ASYMPTOTIC QUANTIZATION FOR PROBABILITY MEASURES ON RIEMANNIAN MANIFOLDS

Mikaela Iacobelli[1,2]

Abstract. In this paper we study the quantization problem for probability measures on Riemannian manifolds. Under a suitable assumption on the growth at infinity of the measure we find asymptotic estimates for the quantization error, generalizing the results on $\mathbb{R}^d$. Our growth assumption depends on the curvature of the manifold and reduces, in the flat case, to a moment condition. We also build an example showing that our hypothesis is sharp.

Mathematics Subject Classification. 49Q20.

1. INTRODUCTION

The problem of quantization of a d-dimensional probability distribution deals with constructive methods to find atomic probability measures supported on a finite number of points, which best approximate a given diffuse probability measure. The quality of this approximation is usually measured in terms of the Wasserstein metric, and up to now this problem has been studied in the flat case and on compact manifolds.

The quantization problem arises in several contexts and has applications in signal compression, stochastic processes, numerical integration, optimal location of service centers, and kinetic theory. For a complete list of references, we refer to the monograph [5] and references therein. In this paper we study it for probability measures on general Riemannian manifolds.

We now introduce the setting of the problem. Let $(\mathcal{M}, g)$ be a complete Riemannian manifold, and fixed $r \geq 1$, consider μ a probability measure on $\mathcal{M}$. Given N points $x^1, \ldots, x^N \in \mathcal{M}$, one wants to find the best approximation of μ, in the Wasserstein distance W_r, by a convex combination of Dirac masses centered at $x^1, \ldots, x^N$. Hence one minimizes

$$\inf \left\{ W_r \left(\sum_i m_i \delta_{x^i}, \mu \right)^r : m_1, \ldots, m_N \geq 0, \ \sum_i m_i = 1 \right\},$$

with

$$W_r(\nu_1, \nu_2) := \inf \left\{ \left(\int_{\mathcal{M} \times \mathcal{M}} d(x,y)^r d\gamma(x,y) \right)^{1/r} : (\pi_1)_\# \gamma = \nu_1, \ (\pi_2)_\# \gamma = \nu_2 \right\},$$

where γ varies among all probability measures on $\mathcal{M} \times \mathcal{M}$, $\pi_i : \mathcal{M} \times \mathcal{M} \to \mathcal{M}$ $(i = 1, 2)$ denotes the canonical projection onto the ith factor, and $d(x,y)$ denotes the Riemannian distance; see [10] for more details on the Wasserstein distance between probability measures.

The best choice of the masses m_i is explicit and can be expressed in terms of the so-called *Voronoi cells* ([5], Chap. 1.4). Also, as shown for instance in ([5], Chap. 1, Lems. 3.1 and 3.4), the following identity holds:

$$\inf \left\{ W_r \left(\sum_i m_i \delta_{x^i}, \mu \right)^r : m_1, \ldots, m_N \geq 0, \ \sum_i m_i = 1 \right\} = F_{N,r}(x^1, \ldots, x^N),$$

where

$$F_{N,r}(x^1, \ldots, x^N) := \int_{\mathcal{M}} \min_{1 \leq i \leq N} d(x^i, y)^r \, d\mu(y).$$

Hence, the main question becomes: Where are the "optimal points" $(x^1, \ldots, x^N)$ located? To answer to this question, at least in the limit as $N \to \infty$, let us first introduce some definitions.

Definition 1.1. Let μ be a probability a probability measure on $\mathcal{M}$, $N \in \mathbb{N}$ and $r \geq 1$. Then, we define the N *quantization error of order* r, $V_{N,r}(\mu)$ as follows:

$$V_{N,r}(\mu) := \inf_{\alpha \subset \mathcal{M}: |\alpha| \leq N} \int_{\mathcal{M}} \min_{a \in \alpha} d(a, y)^r \, d\mu(y), \tag{1.1}$$

where $|\alpha|$ denotes the cardinality of a set α.

Let us notice that, the functional $F_{N,r}$ being decreasing with respect to the number of points N, an equivalent definition of $V_{N,r}$ is:

$$V_{N,r}(\mu) := \inf_{x^1, \ldots, x^N \in \mathcal{M}} F_{N,r}(x^1, \ldots, x^N).$$

Let us observe that the above definitions make sense for general positive measures with finite mass. In the sequel we will sometimes consider this class of measures in order to avoid renormalization constants.

A quantity that plays an important role in our result is the following:

Definition 1.2. Let dx be the Lebesgue measure and $\chi_{[0,1]^d}$ the characteristic function of the unit cube $[0,1]^d$. We set

$$Q_r \left([0,1]^d \right) := \inf_{N \geq 1} N^{r/d} V_{N,r} \left(\chi_{[0,1]^d} dx \right).$$

As proved in ([5], Thm. 6.2), $Q_r([0,1]^d)$ is a positive constant. The following result describes the asymptotic distribution of the minimizing configuration in $\mathbb{R}^d$, answering to our question in the flat case (see [4,5], Chap. 2, Thms. 6.2 and 7.5):

Theorem 1.3. *Let* $\mu = h \, dx + \mu^s$ *be a probability measure on* $\mathbb{R}^d$, *where* μ^s *denotes the singular part of* μ. *Assume that* μ *satisfies*

$$\int_{\mathbb{R}^d} |x|^{r+\delta} \, d\mu(x) < \infty. \tag{1.2}$$

Then

$$\lim_{N \to \infty} N^{r/d} V_{N,r}(\mu) = Q_r \left([0,1]^d \right) \left(\int_{\mathbb{R}^d} h^{d/(d+r)} \, dx \right)^{(d+r)/d}. \tag{1.3}$$

In addition, if $\mu^s \equiv 0$ *and* $x^1, \ldots, x^N$ *minimize the functional* $F_{N,r} : (\mathbb{R}^d)^N \to \mathbb{R}^+$, *then*

$$\frac{1}{N} \sum_{i=1}^{N} \delta_{x^i} \rightharpoonup \frac{h^{d/d+r}}{\int_{\mathbb{R}^d} h^{d/d+r}(y) dy} \, dx \qquad \text{as } N \to \infty. \tag{1.4}$$

It is worth to mention that the problem of the quantization of measure has been studied also with a Γ-convergence approach in [1–3, 7]. It is reasonable that, using the results in [7], the convergence of the empirical measure to a certain power of the measure h in Theorem 1.3 holds whenever the measure μ has an absolutely continuous part. Nevertheless, we do not investigate this question since this paper is focused on the extension of the first statement in Theorem 1.3 to the case of probability measures on general Riemannian manifolds. Such a statement has been generalized to the case of absolutely continuous probability measures on compact Riemannian manifolds in [6]. The aim of this paper is twofold: we first give an alternative proof of Theorem 1.3 for general probability measures on compact manifolds, and then we extend it to arbitrary measures on non-compact manifolds. As we shall see, passing from the compact to the non-compact setting presents nontrivial difficulties. Indeed, while the compact case relies on a localization argument that allows one to mimic the proof in $\mathbb{R}^d$, the non-compact case requires additional new ideas. In particular one needs to find a suitable analogue of the moment condition (1.2) to control the growth at infinity of our given probability measure. We will prove that the needed growth assumption depends on the curvature of the manifold (and more precisely, on the size of the differential of the exponential map).

To state in detail our main result we need to introduce some notation: given a point $x_0 \in \mathcal{M}$, we can consider polar coordinates (ρ, ϑ) on $T_{x_0}\mathcal{M} \simeq \mathbb{R}^d$ induced by the constant metric g_{x_0}, where ϑ denotes a vector on the unit sphere $\mathbb{S}^{d-1}$ and ρ is the the value of the norm in the metric g_{x_0}. Then, we can define the following quantity that measures the size of the differential of the exponential map when restricted to a sphere $\mathbb{S}^{d-1}_\rho \subset T_{x_0}\mathcal{M}$ of radius ρ:

$$A_{x_0}(\rho) := \sup_{v \in \mathbb{S}^{d-1}_\rho, \, w \in T_v\mathbb{S}^{d-1}_\rho, \, |w|_{x_0}=\rho} \left| d_v \exp_{x_0}[w] \right|_{\exp_{x_0}(v)}. \tag{1.5}$$

To prove asymptotic quantization, we shall impose an analogue of (1.2) which involves the above quantity.

