1. Introduction

The geometric barycenter of a set of points is the point which minimizes the sum of the distances at the power 2 to these points. It is the most common estimator in statistics, however it is sensitive to outliers, and it is natural to replace power 2 by \( p \) for some \( p \in [1, 2) \), which leads to the definition of \( p \)-mean. When \( p = 1 \), the minimizer is the median of the set of points, very often used in robust statistics. In many applications, \( p \)-means with some \( p \in (1, 2) \) give the best compromise.

The Fermat-Weber problem concerns finding the median \( e_1 \) of a set of points in an Euclidean space. Numerous authors worked out algorithms for computing \( e_1 \). The first algorithm was proposed by Weiszfeld in \([21]\). It has been extended to sufficiently small domains in Riemannian manifolds with nonnegative curvature by Fletcher and al in \([7]\). A complete generalization to manifolds with positive or negative curvature, including existence and uniqueness results (under some convexity conditions in positive curvature), has been given by one of the authors in \([22]\).

The Riemannian barycenter or Karcher mean of a set of points in a manifold or more generally of a probability measure has been extensively studied, see e.g. \([8]\),
where questions of existence, uniqueness, stability, relation with martingales in manifolds, behaviour when measures are pushed by stochastic flows have been considered. The Riemannian barycenter corresponds to $p = 2$ in the above description. Computation of Riemannian barycenters by gradient descent has been performed by Le in [13].

In [1] Afsari proved existence and uniqueness of $p$-means, $p \geq 1$ on geodesic balls with radius $r < \frac{1}{2} \min \left\{ \text{inj}(M), \frac{\pi}{2\alpha} \right\}$ if $p \in [1, 2)$, and $r < \frac{1}{2} \min \left\{ \text{inj}(M), \frac{\pi}{\alpha} \right\}$ if $p \geq 2$. Here $\text{inj}(M)$ is the injectivity radius of $M$ and $\alpha > 0$ is such that the sectional curvatures in $M$ are bounded above by $\alpha^2$. The point is that in the case $p \geq 2$, the functional to minimize is not convex any more, which makes the situation much more difficult to handle.

In this paper, under the assumptions of [1] we provide in Theorem 2.3 stochastic algorithms which converge almost surely to $p$-means in manifolds, which are easier to implement than gradient descent algorithm since computing the gradient of the function to minimize is not needed. The idea is at each step to go in the direction of a point of the support of $\mu$. The point is chosen at random according to $\mu$ and the size of the step is a well chosen function of the distance to the point, $p$ and the number of the step. For general convergence results on recursive stochastic algorithms, see [14] Theorem 1. However they do not cover the manifold case and nonlinearity of geodesics. Here we give a proof using martingale convergence theorem, and the main point consists in determining and estimating all the geometric quantities, checking that under our curvature conditions all the convergence assumptions are fulfilled, since our processes live in manifolds. See also [3] for convergence in probability of recursive algorithms.

The speed of convergence is studied, and in theorem 2.6 we prove that the renormalized inhomogeneous Markov chain of Theorem 2.3 converges in law to an inhomogeneous diffusion process. This is an invariance principle type result, see e.g. [9], [15], [4], [6] for related works. Here again the main point is to obtain the characteristics of the limiting process from the curvature conditions, the conditions on the support of the measure and estimates on Jacobi fields. Moreover we consider convergence in law for the Skorohod topology, and the limit depends in a crucial way on the decreasing steps of the algorithms.

2. Results

2.1. $p$-means in regular geodesic balls. Let $M$ be a Riemannian manifold with pinched sectional curvatures. Let $\alpha, \beta > 0$ such that $\alpha^2$ is a positive upper bound for sectional curvatures on $M$, and $-\beta^2$ is a negative lower bound for sectional curvatures on $M$. Denote by $\rho$ the Riemannian distance on $M$.

In $M$ consider a geodesic ball $B(a, r)$ with $a \in M$. Let $\mu$ be a probability measure with support included in a compact convex subset $K_{\mu}$ of $B(a, r)$. Fix $p \in [1, \infty)$. We will always make the following assumptions on $(r, p, \mu)$:

**Assumption 2.1.** The support of $\mu$ is not reduced to one point. Either $p > 1$ or the support of $\mu$ is not contained in a line, and the radius $r$ satisfies

\[
(2.1) \quad r < r_{\alpha, p} \quad \text{with} \quad \begin{cases} r_{\alpha, p} = \frac{1}{2} \ min \ \{ \text{inj}(M), \frac{\pi}{2\alpha} \} & \text{if} \ p \in [1, 2) \ \\
 r_{\alpha, p} = \frac{1}{2} \ min \ \{ \text{inj}(M), \frac{\pi}{\alpha} \} & \text{if} \ p \in [2, \infty) \ 
\end{cases}
\]

Note that $B(a, r)$ is convex if $r < \frac{1}{2} \ min \ \{ \text{inj}(M), \frac{\pi}{\alpha} \}$.
Under assumption [2.1] it has been proved in [1] (Theorem 2.1) that the function
\( H_p : M \to \mathbb{R}_+ \)
\( x \mapsto \int_M \rho^p(x, y)\mu(dy) \)
has a unique minimizer \( e_p \) in \( M \), the \( p \)-mean of \( \mu \), and moreover \( e_p \in B(a, r) \). If \( p = 1 \), \( e_1 \) is the median of \( \mu \).

It is easily checked that if \( p \in [1, 2) \), then \( H_p \) is strictly convex on \( B(a, r) \). On the other hand, if \( p \geq 2 \) then \( H_p \) is of class \( C^2 \) on \( B(a, r) \).

**Proposition 2.2.** Let \( K \) be a convex subset of \( B(a, r) \) containing the support of \( \mu \). Then there exists \( C_{p, \mu, K} > 0 \) such that for all \( x \in K \),
\[ H_p(x) - H_p(e_p) \geq \frac{C_{p, \mu, K}}{2} \rho(x, e_p)^2. \]
Moreover if \( p \geq 2 \) then we can choose \( C_{p, \mu, K} \) so that for all \( x \in K \),
\[ \|\nabla_x H_p\|^2 \geq C_{p, \mu, K} (H_p(x) - H_p(e_p)). \]

In the sequel, we fix
\[ K = \overline{B}(a, r - \varepsilon) \quad \text{with} \quad \varepsilon = \frac{\rho(K, B(a, r))}{2}. \]

We now state our main result: we define a stochastic gradient algorithm \( (X_k)_{k \geq 0} \) to approximate the \( p \)-mean \( e_p \) and prove its convergence.

**Theorem 2.3.** Let \( (P_k)_{k \geq 1} \) be a sequence of independent \( B(a, r) \)-valued random variables, with law \( \mu \). Let \( (t_k)_{k \geq 1} \) be a sequence of positive numbers satisfying
\[ \forall k \geq 1, \quad t_k \leq \min \left( \frac{1}{C_{p, \mu, K}}, \frac{\rho(K, B(a, r))}{2p(2r)^{p-1}} \right), \]
\[ \sum_{k=1}^{\infty} t_k = +\infty \quad \text{and} \quad \sum_{k=1}^{\infty} t_k^2 < \infty. \]

Letting \( x_0 \in K \), define inductively the random walk \( (X_k)_{k \geq 0} \) by
\[ X_0 = x_0 \quad \text{and for} \quad k \geq 0 \quad X_{k+1} = \exp_{X_k} \left( -t_{k+1} \nabla_x F_p(\cdot, P_{k+1}) \right) \]
where \( F_p(x, y) = \rho^p(x, y) \), with the convention \( \nabla_x F_p(\cdot, x) = 0 \).

