of the temporal marginal laws in presence of jumps is discussed in detail (it is written in the situation considered there of the circle but can be extended to compact manifolds). Thus we get at any time $t > 0$,

$$\begin{aligned} I'_t \ &=\ 2\int \left(\frac{m_t}{\mu_{\beta_t,\delta_t}} - 1\right)\frac{\partial_t m_t}{\mu_{\beta_t,\delta_t}}\, d\mu_{\beta_t,\delta_t} - 2\int \left(\frac{m_t}{\mu_{\beta_t,\delta_t}} - 1\right)\frac{m_t}{\mu_{\beta_t,\delta_t}}\partial_t \ln(\mu_{\beta_t,\delta_t})\, d\mu_{\beta_t,\delta_t} \\ &\quad + \int \left(\frac{m_t}{\mu_{\beta_t,\delta_t}} - 1\right)^2 \partial_t \ln(\mu_{\beta_t,\delta_t})\, d\mu_{\beta_t,\delta_t} \\ &=\ 2\int \left(\frac{m_t}{\mu_{\beta_t,\delta_t}} - 1\right)\partial_t m_t\, d\lambda - \int \left(\frac{m_t}{\mu_{\beta_t,\delta_t}} - 1\right)^2 \partial_t \ln(\mu_{\beta_t,\delta_t})\, d\mu_{\beta_t,\delta_t} \\ &-\ 2\int \left(\frac{m_t}{\mu_{\beta_t,\delta_t}} - 1\right)\partial_t \ln(\mu_{\beta_t,\delta_t})\, d\mu_{\beta_t,\delta_t} \\ &\leqslant\ 2\int \left(\frac{m_t}{\mu_{\beta_t,\delta_t}} - 1\right)\partial_t m_t\, d\lambda + \left\|\partial_t \ln(\mu_{\beta_t,\delta_t})\right\|_\infty \left(\int \left(\frac{m_t}{\mu_{\beta_t,\delta_t}} - 1\right)^2 d\mu_{\beta_t,\delta_t} + 2\int \left|\frac{m_t}{\mu_{\beta_t,\delta_t}} - 1\right| d\mu_{\beta_t,\delta_t}\right) \\ &\leqslant\ 2\int \left(\frac{m_t}{\mu_{\beta_t,\delta_t}} - 1\right)\partial_t m_t\, d\lambda + \left\|\partial_t \ln(\mu_{\beta_t,\delta_t})\right\|_\infty \left(I_t + 2\sqrt{I_t}\right) \end{aligned}$$

where we used the Cauchy-Schwarz inequality. The last term is easy to deal with:

Lemma 6 *There exists $C_0 \geqslant 0$, depending on κ, such that for any $t \geqslant 0$, we have*

$$\|\partial_t \ln(\mu_{\beta_t,\delta_t})\|_\infty \quad \leqslant \quad C_0 \left(|\beta'_t| + \beta_t \, |\delta'_t| \, \delta_t^{-2}\right).$$

Proof
Since for any $t \geqslant 0$ we have

$$\forall\, x \in M, \qquad \ln(\mu_{\beta_t,\delta_t})(x) \quad = \quad -\beta_t U_{\delta_t}(x) - \ln\left(\int \exp(-\beta_t U_{\delta_t}(y))\, \lambda(dy)\right)$$

it appears that $\forall\, x \in M$,

$$\begin{aligned}
&\partial_t \ln(\mu_{\beta_t,\delta_t})(x) \\
&\quad = \quad \beta'_t \int U_{\delta_t}(y) - U_{\delta_t}(x)\, \mu_{\beta_t,\delta_t}(dy) \\
&\qquad\quad +\beta_t\delta'_t \int\!\!\int\!\!\int (p(\delta_t, y, z)\partial_\delta \ln p(\delta_t, y, z) - p(\delta_t, x, z)\partial_\delta \ln p(\delta_t, x, z))\, \kappa(z, v)\nu(dv)\mu_{\beta_t,\delta_t}(dy)\lambda(dz)
\end{aligned}$$

so that

$$\|\partial_t \ln(\mu_{\beta_t})\|_\infty \quad \leqslant \quad \mathrm{osc}(U_{\delta_t})\, |\beta'_t| + 2\beta_t|\delta'_t|\|\partial_\delta \ln p\|_\infty \cdot \|\kappa\|_\infty.$$

