ESAIM: PS 18 (2014) 185–206
DOI: 10.1051/ps/2013033

**ESAIM: Probability and Statistics**
www.esaim-ps.org

# MEANS IN COMPLETE MANIFOLDS: UNIQUENESS AND APPROXIMATION

Marc Arnaudon[1] and Laurent Miclo[2]

**Abstract.** Let $M$ be a complete Riemannian manifold, $N \in \mathbb{N}$ and $p \geq 1$. We prove that almost everywhere on $x = (x_1, \ldots, x_N) \in M^N$ for Lebesgue measure in $M^N$, the measure $\mu(x) = \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k}$ has a unique $p$–mean $e_p(x)$. As a consequence, if $X = (X_1, \ldots, X_N)$ is a $M^N$-valued random variable with absolutely continuous law, then almost surely $\mu(X(\omega))$ has a unique $p$–mean. In particular if $(X_n)_{n \geq 1}$ is an independent sample of an absolutely continuous law in $M$, then the process $e_{p,n}(\omega) = e_p(X_1(\omega), \ldots, X_n(\omega))$ is well-defined. Assume $M$ is compact and consider a probability measure $\nu$ in $M$. Using partial simulated annealing, we define a continuous semimartingale which converges in probability to the set of minimizers of the integral of distance at power $p$ with respect to $\nu$. When the set is a singleton, it converges to the $p$–mean.

**Mathematics Subject Classification.** 60D05, 58C35, 37A30, 53C21, 60J65.

Received July 13, 2012. Revised November 26, 2012.

## 1. Introduction

Finding the mean of the median or more generaly the $p$–mean $e_p$ of a probability measure in a manifold (the point which minimizes integral with respect to this measure of distance at power $p$) has numerous applications. There is not much to say for the mean in $\mathbb{R}^d$, almost the only case where there is a closed formula, and the most important case as the most useful estimator in statistics when the measure is uniform law on a sample. For medians in $\mathbb{R}^d$ the situation is more complicated. Uniqueness holds as soon as the support of the probability measure is not carried by a line. The first algorithm for computing $e_1$ is due to Weisfeld in [24]. As for the computation of $e_\infty$ (the center of the smallest ball containing the support of the measure), Badŏiu and Clarkson gave a fast and simple algorithm in [6]. For many applications in biology, signal processing, information geometry, extension to other spaces is necessary. The median in Hilbert space is computed in [9]. In nonlinear spaces with convexity assumptions, uniqueness has been established in [18] for the mean, [1] for the $p$–mean. Many algorithms of computation now exist. As far as deterministic algorithms are concerned, one can cite [2, 12, 13, 19] for the mean in Riemannian manifolds, [3] for the mean in Finsler manifolds, [11] and more generally [25] for the

*Keywords and phrases.* Stochastic algorithms, diffusion processes, simulated annealing, homogenization, probability measures on compact Riemannian manifolds, intrinsic p-means, instantaneous invariant measures, Gibbs measures, spectral gap at small temperature.

[1] Laboratoire de Mathématiques et Applications, CNRS: UMR 7348, Université de Poitiers, Téléport 2 – BP 30179, 86962 Futuroscope Chasseneuil Cedex, France. marc.arnaudon@math.univ-poitiers.fr

[2] Institut de Mathématique de Toulouse, CNRS: UMR 5219, 118, route de Narbonne, 31062 Toulouse Cedex 9, France. laurent.miclo@math.univ-toulouse.fr

Article published by EDP Sciences

© EDP Sciences, SMAI 2014

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

median, [5] for $e_\infty$. Stochastic algorithms avoid to compute the gradient of the functional to minimize. They can be found in [4,23]. For other functionals to minimize, see [8].

In this paper we investigate the case of non necessarily convex, complete Riemannian manifolds. Our first result (Thm. 2.1) concerns uniqueness of the $p$–mean of the uniform measure on a finite set $\{x_1,\dots,x_n\}$ of points, almost everywhere on $x=(x_1,\dots,x_n)$ for the Lebesgue measure. This generalizes Bhattacharya and Patrangenaru result on the circle ([7], case $p=2$). See also [10] for more general uniqueness criterions on the circle.

For computation of the $p$–mean, usual deterministic algorithms are not possible any more, due to the fact that the functional to minimize may have many local minima. So restricting to symmetric spaces we use a simulated annealing method with a continuous stochastic process, together with an estimation of the gradient to minimize *via* a drift moving faster and faster. With this method we are able to define a process which converges in distribution to the $p$–mean for $p\in[1,\infty)$ (Thm. 4.3, and Thm. 3.2 for more general but smooth functionals).

The main applications are in signal processing with polarimetric signal, but also for the group of rotations of $\mathbb{R}^n$, so as to determine averages on rotations. Also this solves many problems of optimization which may arise in economy, decision support, operation research. Notice that on the circle, fast computation of the mean has been performed in [17]. In fact this is a case where a closed formula can be found. For general case the situation is much more complicated and the convergence of our processes is slower and weaker. Jump processes and algorithms related to the continuous processes presented here will be investigated in a forthcoming paper.

## 2. Uniqueness of $p$–means for uniform measures with finite support

Let $M$ be a $d$–dimensional complete Riemannian manifold with Riemannian distance denoted by $\rho$. For $\nu$ a probability measure on $M$ and $p\geq 1$, we define

$$H_{p,\nu}: M\to\mathbb{R}_+\cup\{+\infty\},$$
$$y\mapsto\int_M \rho^p(y,z)\,\nu(\mathrm{d}z). \tag{2.1}$$

Either $H_{p,\nu}\equiv\infty$ or for all $y\in M$, $H_{p,\nu}(y)<\infty$. In the latter case we denote by $Q_{p,\nu}$ the set of minimizers of $H_{p,\nu}$. When $Q_{p,\nu}$ has only one element we denote it by $e_{p,\nu}$ and call it the $p$–mean of $\nu$. When there is no possible confusion we let $e_p=e_{p,\nu}$. For $x=(x_1,\dots,x_N)\in M^N$, we let

$$\mu(x)=\frac{1}{N}\sum_{k=1}^N\delta_{x_k}. \tag{2.2}$$

Clearly $H_{p,\mu(x)}$ is finite.

**Theorem 2.1.** *Assume $p>1$ or $\{d>1$ and $N>2\}$. For almost all $x\in M^N$, $Q_{p,\mu(x)}$ has a unique element $e_{p,\mu(x)}$*

**Remark 2.2.** This theorem extends Theorem 4.15 in [26] where the same result has been established for $p=1$ and $M$ compact.

*Proof.* We begin with the case $p>1$.

Since $\mu(x)$ has a finite support, we can assume that $M$ is a compact Riemannian manifold. For this a smooth modification outside a large ball is sufficient. For instance we can choose a radius so that the boundary is smooth, double the ball and finally smoothen the metric locally around the place where the pasting has been performed.

So in the sequel we will assume that $M$ is compact, with diameter $L$. For $y\in M$ we denote by $S_yM\subset T_yM$ the set of unit tangent vectors above $y$. Let

$$\tilde{V}=\{(y,n),\ y\in M,\ n=(n_1,\dots,n_N),\ n_j\in S_yM,\ j=1,\dots N\}\times[0,2L]^N. \tag{2.3}$$

Note $\tilde{V}$ is a compact smooth $(N+1)d$–dimensional manifold with boundary.

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

Define

$$\tilde{\phi} : \tilde{V} \to M^N$$
$$(y, n, r) \mapsto \left(\exp_y(n_1 r_1), \dots, \exp_y(n_N r_N)\right). \tag{2.4}$$

The map $\tilde{\phi}$ is onto. If $x = (x_1, \dots, x_N) \in M^N$, consider $y \in M$ minimizing $H_{p,\mu(x)}$. Then among all $(n, r)$ such that

$$\tilde{\phi}(y, n, r) = x \tag{2.5}$$

we can choose one so that for all $k = 1, \dots, N$ the map $s \mapsto \exp_y(sn_k)$ is a minimal geodesic for $s \in [0, r_k]$. For this choice we have

$$H_{p,\mu(x)}(y) = \frac{1}{N} \sum_{k=1}^{N} r_k^p. \tag{2.6}$$

Let us prove that

$$\sum_{k=1}^{N} r_k^{p-1} n_k = 0. \tag{2.7}$$

For this it is sufficient to check that for all $u \in T_y M$

$$\left\langle \sum_{k=1}^{N} r_k^{p-1} n_k, u \right\rangle = 0. \tag{2.8}$$

For $k = 1, \dots, N$, consider any smooth variation $c_k(s, a)$ of $s \mapsto \exp_y(sn_k)$, $s \in [0, r_k]$, defined on $[0, s_k] \times [-\varepsilon, \varepsilon]$ for some $\varepsilon > 0$, satisfying $c_k(s, 0) = \exp_y(sn_k)$ for $s \in [0, r_k]$, $c_k(0, a) = \exp_y(au)$ for $a \in [-\varepsilon, \varepsilon]$ and $c_k(r_k, a) = \exp_y(r_k n_k) = x_k$, $a \in [-\varepsilon, \varepsilon]$. Denote by $\ell_k(a)$ the length of the path $s \mapsto c_k(s, a)$, $s \in [0, r_k]$. By the variation of arc length formula, we have

$$-\frac{1}{p} \sum_{k=1}^{N} (\ell_k^p)'(0) = \sum_{k=1}^{N} \ell_k^{p-1}(0) \langle n_k, u \rangle = \left\langle \sum_{k=1}^{N} r_k^{p-1} n_k, u \right\rangle. \tag{2.9}$$

Now since $y$ minimizes $H_{p,\mu(x)}$ and by definition

$$H_{p,\mu(x)}(\exp_y(au)) \le \frac{1}{N} \sum_{k=1}^{N} \ell_k^p(a), \qquad H_{p,\mu(x)}(y) = \frac{1}{N} \sum_{k=1}^{N} \ell_k^p(0),$$

we have that 0 minimizes $a \mapsto \frac{1}{p} \sum_{k=1}^{N} \ell_k^p(a)$ and by (2.9) this implies (2.8). So equation (2.7) is proved.

Letting

$$\tilde{W}_p = \left\{ (y, n, r) \in \tilde{V},\ \sum_{k=1}^{N} r_k^{p-1} n_k = 0 \right\} \tag{2.10}$$

and $\tilde{\phi}_p = \tilde{\phi}|_{\tilde{W}_p}$ the restriction of $\tilde{\phi}$ to $\tilde{W}_p$, $\tilde{\phi}_p$ is onto, on $M^N$ by (2.5) and (2.7).

By Sard's theorem, the set $C_1 \subset M^N$ of singular values of $\tilde{\phi}_p$ has measure 0. It is closed since $\tilde{W}_p$ is compact.

