On the Riemannian barycentre of a Markov chain

Salem Said

Abstract

The Riemannian barycentre is one of the most widely used statistical descriptors for probability distributions on Riemannian manifolds. At present, existing algorithms are able to compute the Riemannian barycentre of a probability distribution, only if i.i.d. samples of this distribution are readily available. However, there are many cases where i.i.d. samples are quite difficult to obtain, and have to be replaced with non-independent samples, generated by a Markov chain Monte Carlo method. To overcome this difficulty, the present paper proposes a new Markov chain Monte Carlo algorithm for computing the Riemannian barycentre of a probability distribution on a Hadamard manifold (a simply connected, complete Riemannian manifold with non-positive curvature). This algorithm relies on two original propositions, proved in the paper. The first proposition states that the recursive barycentre of samples generated from a geometrically ergodic Markov chain converges in the mean-square to the Riemannian barycentre of the stationary distribution of this chain. The second proposition provides verifiable conditions which ensure a Metropolis-Hastings Markov chain, with its values in a symmetric Hadamard manifold, is geometrically ergodic. This latter result yields a partial solution, in the context of Riemannian manifolds, to the problem of geometric ergodicity of Metropolis-Hastings chains, which has previously attracted extensive attention when considered in Euclidean space. In addition to these two propositions, the new Markov chain Monte Carlo algorithm, proposed in this paper, is applied to a problem of Bayesian inference, arising from computer vision.

Index Terms

Markov chain, Riemannian barycentre, Metropolis-Hastings, geometric ergodicity, Bayesian inference, symmetric space

I. INTRODUCTION

A Riemannian barycentre of a probability distribution $\pi$ on a Riemannian manifold $M$ is a global minimiser of the so-called variance function [1]–[3]

$$E(z) = \mathbb{E}_\pi d^2(z, x) \quad \text{for} \quad z \in M$$

(1)

where $\mathbb{E}_\pi$ denotes expectation with respect to a random variable $x$ distributed according to $\pi$, and where $d(\cdot, \cdot)$ denotes Riemannian distance. If its variance function $E$ has finite values, then the distribution $\pi$ has at least one Riemannian barycentre. However, in general, this Riemannian barycentre is not unique.

The Riemannian barycentre was originally proposed by Fréchet, as a statistical descriptor for probability distributions on Riemannian manifolds [4]. This was motivated by the fact that, if $M$ is a Euclidean space, then any probability distribution $\pi$ has at most one Riemannian barycentre, identical to its mean (first-order moment). In this way, modulo the issue of its uniqueness, the Riemannian barycentre generalises the concept of mean, from Euclidean space to Riemannian manifolds.

Currently, the Riemannian barycentre has become a widely popular tool for data analysis on Riemannian manifolds [5]–[7]. In fact, the Riemannian manifolds involved in many applications are Hadamard manifolds (simply connected, complete Riemannian manifolds with non-positive sectional curvature). For these manifolds, the Riemannian barycentre of a probability distribution, if it exists, is guaranteed to be unique. In radar signal processing, medical imaging, and remote sensing, Hadamard manifolds arise in the form of spaces of covariance matrices (real, complex, Toeplitz, etc.), and play an increasing role in many applied problems [7]–[9]. These spaces of covariance matrices are examples of symmetric Hadamard manifolds. That is, of Hadamard manifolds which are symmetric spaces [10].

The present paper is concerned with the problem of computing the Riemannian barycentre $b$ of a probability distribution $\pi$ on a Hadamard manifold $M$. Currently [1]–[3], existing algorithms are able to address this problem, only if i.i.d. samples $(x_n; n \geq 1)$ of the distribution $\pi$ are available. Their idea is to construct, from these i.i.d. samples, a sequence of recursive barycentres $(s_n; n \geq 1)$ which converge to $b$.

This construction is best motivated by considering what happens when $M$ is a Euclidean space, so $M = \mathbb{R}^d$ (note that a Euclidean space is a Hadamard manifold with zero sectional curvature). Since $b$ is non other than the mean of the distribution $\pi$, the weak law of large numbers implies that $s_n = (x_1 + \ldots + x_n)/n$ converge to $b$ in the mean square. In addition, it is well-known that one can compute recursively $s_{n+1} = \frac{n}{n+1} s_n + \frac{1}{n+1} x_{n+1}$.
This recursive approach to updating \( s_n \) can be understood from a geometric point of view, as follows: If \( \gamma : [0, 1] \to \mathbb{R}^d \) is the straight line segment with equation \( \gamma(t) = (1-t) s_n + t x_{n+1} \), then \( s_{n+1} = \gamma(\frac{1}{t}) \).

Fortunately, this geometric understanding generalises immediately to any Hadamard manifold \( M \). Indeed \([11]\), recall that any points \( x \) and \( y \) in \( M \) are connected by a unique geodesic curve \( \gamma : [0, 1] \to M \), with \( \gamma(0) = x \) and \( \gamma(1) = y \) (if \( M = \mathbb{R}^d \), this reduces to the segment just described). Then, for each \( t \in [0, 1] \), define the weighted geodesic mean of \( x \) and \( y \) by

\[
x \#_{t} y = \gamma(t)
\]

In \([1][8]\), the sequence of recursive barycentres was constructed, based on \([2]\), by computing \( s_{n+1} = s_n \#_{\frac{1}{n+1}} x_{n+1} \) with the initial value \( s_1 = x_1 \). In \([1]\), the weak law of large numbers on Hadamard manifolds was proved. This states that, if the \( x_n \) are i.i.d. samples of the distribution \( \pi \), then the recursive barycentres \( s_n \) converge to the Riemannian barycentre \( b \) of \( \pi \), in the mean square Riemannian distance of \( M \).

The present paper is motivated by the fact that, in many situations, such as Bayesian inference or stochastic filtering, i.i.d. samples of the target distribution \( \pi \) may be especially difficult to obtain. In particular, this makes it impossible to apply the algorithms in \([1][3]\). To overcome this difficulty, a new Markov chain Monte Carlo algorithm is proposed in Section \( [V] \) below. This algorithm relies on the following original propositions.

Proposition 1 of Section \( [III] \) provides the weak law of large numbers for Markov chains in Hadamard manifolds. This generalises the weak law proved in \([1]\), from the special case where the \( x_n \) are i.i.d. samples of the distribution \( \pi \), to the general case where the \( x_n \) are generated from a geometrically ergodic Markov chain with stationary distribution \( \pi \).

Proposition 2 of Section \( [III] \) provides verifiable conditions which ensure that a Metropolis-Hastings Markov chain, with its values in a symmetric Hadamard manifold, is geometrically ergodic. This generalises the conditions described in \([12][13]\), for Metropolis-Hastings Markov chains in Euclidean space, to the context of symmetric Hadamard manifolds.

Combining Propositions 1 and 2, in order to compute the Riemannian barycentre \( b \) of the distribution \( \pi \), the new algorithm, proposed in Section \( [V] \), generates samples \( x_n \) from a geometrically ergodic Markov chain, with stationary distribution \( \pi \), and simultaneously constructs from these samples the sequence of recursive barycentres \( s_n \). In Section \( [V] \) this algorithm is further applied to a problem of Bayesian inference, arising from computer vision \([6]\).

In the remainder of this paper, it will be useful to keep in mind some general properties of Hadamard manifolds and of probability distributions on these manifolds.

The first among these is the strong convexity property of the squared distance function \([1]\). Specifically, for any points \( x, y \) and \( z \) in a Hadamard manifold \( M \), and for each \( t \in [0, 1] \),

\[
d^2(z, x \#_{t} y) \leq (1-t) d^2(z, x) + t d^2(z, y) - t(1-t) d^2(y, x)
\]

Recalling from \([2]\) that \( x \#_{t} y = \gamma(t) \), it follows from \([3]\) that the squared Riemannian distance \( d^2(z, \gamma(t)) \) is a strongly convex function of \( t \) along the geodesic curve \( \gamma \).

The second property is called the variance inequality \([1]\). If \( \mu \) is a probability distribution on \( M \), denote \( \mathcal{E}_\mu \) the variance function of \( \mu \),

\[
\mathcal{E}_\mu(z) = \mathbb{E}_\mu d^2(z, x) \quad \text{for} \quad z \in M
\]

If this function has finite values, then \( \mu \) has a unique Riemannian barycentre \( b_\mu \). By definition, \( b_\mu \) is the unique global minimiser of \( \mathcal{E}_\mu \). The variance inequality states that, for any \( z \in M \),

\[
\mathcal{E}_\mu(z) \geq \text{Var}_\mu + d^2(z, b_\mu)
\]

where \( \text{Var}_\mu = \mathcal{E}_\mu(b_\mu) \) is the minimum value of the variance function \( \mathcal{E}_\mu \) (this should be called the variance of \( \mu \)).

The third property concerns the behavior of the Riemannian barycentre \( b_\mu \) under isometries of \( M \). To state this, assume that \( \mu(dz) = \mu(z) \text{vol}(dz) \), where \( \mu(z) \) is a probability density function and \( \text{vol}(dz) \) denotes the Riemannian volume of \( M \), and let \( g : M \to M \) be an isometry. Then \([1][11]\),

\[
\mu \circ g = \mu \implies g \cdot b_\mu = b_\mu
\]

where \( g \cdot b_\mu = g(b_\mu) \). In other words, if \( g \) preserves the distribution \( \mu \), then \( g \) fixes the barycentre \( b_\mu \).

Finally, the following property will be needed for the application to Bayesian inference, considered in Section \( [IV] \). Let \( x \) and \( y \) be two points in \( M \) and consider the function

\[
\mathcal{E}_\mu(z) = (1-t) d^2(z, x) + t d^2(z, y) \quad \text{for} \quad z \in M
\]

for some fixed \( t \in [0, 1] \). Then, this function has a unique global minimum over \( M \), achieved at \( z = x \#_{t} y \). In other words, the geodesic mean of two points \( x \) and \( y \) minimises the weighted sum of squared Riemannian distances to these two points.
II. THE WEAK LAW OF LARGE NUMBERS FOR MARKOV CHAINS

Let $M$ be a Hadamard manifold (a simply connected, complete Riemannian manifold with non-positive curvature) \([1]\). Let \((x_n : n \geq 1)\) be a Markov chain in $M$, and consider the sequence of recursive barycentres \((s_n : n \geq 1)\), defined as in \([1,3]\)

\[
s_{n+1} = s_n \# \frac{1}{\pi(x_n)} \ x_{n+1} \quad \text{for} \ s_1 = x_1 \tag{5}
\]

where the geodesic mean operation “\#” was defined in \([2]\). If the distributions $\pi_n$ of the $x_n$ converge to a stationary distribution $\pi$, then one expects the recursive barycentres $s_n$ to converge to the Riemannian barycentre $b$ of $\pi$. Proposition \([4]\) below formulates a sufficient condition for this convergence $s_n \to b$ to hold in mean square Riemannian distance.

Precisely, this is a so-called geometric drift condition, which can be stated as follows \([14]\). Assume there exists a continuous function $V : M \to \mathbb{R}$ which satisfies, for some point $x^* \in M$,

\[
V(z) \geq \max \{ 1, d^2(z, x^*) \} \tag{6a}
\]

for all $z \in M$, and which also satisfies the geometric drift condition

\[
PV(z) \leq \lambda V(z) + bI_C(z) \tag{6b}
\]

where $P$ is the transition kernel of the Markov chain $(x_n)$, and where $\lambda < 1$, $b < \infty$, and $C$ is a small set for $P$ (for extensive background on \((6b)\), the reader may refer to \([14]\)).