Theorem 1.4. *Let $(\mathcal{M}, g)$ be a complete Riemannian manifold without boundary, and let $\mu = h \, \mathrm{dvol} + \mu^s$ be a probability measure on $\mathcal{M}$. Assume there exist a point $x_0 \in \mathcal{M}$ and $\delta > 0$ such that*

$$\int_{\mathcal{M}} d(x, x_0)^{r+\delta} \, \mathrm{d}\mu(x) + \int_{\mathcal{M}} A_{x_0}\big(d(x, x_0)\big)^r \, \mathrm{d}\mu(x) < \infty. \tag{1.6}$$

Then (1.3) holds.

Once this theorem is obtained, by the very same argument as in ([5], Proof of Thm. 7.5) one gets the following:

Corollary 1.5. *Let $(\mathcal{M}, g)$ be a complete Riemannian manifold without boundary, $\mu = h \, \mathrm{dvol}$ an absolutely continuous probability measure on $\mathcal{M}$ and let $x^1, \ldots, x^N$ minimize the functional $F_{N,r} : \mathcal{M}^{\otimes N} \to \mathbb{R}^+$. Assume there exist a point $x_0 \in \mathcal{M}$ and $\delta > 0$ for which (1.6) is satisfied. Then (1.4) holds.*

Notice that the quantity A_{x_0} is related to the curvature of $\mathcal{M}$, being linked to the size of the Jacobi fields (see for instance [8], Chap. 10). In particular, if $\mathcal{M} = \mathbb{H}^d$ is the hyperbolic space then $A_{x_0}(\rho) = \sinh \rho$, while on $\mathbb{R}^d$ we have $A_{x_0}(\rho) = \rho$. Hence the above condition on $\mathbb{H}^d$ reads as

$$\left(1 + \int_{\mathbb{H}^d} d(x, x_0)^{r+\delta} \, \mathrm{d}\mu(x) + \int_{\mathbb{H}^d} \sinh\big(d(x, x_0)\big)^r \, \mathrm{d}\mu(x)\right) \approx \int_{\mathbb{H}^d} \mathrm{e}^{r \, d(x, x_0)} \, \mathrm{d}\mu(x),$$

and on $\mathbb{R}^d$ as

$$\left(1 + \int_{\mathbb{R}^d} d(x, x_0)^{r+\delta} \, \mathrm{d}\mu(x) + \int_{\mathbb{R}^d} d(x, x_0)^r \, \mathrm{d}\mu(x)\right) \approx \int_{\mathbb{R}^d} d(x, x_0)^{r+\delta} \, \mathrm{d}\mu(x).$$

Hence (3.2) holds on $\mathbb{H}^d$ for any probability measure μ satisfying

$$\int_{\mathbb{H}^d} \mathrm{e}^{r \, d(x, x_0)} \, \mathrm{d}\mu(x) < \infty$$

for some $x_0 \in \mathbb{H}^d$, while on $\mathbb{R}^d$ we only need the finiteness of some $(r + \delta)$-moments of μ, therefore recovering the assumption in Theorem 1.3. More in general, thanks to Rauch Comparison Theorem ([8], Thm. 11.9), the size of the Jacobi fields on a manifold $\mathcal{M}$ with sectional curvature bounded from below by $-K$ ($K \geq 0$) is controlled by the Jacobi fields on the hyperbolic space with sectional curvature $-K$. Hence in this case

$$A_{x_0}(\rho) \leq \sinh(Kr) \approx e^{Kr},$$

and Theorem 1.4 yields the following:

Corollary 1.6. *Let $(\mathcal{M}, g)$ be a complete Riemannian manifold without boundary, and let $\mu = h \, \mathrm{dvol} + \mu^s$ be a probability measure on $\mathcal{M}$. Assume that the sectional curvature of $\mathcal{M}$ is bounded from below by $-K$ for some $K \geq 0$, and that there exist a point $x_0 \in \mathcal{M}$ and $\delta > 0$ such that*

$$\int_{\mathcal{M}} d(x, x_0)^{r+\delta} \, \mathrm{d}\mu(x) + \int_{\mathcal{M}} e^{Kr \, d(x,x_0)} \, \mathrm{d}\mu(x) < \infty.$$

Then (1.3) holds. In addition, if $\mu^s \equiv 0$ and $x^1, \ldots, x^N$ minimize the functional $F_{N,r} : (\mathbb{R}^d)^N \to \mathbb{R}^+$, then (1.4) holds.

Let us observe that, by the argument in the proof of Theorem 1.7 in the Corollary above the moment condition does not depend on the choice of the point x_0.

Finally, we show that the moment condition (1.2) required on $\mathbb{R}^d$ is not sufficient to ensure the validity of the result on $\mathbb{H}^d$. Indeed we can provide the following counter example on $\mathbb{H}^2$.

Theorem 1.7. *There exists a measure μ on $\mathbb{H}^2$ such that*

$$\int_{\mathbb{H}^2} d(x, x_0)^p \, \mathrm{d}\mu < \infty \qquad \forall \, p > 0, \, \forall \, x_0 \in \mathbb{H}^2,$$

but

$$N^{r/2} V_{N,r}(\mu) \to \infty \qquad as \ N \to \infty.$$

The paper is structured as follows: first, in Section 2 we prove Theorem 1.4 for compactly supported probability measures. Then, in Section 3 we deal with the non-compact case concluding the proof of Theorem 1.4. Finally, in Section 4 we prove Theorem 1.7.

2. Proof of Theorem 1.4: The compact case

This section is concerned with the study of asymptotic quantization for probability distributions on compact Riemannian manifolds as the number N of points tends to infinity. Although the problem depends *a priori* on the global geometry of the manifold (since $V_{N,r}$ involves the Riemannian distance), we shall now show how a localization argument allows us to prove the result.

2.1. Localization argument

Let $(\mathcal{M}, g)$ be a complete Riemannian manifold without boundary and let μ be a probability measure on $\mathcal{M}$. We consider $\{\mathcal{U}_i, \varphi_i\}_{i \in I}$ an atlas covering $\mathcal{M}$, and $\varphi_i : \mathcal{W}_i \to \mathbb{R}^d$ smooth charts, where $\mathcal{W}_i \supset\supset \mathcal{U}_i$ for all $i \in I$. As we shall see, in order to be able to split our measure as a sum of measures supported on smaller sets, we want to avoid the mass to concentrate on the boundary of the sets $\mathcal{U}_i$. Hence, up to slightly changing the sets $\mathcal{U}_i$, we may assume that

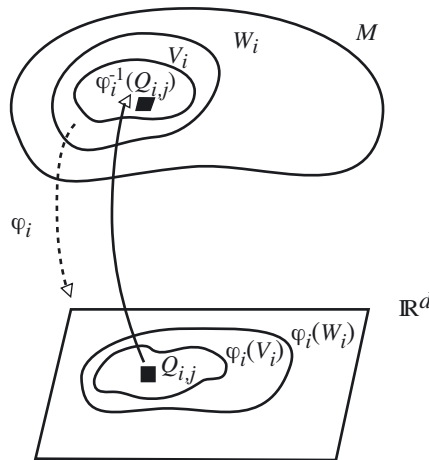$$\mu(\partial \mathcal{U}_i) = 0 \qquad \forall \, i \in I. \tag{2.1}$$

FIGURE 1. We use the map $\varphi_i^{-1} : \varphi_i(\mathcal{W}_i) \subset \mathbb{R}^d \to \mathcal{W}_i \subset \mathcal{M}$ to send the partition in cubes $Q_{i,j}$ of $\varphi_i(\mathcal{V}_i)$ on $\mathcal{M}$.

