The Riemannian barycentre as a proxy for global optimisation

Salem Said and Jonathan H. Manton

1 Laboratoire IMS (CNRS 5218), Université de Bordeaux
   salem.said@u-bordeaux.fr
2 Department of Electrical and Electronic Engineering, The University of Melbourne
   j.manton@ieee.org

Abstract. Let $M$ be a simply-connected compact Riemannian symmetric space, and $U$ a twice-differentiable function on $M$, with unique global minimum at $x^* \in M$. The idea of the present work is to replace the problem of searching for the global minimum of $U$, by the problem of finding the Riemannian barycentre of the Gibbs distribution $P_T \propto \exp(-U/T)$. In other words, instead of minimising the function $U$ itself, to minimise $E_T(x) = \frac{1}{2} \int d^2(x, z) P_T(dz)$, where $d(\cdot, \cdot)$ denotes Riemannian distance.

The following original result is proved: if $U$ is invariant by geodesic symmetry about $x^*$, then for each $\delta < \frac{1}{2} r_{cx}$ (where $r_{cx}$ is the convexity radius of $M$), there exists $T_\delta$ such that $T \leq T_\delta$ implies $E_T$ is strongly convex on the geodesic ball $B(x^*, \delta)$, and $x^*$ is the unique global minimum of $E_T$. Moreover, this $T_\delta$ can be computed explicitly. This result gives rise to a general algorithm for black-box optimisation, which is briefly described, and will be further explored in future work.

Keywords: Riemannian barycentre · black-box optimisation · symmetric space.

It is common knowledge that the Riemannian barycentre $\bar{x}$, of a probability distribution $P$ defined on a Riemannian manifold $M$, may fail to be unique. However, if $P$ is supported inside a geodesic ball $B(x^*, \delta)$ with radius $\delta < \frac{1}{2} r_{cx}$ ($r_{cx}$ the convexity radius of $M$), then $\bar{x}$ is unique and also belongs to $B(x^*, \delta)$. In fact, Afsari has shown this to be true, even when $\delta < r_{cx}$ (see [1, 2]).

Does this statement continue to hold, if $P$ is not supported inside $B(x^*, \delta)$, but merely concentrated on this ball? The answer to this question is positive, assuming that $M$ is a simply-connected compact Riemannian symmetric space, and $P = P_T \propto \exp(-U/T)$, where the function $U$ has unique global minimum at $x^* \in M$. This is given by Proposition 2 in Section 2 below.

Proposition 2 motivates the main idea of the present work: the Riemannian barycentre $\bar{x}_T$ of $P_T$ can be used as a proxy for the global minimum $x^*$ of $U$. In general, $\bar{x}_T$ only provides an approximation of $x^*$, but the two are equal if $U$ is invariant by geodesic symmetry about $x^*$, as stated in Proposition 3 in Section 4 below.

The following Section 1 introduces Proposition 1 which estimates the Riemannian distance between $\bar{x}_T$ and $x^*$, as a function of $T$. 

1 Concentration of the barycentre

Let $P$ be a probability distribution on a complete Riemannian manifold $M$. A (Riemannian) barycentre of $P$ is any global minimiser $\bar{x} \in M$ of the function

$$E(x) = \frac{1}{2} \int_M d^2(x, z) P(dz) \quad \text{for } x \in M$$

(1)

The following statement is due to Karcher, and was improved upon by Afsari [1][2]: if $P$ is supported inside a geodesic ball $B(x^*, \delta)$, where $x^* \in M$ and $\delta < \frac{1}{2}r_{cx}$ ($r_{cx}$ the convexity radius of $M$), then $E$ is strongly convex on $B(x^*, \delta)$, and $P$ has a unique barycentre $\bar{x} \in B(x^*, \delta)$.

On the other hand, the present work considers a setting where $P$ is not supported inside $B(x^*, \delta)$, but merely concentrated on this ball. Precisely, assume $P$ is equal to the Gibbs distribution

$$P_T(dz) = \left( \frac{Z(T)}{\text{vol}(dz)} \right)^{-1} \exp \left( -\frac{U(z)}{T} \right) \text{vol}(dz) \quad T > 0$$

(2)

where $Z(T)$ is a normalising constant, $U$ is a $C^2$ function with unique global minimum at $x^*$, and vol is the Riemannian volume of $M$. Then, let $E_T$ denote the function $E$ in (1), and let $\bar{x}_T$ denote any barycentre of $P_T$.

In this new setting, it is not clear whether $E_T$ is differentiable or not. Therefore, statements about convexity of $E_T$ and uniqueness of $\bar{x}_T$ are postponed to the following Section 2. For now, it is possible to state the following Proposition 1.

In this proposition, $d(\cdot, \cdot)$ denotes Riemannian distance, and $W(\cdot, \cdot)$ denotes the Kantorovich ($L^1$-Wasserstein) distance [3][4]. Moreover, $(\mu_{\text{min}}, \mu_{\text{max}})$ is any open interval which contains the spectrum of the Hessian $\nabla^2 U(x^*)$, considered as a linear mapping of the tangent space $T_x M$.

**Proposition 1.** Assume $M$ is an $n$-dimensional compact Riemannian manifold with non-negative sectional curvature. Denote $\delta_{x^*}$ the Dirac distribution at $x^*$. The following hold,

(i) for any $\eta > 0$,

$$W(P_T, \delta_{x^*}) < \frac{\eta^2}{(4 \text{ diam } M)} \implies d(\bar{x}_T, x^*) < \eta$$

(3)

(ii) for $T \leq T_0$ (which can be computed explicitly)

$$W(P_T, \delta_{x^*}) \leq \sqrt{2\pi} \left( \frac{\pi/2}{\mu_{\text{max}}/\mu_{\text{min}}} \right)^{n/2} \frac{1}{2} \left( T/\mu_{\text{min}} \right)^{1/2}$$

(4)

where $B_n = B(1/2, n/2)$ in terms of the Beta function.

Proposition 1 is motivated by the idea of using $\bar{x}_T$ as an approximation of $x^*$. Intuitively, this requires choosing $T$ so small that $P_T$ is sufficiently close to $\delta_{x^*}$. Just how small a $T$ may be required is indicated by the inequality in (4). This inequality is optimal and explicit, in the following sense.
It is optimal because the dependence on $T^{1/2}$ in its right-hand side cannot be improved. Indeed, by the multi-dimensional Laplace approximation (see [5], for example), the left-hand side is equivalent to $L \cdot T^{1/2}$ (in the limit $T \to 0$). While this constant $L$ is not tractable, the constants appearing in Inequality (4) depend explicitly on the manifold $M$ and the function $U$. In fact, this inequality does not follow from the multi-dimensional Laplace approximation, but rather from volume comparison theorems of Riemannian geometry [6].

In spite of these nice properties, Inequality (4) does not escape the curse of dimensionality. Indeed, for fixed $T$, its right-hand side increases exponentially with the dimension $n$ (note that $B_n$ decreases like $n^{-1/2}$). On the other hand, although $T_o$ also depends on $n$, it is typically much less affected by dimensionality, and decreases slower that $n^{-1}$ as $n$ increases.

2 Convexity and uniqueness

Assume now that $M$ is a simply-connected, compact Riemannian symmetric space. In this case, for any $T$, the function $E_T$ turns out to be $C^2$ throughout $M$. This results from the following lemma.

Lemma 1. let $M$ be a simply-connected compact Riemannian symmetric space. Let $\gamma : I \to M$ be a geodesic defined on a compact interval $I$. Denote $\text{Cut}(\gamma)$ the union of all cut loci $\text{Cut}(\gamma(t))$ for $t \in I$. Then, the topological dimension of $\text{Cut}(\gamma)$ is strictly less than $n = \dim M$. In particular, $\text{Cut}(\gamma)$ is a set with volume equal to zero.

Remark: the assumption that $M$ is simply-connected cannot be removed, as the conclusion does not hold if $M$ is a real projective space.