The random walk \( (X_k)_{k \geq 1} \) converges in \( L^2 \) and almost surely to \( e_p \).

In the following example, we focus on the case \( M = \mathbb{R}^d \) and \( p = 2 \) where drastic simplifications occur.

**Example 2.4.** In the case when \( M = \mathbb{R}^d \) and \( \mu \) is a compactly supported probability measure on \( \mathbb{R}^d \), the stochastic gradient algorithm \( (2.3) \) simplifies into
\[ X_0 = x_0 \quad \text{and for} \quad k \geq 0 \quad X_{k+1} = X_k - t_{k+1} \nabla_x F_p(\cdot, P_{k+1}). \]
If furthermore \( p = 2 \), clearly \( e_2 = E[P_1] \) and \( \nabla_x F_p(\cdot, y) = 2(x - y) \), so that the linear relation
\[ X_{k+1} = (1 - 2t_{k+1})X_k + 2t_{k+1}P_{k+1}, \quad k \geq 0 \]
holds true and an easy induction proves that
\[ X_k = x_0 \prod_{j=0}^{k-1} (1 - 2t_{k-j}) + 2 \sum_{j=0}^{k-1} P_{k-j} t_{k-j} \prod_{\ell=0}^{j-1} (1 - 2t_{k-\ell}), \quad k \geq 1. \]
Now, taking $t_k = \frac{1}{2^k}$, we have
\[
\prod_{j=0}^{k-1} (1 - 2t_{k-j}) = 0 \quad \text{and} \quad \prod_{\ell=0}^{j-1} (1 - 2t_{k-\ell}) = \frac{k-j}{k}
\]
so that
\[
X_k = \sum_{j=0}^{k-1} P_{k-j} \frac{1}{k} = \frac{1}{k} \sum_{j=1}^{k} P_j.
\]

The stochastic gradient algorithm estimating the mean $e_2$ of $\mu$ is given by the empirical mean of a growing sample of independent random variables with distribution $\mu$. In this simple case, the result of Theorem 2.3 is nothing but the strong law of large numbers. Moreover, fluctuations around the mean are given by the central limit theorem and Donsker’s theorem.

2.2. Fluctuations of the stochastic gradient algorithm. The notations are the same as in the beginning of section 2.1. We still make assumption 2.1. Let us define $K$ and $\varepsilon$ as in (2.5) and let
\[
\delta_1 = \min \left( \frac{1}{C_{p,\mu,K}}, \frac{\rho(K, B(a, r)^c)}{2p(2r)^{p-1}} \right).
\]

We consider the time inhomogeneous $M$-valued Markov chain (2.8) in the particular case when
\[
t_k = \min \left( \frac{\delta}{k}, \delta_1 \right), \quad k \geq 1
\]
for some $\delta > 0$. The particular sequence $(t_k)_{k \geq 1}$ defined by (2.11) satisfies (2.6) and (2.7), so Theorem 2.3 holds true and the stochastic gradient algorithm $(X_k)_{k \geq 0}$ converges a.s. and in $L^2$ to the $p$-mean $e_p$.

In order to study the fluctuations around the $p$-mean $e_p$, we define for $n \geq 1$ the rescaled $\varepsilon_p M$-valued Markov chain $(Y_{\varepsilon}^n)_{k \geq 0}$ by
\[
Y_{\varepsilon}^n = \frac{k}{\sqrt{n}} \exp_{\varepsilon}^{-1} X_k.
\]

We will prove convergence of the sequence of process $(Y_{\varepsilon_{(p)}}^n)_{t \geq 0}$ to a non-homogeneous diffusion process. The limit process is defined in the following proposition:

**Proposition 2.5.** Assume that $H_p$ is $C^2$ in a neighborhood of $e_p$, and that $\delta > C_{p,\mu,K}^{-1}$. Define
\[
\Gamma = \mathbb{E} \left[ \text{grad}_{\varepsilon_p} F_p(\cdot, P_1) \otimes \text{grad}_{\varepsilon_p} F_p(\cdot, P_1) \right]
\]
and $G_\delta(t)$ the generator
\[
G_\delta(t)f(y) := \langle d_y f, t^{-1} (y - \delta \nabla dH_p(y, \cdot)^\sharp) \rangle + \frac{\delta^2}{2} \text{Hess}_y f (\Gamma)
\]
where $\nabla dH_p(y, \cdot)^\sharp$ denotes the dual vector of the linear form $\nabla dH_p(y, \cdot)$.

There exists a unique inhomogeneous diffusion process $(\tilde{y}_t(t))_{t>0}$ on $T_{\varepsilon_p}M$ with generator $G_\delta(t)$ and converging in probability to $0$ as $t \to 0^+$. 
The process $y_δ$ is continuous, converges a.s. to 0 as $t \to 0^+$ and has the following integral representation:

$$y_δ(t) = \sum_{i=1}^{d} t^{1-\delta\lambda_i} \int_0^t s^{\delta\lambda_i-1} (\delta \sigma dB_s, e_i) e_i, \quad t \geq 0,$$

where $B_t$ is a standard Brownian motion on $T_{e_p}M$, $\sigma \in \text{End}(T_{e_p}M)$ satisfies $\sigma \sigma^* = \Gamma$, $(e_i)_{1 \leq i \leq d}$ is an orthonormal basis diagonalizing the symmetric bilinear form $\nabla dH_{e_p}(e_p)$ and $(\lambda_i)_{1 \leq i \leq d}$ are the associated eigenvalues.

Note that the integral representation (2.14) implies that $y_δ$ is the centered Gaussian process with covariance

$$E\left[y_δ^i(t_1)y_δ^j(t_2)\right] = \frac{\delta^2 \Gamma (e_i^* \otimes e_j^*)}{\delta (\lambda_i + \lambda_j)^{-1}} t_1^{1-\delta\lambda_i} t_2^{1-\delta\lambda_j} (t_1 \wedge t_2)^{\delta(\lambda_i + \lambda_j)-1},$$

where $y_δ^i(t) = (y_δ(t), e_i), 1 \leq i, j \leq d$ and $t_1, t_2 \geq 0$.

Our main result on the fluctuations of the stochastic gradient algorithm is the following:

**Theorem 2.6.** Assume that either $e_p$ does not belong to the support of $\mu$ or $p \geq 2$. Assume furthermore that $\delta > C_{p,\mu,K}^{-1}$. The sequence of processes $(Y_{\{nt\}}^n)$ weakly converges in $\mathbb{D}((0,\infty), T_{e_p}M)$ to $y_δ$.

**Remark 2.7.** The assumption on $e_p$ implies that $H_p$ is of class $C^2$ in a neighbourhood of $e_p$. In the case $p > 1$, in the “generic” situation for applications, $\mu$ is a discrete measure and $e_p$ does not belong to its support. For $p = 1$ one has to be more careful since if $\mu$ is equidistributed in a random set of points, then with positive probability $e_1$ belongs to the support of $\mu$.

**Remark 2.8.** From section 2.1 we know that, when $p \in (1,2]$, the constant

$$C_{p,\mu,K} = p(2r)^{p-2} (\min (p-1, 2\rho \cot (2\rho r)))$$

is explicit. The constraint $\delta > C_{p,\mu,K}^{-1}$ can easily be checked in this case.

