STOCHASTIC ALGORITHMS FOR COMPUTING MEANS OF PROBABILITY MEASURES

MARC ARNAUDON, CLÉMENT DOMBRY, ANTHONY PHAN, AND LE YANG

Abstract. Consider a probability measure \( \mu \) supported by a regular geodesic ball in a manifold. For any \( p \geq 1 \) we define a stochastic algorithm which converges almost surely to the \( p \)-mean \( e_p \) of \( \mu \). Assuming furthermore that the functional to minimize is regular around \( e_p \), we prove that a natural renormalization of the inhomogeneous Markov chain converges in law into an inhomogeneous diffusion process. We give an explicit expression of this process, as well as its local characteristic.

1. Introduction

Consider a set of points \( \{x_1, \ldots, x_n\} \) in an Euclidean space \( E \) with metric \( d \). The geometric barycenter \( e_2 \) of this set of points is the unique point minimizing the mean square distance to these points, i.e.

\[
e_2 = \arg \min_{x \in E} \frac{1}{n} \sum_{i=1}^n d^2(x, x_i).
\]

It is equal to the standard mean \( e_2 = \frac{1}{n} \sum_{i=1}^n x_i \) and 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 \( p \in [1, 2) \). This leads to the definition of \( p \)-means: for \( p \geq 1 \), a minimizer of the functional

\[
H_p : M \rightarrow \mathbb{R}_+, \quad x \mapsto \frac{1}{n} \sum_{i=1}^n d^p(x, x_i)
\]

is called a \( p \)-mean of the set of points \( \{x_1, \ldots, x_n\} \). When \( p = 1 \), \( e_1 \) is the median of the set of points and is often used in robust statistics. In many applications, \( p \)-means with some \( p \in (1, 2) \) give the best compromise. It is well known that in dimension 1, the median of a set of real numbers may not be uniquely defined. This is however an exceptional case: the \( p \)-mean of a set of points is uniquely defined as soon as \( p = 1 \) and the points are not aligned or as \( p > 1 \). In these cases, uniqueness is due to the strict convexity of the functional \( H_p \).

The notion of \( p \)-mean is naturally extended to probability measures on Riemannian manifolds. Let \( \mu \) be a probability measure on a Riemannian manifold \( M \) with distance \( \rho \). For any \( p \geq 1 \), a \( p \)-mean of \( \mu \) is a minimizer of the functional

\[
H_p : M \rightarrow \mathbb{R}_+, \quad x \mapsto \int_M \rho^p(x, y) \mu(dy)
\]

(1.1)

It should be stressed that unlike the Euclidean case, the functional \( H_p \) may not be convex (if \( p \geq 2 \)) and the \( p \)-mean may not be uniquely defined. In the case \( p = 2 \), we obtain the so-called Riemannian barycenter or Karcher mean of the probability measure \( \mu \). This has been extensively studied, see e.g. \[8\], \[10\], \[11\], \[5\], \[18\].
where questions of existence, uniqueness, stability, relation with martingales in manifolds, behaviour when measures are pushed by stochastic flows have been considered. In the general case $p \geq 1$, Afsari [1] proved existence and uniqueness of $p$-means on “small” geodesic balls. More precisely, let $\text{inj}(M)$ be the injectivity radius of $M$ and $\alpha^2 > 0$ an upper bound for the sectional curvatures in $M$. Existence and uniqueness of the $p$-mean in ensured as soon as the support of the probability measure $\mu$ is contained in a convex compact $K_\mu$ of a geodesic ball $B(a,r)$ with radius

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

The case $p \geq 2$ gives rise to additional difficulties since the functional $H_p$ to minimize is not necessarily convex any more, due to the fact that we can have $r > \frac{\pi}{2\alpha}$. Provided existence and uniqueness of the $p$-mean, the question of its practical determination and computation arises naturally. In the Euclidean setting and when $p = 1$, the problem of finding the median $e_1$ of a set of points is known as the Fermat-Weber problem and numerous algorithms have been designed to solve it. A first one was proposed by Weiszfeld in [21] and was then extended by Fletcher and al in [7] to cover the case of sufficiently small domains in Riemannian manifolds with nonnegative curvature. 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]. In the case $p = 2$, computation of the Riemannian barycenter has been performed by Le in [13] using a gradient descent algorithm.

In this paper, we consider the general case $p \geq 1$ in a Riemannian setting. Under the above mentioned condition of Asfari [1] ensuring existence and uniqueness of the $p$-mean, we provide a stochastic gradient descent algorithm that converges almost surely to $e_p$. This algorithm is easier to implement than the deterministic gradient descent algorithm since it does not require computing the gradient of the functional $H_p$ to minimize. More precisely, we construct a time inhomogeneous Markov chain $(X_k)_{k \geq 0}$ as follows: at each step $k \geq 0$, we draw a random point $P_{k+1}$ with distribution $\mu$ and we move the current point $X_k$ to $X_{k+1}$ along the geodesic from $X_k$ to $X_{k+1}$ by a distance depending on $p$, $X_k$ and on a deterministic parameter $t_{k+1}$. In Theorem 2.3 below, we state that under a suitable condition on the sequence $(t_k)_{k \geq 1}$, the Markov chain $(X_k)_{k \geq 0}$ converges almost surely and in $L^2$ to the $p$-mean $e_p$. Our proof relies on the martingale convergence theorem and the main point consists in determining and estimating all the geometric quantities. For related convergence results on recursive stochastic algorithms, see [14] Theorem 1 or [3].

