ESAIM: COCV 18 (2012) 343–359
DOI: 10.1051/cocv/2010100

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

APPROXIMATION BY FINITELY SUPPORTED MEASURES

Benoît Kloeckner[1]

Abstract. We consider the problem of approximating a probability measure defined on a metric space by a measure supported on a finite number of points. More specifically we seek the asymptotic behavior of the minimal Wasserstein distance to an approximation when the number of points goes to infinity. The main result gives an equivalent when the space is a Riemannian manifold and the approximated measure is absolutely continuous and compactly supported.

Mathematics Subject Classification. 49Q20, 90B85.

Received March 4, 2010. Revised November 9, 2010.
Published online April 13, 2011.

1. Introduction

In this paper we are interested in the following question: given a finite measure μ, at what speed can it be approximated by finitely supported measures? To give a sense to the question, one needs a distance on the space of measures; we shall use the Wasserstein distances W_p, with arbitrary exponent $p \in [1, +\infty)$ (definitions are recalled in Sect. 2).

This problem has been called *Quantization for probability distribution*, the case of exponent $p = 1$ has also been studied under the name of *location problem*, and the case $p = 2$ is linked with *optimal centroidal Voronoi tessellations*. After submission of the present article, we became aware that the previous works cover much more of the material presented than we first thought; see Section 1.2 for detailed references.

This problem could be of interest for a numerical study of transportation problems, where measures can be represented by discrete ones. One would need to know the number of points needed to achieve some precision in the approximation.

We shall restrict our attention to compactly supported Borelian measures on Riemannian manifolds.

1.1. Statement of the results

First we show that the order of convergence is determined by the dimension of the measure (see definitions in Sect. 2); Δ_N denotes the set of all measures supported in at most N points.

Theorem 1.1. *If μ is compactly supported and Alhfors regular of dimension $s > 0$, then*

$$W_p(\mu, \Delta_N) \approx \frac{1}{N^{1/s}}.$$

Keywords and phrases. Measures, Wasserstein distance, quantization, location problem, centroidal Voronoi tessellations.

[1] Institut Fourier, Université Joseph Fourier, BP 53, 38041 Grenoble, France. Benoit.Kloeckner@ujf-grenoble.fr

Article published by EDP Sciences

Here we write $\approx$ to say that one quantity is bounded above and below by positive multiples of the other. Examples of Ahlfors regular measures are given by the volume measures on submanifolds, and self-similar measures (see for example [7]). Theorem 1.1, to be proved in a slightly more general and precise form in Section 4, is simple and unsurprising; it reminds of ball packing and covering, and indeed relies on a standard covering argument.

In the particular case of absolutely continuous measures, one can give much finer estimates. First, it is easily seen that if $\square^d$ denotes the uniform measure on a Euclidean unit cube of dimension d, then there is a constant $\theta(d,p)$ such that

$$W_p(\square^d, \Delta_N) \sim \frac{\theta(d,p)}{N^{1/d}}$$

(Prop. 5.3). Note that determining the precise value of $\theta(d,p)$ seems difficult; known cases are discussed in Section 1.2.

The main result of this paper is the following, where "vol" denotes the volume measure on the considered Riemannian manifold and is the default measure for all integrals.

Theorem 1.2. *If $\mu = \rho\,\mathrm{vol}$ where ρ is a compactly supported function on a Riemannian manifold (M,g), then for all $1 \leqslant p < \infty$ we have*

$$W_p(\mu, \Delta_N) \sim \frac{\theta(d,p)\,|\rho|^{1/p}_{\frac{d}{d+p}}}{N^{1/d}} \tag{1.1}$$

where $|\rho|_\beta = (\int_M \rho^\beta)^{1/\beta}$ is the L^β "norm", here with $\beta < 1$ though.

Moreover, if (μ_N) is a sequence of finitely supported measures such that $\mu_N \in \Delta_N$ minimizes $W_p(\mu,\mu_N)$, then the sequence of probability measures $(\bar\mu_N)$ that are uniform on the support of μ_N converges weakly to the multiple of $\rho^{\frac{d}{p+d}}$ that has mass 1.