**Remark 2.9.** In the case $M = \mathbb{R}^d$, $Y^n_k = \frac{1}{\sqrt{n}} (X_k - e_p)$ and the tangent space $T_{e_p}M$ is identified to $\mathbb{R}^d$. Theorem 2.6 holds and, in particular, when $t_1 = 1$, we obtain a central limit theorem: $\sqrt{n}(X_n - e_p)$ converges as $n \to \infty$ to a centered Gaussian $d$-variate distribution (with covariance structure given by (2.15) with $t_1 = t_2 = 1$). This is a central limit theorem: the fluctuations of the stochastic gradient algorithm are of scale $n^{-1/2}$ and asymptotically Gaussian.

3. Proofs

For simplicity, let us write shortly $e = e_p$ in the proofs.

3.1. **Proof of Proposition 2.2**

For $p = 1$ this is a direct consequence of 2.2 Theorem 3.7.

Next we consider the case $p \in (1,2)$.

Let $K \subset B(a, r)$ be a compact convex set containing the support of $\mu$. Let $x \in K \setminus \{e\}$, $t = \rho(e, x)$, $u \in T_xM$ the unit vector such that $\exp_e(\rho(e, x)u) = x$,
and $\gamma_u$ the geodesic with initial speed $u : \gamma_u(0) = u$. For $y \in K$, letting $h_y(s) = \rho(\gamma_u(s), y)^p$, $s \in [0, t]$, we have since $p > 1$

$$h_y(t) = h_y(0) + th'_y(0) + \int_0^t (t - s)h''_y(s) \, ds$$

with the convention $h''_y(s) = 0$ when $\gamma_u(s) = y$. Indeed, if $y \notin \gamma([0, t])$ then $h_y$ is smooth, and if $y \in \gamma([0, t])$, say $y = \gamma(s_0)$ then $h_y(s) = |s - s_0|^p$ and the formula can easily be checked.

By standard calculation,

$$h''_y(s) \geq pp(\gamma_u(s), y)^{p-2} \times \left( (p - 1)\|\gamma_u(s)^T(y)\|^2 + \|\gamma_u(s)^N(y)\|^2 \alpha \rho(\gamma_u(s), y) \cot (\alpha \rho(\gamma_u(s), y)) \right)$$

with $\gamma_u(s)^T(y)$ (resp. $\gamma_u(s)^N(y)$) the tangential (resp. the normal) part of $\gamma_u(s)$ with respect to $n(\gamma_u(s), y) = \frac{1}{\rho(\gamma_u(s), y)} \exp^{-1}_{\gamma_u(s)}(y)$:

$$\dot{\gamma}_u(s)^T(y) = (\dot{\gamma}_u(s), n(\gamma_u(s), y))n(\gamma_u(s), y), \quad \dot{\gamma}_u(s)^N(y) = \dot{\gamma}_u(s) - \dot{\gamma}_u(s)^T(y).$$

From this we get

$$h''_y(s) \geq pp(\gamma_u(s), y)^{p-2} (\min (p - 1, 2\alpha \cot (2\alpha \rho))).$$

Now

$$H_p(\gamma_u(t'))$$

$$= \int_{B(a,r)} h_y(\gamma_u(t')) \, \mu(dy)$$

$$= \int_{B(a,r)} h_y(0) \, \mu(dy) + t' \int_{B(a,r)} h'_y(0) \, \mu(dy) + \int_0^{t'} (t' - s) \left( \int_{B(a,r)} h_y(s)^{''} \, \mu(dy) \right) \, ds$$

and $H_p(\gamma_u(t'))$ attains its minimum at $t' = 0$, so $\int_{B(a,r)} h'_y(0) \, \mu(dy) = 0$ and we have

$$H_p(x) = H_p(\gamma_u(t)) = H_p(e) + \int_0^t (t - s) \left( \int_{B(a,r)} h_y(s)^{''} \, \mu(dy) \right) \, ds.$$

Using Equation (3.2) we get

$$H_p(x) \geq H_p(e)$$

$$+ \int_0^t \left( (t - s) \int_{B(a,r)} pp(\gamma_u(s), y)^{p-2} (\min (p - 1, 2\alpha \cot (2\alpha \rho))) \, \mu(dy) \right) \, ds.$$

Since $p \leq 2$ we have $\rho(\gamma_u(s), y)^{p-2} \geq (2r)^{p-2}$ and

$$H_p(x) \geq H_p(e) + \frac{t^2}{2} (2r)^{p-2} (\min (p - 1, 2\alpha \cot (2\alpha \rho))).$$

So letting

$$C_{p,\mu,K} = p(2r)^{p-2} (\min (p - 1, 2\alpha \cot (2\alpha \rho)))$$
we obtain

\[(3.5) \quad H_p(x) \geq H_p(e) + \frac{C_{p,u,K}\rho(e, x)^2}{2}.
\]

To finish let us consider the case \(p \geq 2\).

In the proof of [1] Theorem 2.1, it is shown that \(e\) is the only zero of the maps \(x \mapsto \text{grad}_x H_p\) and \(x \mapsto H_p(x) - H_p(e)\), and that \(\nabla dH_p(e)\) is strictly positive. This implies that (2.3) and (2.4) hold on some neighbourhood \(B(e, \varepsilon)\) of \(e\). By compactness and the fact that \(H_p - H_p(e)\) and \(\text{grad} H_p\) do not vanish on \(K \setminus B(e, \varepsilon)\) and \(H_p - H_p(e)\) is bounded, possibly modifying the constant \(C_{p,u,K}\), (2.3) and (2.4) also holds on \(K \setminus B(e, \varepsilon)\).

\[\square\]

3.2. Proof of Theorem [2.3]

Note that, for \(x \neq y\),

\[\text{grad}_x F(\cdot, y) = p\rho^{p-1}(x, y)\frac{\exp_x^{-1}(y)}{\rho(x, y)} = -p\rho^{p-1}(x, y)n(x, y),\]

whith \(n(x, y) := \frac{\exp_x^{-1}(y)}{\rho(x, y)}\) a unit vector. So, with the condition (2.6) on \(t_k\), the random walk \((X_k)_{k \geq 0}\) cannot exit \(K\): if \(X_k \in K\) then there are two possibilities for \(X_{k+1}\):

- either \(X_{k+1}\) is in the geodesic between \(X_k\) and \(P_{k+1}\) and belongs to \(K\) by convexity of \(K\);
- or \(X_{k+1}\) is after \(P_{k+1}\), but since

\[
\|t_{k+1} \text{grad}_x F_p(\cdot, P_{k+1})\| = t_{k+1}p\rho^{p-1}(X_k, P_{k+1}) \\
\leq \frac{\rho(K_{\mu}, B(a, r)_c)}{2p(2r)^{p-1}} - pp^{p-1}(X_k, P_{k+1}) \\
\leq \frac{\rho(K_{\mu}, B(a, r)_c)}{2},
\]

we have in this case

\[
\rho(P_{k+1}, X_{k+1}) \leq \frac{\rho(K_{\mu}, B(a, r)_c)}{2}
\]

which implies that \(X_{k+1} \in K\).

First consider the case \(p \in [1, 2)\).