We then study the speed of convergence to the $p$-mean and its fluctuations: in Theorem 2.6 we prove that the suitably renormalized inhomogeneous Markov chain $(X_k)_{k \geq 0}$ converges in law to an inhomogeneous diffusion process in the Skohorod space. This is an invariance principle type result, see e.g. [9], [15], [4], [6] for related works. Interestingly, the limiting process depends in a crucial way on the sequence $(t_k)_{k \geq 1}$ of the algorithm. The main point is to compute the generator of the rescaled Markov chain and to obtain the characteristics of the limiting process from the curvature conditions and estimates on Jacobi fields.

The paper is organized as follows. Section 2 is devoted to a detailed presentation of the stochastic gradient descent algorithm $(X_k)_{k \geq 0}$ and its properties: almost sure
convergence is stated in Theorem 2.3 and the invariance principle in Theorem 2.6. 
Proofs are gathered in Section 3.

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 equation (1.2).

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 functional \(H_p\) defined by Equation (1.1) has a unique minimizer \(e_p\) in \(M\), the \(p\)-mean of \(\mu\), and moreover \(e_p \in B(a,r)\). If \(p = 1\), \(c_1\) 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 \(B(a,r)\) but not necessarily convex as mentioned in the introduction.

**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

\[
K = B(a,r - \varepsilon) \quad \text{with} \quad \varepsilon = \frac{\rho(K_{\mu}, B(a,r)^c)}{2}.
\]

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

\[
\forall k \geq 1, \quad t_k \leq \min \left( \frac{1}{C_{p,\mu,K}}, \frac{\rho(K_{\mu}, B(a,r)^c)}{2p(2r)^{p-1}} \right),
\]

\[
\sum_{k=1}^{\infty} t_k = +\infty \quad \text{and} \quad \sum_{k=1}^{\infty} t_k^2 < \infty.
\]

Letting \(x_0 \in K\), define inductively the random walk \((X_k)_{k \geq 0}\) by

\[
X_0 = x_0 \quad \text{and for } k \geq 0 \quad X_{k+1} = \exp_{X_k} \left( -t_{k+1} \nabla_x F_p(\cdot, P_{k+1}) \right)
\]

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 (2.6) simplifies into

$$X_0 = x_0 \quad \text{and for } k \geq 0 \quad X_{k+1} = X_k - t_{k+1} \text{grad}_x F_p(\cdot, P_{k+1}).$$

If furthermore $p = 2$, clearly $e_2 = \mathbb{E}[P_1]$ and $\text{grad}_x F_p(\cdot, y) = 2(x - y)$, 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

$$(2.7) \quad X_k = x_0 \prod_{j=0}^{k-1} (1 - 2t_{k-j}) + 2 \sum_{j=0}^{k-1} P_{k-j} t_{k-j} \prod_{\ell=0}^{j-1} (1 - 2t_{k-\ell}), \quad k \geq 1.$$

Now, taking $t_k = \frac{1}{2k}$, 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 (2.3) and let

$$(2.8) \quad \delta_1 = \min \left( \frac{1}{C_{p, \rho, K}}, \frac{\rho(K \mu, B(a, r)^c)}{2p(2r)^{p-1}} \right).$$

We consider the time inhomogeneous $M$-valued Markov chain (2.6) in the particular case when

$$(2.9) \quad t_k = \min \left( \frac{\delta}{K}, \delta_1 \right), \quad k \geq 1$$

for some $\delta > 0$. The particular sequence $(t_k)_{k \geq 1}$ defined by (2.9) satisfies (2.4) and (2.5), 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 $T_{e_p}$-$M$-valued Markov chain $(Y^n_k)_{k \geq 0}$ by

$$(2.10) \quad Y^n_k = \frac{k}{\sqrt{n}} \exp_{e_p}^{-1} X_k.$$ 

We will prove convergence of the sequence of process $(Y^n_{\lfloor nt \rfloor})_{t \geq 0}$ 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 $\delta > C_{\rho,\mu,K}^{-1}$. Define

$$\Gamma = \mathbb{E} \left[ \text{grad}_{e_p} F_p(\cdot, P_1) \otimes \text{grad}_{e_p} F_p(\cdot, P_1) \right]$$

and $G_\delta(t)$ the generator

$$(2.11) \quad G_\delta(t)f(y) := \langle dy_f, t^{-1}(y - \delta \nabla dH_p(y, \cdot)^2) \rangle + \frac{\delta^2}{2} \text{Hess}_y f(\Gamma)$$

where $\nabla dH_p(y, \cdot)^2$ denotes the dual vector of the linear form $\nabla dH_p(y, \cdot)$.

There exists a unique inhomogeneous diffusion process $(y_\delta(t))_{t \geq 0}$ on $T_{e_p}M$ with generator $G_\delta(t)$ and converging in probability to 0 as $t \to 0^+$.

The process $y_\delta$ is continuous, converges a.s. to 0 as $t \to 0^+$ and has the following integral representation:

$$(2.12) \quad y_\delta(t) = \sum_{i=1}^{d} t^{1-\delta \lambda_i} \int_0^t s^{\delta \lambda_i - 1} \langle \delta \sigma dB_s, e_i \rangle e_i, \quad t \geq 0,$$

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 $\nabla dH_p(e_i)$ and $(\lambda_i)_{1 \leq i \leq d}$ are the associated eigenvalues.