**Proposition 1:** Assume that the transition kernel $P$ of the Markov chain $(x_n)$ is irreducible and aperiodic, and that the geometric drift condition \((6)\) is verified. Then, the distributions $\pi_n$ of the $x_n$ converge to a stationary distribution $\pi$, and the Riemannian barycentres $b_n$ of $\pi_n$ and $b$ of $\pi$ are well-defined. Moreover,

(i) the $b_n$ converge geometrically to $b$, in the sense that there exists $\sigma < 1$ such that

\[
d(b_n, b) = O(\sigma^n) \tag{7a}
\]

(ii) the recursive barycentres $s_n$ converge to $b$ in mean square Riemannian distance,

\[
E d^2(s_n, b) \leq \frac{\sup \mathcal{E}_n(b)}{n} + O \left( \frac{1}{n^2} \right) \tag{7b}
\]

where $\mathcal{E}_n$ is the variance function of the probability distribution $\pi_n$, and the supremum is over $n \geq 1$.

**Remark:** the geometric drift condition \((6)\) implies that the Markov chain $(x_n)$ is geometrically ergodic. That is \([14]\), the distributions $\pi_n$ converge geometrically to the stationary distribution $\pi$, in the sense that there exists $\rho < 1$ and $R(x_i) < \infty$ such that

\[
\left| \int_M f(x) [\pi_n(dx) - \pi(dx)] \right| \leq R(x_i) \rho^n \tag{8}
\]

for any function $f : M \to \mathbb{R}$ with $|f| \leq V$. Proposition \([4]\) uses the geometric drift condition \((6)\) only through the geometric ergodicity property \((8)\).

Part (i) of the proposition shows that the geometric convergence of probability distributions $\pi_n \to \pi$ implies the geometric convergence of Riemannian barycentres $b_n \to b$. In fact, if \((8)\) holds for some $\rho < 1$, then \((7a)\) holds for $\sigma = \rho^2$.

On the other hand, Part (ii) establishes the convergence of the recursive barycentres $s_n$, computed along any trajectory of the Markov chain $(x_n)$, to the barycentre $b$ of the stationary distribution $\pi$ of this chain. In \((7b)\), $\mathcal{E}_n$ is the variance function of the distribution $\pi_n$, defined as in \([1]\).

\[
\mathcal{E}_n(z) = \mathbb{E}_{\pi_n} d^2(z, x_n) \quad \text{for} \ z \in M \tag{9}
\]

Moreover, the fact that $\sup \mathcal{E}_n(b) < \infty$ will be obtained as a byproduct of the proof of \((7a)\).

Part (ii) of Proposition \([4]\) generalises the weak law of large numbers obtained in \([1]\), from the i.i.d. case to the Markov chain case. In particular, if the chain $(x_n)$ is an i.i.d. sequence, or even if this chain is stationary (that is, if $\pi_n = \pi$ for all $n$, but the $x_n$ are not required to be independent), then the inequality in \((7b)\) becomes sharper, as the second term on the right-hand side becomes identically zero. In addition, since all $\pi_n$ are equal to $\pi$, $\sup \mathcal{E}_n(b)$ becomes equal to $\mathcal{E}(b) = \text{Var}_\pi$. Therefore, the inequality in \((7b)\) reduces to the one obtained in \([1]\).

\[
E d^2(s_n, b) \leq \frac{\text{Var}_\pi}{n} \tag{10}
\]

Accordingly, Part (ii) of Proposition \([4]\) may be considered as the weak law of large numbers for Markov chains in Hadamard manifolds.
Proof of Proposition 1: the proof of (i) is carried out in Appendix A. The proof of (ii) relies on the following calculation (compare to [[1], Proof of Proposition 4.7]).

By application of the strong convexity property (3) to the definition of the $s_n$ in (5), it follows that for $n > 1$,

$$d^2(b, s_n) \leq \frac{n-1}{n} d^2(b, s_{n-1}) + \frac{1}{n} d^2(b, x_n) - \frac{n-1}{n^2} d^2(x_n, s_{n-1}).$$

Taking expectations in this inequality, it follows from (9),

$$\mathcal{V}_n \leq \frac{n-1}{n} \mathcal{V}_{n-1} + \frac{1}{n} \mathcal{E}_n(b) - \frac{n-1}{n^2} \mathbb{E} d^2(s_{n-1}, x_n)$$

(11a)

where $\mathcal{V}_n = \mathbb{E} d^2(s_n, b)$. But, by applying the variance inequality (4b),

$$\mathbb{E} \mathcal{V}_n d^2(s_{n-1}, x_n) = \mathcal{E}_n(s_{n-1}) \geq \text{Var}_n + d^2(s_{n-1}, b_n)$$

where $\text{Var}_n = \text{Var}_{s_n}$. Then, by taking expectations, it follows that

$$\mathbb{E} d^2(s_{n-1}, x_n) \geq \text{Var}_n + \mathbb{E} d^2(s_{n-1}, b_n)$$

(11b)

Applying this inequality to the third term on the right-hand side of (11a) gives

$$\mathcal{V}_n \leq \frac{n-1}{n} \mathcal{V}_{n-1} + \frac{1}{n} \mathcal{E}_n(b) - \frac{n-1}{n^2} \left[ \text{Var}_n + \mathbb{E} d^2(s_{n-1}, b_n) \right]$$

(11c)

which will be written

$$\mathcal{V}_n \leq \left( \frac{n-1}{n} \right)^2 \mathcal{V}_{n-1} + \frac{1}{n^2} \mathcal{U}_n^{(1)} + \frac{n-1}{n^2} \left[ \mathcal{U}_n^{(3)} - \mathcal{U}_n^{(2)} \right]$$

(12a)

with the notation

$$\mathcal{U}_n^{(1)} = \mathcal{E}_n(b) ; \quad \mathcal{U}_n^{(2)} = \mathbb{E} d^2(s_{n-1}, b_n) - \mathcal{V}_{n-1} ; \quad \mathcal{U}_n^{(3)} = \mathcal{E}_n(b) - \text{Var}_n$$

(12b)

By induction, it follows from (12a) that

$$\mathcal{V}_n \leq \frac{1}{n^2} \mathcal{V}_1 + \frac{1}{n^2} \sum_{m=2}^{n} \mathcal{U}_m^{(1)} + \frac{1}{n^2} \sum_{m=2}^{n} (m-1) \left[ \mathcal{U}_m^{(3)} - \mathcal{U}_m^{(2)} \right]$$

(13a)

The proof of (7b) will follow from (13a) by the following estimates, which are obtained in Appendix A

$$\sup_n \mathcal{U}_n^{(1)} = \sup_n \mathcal{E}_n(b) < \infty \quad (13b)$$

$$\mathcal{U}_n^{(2)} = O(\rho^{n/2}) ; \quad \mathcal{U}_n^{(3)} = O(\rho^{n/2}) \quad (13c)$$

Indeed, it is clear from (13b), that the second term on the right-hand side of (13a) satisfies

$$\frac{1}{n^2} \sum_{m=2}^{n} \mathcal{U}_m^{(1)} \leq \frac{1}{n^2} \left( n \times \sup_n \mathcal{E}_n(b) \right) = \frac{\sup_n \mathcal{E}_n(b)}{n}$$

On the other hand, (13c) implies that the series in the third term on the right-hand side of (13a) converges absolutely. Replacing in (13a), and recalling $\mathcal{V}_n = \mathbb{E} d^2(s_n, b)$, it then follows

$$\mathbb{E} d^2(s_n, b) \leq \frac{\sup_n \mathcal{E}_n(b)}{n} + \frac{1}{n^2} \sum_{m=2}^{\infty} (m-1) |\mathcal{U}_m^{(3)} - \mathcal{U}_m^{(2)}|$$

(14)

after noting that $\mathcal{V}_1 = \mathcal{E}_1(b)$. Finally, (14) is the same as (7b). ■

Remark: if the Markov chain $(x_n)$ is stationary, then all the distributions $\pi_n$ are identical to the stationary distribution $\pi$. This implies that $b_n = b$ and $\text{Var}_n = \text{Var}_x$ for all $n \geq 1$. Replacing this into (12b), it follows from the definition of $\mathcal{V}_n$ that $\mathcal{U}_n^{(2)}$ is identically zero. Similarly, since $\mathcal{E}_n(b) = \mathcal{E}(b) = \text{Var}_x$ for all $n$, it follows that $\mathcal{U}_n^{(3)}$ is also identically zero. With these simplifications, (14) reduces to (10), as discussed after Proposition 1.

In [1], (10) was established in the special case where the chain $(x_n)$ is an i.i.d. sequence. However, the proof of Proposition 1 shows that it holds, more generally, whenever $(x_n)$ is a stationary Markov chain. ■

The proof of Proposition 1 does not employ the differentiable structure of the Hadamard manifold $M$, but relies solely on its metric properties, such as the strong convexity property (3). In fact, Proposition 1 is true, without any change to its statement, even if $M$ is not a differentiable manifold, but more generally a metric space of non-positive curvature (these are discussed in [13, Chapter 9]). Still, the development in the present paper is limited to the case where $M$ has a differentiable manifold structure, as this seems more familiar in applications.
III. GEOMETRIC ERGODICITY OF METROPOLIS CHAINS

Proposition \[1\] of the previous section assumed the existence of a Markov chain \((x_n)\) which verifies the geometric drift condition (6). Thus, in order to apply this proposition, it remains to construct a Markov chain which actually verifies this condition. The present section shows that this can be done using the Metropolis-Hastings method \[16\].

Proposition \[2\] below, states that if \((x_n)\) is a Metropolis chain (precisely, an isotropic Metropolis-Hastings Markov chain), with values in a symmetric Hadamard manifold \(M\), and whose stationary distribution \(\pi\) has sub-Gaussian tails, then \((x_n)\) verifies the geometric drift condition (6). Here, a symmetric Hadamard manifold is a Hadamard manifold which is a symmetric space \[10\] (in particular, this could be a hyperbolic space, or a space of covariance matrices \[7\]).

For the statement of Proposition \[2\] assume that the stationary distribution \(\pi\) can be written \(\pi(dz) = \pi(z)\text{vol}(dz)\), where \(\pi(z)\) is a probability density function and \(\text{vol}(dz)\) denotes the Riemannian volume measure of \(M\). Recall that \((x_n)\) is a Metropolis-Hastings chain if its transition kernel \(P\) is given by \[16\][13],

\[
Pf(z) = \int_M \alpha(z, y) q(z, y) f(y) \text{vol}(dy) + \rho(z) f(z)
\]

for any bounded measurable function \(f : M \to \mathbb{R}\), where \(\alpha(z, y)\) is the probability of accepting a transition from \(z\) to \(dy\), and \(\rho(z)\) is the probability of staying at \(z\), and where \(q(z, y)\) is the proposed transition density, so \(q(z, y) \geq 0\) and

\[
\int_M q(z, y) \text{vol}(dy) = 1 \quad \text{for} \quad z \in M
\]

Then, assume that \((x_n)\) is an isotropic Metropolis-Hastings chain, in the sense that \(q(z, y) = q(d(z, y))\) depends only on the Riemannian distance \(d(z, y)\). In particular, this implies the acceptance probability \(\alpha(z, y)\) is given by \(\alpha(z, y) = \min \{1, \pi(y)/\pi(z)\}\). Now, consider the following additional assumptions.