Clearly $\mathrm{osc}(U_{\delta_t}) \leqslant 2\|\kappa\|_\infty$. To finish the proof we are left to use the bound

$$|\partial_\delta \ln p(\delta, x, y)| \leqslant \frac{C''}{\delta^2} \tag{23}$$

for some $C'' > 0$ (see e.g. [17] which yields estimates for spatial derivatives, combined with the equality $\partial_\delta \ln p(\delta, x, y) = \frac{\Delta_x p(\delta,x,y)}{2p(\delta,x,y)} = \frac{1}{2}\Delta_x \ln p(\delta, x, y) + \frac{1}{2}\|\nabla_x \ln p(\delta, x, y)\|^2$).

■

Denote for any $t > 0$, $f_t \coloneqq m_t/\mu_{\beta_t,\delta_t}$. Up to the classical use of a mollifier (as it is detailed after Remark 39 in [3], arguments which can be extended to the present case), we can assume that f_t is is $\mathcal{C}^2$. So we have by the martingale problem satisfied by the law of X, that

$$\begin{aligned}
\int \left(\frac{m_t}{\mu_{\beta_t,\delta_t}} - 1\right) \partial_t m_t \, d\lambda \quad &= \quad \int L_{\alpha_t,\beta_t,\delta_t}\left[f_t - 1\right]\, dm_t \\
&= \quad \int L_{\alpha_t,\beta_t,\delta_t}\left[f_t - 1\right] f_t \, d\mu_{\beta_t,\delta_t}
\end{aligned}$$

where $L_{\alpha_t,\beta_t,\delta_t}$, described in the previous section, is the instantaneous generator at time $t \geqslant 0$ of X. The interest of the estimate (21) comes from the decomposition of the previous term into

$$\begin{aligned}
&\int L_{\alpha_t,\beta_t,\delta_t}\left[f_t - 1\right](f_t - 1)\, d\mu_{\beta_t,\delta_t} + \int L_{\alpha_t,\beta_t,\delta_t}\left[f_t - 1\right]\, d\mu_{\beta_t,\delta_t} \\
&\quad = \quad \int L_{\alpha_t,\beta_t,\delta_t}\left[f_t - 1\right](f_t - 1)\, d\mu_{\beta_t,\delta_t} + \int (f_t - 1)L^*_{\alpha_t,\beta_t,\delta_t}[\mathbb{1}]\, d\mu_{\beta_t,\delta_t} \\
&\quad \leqslant \quad \int L_{\alpha_t,\beta_t,\delta_t}\left[f_t - 1\right](f_t - 1)\, d\mu_{\beta_t,\delta_t} + \sqrt{I_t}\sqrt{\mu_{\beta_t,\delta_t}[(L^*_{\alpha_t,\beta_t,\delta_t}[\mathbb{1}])^2]}
\end{aligned}$$

It follows from these bounds that to prove Proposition 5, it remains to treat the first term in the above r.h.s. A first step is:

Lemma 7 *There exists a constant $c_3 > 0$, such that for any $\alpha > 0$ and $\beta \geqslant 1$ such that $\alpha\beta^2\delta^{-2} \leqslant 1/2$, we have, for any $f \in \mathcal{C}^2(M)$,*

$$\int L_{\alpha,\beta,\delta}[f-1](f-1)\, d\mu_{\beta,\delta} \quad \leqslant \quad -\left(\frac{1}{2} - c_3\alpha\beta^3\delta^{-3}\right)\int (|\nabla f|)^2\, d\mu_{\beta,\delta} + c_3\alpha\beta^3\delta^{-3}\int (f-1)^2\, d\mu_{\beta,\delta}$$

Later this inequality will be used only when $c_3\alpha\beta^3\delta^{-3} < 1/2$.