Note that the integral representation (2.12) implies that $y_\delta$ is the centered Gaussian process with covariance

$$(2.13) \quad \mathbb{E} \left[ y_\delta(t_1) y_\delta(t_2) \right] = \frac{\delta^2 \Gamma(e_i^* \otimes e_j^*)}{\delta(\lambda_i + \lambda_j)} t_1^{1-\delta \lambda_i} t_2^{1-\delta \lambda_j} (t_1 \wedge t_2)^{(\delta(\lambda_i + \lambda_j) - 1)},$$

where $y_\delta(t) = \langle y_\delta(t), e_i \rangle$, $1 \leq i, j \leq d$ 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 $\delta > C_{\rho,\mu,K}^{-1}$. The sequence of processes $\left( Y_{n(t)}^n \right)_{t \geq 0}$ weakly converges in $D([0, \infty), T_{e_p}M)$ to $y_\delta$.

Remark 2.7. The assumption on $e_p$ implies that $H_p$ is of class $C^2$ in a neighbourhood of $e_p$. For most of the applications $\mu$ is equidistributed on a finite set of data which can be considered as randomly distributed. In this situation, when $p > 1$ then almost surely $e_p$ does not belong to the support of $\mu$. For $p = 1$ one has to be more careful since 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

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

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

Remark 2.9. In the case $M = \mathbb{R}^d$, $Y_k^n = \frac{1}{\sqrt{n}} (X_k - e_p)$ and the tangent space $T_{e_p}M$ is identified to $\mathbb{R}^d$. Theorem 2.6 holds and, in particular, when $t = 1$, 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 (2.13) 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 $e = \epsilon_p$ in the proofs.

3.1. Proof of Proposition 2.2.

For $p = 1$ this is a direct consequence of [22] 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 $x \in K \setminus \{e\}$, $t = \rho(e, x)$, $u \in T_eM$ the unit vector such that $\exp_x(\rho(e, x)u) = x$, and $\gamma_u$ the geodesic with initial speed $u$: $\dot{\gamma}_u(0) = u$. 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 $y \not\in \gamma([0, t])$ 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,

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

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

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

From this we get

$$(3.2) \quad h_y''(s) \geq pp(\gamma_u(s), y)^{p-2} \left( \min (p - 1, 2\alpha \cot (2\alpha)) \right).$$

Now

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

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

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

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

$$H_p(x) = H_p(\gamma_u(t)) = H_p(e) + \int_0^t (t - s) \left( \int_{B(a, r)} h_y(s)^p \mu(dy) \right) \, ds.$$
Using Equation (3.2) we get

\[ H_p(x) \geq H_p(e) + \int_0^t \left( (t - s) \int_{B(a,r)} pp(\gamma_u(s), y)^{p-2} \right) ds. \]

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

\[ H_p(x) \geq H_p(e) + t^2 \frac{p(2r)^{p-2}}{2} \min(p - 1, 2\alpha r \cot(2\alpha r)) \mu(dy). \]

So letting \( C_{p,\mu,K} = p(2r)^{p-2} \min(p - 1, 2\alpha r \cot(2\alpha r)) \)

we obtain

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

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 (2.1) and (2.2) 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 \( C_{p,\mu,K} \), (2.1) and (2.2) also holds on \( K \setminus B(e, \varepsilon) \).

\[ \square \]

3.2. Proof of Theorem 2.3.

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

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

with \( n(x,y) := \frac{\exp_x^{-1}(y)}{\rho(x,y)} \) a unit vector. So, with the condition (2.4) 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

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

we have in this case

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

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 \( \gamma(0) = -\text{grad}_{X_k} F_p(\cdot, P_{k+1}) \). We have for all \( t \in [0, t_{k+1}] \)
\[
E''(t) \leq C(\beta, r, p) := p^2(2r)^{2p-1} \beta \text{coth}(2\beta r)
\]
(see e.g. [22]). By Taylor formula,
\[
\begin{align*}
\rho(X_{k+1}, e)^2 &= 2E(t_{k+1}) \\
&= 2E(0) + 2t_{k+1}E'(0) + t_{k+1}^2 E''(t) \quad \text{for some } t \in [0, t_{k+1}] \\
&\leq \rho(X_k, e)^2 + 2t_{k+1} \langle \text{grad}_{X_k} F_p(\cdot, P_{k+1}), \exp_{X_k}^{-1}(e) \rangle + t_{k+1}^2 C(\beta, r, p).
\end{align*}
\]

Now from the convexity of \( x \mapsto F_p(x, y) \) we have for all \( x, y \in B(a, r) \)
\[
F_p(e, y) - F_p(x, y) \geq \langle \text{grad}_x F_p(\cdot, y), \exp_x^{-1}(e) \rangle.
\]
This applied with \( x = X_k, y = P_{k+1} \) yields
\[
\rho(X_{k+1}, e)^2 \leq \rho(X_k, e)^2 - 2t_{k+1} (F_p(X_k, P_{k+1}) - F_p(e, P_{k+1})) + C(\beta, r, p)t_{k+1}^2.
\]
Letting for \( k \geq 0 \) \( \mathcal{F}_k = \sigma(X_t, 0 \leq t \leq k) \), 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
\[
Y_0 = \rho(X_0, e)^2 \quad \text{and for } k \geq 1 \quad Y_k = \rho(X_k, e)^2 - C(\beta, r, p) \sum_{j=1}^{k} t_j^2
\]
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
\[
(3.10) \quad a = \lim_{k \to \infty} \mathbb{E} \left[ \rho(X_k, e)^2 \right].
\]
We want to prove that \( a = 0 \). We already proved that
\[
(3.11) \quad \mathbb{E} \left[ \rho(X_{k+1}, e)^2 | \mathcal{F}_k \right] \leq \rho(X_k, e)^2 - 2t_{k+1} (H_p(X_k) - H_p(e)) + C(\beta, r, p)t_{k+1}^2.
\]
Taking the expectation and using Proposition 2.2, we obtain
\[
(3.12) \quad \mathbb{E} \left[ \rho(X_{k+1}, e)^2 \right] \leq \mathbb{E} \left[ \rho(X_k, e)^2 \right] - t_{k+1} C_{p, \mu, K} \mathbb{E} \left[ \rho(X_k, e)^2 \right] + C(\beta, r, p)t_{k+1}^2.
\]
An easy induction proves that for \( \ell \geq 1 \),
\[
(3.13) \quad \mathbb{E} \left[ \rho(X_{k+\ell}, e)^2 \right] \leq \prod_{j=1}^{\ell} (1 - C_{p, \mu, K} t_{k+j}) \mathbb{E} \left[ \rho(X_k, e)^2 \right] + C(\beta, r, p) \sum_{j=1}^{\ell} t_{k+j}^2.
\]
Finally using $\ell \to \infty$ 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
\[
(3.14) \quad a \leq C(\beta, r, p) \sum_{j=1}^{\infty} t_{k+j}^2.
\]