(A1) the stationary distribution \(\pi\) has sub-Gaussian tails. That is, the probability density function \(\pi(z)\) is positive and differentiable, and there exists some point \(x^* \in M\) such that \(r(z) = d(z, x^*)\) and \(\ell(z) = \log \pi(z)\) satisfy

\[
\limsup_{r(z) \to \infty} \langle \nabla r, \nabla \ell\rangle_z < 0
\]

where \(\langle \cdot, \cdot \rangle\) denotes the Riemannian metric of \(M\), and \(\nabla\) denotes the gradient with respect to this metric.

(A2) the radial component of the gradient \(\nabla \ell\) is persistent. That is,

\[
\limsup_{r(z) \to \infty} \langle \nabla r, n\rangle_z < 0
\]

where \(n(z) = \nabla \ell(z)/\|\nabla \ell(z)\|\) and \(\|\cdot\|\) denotes the Riemannian norm.

(A3) the proposed transition density \(q(z, y)\) is bounded away from zero near zero. That is, there exist \(\delta_0 > 0\) and \(\epsilon_0 > 0\) such that \(d(z, y) < \delta_0\) implies \(q(z, y) > \epsilon_0\).

Proposition \[2\] generalises, to the context of symmetric Hadamard manifolds, the results about geometric ergodicity of Metropolis-Hastings Markov chains with values in a Euclidean space, which were obtained in \[13\][12].

Proposition 2: Assume \((x_n)\) is an isotropic Metropolis-Hastings Markov chain, with values in a symmetric Hadamard manifold \(M\), and which satisfies Assumptions (A1) – (A3). Then \((x_n)\) verifies the geometric drift condition (6) (and therefore the geometric ergodicity condition (8)).

The proof of Proposition 2 will be carried out in Appendix B. It is a generalisation of the proof, carried out in the special case where \(M\) is a Euclidean space, in \[13\]. The idea is to use Assumptions (A1) – (A3) to show that the following two conditions hold

\[
\limsup_{r(z) \to \infty} \frac{PV(z)}{V(z)} < 1\quad (16a)
\]

\[
\sup_{z \in M} \frac{PV(z)}{V(z)} < \infty\quad (16b)
\]

where \(V(z) = c\pi^{-\frac{1}{2}}(z)\) with \(c\) chosen so \(V(z) \geq 1\) for all \(z \in M\). However, under Assumption (A3), these two conditions are shown to imply (65). On the other hand, (6a) is a straightforward result of Assumption (A1), which implies the existence of strictly positive \(\beta\), \(R\), and \(\pi_R\) such that

\[
r(z) \geq R \implies \pi(z) \leq \pi_R \exp(-\beta r^2(z))
\]

Then, to obtain (6a), it is enough to chose \(c = \max \left\{1, R^2, \pi_R^{-\frac{1}{2}}\right\}\).
Remark: to obtain (17) from Assumption (A1), let $-\delta < 0$ denote the lim sup in (15c). For any $\beta > 0$ such that $2\beta < \delta$ there exists $R > 0$ such that

$$r(z) \geq R \implies \langle \nabla r, \nabla \ell \rangle_z \leq -2\beta r(z)$$

(18a)

Since $M$ is a Hadamard manifold, there exists exactly one unit-speed geodesic $\gamma$ connecting $x^*$ to any $z$ with $r(z) \geq R$. In addition, this geodesic $\gamma$ satisfies the first-order ordinary differential equation $\dot{\gamma} = \nabla r(\gamma)$ (see [11]). If $z^* = \gamma(R)$, then $r(z^*) = R$, so that $r(z^*) = R$, then (18a), along with this differential equation, imply

$$\ell(z) - \ell(z^*) = \int_R^{r(z^*)} \langle \dot{\gamma}, \nabla \ell \rangle_{\gamma(t)} dt = \int_R^{r(z^*)} \langle \nabla r, \nabla \ell \rangle_{\gamma(t)} dt \leq -2\beta \int_R^{r(z^*)} t dt = \beta (R^2 - r^2(z))$$

But, because $\ell(z) = \log \pi(z)$, this means

$$\pi(z) \leq \left( \pi(z^*) e^{\beta R^2} \right) \exp(-\beta r^2(z)) = \pi_R \exp(-\beta r^2(z))$$

(18b)

where $\pi_R$ is the maximum of the expression in parentheses, taken over all $z^*$ such that $r(z^*) = R$. Now, it is clear that (18b) is the same as (17).

IV. MCMC COMPUTATION OF RIEMANNIAN BARYCENTRES

Consider the problem of computing the Riemannian barycentre $b$ of a probability distribution $\pi$ on a Hadamard manifold $M$. It is assumed that $\pi(dz) = \pi(z) \text{vol}(dz)$, where $\pi(z)$ is a probability density function and $\text{vol}(dz)$ the Riemannian volume measure of $M$. Currently [1][3], existing algorithms are able to address this problem, only if i.i.d. samples may be readily generated from the density $\pi(z)$. In many situations, especially in the context of Bayesian inference, such i.i.d. samples are too difficult to obtain, since the density $\pi(z)$ is quite complicated and only partially known (for example, up to normalisation). In order to deal with such situations, a new Markov chain Monte Carlo algorithm is here proposed, based on the above Propositions [1] and [2].

According to Proposition [1] if the $x_n$ are samples from a geometrically ergodic Markov chain with stationary distribution $\pi$, then the recursive barycentres $s_n$ converge to the required Riemannian barycentre $b$ with the rate (7b). Thus, to compute $b$, it is enough to know how to generate samples from a geometrically ergodic Markov chain with stationary distribution $\pi$. However, if $M$ is a symmetric Hadamard manifold, Proposition [2] says that this can be done using an isotropic Metropolis-Hastings Markov chain, at least when Assumptions (A1) – (A3) are satisfied.

These two propositions combine to provide the following algorithm.

**Input:**
- unnormalised density $\omega(z)$
- proposed transition density $q(dz, y)$
- initial guesses $x_1$ and $s_1$
- number of iterations $N - 1$

for $n = 1, 2, \ldots, N - 1$:

1. generate $x_{n+1} \sim q(x_n, x_{n+1})$
2. compute $\rho_{n+1} = 1 - \alpha(x_n, x_{n+1})$
3. reject $x_{n+1}$ with probability $\rho_{n+1}$
4. compute $s_{n+1}$ using (5)

**Output:**
- recursive barycentre $s_N$

In this algorithm, instructions (1) – (3) amount to the Metropolis-Hastings algorithm [16]. Therefore, these instructions generate samples $x_n$ from an isotropic Metropolis-Hastings Markov chain with stationary distribution $\pi$. The last instruction (4) computes the recursive barycentres $s_n$ of these samples $x_n$. In order to guarantee that the output $s_N$ is sufficiently close to the required Riemannian barycentre $b$, at least in the mean square Riemannian distance, taken over many runs of the algorithm, the number of iterations $N$ should be chosen to make the right-hand side of (7b) sufficiently small.

Here, it may be noted that (7b) provides the rate of convergence of the recursive barycentres $s_n$ to the Riemannian barycentre $b$, only in the mean square Riemannian distance, and not in the stronger sense of almost sure convergence. In other words, (7b) does not guarantee that almost all individual runs of the algorithm will produce recursive barycentres $s_N$ which converge to the same limit $b$ as $N$ is made increasingly large. In fact, the rate of almost sure convergence of the $s_n$ to $b$ can be obtained using a somewhat modified version of the arguments in the proof of Proposition [1]. This is not detailed here, but will be pursued in future work.
Consider now the application of the above algorithm to a problem of Bayesian inference arising from computer vision \[7\]. Recall that, when \( M \) is a symmetric Hadamard manifold, so-called Gaussian distributions can be defined on \( M \) \[7\]. Precisely, for each \( z \in M \) and \( \tau^2 > 0 \), there is a Gaussian distribution \( p(z, \tau^2) \), defined by its probability density function, with respect to the Riemannian volume measure \( vol(dy) \),

\[
p(y|z, \tau^2) = (Z(\tau))^{-1} \exp\left[-\frac{d^2(y, z)}{2\tau^2}\right] \quad \text{for } y \in M
\]

In \[7\], a general formula for computing the normalising constant \( Z(\tau) \), for any symmetric Hadamard manifold \( M \), was provided. In addition, methods for generating i.i.d. samples from \( p(z, \tau^2) \) on various symmetric Hadamard manifolds were described and implemented.

Maximum likelihood estimation of the location parameter \( z \) of a Gaussian distribution \( p(z, \tau^2) \) is straightforward. Indeed \[7\], using (19a), it is almost immediate to see that, if \((y_a : 1 \leq a \leq A)\) are i.i.d. samples from \( p(z, \tau^2) \), then the maximum likelihood estimate \( \hat{z}_{ML} \) of the parameter \( z \), based on these samples, is just the global minimiser of

\[
E_A(z) = \frac{1}{2\tau^2} \sum_{a=1}^{A} d^2(z, y_a)
\]

In other words, \( \hat{z}_{ML} \) is the empirical barycentre of the samples \( y_a \) and can be computed using the algorithms in \[1\][3].

In \[6\], Bayesian inference was considered, instead of maximum likelihood estimation. Basically, the prior distribution used for the parameter \( z \) was another Gaussian distribution, \( z \sim p(\bar{z}, \sigma^2) \). Then, it was noted that the posterior density \( \pi(z) \) of \( z \) had the following form

\[
\pi(z) \propto \exp\left[-\frac{d^2(z, \bar{z})}{2\sigma^2} - E_A(z)\right]
\]

Where \( \propto \) indicates a missing normalising constant. From (19b) and (20a), the maximum a posteriori estimate \( \hat{z}_{MAP} \) of \( z \) was found to be the global minimiser of

\[
-\ell(z) = \frac{1}{2\sigma^2} d^2(z, \bar{z}) + \frac{1}{2\tau^2} \sum_{a=1}^{A} d^2(z, y_a)
\]

In other words, \( \hat{z}_{MAP} \) is the weighted empirical barycentre of the prior parameter \( \bar{z} \) and of the samples \( y_a \). Once more, this can be computed using the algorithms in \[1\][3].

This approach, based on maximum a posteriori estimation, while computationally simple, appears to be sub-optimal. Rather, it is clear from definition (1), of the variance function \( E \) of \( \pi \), that the optimal estimate, in the sense of posterior minimum mean square error, is exactly the Riemannian barycentre \( b \) of \( \pi \).

This estimate \( \hat{z}_{MMSE} = b \) cannot be computed using the algorithms in \[1\][3]. Indeed, at present, there exist no known methods which would generate i.i.d. samples from a partially unknown density whose analytical form is given by (20a).

On the other hand, the algorithm proposed in the present section applies directly to computing \( \hat{z}_{MMSE} = b \). Precisely, it is enough to define the inputs of this algorithm by choosing \( \omega(z) \) to be the right-hand side of (20a), and by taking the proposed transition density \( q(d(z, y)) \) to be a Gaussian density of the form (19a), say \( q = p(\bar{z}, \tau^2) \).

Indeed, generating independent samples from this proposed transition density only requires the methods already described in \[7\].