Proof

For any $\alpha > 0$, $\beta \geqslant 0$ and $\delta > 0$, we begin by decomposing the generator $L_{\alpha,\beta,\delta}$ into

$$L_{\alpha,\beta,\delta} \quad = \quad L_{\beta,\delta} + R_{\alpha,\beta,\delta} \tag{24}$$

where

$$L_{\beta,\delta}\cdot \quad \coloneqq \quad \frac{1}{2}(\triangle \cdot - \beta\langle \nabla U_\delta, \nabla \cdot\rangle) \tag{25}$$

and where $R_{\alpha,\beta,\delta}$ is the remaining operator. An immediate integration by parts leads to

$$\begin{aligned} \int L_{\beta,\delta}[f-1](f-1)\, d\mu_{\beta,\delta} &= -\frac{1}{2}\int |\nabla(f-1)|^2\, d\mu_{\beta,\delta} \\ &= -\frac{1}{2}\int |\nabla f|^2\, d\mu_{\beta,\delta} \end{aligned} \tag{26}$$

Thus our main task is to find a constant $c_3 > 0$, such that for any $\alpha > 0$, $\beta \geqslant 1$ and $\delta > 0$ with $\alpha\beta^2\delta^{-2} \leqslant 1/2$, we have, for any $f \in \mathcal{C}^2(M)$,

$$\left|\int R_{\alpha,\beta,\delta}[f-1](f-1)\, d\mu_{\beta,\delta}\right| \quad \leqslant \quad c_3\alpha\beta^3\delta^{-3}\left(\int |\nabla f|^2\, d\mu_{\beta,\delta} + \int (f-1)^2\, d\mu_{\beta,\delta}\right) \tag{27}$$

By definition, we have for any $f \in \mathcal{C}^2(M)$ (but what follows is valid for $f \in \mathcal{C}^1(M)$),

$$\forall\, x \in M, \qquad R_{\alpha,\beta,\delta}[f](x) \quad = \quad \frac{1}{\alpha}\int f(\phi_\delta(\alpha\beta, x, y)) - f(x)\,\nu(dy) + \frac{\beta}{2}\langle \nabla U_\delta(x), \nabla f(x)\rangle$$

To evaluate this quantity, on one hand, recall that we have for any $x \in M$,

$$\nabla U_\delta(x) \quad = \quad \int_M \nabla\kappa_\delta(\cdot, y)(x)\,\nu(dy)$$

and on the other hand, write that for any $x, y \in M$,

$$f(\phi_\delta(\alpha\beta, x, y)) - f(x) \quad = \quad \alpha\beta\int_0^1 \left\langle \nabla f(\phi_\delta(\alpha\beta u, x, y)), -\frac{1}{2}\nabla\kappa_\delta(\cdot, y)(\phi_\delta(\alpha\beta u, x, y))\right\rangle du$$

It follows that

$$\begin{aligned} &\int R_{\alpha,\beta,\delta}[f-1](f-1)\, d\mu_{\beta,\delta} \\ &= \quad \frac{\beta}{2}\int_0^1 du \int \nu(dy)\int \mu_{\beta,\delta}(dx)\left(\langle \nabla f, \nabla\kappa_\delta(\cdot, y)(x)\rangle - \langle \nabla f, \nabla\kappa_\delta(\cdot, y)(\phi_\delta(\alpha\beta u, x, y))\rangle\right)(f(x)-1) \\ &= \quad \frac{\beta}{2}\int_0^1 du \int \nu(dy)\int \lambda(dx)\langle \nabla f, \nabla\kappa_\delta(\cdot, y)(x)\rangle\Big[(f(x)-1)\mu_{\beta,\delta}(x) \\ &- \quad (f(\phi_\delta(-\alpha\beta u, x, y)) - 1)\,\mu_{\beta,\delta}(\phi_\delta(-\alpha\beta u, x, y))|J\phi_\delta(-\alpha\beta u, \cdot, y)|(x)\Big] \end{aligned}$$