Let us prove that the set

$$C_2 := \left\{ (x_1, \dots, x_N) \in M^N,\ \{x_1, \dots, x_N\} \cap Q_{p,\mu(x_1,\dots,x_N)} \neq \emptyset \right\} \tag{2.11}$$

has Lebesgue-measure 0: we can assume that for $i \neq j$, $x_i \neq x_j$ since we exclude 0-measure sets. So the elements we consider are images by $\tilde{\phi}_p$ of

$$\hat{W}_p = \left\{ (y, n, r) \in \tilde{W}_p,\ r_1 = 0,\ \forall k \ge 2\ r_k > 0 \right\}. \tag{2.12}$$

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

The set $\hat{W}_p$ is a submanifold of codimension 1 of $\tilde{W}_p$. Now $\dim \tilde{W}_p = Nd = \dim M^N$ so $\dim \hat{W}_p = \dim M^N - 1$ and its image by $\tilde{\phi}_p$ is of measure 0 in $M^N$. As a conclusion, $C_2$ has measure 0.

Define

$$C_3 := \left\{(x_1, \dots, x_N) \in M^N, \quad \exists i \neq j \text{ s.t. } x_i = x_j\right\} \tag{2.13}$$

and $C = C_1 \cup C_2 \cup C_3$. The set $C$ is closed in $M^N$ and has measure 0. Letting

$$W_p = \left\{(y, n, r) \in \tilde{W}_p, \ \forall k = 1, \dots N, r_k \in (0, 2L)\right\}, \tag{2.14}$$

we proved that $\tilde{\phi}_p|_{W_p}$ is onto on $M^N \backslash C$. Denote $\phi_p = \tilde{\phi}_p|_{W_p}$. Since $W_p$ has same dimension as $M^N$ and $\tilde{W}_p$ is compact, every point $x$ of $M^N \backslash C$ has a neighbourhood $V_x$ such that $\phi_p^{-1}(V_x) = U_{1,x} \cup \ldots \cup U_{m_x,x}$ where the $U_{j,x}$ are disjoint open subsets of $W_p$ and

$$\phi_p|_{U_{j,x}} : U_{j,x} \to \phi_p(U_{j,x}) \tag{2.15}$$

is a diffeomorphism. Now since $M^N \backslash C$ is second countable we can cover it by a countable number of such sets $V_x$. So to prove that the $p$–mean is almost everywhere unique it is sufficient to prove it on $V_x$.

For $x' \in V_x$ denote $x' = (x'_1, \dots, x'_N)$, and for $i \in \{1 \dots m_x\}$, write

$$(\phi|_{U_{i,x}})^{-1}(x') = (y_i(x'), n_1^i(x'), \dots n_d^i(x'), r_1^i(x'), \dots, r_d^i(x')).$$

Let $i, j \in \{1 \dots m_x\}$ satisfy $i \neq j$. If $y_i(x'), y_j(x') \in Q_{p,\mu(x')}$ then we have

$$H_{p,\mu(x')} \circ y_i(x') = H_{p,\mu(x')} \circ y_j(x'). \tag{2.16}$$

We can assume with the same argument as for (2.5) and (2.6) that the maps

$$\gamma_{i,k,x'} : s \mapsto \exp_{y_i(x')}(s n_k^i(x')) \quad \text{and} \quad \gamma_{j,k,x'} : s \mapsto \exp_{y_j(x')}(s n_k^j(x')) \tag{2.17}$$

are minimal geodesics respectively on $[0, r_k^i(x')]$ and $[0, r_k^j(x')]$. So letting $h_p : W_p \to \mathbb{R}$, $(y, n, r) \mapsto \sum_{k=1}^N r_k^p$, we have

$$\frac{1}{N} h_p \circ (\phi_p|_{U_{i,x}})^{-1}(x') = H_{p,\mu(x')} \circ y_i(x'), \quad \frac{1}{N} h_p \circ (\phi_p|_{U_{j,x}})^{-1}(x') = H_{p,\mu(x')} \circ y_j(x').$$

It is sufficient to prove that for all $x' \in V_x$,

$$h_p \circ (\phi_p|_{U_{i,x}})^{-1}(x') = h_p \circ (\phi_p|_{U_{j,x}})^{-1}(x') \tag{2.18}$$

implies

$$\operatorname{grad}_{x'} \left(h_p \circ (\phi_p|_{U_{i,x}})^{-1}\right) \neq \operatorname{grad}_{x'} \left(h_p \circ (\phi_p|_{U_{j,x}})^{-1}\right). \tag{2.19}$$

Indeed with (2.19) we will be able to deduce that the set

$$\left\{(x' \in V_x, \ h_p \circ (\phi_p|_{U_{i,x}})^{-1} = h_p \circ (\phi_p|_{U_{j,x}})^{-1}\right\} \tag{2.20}$$

has codimension $\geq 1$ in $V_x$ and this will imply that

$$\left\{(x' \in V_x, \ H_{p,\mu(x')} \circ y_i(x') = H_{p,\mu(x')} \circ y_j(x')\right\} \tag{2.21}$$

has codimension $\geq 1$ in $V_x$.

Let us prove (2.19). For $k = 1, \dots, N$ let

$$m_k^i(x') = -\dot{\gamma}_{i,k,x'}(r_k^i(x')) \quad \text{and} \quad m_k^j(x') = -\dot{\gamma}_{j,k,x'}(r_k^j(x')).$$

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

These unit vectors satisfy

$$\exp_{x'_k}(r^i_k(x')m^i_k(x')) = y_i(x') \quad \text{and} \quad \exp_{x'_k}(r^j_k(x')m^j_k(x')) = y_j(x').$$

Then noting that $\left(h_p \circ (\phi_p|_{U_{i,x}})^{-1}\right)(x') = \sum_{k=1}^N (r^i_k)^p(x'_k)$ we get

$$\begin{aligned} & d_{x'}\left(h_p \circ (\phi_p|_{U_{i,x}})^{-1}\right)(\cdot) \\ & = \left\langle -p\sum_{k=1}^N (r^i_k)^{p-1}(x')n^i_k(x'), T_{x'}y_i(\cdot)\right\rangle_{T_{y_i(x')}M} \\ & - p\left\langle \left((r^i_1(x'))^{p-1}m^i_1(x'), \ldots, (r^i_N(x'))^{p-1}m^i_N(x')\right), \cdot\right\rangle_{T_{x'}M^N}. \end{aligned}$$

Due to the fact that $(y_i(x'), n^i(x'), r^i(x')) \in W_p$, the first term in the right vanishes. So

$$\operatorname{grad}_{x'}\left(h_p \circ (\phi_p|_{U_{i,x}})^{-1}\right) = -p\left((r^i_1(x'))^{p-1}m^i_1(x'), \ldots, (r^i_N(x'))^{p-1}m^i_N(x')\right) \tag{2.22}$$

and similarly

$$\operatorname{grad}_{x'}\left(h_p \circ (\phi_p|_{U_{j,x}})^{-1}\right) = -p\left((r^j_1(x'))^{p-1}m^j_1(x'), \ldots, (r^j_N(x'))^{p-1}m^j_N(x')\right). \tag{2.23}$$

Since $y_i(x') \neq y_j(x')$ we have $(r^i_1(x'), m^i_1(x')) \neq (r^j_1(x'), m^j_1(x'))$, so $(r^i_1(x'))^{p-1}m^i_1(x') \neq (r^j_1(x'))^{p-1}m^j_1(x')$, from which we conclude that

$$\operatorname{grad}_{x'}\left(h_p \circ (\phi_p|_{U_{i,x}})^{-1}\right) \neq \operatorname{grad}_{x'}\left(h_p \circ (\phi_p|_{U_{j,x}})^{-1}\right).$$

This achieves the proof for the case $p > 1$.

Let us now consider the case $p = 1$. The result is due to Yang in [26], we give the proof here for completeness.

The main difference is that the subset of $M^N$ of points $x = (x_1, \ldots, x_N)$ so that $x_i \in Q_{1,\mu(x)}$ for some $i$ has positive measure.

First consider the open subset $U$ of $M^N$ of points $x$ such that for all $i = 1, \ldots, N$, $x_i \notin Q_{1,\mu(x)}$.

Consider the closed subset $C_0$ of $M^N$ of points $(x_1, \ldots, x_N) = \tilde{\phi}(y, n, r)$, with $(y, n, r) \in \tilde{V}$ such that for all $j, k = 1, \ldots N$, $n_j = \pm n_k$. Since $d > 1$ and $N > 2$ this subset has Lebesgue measure 0.

Replacing $M^N$ by $U$ and $C$ by $C_0 \cup C$, the argument is similar until (2.18). But now we will be able to prove that (2.18) implies (2.19) only in some neighbourhoods $V_{x,x'}$ to be precised later, of $x' \in V_x$ such that the geodesics

$$s \mapsto \exp_{y_i(x')}(sn^i_k(x')) \quad \text{and} \quad s \mapsto \exp_{y_j(x')}(sn^j_k(x'))$$

are minimal respectively on $[0, r^i_k(x')]$ and $[0, r^j_k(x')]$. But this will be sufficient since every compact subset of $V_x$ can be covered by a finite number of these neighbourhoods $V_{x,x'}$.

Making the above assumption on $x'$, the proof is similar until (2.22) and (2.23). Then we have

$$\operatorname{grad}_{x'}\left(h_1 \circ (\phi_1|_{U_{i,x}})^{-1}\right) = -\left(m^i_1(x'), \ldots, m^i_N(x')\right) \tag{2.24}$$

and

$$\operatorname{grad}_{x'}\left(h_1 \circ (\phi_1|_{U_{j,x}})^{-1}\right) = -\left(m^j_1(x'), \ldots, m^j_N(x')\right). \tag{2.25}$$

Assume

$$\operatorname{grad}_{x'}\left(h_1 \circ (\phi_1|_{U_{i,x}})^{-1}\right) = \operatorname{grad}_{x'}\left(h_1 \circ (\phi_1|_{U_{j,x}})^{-1}\right).$$

Then for all $k = 1, \ldots, N$, $m^i_k(x') = m^j_k(x')$. In particular for $k = 1$ this implies (possibly by exchanging $i$ and $j$) that $y_i(x')$ lies in the minimizing geodesic from $x'_1$ to $y_j(x')$. Now since $x' \notin C_0$ there exists $k \in \{1, \ldots N\}$ such

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

that $x'_k \notin \{\exp_{y_i(x')}(sn_1^i(x')),\ s \in [-2L, 2L]\}$. On the other hand since $m_k^i(x') = m_k^j(x')$, $y_j(x')$ (or $y_i(x')$) lies on the minimizing geodesic from $x'_k$ to $y_i(x')$ (or $y_j(x')$). As a consequence there are two minimizing geodesics from $y_i(x')$ to $y_j(x')$. But this is impossible since the geodesic from $x'_1$ to $y_j(x')$ is minimizing, contains $y_i(x')$ and $x'_1 \neq y_i(x')$ by the fact that we have supposed that $x'_1 \notin Q_{1,\mu(x')}$ and $y_i(x') \in Q_{1,\mu(x')}$. So

$$\mathrm{grad}_{x'}\left(h_1 \circ (\phi_1|_{U_{i,x}})^{-1}\right) \neq \mathrm{grad}_{x'}\left(h_1 \circ (\phi_1|_{U_{j,x}})^{-1}\right),$$

and by continuity this is true in a neighbourhood $V_{x,x'}$ of $x'$.