Now, to make sure that the conclusions of Propositions \[1\] and \[2\] are valid in the context of the present application, it must be checked that Assumptions (A1)–(A3) are satisfied. Consider first the case of Assumption (A1). To evaluate the left-hand side of (15c), let \( z^* = \bar{z} \) so \( r(z) = d(z, \bar{z}) \), and the first term on the right-hand side of (20b) is \( r^2(z)/2\sigma^2 \). Thus, computing \( \nabla \ell \) from (20b) gives

\[
\nabla \ell(z) = -\frac{1}{\sigma^2} r(z) \nabla r(z) - \nabla E_A(z)
\]

and therefore, by taking the scalar product with \( \nabla r \),

\[
\langle \nabla r, \nabla \ell \rangle_z = -\frac{1}{\sigma^2} r(z) - \langle \nabla r, \nabla E_A \rangle_z
\]

since \( \nabla r(z) \) is a unit-length vector for any \( z \in M \). Then, (15c) follows by showing that \( \langle \nabla r, \nabla E_A \rangle_z \) is positive for all \( z \) with sufficiently large \( r(z) \). However, this is the case as soon as \( r(z) > \tau \) for all \( 1 \leq a \leq A \). Indeed, it is known from Riemannian geometry that \[1\][17]

\[
\nabla E_A(z) = -\frac{1}{\tau^2} \sum_{a=1}^{A} \text{Exp}_z^{-1}(y_a)
\]
where \( \exp \) is the Riemannian exponential mapping. But [11], since \( r(z) \) is a convex function, if \( r(z) > r(y_a) \) then
\[
(\nabla r(z), \exp^{-1}(y_a)) \leq r(y_a) - r(z) < 0
\]  
(21c)

Thus, (21b) and (21c) imply that, if \( r(z) > r(y_a) \) for all \( 1 \leq a \leq A \) then \( \langle \nabla r, \nabla \mathcal{E}_A \rangle \) is positive, as required.

Using a similar reasoning, it can be checked that Assumption (A2) is satisfied. On the other hand, Assumption (A3) follows easily from [19a], given the choice of proposed transition density \( q = p(z, \tau^2) \). Therefore, all the assumptions required for Propositions 1 and 2 are satisfied.

The algorithm proposed in the present section was applied to computing \( \hat{z}_{\text{MMSE}} = b \) and comparing it to \( \hat{z}_{\text{MAP}} \). This was done in the case where \( A = 1 \), so there is only one sample \( y_a = y \). In this case, from (19b) and (20a),
\[
\pi(z) \propto \exp \left[ -\frac{d^2(z, \bar{z})}{2\sigma^2} - \frac{d^2(z, y)}{2\tau^2} \right]
\]  
(22a)

Moreover, from (20b), by the discussion after (4d),
\[
\hat{z}_{\text{MAP}} = \bar{z} \#_\rho y ; \quad \rho = \frac{\sigma^2}{\sigma^2 + \tau^2}
\]  
(22b)

so \( \hat{z}_{\text{MAP}} \) is a geodesic mean of \( \bar{z} \) and \( y \). There is at least one setting where the two estimates \( \hat{z}_{\text{MMSE}} \) and \( \hat{z}_{\text{MAP}} \) should agree. This is the following:

if \( \sigma^2 = \tau^2 \) (that is, if \( \rho = 1/2 \)) then \( \hat{z}_{\text{MMSE}} = \hat{z}_{\text{MAP}} \).

To see that this is true, let \( s \) denote the geodesic symmetry about \( \hat{z}_{\text{MAP}} \). This is the isometry of \( M \) which fixes the point \( \hat{z}_{\text{MAP}} \) and reverses any geodesic going through this point [10]. In particular, \( s \cdot \bar{z} = y \) and \( s \cdot y = \bar{z} \), as follows from (22b), since \( \rho = 1/2 \). But, in view of (22a), this immediately implies that \( \pi \circ s = \pi \) and therefore \( \hat{z}_{\text{MMSE}} = s \cdot \hat{z}_{\text{MMSE}} \) as follows by the isometry property (4c). However, this amounts to saying that \( \hat{z}_{\text{MMSE}} = \hat{z}_{\text{MAP}} \).

Figure 1a below shows that the proposed algorithm does indeed compute \( \hat{z}_{\text{MMSE}} = \hat{z}_{\text{MAP}} \) in this setting. This figure was obtained by taking \( M \) the hyperbolic plane (in the Poincaré disc model [18]), and replacing \( \sigma^2 = \tau^2 = 0.1 \).

When \( \rho \neq 1/2 \), there is no theoretical argument providing the value of \( \hat{z}_{\text{MMSE}} \). However, this value can still be computed using the proposed algorithm. For \( \rho \) ranging between 0 and 1, the Riemannian distance between \( \hat{z}_{\text{MMSE}} \) and \( \hat{z}_{\text{MAP}} \) did not exceed 0.01, even when \( \sigma^2 \) or \( \tau^2 \) were relatively large, so as to slow down the convergence of the algorithm. Figure 1b corresponds to \( \sigma^2 = 0.1 \) and \( \tau^2 = 1 \).

In Figures 1a and 1b the gray points \( \times \) mark the last 1000 out of \( N = 100000 \) samples \( x_n \) generated using instructions (1) – (3) of the algorithm. The dashed line is the geodesic curve connecting \( \bar{z} = 0 \) to \( y = \Box \). The points \( \hat{z}_{\text{MMSE}} \) and \( \hat{z}_{\text{MAP}} \) appear as identical and are marked by a \( \bullet \). This lies on the geodesic connecting \( \bar{z} \) to \( y \), since \( \hat{z}_{\text{MAP}} \) lies on this geodesic, by its definition (22b).

As a tentative conclusion, it seems that \( \hat{z}_{\text{MMSE}} \approx \hat{z}_{\text{MAP}} \) and that \( \hat{z}_{\text{MAP}} \) (which can be directly computed from (22b)) is a very good substitute for the optimal estimate \( \hat{z}_{\text{MMSE}} \). However, at present, this is only a conjecture, and will require further systematic study.

\[
\text{Fig. 1: Bayesian inference in the Poincaré disc : } \bar{z} = 0 , y = \Box , \hat{z}_{\text{MMSE}} \approx \hat{z}_{\text{MAP}} = \bullet
\]
REFERENCES

[1] K. T. Sturm, “Probability measures on metric spaces of nonpositive curvature,” J. Contemp. Math., vol. 338, 2003.
[2] B. Afsari, “Riemannian $L^p$ center of mass: existence, uniqueness and convexity,” Proc. Amer. Math. Soc., vol. 139, no. 2, pp. 655–673, 2011.
[3] M. Arnaudon, C. Dombry, A. Phan, and L. Yang, “Stochastic algorithms for computing means of probability measures,” Stoch. Proc. Appl., vol. 58, no. 9, pp. 1473–1455, 2012.
[4] M. Fréchet, “Les éléments aléatoires de nature quelconque dans un espace distancié,” Ann. Inst. H. Poincaré, vol. 10, no. 4, pp. 215–310, 1948.
[5] X. Pennec, “Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements,” Journal of Mathematical Imaging and Vision, vol. 25, no. 1, pp. 127–167, 2006.
[6] S. Said, H. Hajri, L. Bombrun, and B. C. Vemuri, “A novel dynamic system in the space of SPD matrices with applications to appearance tracking,” SIAM J. Imaging Sci., vol. 6, no. 1, pp. 592–615, 2013.
[7] M. Arnaudon, F. Barbaresco, and L. Yang, “Riemannian medians and means with applications to Radar signal processing,” IEEE J. Sel. Topics Signal Process, vol. 7, no. 4, pp. 595–604, 2013.
[8] B. Jeuris and R. Vandebril, “The Kähler mean of Block-Toeplitz matrices with Toeplitz structured blocks,” SIAM J. Matrix Anal. Appl., 2016. [Online]. Available: https://link.springer.com/article/10.1007/s00026-016-0393-z
[9] S. Helgason, Differential geometry, Lie groups, and symmetric spaces. Providence RI: American Mathematical Society, 2001.
[10] P. Petersen, Riemannian geometry (second edition). New York: Springer Science+Business Media, LLC, 2006.
[11] G. O. Roberts and R. L. Tweedie, “Geometric ergodicity and central limit theorems for multidimensional Metropolis and Hastings algorithms,” Biometrika, vol. 83, no. 1, pp. 95–110, 1996.
[12] S. F. Jarner and E. Hansen, “Geometric ergodicity of Metropolis algorithms,” Stochastic Process. Appl., vol. 58, no. 2, pp. 341–361, 1998.
[13] G. O. Roberts and J. S. Rosenthal, “General state-space Markov chains and MCMC algorithms,” Probab. Surv., vol. 1, pp. 20–71, 2004.
[14] I. Chavel, Riemannian geometry, a modern introduction. Cambridge, UK: Cambridge University Press, 2006.
[15] R. Benedetti and C. Petronio, Lectures on hyperbolic geometry. Berlin and Heidelberg: Springer-Verlag, 2003.
[16] J. M. Lee, Introduction to smooth manifolds (second edition). New York: Springer Science+Business Media, 2012.
[17] I. Chavel, “Riemannian symmetric spaces of rank one,” Adv. Math., vol. 4, pp. 236–264, 1970.
APPENDIX A
PROOF OF PROPOSITION

Paragraph 1) will give the proof of Part (i) of the proposition. As a byproduct of this proof, the estimates \((13b)\) and \((13c)\), which were used in proving Part (ii), will be recovered in Paragraph 2). In the following, it is admitted that the Markov chain \((x_n)\) verifies the geometric ergodicity condition \((8)\) for some (unique) stationary distribution \(\pi\). In fact \((14)\), this follows from the theory of general state space Markov chains, under the assumption that the transition kernel \(P\) is irreducible and aperiodic, by the geometric drift condition \((6)\).

A further consequence of \((6)\) is that \(E_n(x^*) < \infty\) and \(E(x^*) < \infty\) \((14)\) \((13)\). However, if the variance functions \(E_n\) and \(E\) are finite at one point, they are finite throughout \(M\), and the corresponding Riemannian barycentres \(b_n\) and \(b\) are well-defined \((1)\).

1) Proof of Part (i): the proof is divided into four steps. First, it is proved that the variance functions \(E_n\) converge locally uniformly to \(E\), the variance function of the stationary distribution \(\pi\). Precisely, for each \(z \in M\),

\[
|E_n(z) - E(z)| \leq (2d^2(z, x^*) + 2) R(x_1) \rho^n
\]

where \(\rho < 1\) and \(R(x_1) < \infty\) are the same as in \((8)\). Thus, for any compact \(K \subset M\), if \(R_K(x_1)\) denotes the maximum over \(z \in K\) of \((2d^2(z, x^*) + 2) R(x_1)\), then

\[
\sup_{z \in K} |E_n(z) - E(z)| \leq R_K(x_1) \rho^n
\]

To prove \((23a)\), note from \((1)\) and \((9)\) that

\[
|E_n(z) - E(z)| = \left| \int_M d^2(z, x) [\pi_n(dx) - \pi(dx)] \right|
\]

However, by the triangle inequality,

\[
d^2(z, x) \leq (d(z, x^*) + d(x^*, x))^2 \leq 2d^2(z, x^*) + 2d^2(x, x^*) \leq (2d^2(z, x^*) + 2) V(x)
\]

where the last inequality follows from \((6n)\). Therefore, if the function \(f(x)\) is given by

\[
f(x) = (2d^2(z, x^*) + 2)^{-1} d^2(z, x)
\]

then \(|f| \leq V\) and it is possible to apply \((8)\), which yields

\[
\left| \int_M (2d^2(z, x^*) + 2)^{-1} d^2(z, x) [\pi_n(dx) - \pi(dx)] \right| \leq R(x_1) \rho^n
\]

Now, multiplying this inequality by \((2d^2(z, x^*) + 2)\), \((23a)\) is obtained using \((23c)\).