For \(k \geq 0\) let

\[t \mapsto E(t) := \frac{1}{2} \rho^2(e, \gamma(t)) ,
\]

\(\gamma(t)\in[0, t_{k+1}]\) the geodesic satisfying \(\dot{\gamma}(0) = -\text{grad}_x F_p(\cdot, P_{k+1})\). We have for all \(t \in [0, t_{k+1}]\)

\[
(3.6) \quad E''(t) \leq C(\beta, r, p) := p^2(2r)^{2p-1}\beta \coth(2\beta r)
\]
By Taylor formula,
\[
\begin{align*}
\rho(X_{k+1}, e)^2 &= 2E(t_{k+1}) \\
&= 2E(0) + 2t_{k+1}E'(0) + t_{k+1}^2E''(t) \quad \text{for some } t \in [0, t_{k+1}] \\
&\leq \rho(X_k, e)^2 + 2t_{k+1}\langle \text{grad}_{X_k} F_p(\cdot, P_{k+1}), \exp_{X_k}^{-1}(e) \rangle + t_{k+1}^2 C(\beta, r, p).
\end{align*}
\]

Now from the convexity of \( x \mapsto F_p(x, y) \) we have for all \( x, y \in B(a, r) \)
\begin{equation}
F_p(e, y) - F_p(x, y) \geq \langle \text{grad}_x F_p(\cdot, y), \exp_x^{-1}(e) \rangle.
\end{equation}

This applied with \( x = X_k, \ y = P_{k+1} \) yields
\begin{equation}
\rho(X_{k+1}, e)^2 \leq \rho(X_k, e)^2 - 2t_{k+1} (F_p(X_k, P_{k+1}) - F_p(e, P_{k+1})) + C(\beta, r, p)t_{k+1}^2.
\end{equation}

Letting for \( k \geq 0 \) \( \mathcal{F}_k = \sigma(X_{\ell}, 0 \leq \ell \leq k) \), we get
\[
\begin{align*}
\mathbb{E} \left[ \rho(X_{k+1}, e)^2 | \mathcal{F}_k \right] &\leq \rho(X_k, e)^2 - 2t_{k+1} \int_{B(a, r)} (F_p(X_k, y) - F_p(e, y)) \mu(dy) + C(\beta, r, p)t_{k+1}^2 \\
&= \rho(X_k, e)^2 - 2t_{k+1} (H_p(X_k) - H_p(e)) + C(\beta, r, p)t_{k+1}^2 \\
&\leq \rho(X_k, e)^2 + C(\beta, r, p)t_{k+1}^2
\end{align*}
\]
so that the process \((Y_k)_{k \geq 0}\) defined by
\begin{equation}
Y_0 = \rho(X_0, e)^2 \quad \text{and for } k \geq 1 \quad Y_k = \rho(X_k, e)^2 - C(\beta, r, p) \sum_{j=1}^{k} t_j^2
\end{equation}
is a bounded supermartingale. So it converges in \( L^1 \) and almost surely. Consequently \( \rho(X_k, e)^2 \) also converges in \( L^1 \) and almost surely.

Let
\begin{equation}
a = \lim_{k \to \infty} \mathbb{E} \left[ \rho(X_k, e)^2 \right].
\end{equation}

We want to prove that \( a = 0 \). We already proved that
\begin{equation}
\mathbb{E} \left[ \rho(X_{k+1}, e)^2 | \mathcal{F}_k \right] \leq \rho(X_k, e)^2 - 2t_{k+1} (H_p(X_k) - H_p(e)) + C(\beta, r, p)t_{k+1}^2.
\end{equation}

Taking the expectation and using Proposition 2.2 we obtain
\begin{equation}
\mathbb{E} \left[ \rho(X_{k+1}, e)^2 \right] \leq \mathbb{E} \left[ \rho(X_k, e)^2 \right] - t_{k+1}C(\beta, r, p) \mathbb{E} \left[ \rho(X_k, e)^2 \right] + C(\beta, r, p)t_{k+1}^2.
\end{equation}

An easy induction proves that for \( \ell \geq 1 \),
\begin{equation}
\mathbb{E} \left[ \rho(X_{k+\ell}, e)^2 \right] \leq \prod_{j=1}^{\ell} (1 - C_{p, \mu, K} t_{k+j}) \mathbb{E} \left[ \rho(X_k, e)^2 \right] + C(\beta, r, p) \sum_{j=1}^{\ell} t_{k+j}^2.
\end{equation}

Letting \( \ell \to \infty \) and using the fact that \( \sum_{j=1}^{\infty} t_{k+j} = \infty \) which implies
\[
\prod_{j=1}^{\infty} (1 - C_{p, \mu, K} t_{k+j}) = 0,
\]
we get

\begin{equation}
(3.14) \quad a \leq C(\beta, r, p) \sum_{j=1}^{\infty} t_{k+j}^2.
\end{equation}

Finally using \( \sum_{j=1}^{\infty} t_j^2 < \infty \) we obtain that \( \lim_{k \to \infty} \sum_{j=1}^{\infty} t_{k+j}^2 = 0 \), so \( a = 0 \). This proves \( L^2 \) and almost sure convergence.

Next assume that \( p \geq 2 \).

For \( k \geq 0 \) let

\[ t \mapsto E_p(t) := H_p(\gamma(t)), \]

\( \gamma(t)_{t \in [0, t_{k+1}]} \) the geodesic satisfying \( \dot{\gamma}(0) = -\text{grad}_{X_k} F_p(\cdot, P_{k+1}) \). With a calculation similar to (3.6) we get for all \( t \in [0, t_{k+1}] \)

\begin{equation}
(3.15) \quad E_p''(t) \leq 2C(\beta, r, p) := \frac{p^3}{2}(2r)^{3p-4} (2r\beta \coth(2\beta r) + 2p - 4). \tag{3.15}
\end{equation}

(see e.g. [22]). By Taylor formula,

\[ H_p(X_{k+1}) = E_p(t_{k+1}) \]

\[ = E_p(0) + t_{k+1} E'_p(0) + \frac{t_{k+1}^2}{2} E''_p(t) \quad \text{for some } t \in [0, t_{k+1}] \]

\[ \leq H_p(X_k) + t_{k+1} \langle d_{X_k} H_p, \text{grad}_{X_k} F_p(\cdot, P_{k+1}) \rangle + t_{k+1}^2 C(\beta, r, p). \]

We get

\[ E[H_p(X_{k+1})|\mathcal{F}_k] \]

\[ \leq H_p(X_k) - t_{k+1} \left\langle d_{X_k} H_p, \int_{B(a,r)} \text{grad}_{X_k} F_p(\cdot, y) \mu(dy) \right\rangle + C(\beta, r, p) t_{k+1}^2 \]

\[ = H_p(X_k) - t_{k+1} \left\langle d_{X_k} H_p, \text{grad}_{X_k} H_p(\cdot) \right\rangle + C(\beta, r, p) t_{k+1}^2 \]

\[ = H_p(X_k) - t_{k+1} \left\| \text{grad}_{X_k} H_p(\cdot) \right\|^2 + C(\beta, r, p) t_{k+1}^2 \]

\[ \leq H_p(X_k) - C_{p,\mu, K} t_{k+1} (H_p(X_k) - H_p(e)) + C(\beta, r, p) t_{k+1}^2 \]

(by Proposition 2.2) so that the process \( (Y_k)_{k \geq 0} \) defined by

\begin{equation}
(3.16) \quad Y_0 = H_p(X_0) - H_p(e) \quad \text{and for } k \geq 1 \quad Y_k = H_p(X_k) - H_p(e) - C(\beta, r, p) \sum_{j=1}^{k} t_j^2
\end{equation}

is a bounded supermartingale. So it converges in \( L^1 \) and almost surely. Consequently \( H_p(X_k) - H_p(e) \) also converges in \( L^1 \) and almost surely.