Now we consider the case where $x'_1 \in Q_{1,\mu(x')}$ and $x'_2 \notin Q_{1,\mu(x')}$. We follow the same lines as in the previous part with the difference that now $y_i(x') = x'_1$ and for the definition of $U_{i,x}$ $W_1$ is replaced by

$$W_1^i = \{(y, n, r) \in V,\ r_1 = 0\}.$$

The definition of $U_{j,x}$ remains unchanged. By [25] Theorem 1

$$\left\| \frac{1}{N} \sum_{k=2}^{N} n_k^i(x') \right\| \leq \mu_N(x')(\{x'_1\})$$

which gives

$$\left\| \sum_{k=2}^{N} n_k^i(x') \right\| \leq 1. \tag{2.26}$$

Since $d > 1$ and $N > 2$, the submanifolds of $V_x$ images of

$$\left\{ (y, n, r) \in U_{i,x},\ \left\| \sum_{k=2}^{N} n_k \right\| = 1 \right\}$$

and

$$\left\{ (y, n, r) \in U_{i,x},\ \sum_{k=2}^{N} n_k = 0 \right\}$$

by $\phi_1$ have measure 0, so we can exclude them. On the subset

$$\left\{ (y, n, r) \in U_{i,x},\ 0 < \left\| \sum_{k=2}^{N} n_k \right\| < 1 \right\},$$

the function $h_1$ is smooth and on its image by $\phi_1$,

$$\mathrm{grad}_{x'}\left(h_1 \circ (\phi_1|_{U_{i,x}})^{-1}\right) = -\left(0, m_2^i(x'), \ldots, m_N^i(x')\right). \tag{2.27}$$

Again

$$\mathrm{grad}_{x'}\left(h_1 \circ (\phi_1|_{U_{j,x}})^{-1}\right) = -\left(m_1^j(x'), \ldots, m_N^j(x')\right). \tag{2.28}$$

They are not equal, and this achieves the proof for this case by the same argument as before.

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

Finally we consider the case where $x'_1, x'_2 \in Q_{1,\mu(x')}$ with $x'_1 = y_i(x')$ and $x'_2 = y_j(x')$. We follow the same line as in the previous case, but now for the definition of $U_{j,x}$, $W_1$ is replaced by

$$W_1^j = \{(y, n, r) \in V,\ r_2 = 0\}.$$

Again we can exclude the submanifolds of $V_x$ images of

$$\left\{ (y, n, r) \in U_{j,x},\ \left\| \sum_{k\in\{1,\dots,N\}, k\neq 2} n_k \right\| = 1 \right\}$$

and

$$\left\{ (y, n, r) \in U_{j,x},\ \sum_{k\in\{1,\dots,N\}, k\neq 2} n_k = 0 \right\}$$

by $\phi_1$ and work on

$$\phi_1\left(\left\{ (y, n, r) \in U_{j,x},\ 0 < \left\| \sum_{k\in\{1,\dots,N\}, k\neq 2} n_k \right\| < 1 \right\}\right) \cap \phi_1\left(\left\{ (y, n, r) \in U_{i,x},\ 0 < \left\| \sum_{k=2}^{N} n_k \right\| < 1 \right\}\right).$$

On this set $h_1 \circ (\phi_1|_{U_{i,x}})^{-1}$ and $h_1 \circ (\phi_1|_{U_{j,x}})^{-1}$ are smooth and

$$\mathrm{grad}_{x'}\left(h_1 \circ (\phi_1|_{U_{i,x}})^{-1}\right) = -\left(0, m_2^i(x'), \dots, m_N^i(x')\right). \tag{2.29}$$

$$\mathrm{grad}_{x'}\left(h_1 \circ (\phi_1|_{U_{j,x}})^{-1}\right) = -\left(m_1^j(x'), 0, m_3^j(x'), \dots, m_N^j(x')\right). \tag{2.30}$$

They are not equal, and this achieves the proof. □

**Corollary 2.3.** *Let $p \in [1, \infty)$ and $X = (X_1, \dots, X_N)$ a random variable with values in $M^N$, which has an absolutely continuous law. Then almost-surely $\mu(X(\omega))$ has a unique $p$–mean $e_p(X(\omega))$.*

**Corollary 2.4.** *Let $p \in [1, \infty)$ and $(X_n)_{n\geq 1}$ a sequence of i.i.d. $M$-valued random variables with absolutely continuous laws. Then the process of empirical $p$–means*

$$\Big(e_{p,n}(\omega) := e_p\big(X_1(\omega), \dots, X_n(\omega)\big)\Big)_{n\geq 1}$$

*is well-defined.*

**Remark 2.5.** For $p = 2$ and $M$ a circle, it has been proved in [7] that the assumption can be weakened: the same result holds if the law has no atom.

We believe that it would be interesting to study the behaviour of the process $(e_{p,n})_{n\geq 1}$ in many situations. For instance when the law of $X_1$ is uniform on a compact symmetric space (even the case of the circle is highly non trivial) one would observe a recurrent but irregular and slower and slower process. Again on a compact symmetric space, when the law $\nu$ of $X_1$ has a finite number of $p$–means due to a finite group of symmetries, one would observe an almost stationary behaviour, and at increasingly spaced times jumps between smaller and smaller neighbourhoods of the $p$–means of $\nu$.

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

## 3. Finding the minimizers of some integrated functionals with simulated annealing

Let $M$ be a compact Riemannian manifold. For simplicity and without loss of generality we assume that $M$ has Lebesgue volume 1. On $M$ consider a probability law $\nu$ with a density with respect to Lebesgue measure, also denoted by $\nu$. Assume we are given a continuous function $\kappa : M \times M \to \mathbb{R}_+$, where $\kappa(\theta, y)$ is interpreted as some kind of cost for going from $\theta$ to $y$. Assume furthermore that for all $y \in M$ the function $\theta \mapsto \kappa(\theta, y)$ is smooth and that its first and second derivative in $\theta$ are uniformly bounded in $(\theta, y)$. Consider on $M$ the functional

$$\begin{aligned} U : M &\to \mathbb{R}_+ \\ \theta &\mapsto \int_M \kappa(\theta, y)\nu(\mathrm{d}y) \end{aligned} \tag{3.1}$$

Denote by $\mathcal{M}$ the set of minimizers of $U$. The aim of this section is to find a continuous semimartingale which converges in law to $\mathcal{M}$. Also we try to avoid using the gradient of $U$, which in many cases is difficult or impossible to compute.

For this we will use a sequence $(P_k)_{k\geq 0}$ of independent random variables with law $\nu$, a Poisson process $N_t$ on $\mathbb{N}$ with intensity $\gamma_t^{-1}$ where

$$\gamma_t = (1+t)^{-1}. \tag{3.2}$$

Define

$$c(U) = 2 \sup_{\theta, y \in M} \left( \inf_{\phi \in \mathscr{C}_{\theta,y}} e(\phi) \right), \tag{3.3}$$

$\mathscr{C}_{\theta,y}$ denoting the set of continuous paths $[0,1] \to M$ and for $\phi \in \mathscr{C}_{\theta,y}$, the elevation $e(\phi)$ being defined as

$$e(\phi) = \sup_{0\leq t\leq 1} U(\phi(t)) - U(\theta) - U(y) + \inf_{z\in M} U(z). \tag{3.4}$$

Let

$$\beta_t = \frac{1}{k}\ln(1+t), \tag{3.5}$$

the constant $k$ satisfying $k > c(U)$.

We assume that $(N_t)_{t\geq 0}$ is independent of the sequence $(P_k)_{k\geq 0}$. We let $(B_t)_{t\geq 0}$ be a Brownian motion with values in $\mathbb{R}^r$ for some $r \in \mathbb{N}$, independent of $(N_t)_{t\geq 0}$ and $(P_k)_{k\geq 0}$, and $\sigma$ a smooth section of $TM \otimes (\mathbb{R}^r)^*$: for all $\theta \in M$, $\sigma(\theta)$ is a linear map $\mathbb{R}^r \to T_\theta M$. We assume that for all $\theta \in M$, we have $\sigma(\theta)\sigma(\theta)^* = \mathrm{id}_{T_\theta M}$. We fix $\theta_0 \in M$ and let $\Theta_t$ be the solution started at $\theta_0$ of the Itô equation

$$\mathrm{d}y_t = \sigma(\Theta_t)\,\mathrm{d}B_t - \beta_t \operatorname{grad}_{\Theta_t} \kappa(\cdot, Y_t)\,\mathrm{d}t \quad \text{with} \quad Y_t = P_{N_t}. \tag{3.6}$$

Recall that if $P(\Theta_t) : T_{\theta_0}M \to T_{\Theta_t}M$ is the parallel transport map along $(\Theta_t)$, then

$$\mathrm{d}\Theta_t = P(\Theta_t) d\left( \int_0^{\cdot} P(\Theta_s)^{-1} \circ \mathrm{d}\Theta_s \right)_t. \tag{3.7}$$

Also define $\Theta_t^0$ the solution started at $\theta_0$ of the Itô equation

$$\mathrm{d}\Theta_t^0 = \sigma(\Theta_t^0)\,\mathrm{d}B_t - \beta_t \left( \int_M \operatorname{grad}_{\Theta_t^0} \kappa(\cdot, y)\,\nu(y)\mathrm{d}y \right) \mathrm{d}t. \tag{3.8}$$

Note (3.8) rewrites as

$$\mathrm{d}\Theta_t^0 = \sigma(\Theta_t^0)\,\mathrm{d}B_t - \beta_t \operatorname{grad}_{\Theta_t^0} U\,\mathrm{d}t, \tag{3.9}$$

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

so that the same equation with fixed $\beta$ instead of $\beta_t$ has an invariant law with density

$$\mu_\beta(\theta) = \frac{1}{Z_\beta} \mathrm{e}^{-2\beta U(\theta)}, \quad \text{with} \quad Z_\beta = \int_M \mathrm{e}^{-2\beta U(\theta')}\, \mathrm{d}\theta'. \tag{3.10}$$

The process $\Theta_t^0$ is an inhomogeneous diffusion with generator

$$L_t^0(\theta) = \frac{1}{2}\Delta(\theta) - \beta_t \operatorname{grad}_\theta U \tag{3.11}$$

(here and in the sequel a vector field $A$ is identified with the map $f \mapsto A(f) = \langle \operatorname{grad} f, A\rangle$ which acts on $C^1$ functions on $M$). Denote by $m_t(\theta)$ the density of $\Theta_t$.