The second step of the proof is to note that the functions \(E_n\) and \(E\) are uniformly coercive. Precisely, for \(n \geq 1\),

\[
E_n(z) > d(z, x^*) \text{ whenever } d(z, x^*) > 1 + 2 (E(x^*) + R(x_1))^{1/2}
\]

and the same condition is verified by \(E\). To prove \((24a)\), apply the triangle inequality to \((9)\), to write

\[
E_n(z) = E_{n+1} d^2(z, x_n) \geq E_{n+1} (d(z, x^*) - d(x^*, x_n))^2
\]

Then, it follows easily that

\[
E_n(z) > E_{n+1} (d(z, x^*) - 2d(z, x^*)d(x^*, x_n)) = d(z, x^*) E_{n+1} (d(z, x^*) - 2d(x^*, x_n))
\]

which, after an application of Jensen’s inequality, becomes

\[
E_n(z) > d(z, x^*) \left( d(z, x^*) - 2 E_n^{1/2}(x^*) \right)
\]

Now, to obtain \((24a)\), note that

\[
E_n^{1/2}(x^*) \leq (E(x^*) + 2R(x_1))^{1/2} \leq 2 (E(x^*) + R(x_1))^{1/2}
\]

where the first inequality follows by putting \(z = x^*\) in \((23a)\), and taking square roots. Thus, the condition

\[
d(z, x^*) > 1 + 2 (E(x^*) + R(x_1))^{1/2}
\]

implies that

\[
d(z, x^*) > 1 + 2 E_n^{1/2}(x^*)
\]

which means the expression inside the parentheses on the right-hand side of \((24b)\) is \(> 1\). The fact that \(E\) also verifies \((24a)\) can be seen by repeating the same calculations, with \(E\) instead of \(E_n\).
The third step of the proof is to note that the Riemannian barycentres $b_n$ and $b$ all belong to one and the same compact $K, \subset M$. Indeed, $b_n$ and $b$ are the unique global minima of the variance functions $\mathcal{E}_n$ and $\mathcal{E}$, respectively. In particular $\mathcal{E}_n(b_n) \leq \mathcal{E}_n(x^*)$ and $\mathcal{E}(b) \leq \mathcal{E}(x^*)$. Therefore,

$$
\mathcal{E}_n(b_n) \leq \mathcal{E}_n(x^*) \leq \mathcal{E}(x^*) + 2R(x_i) \leq (\delta - 1)^2
$$

(25a)

where the second inequality follows by putting $z = x^*$ in (23a), and where $\delta = 1 + 2(\mathcal{E}(x^*) + R(x_i))^{\frac{1}{2}}$. Now, it is possible to see that

$$
b_n \in B(x^*, L) \quad \text{where} \quad L = \max \{\delta, (\delta - 1)^2\}
$$

(25b)

where $B(x^*, L)$ denotes the closed ball of centre $x^*$ and radius $L$. Assume this is not true, so $d(b_n, x^*) > L \geq \delta$. Then, $\mathcal{E}_n(b_n) > L \geq (\delta - 1)^2$, in contradiction with (25a). By an almost identical argument, it can be seen that if $b \in B(x^*, L)$. Putting $K = B(x^*, L)$, the Hopf-Rinow theorem implies that $K$ is compact $[11]$. Finally, it is clear that $K$ does not depend on $n$.

The fourth, and final, step of the proof is to obtain (7a), from the following statement

$$
d(b_n, b) \leq \eta_n \quad \text{whenever} \quad \eta_n \leq 2L
$$

(26a)

where $\eta_n$ is given by

$$
\eta_n = (3R_\ast(x_i) \rho^n) \frac{1}{4} \quad \text{for} \quad R_\ast(x_i) = (18L^2 + 2) R(x_i)
$$

To prove (26a), note from (25b), using the triangle inequality, that all $b_n$ lie within $K = B(b, 2L)$, which is a compact set, by the Hopf-Rinow theorem. Using the triangle inequality, again, it is possible to show that the maximum over $z \in K$ of $(2d^2(z, x^*) + 2) R(x_i)$ is $\leq R_\ast(x_i)$. Thus, (23b) can be written

$$
\sup_{z \in K} |\mathcal{E}_n(z) - \mathcal{E}(z)| \leq R_\ast(x_i) \rho^n
$$

(26b)

However, if $\eta_n \leq 2L$ then $\hat{B}(b, \eta_n) \subset K$, so that (26b) implies the following inequalities

$$
\inf_{z \in \hat{B}(b, \eta_n)} \mathcal{E}_n(z) \leq \inf_{z \in \hat{B}(b, \eta_n)} \mathcal{E}(z) + R_\ast(x_i) \rho^n
$$

(26c)

and

$$
\inf_{z \in K - \hat{B}(b, \eta_n)} \mathcal{E}_n(z) \geq \inf_{z \in K - \hat{B}(b, \eta_n)} \mathcal{E}(z) - R_\ast(x_i) \rho^n
$$

(26d)

Since the global minimum of $\mathcal{E}$ is at $b \in \hat{B}(b, \eta_n)$, it is possible to write in (26c),

$$
\inf_{z \in \hat{B}(b, \eta_n)} \mathcal{E}_n(z) \leq \mathcal{E}(b) + R_\ast(x_i) \rho^n
$$

(26e)

On the other hand, for any $z \in K - \hat{B}(b, \eta_n)$,

$$
\mathcal{E}(z) \geq \mathcal{E}(b) + d^2(z, b) \geq \mathcal{E}(b) + \eta_n^2
$$

where the first inequality follows from the triangle inequality (4b). Replacing this into (26d) gives

$$
\inf_{z \in K - \hat{B}(b, \eta_n)} \mathcal{E}_n(z) \geq \mathcal{E}(b) + \eta_n^2 - R_\ast(x_i) \rho^n = \mathcal{E}(b) + 2R(x_i) \rho^n
$$

(26f)

after using the definition of $\eta_n$. Then, it is clear from (26e) and (26f) that

$$
\inf_{z \in K - \hat{B}(b, \eta_n)} \mathcal{E}_n(z) \geq \inf_{z \in \hat{B}(b, \eta_n)} \mathcal{E}_n(z)
$$

so the global minimum $b_n$ of $\mathcal{E}_n$ must belong to $\hat{B}(b, \eta_n)$. In other words, $d(b_n, b) \leq \eta_n$. This proves (26a), which, if $\sigma = \rho^2$, is equivalent to $d(b_n, b) = O(\sigma^n)$, as in (7a).

2) Proof of (13b) and (13c): the proof of (13b) relies on (23a). Indeed, putting $z = b$ in (23a) yields, after using the fact that $\rho < 1$,

$$
\mathcal{E}_n(b) \leq \mathcal{E}(b) + (2d^2(b, x^*) + 2) R(x_i)
$$

Since the right-hand side does not depend on $n$, it becomes clear that

$$
\sup_n \mathcal{E}_n(b) \leq \mathcal{E}(b) + (2d^2(b, x^*) + 2) R(x_i) < \infty
$$

which is the same as (13b).

Consider now the proof of (13c). First, it must be proved that

$$
\mathbb{E}d^2(s_{n-1}, b_n) - \mathcal{V}_{n-1} = O(\rho^{n/2})
$$

(27a)

where $\mathcal{V}_{n-1} = \mathbb{E}d^2(s_{n-1}, b)$, Indeed, by the definition of $\mathcal{U}_{n-1}$ in (12b), this is the first estimate in (13c). Note that the left-hand side of (27a) is

$$
\mathbb{E}[d^2(s_{n-1}, b_n) - d^2(s_{n-1}, b)] = \mathbb{E}[(d(s_{n-1}, b_n) + d(s_{n-1}, b)) (d(s_{n-1}, b_n) - d(s_{n-1}, b))]
$$

But, by the triangle inequality, $d(s_{n-1}, b_n) - d(s_{n-1}, b) \leq d(b_n, b)$, so the left-hand side of (27a) verifies
\[ \mathbb{E} d^2(s_{n-1}, b_n) - \mathcal{V}_{n-1} \leq \mathbb{E} [d(s_{n-1}, b_n) + d(s_{n-1}, b)] d(b_n, b) \]
\[ = \mathbb{E} [d(s_{n-1}, b_n) + d(s_{n-1}, b)] O(\rho^{n/2}) \]  
(27b)

where the equality on the second line follows by (26a) as in (27b). Now, (27c) will follow by showing

\[ \sup_n \mathbb{E} [d(s_{n-1}, b_n) + d(s_{n-1}, b)] < \infty \]  
(27c)

However, by the triangle inequality, applied under the expectation,

\[ \mathbb{E} [d(s_{n-1}, b_n) + d(s_{n-1}, b)] \leq d(b_n, b) + 2\mathbb{E} d(s_{n-1}, b) \]
\[ = O(\rho^{n/2}) + 2\mathbb{E} d(s_{n-1}, b) \]

where the last equality follows by using (26a) as in (27b). Now, (27c) will follow by showing

\[ \mathbb{E} d(s_n, b) \leq \max_{k=1,...,n} \mathcal{E}_k^*(b) \]  
(27d)

since \( \sup \mathcal{E}_n(b) < \infty \) by (13b). To obtain (27d), note that,

\[ (\mathbb{E} d(s_n, b))^2 \leq \mathbb{E} d^2(s_{n-1} \# x_n, b) \]  
(27e)

which follows by using Jensen’s inequality, and replacing from the definition of the \( s_n \) in (5). Then, an application of the strong convexity property (3) to (27e) yields

\[ (\mathbb{E} d(s_n, b))^2 \leq \frac{n-1}{n} \mathbb{E} d^2(s_{n-1}, b) + \frac{1}{n} \mathbb{E} d^2(x_n, b) \]
\[ = \frac{n-1}{n} \mathbb{E} d^2(s_{n-1}, b) + \frac{1}{n} \mathcal{E}_n(b) \]

where the last equality uses the definition of \( \mathcal{E}_n \) in (5). From this last inequality, (27d) can be proved by induction, since \( s_i = x_i \) so that \( \mathbb{E} d^2(s_i, b) = \mathcal{E}_i(b) \).