Finally using $\sum_{j=1}^{\infty} t_{k+j}^2 < \infty$ 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
\[
t \mapsto E_p(t) := H_p(\gamma(t)),
\]
where $\gamma(t)_{t \in [0,t_k+1]}$ the geodesic satisfying $\dot{\gamma}(0) = -\operatorname{grad}_{X_k} F_p(\cdot, P_{k+1})$. With a calculation similar to (3.6) we get for all $t \in [0,t_k+1]$
\[
(3.15) \quad E''_p(t) \leq 2C(\beta, r, p) := p^3 (2r)^{p-4} (2r \beta \coth(2\beta r) + p - 2).
\]

(see e.g. [22]). By Taylor formula,
\[
H_p(X_{k+1}) = E_p(t_{k+1})
\]
\[
= E_p(0) + t_{k+1} E'_p(0) + \frac{t_{k+1}^2}{2} E''_p(t) \quad \text{for some } t \in [0,t_{k+1}]
\]
\[
\leq H_p(X_{k}) + t_{k+1} \langle d_{X_k} H_p, \operatorname{grad}_{X_k} F_p(\cdot, P_{k+1}) \rangle + t_{k+1}^2 C(\beta, r, p).
\]

We get
\[
\mathbb{E} [H_p(X_{k+1}) | \mathcal{F}_k]
\]
\[
\leq H_p(X_{k}) - t_{k+1} \langle d_{X_k} H_p, \int_{B(a, r)} \operatorname{grad}_{X_k} F_p(\cdot, y) \mu(dy) \rangle + C(\beta, r, p) t_{k+1}^2
\]
\[
= H_p(X_{k}) - t_{k+1} \langle \operatorname{grad}_{X_k} F_p(\cdot) \rangle + C(\beta, r, p) t_{k+1}^2
\]
\[
= H_p(X_{k}) - t_{k+1} \| \operatorname{grad}_{X_k} H_p(\cdot) \| \|^2 + C(\beta, r, p) t_{k+1}^2
\]
\[
\leq H_p(X_{k}) - C_{p,\mu,K} t_{k+1} (H_p(X_{k}) - H_p(e)) + C(\beta, r, p) t_{k+1}^2
\]
(by Proposition 1.1) so that the process $(Y_k)_{k \geq 0}$ defined by
\[
(3.16) \quad Y_0 = H_p(X_0) - H_p(e) \quad \text{and for } k \geq 1 \quad Y_k = H_p(X_k) - H_p(e) - C(\beta, r, p) \sum_{j=1}^{k} t_j^2
\]
is a bounded supermartingale. Now the argument is exactly the same as in the first part to prove that $H_p(X_k) - H_p(e)$ also converges in $L^1$ and almost surely to 0.

Finally (2.1) proves that $\rho(X_k, e)^p$ converges in $L^1$ and almost surely to 0. \qed
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

\begin{equation}
    dy_t = \frac{1}{t}L_\delta(y_t)\,dt + \delta\sigma\,dB_t
\end{equation}

where $L_\delta(y) = y - \delta\nabla dH_\rho(y, \cdot)^2$ and $B_t$ and $\sigma$ are as in Proposition 2.5. This sde can be solved explicitely on $[\varepsilon, \infty)$. The symmetric endomorphism $y \mapsto \nabla dH_\rho(y, \cdot)^2$ is diagonalisable in the orthonormal basis $(e_i)_{1 \leq i \leq d}$ with eigenvalues $(\lambda_i)_{1 \leq i \leq d}$. The endomorphism $L_\delta = \text{id} - \delta\nabla dH_\rho(\text{id}, \cdot)^2$ is also diagonalisable in this basis with eigenvalues $(1 - \delta\lambda_i)_{1 \leq i \leq d}$. The solution $y_t = \sum_{i=1}^{d} y_t^i e_i$ of (3.17) started at

\begin{equation}
    y_\varepsilon = \sum_{i=1}^{d} y_\varepsilon^i e_i
\end{equation}

is given by

\begin{equation}
    y_t = \sum_{i=1}^{d} \left( y_\varepsilon^i e_i \delta^{\lambda_i - 1} + \int_{\varepsilon}^{t} s^{\delta\lambda_i - 1} \langle \delta\sigma\,dB_s, e_i \rangle \right) t^{1 - \delta\lambda_i} e_i, \quad t \geq \varepsilon.
\end{equation}