We want to cover $\mathcal{M}$ with an atlas of disjoint sets, up to sets of μ-measure zero. To do that we define

$$\mathcal{V}_i := \mathcal{U}_i \setminus \left(\bigcup_{j=1}^{i-1} \mathcal{U}_j \right).$$

Notice that we still have $\mathcal{V}_i \subset\subset \mathcal{W}_i$.

Given an open subset of $\mathbb{R}^d$, we can cover it with a countable partition of half-open disjoint cubes such that the maximum length of the edges is a given number δ. We now apply this observation to each open subset $\varphi_i(\overset{\circ}{\mathcal{V}_i}) \subset \mathbb{R}^d$ and we cover it with a family $\mathcal{G}_i$ of half-open cubes $\{Q_{i,j}\}_{j\in\mathbb{N}}$ with edges of length $\ell_j \leq \delta$.

We notice that the "cubes" $\varphi_i^{-1}(Q_{i,j}) \subset \mathcal{M}$ are disjoint and

$$\bigcup_{i\in I} \bigcup_{Q_{i,j}\in\mathcal{G}_i} Q_{i,j} = \mathcal{M} \setminus \left(\bigcup_i \partial\mathcal{U}_i \right).$$

Since by (2.1) the set $\cup_i \partial\mathcal{U}_i$ has zero μ-measure, we can decompose the measure μ as

$$\mu = \sum_{i\in I} \mu \mathbf{1}_{\mathcal{V}_i} = \sum_{i\in I} \sum_{Q_{i,j}\in\mathcal{G}_i} \mu \mathbf{1}_{\mathcal{V}_i \cap \varphi_i^{-1}(Q_{i,j})}.$$

We now set

$$\alpha_{ij} := \int_{\mathcal{V}_i \cap \varphi_i^{-1}(Q_{i,j})} \mathrm{d}\mu, \qquad \mu_{ij} := \frac{\mu \mathbf{1}_{\mathcal{V}_i \cap \varphi_i^{-1}(Q_{i,j})}}{\alpha_{ij}},$$

so that

$$\mu = \sum_{ij} \alpha_{ij} \, \mu_{ij}, \qquad \int_{\mathcal{M}} \mathrm{d}\mu_{ij} = 1, \qquad \mathrm{supp}(\mu_{ij}) \subset \mathcal{V}_i \cap \varphi_i^{-1}(Q_{i,j}),$$

where, to simplify the notation, in the above formula the indices i, j implicitly run over $i \in I, Q_{i,j} \in \mathcal{G}_i$. We will keep using this convention also later on.

The idea is now the following: by choosing δ small enough, each measure μ_{ij} is supported on a very small set where the metric is essentially constant and allows us to reduce ourselves to the flat case and apply Theorem 1.3 to each of these measures. A "gluing argument" then gives the result when $\mu = \sum_{ij} \alpha_{ij}\mu_{ij}$ is compactly supported, $\alpha_{ij} \neq 0$ for at most finitely many indices, and μ_{ij} has constant density on $\varphi_i^{-1}(Q_{i,j})$. Finally, an approximation argument yields the result for general compactly supported measures.

2.2. The local quantization error

The goal of this section is to understand the behavior of $V_{N,r}(\mu)$ when

$$\mu = \lambda \mathbf{1}_{\varphi^{-1}(Q)} \, \mathrm{dvol}, \tag{2.2}$$

where $\lambda := \frac{1}{\mathrm{vol}(\varphi^{-1}(Q))}$ (so that μ has mass 1), Q is a δ-cube in $\mathbb{R}^d$, $\varphi : \mathcal{W} \to \mathbb{R}^d$ is a diffeomorphism defined on a neighborhood $\mathcal{W} \subset \mathcal{M}$ of $\varphi^{-1}(Q)$.

We observe that, in the computation of $V_{N,r}(\mu)$, if the size of the cube is sufficiently small then we can assume that all the points belong to a $K\delta$-neighborhood of $\varphi^{-1}(Q)$, with K a large universal constant, that we denote by $\mathcal{Z}_{K\delta}$. Indeed, if $\mathrm{dist}(b, \varphi^{-1}(Q)) > K\delta$ then $\mathrm{dist}(x, b) > \mathrm{dist}(x, y)$ for all $x, y \in \varphi^{-1}(Q)$, which implies that, in the definition of $V_{N,r}(\mu)$, it is better to substitute b with an arbitrary point inside $\varphi^{-1}(Q)$.

Hence, denoting by β a family of N points inside a $\mathcal{Z}_{K\delta}$, and by α a family of N points inside $\varphi(\mathcal{Z}_{K\delta})$, we have

$$V_{N,r}(\mu) = \lambda \inf_{\beta} \int_Q \min_{a \in \alpha} d\big(\varphi^{-1}(x), \varphi^{-1}(a)\big)^r \sqrt{\det g_{k\ell}(x)} \, \mathrm{d}x. \tag{2.3}$$

We now begin by showing that $d\big(\varphi^{-1}(x), \varphi^{-1}(a)\big)$ can be approximated with a constant metric. We use the notation $g_{k\ell}$ to denote the metric in the chart, that is

$$\sum_{k\ell} g_{k\ell}(x) v^k v^\ell := g_{\varphi^{-1}(x)}\Big(d\varphi^{-1}(x)[v], d\varphi^{-1}(x)[v]\Big), \qquad \forall x \in \varphi(\mathcal{W}), \, v \in \mathbb{R}^d. \tag{2.4}$$

Lemma 2.1. *Let p be the center of the cube Q and let A be the matrix with entries $A_{k\ell} := g_{k\ell}(p)$. There exists a universal constant $\hat{C}$ such that, for all $x \in Q$ and $a \in \varphi(\mathcal{Z}_{K\delta})$, it holds*

$$(1 - \hat{C}\delta) \, \langle A(x - a), x - a \rangle \leq d\big(\varphi^{-1}(x), \varphi^{-1}(a)\big)^2 \leq (1 + \hat{C}\delta) \, \langle A(x - a), x - a \rangle.$$

Proof. We begin by recalling that [3]

$$d\big(\varphi^{-1}(x), \varphi^{-1}(a)\big)^2 = \inf_{\substack{\gamma(0) = \varphi^{-1}(x), \\ \gamma(1) = \varphi^{-1}(a)}} \int_0^1 g_{\gamma(t)}\big(\dot{\gamma}(t), \dot{\gamma}(t)\big) \, \mathrm{d}t.$$

Let $\bar{\gamma} : [0, 1] \to \mathcal{M}$ denote a minimizing geodesic [4]. Then the speed of $\bar{\gamma}$ is constant and equal to the distance between the two points, that is

$$\|\dot{\bar{\gamma}}(t)\|_g := \sqrt{g_{\bar{\gamma}(t)}\big(\dot{\bar{\gamma}}(t), \dot{\bar{\gamma}}(t)\big)} = d\big(\varphi^{-1}(x), \varphi^{-1}(a)\big). \tag{2.5}$$

We can bound from above $d\big(\varphi^{-1}(x), \varphi^{-1}(a)\big)$ by choosing a curve γ obtained by the image via φ^{-1} of a segment:

$$d\big(\varphi^{-1}(x), \varphi^{-1}(a)\big)^2 \leq \int_0^1 g_{\sigma(t)}\big(\dot{\sigma}(t), \dot{\sigma}(t)\big) \, \mathrm{d}t, \qquad \sigma(t) := \varphi^{-1}\big((1 - t)x + ta\big).$$

---

[3] Recall that there are two equivalent definition of the distance between two points:

$$d(x, y) = \inf_{\substack{\gamma(0) = x, \\ \gamma(1) = y}} \int_0^1 \sqrt{g_{\gamma(t)}\big(\dot{\gamma}(t), \dot{\gamma}(t)\big)} \, \mathrm{d}t = \inf_{\substack{\gamma(0) = x, \\ \gamma(1) = y}} \sqrt{\int_0^1 g_{\gamma(t)}\big(\dot{\gamma}(t), \dot{\gamma}(t)\big) \, \mathrm{d}t}.$$

In this paper we will make use of both definitions.