To finish the proof of (13c), it must be proved that

\[ \mathcal{E}_n(b) - \mathcal{E}_n(b_n) = O(\rho^{n/2}) \]  
(28a)

Indeed, by definition, \( \text{Var}_n = \mathcal{E}_n(b_n) \). Therefore, by the definition of \( \mathcal{U}_n^{(3)} \) in (12b), it is clear that (28a) is the same as the second estimate in (13c). Now, the left-hand side of (28a) is positive (since \( b_n \) is the global minimum of \( \mathcal{E}_n \)) and verifies the inequality

\[ \mathcal{E}_n(b) - \mathcal{E}_n(b_n) \leq \lvert \mathcal{E}_n(b) - \mathcal{E}(b) \rvert + \lvert \mathcal{E}(b) - \mathcal{E}(b_n) \rvert + \lvert \mathcal{E}(b_n) - \mathcal{E}_n(b_n) \rvert \]  
(28b)

Recall that all \( b_n \) and \( b \) belong to the compact set \( K_* = B(x^*, L) \) given by (25b). It follows by (23b),

\[ \lvert \mathcal{E}_n(b) - \mathcal{E}(b) \rvert + \lvert \mathcal{E}(b_n) - \mathcal{E}_n(b_n) \rvert \leq 2R_{K_*}(x^*) \rho^n = O(\rho^n) \]  
(28c)

On the other hand, by (1),

\[ \mathcal{E}(b) - \mathcal{E}(b_n) = \int_M (d^2(b, x) - d^2(b_n, x)) \pi(dx) \]
\[ = \int_M (d(b, x) - d(b_n, x))(d(b, x) + d(b_n, x)) \pi(dx) \]

But, from the triangle inequality, \( |d(b, x) - d(b_n, x)| \leq d(b_n, b) \), and this is \( O(\rho^{n/2}) \) by (26a). Therefore, by taking the absolute value,

\[ \lvert \mathcal{E}(b) - \mathcal{E}(b_n) \rvert \leq \int_M (d(b, x) + d(b_n, x)) \pi(dx) O(\rho^{n/2}) \]  
(28d)

However, the integral inside square brackets verifies

\[ \int_M (d(b, x) + d(b_n, x)) \pi(dx) \leq \mathcal{E}^+(b) + \mathcal{E}^+(b_n) \leq 2 \sup_{x \in K_*} \mathcal{E}^+(x) < \infty \]

where the first inequality follows by Jensen’s inequality. Replacing this into (28d) yields

\[ \lvert \mathcal{E}(b) - \mathcal{E}(b_n) \rvert = O(\rho^{n/2}) \]  
(28e)

Finally, it is clear that (28a) follows from (28b), after adding together (28c) and (28e). Accordingly, the proof of (13c) is now complete.
The first, and major, step in the proof is to show that Assumptions (A1) – (A3) imply the two conditions in (16). This follows from Propositions 3 and 4 below. In these two propositions, as in Proposition 2, (19) is equal to

\[
\pi(y) \leq \pi(z) \quad \text{for all } y \in M.
\]

Then, \(y \in A(z)\), so it is clear that (18a) holds. However, by definition of \(r(y)\), it is clear that \(r(y) = c(\xi)\), so \(c(\xi) > R\). Thus, for any \(c > c(\xi)\), it follows from (18a) that

\[
\log(\pi \circ \gamma_\xi) - \log(\pi \circ \gamma_\xi)(c(\xi)) = \int_{\gamma_\xi} \langle \dot{\gamma}_\xi, \nabla \ell \rangle_{\gamma_\xi} dt \leq -2\beta \int_{\gamma_\xi} \ell dt < 0.
\]

In other words, for each \(\xi \in S_xM\) and \(y = \gamma_\xi(c)\), it follows that \(\pi(y) > \pi(z)\) if \(c > c(\xi)\). On the other hand, the definition of \(c(\xi)\), it follows that \(\pi(y) > \pi(z)\) if \(c > c(\xi)\).

Once Propositions 3 and 4 have been proved, the proof of Proposition 2 can be completed by using the following lemma.

**Lemma 1:** Let \((x_n)\) be an isotropic Metropolis-Hastings Markov chain, with values in a Hadamard manifold \(M\), and which satisfies Assumption (A3). Moreover, assume the stationary distribution \(\pi\) has positive and continuous probability density function \(\pi(z)\). If the two conditions in (16) are verified, then \((x_n)\) verifies the geometric drift condition (6).

Lemma 1 is here given without proof because it can be proved by repeating, almost word for word, the proofs for random-walk Metropolis chains in Euclidean space [12][13]. The main point is that Assumption (A3) implies that every non-empty bounded subset of \(M\) is a small set for the transition kernel \(P\) given in (15). With this in mind, the geometric drift condition (6) follows almost directly from the two conditions in (16).

Indeed, (16a) implies there exists \(\lambda < 1\) and \(R > 0\) such that

\[
r(z) \geq R \implies PV(z) \leq \lambda V(z)
\]

That is, (6b) is verified on \(M - C\), where \(C\) is the open ball \(B(x^*, R)\). In addition, by (16b),

\[
\lim_{r(z) \to \infty} Q(z, A(z)) > 0
\]
\[ b = \left[ \sup_{z \in B(z^*, R)} V(z) \right] \left[ \sup_{z \in M} \frac{PV(z)}{V(z)} \right] < \infty \quad (32b) \]

Therefore, (32b) is also verified on \( C \), since for \( z \in C \)
\[ PV(z) \leq b \leq \lambda V(z) + b \]
Thus (32b) is verified throughout \( M \). It remains to note that \( C \) is a small set, since it is bounded.

**Proof of Proposition 3:** by Assumption (A2), there exist \( \delta > 0 \) and \( R > 0 \) such that
\[ r(y) \geq R \implies \langle \nabla r, n \rangle_y < -\delta \quad (33a) \]
Since \( M \) is a symmetric Hadamard manifold, its sectional curvature is negative and bounded below \([10]\). Precisely, let \(-\kappa^2\) be a lower bound on the sectional curvature of \( M \), and let \( \Lambda \) be a positive number with
\[ (\dim M)^\frac{1}{2} \Lambda \leq \frac{\delta}{2\kappa} \tanh(\kappa R) \quad (33b) \]
where \( \dim M \) is the dimension of \( M \). Now, for any \( z \in M \) with \( r(z) \geq R + \Lambda \), consider the set
\[ \Omega(z) = \left\{ \Exp (-a \zeta) : a \in (0, \Lambda), \zeta \in S, M, \|\nabla r(z) - \zeta\| \leq \frac{\delta}{2} \right\} \]
Let \( y = \Exp (-a \zeta) \) be a point in \( \Omega(z) \), and denote by \( \eta(t) \) the unit-speed geodesic with \( \eta(0) = z \) and \( \eta(a) = y \). It is first proved that
\[ \langle \dot{\eta}(t), n \rangle_{\eta(t)} > 0 \quad \text{for } t \in (0, a) \quad (33c) \]
Indeed, the left-hand side of (33c) may be written
\[ \langle \dot{\eta}(t), n \rangle_{\eta(t)} = -\langle \nabla r, n \rangle_{\eta(t)} + \langle \dot{\eta}(t) + \nabla r, n \rangle_{\eta(t)} \]
Then, if \( \Pi_t \) denotes the parallel transport along \( \eta \) from \( \eta(0) = z \) to \( \eta(t) \),
\[ \langle \dot{\eta}(t), n \rangle_{\eta(t)} = -\langle \nabla r, n \rangle_{\eta(t)} + \langle \Pi_t(\nabla r(z) - \zeta), n \rangle_{\eta(t)} + \langle \nabla r - \Pi_t(\nabla r(z)), n \rangle_{\eta(t)} \quad (34a) \]
as may be checked easily, by adding the three terms on the right-hand side, and then noting that \( \dot{\eta}(t) = \Pi_t(-\zeta) \), since \( \dot{\eta}(0) = \zeta \) and \( \dot{\eta}(t) \) is self-parallel. But, by the triangle inequality,
\[ r(\eta(t)) \geq r(z) - d(z, \eta(t)) > (R + \Lambda) - \Lambda = R \]
since \( d(z, x^*) = r(z) \geq R + \Lambda \) and \( d(z, \eta(t)) \leq a < \Lambda \). Thus, it follows from (33a),
\[ -\langle \nabla r, n \rangle_{\eta(t)} > \delta \quad (34b) \]
Moreover, since the parallel transport \( \Pi_t \) preserves Riemannian norms, and \( \|\nabla r(z) - \zeta\| \leq \delta/2 \) from the definition of \( \Omega(z) \), it follows by the Cauchy-Schwarz inequality
\[ \langle \Pi_t(\nabla r(z) - \zeta), n \rangle_{\eta(t)} \geq -\|\Pi_t(\nabla r(z) - \zeta)\| = -\|\nabla r(z) - \zeta\| \geq -\frac{\delta}{2} \quad (34c) \]
On the other hand, let \( (e_i(t)) : i = 1, \ldots, d \) be a parallel orthonormal basis along the unit-speed geodesic \( \eta(t) \) (here, \( d = \dim M \)). Then,
\[ \langle \nabla r - \Pi_t(\nabla r(z)), e_i \rangle_{\eta(t)} = \int_0^t \langle \nabla^2 r \cdot \dot{\eta}, e_i \rangle_{\eta(s)} \, ds \]
where \( \nabla^2 r \) denotes the Riemannian Hessian of the function \( r \). But, according to the Hessian comparison theorem \([11]\) (see Page 175),
\[ \int_0^t \langle \nabla^2 r \cdot \dot{\eta}, e_i \rangle_{\eta(s)} \, ds \leq \int_0^t \kappa \coth(\kappa r(\eta(s))) \, ds \leq \Lambda \kappa \coth(\kappa R) \]
Thus, using (33b), it then follows by the Cauchy-Schwarz inequality
\[ \langle \nabla r - \Pi_t(\nabla r(z)), n \rangle_{\eta(t)} \geq -\frac{\delta}{2} \quad (34d) \]
Finally, by adding (34b) to (34c) and (34d), it follows from (34a)
\[ \langle \dot{\eta}(t), n \rangle_{\eta(t)} > \delta - \frac{\delta}{2} - \frac{\delta}{2} = 0 \]
which is the same as (33c). Moving on, from (33c), it is possible to prove that
\[ \Omega(z) \subset A(z) \quad (35a) \]
for all \( z \) such that \( r(z) \geq R + \Lambda \), where \( A(z) \) is the acceptance region of \( z \), defined after (29).
To prove (35a), consider \( y \in \Omega(z) \) and \( \eta(t) \) as before, with \( \eta(0) = z \) and \( \eta(a) = y \). Now, assume that \( y \in C_z \) (the contour manifold of \( z \), given by (31)). Then, \( \pi(\eta(0)) = \pi(\eta(a)) \), so that, by the mean value theorem, there exists \( t \in (0, a) \) such that

\[
\frac{d}{dt} \pi(\eta(t)) = \langle \dot{\eta}(t), \nabla \pi \rangle_{\eta(t)} = 0
\]

But, from the definition of \( n(z) \), this implies

\[
\langle \dot{\eta}(t), n(\eta(t)) \rangle = \| \nabla \pi(z) \|^{-2} \langle \dot{\eta}(t), \nabla \pi \rangle_{\eta(t)} = 0
\]

in contradiction with (33c). Thus, the assumption that \( y \in C_z \) cannot hold. Since \( y \in \Omega(z) \) is arbitrary, this means that the intersection of \( \Omega(z) \) and \( C_z \) is empty, \( \Omega(z) \cap C_z = \emptyset \) (35b).

However, note that \( y_* = \text{Exp}_z (-a \nabla r(z)) \) belongs to \( \Omega(z) \), as can be seen from the definition of \( \Omega(z) \). Also, since \( r(y_*) = r(z) - a \), it follows that \( y_* \) belongs to \( A(z) \), because \( A(z) \) is equal to the region inside of \( C_z \). Therefore, the intersection of \( \Omega(z) \) and \( A(z) \) is non-empty. Finally, it is enough to note that the set \( \Omega(z) \) is connected, since it is the image under \( \text{Exp}_z \) of a connected set. This implies that, if the intersection of \( \Omega(z) \) and \( R(z) \), the complement of \( A(z) \), were non-empty, then \( \Omega(z) \) would also intersect \( C_z \). Clearly, this would be in contradiction with (35b).