The proof of Lemma 1 uses the structure of Riemannian symmetric spaces, as well as some results from topological dimension theory [7] (Chapter VII). The notion of topological dimension arises because it is possible $\text{Cut}(\gamma)$ is not a manifold. The lemma immediately implies, for all $t$,

$$E_T(\gamma(t)) = \frac{1}{2} \int_M d^2(\gamma(t),z)P_T(dz) = \frac{1}{2} \int_{M-\text{Cut}(\gamma)} d^2(\gamma(t),z)P_T(dz)$$

Then, since the domain of integration avoids the cut loci of all the $\gamma(t)$, it becomes possible to differentiate under the integral. This is used in obtaining the following (the assumptions are the same as in Lemma 1).

Corollary 1. for $x \in M$, let $G_x(z) = \nabla f_z(x)$ and $H_x(z) = \nabla^2 f_z(x)$, where $f_z$ is the function $x \mapsto \frac{1}{2} d^2(x,z)$. The following integrals converge for any $T$

$$G_x = \int_{M-\text{Cut}(x)} G_x(z) P_T(dz) \quad H_x = \int_{M-\text{Cut}(x)} H_x(z) P_T(dz)$$

and both depend continuously on $x$. Moreover,

$$\nabla E_T(x) = G_x \quad \nabla^2 E_T(x) = H_x$$

so that $E_T$ is $C^2$ throughout $M$. 

With Corollary 1 at hand, it is possible to obtain Proposition 2, which is concerned with the convexity of $\mathcal{E}_t$ and uniqueness of $\bar{x}_t$. In this proposition, the following notation is used

$$f(T) = (2/\pi) (\pi/8)^{\kappa/2} (\mu_{\text{max}}/T)^{\kappa/2} \exp (-U_\delta/T)$$

where $U_\delta = \inf \{ U(x) - U(x^*) ; x \notin B(x^*, \delta) \}$ for positive $\delta$. The reader may wish to note the fact that $f(T)$ decreases to 0 as $T$ decreases to 0.

**Proposition 2.** Let $M$ be a simply-connected compact Riemannian symmetric space. Let $\kappa^2$ be the maximum sectional curvature of $M$, and $r_{\text{ex}} = \kappa^{-1} \pi$ its convexity radius. If $T \leq T_o$ (see (ii) of Proposition 1), then the following hold for any $\delta < \frac{1}{2} r_{\text{ex}}$.

(i) for all $x$ in the geodesic ball $B(x^*, \delta)$,

$$\nabla^2 \mathcal{E}_t(x) \geq \text{Ct}(2\delta) (1 - \text{vol}(M)) f(T) - \pi A_M f(T)$$

where $\text{Ct}(2\delta) = 2\kappa\delta \cot(2\kappa\delta) > 0$ and $A_M > 0$ is a constant given by the structure of the symmetric space $M$.

(ii) there exists $T_o$ (which can be computed explicitly), such that $T \leq T_o$ implies $\mathcal{E}_t$ is strongly convex on $B(x^*, \delta)$, and has a unique global minimum $\bar{x}_t \in B(x^*, \delta)$. In particular, this means $\bar{x}_t$ is the unique barycentre of $P_t$.

Note that (ii) of Proposition 2 generalises the statement due to Karcher [1], which was recalled in Section 1.

### 3 Finding $T_o$ and $T_\delta$

Propositions 1 and 2 claim that $T_o$ and $T_\delta$ can be computed explicitly. This means that, with some knowledge of the Riemannian manifold $M$ and the function $U$, $T_o$ and $T_\delta$ can be found by solving scalar equations. The current section gives the definitions of $T_o$ and $T_\delta$.

In the notation of Proposition 1, let $\rho > 0$ be small enough, so that,

$$\mu_{\text{min}} d^2 (x, x^*) \leq 2 (U(x) - U(x^*)) \leq \mu_{\text{max}} d^2 (x, x^*)$$

whenever $d(x, x^*) \leq \rho$, and consider the quantity

$$f(T, m, \rho) = (2/\pi)^{1/2} (\mu_{\text{max}}/T)^{m/2} \exp (-U_\rho/T)$$

where $U_\rho$ is defined as in (6). Note that $f(T, m, \rho)$ decreases to 0 as $T$ decreases to 0, for fixed $m$ and $\rho$. Now, it is possible to define $T_o$ as

$$T_o = \min \{ T_o^1, T_o^2 \}$$

where

$$T_o^1 = \inf \left\{ T > 0 : f(T, m - 2, \rho) > \rho^{2-n} A_{n-1} \right\}$$

$$T_o^2 = \inf \left\{ T > 0 : f(T, m + 1, \rho) > (\mu_{\text{max}} / \mu_{\text{min}})^{n/2} C_n \right\}$$

Here, $A_n = E|X|^n$ for $X \sim N(0, 1)$, and $C_n = \omega_n A_n / (\text{diam} M \times \text{vol} M)$, where $\omega_n$ is the surface area of a unit sphere $S^{n-1}$. 
With regard to Proposition 2, define $T_i$ as follows,

$$T_i = \min \left\{ T_1^i, T_2^i \right\} - \varepsilon$$

for some arbitrary $\varepsilon > 0$. Here, in the notation of (4), (5) and (7),

$$T_1^i = \inf \left\{ T \leq T_0 : \sqrt{2\pi (T/\mu_{\min})^{1/2}} > \delta^2 (\mu_{\min}/\mu_{\max})^{n/2} D_n \right\}$$

$$T_2^i = \inf \left\{ T \leq T_0 : f(T) > C t(2\delta) (C t(2\delta) \text{vol}(M) + \pi A_M)^{-1} \right\}$$

where $D_n = (2/\pi)^{n-1} B_n/(4 \text{ diam} M)$.

### 4 Black-box optimisation

Consider the problem of searching for the unique global minimum $x^*$ of $U$. In black-box optimisation, it is only possible to evaluate $U(x)$ for given $x \in M$, and the cost of this evaluation precludes numerical approximation of derivatives. Then, the problem is to find $x^*$ using successive evaluations of $U(x)$ (hopefully, as few of these evaluations as possible).

Here, a new algorithm for solving this problem is described. The idea of this algorithm is to find $\bar{x}_T$ using successive evaluations of $U(x)$, in the hope that $\bar{x}_T$ will provide a good approximation of $x^*$. While the quality of this approximation is controlled by Inequalities (3) and (4) of Proposition 1 in some cases of interest, $\bar{x}_T$ is exactly equal to $x^*$, for correctly chosen $T$, as in the following proposition 3.

To state this proposition, let $s_\cdot$ denote geodesic symmetry about $x^*$ (see [7]). This is the transformation of $M$, which leaves $x^*$ fixed, and reverses the direction of geodesics passing through $x^*$.

**Proposition 3.** Assume that $U$ is invariant by geodesic symmetry about $x^*$, in the sense that $U \circ s_{\cdot} = U$. If $T \leq T_i$ (see (ii) of Proposition 2), then $\bar{x}_T = x^*$ is the unique barycentre of $P_T$.

Proposition 3 follows rather directly from Proposition 2. Precisely, by (ii) of Proposition 2, the condition $T \leq T_i$ implies $E_T$ is strongly convex on $B(x^*, \delta)$, and $\bar{x}_T \in B(x^*, \delta)$. Thus, $\bar{x}_T$ is the unique stationary point of $E_T$ in $B(x^*, \delta)$. But, using the fact that $U$ is invariant by geodesic symmetry about $x^*$, it is possible to prove that $x^*$ is a stationary point of $E_T$, and this implies $\bar{x}_T = x^*$.

The two following examples verify the conditions of Proposition 3.