Let

\begin{equation}
(3.17) \quad a = \lim_{k \to \infty} E[H_p(X_k) - H_p(e)].
\end{equation}

We want to prove that \( a = 0 \). We already proved that

\begin{equation}
(3.18) \quad E[H_p(X_{k+1}) - H_p(e)|\mathcal{F}_k] \leq H_p(X_k) - H_p(e) - C_{p,\mu, K} t_{k+1} (H_p(X_k) - H_p(e)) + C(\beta, r, p) t_{k+1}^2.
\end{equation}
Taking the expectation we obtain
\[(3.19) \quad \mathbb{E} [H_p(X_{k+1}) - H_p(e)] \leq (1 - t_{k+1}C_{p,\mu,K}) \mathbb{E} [H_p(X_k) - H_p(e)] + C(\beta, r, p)t_{k+1}^2\]
so that proving that \(a = 0\) is similar to the previous case.

Finally (2.3) proves that \(\rho(X_k, e)^2\) converges in \(L^1\) and almost surely to 0. \(\square\)

### 3.3. Proof of Proposition 2.5

Fix \(\varepsilon > 0\). Any diffusion process on \([\varepsilon, \infty)\) with generator \(G_\delta(t)\) is solution of a sde of the type
\[
\text{(3.20)} \quad dy_t = \frac{1}{t} L_\delta(y_t) \, dt + \sigma^t \, dB_t
\]
where \(L_\delta(y) = y - \delta \nabla dH_p(y, \cdot)^2\) and \(B_t\) and \(\sigma\) are as in Proposition 2.5. This sde can be solved explicitely on \([\varepsilon, \infty)\). The symmetric endomorphism \(y \mapsto \nabla dH_p(y, \cdot)^2\) is diagonalisable in the orthonormal basis \((e_i)_{1 \leq i \leq d}\) with eigenvalues \((\lambda_i)_{1 \leq i \leq d}\). The endomorphism \(L_\delta = \text{id} - \delta \nabla dH_p(e)(\text{id}, \cdot)^2\) is also diagonalisable in this basis with eigenvalues \((1 - \delta \lambda_i)_{1 \leq i \leq d}\). The solution \(y_t = \sum_{i=1}^d y^i_t e_i\) of (3.20) started at \(y_\varepsilon = \sum_{i=1}^d y^i_\varepsilon e_i\) is given by
\[
\text{(3.21)} \quad y_t = \sum_{i=1}^d \left( y^i_\varepsilon e_i + \int_\varepsilon^t 1_{\lambda_i - 1} (\delta \sigma dB_s, e_i) \right) t^{1-\delta \lambda_i} e_i, \quad t \geq \varepsilon.
\]
Now by definition of \(C_{p,\mu,K}\) we clearly have
\[
\text{(3.22)} \quad C_{p,\mu,K} \leq \min_{1 \leq i \leq d} \lambda_i.
\]
So the condition \(\delta C_{p,\mu,K} > 1\) implies that for all \(i\), \(\delta \lambda_i - 1 > 0\), and as \(\varepsilon \to 0\),
\[
\text{(3.23)} \quad \int_\varepsilon^t s^{\delta \lambda_i - 1} (\delta \sigma dB_s, e_i) \to \int_0^t s^{\delta \lambda_i - 1} (\delta \sigma dB_s, e_i) \quad \text{in probability}.
\]
Assume that a continuous solution \(y_t\) converging in probability to 0 as \(t \to 0^+\) exists. Since \(y^i_\varepsilon e_i\) converges to 0 in probability as \(\varepsilon \to 0\), we necessarily have using (3.22)
\[
\text{(3.24)} \quad y_t = \sum_{i=1}^d t^{1-\delta \lambda_i} \int_0^t s^{\delta \lambda_i - 1} (\delta \sigma dB_s, e_i) e_i, \quad t \geq 0.
\]
Note \(y^i_\delta\) is Gaussian with variance \(\frac{t \delta^2 \Gamma(e^*_i \otimes e^*_i)}{2\delta \lambda_i - 1}\), so it converges in \(L^2\) to 0 as \(t \to 0\).

Conversely, it is easy to check that equation (3.24) defines a solution to (3.20).

To prove the a.s. convergence to 0 we use the representation
\[
\int_0^t s^{\delta \lambda_i - 1} (\delta \sigma dB_s, e_i) = B^i_{\varphi_i(t)}
\]
where \(B^i_s\) is a Brownian motion and \(\varphi_i(t) = \frac{\delta^2 \Gamma(e^*_i \otimes e^*_i)}{2\delta \lambda_i - 1} t^{2\delta \lambda_i - 1}\). Then by the law of iterated logarithm
\[
\limsup_{t \downarrow 0} t^{1-\delta \lambda_i} B^i_{\varphi_i(t)} \leq \limsup_{t \downarrow 0} t^{1-\delta \lambda_i} \sqrt{2 \varphi_i(t) \ln \ln (\varphi_i^{-1}(t))}
\]
But for \( t \) small we have
\[
\sqrt{2\varphi_i(t) \ln (\varphi_i^{-1}(t))} \leq t^d \lambda_i^{-3/4}
\]
so
\[
\limsup_{t \downarrow 0} t^{1-\delta} B_{\varphi_i(t)}^i \leq \lim_{t \downarrow 0} t^{1/4} = 0.
\]
This proves a.s. convergence to 0. Continuity is easily checked using the integral representation (3.24). \( \square \)

### 3.4. Proof of Theorem 2.6

Consider the time homogeneous Markov chain \( (Z^n_k)_{k \geq 0} \) with state space \([0, \infty) \times T_e M\) defined by
\[
Z^n_k = \left( \frac{k}{n}, Y^n_k \right).
\]
The first component has a deterministic evolution and will be denoted by \( t^n_k \); it satisfies
\[
t^n_{k+1} = t^n_k + \frac{1}{n}, \quad k \geq 0.
\]
Let \( k_0 \) be such that
\[
\frac{\delta}{k_0} < \delta_1.
\]
Using equations (2.8), (2.12) and (2.11), we have for \( k \geq k_0,\)
\[
Y^n_{k+1} = \frac{nt^n_k + 1}{\sqrt{n}} \exp^{-1} \left( \exp_{\exp_{\frac{\delta}{nt^n_k + 1}} Y^n_k} \left( -\frac{\delta}{nt^n_k + 1} \grad_{\exp_{\frac{\delta}{nt^n_k + 1}}} F_p(\cdot, P_{k+1}) \right) \right).
\]
Consider the transition kernel \( P^n(z, dz') \) on \((0, \infty) \times T_e M\) defined for \( z = (t, y) \) by
\[
P^n(z, A) = \mathbb{P} \left[ \left( t + \frac{1}{n} \frac{nt + 1}{\sqrt{n}} \exp^{-1} \left( \exp_{\exp_{\frac{\delta}{nt + 1}} y} \left( -\frac{\delta}{nt + 1} \grad_{\exp_{\frac{\delta}{nt + 1}}} F_p(\cdot, P_1) \right) \right) \right) \in A \right]
\]
where \( A \in \mathcal{B}(\{0, \infty\} \times T_e M)\). Clearly this transition kernel drives the evolution of the Markov chain \( (Z^n_k)_{k \geq k_0}\).