The process $(\Theta_t, Y_t)$ is Markovian with generator $L_t$ given by

$$\begin{aligned} L_t f(\theta, y) &= \left(\frac{1}{2}\Delta(\theta) - \beta_t \operatorname{grad}_\theta \kappa(\cdot, y)\right) f(\cdot, y) + \gamma_t^{-1} \int_M (f(\theta, z) - f(\theta, y))\ \nu(\mathrm{d}z) \\ &= L_{1,t} f(\cdot, y)(\theta) + L_{2,t} f(\theta, \cdot)(y). \end{aligned} \tag{3.12}$$

We know that for all neighbourhood $\mathcal{N}$ of $\mathcal{M}$, $\int_{\mathcal{N}} \mu_\beta(\theta)\, \mathrm{d}\theta$ converges to 1 as $\beta \to \infty$. So to prove that $\int_{\mathcal{N}} m_t(\theta)\, \mathrm{d}\theta$ converges to 1 it is sufficient to prove the following proposition:

**Proposition 3.1.** *The entropy*

$$J_t := \int_M \ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right) m_t(\theta)\, \mathrm{d}\theta \tag{3.13}$$

*converges to* 0 *as* $t \to \infty$.

*Proof.* There will be 3 steps.

In the sequel we will denote by $m_t(\theta, y)$ the joint density of $(\Theta_t, Y_t)$, and $m_t(y|\theta)$ the density of $Y_t$ conditioned by $\Theta_t = \theta$.

**Step 1.** Let us prove that

$$\frac{\mathrm{d}J_t}{\mathrm{d}t} \le \frac{4\|\kappa\|_\infty}{k(1+t)} - c_2(\beta_t \vee 1)^{-p} \exp\left(-c(U)\beta_t\right) J_t + \beta_t^2 32 K^2 I_t \tag{3.14}$$

with

$$I_t = \int_{M\times M} \ln\left(\frac{m_t(y|\theta)}{\nu(y)}\right) m_t(\theta, y)\, \mathrm{d}\theta \mathrm{d}y \tag{3.15}$$

and $c_2, p, K > 0$ defined below (in (3.20) and (3.22)).

We compute

$$\frac{\mathrm{d}J_t}{\mathrm{d}t} = \int_M \frac{dm_t(\theta)}{\mathrm{d}t}\, \mathrm{d}\theta - \int_M \frac{\mathrm{d}\ln \mu_{\beta_t}(\theta)}{\mathrm{d}t} m_t(\theta)\, \mathrm{d}\theta + \int_M \ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right) \frac{dm_t(\theta)}{\mathrm{d}t}\, \mathrm{d}\theta. \tag{3.16}$$

Since for all $t$ $m_t(\theta)$ is a probability density, the first term in the right vanishes. So we get

$$\frac{\mathrm{d}J_t}{\mathrm{d}t} = 2\beta_t' \int_M U(\theta)(m_t(\theta) - \mu_{\beta_t}(\theta))\, \mathrm{d}\theta + \int_{M\times M} L_t\left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right)\right] m_t(\theta, y)\, \mathrm{d}\theta \mathrm{d}y \tag{3.17}$$

where the last term comes from Dynkin formula. For the first term in the right we have using (3.5)

$$2\beta_t' \int_M U(\theta)(m_t(\theta) - \mu_{\beta_t}(\theta))\, \mathrm{d}\theta \le 4\|U\|_\infty |\beta_t'| \le \frac{4\|\kappa\|_\infty}{k(1+t)}. \tag{3.18}$$

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

Now by writing $L_t = L_t^0 + R_t$, we split the second term in the right of (3.17) into

$$\begin{aligned}&\int_{M\times M} L_t \left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right)\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y\\&= \int_M L_t^0 \left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right)\right] m_t(\theta)\,\mathrm{d}\theta + \int_{M\times M} R_t(\theta,y)\left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right)\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y. \qquad (3.19)\end{aligned}$$

We have

$$\begin{aligned}&\int_M L_t^0 \left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right)\right] m_t(\theta)\,\mathrm{d}\theta\\&= \int_M L_t^0 \left[\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right)\right] \mu_{\beta_t}(\theta)\,\mathrm{d}\theta - \frac{1}{2}\int_M \left\|\mathrm{grad}_\theta \ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right)\right\|^2 m_t(\theta)\,\mathrm{d}\theta\\&= -2\int_M \left\|\mathrm{grad}_\theta \sqrt{\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}}\right\|^2 \mu_{\beta_t}(\theta)\,\mathrm{d}\theta\\&\le -2c_2(\beta_t\vee 1)^{-p}\exp\left(-c(U)\beta_t\right) J_t \qquad (3.20)\end{aligned}$$

for some $c_2 > 0$ and integer $p > 0$ by logarithmic Sobolev inequality ([15] and [16], for more details see [20]). Note we used again Dynkin formula to prove the vanishing of the first term in the right of the second line.

As for the second term in the right of (3.19) we have

$$\begin{aligned}&\int_{M\times M} R_t(\theta,y)\left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right)\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y\\&= \int_{M\times M} -\beta_t \left\langle \mathrm{grad}_\theta \ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right), \mathrm{grad}_\theta\,\kappa(\cdot,y) - \int_M \mathrm{grad}_\theta\,\kappa(\cdot,z)\,\nu(\mathrm{d}z)\right\rangle m_t(\theta,y)\,\mathrm{d}\theta\,\mathrm{d}y\\&= -\beta_t \int_M \left\langle \mathrm{grad}_\theta \ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right), \int_M \mathrm{grad}_\theta\,\kappa(\cdot,y)\,(m_t(y|\theta)-\nu(y))\,\mathrm{d}y\right\rangle m_t(\theta)\,\mathrm{d}\theta\\&= 2\beta_t \int_M \sqrt{\frac{\mu_{\beta_t}}{m_t}(\theta)}\left\langle \mathrm{grad}_\theta \sqrt{\frac{m_t}{\mu_{\beta_t}}(\theta)}, R_t(\theta)\right\rangle m_t(\theta)\,\mathrm{d}\theta\end{aligned}$$

with

$$R_t(\theta) = -\int_M \mathrm{grad}_\theta\,\kappa(\cdot,y)(m_t(y|\theta)-\nu(y))\,\mathrm{d}y.$$

So by Cauchy–Schwarz inequality

$$\begin{aligned}&\int_{M\times M} R_t(\theta,y)\left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right)\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y\\&\le 2\beta_t\left(\int_M \left\|\mathrm{grad}_\theta \sqrt{\frac{m_t}{\mu_{\beta_t}}(\theta)}\right\|^2 \mu_{\beta_t}(\theta)\,\mathrm{d}\theta\right)^{1/2}\left(\int_M \|R_t(\theta)\|^2 m_t(\theta)\,\mathrm{d}\theta\right)^{1/2}\\&\le \beta_t^2\int_M \|R_t(\theta)\|^2 m_t(\theta)\,\mathrm{d}\theta + \int_M \left\|\mathrm{grad}_\theta \sqrt{\frac{m_t}{\mu_{\beta_t}}(\theta)}\right\|^2 \mu_{\beta_t}(\theta)\,\mathrm{d}\theta.\end{aligned}$$

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

Summing with (3.20) and using (3.19) we get

$$\begin{aligned}
&\int_{M\times M} L_t\left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right)\right] m_t(\theta,y))\,\mathrm{d}\theta\mathrm{d}y\\
&\le \beta_t^2\int_M \|R_t(\theta)\|^2 m_t(\theta)\,\mathrm{d}\theta - \int_M \left\|\mathrm{grad}_\theta\sqrt{\frac{m_t}{\mu_{\beta_t}}(\theta)}\right\|^2 \mu_{\beta_t}(\theta)\,\mathrm{d}\theta\\
&\le \beta_t^2\int_M \|R_t(\theta)\|^2 m_t(\theta)\,\mathrm{d}\theta - c_2(\beta_t\vee 1)^{-p}\exp\left(-c(U)\beta_t\right) J_t. \qquad (3.21)
\end{aligned}$$

Defining

$$K = \sup_{\theta,y\in M} \|\,\mathrm{grad}_\theta\,\kappa(\cdot,y)\|, \qquad (3.22)$$

let us now bound

$$\begin{aligned}
\int_M \|R_t(\theta)\|^2 m_t(\theta)\,\mathrm{d}\theta &= \int_M \left\|\int_M \mathrm{grad}_\theta\,\kappa(\cdot,y)(m_t(y|\theta)-\nu(y))\,\mathrm{d}y\right\|^2 m_t(\theta)\,\mathrm{d}\theta\\
&\le \int_M \left\|K\int_M |m_t(y|\theta)-\nu(y)|\,\mathrm{d}y\right\|^2 m_t(\theta)\,\mathrm{d}\theta\\
&\le 32K^2\int_M\left(\int_M \ln\left(\frac{m_t(y|\theta)}{\nu(y)}\right) m_t(y|\theta)\,\mathrm{d}y\right) m_t(\theta)\,\mathrm{d}\theta\\
&= 32K^2 I_t \qquad (3.23)
\end{aligned}$$

where $I_t$ is defined in (3.15).

We also used classical bound of total variation by entropy ([16]):

$$\int_M |m_t(y|\theta)-\nu(y)|\,\mathrm{d}y \le 4\sqrt{2}\left(\int_M \ln\left(\frac{m_t(y|\theta)}{\nu(y)}\right) m_t(y|\theta)\,\mathrm{d}y\right)^{1/2}.$$

At this stage, combining (3.17), (3.18), (3.21), (3.23) and (3.15), we proved (3.14).