Now by definition of $C_{\mu, K}$ we clearly have

\begin{equation}
    C_{\mu, K} \leq \min_{1 \leq i \leq d} \lambda_i.
\end{equation}

So the condition $\delta C_{\mu, K} > 1$ implies that for all $i$, $\delta\lambda_i - 1 > 0$, and as $\varepsilon \to 0$,

\begin{equation}
    \int_{\varepsilon}^{t} s^{\delta\lambda_i - 1} \langle \delta\sigma\,dB_s, e_i \rangle \to \int_{0}^{t} s^{\delta\lambda_i - 1} \langle \delta\sigma\,dB_s, e_i \rangle \quad \text{in probability.}
\end{equation}

Assume that a continuous solution $y_t$ converging in probability to 0 as $t \to 0^+$ exists. Since $y_\varepsilon^i e_i \delta^{\lambda_i - 1} \to 0$ in probability as $\varepsilon \to 0$, we necessarily have using (3.20)

\begin{equation}
    y_t = \sum_{i=1}^{d} t^{1 - \delta\lambda_i} \int_{0}^{t} s^{\delta\lambda_i - 1} \langle \delta\sigma\,dB_s, e_i \rangle e_i, \quad t \geq 0.
\end{equation}

Note $y_t^i$ is Gaussian with variance $\frac{t\delta^2 \Gamma(e_i^\ast \otimes e_i^\ast)}{2\delta\lambda_i - 1}$, so it converges in $L^2$ to 0 as $t \to 0$. Conversely, it is easy to check that equation (3.21) defines a solution to (3.17).

To prove the a.s. convergence to 0 we use the representation

\begin{equation*}
    \int_{0}^{t} s^{\delta\lambda_i - 1} \langle \delta\sigma\,dB_s, e_i \rangle = B^i_{\varphi_i(t)}
\end{equation*}

where $B^i_{\varphi_i}$ is a Brownian motion and $\varphi_i(t) = \frac{t\delta^2 \Gamma(e_i^\ast \otimes e_i^\ast)}{2\delta\lambda_i - 1} t^{2\delta\lambda_i - 1}$. Then by the law of iterated logarithm

\begin{equation*}
    \limsup_{t \to 0} t^{1 - \delta\lambda_i} B^i_{\varphi_i(t)} \leq \limsup_{t \to 0} t^{1 - \delta\lambda_i} \sqrt{2 \varphi_i(t) \ln(\varphi_i^{-1}(t))}
\end{equation*}

But for $t$ small we have

\begin{equation*}
    \sqrt{2 \varphi_i(t) \ln(\varphi_i^{-1}(t))} \leq t^{\delta\lambda_i - 3/4}
\end{equation*}

so

\begin{equation*}
    \limsup_{t \to 0} t^{1 - \delta\lambda_i} B^i_{\varphi_i(t)} \leq \lim_{t \to 0} t^{1/4} = 0.
\end{equation*}
This proves a.s. convergence to 0. Continuity is easily checked using the integral representation (3.21).

3.4. Proof of Theorem 2.6. Consider the time homogeneous Markov chain \((Z^n_k)_{k \geq 0}\) with state space \([0, \infty) \times T_\epsilon M\) 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 (2.6), (2.10) and (2.9), we have for \(k \geq k_0,
\]

\[
Y^n_{k+1} = \frac{nt^n_k + 1}{\sqrt{n}} \exp_{\epsilon}^{-1} \left( \exp_{\epsilon} \frac{1}{\sqrt{n}} \left( -\frac{\delta}{nt^n_k + 1} \text{grad}_{\epsilon} y^n_k F_p(\cdot, P_{k+1}) \right) \right).
\]

Consider the transition kernel \(P^n(z, dz')\) on \((0, \infty) \times T_\epsilon M\) defined for \(z = (t, y)\) by

\[
P^n(z, A) = \mathbb{P} \left[ \left( t + \frac{1}{n}, \frac{nt^n_k + 1}{\sqrt{n}} \exp_{\epsilon}^{-1} \left( \exp_{\epsilon} \frac{1}{\sqrt{n}} \left( -\frac{\delta}{nt^n_k + 1} \text{grad}_{\epsilon} y^n_k F_p(\cdot, P_{k+1}) \right) \right) \right) \in A \right]
\]

where \(A \in \mathcal{B}((0, \infty) \times T_\epsilon M)\). 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 \(e\) does not belong to the support \(\text{supp}(\mu)\) 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 \(e\)). Fix \(\delta > 0\). Let \(B\) be a bounded set in \(T_\epsilon M\) and let \(0 < e < T\). We have for all \(C^2\) function \(f\) on \(T_\epsilon M\)

\[
n \left( f \left( \frac{nt^n_k + 1}{\sqrt{n}} \exp_{\epsilon}^{-1} \left( \exp_{\epsilon} \frac{1}{\sqrt{n}} \left( -\frac{\delta}{nt^n_k + 1} \text{grad}_{\epsilon} y^n_k F_p(\cdot, x) \right) \right) \right) - f(y) \right) = \left( d_y f, \frac{y}{T} \right) - \sqrt{n} \left( d_y f, \delta \text{grad}_{\epsilon} F_p(\cdot, x) \right) - \delta \nabla dF_p(\cdot, x) \left( \text{grad}_{\epsilon} f, \frac{y}{T} \right)
\]

\[
+ \frac{\delta^2}{2} \text{Hess}_y f (\text{grad}_{\epsilon} F_p(\cdot, x) \otimes \text{grad}_{\epsilon} F_p(\cdot, x)) + O \left( \frac{1}{\sqrt{n}} \right)
\]

uniformly in \(y \in B, x \in \text{supp}(\mu), t \in [e, T]\).