Using (35a), it is now possible to prove (30). Indeed, for \( z \) such that \( r(z) \geq R + \Lambda \), it follows from (35a) that

\[
Q(z, A(z)) \geq Q(z, \Omega(z)) = \int_{\Omega(z)} q(z, y) \vol(dy)
\]

where the last equality follows from (29). However, by Assumption (A3),

\[
\int_{\Omega(z)} q(z, y) \vol(dy) \geq \int_{\Omega(z)} q(z, y) 1_{B(z, \delta_q)}(y) \vol(dy) > \epsilon_q \times \vol(\Omega(z) \cap B(z, \delta_q))
\]

Now, to prove (30), it only remains to show that

\[
\vol(\Omega(z) \cap B(z, \delta_q)) \geq c > 0
\]

where the constant \( c \) does not depend on \( z \). Indeed, it is then clear from (36a) and (36b) that

\[
\liminf_{r(z) \to \infty} Q(z, A(z)) > \epsilon_q \times c > 0
\]

To obtain (36c), consider the mapping \( \varphi(r, \zeta) = \text{Exp}_z (-r \zeta) \) for \( r > 0 \) and \( \zeta \in S_z M \). Because \( M \) is a Hadamard manifold, \( \varphi \) is a diffeomorphism onto \( M \). Thus, it is possible to write

\[
\vol(\Omega(z) \cap B(z, \delta_q)) = \int_0^\tau \int_{S_z M} 1 \{ ||\nabla r(z) - \zeta || \leq \delta/2 \} \varphi^* \vol(dr d\zeta)
\]

where \( \tau = \min\{ \Lambda, \delta_q \} \) and \( \varphi^* \vol \) denotes the pullback of the Riemannian volume measure \( \vol \) under \( \varphi \). Here, let

\[
\varphi^* \vol(dr d\zeta) = \lambda(r, \zeta) \omega(d\zeta) dr
\]

where \( \omega(d\zeta) \) denotes the area measure on the unit sphere \( S_z M \). By the Riemannian volume comparison theorem in [17] (Page 128), \( \lambda(r, \zeta) \geq r^{d-1} \). Therefore, (37a) and (37b) imply

\[
\vol(\Omega(z) \cap B(z, \delta_q)) \geq \frac{r^d}{d} \times \omega \{ ||\nabla r(z) - \zeta || \leq \delta/2 \}
\]

However, since the area measure \( \omega \) is invariant by rotation, the area

\[
\omega \{ ||\nabla r(z) - \zeta || \leq \delta/2 \} = a
\]

does not depend on \( z \). Precisely, \( a \) is equal to the area of a spherical cap, with angle \( 2 \acos(1 - \delta^2/8) \). Finally, (36c) can be immediately obtained, by letting \( c = (r^d/d) \times a \). This completes the proof of (30).

**Proof of Proposition 4:** let \( V(z) = c \pi^{-\frac{1}{2}}(z) \) as in the proposition. In order to check the two conditions in (16), recall that the transition kernel \( P \) is given by (15). In (15a), one should have

\[
\rho(z) = \int_M (1 - \alpha(z, y)) q(z, y) \vol(dy)
\]
since the right-hand side of (15a) should integrate to 1 when \( f(z) \) is the constant function equal to 1 for all \( z \in M \). But, since \( \alpha(z, y) = \min\{1, \pi(y)/\pi(z)\} \), it follows that \( 1 - \alpha(z, y) = 0 \) when \( y \in A(z) \), the acceptance region of \( z \), defined after (29). Thus,

\[
\rho(z) = \int_{R(z)} \left[ 1 - \frac{\pi(y)}{\pi(z)} \right] q(z, y) \, \text{vol}(dy)
\]

where \( R(z) \) is the rejection region of \( z \). With this expression of \( \rho(z) \), putting \( f(z) = V(z) \) in (15a), it follows by a direct calculation

\[
\frac{PV(z)}{V(z)} = \int_{A(z)} q(z, y) \left( \frac{\pi(y)}{\pi(z)} \right) \frac{1}{2} \, \text{vol}(dy) + \int_{R(z)} q(z, y) \left[ 1 - \frac{\pi(y)}{\pi(z)} + \frac{\pi(y)}{\pi(z)} \right] \frac{1}{2} \, \text{vol}(dy)
\]

(38)

Here, all the ratios are less than or equal to 1, so that (15b) immediately implies (16b).

In order to prove (16a), it is enough to prove that

\[
\lim_{r(z) \to \infty} \int_{A(z)} q(z, y) \left( \frac{\pi(y)}{\pi(z)} \right) \frac{1}{2} \, \text{vol}(dy) = 0
\]

(39a)

and

\[
\lim_{r(z) \to \infty} \int_{R(z)} q(z, y) \left[ 1 - \frac{\pi(y)}{\pi(z)} + \frac{\pi(y)}{\pi(z)} \right] \frac{1}{2} \, \text{vol}(dy) = 0
\]

(39b)

Indeed, if these two limits are replaced in (38), it will follow that

\[
\lim_{r(z) \to \infty} \sup PV(z) = \lim_{r(z) \to \infty} \sup Q(z, R(z)) = \lim_{r(z) \to \infty} 1 - Q(z, A(z)) < 1
\]

where the inequality is obtained using (30). However, this is the same as (16a). Thus, to complete the proof, it is enough to prove (39a) and (39b). The proofs of (39a) and (39b) being very similar, only the proof of (39a) is presented.

**Proof of (39a)**: this is divided into three steps. First, it is proved that

\[
\lim_{L \to \infty} \int_{A(z) \setminus B(z, L)} q(z, y) \left( \alpha(y, z) \right)^{1/2} \, \text{vol}(dy) = 0 \quad \text{uniformly in } z
\]

(40a)

where \( \alpha(y, z) = \pi(z)/\pi(y) \). To prove (40a), note that \( \alpha(y, z) \leq 1 \) for \( y \in A(z) \), and that \( A(z) \setminus B(z, L) \subset M \setminus B(z, L) \). It follows that, for any \( z \in M \),

\[
\int_{A(z) \setminus B(z, L)} q(z, y) \left( \alpha(y, z) \right)^{1/2} \, \text{vol}(dy) \leq \int_{M \setminus B(z, L)} q(z, y) \, \text{vol}(dy)
\]

(40b)

Since \( M \) is a symmetric Hadamard manifold, there exists an isometry \( g : M \to M \) such that \( g \cdot x^* = z \) (here, \( g \cdot x^* = g(x^*) \)). Since \( g \) is an isometry, it leaves invariant the Riemannian volume, so it is possible to perform a change of variables,

\[
\int_{M \setminus B(z, L)} q(z, y) \, \text{vol}(dy) = \int_{M \setminus B(x^*, L)} q(z, g \cdot y) \, \text{vol}(dy)
\]

But, \( q(z, y) = q(d(z, y)) \) depends only on the Riemannian distance \( d(z, y) \). This implies that, \( q(z, g \cdot y) = q(x^*, y) \), since \( g \) is an isometry, and therefore preserves Riemannian distance. Thus,

\[
\int_{M \setminus B(z, L)} q(z, y) \, \text{vol}(dy) = \int_{M \setminus B(x^*, L)} q(x^*, y) \, \text{vol}(dy)
\]

(40b)

Here, the right-hand side does not depend on \( z \), and tends to zero as \( L \to \infty \), as seen by putting \( z = x^* \) in (15b).

Now, (40a) follows directly from (40b).

Second, assume that \( r(z) \) is so large that the level set \( C_z \) verifies (31) and \( A(z) \) is equal to the region inside of \( C_z \). It is then proved that, for any \( L > 0 \),

\[
\lim_{r(z) \to \infty} \int_{A(z) \cap B(z, L) - C_z} q(z, y) \left( \alpha(y, z) \right)^{1/2} \, \text{vol}(dy) = 0
\]

(41a)

where \( C_z(\epsilon) \) is the tubular neighborhood of \( C_z \) given by

\[
C_z(\epsilon) = \{ \text{Exp}_y (s \nabla r(y)) : y \in C_z, |s| < \epsilon \}
\]
Because of (15b), to prove (41a) it is enough to prove that
\[ \lim_{r(z) \to \infty} \alpha(y, z) = 0 \quad \text{uniformly in } y \in A(z) \cap B(z, L) - C_z(\varepsilon) \] (41b)

However, this follows by Assumption (A1). Indeed, this assumption guarantees the existence of some \( \beta > 0 \) and \( R > 0 \) as in (17). Then, take \( r(z) > R + \varepsilon \) and note that, for \( y \) as in (41b), if \( r(y) \leq R \) then it follows from (17)
\[ \alpha(y, z) \leq \frac{\pi_r \exp (-\beta r^2(z))}{\pi(y)} \leq \frac{\pi_r \exp (-\beta r^2(z))}{\min_{r(y) \leq R} \pi(y)} \] (41c)

where the right-hand side converges to zero as \( r(z) \to \infty \), uniformly in \( y \). On the other hand, if \( r(y) > R \), let \( \gamma \) be the unit-speed geodesic connecting \( x^* \) to \( y \). Since \( y \in A(z) \) (so \( y \) lies inside of \( C_z \)) there exists some \( r \geq r(y) \) such that \( \gamma(r) \in C_z \). Moreover, since \( y \notin C_z(\varepsilon) \), it follows that \( r > r(y) + \varepsilon \). Then, using the same steps which lead from (18a) to (18b)
\[ \alpha(y, z) = \frac{\pi(\gamma(r))}{\pi(\gamma(r(y)))} = \exp \left( \int_{r(y)}^r \langle \gamma_r(\gamma(t)), dt \rangle \right) \leq \exp \left( -2\beta \int_{r(y)}^r t \, dt \right) \]

By a direct calculation, this implies
\[ \alpha(y, z) \leq \exp \left( -2\beta \varepsilon r + \beta \varepsilon^2 \right) \leq \exp \left( -2\beta \varepsilon r(w) + \beta \varepsilon^2 \right) \] (41d)

where \( w \in C_z \) is such that \( r(w) \) is the minimum of \( r(w') \) taken over all \( w' \in C_z \). Note that the right-hand side of (41d) does not depend on \( y \). Moreover, \( r(w) \) tends to zero as \( r(z) \to \infty \), since \( \pi(w) = \pi(z) \) and \( \pi(z) \) tends to zero as \( r(z) \to \infty \). Therefore, because \( \pi(w) \) is positive, it follows that \( r(w) \to \infty \) as \( r(z) \to \infty \). But, this implies that the right-hand side of (41d) converges to zero as \( r(z) \to \infty \), uniformly in \( y \). Now, (41b) follows from (41c) and (41d).