**Step 2.** Let us prove that

$$\frac{\mathrm{d}I_t}{\mathrm{d}t} \le 4\|\kappa\|_\infty\beta_t' + K'(\beta_t\vee 1)\beta_t - \frac{\mathrm{d}J_t}{\mathrm{d}t} - \gamma_t^{-1} I_t \qquad (3.24)$$

with

$$K' = \sup_{\theta,y\in M} |\Delta_\theta\kappa(\cdot,y)| + 2K^2. \qquad (3.25)$$

As before

$$\begin{aligned}
\frac{\mathrm{d}I_t}{\mathrm{d}t} &= \int_{M\times M} L_t\left[\ln\left(\frac{m_t(y|\theta)}{\nu(y)}\right)\right] m_t(y,\theta)\,\mathrm{d}\theta\mathrm{d}y\\
&= \int_{M\times M} (L_{2,t}+L_{1,t})\left[\ln\left(\frac{m_t(y|\theta)}{\nu(y)}\right)\right] m_t(y,\theta)\,\mathrm{d}\theta\mathrm{d}y. \qquad (3.26)
\end{aligned}$$

We begin with the first term:

$$\begin{aligned}
&\int_{M\times M} L_{2,t}\left[\ln\left(\frac{m_t(y|\theta)}{\nu(y)}\right)\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y\\
&= \gamma_t^{-1}\int_{M\times M}\int_M \left[\ln\left(\frac{m_t(z|\theta)}{\nu(z)}\right) - \ln\left(\frac{m_t(y|\theta)}{\nu(y)}\right)\right]\nu(\mathrm{d}z) m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y\\
&= \gamma_t^{-1}\int_{M\times M} \ln\left(\frac{m_t(y|\theta)}{\nu(y)}\right)(\nu(y)-m_t(y|\theta))\,m_t(\theta)\,\mathrm{d}\theta\mathrm{d}y.
\end{aligned}$$

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

By Jensen inequality we have

$$\begin{aligned}
&\int_{M\times M} \ln\left(\frac{m_t(y|\theta)}{\nu(y)}\right)\nu(y)m_t(\theta)\,\mathrm{d}y\mathrm{d}\theta\\
&= \int_M \left(\int_M \ln\left(\frac{m_t(y|\theta)}{\nu(y)}\right)\nu(y)\,\mathrm{d}y\right) m_t(\theta)\,\mathrm{d}\theta\\
&\leq \int_M \ln\left(\int_M \frac{m_t(y|\theta)}{\nu(y)}\nu(y)\,\mathrm{d}y\right) m_t(\theta)\,\mathrm{d}\theta\\
&= \int_M \ln(1)m_t(\theta)\,\mathrm{d}\theta = 0
\end{aligned}$$

Consequently

$$\int_{M\times M} L_{2,t}\left[\ln\left(\frac{m_t(y|\theta)}{\nu(y)}\right)\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y \leq -\gamma_t^{-1}\int_{M\times M}\ln\left(\frac{m_t(y|\theta)}{\nu(y)}\right)m_t(y|\theta)m_t(\theta)\,\mathrm{d}\theta\mathrm{d}y$$

which rewrites as

$$\int_{M\times M} L_{2,t}\left[\ln\left(\frac{m_t(y|\theta)}{\nu(y)}\right)\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y \leq -\gamma_t^{-1}I_t. \tag{3.27}$$

Let us now consider the second term in the right of (3.26). Since

$$\ln\left(\frac{m_t(y|\theta)}{\nu(y)}\right) = \ln\left(\frac{m_t(\theta|y)}{m_t(\theta)}\right)$$

(recall that $Y_t$ has law $\nu$) it rewrites as

$$\int_{M\times M} L_{1,t}\left[\ln\left(\frac{m_t(\theta|y)}{m_t(\theta)}\right)\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y = \int_{M\times M} L_{1,t}\left[\ln(m_t(\theta|y)) - \ln(m_t(\theta))\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y. \tag{3.28}$$

But

$$\begin{aligned}
&\int_{M\times M} L_{1,t}\ln(m_t(\theta|y))m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y\\
&= \frac{1}{2}\int_{M\times M} \Delta\ln(m_t(\theta|y))m_t(\theta,y)\mathrm{d}\theta\mathrm{d}y - \beta_t\int_{M\times M}\langle \mathrm{grad}_\theta \ln m_t(\cdot|y), \mathrm{grad}_\theta\,\kappa(\cdot,y)\rangle\, m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y.
\end{aligned} \tag{3.29}$$

We compute

$$\begin{aligned}
&\int_{M\times M} \Delta\ln(m_t(\theta|y))m_t(\theta,y)\mathrm{d}\theta\mathrm{d}y\\
&\int_M\left(\int_M \Delta m_t(\theta|y)\,\mathrm{d}\theta\right)\nu(y)\,\mathrm{d}y - \int_{M\times M}\|\mathrm{grad}_\theta \ln m(\theta|y)\|^2\, m(\theta,y)\,\mathrm{d}\theta\mathrm{d}y\\
&= -4\int_{M\times M}\left\|\mathrm{grad}_\theta\sqrt{m(\theta|y)}\right\|^2\nu(y)\,\mathrm{d}\theta\mathrm{d}y
\end{aligned}$$

where we used the fact that the first term in the right of the first equality vanishes. Consequently,

$$\begin{aligned}
&\int_{M\times M} L_{1,t}\ln(m_t(\theta|y))m_t(y,\theta)\,\mathrm{d}\theta\mathrm{d}y\\
&= -2\int_{M\times M}\left\|\mathrm{grad}_\theta\sqrt{m_t(\theta|y)}\right\|^2\mathrm{d}\theta\nu(\mathrm{d}y)\\
&- \beta_t\int_{M\times M}\langle \mathrm{grad}_\theta\, m_t(\cdot|y), \mathrm{grad}_\theta\,\kappa(\cdot,y)\rangle\,\nu(y)\,\mathrm{d}\theta\mathrm{d}y.
\end{aligned} \tag{3.30}$$

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

Let us bound the absolute value of the last term:

$$\begin{aligned}
&\left|-\beta_t \int_{M\times M} \langle \mathrm{grad}_\theta\, m_t(\cdot|y), \mathrm{grad}_\theta\, \kappa(\cdot,y)\rangle\, \nu(y)\, \mathrm{d}\theta\mathrm{d}y\right| \\
&= \left|2\beta_t \int_{M\times M} \left\langle \mathrm{grad}_\theta \sqrt{m_t(\theta|y)}, \mathrm{grad}_\theta\, \kappa(\theta,y)\right\rangle \sqrt{m_t(\theta|y)}\nu(\mathrm{d}y)\, \mathrm{d}\theta\mathrm{d}y\right| \\
&\leq 2\beta_t K \int_{M\times M} \left\|\mathrm{grad}_\theta \sqrt{m_t(\theta|y)}\right\| \sqrt{m_t(\theta|y)}\nu(y)\, \mathrm{d}\theta\mathrm{d}y \\
&\leq \int_{M\times M} \left(\frac{1}{2}\beta_t^2 K^2 m_t(\theta|y) + 2\left\|\mathrm{grad}_\theta \sqrt{m_t(\theta|y)}\right\|^2\right) \nu(y)\, \mathrm{d}\theta\mathrm{d}y \\
&= \frac{1}{2}\beta_t^2 K^2 + 2\int_{M\times M} \left\|\mathrm{grad}_\theta \sqrt{m_t(\theta|y)}\right\|^2 \nu(y)\, \mathrm{d}\theta\mathrm{d}y. \qquad (3.31)
\end{aligned}$$

This yields

$$\int_{M\times M} L_{1,t} \ln(m_t(\theta|y)) m_t(\theta,y)\, \mathrm{d}\theta\mathrm{d}y \leq \frac{1}{2}\beta_t^2 K^2 \qquad (3.32)$$

We also have to bound the last term in (3.28):

$$-L_{1,t} \ln(m_t(\theta)) = -L_{1,t} \ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right) - L_{1,t} \ln(\mu_{\beta_t}(\theta)). \qquad (3.33)$$

From (3.17) we get

$$\frac{\mathrm{d}J_t}{\mathrm{d}t} \leq 4\|\kappa\|_\infty \beta_t' + \int_{M\times M} L_{1,t} \ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right) m_t(\theta,y)\, \mathrm{d}\theta\mathrm{d}y$$

or equivalently

$$-\int_{M\times M} L_{1,t} \ln\left(\frac{m_t(\theta)}{\mu_{\beta_t}(\theta)}\right) m_t(\theta,y)\, \mathrm{d}\theta\mathrm{d}y \leq -\frac{\mathrm{d}J_t}{\mathrm{d}t} + 4\|\kappa\|_\infty \beta_t'. \qquad (3.34)$$

For the second term we have

$$\begin{aligned}
-L_{1,t} \ln(\mu_{\beta_t}(\theta)) &= 2\beta_t L_{1,t} U(\theta) \\
&= \beta_t \Delta U(\theta) + 2\beta_t^2 \langle dU, \mathrm{grad}_\theta\, \kappa(\cdot,y)\rangle \\
&= \beta_t \int_M \Delta_\theta \kappa(\theta,y)\nu(\mathrm{d}y) + 2\beta_t^2 \int_M \|\mathrm{grad}_\theta\, \kappa(\theta,y)\|^2\, \nu(\mathrm{d}y) \\
&\leq K'(\beta_t \vee 1)\beta_t \qquad (3.35)
\end{aligned}$$

with $K'$ defined in (3.25).

Finally we obtain (3.24).

**Step 3.** We finally prove that

$$\lim_{t\to\infty} J_t = 0. \qquad (3.36)$$

With inequalities (3.24) and (3.14) we can use the end of the proof of theorem 1 in [21] to obtain that under assumptions (3.5) and (3.3) then (3.36) holds (notice that in Sect. 4 we will prove this in a more general context). □

**Theorem 3.2.** *Assume*

$$\beta_t = \frac{1}{k}\ln(1+t), \quad \textit{and} \quad \gamma_t = (1+t)^{-1}, \qquad (3.37)$$

*where $k > c(U)$, ($c(U)$ defined in (3.3)). Then for any neighbourhood $\mathcal{N}$ of $\mathcal{M}$,*

$$\lim_{t\to\infty} \mathbb{P}\left[\Theta_t \in \mathcal{N}\right] = 1. \qquad (3.38)$$

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

*Proof.* We use Proposition 3.1 together with the fact that

$$\|m_t - \mu_{\beta_t}\| \leq 4\sqrt{2J_t}$$

and

$$\lim_{t\to\infty} \mu_{\beta_t}(\mathcal{N}) = 1. \qquad \square$$

## 4. Application to location of $p$–means in symmetric spaces

In this section we assume that $M$ is a compact symmetric space endowed with the canonical Riemannian metric of volume 1. Denote by $\rho$ the Riemannian distance in $M$, $D$ its diameter. We fix $p \geq 1$ and consider a probability measure $\nu$ on $M$. We aim to find at least one element of $Q_{p,\nu}$ by using the result of the previous section. In particular if $\nu$ has a unique $p$–mean $e_p$, then we will be able to construct a process which converges in probability to $e_p$ as $t \to \infty$.