**Example 1.** Assume $M = \text{Gr}(k, \mathbb{C}^n)$ is a complex Grassmann manifold. In particular, $M$ is a simply-connected, compact Riemannian symmetric space. Identify $M$ with the set of Hermitian projectors $x : \mathbb{C}^n \to \mathbb{C}^n$ such that $\text{tr}(x) = k$, where $\text{tr}$ denotes the trace. Then, define $U(x) = -\text{tr}(C x)$ for $x \in \text{Gr}(k, \mathbb{C}^n)$, where $C$ is a Hermitian positive-definite matrix with distinct eigenvalues. Now, the unique global minimum of $U$ occurs at $x^*$, the projector onto the principal $k$-subspace of $C$. Also, the geodesic symmetry $s_{\cdot}$ is given by $s_{\cdot} : x = r_{\cdot} x r_{\cdot}$, where $r_{\cdot} : \mathbb{C}^n \to \mathbb{C}^n$ denotes reflection through the image space of $x^*$. It is elementary to verify that $U$ is invariant by this geodesic symmetry.
Example 2: Let $M$ be a simply-connected, compact Riemannian symmetric space, and $U_o$ a function on $M$ with unique global minimum at $o \in M$. Assume moreover that $U_o$ is invariant by geodesic symmetry about $o$. For each $x^* \in M$, there exists an isometry $g$ of $M$, such that $x^* = g \cdot o$. Then, $U(x) = U_o(g^{-1} \cdot x)$ has unique global minimum at $x^*$, and is invariant by geodesic symmetry about $x^*$.

Example 1 describes the standard problem of finding the principal subspace of the covariance matrix $C$. In Example 2, the function $U_o$ is a known template, which undergoes an unknown transformation $g$, leading to the observed pattern $U$. This is a typical situation in pattern recognition problems.

Of course, from a mathematical point of view, Example 2 is not really an example, since it describes the completely general setting where the conditions of Proposition 3 are verified. In this setting, consider the following algorithm.

Description of the algorithm:

- Input: $T \leq T_\delta$  
  - to find such $T$, see Section 3
  - $Q(x, dz) = q(x, z) \text{vol}(dz)$  
    - symmetric Markov kernel
  - $\hat{x}_0 = z_0 \in M$  
    - initial guess for $x^*$

- Iterate: For $n = 1, 2, \ldots$

  1. Sample $z_n \sim q(z_{n-1}, z)$
  2. Compute $r_n = 1 - \min\{1, \exp[(U(z_{n-1}) - U(z_n))/T]\}$
  3. Reject $z_n$ with probability $r_n$  
    - then, $z_n = z_{n-1}$
  4. $\hat{x}_n = \hat{x}_{n-1} \#_{1/n} z_n$  
    - see definition (10) below

- Until: $\hat{x}_n$ does not change sensibly

- Output: $\hat{x}_n$  
  - approximation of $x^*$

The above algorithm recursively computes the Riemannian barycentre $\hat{x}_n$ of the samples $z_n$ generated by a symmetric Metropolis-Hastings algorithm (see [8]). Here, the Metropolis-Hastings algorithm is implemented in lines (1)--(3). On the other hand, line (4) takes care of the Riemannian barycentre. Precisely, if $\gamma: [0, 1] \to M$ is a length-minimising geodesic connecting $\hat{x}_{n-1}$ to $z_n$, let

$$\hat{x}_{n-1} \#_{1/n} z_n = \gamma(1/n)$$  
(10)

This geodesic $\gamma$ need not be unique.

The point of using the Metropolis-Hastings algorithm is that the generated $z_n$ eventually sample from the Gibbs distribution $P_T$. The convergence of the distribution $P_n$ of $z_n$ to $P_T$ takes place exponentially fast. Indeed, it may be inferred from [8] (see Theorem 8, Page 36)

$$\|P_n - P_T\|_{TV} \leq (1 - p_T)^n$$  
(11)

where $\| \cdot \|_{TV}$ is the total variation norm, and $p_T \in (0, 1)$ verifies

$$p_T \leq (\text{vol } M) \inf_{x,z} q(x, z) \exp(-\sup_x U(x)/T)$$

so the rate of convergence is degraded when $T$ is small.
Accordingly, the intuitive justification of the above algorithm is the following. Since the \( z_n \) eventually sample from the Gibbs distribution \( P_T \), and the desired global minimum \( x^* \) of \( U \) is equal to the barycentre \( \bar{x}_T \) of \( P_T \) (by Proposition 3), then the barycentre \( \hat{x}_n \) of the \( z_n \) is expected to converge to \( x^* \).

It should be emphasised that, in the present state of the literature, there is no rigorous result which confirms this convergence \( z_n \to x^* \). It is therefore an open problem, to be confronted in future work.

For a basic computer experiment, consider \( M = S^2 \subset \mathbb{R}^3 \), and let

\[
U(x) = -P_9(x^3) \quad \text{for} \quad x = (x^1, x^2, x^3) \in S^2
\]

where \( P_9 \) is the Legendre polynomial of degree 9 [9]. The unique global minimiser of \( U \) is \( x^* = (0, 0, 1) \), and the conditions of Proposition 3 are verified, since \( U \) is invariant by reflection in the \( x^3 \) axis, which is geodesic symmetry about \( x^* \).

![Fig. 1. graph of \(-P_9(x^3)\)](image1)

![Fig. 2. \( \hat{x}_n^3 \) versus \( n \)](image2)

Figure 1 shows the dependence of \( U(x) \) on \( x^3 \), displaying multiple local minima and maxima. Figure 2 shows the algorithm overcoming these local minima and maxima, and converging to the global minimum \( x^* = (0, 0, 1) \), within \( n = 5000 \) iterations. The experiment was conducted with \( T = 0.2 \), and the Markov kernel \( Q \) obtained from the von Mises-Fisher distribution (see [10]). The initial guess \( \hat{x}_0 = (0, 0, -1) \) is not shown in Figure 2.

In comparison, a standard simulated annealing method offered less robust performance, which varied considerably with the choice of annealing schedule.
5 Proofs

This section is devoted to the proofs of the results stated in previous sections.

As of now, assume that $U(x^*) = 0$. There is no loss of generality in making this assumption.

5.1 Proof of Proposition 1

Proof of (i): denote $f_x(z) = \frac{1}{2} d^2(x, z)$. By the definition of $E_T$

$$E_T(x) = \int_M f_x(z) P_T(dz) \tag{13a}$$

Moreover, let $E_0$ be the function

$$E_0(x) = \int_M f_x(z) \delta_{x^*}(dz) = \frac{1}{2} d^2(x, x^*) \tag{13b}$$

For any $x$, it is elementary that $f_x(z)$ is Lipschitz continuous, with respect to $z$, with Lipschitz constant $\text{diam } M$. Then, from the Kantorovich-Rubinshtein formula [4],

$$|E_T(x) - E_0(x)| \leq (\text{diam } M) W(P_T, \delta_{x^*}) \tag{13c}$$

a uniform bound in $x \in M$. It now follows that

$$\inf_{x \in B(x^*, \eta)} E_T(x) - \inf_{x \in B(x^*, \eta)} E_0(x) \leq (\text{diam } M) W(P_T, \delta_{x^*}) \quad \text{and} \quad (13d)$$

$$\inf_{x \notin B(x^*, \eta)} E_0(x) - \inf_{x \notin B(x^*, \eta)} E_T(x) \leq (\text{diam } M) W(P_T, \delta_{x^*}) \tag{13e}$$

However, from (13b), it is clear that

$$\inf_{x \in B(x^*, \eta)} E_0(x) = 0 \quad \text{and} \quad \inf_{x \notin B(x^*, \eta)} E_0(x) = \frac{\eta^2}{2}$$

To complete the proof, replace this into (13d) and (13e). Then, assuming the condition in (3) is verified,

$$\inf_{x \in B(x^*, \eta)} E_T(x) < \frac{\eta^2}{4} < \inf_{x \notin B(x^*, \eta)} E_T(x) \tag{13f}$$

This means that any global minimum $\bar{x}_T$ of $E_T$ must belong to the open ball $B(x^*, \eta)$. In other words, $d(\bar{x}_T, x^*) < \eta$. This completes the proof of (3).