Theorem 1.2 is proved in Section 5. Note that the hypothesis that μ has compact support is obviously needed: otherwise, $|\rho|_{d/(d+p)}$ could be infinite. Even when μ is in $L^{d/(d+p)}$, there is the case where it is supported on a sequence of small balls going to infinity. Then the location of the balls is important in the quality of approximation and not only the profile of the density function. However, this hypothesis could probably be relaxed to a moment condition.

Theorem 1.2 has no real analog for measures of fractional dimension.

Theorem 1.3. *There is a s-dimensional Ahlfors regular measure κ on $\mathbb{R}$ (namely, κ is the Cantor dyadic measure) such that $W_p(\kappa,\Delta_N)N^{1/s}$ has no limit.*

Section 6 is devoted to this example.

Part of the interest of Theorem 1.2 comes from the following observation, to be discussed in Section 7: when $p = 2$, the support of a distance minimizing $\mu_N \in \Delta_N$ generates a centroidal Voronoi tessellation, that is, each point is the center of mass (with respect to μ) of its Voronoi cell. We thus get the asymptotic repartition of an important family of centroidal Voronoi tessellations, which enables us to prove some sort of energy equidistribution principle.

1.2. **Discussion of previously known results**

There are several previous works closely related to the content of this paper.

1.2.1. *Foundations of Quantization for Probability Distributions*

The book [10] by Graf and Luschgy (see also the references therein), that we only discovered recently, contains many results on the present problem. Theorem 1.1 is proved there in Section 12, but our proof seems more direct. Theorem 1.2 is proved in the Euclidean case in Sections 6 and 7 (with a weakening of the compact support assumption). A generalization of Theorem 1.3 is proved in Section 14, yet we present a proof for the sake of self-completeness.

The case $p = 1$, $M = \mathbb{R}^n$ is usually called the *location problem*. In this setting, Theorem 1.2 has also been proved by Bouchitté *et al.* [1] under the additional assumption that ρ is lower semi-continuous.

Our main motivation to publish this work despite these overlaps is that the case of measures on manifold should find applications; for example, good approximations of the curvature measure of a convex body by discrete measures should give good approximations of the body by polyhedra.

It seems that the quantization, the location problem and the study of optimal CVTs, although the last two are particular cases of the first one, have been only studied independently. We hope that noticing this proximity will encourage progress on each question to be translated in the others.

1.2.2. *Around the main theorem*

Mosconi and Tilli in [15] have studied (for any exponent p, in $\mathbb{R}^n$) the *irrigation problem*, where the approximating measures are supported on connected sets of length $< \ell$ (the length being the 1-dimensional Hausdorff measure) instead of being supported on N points; the order of approximation is then $\ell^{1/(d-1)}$.

Brancolini *et al.* compare in [2] the location problem with its "short-term planning" version, where the support of $\operatorname{supp} \mu_N$ is constructed by adding one point to that of μ_{N-1}, minimizing the cost only locally in N.

1.2.3. *Approximation constants for cubes*

Some values of $\theta(d, p)$ have been determined. First, it is easy to compute them in dimension 1:

$$\theta(1, p) = \frac{(p + 1)^{-1/p}}{2}\cdot$$

The case $d = 2$ has been solved by Fejes Tóth [8,9] (and by Newman [16] for $p = 2$ and Morgan and Bolton [14] for $p = 1$), see also [10], Section 8. In particular

$$\theta(2, 2) = 5\sqrt{3}/54, \qquad \theta(2, 1) = 2^{-2/3}3^{-7/4}(4 + \ln 27).$$

When $d = 2$ and for all p, the hexagonal lattice is optimal (that is, the given θ is the distance between the uniform measure on a regular hexagon and a Dirac mass at its center). All other cases are open to our knowledge. For numerical evidence in the case $p = 2, d = 3$ see Du and Wang [5]. Note that in the limit case $p = \infty$, determining θ amounts to determining the minimal density of a ball covering of $\mathbb{R}^d$, which is arguably as challenging as determining the maximal density of a ball packing, a well-known open problem if $d > 3$.

1.2.4. *Random variations*

Concerning the order of convergence, it is worth comparing with the problem of estimating the distance from a measure μ to empirical measures $\bar{\mu}_N = N^{-1} \sum_k \delta_{X_k}$ where $X_1, \dots, X_N$ are independent random variables of law μ. It seems that $\bar{\mu}_N$ is almost optimal in the sense that $W_2(\mu, \bar{\mu}_N) \sim C\, N^{-1/d}$ almost surely (under moment conditions, but here we take μ compactly supported so this is not an issue); Horowitz and Karandikar have shown in [12] that $W_2(\mu, \bar{\mu}_N)$ has the order at most $N^{-1/(d+4)}$ and the better exponent above is suggested in the Mathematical Review of that paper.