For the sake of clarity, we divide the proof of Theorem 2.6 into four lemmas.

**Lemma 3.1.** Assume that either \( p \geq 2 \) or \( c \) does not belong to the support supp(\( \mu \)) of \( \mu \) (note this implies that for all \( x \in \text{supp}(\mu) \) the function \( F_p(\cdot, x) \) is of class \( C^2 \) in a neighbourhood of \( c \)). Fix \( \delta > 0 \). Let \( B \) be a bounded set in \( T_e M \) and let \( 0 < \varepsilon < T \). We have for all \( C^2 \) function \( f \) on \( T_e M \)
\[
n \left( f \left( \frac{nt + 1}{\sqrt{n}} \exp^{-1} \left( \exp_{\exp_{\frac{\delta}{nt + 1}} y} \left( -\frac{\delta}{nt + 1} \grad_{\exp_{\frac{\delta}{nt + 1}}} F_p(\cdot, x) \right) \right) \right) - f(y) \right)
\]
\[
= \left< \frac{d_y f}{t}, \frac{y}{t} \right> - \sqrt{n} \left< d_y f, \delta \grad_{\exp_{\frac{\delta}{nt + 1}}} F_p(\cdot, x) \right> - \delta \nabla F_p(\cdot, x) \left< \grad_y f, \frac{y}{t} \right>
\]
\[
+ \frac{\delta^2}{2} \Hess_y f (\grad_{\exp_{\frac{\delta}{nt + 1}}} F_p(\cdot, x) \otimes \grad_{\exp_{\frac{\delta}{nt + 1}}} F_p(\cdot, x)) + O \left( \frac{1}{\sqrt{n}} \right)
\]
unifromly in \( y \in B, x \in \text{supp}(\mu), t \in [\varepsilon, T] \).
Proof. Let \( x \in \text{supp}(\mu) \), \( y \in T_xM \), \( u, v \in \mathbb{R} \) sufficiently close to 0, and \( q = \exp_c \left( \frac{uv}{t} \right) \). For \( s \in [0, 1] \) denote by \( a \mapsto c(a, s, u, v) \) the geodesic with endpoints \( \{0, (s, u, v)\} = e \) and

\[
c(1, s, u, v) = \exp_{\exp_c} \left( -sv \text{grad}_{\exp_c} \left( \frac{uv}{t} \right) F_p(\cdot, x) \right) :
\]

\[
c(a, s, u, v) = \exp_{\exp_c} \left\{ a \exp_{\exp_c}^{-1} \left[ \exp_{\exp_c} \left( \frac{uv}{t} \right) \left( -sv \text{grad}_{\exp_c} \left( \frac{uv}{t} \right) F_p(\cdot, x) \right) \right] \right\}.
\]

This is a \( C^2 \) function of \( (a, s, u, v) \) \( \in [0, 1]^2 \times (-\eta, \eta)^2 \), \( \eta \) sufficiently small. It also depends in a \( C^2 \) way of \( x \) and \( y \). Letting \( c(a, s) = c \left( a, s, -\frac{1}{\sqrt{n}}, -\frac{2}{nt+1} \right) \), we have

\[
\exp_{\exp_c}^{-1} \left( \exp_{\exp_c} \left( \frac{uv}{t} \right) \left( -\frac{2}{nt+1} \text{grad}_{\exp_c} \left( \frac{uv}{t} \right) F_p(\cdot, x) \right) \right) = \partial_a c(0, 1).
\]

So we need a Taylor expansion up to order \( n^{-1} \) of \( \frac{nt+1}{\sqrt{n}} \partial_a c(0, 1) \).

We have \( c(a, s, 0, 1) = \exp_{\exp_c} \left( -as \text{grad}_c F_p(\cdot, x) \right) \) and this implies

\[
\partial^2_{a} \partial_a c(0, s, 0, 1) = 0, \quad \text{so} \quad \partial^2_{a} \partial_a c(0, s, u, v) = O(u).
\]

On the other hand the identities \( c(a, s, u, v) = c(a, sv, u, 1) \) yields \( \partial^2_{a} \partial_a c(a, s, u, v) = v^2 \partial^2_{a} \partial_a c(a, s, u, 1) \), so we obtain

\[
\partial^2_{a} \partial_a c(0, s, u, v) = O(u v^2)
\]

and this yields

\[
\partial^2_{a} \partial_a c(0, s) = O(n^{-5/2}),
\]

uniformly in \( s, x, y, t \). But since

\[
\| \partial_a c(0, 1) - \partial_a c(0, 0) - \partial_s \partial_a c(0, 0) \| \leq \frac{1}{2} \sup_{s \in [0, 1]} \| \partial^2_{a} \partial_a c(0, s) \|
\]

we only need to estimate \( \partial_a c(0, 0) \) and \( \partial_s \partial_a c(0, 0) \).

Denoting by \( J(a) \) the Jacobi field \( \partial_a c(a, 0) \) we have

\[
\frac{nt+1}{\sqrt{n}} \partial_a c(0, 1) = \frac{nt+1}{\sqrt{n}} \partial_a c(0, 0) + \frac{nt+1}{\sqrt{n}} \dot{J}(0) + O \left( \frac{1}{n^2} \right).
\]

On the other hand

\[
\frac{nt+1}{\sqrt{n}} \partial_a c(0, 0) = \frac{nt+1}{\sqrt{n}} \frac{y}{\sqrt{nt}} = y + \frac{y}{nt}
\]

so it remains to estimate \( \dot{J}(0) \).

The Jacobi field \( a \mapsto J(a, u, v) \) with endpoints \( J(0, u, v) = 0 \) and

\[
J(1, u, v) = -v \text{grad}_{\exp_c} \left( \frac{uv}{t} \right) F_p(\cdot, x)
\]

satisfies

\[
\nabla^2_a J(a, u, v) = -R(J(a, u, v), \partial_a c(a, 0, u, v)) \partial_a c(a, 0, u, v) = O(u v^2).
\]

This implies that

\[
\nabla^2_a J(a) = O(n^{-2}).
\]

Consequently, denoting by \( P_{x_1, x_2} : T_{x_1}M \to T_{x_2}M \) the parallel transport along the minimal geodesic from \( x_1 \) to \( x_2 \) (whenever it is unique) we have

\[
P_{c(1, 0), c} J(1) = J(0) + \dot{J}(0) + O(n^{-2}) = \dot{J}(0) + O(n^{-2}).
\]
Lemma 3.2. The following property holds:

\[ b_n(z) = n \int_{|z' - z| \leq 1} (z' - z)^n P^n(z, dz') \]

and

\[ a_n(z) = n \int_{|z' - z| \leq 1} (z' - z) \otimes (z' - z)^n P^n(z, dz'). \]

The following property holds:

**Lemma 3.2.** Assume that either \( p \geq 2 \) or \( e \) does not belong to the support \( \text{supp}(\mu) \).

1. For all \( R > 0 \) and \( \varepsilon > 0 \), there exists \( n_0 \) such that for all \( n \geq n_0 \) and \( z \in [\varepsilon, T] \times B(0, R) \), where \( B(0, R) \) is the open ball in \( T_e M \) centered at the origin with radius \( R \),

\[ \int_{1(|z' - z| > 1)} P^n(z, dz') = 0. \]