Proof of (ii): let $\rho \leq \min \{\text{inj } x^*, \kappa^{-1} \frac{\eta}{2}\}$ where $\text{inj } x^*$ is the injectivity radius of $M$ at $x^*$, and $\kappa^2$ is an upper bound on the sectional curvature of $M$. Assume, in addition, $\rho$ is small enough so

$$\mu_{\text{min}} d^2(x, x^*) \leq 2 (U(x) - U(x^*)) \leq \mu_{\text{max}} d^2(x, x^*) \tag{14a}$$
whenever \( d(x, x^*) \leq \rho \). Further, consider the truncated distribution

\[
P_T^\rho(dz) = \frac{1_{B_\rho}(z)}{P_T(B_\rho)} \cdot P_T(dz)
\]  

where \( 1 \) denotes the indicator function, and \( B_\rho \) stands for the open ball \( B(x^*, \rho) \). Of course, by the triangle inequality,

\[
W(P_T, \delta_{x^*}) \leq W(P_T, P_T^\rho) + W(P_T^\rho, \delta_{x^*}) \tag{14c}
\]

The proof relies on the following estimates, which use the notation of Section 3.

**First estimate:** if \( T \leq T_o^1 \), then

\[
W(P_T, P_T^\rho) \leq \left( \text{diam } M \times \text{vol } M \right) \frac{2}{\pi} \left( \frac{\pi}{8} \right)^{n/2} \left( \frac{\mu_{\text{max}}}{T} \right)^{n/2} \exp \left( -\frac{U_\rho}{T} \right) \tag{14d}
\]

**Second estimate:** if \( T \leq T_o^1 \), then

\[
W(P_T^\rho, \delta_{x^*}) \leq 2 \sqrt{2\pi} \left( \frac{\pi}{2} \right)^{n-1} B_n^{-1} \left( \frac{\mu_{\text{max}}}{\mu_{\text{min}}} \right)^{n/2} \left( \frac{T}{\mu_{\text{min}}} \right)^{1/2} \tag{14e}
\]

These two estimates are proved below. Assume now they hold true, and \( T \leq T_o^1 \).

In particular, since \( T \leq T_o^2 \), the definition of \( T_o^2 \) implies

\[
f(T, n+1, \rho) \leq \left( \mu_{\text{max}} / \mu_{\text{min}} \right)^{n/2} C_n
\]

Recall the definition of \( C_n \), and express \( \omega_n \) and \( A_n \) in terms of the Gamma function [9]. The last inequality becomes

\[
(d \text{diam } M \times \text{vol } M) \ f(T, n+1, \rho) \leq 2 \left( 2\pi \right)^{n/2} B_n^{-1} \left( \mu_{\text{max}} / \mu_{\text{min}} \right)^{n/2}
\]

This is the same as

\[
(d \text{diam } M \times \text{vol } M) \ \frac{1}{\pi} \left( \frac{\pi}{8} \right)^{n/2} f(T, n+1, \rho) \leq \left( \frac{\pi}{2} \right)^{n-1} B_n^{-1} \left( \mu_{\text{max}} / \mu_{\text{min}} \right)^{n/2}
\]

By the definition of \( f(T, n+1, \rho) \), it now follows the right-hand side of \( (14d) \) is less than half the right-hand side of \( (14e) \). In this case, \( \boxed{14c} \) follows from the triangle inequality \( \boxed{14e} \).

**Proof of first estimate:** consider the coupling of \( P_T \) and \( P_T^\rho \), provided by the probability distribution \( K \) on \( M \times M \),

\[
K(dz_1 \times dz_2) = P_T^\rho(dz_1) \left[ P_T(B_\rho) \delta_{z_1}(dz_2) + 1_{B_\rho^c}(z_2) P_T(dz_2) \right] \tag{15a}
\]

where \( B_\rho^c \) denotes the complement of \( B_\rho \). Recall the definition of the Kantorovich distance (see [4]). Replacing \( \boxed{15a} \) into this definition, it follows that

\[
W(P_T, P_T^\rho) \leq (d \text{diam } M) \ P_T(B_\rho^c) \tag{15b}
\]
Then, from the definition 2 of $P_T$,
\[ P_T(B^\rho) \leq (Z(T))^{-1} \text{vol } M \exp \left( -\frac{U}{T} \right) \quad (15c) \]

Now, (14d) follows directly from (15b) and (15c), if the following lower bound on $Z(T)$ can be proved,
\[ Z(T) \geq \frac{\pi}{2} \left( \frac{8}{\pi} \right)^{n/2} \left( \frac{T}{\mu_{\text{max}}} \right)^{n/2} \quad \text{for } T \leq T^1 \quad (15d) \]

To prove this lower bound, note that
\[ Z(T) = \int_M e^{-\frac{U(z)}{T}} \text{vol}(dz) \geq \int_{B^\rho} e^{-\frac{U(z)}{T}} \text{vol}(dz) \]

Using this last inequality and (14a), it is possible to write
\[ Z(T) \geq \int_{B^\rho} e^{-\frac{U(z)}{T}} \text{vol}(dz) \geq \int_{B^\rho} e^{-\frac{\mu_{\text{max}}}{2T} r^2} \text{vol}(dz) \quad (15e) \]

Writing this last integral in Riemannian spherical coordinates,
\[ \int_{B^\rho} e^{-\frac{\mu_{\text{max}}}{2T} r^2} d^2(x,x^*) \text{vol}(dz) = \int_0^\rho \int_{S^{n-1}} e^{-\frac{\mu_{\text{max}}}{2T} r^2} \lambda(r,s) dr \omega_n(ds) \quad (15f) \]

where $\lambda(r,s)$ is the volume density in the Riemannian spherical coordinates, $r \geq 0$ and $s \in S^{n-1}$, and where $\omega_n(ds)$ is the area element of $S^{n-1}$. From the volume comparison theorem in [6] (see Page 129),
\[ \lambda(r,s) \geq \left( \kappa^{-1} \sin(\kappa r) \right)^{n-1} \geq ((2/\pi) r)^{n-1} \quad (15g) \]

where the second inequality follows since $x \mapsto \sin(x)$ is concave for $x \in (0, \pi)$. Now, it follows from (15e) and (15f),
\[ Z(T) \geq \omega_n \left( \frac{2}{\pi} \right)^{n-1} \int_0^\rho e^{-\frac{\mu_{\text{max}}}{2T} r^2} r^{n-1} dr \quad (15h) \]

where $\omega_n$ is the surface area of $S^{n-1}$. Thus, the required lower bound (15d) follows by noting that
\[ \int_0^\rho e^{-\frac{\mu_{\text{max}}}{2T} r^2} r^{n-1} dr = \left( 2\pi \right)^{1/2} \left( \frac{T}{\mu_{\text{max}}} \right)^{n/2} \pi^{n/2} \Gamma(n/2) - \int_\rho^\infty e^{-\frac{\mu_{\text{max}}}{2T} r^2} r^{n-1} dr \]

where $\Gamma(n)$ is the Gamma function. And that
\[ \int_\rho^\infty e^{-\frac{\mu_{\text{max}}}{2T} r^2} r^{n-1} dr \leq \rho^{n-2} \left( \frac{T}{\mu_{\text{max}}} \right)^{n/2} e^{-\frac{\mu_{\text{max}}}{2T} \rho^2} \leq \rho^{n-2} \left( \frac{T}{\mu_{\text{max}}} \right)^{n/2} e^{-\frac{\mu_{\text{max}}}{T} \rho^2} \]
Indeed, taken together, these give

\[ Z(T) \geq \omega_n \left( \frac{2}{\pi} \right)^{n-1} \left[ (2\pi)^{1/2} \left( \frac{T}{\mu_{\text{max}}} \right)^{n/2} A_{n-1} - \rho^{n-2} \frac{T}{\mu_{\text{max}}} e^{\frac{\nu}{T}} \right] \]

Finally, (15d) can be obtained by noting the second term in square brackets is negligible compared to the first, as \( T \) decreases to 0, and by expressing \( \omega_n \) and \( A_{n-1} \) in terms of the Gamma function [9].

\[ \square \]

**Proof of second estimate:** the Kantorovich distance between \( P_T^\rho \) and the Dirac distribution \( \delta_{x^*} \) is equal to the expectation of the distance to \( x^* \), with respect to \( P_T^\rho \) [4]. Precisely,

\[ W(P_T^\rho, \delta_{x^*}) = \int_M d(x^*, z) P_T^\rho(dz) \]