Denote by $p(s,x,y)$ the heat kernel on $M$, and for $s > 0$ let $\nu_s$ be the probability measure with density

$$\nu_s(y) = \int_M p(s,y,z)\nu(\mathrm{d}z), \tag{4.1}$$

and let

$$\begin{aligned} \kappa_s : M \times M &\to \mathbb{R}_+ \\ (\theta, y) &\mapsto \int_M p(s,\theta,z)\rho^p(z,y)\,\mathrm{d}z, \end{aligned} \tag{4.2}$$

and

$$\begin{aligned} U_{s_1,s_2} : M &\to \mathbb{R}_+ \\ \theta &\mapsto \int_M \kappa_{s_1}(\theta,y)\nu_{s_2}(y)\,\mathrm{d}y. \end{aligned} \tag{4.3}$$

Also let $U = H_{p,\nu}$. Clearly $\nu_{s_1}$ and $\kappa_{s_2}$ satisfy the assumption of the previous section. Moreover, denoting by $\mathcal{M}_{s_1,s_2}$ the set of minimizers of $U_{s_1,s_2}$ then as $s_1, s_2 \to 0$ we have $\mathcal{M}_{s_1,s_2} \to Q_{p,\nu}$ is the sense that for any neighbourhood $\mathcal{N}$ of $Q_{p,\nu}$, we have $\mathcal{M}_{s_1,s_2} \subset \mathcal{N}$ for all $s_1, s_2$ sufficiently small. This is due to the fact that as $s_1, s_2 \to 0$, $U_{s_1,s_2}(\theta) \to U(\theta)$ uniformly in $\theta$.

**Lemma 4.1.** *For all $s_1, s_2 > 0$ we have*

$$U_{s_1,s_2}(\theta) = U_{0,s_1+s_2}(\theta) = \int_M \rho^p(\theta,y)\nu_{s_1+s_2}(y)\,\mathrm{d}y. \tag{4.4}$$

*Proof.* Fix $\theta, y \in M$, let $m$ be the middle point of a minimal geodesic from $\theta$ to $y$ and $i_m$ the symmetry centered at $m$. We have

$$\begin{aligned} \int_M p(s_1,\theta,z)\rho^p(z,y)\,\mathrm{d}z &= \int_M p(s_1, i_m(\theta), i_m(z))\rho^p(i_m(z), i_m(y))\,\mathrm{d}z \\ &= \int_M p(s_1, i_m(\theta), z')\rho^p(z', i_m(y))\,\mathrm{d}z' \\ &= \int_M p(s_1, y, z')\rho^p(z', \theta)\,\mathrm{d}z' \\ &= \int_M \rho^p(\theta, z')p(s_1, z', y)\,\mathrm{d}z' \end{aligned}$$

where we first used the invariance by isometry of the heat kernel and then did the change of variable $z' = i_m(z)$ in the integral and finally used the symmetry of the heat kernel. To finish the proof we are left to use the convolution property of the heat semigroup. $\square$

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

**Corollary 4.2.** *We have for all* $s_1, s_2 > 0$, $\theta, y \in M$,

$$\|\operatorname{grad}_\theta \kappa_{s_1}(\cdot, y)\| \leq pD^{p-1} =: K \quad \textit{and} \quad \|\operatorname{grad}_\theta U_{s_1,s_2}\| \leq K. \tag{4.5}$$

With all these properties we would like to find $s_1(t) \searrow 0$ and $s_2(t) \searrow 0$ such that the process $\Theta_t$ started at $\theta_0$ and solution to

$$\mathrm{d}\Theta_t = \sigma(\Theta_t)\,\mathrm{d}B_t - \beta_t \operatorname{grad}_{\Theta_t} \kappa_{s_1(t)}(\cdot, Y_t^{s_2})\,\mathrm{d}t \tag{4.6}$$

converges in law to $e_p$, where by definition $Y_t^{s_2} = Y_{T_{N_t}}$ and $Y_n$ is the second coordinate of a Poisson point process $(T_n, Y_n)_{n\geq 1}$ taking its values in $[0,\infty) \times M$ with intensity $\gamma(t)^{-1}\nu_{s_2(t)}(y)\,\mathrm{d}t\,\mathrm{d}y$, independent of $(B_t)$. The process $N_t$ is the counting function of $T_1 < T_2 < \dots$ So $N_t$ is a Poisson process on $\mathbb{N}$ with intensity $\gamma_t^{-1}$, and conditioned by $(N_s)_{s\geq 0}$, $Y_n$ has law $\nu_{s_2(T_n)}$, consequently $Y_t^{s_2}$ has law $\nu_{s_2(T_{N_t})}$.

We also need to define $T_0 = 0$ and to let $Y_0$ be a random variable with law $\nu_1$, independent of all the other random variables and processes.

This convergence in law is the object of the next theorem in which we will take

$$s_1(t) = s_2(t) = s_t = (\ln(1+t))^{-1}.$$

So define $\Theta_t^0$ the solution started at $\theta_0$ of the Itô equation

$$\mathrm{d}\Theta_t^0 = \sigma(\Theta_t^0)\,\mathrm{d}B_t - \beta_t \left( \int_M \operatorname{grad}_{\Theta_t^0} \kappa_{s_t}(\cdot, y)\,\nu_{s_t}(y)\mathrm{d}y \right) \mathrm{d}t. \tag{4.7}$$

Notice that using Lemma 4.1, (4.7) rewrites as

$$\mathrm{d}\Theta_t^0 = \sigma(\Theta_t^0)\,\mathrm{d}B_t - \beta_t \operatorname{grad}_{\Theta_t^0} U_{2s_t}\,\mathrm{d}t, \tag{4.8}$$

where $U_{2s_t} := U_{0,2s_t}$, so that the same equation with fixed $(\beta, s)$ instead of $(\beta_t, s_t)$ has an invariant law with density

$$\mu_{\beta,s}(\theta) = \frac{1}{Z_{\beta,s}} \mathrm{e}^{-2\beta U_{2s}(\theta)}, \quad \text{with} \quad Z_{\beta,s} = \int_M \mathrm{e}^{-2\beta U_{2s}(\theta')}\,\mathrm{d}\theta'. \tag{4.9}$$

The process $\Theta_t^0$ is an inhomogeneous diffusion with generator

$$L_t^0(\theta) = \frac{1}{2}\Delta(\theta) - \beta_t \operatorname{grad}_\theta U_{2s_t}. \tag{4.10}$$

Denote by $m_t(\theta)$ the density of $\Theta_t$.

Let $Y_t := Y_t^s$. The process $(\Theta_t, Y_t)$ is Markovian with generator $L_t$ given by

$$\begin{aligned} L_t f(\theta, y) &= \left( \frac{1}{2}\Delta(\theta) - \beta_t \operatorname{grad}_\theta \kappa_{s_t}(\cdot, y) \right) f(\cdot, y) + \gamma_t^{-1} \int_M (f(\theta, z) - f(\theta, y))\ \nu_{s_t}(\mathrm{d}z) \\ &= L_{1,t} f(\cdot, y)(\theta) + L_{2,t} f(\theta, \cdot)(y). \end{aligned} \tag{4.11}$$

We know that for all neighbourhood $\mathcal{N}$ of $Q_{p,\nu}$, $\int_{\mathcal{N}} \mu_{\beta,s}(\theta)\,\mathrm{d}\theta$ converges to 1 as $\beta \to \infty$, uniformly in $s$ sufficiently small (depending on $\mathcal{N}$). Again define

$$J_t := \int_M \ln\left( \frac{m_t(\theta)}{\mu_{\beta_t,s_t}(\theta)} \right) m_t(\theta)\,\mathrm{d}\theta. \tag{4.12}$$

**Theorem 4.3.** *Assume*

$$\beta_t = \frac{1}{k}\ln(1+t),\ \gamma_t = (1+t)^{-1},\ s_1(t) = s_2(t) = s(t) = (\ln(1+t))^{-1}. \tag{4.13}$$

*where* $k > c(U)$, ($c(U)$ *defined in* (3.3)). *Then for any neighbourhood* $\mathcal{N}$ *of* $Q_{p,\nu}$, *the process* $\Theta_t$ *defined in equation* (4.6) *satisfies*

$$\lim_{t\to\infty} \mathbb{P}\left[\Theta_t \in \mathcal{N}\right] = 1. \tag{4.14}$$

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

*Proof.* We use Proposition 4.4 below together with the fact that

$$\|m_t - \mu_{\beta_t,s_t}\| \leq 4\sqrt{2J_t}$$

and

$$\lim_{t\to\infty} \mu_{\beta_t,s_t}(\mathcal{N}) = 1. \qquad \square$$

**Proposition 4.4.** *The entropy*

$$J_t = \int_M \ln\left(\frac{m_t(\theta)}{\mu_{\beta_t,s_t}(\theta)}\right) m_t(\theta)\,\mathrm{d}\theta \tag{4.15}$$

*converges to* 0 *as* $t \to \infty$.

*Proof.* As for proposition 3.1, we split the proof into 3 steps.

**Step 1.** Let us establish

$$\begin{aligned}\frac{\mathrm{d}J_t}{\mathrm{d}t} &\leq \frac{C}{(1+t)k}\,(1+\ln(1+t)) \\ &\quad - c_2(\beta_t \vee 1)^{-p}\exp\left(-c(U_{2s_t})\beta_t\right) J_t + \beta_t^2 32K^2 I_t\end{aligned} \tag{4.16}$$

with $c_2, K$ defined in (4.26) and (4.5), and where

$$I_t = \int_{M\times M} \ln\left(\frac{m_t(y|\theta)}{\nu_{s_t}(y)}\right) m_t(\theta,y)\,\mathrm{d}y, \tag{4.17}$$

$m_t(y|\theta)$ being the density of $Y^s$ conditioned by $\Theta_t = \theta$.

Let us compute as before

$$\frac{\mathrm{d}J_t}{\mathrm{d}t} = -\int_M \partial_t \ln(\mu_{\beta_t,s_t}(\theta)) m_t(\theta))\,\mathrm{d}\theta + \int_M L_t\left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t,s_t}(\theta)}\right)\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y. \tag{4.18}$$

For the first term in the right we have using (4.9)

$$\begin{aligned}&\partial_t \ln(\mu_{\beta_t,s_t}(\theta)) \\ &= -2\beta_t' U_{2s_t} - 2\beta_t \int_{M\times M} 2s_t' \partial_s \ln p(2s_t,\theta,z) p(2s_t,\theta,z)\rho^p(z,y)\,\nu(\mathrm{d}y)\mathrm{d}z \\ &\quad + 2\beta_t' \int_M U_{2s_t}(\theta')\mu_{\beta_t,s_t}(\theta')\,\mathrm{d}\theta' \\ &\quad + 2\beta_t \int_M \left(\int_{M\times M} 2s_t' \partial_s \ln p(2s_t,\theta',z) p(2s_t,\theta',z)\rho^p(z,y)\,\mathrm{d}z\nu(\mathrm{d}y)\right)\mu_{\beta_t,s_t}(\theta')\,\mathrm{d}\theta'.\end{aligned} \tag{4.19}$$

It is known that there exists $C_0 > 0$ such that $\forall s \in (0,1]$

$$|\partial_s \ln p(s,\theta,z)| \leq \frac{C_0}{s^2}, \tag{4.20}$$

see *e.g.* [14,22] where bounds of the type $|\operatorname{grad}_\theta \ln p(s,\theta,z)| \leq \frac{C_1}{s}$ and $|\operatorname{grad}_\theta^2 \ln p(s,\theta,z)| \leq \frac{C_2}{s^2}$ are given. Here we use

$$|\partial_s \ln p(s,\theta,z)| = \frac{1}{2}\left|\frac{\Delta_\theta p(s,\theta,z)}{p(s,\theta,z)}\right| \leq \frac{\dim M}{2}\left(|\operatorname{grad}_\theta^2 \ln p(s,\theta,z)| + |\operatorname{grad}_\theta \ln p(s,\theta,z)|^2\right).$$