2. For all \( R > 0 \) and \( \varepsilon > 0 \),

\[ \lim_{n \to \infty} \sup_{z \in [\varepsilon, T] \times B(0, R)} |b_n(z) - b(z)| = 0 \]
We use the notation

\[ b(z) = \left(1, \frac{1}{t} L_\delta(y) \right) \quad \text{and} \quad L_\delta(y) = y - \delta \nabla dH(y, \cdot)^2. \]

(3) For all \( R > 0 \) and \( \varepsilon > 0 \),

\[ \lim_{n \to \infty} \sup_{z \in [e, T] \times B(0, R)} |a_n(z) - a(z)| = 0 \]

with

\[ a(z) = \delta^2 \text{diag}(0, \Gamma) \quad \text{and} \quad \Gamma = \mathbb{E} [\text{grad}_e F_p(\cdot, P_1) \otimes \text{grad}_e F_p(\cdot, P_1)]. \]

Proof. (1) We use the notation \( z = (t, y) \) and \( z' = (t', y') \). We have

\[ \int 1_{|z'-z|>1} P^n(z, dz') \]

\[ = \int 1_{\max(|t'-t|, |y'-y|)>1} P^n(z, dz') \]

\[ = \int 1_{\max(\frac{1}{n^2}, |y'-y|)>1} P^n(z, dz') \]

\[ = \mathbb{P} \left[ \frac{nt + 1}{\sqrt{n}} \exp_y e^{-1} \left( \exp_y \left( \frac{\delta}{nt + 1} \text{grad}_y \left( \frac{1}{\sqrt{n}} F_p(\cdot, P_1) \right) \right) \right) - y \mid t > 1 \right]. \]

On the other hand, since \( F_p(\cdot, x) \) is of class \( C^2 \) in a neighbourhood of \( e \), we have by (3.32)

\[ \left| nt + 1 \sqrt{n} \exp_y e^{-1} \left( \exp_y \left( \frac{\delta}{nt + 1} \text{grad}_y \left( \frac{1}{\sqrt{n}} F_p(\cdot, P_1) \right) \right) \right) - y \right| \leq C\delta \sqrt{n} \varepsilon \]

for some constant \( C > 0 \).

(2) Equation (3.35) implies that for \( n \geq n_0 \)

\[ b_n(z) \]

\[ = n \int (z'-z) P^n(z, dz') \]

\[ = n \left( \frac{1}{\sqrt{n}}, \mathbb{E} \left[ \frac{nt + 1}{\sqrt{n}} \exp_y e^{-1} \left( \exp_y \left( \frac{\delta}{nt + 1} \text{grad}_y \left( \frac{1}{\sqrt{n}} F_p(\cdot, P_1) \right) \right) \right) \right] - y \right). \]

We have by lemma (3.1)

\[ n \left( \frac{nt + 1}{\sqrt{n}} \exp_y e^{-1} \left( \exp_y \left( \frac{\delta}{nt + 1} \text{grad}_y \left( \frac{1}{\sqrt{n}} F_p(\cdot, P_1) \right) \right) \right) - y \right) \]

\[ = \frac{1}{t} y - \delta \sqrt{n} \text{grad}_e F_p(\cdot, P_1) - \delta \nabla dF_p(\cdot, P_1) \left( \frac{1}{t} y, \cdot \right) + o \left( \frac{1}{n^{1/2}} \right) \]

a.s. uniformly in \( n \), and since

\[ \mathbb{E} [\delta \sqrt{n} \text{grad}_e F_p(\cdot, P_1)] = 0, \]

this implies that

\[ n \left( \mathbb{E} \left[ \frac{nt + 1}{\sqrt{n}} \exp_y e^{-1} \left( \exp_y \left( \frac{\delta}{nt + 1} \text{grad}_y \left( \frac{1}{\sqrt{n}} F_p(\cdot, P_1) \right) \right) \right) \right] - y \right) \]
converges to

\[
\frac{1}{t} y - E\left[ \delta \nabla dF_p(\cdot, P_1) \left( \frac{1}{t} y, \cdot \right) \right] = \frac{1}{t} y - \delta \nabla dH_p \left( \frac{1}{t} y, \cdot \right).
\]

Moreover the convergence is uniform in \( z \in [\varepsilon, T] \times B(0, R) \), so this yields (3.38).

(3) In the same way, using lemma 3.1

\[
n \int (y' - y) \otimes (y' - y) P^n(z, dz)
= \frac{1}{n} E \left[ \left( -\sqrt{n} \delta \nabla e, F_p(\cdot, P_1) \right) \otimes \left( -\sqrt{n} \delta \nabla e, F_p(\cdot, P_1) \right) \right] + o(1)
= \delta^2 E \left[ \nabla e, F_p(\cdot, P_1) \right] \otimes \nabla e, F_p(\cdot, P_1) + o(1)
\]

uniformly in \( z \in [\varepsilon, T] \times B(0, R) \), so this yields (3.38).

\[\square\]

**Lemma 3.3.** Suppose that \( t_n = \frac{\delta}{n} \) for some \( \delta > 0 \). For all \( \delta > C_{p, \mu, K}^{-1}, \)

\[
\sup_{n \geq 1} n E \left[ \rho^2(e, X_n) \right] < \infty.
\]

**Proof.** First consider the case \( p \in [1, 2) \).

We know by (3.12) that there exists some constant \( C(\beta, r, p) \) such that

\[
E \left[ \rho^2(e, X_{k+1}) \right] \leq E \left[ \rho^2(e, X_k) \right] \exp \left( -C_{p, \mu, K} t_{k+1} \right) + C(\beta, r, p) t_{k+1}^2.
\]

From this (3.12) is a consequence of Lemma 0.0.1 (case \( \alpha > 1 \) in [10]. We give the proof for completeness. We deduce easily by induction that for all \( k \geq k_0, \)

\[
E \left[ \rho^2(e, X_k) \right] \leq E \left[ \rho^2(e, X_{k_0}) \right] \exp \left( -C_{p, \mu, K} \sum_{j=k_0+1}^k t_j \right) + C(\beta, r, p) \sum_{i=k_0+1}^k t_i^2 \exp \left( -C_{p, \mu, K} \sum_{j=i+1}^k t_j \right),
\]

where the convention \( \sum_{j=k_0+1}^k t_j = 0 \) is used. With \( t_n = \frac{\delta}{n} \), the following inequality holds for all \( i \geq k_0 \) and \( k \geq i: \)

\[
\frac{1}{n} \sum_{j=i+1}^k t_j = \delta \frac{1}{n} \sum_{j=i+1}^k \frac{1}{j} \geq \delta \int_{i+1}^{k+1} \frac{dt}{t} \geq \delta \ln \frac{k+1}{i+1}.
\]

Hence,

\[
E \left[ \rho^2(e, X_k) \right] \leq E \left[ \rho^2(e, X_{k_0}) \right] \left( \frac{k_0 + 1}{k + 1} \right)^\delta C_{p, \mu, K} + \frac{\delta^2 C(\beta, r, p)}{(k + 1)^\delta C_{p, \mu, K}} \sum_{i=k_0+1}^k \frac{(i + 1)^\delta C_{p, \mu, K}}{i^2}.
\]

For \( \delta C_{p, \mu, K} > 1 \) we have as \( k \to \infty \)

\[
\frac{\delta^2 C(\beta, r, p)}{(k + 1)^\delta C_{p, \mu, K}} \sum_{i=k_0+1}^k \frac{(i + 1)^\delta C_{p, \mu, K}}{i^2} \sim \frac{\delta^2 C(\beta, r, p)}{(k + 1)^\delta C_{p, \mu, K}} \frac{k^{\delta C_{p, \mu, K} - 1}}{\delta C_{p, \mu, K} - 1} \sim \frac{\delta^2 C(\beta, r, p)}{\delta C_{p, \mu, K} - 1} k^{\delta C_{p, \mu, K} - 1}\]
Lemma 3.4. Assume bounded. We conclude with (2.3). □

\[(3.48)\]

Proof. Denote by \(k\) that from any subsequence \((\tilde{Y}_\varepsilon^n)_{n \geq 1}\), we can extract a further subsequence \((\tilde{Y}_\varepsilon^{\phi(n)})_{n \geq 1}\) that weakly converges in \(\mathbb{D}([\varepsilon, 1], \mathbb{R}^d)\).