Proof. Let \(x \in \text{supp}(\mu), y \in T_\epsilon M, u, v \in \mathbb{R}\) sufficiently close to 0, and \(q = \exp_{\epsilon} \left( \frac{uy}{T} \right)\). For \(s \in [0, 1]\) denote by \(a \mapsto c(a, s, u, v)\) the geodesic with endpoints \(c(0, s, u, v) = e\) and

\[
c(1, s, u, v) = \exp_{\epsilon} \left( \frac{sv}{T} \right) \left( -us \text{grad}_{\epsilon} \left( \frac{s}{T} \right) F_p(\cdot, x) \right)
\]:
c(a, s, u, v) = \exp_e \left\{ a \exp_e^{-1} \left[ \exp_{\exp_e(a)} \left( -s \nabla \exp_e(a) 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 $c(a, s) = c \left( a, s, \frac{1}{\sqrt{n}} \frac{\delta}{nt + 1} \right)$, we have

\[
\exp^{-1}_e \left( \exp_{\exp_e(a)} \left( - \frac{\delta}{nt + 1} \nabla \exp_e(a) F_p(\cdot, x) \right) \right) = \partial_a c(0, 1).
\]

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

We have $c(a, s, 0, 1) = \exp_e \left( -as \nabla \exp_e F_p(\cdot, x) \right)$ and this implies

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

On the other hand the identities $c(a, s, u, v) = c(a, s, u, 1)$ yields $\partial_a^2 \partial_a c(a, s, u, 1) = v^2 \partial_a^2 \partial_a c(a, s, u, 1)$, so we obtain

\[
\partial_a^2 \partial_a c(0, s, u, v) = O(u v^2)
\]

and this yields

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

uniformly in $s, x, y, t$. But since

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

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

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

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

On the other hand

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

so it remains to estimate $J(0)$.

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

\[
J(1, u, v) = -v \nabla \exp_e \left( \frac{uv}{\sqrt{t}} \right) F_p(\cdot, x)
\]

satisfies

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

This implies that

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

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

(3.28) \quad $P_{c(1,0),e} J(1) = J(0) + \dot{J}(0) + O(n^{-2}) = \dot{J}(0) + O(n^{-2})$.

But we also have

\[
P_{c(1,0,0,v),e} J(1, u, v) = P_{c(1,0,0,v),e} \left( -v \nabla \partial_a c(1,0,0,u,v) F_p(\cdot, x) \right)
\]

\[
= -v \nabla \partial_a F_p(\cdot, x) - v \nabla \partial_a c(0,0,u,v) \nabla F_p(\cdot, x) + O(vu^2)
\]

\[
= -v \nabla \partial_a F_p(\cdot, x) - v \nabla dF_p(\cdot, x) \left( \frac{uv}{t} \right)^{1/2} + O(vu^2)
\]
where we used \( \partial_a c(0,0,u,v) = \frac{uv}{T} \) and for vector fields \( A,B \) on \( TM \) and a \( C^2 \) function \( f_1 \) on \( M \)

\[
\langle \nabla_{A_e} \text{grad} f_1, B_e \rangle = A_e \langle \text{grad} f_1, B_e \rangle - \langle \text{grad} f_1, \nabla_{A_e} B \rangle = \nabla f_1(A_e, B_e)
\]

which implies

\[
\nabla_{A_e} \text{grad} f_1 = \nabla d f_1(A_e, \cdot)\]

We obtain

\[
P_{e(1,0),e} J(1) = -\frac{\delta}{nt+1} \text{grad}_e F_p(\cdot,x) - \frac{\delta}{\sqrt{n}(nt+1)} \nabla d F_p(\cdot,x) \left( \frac{y}{\sqrt{t}}, \cdot \right) + O(\frac{1}{n^2}).
\]

Combining with (3.28) this gives

\[
\dot{J}(0) = -\frac{\delta}{nt+1} \text{grad}_e F_p(\cdot,x) - \frac{\delta}{nt+1} \nabla d F_p(\cdot,x) \left( \frac{y}{\sqrt{nt}}, \cdot \right) + O\left( \frac{1}{n^2} \right)
\]

So finally

\[
\frac{nt+1}{\sqrt{n}} \partial_{ac}(0,1) = y + \frac{y}{nt} - \frac{\delta}{\sqrt{n}} \text{grad}_e F_p(\cdot,x) - \delta \nabla d F_p(\cdot,x) \left( \frac{y}{\sqrt{nt}}, \cdot \right) + O\left( \frac{1}{n^{3/2}} \right).
\]

To get the final result we are left to make a Taylor expansion of \( f \) up to order 2. □

Define the following quantities:

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

and

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

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 \( z \in [\varepsilon,T] \times B(0,e,R) \), where \( B(0,e,R) \) is the open ball in \( T_xM \) centered at \( 0 \) with radius \( R \),

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

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

\[
(3.33) \quad \lim_{n \to \infty} \sup_{z \in [\varepsilon,T] \times B(0,e,R)} |b_n(z) - b(z)| = 0
\]

with

\[
(3.34) \quad b(z) = \left( 1, \frac{1}{t} L_\delta(y) \right) \quad \text{and} \quad L_\delta(y) = y - \delta \nabla d H(y, \cdot)^2.
\]
(3) For all $R > 0$ and $\varepsilon > 0$, \( (3.35) \)

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

with

\( (3.36) \)

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

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

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

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

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

\[
= \mathbb{P} \left[ \left| \frac{nt + 1}{\sqrt{n}} \exp^{-1} \left( \exp_{\exp_e} \frac{\delta}{\sqrt{n}} y \left( -\frac{\delta}{nt + 1} \text{grad}_{\exp_e} \frac{\delta}{\sqrt{n}} y F_p(\cdot, P_1) \right) \right) - y \right| > 1 \right].
\]

On the other hand, since $F_p(\cdot, x)$ is of class $C^2$ in a neighbourhood of $e$, we have by (3.29)

\( (3.37) \)

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

for some constant $C > 0$.