According to (2) and (14b), this is

\[ W(P_T^\rho, \delta_{x^*}) = (P_T(B_\rho)Z(T))^{-1} \int_{B_\rho} d(x^*, z) e^{-\frac{U(z)}{T}} \text{vol}(dz) \]

Using (2) to express the probability \( P_T(B_\rho) \), this becomes

\[ W(P_T^\rho, \delta_{x^*}) = \frac{\int_{B_\rho} d(x^*, z) e^{-\frac{U(z)}{T}} \text{vol}(dz)}{\int_{B_\rho} e^{-\frac{U(z)}{T}} \text{vol}(dz)} \quad (16a) \]

A lower bound on the denominator can be found from (15c) and subsequent inequalities, which were used to prove (15d). Precisely, these inequalities provide

\[ \int_{B_\rho} e^{-\frac{U(z)}{T}} \text{vol}(dz) \geq \frac{1}{2} \omega_n \left( \frac{2}{\pi} \right)^{n-1} (2\pi)^{1/2} A_{n-1} \left( \frac{T}{\mu_{\text{max}}} \right)^{n/2} \quad (16b) \]

whenever \( T \leq T_0^1 \). For the numerator in (16a), it will be shown that, for any \( T \),

\[ \int_{B_\rho} d(x^*, z) e^{-\frac{U(z)}{T}} \text{vol}(dz) \leq \omega_n (2\pi)^{1/2} A_n \left( \frac{T}{\mu_{\text{min}}} \right)^{(n+1)/2} \quad (16c) \]

Then, (14e) follows by dividing (16c) by (16b), and replacing in (16a), after noting that \( A_n/A_{n-1} = \sqrt{2\pi} B_n^{-1} \). Thus, it only remains to prove (16c). Using (14a), it is seen that

\[ \int_{B_\rho} d(x^*, z) e^{-\frac{U(z)}{T}} \text{vol}(dz) \leq \int_{B_\rho} d(x^*, z) e^{-\frac{\mu_{\text{min}}}{2T} d^2(x, x^*)} \text{vol}(dz) \]

By expressing this last integral in Riemannian spherical coordinates, as in (15c),

\[ \int_{B_\rho} d(x^*, z) e^{-\frac{U(z)}{T}} \text{vol}(dz) \leq \int_0^r \int_{S^{n-1}} r e^{-\frac{\mu_{\text{min}}}{2} r^2 \lambda(r, s)} d\omega_n(ds) \quad (16d) \]
From the volume comparison theorem in [6] (see Page 130), \( \lambda(r,s) \leq r^{n-1} \). Therefore, (16d) becomes

\[
\int_{B_r} d(x^*, z) e^{-\frac{U(z)}{r}} \vol(dz) \leq \omega_n \int_0^r e^{-\frac{\mu_{\min}}{2r^2}} r^n dr \leq \omega_n \int_0^\infty e^{-\frac{\mu_{\min}}{2r^2}} r^n dr
\]

The right-hand side is half the \( n \)th absolute moment of a normal distribution. Expressing this in terms of \( A_n \), and replacing in (16d), gives (16c).

6 Proof of Lemma 1

Denote \( G \) the connected component at identity of the group of isometries of \( M \). It will be assumed that \( G \) is simply-connected and semisimple [7]. Any geodesic \( \gamma : I \to M \) is of the form [7][11],

\[
\gamma(t) = \exp(tY) \cdot x
\]

for some \( x \in M \) and \( Y \in \mathfrak{g} \), the Lie algebra of \( G \), where \( \exp : \mathfrak{g} \to G \) denotes the Lie group exponential mapping, and the dot denotes the action of \( G \) on \( M \).

For each \( t \in I \), the cut locus \( \Cut(\gamma(t)) \) of \( \gamma(t) \) is given by

\[
\Cut(\gamma(t)) = \exp(tY) \cdot \Cut(x)
\]

This is due to a more general result: let \( M \) be a Riemannian manifold and \( g : M \to M \) be an isometry of \( M \). Then, \( \Cut(g \cdot x) = g \cdot \Cut(x) \) for all \( x \in M \).

This is because \( y \in \Cut(x) \) if and only if \( y \) is conjugate to \( x \) along some geodesic, or there exist two different geodesics connecting \( x \) to \( y \) [6][11]. Both of these properties are preserved by the isometry \( g \).

In order to describe the set \( \Cut(x) \), denote \( K \) the isotropy group of \( x \) in \( G \), and \( \mathfrak{k} \) the Lie algebra of \( K \). Let \( \mathfrak{g} = \mathfrak{k} + \mathfrak{p} \) be an orthogonal decomposition, with respect to the Killing form of \( \mathfrak{g} \), and let \( \mathfrak{a} \) be a maximal Abelian subspace of \( \mathfrak{p} \). Define \( S = K/C_\mathfrak{a} \) (\( C_\mathfrak{a} \) the centraliser of \( \mathfrak{a} \) in \( K \)), and consider the mapping

\[
\phi(s,a) = \exp(\Ad(s) a) \cdot x \quad \text{for} \quad (s,a) \in S \times \mathfrak{a}
\]

The set \( \Cut(x) \) is the image under \( \phi \) of a certain set \( S \times \partial Q \), which is now described, following [7][12].

Let \( \Delta_+ \) be the set of positive restricted roots associated to the pair \((G, K)\), (each \( \lambda \in \Delta_+ \) is a linear form \( \lambda : \mathfrak{a} \to \mathbb{R} \)). Then, let \( Q \) be the set of \( a \in \mathfrak{a} \) such that \( |\lambda(a)| \leq \pi \) for all \( \lambda \in \Delta_+ \), and \( \partial Q \) the boundary of \( Q \). Then

\[
\Cut(x) = \phi(S \times \partial Q)
\]

Recapitulating (17b) and (17d),

\[
\Cut(\gamma) = \Phi(I \times S \times \partial Q) \quad \text{where} \quad \Phi(t,s,a) = \exp(tY) \cdot \phi(s,a)
\]

Lemma 1 states that the topological dimension of \( \Cut(\gamma) \) is strictly less than \( \dim M \). This is proved using results from topological dimension theory [7][13].
Note that both $I$ and $S$ are compact. Indeed, $S$ is compact since it is the continuous image of the compact group $K$ under the projection $K \to K/C_a$. Also, $\partial Q$ is compact in $a$, and $\partial Q = \cup_\lambda \partial Q_\lambda$ where $\partial Q_\lambda = \partial Q \cap \{\lambda(a) = \pm \pi\}$ for $\lambda \in \Delta_+$. Since $\{\lambda(a) = \pm \pi\}$ is the union of two (closed) hyperplanes in $a$, $\partial Q_\lambda$ is compact. Now, each $I \times S \times \partial Q_\lambda$ is compact, and therefore closed. It follows from (17e) that (see [13], Page 30),

$$\dim \text{Cut}(\gamma) = \dim \bigcup_\lambda \Phi(I \times S \times \partial Q_\lambda) \leq \max_\lambda \dim \Phi(I \times S \times \partial Q_\lambda) \quad (17f)$$

But, for each $\lambda$,

$$\Phi(I \times S \times \partial Q_\lambda) = \Phi(I \times S_\lambda \times \partial Q_\lambda) \subset \Phi(\mathbb{R} \times S_\lambda \times \{\lambda(a) = \pm \pi\})$$

where $S_\lambda = K/C_\lambda (C_\lambda$ the centraliser of $\{\lambda(a) = \pm \pi\}$ in $K$). The above inclusion implies (by [13], Page 26),

$$\dim \Phi(I \times S \times \partial Q_\lambda) \leq \dim \Phi(\mathbb{R} \times S_\lambda \times \{\lambda(a) = \pm \pi\}) \quad (17g)$$

To conclude, note that the set $\mathbb{R} \times S_\lambda \times \{\lambda(a) = \pm \pi\}$ is a differentiable manifold. It follows that (see [7], Page 345),

$$\dim \Phi(\mathbb{R} \times S_\lambda \times \{\lambda(a) = \pm \pi\}) \leq \dim (\mathbb{R} \times S_\lambda \times \{\lambda(a) = \pm \pi\}) \quad (17h)$$

The right-hand side of this inequality is

$$\dim (\mathbb{R} \times S_\lambda \times \{\lambda(a) = \pm \pi\}) = 1 + \dim S_\lambda + \dim a - 1$$

since the dimension of a hyperplane in $a$ is $\dim a - 1$. In addition, according to [7] (Page 296), $\dim S_\lambda \leq \dim S$. Thus,

$$\dim (\mathbb{R} \times S_\lambda \times \{\lambda(a) = \pm \pi\}) = \dim S_\lambda + \dim a < \dim M$$

since $\dim M = \dim S + \dim a$ [7]. Replacing this into (17h), it follows from (17f) and (17g) that $\dim \text{Cut}(\gamma) < \dim M$, as required.