where we used the change of variable $z \mapsto \phi_\delta(-\alpha\beta u, z, y)$ for the second term in the right. So

$$\begin{aligned}&\int R_{\alpha,\beta,\delta}\,[f-1]\,(f-1)\,d\mu_{\beta,\delta}\\ &\quad= \quad \frac{\beta}{2}\int_0^1 du\int \nu(dy)\int \lambda(dx)\langle \nabla f, \nabla\kappa_\delta(\cdot,y)(x)\rangle I_\delta(\alpha\beta u, x, y)\end{aligned}$$

where

$$\begin{aligned}I_\delta(s,x,y) &= |J\phi_\delta(-s,\cdot,y)|(x)\left\{(f(x)-1)\mu_{\beta,\delta}(x) - (f(\phi_\delta(-s,x,y))-1)\mu_{\beta,\delta}(\phi_\delta(-s,x,y))\right\}\\ &\quad + (f(x)-1)\mu_{\beta,\delta}(x)(1-|J\phi_\delta(-s,\cdot,y)|(x)).\end{aligned}$$

Write

$$\int R_{\alpha,\beta,\delta}\,[f-1]\,(f-1)\,d\mu_{\beta,\delta} = J_1 + J_2$$

where J_1 is the integral containing the term $(1-|J\phi_\delta(-s,\cdot,y)|(x))$. From the validity of

$$\phi_\delta(-s,x,y) = \exp_x\left(s\int_0^1 (\log_x\circ\phi_\delta)'(-vs,x,y)\,dv\right)$$

for s small enough (here again we use Taylor formula with remainder in integral form), we get that for any s,

$$|1-|J\phi_\delta(-s,\cdot,y)|(x)| \leqslant c_4 s\delta^{-2}$$

for some $c_4>0$, which yields

$$\begin{aligned}J_1 &\leqslant \frac{1}{2}c_4\alpha\beta^2\delta^{-2}\int_0^1 du\int\nu(dy)\int \mu_{\beta,\delta}(dx)\|\nabla_1\kappa_\delta\|_\infty|\nabla f|(x)\cdot|f(x)-1|\\ &\leqslant \frac{1}{2}c_4\alpha\beta^2\delta^{-2}\|\nabla_1\kappa_\delta\|_\infty\|\nabla f\|_{\mathbb{L}^2(\mu_\beta)}\sqrt{\int (f-1)^2\,d\mu_\beta}\\ &\leqslant \frac{1}{4}c_4\alpha\beta^2\delta^{-3}\left(\|\nabla f\|^2_{\mathbb{L}^2(\mu_{\beta,\delta})} + \int (f-1)^2\,d\mu_{\beta,\delta}\right)\\ &= c_5\alpha\beta^2\delta^{-3}\left(\|\nabla f\|^2_{\mathbb{L}^2(\mu_{\beta,\delta})} + \int (f-1)^2\,d\mu_{\beta,\delta}\right)\end{aligned}$$