(2) Equation (3.32) implies that for $n \geq n_0$

\[
b_n(z)
\]

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

\[
= n \left( \frac{1}{n} \mathbb{E} \left[ \left| \frac{nt + 1}{\sqrt{n}} \exp^{-1} \left( \exp_{\exp_e} \frac{\delta}{\sqrt{n}} y \left( -\frac{\delta}{nt + 1} \text{grad}_{\exp_e} \frac{\delta}{\sqrt{n}} y F_p(\cdot, P_1) \right) \right) - y \right| \right] \right).
\]

We have by lemma 3.1

\[
n \left( \frac{nt + 1}{\sqrt{n}} \exp^{-1} \left( \exp_{\exp_e} \frac{\delta}{\sqrt{n}} y \left( -\frac{\delta}{nt + 1} \text{grad}_{\exp_e} \frac{\delta}{\sqrt{n}} y F_p(\cdot, P_1) \right) \right) - y \right)
\]

\[
= \frac{1}{\varepsilon} \mathbb{E} \left[ \varepsilon \text{grad}_e F_p(\cdot, P_1) - \Delta \varepsilon dF_p(\cdot, P_1) \left( \frac{1}{\varepsilon} y, \cdot \right) \right] + O \left( \frac{1}{n^{1/2}} \right)
\]

a.s. uniformly in $n$, and since

\[
\mathbb{E} \left[ \varepsilon \text{grad}_e F_p(\cdot, P_1) \right] = 0,
\]

this implies that

\[
n \left( \mathbb{E} \left[ \frac{nt + 1}{\sqrt{n}} \exp^{-1} \left( \exp_{\exp_e} \frac{\delta}{\sqrt{n}} y \left( -\frac{\delta}{nt + 1} \text{grad}_{\exp_e} \frac{\delta}{\sqrt{n}} y F_p(\cdot, P_1) \right) \right) - y \right] \right)
\]

converges to

\( (3.38) \)

\[
\frac{1}{\varepsilon} y - \mathbb{E} \left[ \Delta \varepsilon dF_p(\cdot, P_1) \left( \frac{1}{\varepsilon} y, \cdot \right) \right] = \frac{1}{\varepsilon} y - \Delta \varepsilon dF_p \left( \frac{1}{\varepsilon} y, \cdot \right).
\]

Moreover the convergence is uniform in $z \in [\varepsilon, T] \times B(0, R)$, so this yields (3.33).
(3) In the same way, using lemma 3.1,
\[ n \int (y' - y) \otimes (y' - y) P^n(z, dz') \]
\[ = \frac{1}{n} \mathbb{E} \left[ \left( -\sqrt{n} \delta \text{grad}_e F_p(\cdot, P_1) \right) \otimes \left( -\sqrt{n} \delta \text{grad}_e F_p(\cdot, P_1) \right) \right] + o(1) \]
\[ = \delta^2 \mathbb{E} \left[ \text{grad}_e F_p(\cdot, P_1) \otimes \text{grad}_e F_p(\cdot, P_1) \right] + o(1) \]
uniformly in \( z \in [\epsilon, T] \times B(0, R) \), so this yields (3.35).

\[ \text{□} \]

**Lemma 3.3.** Suppose that \( t_n = \frac{\delta}{n} \) for some \( \delta > 0 \). For all \( \delta > C_{p,\mu,K}^{-1} \),
\[ (3.39) \quad \sup_{n \geq 1} n \mathbb{E} \left[ \rho^2(e, X_n) \right] < \infty. \]

**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
\[ (3.40) \quad \mathbb{E} \left[ \rho^2(e, X_{k+1}) \right] \leq \mathbb{E} \left[ \rho^2(e, X_k) \right] \exp \left( -C_{p,\mu,K} t_{k+1} \right) + C(\beta, r, p) t_{k+1}^2. \]

From this (3.39) is a consequence of Lemma 0.0.1 (case \( \alpha > 1 \)) in [16]. We give the proof for completeness. We deduce easily by induction that for all \( k \geq k_0 \),
\[ (3.41) \quad \mathbb{E} \left[ \rho^2(e, X_k) \right] \]
\[ \leq \mathbb{E} \left[ \rho^2(e, X_{k_0}) \right] \exp \left( -C_{p,\mu,K} \sum_{j=k_0+1}^{k} t_j \right) + C(\beta, r, p) \sum_{j=k_0}^{k} t_j^2 \exp \left( -C_{p,\mu,K} \sum_{j=k_0+1}^{k} t_j \right), \]

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 \( k \geq i \):
\[ (3.42) \quad \sum_{j=i+1}^{k} t_j = \delta \sum_{j=i+1}^{k} \frac{1}{j} \geq \delta \int_{i+1}^{k+1} dt \geq \delta \ln \frac{k+1}{i+1}. \]