## 7 Proof of Corollary 1

The corollary can be split into the two following claims, which will be proved separately.

**First claim:** both integrals $G_x$ and $H_x$ converge for any value of $T$.

**Second claim:** $\mathcal{E}_T$ is $C^2$ throughout $M$, with derivatives given by (5).

The fact that $G_x$ and $H_x$ depend continuously on $x$ is contained in the second claim, since (5) states that $G_x$ and $H_x$ are the gradient and Hessian of $\mathcal{E}_T$ at $x$.

In the following proofs, the notation $D(x) = M - \text{Cut}(x)$ will be used, in order to avoid cumbersome expressions.
Proof of first claim: The convergence of the integral $G_x$ is straightforward, since the integrand $G_x(z)$ is a smooth and bounded function, from $D(x)$ to $T_xM$. This is because, by definition, $G_x(z)$ is given by

$$G_x(z) = -\text{Exp}_x^{-1}(z)$$

(18)

where $\text{Exp}$ is the Riemannian exponential mapping \[6\]. Therefore, $G_x(z)$ is smooth. In addition, $G_x(z)$ is bounded, in Riemannian norm, by $\text{diam } M$.

The convergence of the integral $H_x$ is more difficult. While the integrand $H_x(z)$ is smooth on $D(x)$, it is not bounded. It will be seen that $H_x$ is an absolutely convergent improper integral.

Recall the mapping $\phi$ defined in (17c). Let $D_+$ be the set of points $a \in a$ which belong to the interior of $Q$, and which verify $\lambda(a) \geq 0$ for each $\lambda \in \Delta_+$. Let $D_+$ be the interior of $D_+$. Then, $\phi$ maps $S \times D_+$ onto $D(x)$, and is a diffeomorphism of $S \times D_+$ onto its image in $D(x)$ \[7\] \[12\] (see Chapter VII in \[7\]). Using Sard’s theorem \[14\], it follows from the definition of $H_x$ that

$$H_x = \int_S \int_{D_+} H_x(\phi(s,a)) \ p_T(\phi(s,a)) \ J(a) \ da \omega(ds)$$

(19a)

where $p_T$ denotes the density of $P_T$ with respect to the Riemannian volume of $M$, and $J(a)$ is the Jacobian determinant of $\phi$, given by \[7\]

$$J(a) = \prod_{\lambda \in \Delta_+} (\sin \lambda(a))^{m_{\lambda}}$$

(19b)

with $m_{\lambda}$ the multiplicity of the restricted root $\lambda$, and where $\omega(ds)$ is the invariant Riemannian volume induced on $S$ from $K$.

Now, $H_x(\phi(s,a))$ can be expressed as follows (cot is the cotangent function)

$$H_x(\phi(s,a)) = H_0(s) + \sum_{\lambda \in \Delta_+} \lambda(a) \cot \lambda(a) \ H_\lambda(s)$$

(19c)

where $H_0(s)$ and the $H_\lambda(s)$ denote orthogonal projectors, onto the respective eigenspaces of $H_x(\phi(s,a))$.

According to this expression, $H_x(\phi(s,a))$ diverges to $-\infty$ whenever $\lambda(a) = \pi$. However, the product

$$H_x(\phi(s,a)) \ p_T(\phi(s,a)) \ J(a)$$

which appears under the integral in (19a), is clearly continuous and bounded on the domain of integration. Thus, the absolute convergence of the integral $H_x$ follows immediately from (19a). It now remains to provide a proof of (19c). This is here only briefly indicated. Expression (19c) is a slight improvement of the one in \[15\] (see Theorem IV.1, Page 636), where it is enough to note that if $R$ is the curvature tensor of $M$, then the operator $R_v(u) = R(v, u)v$ has the eigenvalues 0 and $(\lambda(a))^2$ for each $\lambda \in \Delta_+$, whenever $v, u \in T_xM \simeq p$ with $v = \text{Ad}(s) a$ \[7\] \[12\]. It is well-known, by properties of the Jacobi equation \[6\], that $H_x(\phi(s,a))$ has the same eigenspace decomposition as $R_v$, in this case. ■
Proof of second claim: the proof of this claim relies in a crucial way on Lemma 1. To compute the gradient and Hessian of the function $E_T$ at $x \in M$, consider any geodesic $\gamma : I \to M$, defined on a compact interval $I = [-\tau, \tau]$, such that $\gamma(0) = x$. For each $t \in I$, by definition of the function $E_T$,

$$E_T(\gamma(t)) = \frac{1}{2} \int_M d^2(\gamma(t), z) P_T(dz) \quad (20a)$$

However, Lemma 1 states that the set $\text{Cut}(\gamma) = \bigcup_{t \in I} \text{Cut}(\gamma(t))$ has Riemannian volume equal to zero. From (2), it is clear that $P_T$ is absolutely continuous with respect to Riemannian volume. Therefore, $\text{Cut}(\gamma)$ can be removed from the domain of integration in (20a). Then,

$$E_T(\gamma(t)) = \frac{1}{2} \int_{D(\gamma)} d^2(\gamma(t), z) P_T(dz) \quad (20b)$$

where $D(\gamma) = M - \text{Cut}(\gamma)$. Now, for each $z \in D(\gamma)$, the function

$$t \mapsto f_z(t) = \frac{1}{2} d^2(\gamma(t), z)$$

is twice continuously differentiable with respect to $t \in I$, with

$$\frac{df_z}{dt} = \langle G_{\gamma(t)}(z), \dot{\gamma} \rangle \quad \text{and} \quad \frac{d^2 f_z}{dt^2} = H_{\gamma(t)}(z)(\dot{\gamma}, \dot{\gamma}) \quad (20c)$$

where $\langle \cdot, \cdot \rangle$ denotes the Riemannian metric of $M$, and $\dot{\gamma}$ the velocity of the geodesic $\gamma$. Indeed, this holds because the geodesic $\gamma$ does not intersect the cut locus $\text{Cut}(z)$ (see [6]).

The claim that $E_T$ is twice differentiable, and has derivatives given by (5), follows from (20b) and (20c), by differentiation under the integral sign, provided it can be shown that the families of functions

$$\{ z \mapsto G_{\gamma(t)}(z) ; t \in I \} \quad \text{and} \quad \{ z \mapsto H_{\gamma(t)}(z) ; t \in I \}$$

which all have the common domain of definition $D(\gamma)$, are uniformly integrable with respect to $P_T$ [14]. Roughly, uniform integrability means that the rate of absolute convergence of the following integrals does not depend on $t$,

$$G_{\gamma(t)} = \int_{D(\gamma)} G_{\gamma(t)}(z) P_T(dz) \quad \text{and} \quad H_{\gamma(t)} = \int_{D(\gamma)} H_{\gamma(t)}(z) P_T(dz)$$

This is clear for the integrals $G_{\gamma(t)}$ because $G_{\gamma(t)}(z)$ is bounded in Riemannian norm by $\text{diam} M$, uniformly in $t$ and $z$ (see the proof of the first claim).
Then, consider the integral \( H_x = H_x(0) \), and recall Formulae \([19a]\) and \([19c]\). Each \( z \in D(\gamma) \) can be written under the form \( z = \phi(s, a) \) where \((s, a) \in \mathcal{S} \times \Delta^+\). Accordingly, it follows from \([19c]\) that
\[
\|H_x(z)\|_{F} \leq (\dim M)^{\frac{1}{2}} \max \{1, |\kappa(a) \cot \kappa(a)|\} \tag{20d}
\]
where \(\| \cdot \|_F\) is the Frobenius norm with respect to the Riemannian metric of \(M\), and \(\kappa \in \Delta^+\) is the highest restricted root \([7]\) \((\kappa(a) \geq \lambda(a) \text{ for } \lambda \in \Delta^+, a \in \Delta^+)\).