with $c_5 = \frac{1}{4}c_4$, using again (17). Moreover, we have

$$\begin{aligned}J_2 &= \frac{\beta}{2}\int_0^1 du\int\nu(dy)\int\sqrt{\mu_{\beta,\delta}(x)}\lambda(dx)\langle\nabla f,\nabla\kappa_\delta(\cdot,y)\rangle(x)|J\phi_\delta(-\alpha\beta u,\cdot,y)|(x)(-\alpha\beta u)\int_0^1 dv\\ &\left(\langle\nabla f, -\frac{1}{2}\nabla\kappa_\delta(\cdot,y)\rangle(\phi_\delta(-\alpha\beta uv,x,y))\sqrt{\mu_{\beta,\delta}(\phi_\delta(-\alpha\beta uv,x,y))}\sqrt{\frac{\mu_{\beta,\delta}(\phi_\delta(-\alpha\beta uv,x,y))}{\mu_{\beta,\delta}(x)}}\right.\\ &\quad + f(\phi_\delta(-\alpha\beta uv,x,y))-1)\langle\nabla\ln\mu_{\beta,\delta}, -\frac{1}{2}\nabla\kappa_\delta(\cdot,y)\rangle(\phi_\delta(-\alpha\beta uv,x,y))\\ &\left.\sqrt{\mu_{\beta,\delta}(\phi_\delta(-\alpha\beta uv,x,y))}\sqrt{\frac{\mu_{\beta,\delta}(\phi_\delta(-\alpha\beta uv,x,y))}{\mu_{\beta,\delta}(x)}}\right).\end{aligned}$$

Recalling (20), notice that $|J\phi_\delta(-\alpha\beta u,\cdot,y)|(x)$ and $\sqrt{\frac{\mu_{\beta,\delta}(\phi_\delta(-\alpha\beta uv,x,y))}{\mu_{\beta,\delta}(x)}}$ are uniformly bounded, since $\alpha\beta^2\delta^{-2}\leqslant 1/2$. Exchanging the orders of integration to have the integral with respect to v

on the left, using Cauchy-Schwartz inequality for the integral in the right and making the change of variable $z = \phi_\delta(-\alpha\beta uv, x, y)$ we get

$$\begin{aligned} J_2 &\leqslant c_6\alpha\beta^2\|\nabla_1\kappa_\delta\|_\infty^2\|\nabla f\|_{\mathbb{L}^2(\mu_{\beta,\delta})}\left(\|\nabla f\|_{\mathbb{L}^2(\mu_{\beta,\delta})} + \beta\|\nabla_1\kappa_\delta\|_\infty\sqrt{\int (f-1)^2\,d\mu_{\beta,\delta}}\right) \\ &\leqslant c_6\alpha\beta^2\|\nabla_1\kappa_\delta\|_\infty^2\left[\left(1+\frac{1}{2}\beta\|\nabla_1\kappa_\delta\|_\infty\right)\|\nabla f\|^2_{\mathbb{L}^2(\mu_{\beta,\delta})} + \frac{1}{2}\beta\|\nabla_1\kappa_\delta\|_\infty\int (f-1)^2\,d\mu_{\beta,\delta}\right] \end{aligned} \tag{28}$$

where $c_6 \geqslant 0$ is a constant independent from α, β, δ. Up to a change of this constant, we obtain

$$J_2 \leqslant c_6\alpha\beta^3\delta^{-3}\left(\|\nabla f\|^2_{\mathbb{L}^2(\mu_{\beta,\delta})} + \int (f-1)^2\,d\mu_{\beta,\delta}\right) \tag{29}$$

where we used (17).

So putting together (26) with the bounds for J_1 and J_2 we get the wanted result.

■

To conclude the proof of Proposition 5, we must be able to compare, for any $\beta \geqslant 0$ and any $f \in \mathcal{C}^1(M)$, the energy $\mu_{\beta,\delta}[|\nabla f|^2]$ and the variance $\mathrm{Var}(f, \mu_{\beta,\delta})$. This task was already done by Holley, Kusuoka and Stroock [15], let us recall their result:

Proposition 8 *Let $\widetilde{U}$ be a $\mathcal{C}^1$ function on a compact Riemannian manifold M of dimension $m \geqslant 1$. Let $b(\widetilde{U}) \geqslant 0$ be the associated constant as in (6). For any $\beta \geqslant 0$, consider the Gibbs measure $\widetilde{\mu}_\beta$ given similarly to (10). Then there exists a constant $C_M > 0$, depending only on M, such that the following Poincaré inequalities are satisfied:*

$$\forall\ \beta \geqslant 0,\ \forall\ f \in \mathcal{C}^1(M), \qquad \mathrm{Var}(f, \widetilde{\mu}_\beta) \ \leqslant\ C_M[1 \vee (\beta\left\|\nabla\widetilde{U}\right\|_\infty)]^{5m-2}\exp(b(\widetilde{U})\beta)\widetilde{\mu}_\beta[|\nabla f|^2]$$