[4] Notice that the hypothesis of completeness on $\mathcal{M}$ ensures the existence of minimizing geodesics.

Observe that this formula makes sense since $(1 - t)x + ta \in \varphi(\mathcal{W})$ provided δ is sufficiently small.

Since

$$\sqrt{\int_0^1 g_{\sigma(t)}\big(\dot{\sigma}(t), \dot{\sigma}(t)\big)\, \mathrm{d}t} \leq C'|x - a| \tag{2.6}$$

for some universal constant C', combining (2.5) and (2.6) we deduce that

$$\|\dot{\bar{\gamma}}(t)\|_g \leq C'|x - a| \leq C''\delta \qquad \forall\, t \in [0, 1].$$

In particular

$$d(\bar{\gamma}(t), x) = d(\bar{\gamma}(t), \bar{\gamma}(0)) \leq C''\delta \qquad \text{for all } t \in [0, 1],$$

which implies that $\bar{\gamma}$ belongs to the $K\delta$-neighborhood of $\varphi^{-1}(Q)$, that is $\bar{\gamma} \subset \mathcal{Z}_{C''\delta}$.

Thanks to this fact we deduce that in the definition of the distance we can restrict to curves contained inside $\mathcal{Z}_{C''\delta}$. Since $\mathcal{Z}_{C''\delta} \subset \mathcal{W}$ for δ sufficiently small, all such curves can be seen as the image through φ^{-1} of a curve contained inside $\varphi(\mathcal{W}) \subset \mathbb{R}^d$. Notice that, by (2.4), if $\sigma(t) := \varphi(\gamma(t)) = \big(\sigma^1(t), \ldots, \sigma^n(t)\big) \in \mathbb{R}^d$ then

$$g_{\gamma(t)}\big(\dot{\gamma}(t), \dot{\gamma}(t)\big) = \sum_{k\ell} g_{k\ell}(\sigma(t))\dot{\sigma}^k(t)\dot{\sigma}^\ell(t),$$

therefore

$$d\big(\varphi^{-1}(x), \varphi^{-1}(a)\big)^2 \leq \big(1 + \hat{C}\delta\big) \inf_{\substack{\sigma(0)=x,\, \sigma(1)=a, \\ \sigma \subset \varphi(\mathcal{Z}_{C''\delta})}} \int_0^1 \sum_{k\ell} A_{k\ell}\dot{\sigma}^k(t)\dot{\sigma}^\ell(t)\, \mathrm{d}t,$$

where we used that, by the Lipschitz regularity of the metric and the fact that $g_{k\ell}$ is positive definite, we have

$$\sum_{k\ell} g_{k\ell}(z)v^k v^\ell \leq (1 + \hat{C}\delta) \sum_{k\ell} A_{k\ell}v^k v^\ell \qquad \forall\, z \in \varphi(\mathcal{Z}_{C''\delta}), \qquad \forall\, v \in \mathbb{R}^d.$$

Using now that the minimizer for the problem

$$\inf_{\sigma(0)=x,\, \sigma(1)=a} \int_0^1 \sum_{k\ell} A_{k\ell}\dot{\sigma}^k(t)\dot{\sigma}^\ell(t)\, \mathrm{d}t$$

is given by a straight segment, and since this segment is contained inside $\varphi(\mathcal{Z}_{C''\delta})$, we obtain

$$\inf_{\substack{\sigma(0)=x,\, \sigma(1)=a, \\ \sigma \subset \varphi(\mathcal{Z}_{C''\delta})}} \int_0^1 \sum_{k\ell} A_{k\ell}\dot{\sigma}^k(t)\dot{\sigma}^\ell(t)\, \mathrm{d}t = \langle A(x - a), x - a \rangle,$$

which proves

$$d\big(\varphi^{-1}(x), \varphi^{-1}(a)\big)^2 \leq \big(1 + \hat{C}\delta\big)\langle A(x - a), x - a \rangle.$$

The lower bound is proved analogously using that

$$\sum_{k\ell} g_{k\ell}(z)v^k v^\ell \geq (1 - \hat{C}\delta) \sum_{k\ell} A_{k\ell}v^k v^\ell \qquad \forall\, z \in \varphi(\mathcal{Z}_{C''\delta}), \qquad \forall\, v \in \mathbb{R}^d,$$

concluding the proof. $\qquad\square$

Applying now this lemma, we can estimate $V_{N,r}(\mu)$ both from above and below. Since the argument in both cases is completely analogous, we just prove the upper bound.

Notice that, by the Lipschitz regularity of the metric and the fact that $\det g_{k\ell}$ is bounded away from zero, we have

$$\sqrt{\det g_{k\ell}(x)} \le (1 + C\delta)\sqrt{\det g_{k\ell}(p)} = (1 + C\delta)\sqrt{\det A} \qquad \forall\, x \in Q.$$

Combining this estimate with (2.3) and Lemma 2.1, we get

$$V_{N,r}(\mu) \le (1 + C'\delta)\,\lambda \inf_\alpha \int_Q \min_{a \in \alpha} \langle A(x - a), x - a \rangle^{r/2} \sqrt{\det A}\,\mathrm{d}x$$

$$= (1 + C'\delta)\,\lambda \inf_\alpha \int_{A^{1/2}(Q)} \min_{a \in \alpha} |z - a|^r\,\mathrm{d}z,$$

where $|\cdot|$ denotes the Euclidean norm.

We now apply Theorem 1.3 to the probability measure $\frac{1}{|A^{1/2}(Q)|}\mathbf{1}_{A^{1/2}(Q)}\,\mathrm{d}z$ to get

$$\limsup_{N \to \infty} N^{r/d} V_{N,r}(\mu) \le (1 + C'\delta)\,\lambda\, Q_r\big([0,1]^d\big)\,|A^{1/2}(Q)|^{(d+r)/d}.$$

Observing that

$$|A^{1/2}(Q)| = \int_Q \sqrt{\det A}\,\mathrm{d}x \le (1 + C\delta) \int_Q \sqrt{\det g_{k\ell}(x)}\,\mathrm{d}x = (1 + C\delta)\,\frac{1}{\lambda},$$

we conclude that

$$\limsup_{N \to \infty} N^{r/d} V_{N,r}(\mu) \le \big(1 + \bar{C}\,\delta\big)\, Q_r\big([0,1]^d\big)\,\mathrm{vol}(\varphi^{-1}(Q))^{r/d}. \tag{2.7}$$

Arguing similarly for the lower bound, we also have

$$\liminf_{N \to \infty} N^{r/d} V_{N,r}(\mu) \ge \big(1 - \bar{C}\delta\big)\, Q_r\big([0,1]^d\big)\,\mathrm{vol}(\varphi^{-1}(Q))^{r/d}, \tag{2.8}$$

which concludes the local analysis of the quantization error for μ as in (2.2).

2.3. Upper bound for $V_{N,r}$

We consider a compactly supported measure $\mu = \sum_{ij} \alpha_{ij}\mu_{ij}$ where $\alpha_{ij} \ne 0$ for at most finitely many indices, and μ_{ij} is of the form $\lambda_{ij}\mathbf{1}_{\varphi_i^{-1}(Q_{i,j})}d\,\mathrm{vol}$ with

$$\varphi_i^{-1}(Q_{i,j}) \cap \varphi_{i'}^{-1}(Q_{j'}) = \emptyset, \qquad \forall\, i, i', \quad \forall\, j \ne j',$$

and $\lambda_{ij} := \frac{1}{\mathrm{vol}(\varphi_i^{-1}(Q_{i,j}))}$ (so that each measure μ_{ij} has mass 1).