Let us also briefly note that the optimal matching problem for random data is related to our problem. Simply put, one can say that if $\bar{\mu}'_N$ is another empirical measure of μ, then $W_2(\bar{\mu}_N, \bar{\mu}'_N)$ also has the order $N^{-1/d}$ if $d \geqslant 3$ (see for example Dobrić and Yukich [4]). In the same flavour, other optimisation problems for random data have been studied (minimal length covering tree, traveling salesperson problem, bipartite version of those, etc.)

1.2.5. *Centroidal Voronoi Tessellations*

In the case $p = 2$, the problem is linked to (optimal) *Centroidal Voronoi Tessellation*, see Section 7 and [6]. In that paper (Sect. 6.4.1), the principle of energy equidistribution is given in the 1-dimensional case for smooth density ρ. Our Corollary 7.1 in the last section generalize this to non regular densities, all exponents, and all dimensions; it is however quite a direct consequence of Theorem 1.2.

1.3. Related open questions

The number N of points of the support may be the first measure of complexity of a finitely supported measure that one comes up with, but it is not necessarily the most relevant. Concerning the problem of numerical analysis of transportation problems, numbers are usually encoded in a computer by floating numbers. One could therefore define the complexity of a measure supported on points of decimal coordinates, with decimal quantity of mass at each point as the memory size needed to describe it, and search to minimize the distance to a given μ among measures of given complexity.

Another possible notion of complexity is entropy: one defines

$$h\left(\sum_i m_i\delta_{x_i}\right) = -\sum_i m_i \ln(m_i).$$

A natural question is to search a μ_h that minimizes the distance to μ among the finitely supported measures of entropy at most h, and to study the behavior of μ_h when we let $h \to \infty$.

2. Recalls and definitions

2.1. Notations

Given two sequences (u_n), (v_n) of non-negative real numbers, we shall write:

- $u_n \lesssim v_n$ to mean that there exists a positive real a and an integer N_0 such that $u_n \leqslant av_n$ for all $n \geqslant N_0$,
- $u_n \approx v_n$ to mean $u_n \lesssim v_n$ and $u_n \gtrsim v_n$.

From now on, M is a given Riemannian manifold of dimension d. By a *domain* of M we mean a compact domain with piecewise smooth boundary (and possibly corners) and finitely many connected components.

2.2. Ahlfors regularity and a covering result

We denote by $B(x,r)$ the closed ball of radius r and center x; sometimes, when $B = B(x,r)$ and $k \in \mathbb{R}$, we denote by kB the ball $B(x,kr)$.

Let μ be a finite, compactly supported measure on a manifold M of dimension d, and let $s \in (0,+\infty)$. One says that μ is *Ahlfors regular* of dimension s if there is a constant C such that for all $x \in \operatorname{supp}\mu$ and for all $r \leqslant \operatorname{diam}(\operatorname{supp}\mu)$, one has

$$C^{-1}r^s \leqslant \mu(B(x,r)) \leqslant Cr^s.$$

This is a strong condition, but is satisfied for example by auto-similar measures, see [7,13] for definitions and Section 6 for the most famous example of the Cantor measure.

Note that if μ is Ahlfors regular of dimension s, then s is the Hausdorff dimension of $\operatorname{supp}\mu$ (and therefore $s \leqslant d$), see [11], Section 8.7.

We shall need the following classical covering result.

Proposition 2.1 (5δ covering)**.** *If X is a closed set and $\mathcal{F}$ is a family of balls of uniformly bounded diameter such that $X \subset \bigcup_{\mathcal{F}} B$, then there is a subfamily $\mathcal{G}$ of $\mathcal{F}$ such that:*

- *$X \subset \bigcup_{\mathcal{G}} 5B$;*
- *$B \cap B' = \varnothing$ whenever $B \neq B' \in \mathcal{G}$.*

2.3. Wasserstein distances

Here we recall some basic facts on optimal transportation and Wasserstein distances. For more information, the reader is suggested to look for example at Villani's book [17] which provides a very good introduction to this topic.