We can now come back to the study of the evolution of the quantity $I_t = \mathrm{Var}(f_t, \mu_{\beta_t,\delta_t})$, for $t > 0$. Indeed applying Lemma 7 and Proposition 8 with $\alpha = \alpha_t$, $\beta = \beta_t$, $\delta = \delta_t$ and $f = f_t$, we get at any time $t > 0$ such that $\beta_t \geqslant 1$, $\delta_t \in (0, 1]$ and $\alpha_t\beta_t^3\delta_t^{-3} \leqslant 1/(2c_3)$,

$$\begin{aligned} &\int L_{\alpha_t,\beta_t,\delta_t}[f_t - 1]\,(f_t - 1)\,d\mu_{\beta_t,\delta_t} \\ &\quad\leqslant\ -c_7(\beta_t\delta_t^{-1})^{2-5m}\exp(-b(U_{\delta_t})\beta_t)\left(1 - 2c_3\alpha_t\beta_t^3\delta_t^{-3}\right)I_t + c_3\alpha_t\beta_t^3\delta_t^{-3}I_t \\ &\quad\leqslant\ -\left[c_7(\beta_t\delta_t^{-1})^{2-5m}\exp(-b(U_{\delta_t})\beta_t) - c_8\alpha_t\beta_t^3\delta_t^{-3}\right]I_t \end{aligned}$$

for some constants $c_7, c_8 > 0$.
Taking into account Lemma 6, the computations preceding Lemma 7 and (21), one can find constants $c_1, c_2 > 0$ such that Proposition 5 is satisfied. This achieves the proof of Proposition 5.

■

This result leads immediately to conditions insuring the convergence toward 0 of the quantity I_t for large times $t > 0$:

Proposition 9 *Let $\alpha, \delta : \mathbb{R}_+ \to \mathbb{R}_+^*$ and $\beta : \mathbb{R}_+ \to \mathbb{R}_+$ be schemes as at the beginning of this section and assume:*

$$\begin{aligned} \lim_{t\to+\infty}\alpha_t &= 0 \\ \lim_{t\to+\infty}\beta_t &= +\infty \\ \lim_{t\to+\infty}\delta_t &= 0 \\ \int_0^{+\infty}(1 \vee (\beta_t\delta_t^{-1}))^{2-5m}\exp(-b(U_{\delta_t})\beta_t)\,dt &= +\infty \end{aligned}$$

and that for large times $t > 0$,

$$\max\{\alpha_t\beta_t^4\delta_t^{-4},\, |\beta_t'|\,,\, \beta_t\delta_t^{-2}\,|\delta_t'|\} \quad \ll \quad (\beta_t\delta_t^{-1})^{2-5m}\exp(-b(U_{\delta_t})\beta_t)$$

Then we are assured of

$$\lim_{t\to+\infty} I_t \quad = \quad 0$$

Proof

The differential equation of Proposition 5 can be rewritten under the form

$$F_t' \quad \leqslant \quad -\eta_t F_t + \epsilon_t \tag{30}$$

where for any $t > 0$,

$$\begin{aligned}
F_t &\coloneqq \sqrt{I_t}\\
\eta_t &\coloneqq c_1((\beta_t\delta_t^{-1})^{2-5m}\exp(-b(U_{\delta_t})\beta_t) - \alpha_t\beta_t^3\delta_t^{-3} - |\beta_t'| - \beta_t\,|\delta_t'|\,\delta_t^{-2})/2\\
\epsilon_t &\coloneqq c_2(\alpha_t\beta_t^4\delta_t^{-4} + |\beta_t'| + \beta_t\delta_t^{-2}\,|\delta_t'|)/2
\end{aligned}$$