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

So (4.19) and (4.20) yield

$$|\partial_t \ln(\mu_{\beta_t,s_t}(\theta))| \leq D^p \left(4\beta_t' + \frac{C_0\beta_t|s_t'|}{s_t^2}\right). \tag{4.21}$$

which implies

$$|\partial_t \ln(\mu_{\beta_t,s_t}(\theta))| \leq C \left(\beta_t' + \frac{\beta_t|s_t'|}{s_t^2}\right). \tag{4.22}$$

with

$$C = D^p(4 + C_0). \tag{4.23}$$

Evaluating with (4.13) and integrating on $M$ we get

$$\left| - \int_M \partial_t \ln(\mu_{\beta_t,s_t}(\theta)) m_t(\theta)\, \mathrm{d}\theta \right| \leq \frac{C}{(1+t)k}\,(1 + \ln(1+t)). \tag{4.24}$$

Now we split the second term in the right of (4.18) into

$$\begin{aligned}
&\int_M L_t \left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t,s_t}(\theta)}\right)\right] m_t(\theta,y)\, \mathrm{d}\theta \mathrm{d}y \\
&= \int_M L_t^0 \left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t,s_t}(\theta)}\right)\right] m_t(\theta)\, \mathrm{d}\theta + \int_M R_t(\theta,y) \left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t,s_t}(\theta)}\right)\right] m_t(\theta,y)\, \mathrm{d}\theta \mathrm{d}y.
\end{aligned} \tag{4.25}$$

We have as for (3.20)

$$\begin{aligned}
\int_M L_t^0 \left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t,s_t}(\theta)}\right)\right] m_t(\theta)\, \mathrm{d}\theta &= -2 \int_M \left\| \mathrm{grad}_\theta \sqrt{\frac{m_t(\theta)}{\mu_{\beta_t,s_t}(\theta)}} \right\|^2 \mu_{\beta_t,s_t}(\theta)\, \mathrm{d}\theta \\
&\leq -2c_2(\beta_t \vee 1)^{-p} \exp\left(-c(U_{2s_t})\beta_t\right) J_t
\end{aligned} \tag{4.26}$$

for some $c_2 > 0$ and integer $p > 0$ by logarithmic Sobolev inequality [20].

The computation for the second term is similar to the one after (3.20) and we get

$$\int_M R_t(\theta,y) \left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t,s_t}(\theta)}\right)\right] m_t(\theta,y)\, \mathrm{d}\theta \mathrm{d}y = 2\beta_t \int_M \sqrt{\frac{\mu_{\beta_t,s_t}}{m_t}(\theta)} \left\langle d\sqrt{\frac{m_t}{\mu_{\beta_t,s_t}}}(\theta), R_t(\theta) \right\rangle m_t(\theta)\, \mathrm{d}\theta$$

with

$$R_t(\theta) = -\int_M \mathrm{grad}_\theta\, \kappa_{s_t}(\cdot,y)(m_t(y|\theta) - \nu_{s_t}(y))\, \mathrm{d}y,$$

and again

$$\begin{aligned}
&\int_M R_t(\theta,y) \left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t,s_t}(\theta)}\right)\right] m_t(\theta)\, \mathrm{d}\theta \\
&\leq \beta_t^2 \int_M \|R_t(\theta)\|^2 m_t(\theta)\, \mathrm{d}\theta + \int_M \left\| \mathrm{grad}_\theta \sqrt{\frac{m_t}{\mu_{\beta_t,s_t}}(\theta)} \right\|^2 \mu_{\beta_t,s_t}(\theta)\, \mathrm{d}\theta.
\end{aligned}$$

Summing with (4.26) we get

$$\begin{aligned}
&\int_{M\times M} L_t \left[\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t,s_t}(\theta)}\right)\right] m_t(\theta,y))\, \mathrm{d}\theta \mathrm{d}y \\
&\leq \beta_t^2 \int_M \|R_t(\theta)\|^2 m_t(\theta)\, \mathrm{d}\theta - c_2(\beta_t \vee 1)^{-p} \exp\left(-c(U_{2s_t})\beta_t\right) J_t.
\end{aligned} \tag{4.27}$$

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

Here again

$$\int_M \|R_t(\theta)\|^2 m_t(\theta)\,\mathrm{d}\theta \le 32K^2 I_t$$

where $I_t$ is defined in (4.17). At this stage we proved (4.16).

**Step 2.** Let us establish

$$\frac{\mathrm{d}I_t}{\mathrm{d}t} \le \frac{C_0}{(1+t)} + \frac{K'}{k^2}(\ln(1+t)\vee k)(\ln(1+t))^3 - \frac{\mathrm{d}J_t}{\mathrm{d}t} - (1+t)I_t \tag{4.28}$$

for some $K'$ defined below.

As before

$$\frac{\mathrm{d}I_t}{\mathrm{d}t} = -\int_{M\times M} \partial_t \ln(\nu_{s_t}(y)) m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y + \int_{M\times M} L_t\left[\ln\left(\frac{m_t(y|\theta)}{\nu_{s_t}(y)}\right)\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y \tag{4.29}$$

and

$$\begin{aligned}
\left|\int_{M\times M} \partial_t \ln(\nu_{s_t}(y)) m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y\right| &= \left|\int_{M\times M} \partial_t\left(\int_M p(s_t,y,z)\nu(\mathrm{d}z)\right)\frac{m_t(\theta,y)}{\nu_{s_t}(y)}\,\mathrm{d}\theta\mathrm{d}y\right| \\
&\le \int_{M\times M}\left(\int_M |\partial_t \ln p(s_t,y,z)|\, p(s_t,y,z)\nu(\mathrm{d}z)\right)\frac{m_t(\theta,y)}{\nu_{s_t}(y)}\,\mathrm{d}\theta\mathrm{d}y \\
&\le \frac{|s_t'|C_0}{s_t^2}\int_{M\times M}\left(\int_M p(s_t,y,z)\nu(\mathrm{d}z)\right)\frac{m_t(\theta,y)}{\nu_{s_t}(y)}\,\mathrm{d}\theta\mathrm{d}y \\
&= \frac{|s_t'|C_0}{s_t^2}\int_{M\times M}\nu_{s_t}(y)\frac{m_t(\theta,y)}{\nu_{s_t}(y)}\,\mathrm{d}\theta\mathrm{d}y \\
&= \frac{|s_t'|C_0}{s_t^2} = \frac{C_0}{1+t}
\end{aligned} \tag{4.30}$$

where we used (4.20) for the last inequality. Now

$$\int_{M\times M} L_t\left[\ln\left(\frac{m_t(y|\theta)}{\nu_{s_t}(y)}\right)\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y = \int_{M\times M}(L_{2,t}+L_{1,t})\left[\ln\left(\frac{m_t(y|\theta)}{\nu_{s_t}(y)}\right)\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y. \tag{4.31}$$

We begin with the first term:

$$\int_{M\times M} L_{2,t}\left[\ln\left(\frac{m_t(y|\theta)}{\nu_{s_t}(y)}\right)\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y = \gamma_t^{-1}\int_{M\times M}\ln\left(\frac{m_t(y|\theta)}{\nu_{s_t}(y)}\right)(\nu_{s_t}(y) - m_t(y|\theta))\, m_t(\theta)\,\mathrm{d}\theta$$

and estimate it as for (3.27):

$$\int_{M\times M} L_{2,t}\left[\ln\left(\frac{m_t(y|\theta)}{\nu_{s_t}(y)}\right)\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y \le -\gamma_t^{-1} I_t. \tag{4.32}$$

For the second term in the right of (4.29) we need to introduce the density $f_t(y)$ of $Y_t^s$. Since

$$\ln\left(\frac{m_t(y|\theta)}{\nu_{s_t}(y)}\right) = \ln\left(\frac{m_t(\theta|y)}{m_t(\theta)}\right) + \ln\left(\frac{f_t(y)}{\nu_{s_t}(y)}\right)$$

and the last term does not depend on $\theta$, it rewrites as

$$\int_{M\times M} L_{1,t}\left[\ln\left(\frac{m_t(\theta|y)}{m_t(\theta)}\right)\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y = \int_{M\times M} L_{1,t}\left[\ln(m_t(\theta|y)) - \ln(m_t(\theta))\right] m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y. \tag{4.33}$$

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

Similarly to (3.29)

$$\begin{aligned}
&\int_{M\times M} L_{1,t}\ln(m_t(\theta|y))m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y\\
&=\frac{1}{2}\int_{M\times M}\Delta\ln(m_t(\theta|y))m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y-\beta_t\int_{M\times M}\langle\mathrm{grad}_\theta\ln m_t(\cdot|y),\mathrm{grad}_\theta\,\kappa_{s_t}(\cdot,y)\rangle\,m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y\\
&=-2\int_{M\times M}\left\|\mathrm{grad}_\theta\sqrt{m_t(\theta|y)}\right\|^2\mathrm{d}\theta f_t(y)\mathrm{d}y\\
&\quad-\beta_t\int_{M\times M}\langle\mathrm{grad}_\theta\,m_t(\cdot|y),\mathrm{grad}_\theta\,\kappa_{s_t}(\cdot,y)\rangle\,f_t(y)\,\mathrm{d}\theta\mathrm{d}y. \qquad (4.34)
\end{aligned}$$

For the absolute value of the last term:

$$\begin{aligned}
&\left|-\beta_t\int_{M\times M}\langle\mathrm{grad}_\theta\,m_t(\cdot|y),\mathrm{grad}_\theta\,\kappa(\cdot,y)\rangle\;f_t(y)\,\mathrm{d}\theta\mathrm{d}y\right|\\
&\le\frac{1}{2}\beta_t^2K^2+2\int_{M\times M}\left\|\mathrm{grad}_\theta\sqrt{m_t(\theta|y)}\right\|^2 f_t(y)\,\mathrm{d}\theta\mathrm{d}y. \qquad (4.35)
\end{aligned}$$

We get as in (3.32)

$$\int_{M\times M} L_{1,t}\ln(m_t(\theta|y))m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y\le\frac{1}{2}\beta_t^2K^2 \qquad (4.36)$$

Then we bound the last term in (4.33):

$$-L_{1,t}\ln(m_t(\theta))=-L_{1,t}\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t,s_t}(\theta)}\right)-L_{1,t}\ln(\mu_{\beta_t,s_t}(\theta)). \qquad (4.37)$$

We already know by (4.18) and (4.24) that

$$-\int_{M\times M}L_{1,t}\ln\left(\frac{m_t(\theta)}{\mu_{\beta_t,s_t}(\theta)}\right)m_t(\theta,y)\,\mathrm{d}\theta\mathrm{d}y\le-\frac{\mathrm{d}J_t}{\mathrm{d}t}+\frac{C}{(1+t)k}\left(1+\ln(1+t)\right). \qquad (4.38)$$

For the second term we have

$$\begin{aligned}
L_{1,t}\ln(\mu_{\beta_t,s_t}(\theta))&=-2\beta_tL_{1,t}U_{2s_t}(\theta)\\
&=-\beta_t\Delta U_{2s_t}(\theta)+2\beta_t^2\,\langle dU_{2s_t},\mathrm{grad}_\theta\,\kappa_{s_t}(\cdot,y)\rangle\\
&\le K'(\beta_t\vee1)\beta_ts_t^{-2} \qquad (4.39)
\end{aligned}$$

for some $K'>0$, where we used

$$\Delta U_{2s}=\int_M\left(\Delta_\theta\ln p(2s,\theta,y)+\|\,\mathrm{grad}_\theta\ln p(2s,\theta,y)\|^2\right)p(2s,\theta,y)\rho^p(y,z)\,\nu(\mathrm{d}z)$$

and standard bounds for the first and second derivatives of the heat kernel ([14] and [22] and the explanation after (4.20)).