The third, and final, step is to show that, for any \( L > 0 \),
\[ \lim_{\varepsilon \to 0} \limsup_{r(z) \to \infty} \int_{A(z) \cap B(z, L) \cap C_z(\varepsilon)} q(z, y) \left( \alpha(y, z) \right) \frac{1}{2} \, vol(dy) = 0 \] (42a)

Since \( \alpha(y, z) \leq 1 \) for \( y \in A(z) \), to prove (42a), it is enough to prove
\[ \lim_{\varepsilon \to 0} \limsup_{r(z) \to \infty} \int_{A(z) \cap B(z, L) \cap C_z(\varepsilon)} q(z, y) \, vol(dy) = 0 \] (42b)

For brevity, the proof is carried out under the assumption that \( q(z, y) = q(d(z, y)) \) is uniformly bounded, in \( z \) and \( y \). This assumption is verified in all practical situations. If it is admitted, then (42b) follows immediately by showing
\[ \lim_{\varepsilon \to 0} \limsup_{r(z) \to \infty} \int_{A(z) \cap B(z, L) \cap C_z(\varepsilon)} q(z, y) \, vol(dy) = 0 \] (42c)

To show (42c), consider the following sets
\[ T(z) = \{ \xi \in S_x, M : Exp_x(r \xi) \in B(z, L) \quad \text{for some } r \geq 0 \} \]
\[ S(z) = \{ Exp_x(r \xi) : \xi \in T(z) \text{ and } |r - r(z)| \leq L \} \]

Using the triangle inequality, it is possible to show that
\[ B(z, L) \subset S(z) \subset B(z, 3L) \] (43a)

To estimate the volume in (42c), consider the mapping \( \varphi(r, \xi) = Exp_x(r \xi) \) for \( r > 0 \) and \( \xi \in S_x, M \). Since \( M \) is a Hadamard manifold, \( \varphi \) is a diffeomorphism onto \( M \) [11]. Now, the first inclusion in (43a) implies
\[ vol(B(z, L) \cap C_z(\varepsilon)) \leq vol(S(z) \cap C_z(\varepsilon)) = \int_{r(z)-L}^{r(z)+L} \int_{S_x, M} 1_{C_z(\varepsilon)}(\varphi(r, \xi)) \lambda(r, \xi) \omega(d\xi) \, dr \]
where \( \varphi^*vol(dr \, d\xi) = \lambda(r, \xi) \omega(d\xi) dr \) denotes the pullback of the Riemannian volume measure \( vol \) under \( \varphi \), with \( \omega(d\xi) \) the area measure on the unit sphere \( S_x, M \). Bounding the last integral from above,
\[ vol(B(z, L) \cap C_z(\varepsilon)) \leq 2\varepsilon \omega(S_x, M) \sup_{\varphi(r, \xi) \in B(z, 3L)} \lambda(r, \xi) \] (44a)

Similarly, the second inclusion in (43a) implies
\[ vol(B(z, 3L)) \geq vol(S(z)) = \int_{r(z)-L}^{r(z)+L} \int_{S_x, M} \lambda(r, \xi) \omega(d\xi) \, dr \]
and bounding the last integral from below gives,
\[
\text{vol}(B(z, 3L)) \geq 2L \omega(S_\ast M) \inf_{\phi(r, \xi) \in B(z, 3L)} \lambda(r, \xi)
\]
(44b)

From (44a) and (44b) it follows that
\[
\text{vol}(B(z, L) \cap C_x(\varepsilon)) \leq \varepsilon \text{vol}(B(z, 3L)) \sup_{\phi(r, \xi) \in B(z, 3L)} \frac{\lambda(r, \xi)}{\inf_{\phi(r, \xi) \in B(z, 3L)} \lambda(r, \xi)}
\]
(44c)

However, since $M$ is a symmetric Hadamard manifold (in particular, then, a Riemannian homogeneous space [10]), vol $(B(z, 3L))$ does not depend on $z$. Therefore, by (44c), to prove (42c), it is enough to show that
\[
\limsup_{r(z) \to \infty} \sup_{\phi(r, \xi) \in B(z, 3L)} \frac{\lambda(r, \xi)}{\inf_{\phi(r, \xi) \in B(z, 3L)} \lambda(r, \xi)} < \infty
\]
(45)

Once this is done, (42b) follows immediately from (42c), and this completes the proof of (42a).

**Conclusion:** finally, (39a) can be obtained by combining (40a), (41a) and (42a). Precisely, the integral under the limit in (39a) can be decomposed into the sum of three integrals
\[
\left( \int_{A(z) - B(z, L)} + \int_{A(z) \cap B(z, L) - C_x(\varepsilon)} + \int_{A(z) \cap B(z, L) \cap C_x(\varepsilon)} \right) q(z, y) (\alpha(y, z))^\frac{1}{2} \text{vol}(dy)
\]
By (40a), for any $\Delta > 0$, it is possible to chose $L$ to make the first integral less than $\Delta/3$, irrespective of $z$ and $\varepsilon$. By (42a), it is possible to chose $\varepsilon$ to make the third integral less than $\Delta/3$, for all $z$ with sufficiently large $r(z)$. With $L$ and $\varepsilon$ chosen in this way, by (41a), if $r(z)$ is sufficiently large, then the second integral is less than $\Delta/3$. Then, the sum of the three integrals is $< \Delta$, and (39a) follows since $\Delta$ is arbitrary.

**Proof of (45):** let $R$ denote the Riemann curvature tensor of $M$, and consider for each $\xi \in S_\ast M$ the linear operator $R_\xi : T_\ast M \to T_\ast M$ given by
\[
R_\xi(\zeta) = - R(\xi, \zeta) \zeta ; \quad \zeta \in T_\ast M
\]
Then, $R_\xi$ is self-adjoint with respect to the restriction of the Riemannian metric of $M$ to $T_\ast M$, and all of its eigenvalues are positive [10][20]. Now, if $\kappa^2(\xi)$ runs through the eigenvalues of $R_\xi$ then $\lambda(r, \xi)$ has the following expression
\[
\lambda(r, \xi) = \prod_{\kappa(\xi)} \left( \frac{\sinh(\kappa(\xi) r)}{\kappa(\xi)} \right)^{m_{\kappa(\xi)}}
\]
(46a)
where $m_{\kappa(\xi)}$ denotes the multiplicity of the eigenvalue $\kappa^2(\xi)$ of $R_\xi$. The expression in (46a) follows from the solution of the Jacobi equation, valid for any locally symmetric space, given in [20].

This expression may be written under the following, different form. Let $M = G/K$ where $(G, K)$ is a Riemannian symmetric pair of non-compact type [10]. Let $\mathfrak{g}$ and $\mathfrak{k}$ be the Lie algebras of $G$ and $K$, and $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$ the corresponding Cartan decomposition. If $\mathfrak{a}$ is a maximal Abelian subspace of $\mathfrak{p}$, then for each $r > 0$ and $\xi \in S_\ast M$ it is possible to write $r \xi = \text{Ad}(k) \cdot a$ for some $k \in K$ and $a \in \mathfrak{a}$, where $\text{Ad}$ stands for the adjoint representation. Using this notation [10], $r = \|a\|$ where $\| \cdot \|$ denotes the Riemannian norm (here, $\mathfrak{p}$ is naturally identified with $T_\ast M$). Moreover, each $\kappa(\xi)$ can be written $\kappa(\xi) = \alpha(\alpha)/\|a\|$ where $\alpha$ is a positive restricted root associated to the pair $(G, K)$, with multiplicity $m_\alpha = m_{\kappa(\xi)}$. Replacing into (46a), this gives
\[
\lambda(r, \xi) = \prod_{\alpha} \left( \frac{\|a\|}{\alpha(\alpha)} \sinh(\alpha(\alpha)) \right)^{m_\alpha}
\]
(46b)
Here, if the right-hand side is denoted by $f(a)$, then it is elementary that $\log f(a)$ is a Lipschitz function on the complement of any bounded subset of $\mathfrak{a}$ which contains the origin of $\mathfrak{a}$.

Returning to (45), let the supremum in the numerator be achieved at $(r_{\text{max}}, \xi_{\text{max}})$ and the infimum in the denominator be achieved at $(r_{\text{min}}, \xi_{\text{min}})$. Also, let $(k_{\text{max}}, a_{\text{max}})$ and $(k_{\text{min}}, a_{\text{min}})$ be the corresponding values of $k$ and $a$. Note that for all $(r, \xi)$ such that $\phi(r, \xi) \in B(z, 3L)$ it holds that $r \geq r(z) - 3L$. But, since $r = \|a\|$, this also means $\|a\| \geq r(z) - 3L$. If $r(z) > 3L$ then, as stated above, $\log f(a)$ is Lipschitz on the set of $a$ such that $\|a\| \geq r(z) - 3L$. Thus, if $C$ is a corresponding Lipschitz constant, it follows that
\[
\sup_{\phi(r, \xi) \in B(z, 3L)} \frac{\lambda(r, \xi)}{\inf_{\phi(r, \xi) \in B(z, 3L)} \lambda(r, \xi)} = \frac{f(a_{\text{max}})}{f(a_{\text{min}})} \leq e^{C \|a_{\text{max}} - a_{\text{min}}\|}
\]
(46c)
Now, (45) will follow by showing that $\|a_{\text{max}} - a_{\text{min}}\| \leq 6L$ whenever $r(z) > 3L$. To do so, let $y_{\text{max}} = \phi(r_{\text{max}}, \xi_{\text{max}})$ and $y_{\text{min}} = \phi(r_{\text{min}}, \xi_{\text{min}})$. Since both $y_{\text{max}}$ and $y_{\text{min}}$ belong to the closure of $B(z, 3L)$, it follows from the triangle inequality that $d(y_{\text{max}}, y_{\text{min}}) \leq 6L$. If $\eta(t)$ is a geodesic with $\eta(0) = y_{\text{min}}$ and $\eta(1) = y_{\text{max}}$ then
\[
\int_0^1 \|\dot{\eta}(t)\| \, dt = d(y_{\text{max}}, y_{\text{min}}) \leq 6L \tag{47a}
\]

On the other hand, if \( \eta(t) = \varphi(r(t), \xi(t)) \) then it is possible to write \( r(t) \xi(t) = \text{Ad}(k(t)) \cdot a(t) \) where \( k(t) \) and \( a(t) \) are differentiable curves in \( K \) and \( a \), respectively. Using once more the solution of the Jacobi equation given in [20], the following expression of the Riemannian norm can be obtained

\[
\|\dot{\eta}(t)\|^2 = \|\dot{a}(t)\|^2 + \sum_a \left( \sinh(\alpha(a(t))) \|\dot{k}_\alpha(t)\| \right)^2 \geq \|\dot{a}(t)\|^2 \tag{47b}
\]

where \( \dot{k}_\alpha(t) \) denotes the orthogonal projection of the Lie bracket \( [\dot{k}(t), a(t)] \) onto the eigenspace of the linear operator \( R_{r\xi(t)} \) corresponding to the eigenvalue \( \kappa^2_{r\xi(t)} = \alpha^2(a(t)) \) (this orthogonal projection is well-defined since the Lie bracket just mentioned lies in \( p \) for any value of \( t \) [10]). Finally, from (47a) and (47b), it follows that

\[
\|a_{\text{max}} - a_{\text{min}}\| \leq \int_0^1 \|\dot{a}(t)\| \, dt \leq \int_0^1 \|\dot{\eta}(t)\| \, dt \leq 6L
\]

where the first inequality follows because \( \|a_{\text{max}} - a_{\text{min}}\| \) is the length of a straight line in \( a \) from \( a_{\text{min}} \) to \( a_{\text{max}} \), while the differentiable curve \( a(t) \) also connects \( a_{\text{min}} \) to \( a_{\text{max}} \). Now, it is possible to replace in (46c), obtaining

\[
\sup_{\varphi(r,\xi) \in B(z, 3L)} \lambda(r, \xi) \lambda(r, \xi) \leq e^{6LC} \tag{47c}
\]

for all \( z \) such that \( r(z) > 3L \). However, this immediately implies (45). \( \blacksquare \)