The assumptions of the above proposition imply that for $t \geqslant 0$ large enough, $\beta_t \geqslant 1$ and $\alpha_t\beta_t^2\delta_t^{-2} \leqslant 1/2$ and $\alpha_t\beta_t^3\delta_t^{-3} \ll (\beta_t\delta_t^{-1})^{2-5m}\exp(-b(U_{\delta_t})\beta_t)$. This insures that there exists $T > 0$ such that (30) is satisfied for any $t \geqslant T$ (and also $F_T < +\infty$). We deduce that for any $t \geqslant T$,

$$F_t \quad \leqslant \quad F_T\exp\left(-\int_T^t \eta_s\,ds\right) + \int_T^t \epsilon_s\exp\left(-\int_s^t \eta_u\,du\right)\,ds \tag{31}$$

It appears that $\lim_{t\to+\infty} F_t = 0$ as soon as

$$\begin{aligned}
\int_T^{+\infty} \eta_s\,ds &= +\infty\\
\lim_{t\to+\infty} \epsilon_t/\eta_t &= 0
\end{aligned}$$

The above assumptions were chosen to insure these properties.

■

In particular, the schemes given in (8) satisfy the hypotheses of the previous proposition (notice that $b(U_\delta) \to b(U)$ as $\delta \to 0$, due to the uniform convergence of U_δ to U), so that under the conditions of Theorem 1, we get

$$\lim_{t\to+\infty} I_t \quad = \quad 0$$

Let us deduce (9) for any neighborhood $\mathcal{N}$ of the set $\mathcal{M}$ of the global minima of U. From Cauchy-Schwartz inequality we have for any $t > 0$,

$$\begin{aligned}
\|m_t - \mu_{\beta_t,\delta_t}\|_{\mathrm{tv}} &= \int |f_t - 1|\;\mu_{\beta_t,\delta_t}\\
&\leqslant \sqrt{I_t}
\end{aligned}$$

An equivalent definition of the total variation norm states that

$$\|m_t - \mu_{\beta_t}\|_{\mathrm{tv}} \quad = \quad 2\max_{A\in\mathcal{T}} |m_t(A) - \mu_{\beta_t,\delta_t}(A)|$$

where $\mathcal{T}$ is the Borelian σ-algebra of M. It follows that (9) reduces to

$$\lim_{\beta,\delta^{-1}\to+\infty} \mu_{\beta,\delta}(\mathcal{N}) \quad = \quad 1$$

for any neighborhood $\mathcal{N}$ of $\mathcal{M}$ and δ sufficiently small, a property which is immediate from the definition (10) of the Gibbs measures $\mu_{\beta,\delta}$ for $\beta \geqslant 0$ and $\delta > 0$.

Remark 10 Similarly to the approach presented for instance in [21, 23], we could have studied the evolution of $(E_t)_{t>0}$, which are the relative entropies of the time marginal laws with respect to the corresponding instantaneous Gibbs measures, namely

$$\forall\ t > 0, \qquad E_t \ \ \coloneqq \ \ \int \ln\left(\frac{m_t}{\mu_{\beta_t,\delta_t}}\right) dm_t$$

To get a differential inequality satisfied by these functionals, the spectral gap estimate of Holley, Kusuoka and Stroock [15] recalled in Proposition 8 must be replaced by the corresponding logarithmic Sobolev constant estimate.

□

References

[1] B. Afsari, R. Tron and R. Vidal, *On the convergence of gradient descent for finding the Riemannian center of mass* SIAM J. Control Optim. 51 (2013), no. 3, 2230-2260.

[2] M. Arnaudon and L. Miclo. Means in complete manifolds: uniqueness and approximation. ESAIM Probability and Statistics, Vol 18 (January 2014), pp. 185-206.

[3] M. Arnaudon and L. Miclo. A stochastic algorithm finding p-means on the circle. Available on `http://hal.archives-ouvertes.fr/hal-00781715`, 2013-01.