First consider the case $p < \infty$, which shall attract most of our attention. A finite measure μ on M is said to have *finite p-th moment* if for some (hence all) $x_0 \in M$ the following holds:

$$\int_{\mathbb{R}^d} d(x_0, x)^p \mu(\mathrm{d}x) < +\infty.$$

In particular, any compactly supported finite measure has finite p-th moment for all p.

Let μ_0, μ_1 be two finite measures having finite p-th moment and the same mass. A *transport plan* between μ_0 and μ_1 is a measure Π on $M \times M$ that has μ_0 and μ_1 as marginals, that is : $\Pi(A \times M) = \mu_0(A)$ and $\Pi(M \times A) = \mu_1(A)$ for all Borelian set A. One shall think of Π has a assignment of mass: $\Pi(A \times B)$ represents the mass sent from A to B.

The L^p *cost* of a transport plan is defined as

$$c_p(\Pi) = \int_{M \times M} d(x, y)^p \Pi(\mathrm{d}x\mathrm{d}y).$$

One defines the L^p *Wasserstein distance* by

$$W_p(\mu_0, \mu_1) = \inf_{\Pi} c_p(\Pi)^{1/p}$$

where the infimum is on all transport plan between μ_0 and μ_1. One can show that there is always a transport plan that achieves this infimum, and that W_p defines a distance on the set of measures with finite p-th moment and given mass.

Moreover, if M is compact W_p metrizes the weak topology. If M is non-compact, it defines a finer topology.

Most of the time, one restricts itself to probability measures. Here, we shall use extensively mass transportation between submeasures of the main measures under study, so that we need to consider measures of arbitrary mass. Given positive measures μ and ν, we write that $\mu \leqslant \nu$ if $\mu(A) \leqslant \nu(A)$ for all Borelian set A, which means that $\nu - \mu$ is also a positive measure.

It is important to notice that $c_p(\Pi)$ is homogeneous of degree 1 in the total mass and of degree p on distances, so that in the case $M = \mathbb{R}^d$ if φ is a similitude of ratio r, we have $W_p(m\,\varphi_\# \mu_0, m\,\varphi_\# \mu_1) = m^{1/p}\, r\, W_p(\mu_0, \mu_1)$.

The case $p = \infty$ is obtained as a limit of the finite case, see [3]. Let μ_0 and μ_1 be compactly supported measures of the same mass and let Π be a transport plan between μ_0 and μ_1. The L^∞ *length* of Π is defined as

$$\ell_\infty(\Pi) = \sup\{d(x, y) \mid x, y \in \operatorname{supp} \Pi\}$$

that is, the maximal distance moved by some infinitesimal amount of mass when applying Π. The L^∞ distance between μ_0 and μ_1 then is

$$W_\infty(\mu_0, \mu_1) = \inf_{\Pi} \ell_\infty(\Pi)$$

where the infimum is on all transport plan from μ_0 to μ_1. In a sense, the L^∞ distance is a generalisation to measures of the Hausdorff metric on compact sets. We shall use ℓ_∞, but not d_∞. The problem of minimizing $W_\infty(\mu, \Delta_N)$ is a matter of covering $\operatorname{supp} \mu$ (independently of μ itself), a problem with quite a different taste than our.

3. Preparatory results

The following lemmas are useful tools we shall need; the first two at least cannot pretend to any kind of originality by themselves.

Lemma 3.1 (monotony)**.** *Let μ and ν be finite measures of equal mass and $\tilde{\mu} \leqslant \mu$. Then there is a measure $\tilde{\nu} \leqslant \nu$ (in particular,* $\operatorname{supp} \tilde{\nu} \subset \operatorname{supp} \nu$*) such that*

$$W_p(\tilde{\mu}, \tilde{\nu}) \leqslant W_p(\mu, \nu).$$

Proof. Let Π be an optimal transportation plan from μ to ν. We construct a low-cost transportation plan from $\tilde{\mu}$ to $\tilde{\nu}$ by disintegrating Π.