Finally we obtain (4.28). This together with (4.16) yields:

$$\frac{\mathrm{d}J_t}{\mathrm{d}t}\le\frac{C}{(1+t)k}\left(1+\ln(1+t)\right)-c_2(\beta_t\vee1)^{-p}\exp\left(-c(U_{2s_t})\beta_t\right)J_t+2\beta_t^232K^2I_t \qquad (4.40)$$

which rewrites as

$$\frac{\mathrm{d}I_t}{\mathrm{d}t}\le k_1(\ln(1+t))^4-\frac{\mathrm{d}J_t}{\mathrm{d}t}-(1+t)I_t \qquad (4.41)$$

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

and

$$\frac{\mathrm{d}J_t}{\mathrm{d}t} \leq c_1 \left( \frac{\ln(1+t)}{1+t} + (\ln(1+t))^2 I_t \right) - c_2 (\ln(1+t))^{-p} (1+t)^{-\frac{c(U_{2s_t})}{k}} J_t \tag{4.42}$$

for some constants $c_1, k_1 > 0$, as soon as $t \geq 2$.

**Step 3.** Let us finally prove that

$$\lim_{t\to\infty} J_t = 0. \tag{4.43}$$

We can use a similar computation to the end of the proof of theorem 1 in [21] to obtain it under assumptions (4.13) and (3.3). However we will do the calculation for completeness, and because there are some small differences. Recall $U_s \to U$ uniformly as $s \to 0$. Moreover $2s_t \to 0$ as $t \to \infty$, so we get

$$\limsup_{t\to\infty} c(U_{2s_t}) \leq c(U).$$

As a consequence, for $t$ sufficiently large we have

$$\frac{c(U_{2s_t})}{k} \leq 1 - \varepsilon \tag{4.44}$$

for some $\varepsilon > 0$. Let

$$\ell_t = \frac{c_1 (\ln(1+t))^2}{1 + t + c_1 (\ln(1+t))^2 - c_2 (\ln(1+t))^{-p} (1+t)^{-(1-\varepsilon)}} \tag{4.45}$$

where $\varepsilon > 0$ is defined in (4.44). It is easily checked that for $t$ sufficiently large $\ell_t$ is positive and decreasing, and that it converges to 0 as $t \to \infty$. Define

$$K_t = J_t + \ell_t I_t. \tag{4.46}$$

We will prove that $K_t \to 0$ as $t \to \infty$ and from this we will get (4.43).

for $t$ sufficiently large,

$$\frac{\mathrm{d}K_t}{\mathrm{d}t} \leq \frac{\mathrm{d}J_t}{\mathrm{d}t} + \ell_t \frac{\mathrm{d}I_t}{\mathrm{d}t} \tag{4.47}$$

and this yields with (4.41) and (4.42)

$$\begin{aligned} \frac{\mathrm{d}K_t}{\mathrm{d}t} \leq{}& (1-\ell_t) c_1 \frac{\ln(1+t)}{1+t} + c_1 (\ln(1+t))^2 I_t \\ & - \ell_t c_1 (\ln(1+t))^2 I_t - (1-\ell_t) c_2 (\ln(1+t))^{-p} (1+t)^{-\frac{c(U_{2s_t})}{k}} J_t \\ & + \ell_t k_1 (\ln(1+t))^4 - (1+t)\ell_t I_t. \end{aligned}$$

Replacing $c_1(\ln(1+t))^2$ at the end of the first line by

$$\ell_t \left( 1 + t + c_1 (\ln(1+t))^2 - c_2 (\ln(1+t))^{-p} (1+t)^{-(1-\varepsilon)} \right)$$

by the help of (4.45) we obtain

$$\begin{aligned} \frac{\mathrm{d}K_t}{\mathrm{d}t} \leq{}& c_1 \frac{\ln(1+t)}{1+t} - c_2 \ell_t (\ln(1+t))^{-p} (1+t)^{-\frac{c(U_{2s_t})}{k}} I_t \\ & - (1-\ell_t) c_2 (\ln(1+t))^{-p} (1+t)^{-\frac{c(U_{2s_t})}{k}} J_t + \ell_t k_1 (\ln(1+t))^4 \end{aligned}$$

and this yields using $-(1+t)^{-\frac{c(U_{2s_t})}{k}} \leq -(1+t)^{-(1-\varepsilon)}$:

$$\frac{\mathrm{d}K_t}{\mathrm{d}t} \leq A_t - B_t K_t \tag{4.48}$$

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

with

$$A_t = c_1 \frac{\ln(1+t)}{1+t} + \ell_t k_1 \ln(1+t))^4 \tag{4.49}$$

and

$$B_t = (1-\ell_t)c_2(\ln(1+t))^{-p}(1+t)^{-(1-\varepsilon)}. \tag{4.50}$$

A sufficient condition for $K_t$ to converge to 0 as $t \to \infty$ is

$$\int_{\cdot}^{\infty} B_t \, \mathrm{d}t = +\infty \tag{4.51}$$

and

$$\lim_{t\to\infty} \frac{A_t}{B_t} = 0. \tag{4.52}$$

Condition (4.51) clearly is realized. As for condition (4.52) we easily see that

$$\frac{c_1 \frac{\ln(1+t)}{1+t}}{(1-\ell_t)c_2(\ln(1+t))^{-p}(1+t)^{-(1-\varepsilon)}} \to 0$$

and also

$$\frac{\ell_t k_1 (\ln(1+t))^4}{(1-\ell_t)c_2(\ln(1+t))^{-p}(1+t)^{-(1-\varepsilon)}} \to 0$$

from the fact that

$$\ell_t \leq \frac{c(\ln(1+t))^2}{1+t}$$

for some $c > 0$. □

## References


[1] B. Afsari, Riemannian $L^p$ center of mass: existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* S 0002–9939 (2010) 10541-5. (electronic)

[2] B. Afsari, R. Tron and R. Vidal, On the convergence of gradient descent for finding the Riemannian center of mass. arXiv:1201.0925.

[3] M. Arnaudon and F. Nielsen, Medians and means in Finsler geometry. *LMS J. Comput. Math.* **15** (2012) 23–37.

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

[5] M. Arnaudon and F. Nielsen, On computing the Riemannian 1-Center. *Comput. Geom.* **46** (2013) 93–104.

[6] M. Bădoiu and K.L. Clarkson, Smaller core-sets for balls, *Proc. of the fourteenth Annual ACM-SIAM Symposium on Discrete algorithms*. Soc. Industrial Appl. Math. Philadelphia, PA, USA (2003) 801–802.

[7] R. Bhattacharya and V. Patrangenaru, Large sample theory of intrinsic and extrinsic sample means on manifolds (i). *Ann. Statis.* **31**(2003) 1–29.

[8] S. Bonnabel, *Convergence des méthodes de gradient stochastique sur les variétés riemanniennes*. In GRETSI, Bordeaux (2011).

[9] H. Cardot, P. Cénac and P.-A. Zitt, *Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm*, Bernoulli.

[10] B. Charlier, Necessary and sufficient condition for the existence of a Fréchet mean on the circle. arXiv:1109.1986.

[11] P.T. Fletcher, S. Venkatasubramanian and S. Joshi, The geometric median on Riemannian manifolds with application to robust atlas estimation. *NeuroImage* **45** (2009) S143–S152.

[12] D. Groisser, Newton's method, zeroes of vector fields, and the Riemannian center of mass. *Adv. Appl. Math.* **33** (2004) 95–135.

[13] D. Groisser, On the convergence of some Procrustean averaging algorithms. *Stochastics* **77** (2005) 31–60.

[14] E.P. Hsu, Estimates of derivatives of the heat kernel on a compact Riemannian manifold. *Proc. Amer. Math. Soc.* **127** (1999) 3739–3744.

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

[16] R. Holley and D. Stroock, Annealing via Sobolev inequalities. *Commun. Math. Phys.* **115** (1988) 553–569.

[17] T. Hotz and S. Huckemann, Intrinsic mean on the circle: Uniqueness, Locus and Asymptotics. arXiv:org1108:2141.


https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press

[18] W.S. Kendall, Probability, convexity and harmonic maps with small image I: uniqueness and fine existence. *Proc. London Math. Soc.* **61** (1990) 371–406.
[19] H. Le, Estimation of Riemannian barycentres. *LMS J. Comput. Math.* **7** (2004) 193–200.
[20] L. Miclo, Recuit simulé sans potentiel sur une variété compacte. *Stoch. and Stochastic Reports* **41** (1992) 23–56.
[21] L. Miclo, Recuit simulé partiel, *Stoch. Process. Appl.* **65** (1996) 281–298.
[22] S.J. Sheu, Some estimates of the transition density function of a nondegenerate diffusion Markov process. *Ann. Probab.* **19** (1991) 538–561.
[23] K.T. Sturm, Probability measures on metric spaces of nonpositive curvature, Heat kernels and analysis on manifolds, graphs, and metric spaces (Paris, 2002). *Contemp. Math. Amer. Math. Soc.* **338** (2003) 357–390.
[24] E. Weiszfeld, Sur le point pour lequel la somme des distances de $n$ points donnés est minimum. *Tohoku Math. J.* **43** (1937) 355–386.
[25] L.Yang, Riemannian median and its estimation. *LMS J. Comput. Math.* **13** (2010) 461–479.
[26] L.Yang, *Some properties of Frechet medians in Riemannian manifolds*. Preprint.

https://doi.org/10.1051/ps/2013033 Published online by Cambridge University Press