[4] M. Arnaudon, F. Nielsen, *Medians and means in Finsler geometry*, LMS Journal of Computation and Mathematics volume 15, pp. 23-37 (2012)

[5] M. Arnaudon,C. Dombry, A. Phan, A., L. Yang, *Stochastic algorithms for computing means of probability measures* Stoch. Proc. Appl. 122 (2012), pp. 1437-1455.

[6] M. Arnaudon, F. Nielsen, *On computing the Riemannian 1-Center*, Computational Geometry: Theory and Applications, 46 (2013), no. 1, 93–104.

[7] M. Bădoiu,K. L. Clarkson, 2003, *Smaller core-sets for balls,* Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, pp. 801–802.

[8] R. Bhattacharya and V. Patrangenaru, *Large sample theory of intrinsic and extrinsic sample means on manifolds (i)*, Annals of Statistics 31 (1), pp. 1–29 (2003)

[9] S. Bonnabel, *Stochastic Gradient Descent on Riemannian Manifolds* IEEE Transactions on Automatic Control, 58(9) (2013), 2217-2229.

[10] H. Cardot, P. Cénac, P.-A. Zitt, *Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm*, Bernoulli 19 (2013), no. 1, 18-43.

[11] B. Charlier, *Necessary and sufficient condition for the existence of a Fréchet mean on the circle*, ESAIM Probab. Stat. 17 (2013), 635-649.

[12] S.N. Ethier and T. G. Kurtz. *Markov processes.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.

[13] D. Groisser, *Newton's method, zeroes of vector fields, and the Riemannian center of mass* Adv. in Appl. Math. 33 (2004), no. 1, 95-135.

[14] D. Groisser, *On the convergence of some Procrustean averaging algorithms* Stochastics 77 (2005), no. 1, 31-60.

[15] R. A. Holley, S. Kusuoka, and D. W. Stroock. Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.*, 83(2):333–347, 1989.

[16] T. Hotz and S. Huckemann. Intrinsic Means on the Circle: Uniqueness, Locus and Asymptotics. *ArXiv e-prints*, August 2011.

[17] E.P. Hsu Estimates of derivatives of the heat kernel on a compact Riemannian manifold *Proceedings of the American Mathematical Society.* Vol. 127 (1999), n. 12, pp 3739–3744.

[18] C. R. Hwang. Laplace's method revisited: weak convergence of probability measures. *Ann. Probab.*, 8(6):1177–1182, 1980.

[19] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library.* North-Holland Publishing Co., Amsterdam, second edition, 1989.

[20] H. Le, *Estimation of Riemannian barycentres*, LMS J. Comput. Math. 7 (2004), pp. 193–200

[21] L. Miclo. Recuit simulé sans potentiel sur une variété riemannienne compacte. *Stochastics Stochastics Rep.*, 41(1-2):23–56, 1992.

[22] L. Miclo. Remarques sur l'ergodicité des algorithmes de recuit simulé sur un graphe. *Stochastic Process. Appl.*, 58(2):329–360, 1995.

[23] L. Miclo. Recuit simulé partiel. *Stochastic Process. Appl.*, 65(2):281–298, 1996.

[24] X. Pennec. Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *J. Math. Imaging Vision*, 25(1):127–154, 2006.

[25] K.T. Sturm, *Probability measures on metric spaces of nonpositive curvature*, Heat kernels and analysis on manifolds, graphs, and metric spaces (Paris, 2002), 357-390, Contemp. Math., 338, Amer. Math. Soc., Providence, RI, 2003.

[26] L.Yang, *Riemannian median and its estimation*, LMS Journal of Computation and Mathematics, Vol 13 (2010), pp 461–479.

† marc.arnaudon@math.u-bordeaux1.fr
Institut de Mathématiques de Bordeaux
Université de Bordeaux 1
351, Cours de la Libération
F33405 Talence Cedex, France

‡ miclo@math.univ-toulouse.fr
Institut de Mathématiques de Toulouse
Université Paul Sabatier
118, route de Narbonne
31062 Toulouse Cedex 9, France