Hence,
\[ (3.43) \quad \mathbb{E} \left[ \rho^2(e, X_k) \right] \]
\[ \leq \mathbb{E} \left[ \rho^2(e, X_{k_0}) \right] \left( \frac{k_0 + 1}{k + 1} \right)^{\delta C_{p,\mu,K}^{-1}} + \frac{\delta^2 C(\beta, r, p)}{(k + 1)^{\delta C_{p,\mu,K}^{-1}}} \sum_{i=k_0+1}^{k} \frac{(i + 1)^{\delta C_{p,\mu,K}^{-1}}}{i^2}. \]

For \( \delta C_{p,\mu,K}^{-1} > 1 \) we have as \( k \to \infty \)
\[ (3.44) \quad \frac{\delta^2 C(\beta, r, p)}{(k + 1)^{\delta C_{p,\mu,K}^{-1}}} \sum_{i=k_0+1}^{k} \frac{(i + 1)^{\delta C_{p,\mu,K}^{-1}}}{i^2} \sim \frac{\delta^2 C(\beta, r, p)}{(k + 1)^{\delta C_{p,\mu,K}^{-1}}} \frac{k^{\delta C_{p,\mu,K}^{-1}-1}}{\delta C_{p,\mu,K}^{-1} - 1} \]
and
\[ \mathbb{E} \left[ \rho^2(e, X_{k_0}) \right] \left( \frac{k_0 + 1}{k + 1} \right)^{\delta C_{p,\mu,K}^{-1}} \sim o(k^{-1}). \]

This implies that the sequence \( k \mathbb{E} \left[ \rho^2(e, X_k) \right] \) is bounded.

Next consider the case \( p \geq 2 \).
From the proof of Theorem 2.3 we have
\begin{equation}
E[H_p(X_{k+1}) - H_p(e)] \leq (1 - t_{k+1}C_{p,\mu,K})E[H_p(X_k) - H_p(e)] + C(\beta,r,p)t_{k+1}^2
\end{equation}
which implies
\begin{equation}
E[H_p(X_{k+1}) - H_p(e)] \leq E[H_p(X_k) - H_p(e)] \exp(-C_{p,\mu,K}t_{k+1}) + C(\beta,r,p)t_{k+1}^2.
\end{equation}
From this, arguing similarly, we obtain that the sequence \( kE[H_p(X_k) - H_p(e)] \) is bounded. We conclude with (2.1).

**Lemma 3.4.** Assume \( \delta > C^{-1} p,\mu,K \) and that \( H_p \) is \( C^2 \) in a neighbourhood of \( e \). For all \( 0 < \varepsilon < T \), the sequence of processes \( \left( Y^n_{\varepsilon,nt} \right)_{\varepsilon \leq t \leq T} \) is tight in \( \mathbb{D}([\varepsilon,1],\mathbb{R}^d) \).

**Proof.** Denote by \( \left( \tilde{Y}^n_{\varepsilon} = \left( Y^n_{\varepsilon,nt} \right)_{\varepsilon \leq t \leq T} \right)_{n \geq 1} \), the sequence of processes. We prove that from any subsequence \( \left( \tilde{Y}^\phi_{\varepsilon}(n) \right)_{n \geq 1} \) we can extract a further subsequence \( \left( \tilde{Y}^\psi_{\varepsilon}(n) \right)_{n \geq 1} \) that weakly converges in \( \mathbb{D}([\varepsilon,1],\mathbb{R}^d) \).

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

\[
\left\| \tilde{Y}^\phi_{\varepsilon}(n)(\varepsilon) \right\|^2 = \left\| \frac{\phi(n)\varepsilon}{\phi(n)} \right\| E[ \rho^2(e,X_{\phi(n)\varepsilon})] \leq \varepsilon \sup_{n \geq 1} \left\{ nE[\rho^2(e,X_n)] \right\}
\]
and the last term is bounded by lemma 3.3.

Consequently \( \left( \tilde{Y}^\phi_{\varepsilon}(n)(\varepsilon) \right)_{n \geq 1} \) is tight. So there is a subsequence \( \left( \tilde{Y}^\phi_{\varepsilon}(n)(\varepsilon) \right)_{n \geq 1} \) that weakly converges in \( T,M \) 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 \( \left( \tilde{Y}^\phi_{\varepsilon}(n) \right)_{n \geq 1} \) weakly converges to a diffusion \( (y_\varepsilon)_{\varepsilon \leq t \leq T} \) with generator \( G_\varepsilon(t) \) given by (2.11) and such that \( y_\varepsilon \) has law \( \nu_\varepsilon \). This achieves the proof of lemma 3.4. \( \square \)

**Proof of Theorem 2.6.** Let \( \tilde{Y}^n = \left( Y^n_{\varepsilon,nt} \right)_{0 \leq t \leq T} \). It is sufficient to prove that any subsequence of \( \left( \tilde{Y}^n \right)_{n \geq 1} \) has a further subsequence which converges in law to \( (y_\varepsilon(t))_{0 \leq t \leq T} \). So let \( \left( \tilde{Y}^\phi(n) \right)_{n \geq 1} \) 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 \( \tilde{Y}^\eta(n) \). 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

\[
\left\| \tilde{Y}^\eta(n)(t) \right\|^2 \leq Ct
\]
for some \( C > 0 \). So \( \left\| \tilde{Y}^\eta(n)(t) \right\|^2 \to 0 \) as \( t \to 0 \), which in turn implies \( \left\| y'(t) \right\|^2 \to 0 \) as \( t \to 0 \). The unicity statement in Proposition 2.5 implies that \( (y'(t))_{t \in (0,T]} \) and \( (y_\varepsilon(t))_{t \in (0,T]} \) are equal in law. This achieves the proof. \( \square \)
REFERENCES