The required uniform integrability is equivalent to the statement that
\[
\lim_{K \to \infty} \int_{D(\gamma)} \|H_x(z)\|_F \mathbf{1}\{\|H_x(z)\|_F > K\} \ P_T(dz) = 0 \tag{20e}
\]
where the rate of convergence to this limit does not depend on \(x\). But, according to \([20d]\), if \(K > 1\), there exists \(\epsilon > 0\) such that
\[
\{\|H_x(z)\|_F > K\} = \{\kappa(a) > \pi - \epsilon\}
\]
and \(\epsilon \to 0\) as \(K \to \infty\). In this case, the integral in \([20e]\) is less than
\[
(\dim M)^{\frac{1}{2}} \left(\sup_z p_T(z)\right) \int_{D(\gamma)} |\kappa(a) \cot \kappa(a)| \ \mathbf{1}\{\kappa(a) > \pi - \epsilon\} \ vol(dz) \tag{20f}
\]
Now, using the same integral formula as in \([19a]\), this last integral is equal to
\[
\int_{S} \int_{D_+} |\kappa(a) \cot \kappa(a)| \ \mathbf{1}\{\kappa(a) > \pi - \epsilon\} \ J(a) \ da \ \omega(ds) =
\omega(S) \int_{D_+} [\kappa(a) \cot \kappa(a)] \ J(a) \ \mathbf{1}\{\kappa(a) > \pi - \epsilon\} \ da
\]
In view of \([19b]\), since \(\kappa \in \Delta^+\), the function in square brackets is bounded on the closure of \(D_+\). In fact \([7]\), its supremum is \(\kappa^2 = (\kappa, \kappa)\) where \((\cdot, \cdot)\) is the scalar product induced on \(\mathfrak{a}^*\) (the dual space of \(\mathfrak{a}\)) by the Killing form of \(\mathfrak{g}\). Finally, by \([20f]\), the integral in \([20e]\) is less than
\[
(\dim M)^{\frac{1}{2}} \left(\sup_z p_T(z)\right) \omega(S) \kappa^2 \int_{D_+} \mathbf{1}\{\kappa(a) > \pi - \epsilon\} \ da
\]
Since, \(\kappa(a) \in [0, \pi]\) for \(a \in D_+\), this last integral converges to 0 as \(\epsilon \to 0\), at a rate which does not depend on \(x\). This proves the required uniform integrability, so the proof is now complete. \(\blacksquare\)

8 Proof of Proposition 2

Remark: in the statement of Proposition 2 the notation \(\kappa^2\) is used for the maximum sectional curvature of \(M\). In the previous proof of Corollary 1 the same notation \(\kappa^2\) was used for the squared norm of the highest restricted root. This is not an abuse of notation, since the two quantities are in fact equal \([7]\) (see Page 334).
Proof of (i): Let \( x \in B(x^*, \delta) \). By (5) of Corollary 1, \( \nabla^2 E_{r}(x) \) is equal to \( H_x \).

To obtain (7), decompose \( H_x \) into two integrals

\[
H_x = \int_{B(x, r_{cx})} H_x(z) P_T(dz) + \int_{D(x) - B(x, r_{cx})} H_x(z) P_T(dz)
\]  

(21a)

This is possible since \( B(x, r_{cx}) \subset D(x) \), where \( D(x) = M - \text{Cut}(x) \). The first integral in (21a) will be denoted \( I_1 \), and the second integral \( I_2 \).

With regard to \( I_1 \), note the inclusions \( B(x^*, \delta) \subset B(x, 2\delta) \subset B(x, r_{cx}) \), which follow from the triangle inequality. In addition, note that \( H_x(z) \geq 0 \) (in the Loewner order [16]), for \( z \in B(x, r_{cx}) \). Therefore,

\[
I_1 \geq \int_{B(x^*, \delta)} H_x(z) P_T(dz)
\]  

(21b)

However, from (19c) and the definition of \( \kappa \in \Delta_+ \),

\[
H_x(z) \geq \kappa(a) \cot \kappa(a)
\]  

(21c)

for \( z = \phi(s, a) \in D(x) \). Using the Cauchy-Schwarz inequality, \( \kappa(a) \leq \kappa ||a|| \). Moreover, (17c) implies \( ||a|| = d(x, z) \), since \( \text{Ad}(s) \) is an isometry. Accordingly, if \( z \in B(x, 2\delta) \), it follows from (21c)

\[
H_x(z) \geq \kappa(a) \cot \kappa(a) \geq 2\kappa\delta \cot(2\kappa\delta) = C_t(2\delta) > 0
\]  

(21d)

where the last inequality is because \( 2\delta < r_{cx} = \kappa^{-1} \frac{\pi}{2} \). Replacing in (21b) gives

\[
I_1 \geq C_t(2\delta) P_T(B(x^*, \delta)) = C_t(2\delta) [1 - P_T(B^c(x^*, \delta))]
\]  

Finally, (15c) and (15d) imply that \( P_T(B^c(x^*, \delta)) \leq \text{vol}(M) f(T) \), where \( f(T) \) was defined in (6) - Precisely, this follows after replacing \( \rho \) by \( \delta \) in (15c). Thus,

\[
I_1 \geq C_t(2\delta) (1 - \text{vol}(M) f(T))
\]  

(21e)

The proof of (7) will be completed by showing

\[
I_2 \geq -\pi A_M f(T)
\]  

(22a)

To show this, note using (21c) that

\[
I_2 \geq \int_{D(x) - B(x, r_{cx})} \kappa(a) \cot \kappa(a) P_T(dz)
\]  

(22b)

Now, \( \kappa(a) \cot \kappa(a) \) is negative if and only if \( \kappa(a) \geq \frac{\pi}{2} \). However, the set of \( z = \phi(s, a) \) where \( \kappa(a) \geq \frac{\pi}{2} \) is a subset of \( D(x) - B(x, r_{cx}) \). Indeed, \( \kappa(a) \geq \frac{\pi}{2} \) implies \( ||a|| \geq \kappa^{-1} \frac{\pi}{2} = r_{cx} \), by the Cauchy-Schwarz inequality, and this is the same as \( d(x, z) \geq r_{cx} \), since \( ||a|| = d(x, z) \). Therefore, it follows from (22b),

\[
I_2 \geq \int_{D(x)} 1\{\kappa(a) \geq \pi/2\} \kappa(a) \cot \kappa(a) P_T(dz)
\]  

(22c)
Using the same integral formula as in (19a), this last integral is equal to
\[
\int_S \int_{D_+} 1\{k(a) \geq \pi/2\} k(a) \cot(k(a)) \rho_T(\phi(s,a), a) \, da \, \omega(ds) \geq \\
- \int_S \int_{D_+} 1\{k(a) \geq \pi/2\} k(a) \rho_T(\phi(s,a)) \, da \, \omega(ds)
\]
because the product \(\cot(k(a)) J(a) \geq -1\) for all \(a \in D_+\). Using this last inequality, and the fact that \(k(a) \leq \pi\) for all \(a \in D_+\), it follows from (22c),
\[
I_2 \geq -\pi \int_S \int_{D_+} 1\{k(a) \geq \pi/2\} k(a) \rho_T(\phi(s,a)) \, da \, \omega(ds)
\]
Recall that \(\{k(a) \geq \pi/2\} \subset B_c(x, r_{cx})\), as discussed before (22c). In particular, this implies \(\{k(a) \geq \pi/2\} \subset B_c(x^*, \delta)\). However, by (2) and (15d), \(\rho_T(z) \leq f(T)\) for all \(z \in B_c(x^*, \delta)\). Returning to (22d), this gives
\[
I_2 \geq -\pi f(T) \int_S \int_{D_+} da \, \omega(ds)
\]
The double integral on the right-hand side is a constant which depends only on the structure of the symmetric space \(M\). Denoting this constant by \(A_M\) gives the required lower bound (22a), and completes the proof of (7).