To estimate $V_{N,r}(\mu)$ we first observe that, for any choice of N_{ij} such that $\sum_{ij} N_{ij} \le N$ the following inequality holds:

$$V_{N,r}(\mu) \le \sum_{ij} \alpha_{ij}\, V_{N_{ij},r}(\mu_{ij}).$$

We want to chose the N_{ij} in an optimal way. As it will be clear from the estimates below, the best choice is to set [5]

$$t_{ij} := \frac{\big(\alpha_{ij}\,\mathrm{vol}(\varphi_i^{-1}(Q_{i,j}))^{r/d}\big)^{d/(d+r)}}{\sum_{k\ell}\big(\alpha_{k\ell}\,\mathrm{vol}(\varphi_k^{-1}(Q_\ell))^{r/d}\big)^{d/(d+r)}},$$

---

[5]Notice that, if we were on $\mathbb{R}^d$ and φ_i were just the identity map, then the formula for t_{ij} simplifies to

$$t_{ij} = \frac{(\alpha_{ij})^{d/(d+r)}}{\sum_{k\ell}(\alpha_{k\ell})^{d/(d+r)}},$$

that is the exact same formula used in ([5], Proof of Thm. 6.2, Step 2).

and define
$$N_{ij} := [t_{ij}N].$$

Notice that N_{ij} satisfy $\sum_{ij} N_{ij} \leq N$ and $\sum_{ij} \frac{N_{ij}}{N} \to 1$ as $N \to \infty$. We observe that each measure μ_{ij} is a probability measure supported in only one "cube" with constant density. Hence we can apply the local quantization error (2.7) to each measure μ_{ij} to get that

$$\limsup_{N_{ij}\to\infty} N_{ij}^{r/d} V_{N_{ij},r}(\mu_{ij}) \leq (1 + \bar{C}\delta) \, Q_r\big([0,1]^d\big) \, \mathrm{vol}(\varphi_i^{-1}(Q_{i,j}))^{r/d}.$$

Recalling our choice of N_{ij},

$$\limsup_{N\to\infty} N^{r/d} V_{N,r}(\mu) \leq (1 + \bar{C}\delta) \, Q_r\big([0,1]^d\big) \sum_{ij} \alpha_{ij} t_{ij}^{-r/d} \, \mathrm{vol}(\varphi_i^{-1}(Q_{i,j}))^{r/d},$$

and observing that

$$\sum_{ij} \alpha_{ij} \, t_{ij}^{-r/d} \, \mathrm{vol}(\varphi_i^{-1}(Q_{i,j}))^{r/d} = \left(\int_{\mathcal{M}} h^{d/(d+r)} \, \mathrm{dvol}\right)^{(d+r)/d},$$

we get

$$\limsup_{N\to\infty} N^{r/d} V_{N,r}(\mu) \leq (1 + \bar{C}\delta) \, Q_r\big([0,1]^d\big) \left(\int_{\mathcal{M}} h^{d/(d+r)} \, \mathrm{dvol}\right)^{(d+r)/d}.$$

2.4. Lower bound for $V_{N,r}$

We consider again a compactly supported measure $\mu = \sum_{ij} \alpha_{ij} \mu_{ij}$ where $\alpha_{ij} \neq 0$ for at most finitely many indices, and μ_{ij} is of the form $\lambda_{ij} \mathbf{1}_{\varphi_i^{-1}(Q_{i,j})} d\,\mathrm{vol}$ with

$$\varphi_i^{-1}(Q_{i,j}) \cap \varphi_{i'}^{-1}(Q_{j'}) = \emptyset, \qquad \forall \, i, i', \quad \forall \, j \neq j',$$

and $\lambda_{ij} := \frac{1}{\mathrm{vol}(\varphi_i^{-1}(Q_{i,j}))}$ (so that $\int_{\mathcal{M}} \mu_{ij} = 1$). Fix $\varepsilon > 0$ with $\varepsilon \ll \delta$, and consider the cubes $Q_{j,\varepsilon}$ given by

$$Q_{j,\varepsilon} := \{y \in Q_{i,j} : \mathrm{dist}(y, \partial Q_{i,j}) > \varepsilon\}.$$

Also, consider a set γ_{ij} consisting of K_{ij} points such that

$$\min_{a \in \gamma_{ij}} d(x,a) \leq \inf_{z \in \mathcal{M} \setminus \varphi_i^{-1}(Q_{i,j})} d(x,z) \qquad \forall \, x \in \varphi_i^{-1}(Q_{j,\varepsilon}) \text{ s.t. } \varphi_i^{-1}(Q_{i,j}) \cap \mathrm{supp}(\mu) \neq \emptyset.$$

Notice that the property of μ being compactly supported ensures that

$$K := \max\left\{ K_{ij} : \varphi_i^{-1}(Q_{i,j}) \cap \mathrm{supp}(\mu) \neq \emptyset \right\} < \infty.$$

Then, if β is a set of N points optimal for $V_{N,r}(\mu)$ and $\beta_{ij} := \beta \cap \varphi_i^{-1}(Q_j)$,

$$V_{N,r}(\mu) \geq \sum_{ij} \int_{\varphi_i^{-1}(Q_{j,\varepsilon})} \min_{b \in \beta \cup \gamma_{ij}} d(x,b)^r \, d\mu \geq \sum_{ij} \alpha_{ij}^\varepsilon \, V_{N_{ij}+K_{ij},r}(\mu_{ij}^\varepsilon), \tag{2.9}$$

where

$$\alpha_{ij}^\varepsilon := \int_{\mathcal{V}_i \cap \varphi_i^{-1}(Q_{j,\varepsilon})} d\mu, \qquad \mu_{ij}^\varepsilon := \frac{\mathbf{1}_{\mathcal{V}_i \cap \varphi_i^{-1}(Q_{j,\varepsilon})} \, \mathrm{dvol}}{\mathrm{vol}\big(\varphi_i^{-1}(Q_{j,\varepsilon})\big)}, \qquad N_{ij} := \#\beta_{ij}.$$

We notice that $\alpha_{ij}^\varepsilon \to \alpha_{ij}$ as $\varepsilon \to 0$.

Let $L := \liminf_{N\to\infty} N^{r/d} V_{N,r}(\mu)$. Notice that $L < \infty$ by the upper bound proved in the previous step. Choose a subsequence $N(k)$ such that

$$N(k)^{r/d} V_{N(k),r}(\mu) \to L \qquad \text{as } k \to \infty$$

and, for all i, j,

$$\frac{N_{ij}(k)}{N(k)} \to v_{ij} \in [0,1] \qquad \text{as } k \to \infty.$$

Since $\sum_{ij} N_{ij}(k) = N(k)$ we have $\sum_{ij} v_{ij} = 1$.

Moreover $N_{ij}(k) \to \infty$ for every i, j. Indeed, if not, it would contradict the finiteness of L, see also ([5], Thm. 6.2, Step 2).

Thanks to this fact, we can now apply the local quantization error (2.8) to deduce that

$$\liminf_{k\to\infty} N_{ij}(k)^{r/d} V_{N_{ij}(k)+K_{ij}(k),r}(\mu_{ij}^{\varepsilon}) \geq (1 - \bar{C}\delta)\, Q_r\big([0,1]^d\big)\, \text{vol}(\varphi_i^{-1}(Q_{j,\varepsilon}))^{r/d},$$

which implies that (recalling (2.9))

$$L \geq (1 - \bar{C}\delta)\, Q_r\big([0,1]^d\big) \sum_{ij} \alpha_{ij}^{\varepsilon}\, v_{ij}^{-r/d} \text{vol}(\varphi_i^{-1}(Q_{j,\varepsilon}))^{r/d}.$$

Letting $\varepsilon \to 0$ and applying ([5], Lem. 6.8) we finally obtain

$$L \geq (1 - \bar{C}\delta)\, Q_r\big([0,1]^d\big) \left(\int_{\mathcal{M}} h^{d/(d+r)}\, \text{dvol}\right)^{(d+r)/d}.$$

2.5. Approximation argument: general compactly supported measures

In the previous two sections we proved that if μ is compactly supported and it is of the form

$$\mu = \sum_{ij} \alpha_{ij} \frac{\mathbf{1}_{\varphi_i^{-1}(Q_{i,j})}}{\text{vol}\big(\varphi_i^{-1}(Q_{i,j})\big)}\, \text{d vol}$$

where $Q_{i,j}$ is a family of cubes in $\mathbb{R}^d$ of size at most δ and $\alpha_{ij} \neq 0$ for finitely many indices, then

$$(1 - \bar{C}\delta)\, Q_r\big([0,1]^d\big) \left(\int_{\mathcal{M}} h^{d/(d+r)}\, \text{dvol}\right)^{(d+r)/d} \leq \liminf_{N\to\infty} N^{r/d} V_{N,r}(\mu)$$

$$\leq \limsup_{N\to\infty} N^{r/d} V_{N,r}(\mu) \leq (1 + \bar{C}\delta)\, Q_r\big([0,1]^d\big) \left(\int_{\mathcal{M}} h^{d/(d+r)}\, \text{dvol}\right)^{(d+r)/d}. \quad (2.10)$$

To prove the quantization result for general measures with compact support, we need three approximation steps.