Let us first prove that \((\tilde{Y}_\varepsilon^{\phi(n)}(\varepsilon))_{n \geq 1}\) is bounded in \(L^2\).

\[\left\|\tilde{Y}_\varepsilon^{\phi(n)}(\varepsilon)\right\|^2_2 = \frac{\left(\phi(n)\varepsilon^2\right)^2}{\phi(n)} \mathbb{E}\left[\rho^2(e, X_{\phi(n)\varepsilon})\right] \leq \varepsilon \sup_{n \geq 1} \left(n \mathbb{E}\left[\rho^2(e, X_n)\right]\right)\]

and the last term is bounded by lemma 3.3.

Consequently \((\tilde{Y}_\varepsilon^{\phi(n)}(\varepsilon))_{n \geq 1}\) is tight. So there is a subsequence \((\tilde{Y}_\varepsilon^{\phi(n)}(\varepsilon))_{n \geq 1}\) that weakly converges in \(T_{\varepsilon} \tilde{M}\) to the distribution \(\nu_\varepsilon\). Thanks to Skorohod theorem which allows to realize it as an a.s. convergence and to lemma 3.2 we can apply Theorem 11.2.3 of [20], and we obtain that the sequence of processes \((\tilde{Y}_\varepsilon^{\phi(n)})_{n \geq 1}\) weakly converges to a diffusion \((y_\varepsilon)_{\varepsilon \leq t \leq T}\) with generator \(G_\phi(t)\) given by (2.13) and such that \(y_\varepsilon\) has law \(\nu_\varepsilon\). This achieves the proof of lemma 3.4. □

Proof of Theorem 2.6. Let \(\tilde{Y}^n = \left(Y_{[n]}^n\right)_{0 \leq t \leq T}\). It is sufficient to prove that any subsequence of \((\tilde{Y}^n)_{n \geq 1}\) has a further subsequence which converges in law to \((y_\delta(t))_{0 \leq t \leq T}\). So let \((\tilde{Y}_\varepsilon^{\phi(n)})_{n \geq 1}\) a subsequence. By lemma 3.4 with \(\varepsilon = 1/m\) there exists a subsequence which converges in law on \([1/m, T]\). Then we extract a sequence indexed by \(m\) of subsequence and take the diagonal subsequence \(\tilde{Y}^n(\varepsilon)\). This subsequence converges in \(\mathbb{D}((0, T], \mathbb{R}^d)\) to \((y'(t))_{t \in (0, T]}\). On the other hand, as in the proof of lemma 3.4 we have

\[\|\tilde{Y}^n(\varepsilon)(t)\|_2^2 \leq Ct\]

for some \(C > 0\). So \(\|\tilde{Y}^n(\varepsilon)(t)\|_2^2 \to 0\) as \(\varepsilon \to 0\), which in turn implies \(\|y'(t)\|_2^2 \to 0\) as \(t \to 0\). The unicity statement in Proposition 2.4 implies that \((y'(t))_{t \in (0, T]}\) and \((y_\delta(t))_{t \in (0, T]}\) are equal in law. This achieves the proof. □
References

[1] B. Afsari, *Riemannian $L^p$ center of mass: existence, uniqueness, and convexity*, Proceedings of the American Mathematical Society, S 0002-9939(2010)10541-5, Article electronically published on August 27, 2010.

[2] M. Arnaudon and X.M. Li, *Barycenters of measures transported by stochastic flows*, The Annals of Probability, 33 (2005), no. 4, 1509–1543

[3] A. Benveniste, M. Goursat and G. Ruget, *Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms with applications to data communication algorithm*, IEEE Transactions on Automatic Control, Vol. AC-25, no. 6, December 1980.

[4] E. Berger, *An almost sure invariance principle for stochastic approximation procedures in linear filtering theory*, the Annals of applied probability, vol. 7, no. 2 (May 1997), pp. 444–459.

[5] M. Emery and G. Mokobodzki, *Sur le barycentre d’une probabilité dans une variété*, Séminaire de Probabilités XXV, Lecture Notes in Mathematics 1485 (Springer, Berlin, 1991), pp. 220–233

[6] S. Gouëzel, *Almost sure invariance principle for dynamical systems by spectral methods* The annals of Probability, Vol 38, no. 4 (2010), pp. 1639–1671

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

[8] H. Karcher, *Riemannian center of mass and mollifier smoothing*, Communications on Pure and Applied Mathematics, vol XXX (1977), 509–541

[9] R.Z. Khas’Minskii, *On stochastic processes defined by differential equations with a small parameter*, Theory Prob. Appl., vol. XI-2 (1968), pp. 211–228

[10] W.S. Kendall, *Probability, convexity and harmonic maps with small image I: uniqueness and fine existence*, Proc. London Math. Soc. (3) 61 no. 2 (1990) pp. 371–406

[11] W.S. Kendall, *Convexity and the hemisphere*, J. London Math. Soc., (2) 43 no.3 (1991), pp. 567–576

[12] H.W. Kuhn, *A note on Fermat’s problem*, Mathematical Programming, 4 (1973), pp. 98–107

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

[14] L. Ljung, *Analysis of recursive stochastic algorithms*, IEEE Transactions on Automatic Control, Vol. AC-22, no. 4, August 1977.

[15] D. L. McLeish, *Dependent central limit theorems and invariance principles*, the Annals of Probability, vol. 2, no. 4 (1974), pp. 620–628

[16] A. Nedic and D.P. Bertsekas, *Convergence rate of incremental subgradient algorithms*, Stochastic optimization : algorithms and applications (S. Uryasev and P.M. Pardalos, Editors), Kluwer Academic Publishers (2000), pp. 263–304

[17] L.M.JR. Ostresh, *On the convergence of a class of iterative methods for solving Weber location problem*, Operation. Research, 26 (1978), no. 4

[18] J. Picard, *Barycentres et martingales dans les variétés*, Ann. Inst. H. Poincaré Probab. Statist. 30 (1994), pp. 647–702

[19] A. Sahib, *Espérance d’une variable aléatoire à valeurs dans un espace métrique*, Thése de l’université de Rouen (1998)

[20] D.W. Stroock and S.R.S. Varadhan, *Multidimensional diffusion processes*, Grundlehren der mathematischen Wissenschaften 233, Springer, 1979.

[21] E. Weissfeld, *Sur le point pour lequel la somme des distances de n points donnés est minimum*, Tohoku Math. J. 43 (1937), pp. 355–386

[22] L. Yang, *Riemannian median and its estimation*, to appear in LMS Journal of Computation and Mathematics