Proof of (ii): fix \(\delta < \frac{1}{2} r_{cx}\), and let \(T_\delta\) be given by (9). If \(T \leq T_\delta\), then \(T < T_1\), so the definition of \(T_1\) implies
\[
f(T) < \frac{Ct(2\delta)}{Ct(2\delta) \text{vol}(M) + \pi A_M}
\]
Now, by (7),
\[
\nabla^2 \mathcal{E}_T(x) \geq Ct(2\delta) (1 - \text{vol}(M) f(T)) - \pi A_M f(T)
\]
for all \(x \in B(x^*, \delta)\). However, it is clear from (23a), that the right-hand side of this inequality is strictly positive. It follows that \(\mathcal{E}_T\) is strongly convex on \(B(x^*, \delta)\). Thus, to complete the proof, it only remains to show that any global minimum \(\bar{x}_T\) of \(\mathcal{E}_T\) must belong to \(B(x^*, \delta)\). Indeed, since \(\mathcal{E}_T\) is strongly convex on \(B(x^*, \delta)\), it has only one global minimum in \(B(x^*, \delta)\). Therefore, \(\mathcal{E}_T\) can have only one global minimum \(\bar{x}_T\).

By (i) of Proposition 1 to prove that \(\bar{x}_T \in B(x^*, \delta)\), it is enough to prove
\[
W(P_T, \delta) < \frac{\delta^2}{4 \text{diam}(M)}
\]
However, if \(T \leq T_\delta\), then \(T < T_o\). Therefore, by (ii) of Proposition 1 \(W(P_T, \delta)\) satisfies inequality (4). Furthermore, because \(T < T_1\), it follows from the definition of \(T_1\) that
\[
\sqrt{2\pi} \left(\frac{T}{\mu_{\text{min}}}\right)^{1/2} < \delta^2 \left(\frac{\mu_{\text{min}}}{\mu_{\text{max}}}\right)^{n/2} D_n
\]
or, by replacing the expression of $D_n$, and simplifying
\[ \sqrt{2\pi} \left(\frac{\pi}{2}\right)^{n-1} B_n^{-1} \left(\frac{\mu_{\max}}{\mu_{\min}}\right)^{n/2} \left(\frac{T}{\mu_{\min}}\right)^{1/2} < \frac{\delta^2}{(4 \text{ diam } M)} \] (23d)
Thus, (23c) follows from (4) and (23d). This proves that $\bar{x}_T$ belongs to $B(x^*, \delta)$, and therefore $\bar{x}_T$ is the unique global minimum of $E_T$. But this is equivalent to saying that $\bar{x}_T$ is the unique barycentre of $P_T$.

\[\square\]

9 Proof of Proposition 3

Fix $\delta < \frac{1}{2} r_{cx}$, and let $T_0$ be given by (9). By (ii) of Proposition 2, if $T \leq T_0$, then $E_T$ is strictly convex on $B(x^*, \delta)$, with unique global minimum $\bar{x}_T \in B(x^*, \delta)$. By definition, this unique global minimum $\bar{x}_T$ is the unique barycentre of $P_T$.

Accordingly, to prove that $\bar{x}_T = x^*$, it is enough to prove that $x^*$ is a stationary point of $E_T$. Indeed, as $E_T$ is strictly convex on $B(x^*, \delta)$, it can have only one stationary point in $B(x^*, \delta)$. This stationary point is then identical to $\bar{x}_T$.

The fact that $x^*$ is a stationary point of $E_T$ will follow because $U$ is invariant by geodesic symmetry about $x^*$. This invariance will be seen to imply
\[ ds_{x^*} \cdot G_{x^*} = G_{x^*} \] (24a)
which is equivalent to $G_{x^*} = 0$, since the derivative $ds_{x^*}$ is equal to minus the identity, on the tangent space $T_{x^*} M$ [7]. By (9) of Corollary 1, this shows that $\nabla E_T(x^*) = 0$, so $x^*$ is indeed a stationary point of $E_T$.

To obtain (24a), it is possible to write, from the definition of $G_{x^*}$,
\[ ds_{x^*} \cdot G_{x^*} = ds_{x^*} \cdot \int_{D(x)} G_{x^*}(z) P_T(dz) \] (24b)
where $D(x) = M - \text{Cut}(x)$. From (18), since $s_{x^*}$ is an isometry, and reverses geodesics passing through $x^*$,
\[ ds_{x^*} \cdot G_{x^*}(z) = G_{x^*}(s_{x^*}(z)) \]
Replacing this into (24b), and using $w = s_{x^*}(z)$ as a new variable of integration, it follows that
\[ ds_{x^*} \cdot G_{x^*} = \int_{D(x)} G_{x^*}(w) (P_T \circ s_{x^*})(dw) \] (24c)
because $s_{x^*}^{-1} = s_{x^*}$ and $s_{x^*}$ maps $D(x)$ onto itself. Now, note that $P_T \circ s_{x^*} = P_T$.

This is clear, since from (9),
\[ (P_T \circ s_{x^*})(dw) = (Z(T))^{-1} \exp \left[ \frac{(U \circ s_{x^*})(w)}{T} \right] (\text{vol} \circ s_{x^*})(dw) \]
However, by assumption, $U \circ s_{x^*}(w) = U(w)$. Moreover, since $s_{x^*}$ is an isometry, it preserves Riemannian volume, so $(\text{vol} \circ s_{x^*})(dw) = \text{vol}(dw)$. Thus, (24c) reads
\[ ds_{x^*} \cdot G_{x^*} = \int_{D(x)} G_{x^*}(w) P_T(dw) \]
By definition, the right-hand side is $G_{x^*}$, so (24a) is obtained. \[\square\]
References

1. Karcher, H.: Riemannian centre of mass and mollifier smoothing. Comm. Pure. Appl. Math. 30(5), 509–541 (1977).
2. Afsari, B.: Riemannian $L^p$ center of Mass: existence, uniqueness, and convexity. Proc. Am. Math. Soc. 139(2), 655–673 (2010).
3. Kantorovich, L.V., Akilov, G.P.: Functional Analysis (Second Edition). Pergamon Press, Oxford (1982).
4. Villani, C.: Optimal transport, old and new. 2nd edn. Springer-Verlag, Berlin-Heidelberg (2009).
5. Wong, R.: Asymptotic approximations of Integrals. Society for Industrial and Applied Mathematics (2001).
6. Chavel, I.: Riemannian Geometry, a modern introduction. Cambridge University Press, Cambridge (2006).
7. Helgason, S.: Differential geometry, Lie groups, and symmetric spaces. American Mathematical Society (1978).
8. Roberts, G. O., Rosenthal, J. S.: General state space Markov chains and MCMC algorithms. Probab. Surveys. 1, 20–71 (2004).
9. Beals, R., Wong, R.: Special functions, a graduate text. Cambridge University Press, Cambridge (2010).
10. Mardia, K. V., Jupp, P. E.: Directional statistics. Academic Press, Inc., London (1972).
11. Kobayashi, S., Nomizu, K.: Foundations of differential geometry, Vol. II. Interscience Publishers (1969).
12. Crittenden, R.: Minimum and conjugate points in symmetric spaces. Canad. J. Math. 14, 320–328 (1962).
13. Hurewicz, W., Wallman, H.: Dimension Theory. Princeton University Press (1941).
14. Bogachev, V., I.: Measure theory, Vol. I. Springer-Verlag, Berlin-Heidelberg (2007).
15. Ferreira, R., Xavier, J., Costeira, J. P., Barroso, V.: Newton algorithms for Riemannian distance related problems on connected locally-symmetric manifolds. IEEE J. Sel. Topics Signal Process. 7(4), 634–645 (2013).
16. Zhan, X.: Matrix inequalities (Lecture Notes in Mathematics 1790). Springer-Verlag, Berlin-Heidelberg (2008).