First, given a compactly supported measure $\mu = h\,\text{dvol}$, we can approximate it with a sequence $\{\mu_k\}_{k\in\mathbb{N}}$ of measures as above where the size of the cubes $\delta_k \to 0$, and this allows us to prove that

$$N^{r/d} V_{N,r}(\mu) \to Q_r\big([0,1]^d\big) \left(\int_{\mathcal{M}} h^{d/(d+r)}\, \text{dvol}\right)^{(d+r)/d} \quad (2.11)$$

for any compactly supported measure of the form $h\,\text{dvol}$. Then, given a singular measure with compact support $\mu = \mu^s$, we show that

$$N^{r/d} V_{N,r}(\mu) \to 0.$$

Finally, given an arbitrary measure with compact support $\mu = h\,\text{dvol} + \mu^s$, we show that (2.11) still holds true.

The proofs of these three steps is performed in detail in ([5], Thm. 6.2, Step 3, Step 4, Step 5) for the case of $\mathbb{R}^d$. As it can be easily checked, such a proof applies immediately also in our case, so we will not repeat here for the sake of conciseness.

This concludes the Proof of Theorem 1.4 when μ is compactly supported (in particular, whenever $\mathcal{M}$ is compact).

3. PROOF OF THEOREM 1.4: THE NON-COMPACT CASE

The aim of this section is to study the case of non-compactly supported measures. As we shall see, this situation is very different with respect to the flat case as we need to deal with the growth at infinity of μ.

To state our result, let us recall the notation we already presented in the introduction: given a point $x_0 \in \mathcal{M}$, we can consider polar coordinates (ρ, ϑ) on $T_{x_0}\mathcal{M} \simeq \mathbb{R}^d$ induced by the constant metric g_{x_0}, where ϑ denotes a vector on the unit sphere $\mathbb{S}^{d-1}$ and ρ is the value of the norm in the metric g_{x_0}. Then we define the quantity $A_{x_0}(\rho)$ as in (1.5). Our goal is to prove the following result which implies Theorem 1.4.

Theorem 3.1. *Let $(\mathcal{M}, g)$ be a complete Riemannian manifold, and let $\mu = h\,\mathrm{dvol} + \mu^s$ be a probability measure on $\mathcal{M}$. Then, for any $x_0 \in \mathcal{M}$ and $\delta > 0$, there exists a constant $C = C(\delta) > 0$ such that*

$$N^r V_{N^d,r}(\mu) \leq C\left(1 + \int_{\mathcal{M}} d(x, x_0)^{r+\delta}\,\mathrm{d}\mu(x) + \int_{\mathcal{M}} A_{x_0}\big(d(x, x_0)\big)^r\,\mathrm{d}\mu(x)\right). \tag{3.1}$$

If there exists a point $x_0 \in \mathcal{M}$ and $\delta > 0$ for which the right hand side is finite, we have

$$N^{r/d} V_{N,r}(\mu) \to Q_r\big([0,1]^d\big) \left(\int_{\mathcal{M}} h^{d/(d+r)}\,\mathrm{dvol}\right)^{(d+r)/d}. \tag{3.2}$$

3.1. Proof of Theorem 3.1

We begin by the proof of (3.1). For this we will need the following result, whose proof is contained in ([5], Lem. 6.6).

Lemma 3.2. *Let ν be a probability measure on $\mathbb{R}$. Then*

$$N^r V_{N,r}(\nu) \leq C\left(1 + \int_{\mathbb{R}} |t|^{r+\delta}\,\mathrm{d}\nu(t)\right). \tag{3.3}$$

To simplify the notation, given $v \in T_{x_0}\mathcal{M}$ we use $|v|_{x_0}$ to denote $\sqrt{g_{x_0}(v, v)}$.

In order to construct a family of N^d points on $\mathcal{M}$, we argue as follows: first of all we consider polar coordinates (ρ, ϑ) on $T_{x_0}\mathcal{M} \simeq \mathbb{R}^d$ induced by the constant metric g_{x_0}, where ϑ denotes a vector on the unit sphere $\mathbb{S}^{d-1}$, and then we consider a family of "radii" $0 < \rho_1 < \ldots < \rho_N < \infty$ and a set of N^{d-1} points $\{\vartheta_1, \ldots, \vartheta_{N^{d-1}}\} \subset \mathbb{S}^{d-1}$ distributed in a "uniform" way on the sphere so that

$$\min_k d_\theta(\vartheta, \vartheta_k) \leq \frac{C}{N} \qquad \forall \vartheta \in \mathbb{S}^{d-1}, \tag{3.4}$$

where $d_\theta(\vartheta, \vartheta_k)$ denotes the distance on the sphere induced by g_{x_0}.

We then define the family of points $p_{i,k}$ on the tangent space $T_{x_0}\mathcal{M}$ that, in polar coordinates, are given by $p_{i,k} := (\rho_i, \vartheta_k)$, and we take the family of points on $\mathcal{M}$ given by

$$x_{i,k} := \exp_{x_0}(p_{i,k}) \qquad i = 1, \ldots, N; \quad k = 1, \ldots, N^{d-1}.$$

We notice the following estimate: given a point $x \in \mathcal{M}$, we consider the vector $p = (\rho, \vartheta) \in T_{x_0}\mathcal{M}$ defined as $p := \dot{\gamma}(0)$ where $\gamma : [0, 1] \to \mathcal{M}$ is a constant speed minimizing geodesic arriving at x. By the definition of the
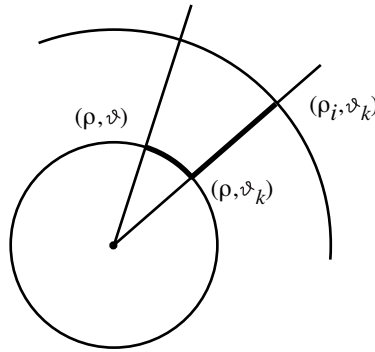
FIGURE 2. The bold curve joining (ρ, ϑ) and (ρ_i, ϑ_k) provides an upper bound for the distance between the two points.

exponential map we notice that $x = \exp_{x_0}(p)$ and $\rho = |p|_{x_0} = d(x, x_0)$. Then, we can estimate the distance between $x := \exp_{x_0}(p)$ and $x_{i,k}$ as follows: first we consider $\sigma : [0, 1] \to \mathbb{S}^{d-1} \subset T_{x_0}\mathcal{M}$ a geodesic (on the unit sphere) connecting ϑ to ϑ_k and we define $\eta := \exp_{x_0}(\rho\,\sigma)$, and then we connect $\exp_{x_0}((\rho, \vartheta_k))$ to $x_{i,k}$ considering $\gamma|_{[\rho,\rho_i]}$, where $\gamma(s) := \exp_{x_0}((s, \vartheta))$ is a unit speed geodesic (see Fig. 2).

Setting $\eta := \exp_{x_0}(\rho\,\sigma)$, this gives the bound

$$d(x, x_{i,k}) \leq \int_0^1 \left| \dot{\eta}(t) \right|_{\eta(t)} \mathrm{d}t + \left| \int_\rho^{\rho_i} \left| \dot{\gamma}(s) \right|_{\gamma(s)} \mathrm{d}s \right|$$
$$\leq A_{x_0}\big(d(x, x_0)\big)\, d_\theta(\vartheta_k, \vartheta) + |d(x, x_0) - \rho_i|,$$

where $A_{x_0}(\rho)$ is defined in (1.5), and we used that $\sigma(t)$ is a geodesic (on the sphere) from ϑ_k to ϑ and that $\rho = d(x, x_0)$.

Notice that, thanks to the estimate above and by (3.4),

$$\min_{i,k} d(x, x_{i,k})^r \leq \min_i \left[A_{x_0}\big(d(x, x_0)\big) \frac{C}{N} + |d(x, x_0) - \rho_i| \right]^r.$$

We can now estimate the quantization error:

$$N^r V_{N^d, r}(\mu) \leq N^r \int_\mathcal{M} \min_{i,k} \left[A_{x_0}\big(d(x, x_0)\big) \frac{C}{N} + |d(x, x_0) - \rho_i| \right]^r \mathrm{d}\mu(x).$$

Using that $(a + b)^r \leq 2^{r-1}(a^r + b^r)$ for $a, b > 0$ we get

$$N^r V_{N^d, r}(\mu) \leq N^r 2^{r-1} \int_\mathcal{M} \min_i |d(x, x_0) - \rho_i|^r \, \mathrm{d}\mu(x) + C^r 2^{r-1} \int_\mathcal{M} A_{x_0}\big(d(x, x_0)\big)^r \, \mathrm{d}\mu(x).$$

Let us now consider the map $d_{x_0} : \mathcal{M} \to \mathbb{R}$ defined as $d_{x_0}(x) := d(x, x_0)$, and define the probability measure on $\mathbb{R}$ given by $\mu_1 := (d_{x_0})_\# \mu$. In this way

$$\int_\mathcal{M} \min_i |d(x, x_0) - \rho_i|^r \, \mathrm{d}\mu(x) = \int_\mathbb{R} \min_i |s - \rho_i|^r \, \mathrm{d}\mu_1(s).$$

We now choose the radii ρ_i to be optimal for the quantization problem in one dimension for μ_1. Then the above estimate and Lemma 3.2 yield

$$N^r V_{N^d, r}(\mu) \leq C'\left(1 + \int_0^\infty s^{r+\delta}\, \mathrm{d}\mu_1(s) + \int_{\mathcal{M}} A_{x_0}\big(d(x, x_0)\big)^r \mathrm{d}\mu(x)\right)$$

$$= C'\left(1 + \int_{\mathcal{M}} d(x, x_0)^{r+\delta}\, \mathrm{d}\mu(x) + \int_{\mathcal{M}} A_{x_0}\big(d(x, x_0)\big)^r \mathrm{d}\mu(x)\right),$$

that concludes the proof of (3.1).

To show why this bound implies (3.2) (and hence Thm. 1.4 in the general non-compact case), we first notice that by (3.1) it follows that, for any $M \geq 1$,

$$M^{r/d} V_{M, r}(\mu) \leq C\left(1 + \int_{\mathcal{M}} d(x, x_0)^{r+\delta}\, \mathrm{d}\mu(x) + \int_{\mathcal{M}} A_{x_0}\big(d(x, x_0)\big)^r \mathrm{d}\mu(x)\right). \tag{3.5}$$

Indeed, for any $M \geq 1$ there exists $N \geq 1$ such that $N^d \leq M < (N+1)^d$, hence (since $V_{M, r}$ is decreasing in M)

$$M^{r/d} V_{M, r}(\mu) \leq (N+1)^r V_{N^d, r}(\mu) = \left(1 + \frac{1}{N}\right)^r N^r V_{N^d, r}(\mu)$$

$$\leq C\left(1 + \int_{\mathcal{M}} d(x, x_0)^{r+\delta}\, \mathrm{d}\mu(x) + \int_{\mathcal{M}} A_{x_0}\big(d(x, x_0)\big)^r \mathrm{d}\mu(x)\right),$$

which proves (3.5).

We now prove (3.2). Observe that, as shown in ([5], Proof of Thm. 6.2, Step 5), once the asymptotic quantization is proved for compactly supported probability measures, by the monotone convergence theorem one always has

$$\liminf_{N \to \infty} N^{r/d} V_{N, r}(\mu) \geq Q_r\big([0, 1]^d\big)\left(\int_{\mathcal{M}} h^{d/(d+r)}\, \mathrm{dvol}\right)^{(d+r)/d},$$

hence one only have to prove the limsup inequality.

For that, one splits the measure μ as the sum of $\mu_R^1 := \chi_{B_R(x_0)}\mu$ and $\mu_R^2 := \chi_{\mathcal{M} \setminus B_R(x_0)}\mu$, where $R \gg 1$. Then one applies ([5], Lem. 6.5(a)) to bound from above $N^{r/d} V_{N, r}(\mu)$ in terms of $N^{r/d} V_{N, r}(\mu_R^1)$ and $N^{r/d} V_{N, r}(\mu_R^2)$, and uses the result in the compact case for $N^{r/d} V_{N, r}(\mu_R^1)$, to obtain that, for any $\varepsilon \in (0, 1)$

$$\limsup_{N \to \infty} N^{r/d} V_{N, r}(\mu) \leq (1 - \varepsilon)^{-r/d} Q_r\big([0, 1]^d\big)\left(\int_{B_R(x_0)} h^{d/(d+r)}\, \mathrm{dvol}\right)^{(d+r)/d}$$

$$+ \mu(\mathcal{M} \setminus B_R(x_0))\, \varepsilon^{-r/d} \limsup_{N \to \infty} N^{r/d} V_{N, r}\left(\frac{1}{\mu(\mathcal{M} \setminus B_R(x_0))}\mu_R^2\right).$$

Thanks to (3.5), we can bound the limsup in the right hand side by

$$\varepsilon^{-r/d}\left(\mu(\mathcal{M} \setminus B_R(x_0)) + \int_{\mathcal{M}} d(x, x_0)^{r+\delta}\, \mathrm{d}\mu_R^2(x) + \int_{\mathcal{M}} A_{x_0}\big(d(x, x_0)\big)^r \mathrm{d}\mu_R^2(x)\right),$$

that tends to 0 as $R \to \infty$ by dominated convergence. Hence, letting $R \to \infty$ we deduce that

$$\limsup_{N \to \infty} N^{r/d} V_{N, r}(\mu) \leq (1 - \varepsilon)^{-r/d} Q_r\big([0, 1]^d\big)\left(\int_{\mathcal{M}} h^{d/(d+r)}\, \mathrm{dvol}\right)^{(d+r)/d},$$

and the result follows letting $\varepsilon \to 0$.

4. Proof of Theorem 1.7

We begin by noticing that if

$$\int_{\mathbb{H}^2} d(x, x_0)^p \, d\mu < \infty$$

for some $x_0 \in \mathbb{H}^2$, then this holds for any other point: indeed, given $x_1 \in \mathbb{H}^2$,

$$\int_{\mathbb{H}^2} d(x, x_1)^p \, d\mu \leq 2^{p-1} \int_{\mathbb{H}^2} \left[d(x, x_0)^p + d(x_0, x_1)^p \right] d\mu < \infty.$$

In particular, it suffices to check the moment condition at only one point.

We fix a point $x_0 \in \mathbb{H}^2$ and we use the exponential map at x_0 to identify $\mathbb{H}^2$ with $(\mathbb{R}^2, d^2\rho + \sinh \rho \, d^2 \vartheta)$. Then, we define the measure

$$\mu := \sum_{k \in \mathbb{N}} e^{-(1+\varepsilon)k} \mathcal{H}^1 \llcorner \mathbb{S}_k^1,$$

where $\mathcal{H}^1 \llcorner \mathbb{S}_R^1$ denotes the 1-dimensional Haudorff measure restricted to the circle around the origin of radius R, and $\varepsilon > 0$ is a constant to be fixed.

We begin by noticing that

$$\begin{aligned}
\int_{\mathbb{H}^2} d(x, x_0)^p \, d\mu &= \sum_{k \in \mathbb{N}} e^{-(1+\varepsilon)k} \int_{\mathbb{S}_k^1} \rho^p \, d\mathcal{H}^1 \\
&= \sum_{k \in \mathbb{N}} e^{-(1+\varepsilon)k} k^p \, 2\pi \sinh(k) \approx \sum_{k \in \mathbb{N}} e^{-\varepsilon k} k^p < \infty
\end{aligned}$$

for all $p > 0$.

An important ingredient of the proof will be the following estimate on the quantization error for the uniform measure on a circle around the origin.

Lemma 4.1. *For any $R \geq 1$ and $M \in \mathbb{N}$ we have*

$$V_{M,r}\left(\mathcal{H}^1 \llcorner \mathbb{S}_R^1\right) \gtrsim \left(\frac{e^R}{2R} - M \right)_+ R.$$

Proof. To prove the above estimate, we find a good competitor for the minimization problem. Let us denote with $[\cdot]$ the integer part, and define

$$L := \left[\frac{e^R}{2R} \right].$$

We split $\mathbb{S}_R^1$ in $2L$ arcs $\Sigma_{i,R}$ of equal length. Notice that the following estimate holds: there exists a positive constant c, independent of R, such that

$$d(\Sigma_{2j,R}, \Sigma_{2j',R}) > c \qquad \forall \, j \neq j' \in \{1, \ldots, L\}. \tag{4.1}$$

To show this fact, one argues as follows: consider a geodesic connecting a point $x_1 \in \Sigma_{2j,R}$ to $x_2 \in \Sigma_{2j',R}$. Because $j \neq j'$ any curve connecting them has to rotate by an angle of order at least R/e^R. Now, two cases arise: either the geodesic $\gamma : [0,1] \to \mathbb{H}^2$ is always contained inside $\mathbb{R}^d \setminus B_{R-1}(0)$, or not. In the first case we exploit that the metric is always larger than $\sinh^2(R-1)d^2\vartheta$. More precisely, if we denote by (e_ρ, e_θ) a basis of tangent vectors in polar coordinates

$$d(x, y) = \int_0^1 \sqrt{\left(\dot{\gamma}(t) \cdot e_\rho \right)^2 + \sinh^2(\rho)\left(\dot{\gamma}(t) \cdot e_\theta \right)^2} \, dt \gtrsim e^{R-1} \frac{R}{e^R} \approx R \geq 1,$$

where we used that γ has to rotate by an angle of order at least R/e^R. In the second case, to enter inside the ball $B_{R-1}(0)$ the geodesic has to travel a distance at least 1, so its length is greater that 1. This proves the validity of (4.1). We pick now a family of M points $\{x_\ell\}_{\ell=1}^M$. Then, by (4.1) and triangle inequality, we have that for every index ℓ there exists at most one index $j(\ell)$ such that

$$d(x_\ell, \Sigma_{2j,R}) > \frac{c}{2} \qquad \forall\, j \neq j(\ell).$$

Therefore there exists a family of indices $J \in \{1, \ldots, L\}$ of cardinality at least $(L-M)_+$ such that

$$d(x_\ell, \Sigma_{2j,R}) > \frac{c}{2} \qquad \forall j \in J, \quad \forall \ell = 1, \ldots, M.$$

We can now estimate the quantization error:

$$V_{M,r}\big(\mathcal{H}^1 \llcorner \mathbb{S}_R^1\big) \geq \min_{\alpha \subset \mathbb{H}^2 : |\alpha| = M} \sum_{j=1}^L \int_{\Sigma_{2j,R}} \min_{x_\ell \in \alpha} d(x, x_\ell)^r \, d\mathcal{H}^1$$

$$\geq \sum_{j \in J} \int_{\Sigma_{2j,R}} \left(\frac{c}{2}\right)^r d\mathcal{H}^1 \gtrsim (L-M)_+ R,$$

where at the last step we used that $\mathcal{H}^1(\Sigma_{2j,R}) \approx R$. $\qquad\square$

We can now conclude the proof. Indeed, given a set of points $\{x_\ell\}_{1 \leq \ell \leq N^2}$ optimal for μ, these points are admissible for the quantization problem of each measure $\mathcal{H}^1 \llcorner \mathbb{S}_k^1$, therefore

$$V_{N^2,r}(\mu) = \sum_{k \in \mathbb{N}} \mathrm{e}^{-(1+\varepsilon)k} \int_{\mathbb{S}_k^1} \min_\ell d(x, x_\ell)^r \, d\mathcal{H}^1(x)$$

$$\geq \sum_{k \in \mathbb{N}} \mathrm{e}^{-(1+\varepsilon)k} V_{N^2,r}\big(\mathcal{H}^1 \llcorner \mathbb{S}_k^1\big) \gtrsim \sum_{k \in \mathbb{N}} \mathrm{e}^{-(1+\varepsilon)k} \left(\frac{\mathrm{e}^k}{2k} - N^2\right)_+ k,$$

where at the last step we used Lemma 4.1. Noticing that, for N large,

$$\frac{\mathrm{e}^k}{2k} - N^2 \geq \frac{1}{4}\frac{\mathrm{e}^k}{k} \qquad \text{for } k \geq \log(N^4),$$

we conclude that

$$N^r V_{N^2,r}(\mu) \gtrsim \frac{N^r}{4} \sum_{k \geq \log(N^4)} \mathrm{e}^{-(1+\varepsilon)k} \frac{\mathrm{e}^k}{k} k = \frac{N^r}{4} \sum_{k \geq \log(N^4)} \mathrm{e}^{-\varepsilon k}$$

$$\gtrsim N^r \int_{\log(N^4)}^\infty \mathrm{e}^{-\varepsilon t} \, dt \approx \frac{N^r N^{-4\varepsilon}}{\varepsilon} \to \infty$$

as $N \to \infty$ provided we choose $\varepsilon < r/4$.

References

[1] G. Bouchitté, C. Jimenez and M. Rajesh, Asymptotique d'un problème de positionnement optimal. *C. R. Math. Acad. Sci. Paris* **335** (2002) 853–858.

[2] G. Bouchitté, C. Jimenez and M. Rajesh, Asymptotic analysis of a class of optimal location problems. *J. Math. Pures Appl.* **95** (2011) 382–419.

[3] A. Brancolini, G. Buttazzo, F. Santambrogio and E. Stepanov, Long-term planning *versus* short-term planning in the asymptotical location problem. *ESAIM: COCV* **15** (2009) 509–524.

[4] J. Bucklew and G. Wise, Multidimensional Asymptotic Quantization Theory with rth Power Distortion Measures. *IEEE Inform. Theory* **28** (1982) 239–247.

[5] S. Graf and H. Luschgy, Foundations of Quantization for Probability Distributions, Vol. 1730 of *Lect. Notes Math.* Springer-Verlag, Berlin Heidelberg (2000).

[6] B. Kloeckner, Approximation by finitely supported measures. *ESAIM: COCV* **18** (2012) 343–359.

[7] S. Mosconi, P. Tilli, Γ-Convergence for the Irrigation Problem. *J. Conv. Anal.* **12** (2005) 145–158.

[8] J.M. Lee, Riemannian manifolds. An introduction to curvature. Vol. 176 of *Grad. Texts Math.* Springer-Verlag, New York (1997).

[9] J.G. Ratcliffe, Foundations of hyperbolic manifolds, 2nd edn. Vol. 149 of *Grad. Texts Math.* Springer, New York (2006).

[10] C. Villani, Topics in Optimal Transportation, Vol. 58 of *Grad. Studies Math.* American Math. Soc., Providence RI (2003).