There is a family of finite measures $(\eta_x)_{x\in M}$ such that $\Pi = \int \eta_x \mu(\mathrm{d}x)$, that is

$$\Pi(A\times B) = \int_A \eta_x(B)\mu(\mathrm{d}x)$$

for all Borelian A and B. Define

$$\tilde{\Pi}(A\times B) = \int_A \eta_x(B)\tilde{\mu}(\mathrm{d}x)$$

and let $\tilde{\nu}$ be the second factor projection of $\tilde{\Pi}$. Since $\tilde{\Pi} \leqslant \Pi$, we have $\tilde{\nu} \leqslant \nu$ and $c_p(\tilde{\Pi}) \leqslant c_p(\Pi)$; moreover $\tilde{\Pi}$ is a transport plan from $\tilde{\mu}$ to $\tilde{\nu}$ by definition of $\tilde{\nu}$. □

Lemma 3.2 (summing)**.** *Let (μ,ν) and $(\tilde{\mu},\tilde{\nu})$ be finite measures with pairwise equal masses. Then*

$$W_p^p(\mu+\tilde{\mu},\nu+\tilde{\nu}) \leqslant W_p^p(\mu,\nu) + W_p^p(\tilde{\mu},\tilde{\nu}).$$

Proof. Let Π and $\tilde{\Pi}$ be optimal transport plans between respectively μ and ν, $\tilde{\mu}$ and $\tilde{\nu}$. Then $\Pi+\tilde{\Pi}$ is a transport plan between $\mu+\tilde{\mu}$ and $\nu+\tilde{\nu}$ whose cost is $c_p(\Pi+\tilde{\Pi}) = c_p(\Pi)+c_p(\tilde{\Pi})$. □

This very simple results have a particularly important consequence concerning our question.

Lemma 3.3 (L^1 stability)**.** *Let μ and $\tilde{\mu}$ be finite compactly supported measures on M, $\varepsilon \in (0,1)$ and (μ_N) be any sequence of N-supported measures.*

There is a sequence of N-supported measures $\tilde{\mu}_N$ such that there are at most εN points in $\operatorname{supp}\tilde{\mu}_N \setminus \operatorname{supp}\mu_N$ *and*

$$W_p^p(\tilde{\mu},\tilde{\mu}_N) \leqslant W_p^p(\mu,\mu_{N_1}) + O\left(\frac{|\mu-\tilde{\mu}|_{TV}}{(\varepsilon N)^{p/d}}\right)$$

where N_1 is equivalent to (and at least) $(1-\varepsilon)N$, $|\cdot|_{TV}$ is the total variation norm and the constant in the O depends only on the geometry of a domain where both μ and $\tilde{\mu}$ are concentrated.

In particular we get

$$W_p^p(\mu,\Delta_N) \leqslant W_p^p(\tilde{\mu},\Delta_{N_1}) + O\left(\frac{|\mu-\tilde{\mu}|_{TV}}{(\varepsilon N)^{p/d}}\right).$$

The name of this result has been chosen to emphasize that the total variation distance between two absolutely continuous measures is half the L^1 distance between their densities.

Proof. We can write $\tilde{\mu} = \mu' + \nu$ where $\mu' \leqslant \mu$ and ν is a positive measure of total mass at most $|\mu-\tilde{\mu}|_{VT}$. If D is a domain supporting both μ and $\tilde{\mu}$, it is a classical fact that there is a constant C (depending only on D) such that for all integer K, there are points $x_1,\dots,x_{K^d} \in D$ such that each point of D is at distance at most C/K from one of the x_i. For example if D is a Euclidean cube of side length L, by dividing it regularly one can achieve $C = L\sqrt{d}/2$.

Take $K = \lfloor (\varepsilon N)^{1/d} \rfloor$; then by sending each point of D to a closest x_i, one constructs a transport plan between ν and a K^d-supported measure ν_N whose cost is at most $|\mu-\tilde{\mu}|_{VT}(C/K)^p$.

Let $N_1 = N - K^d$. The monotony lemma gives a measure $\mu'_N \leqslant \mu_{N_1}$ (in particular, μ'_N is N_1-supported) such that

$$W_p(\mu',\mu'_N) \leqslant W_p(\mu,\mu_{N_1}).$$

The summing lemma now shows that

$$W_p^p(\tilde{\mu},\mu'_N+\nu_N) \leqslant W_p^p(\mu,\mu_{N_1}) + \frac{C^p|\mu-\tilde{\mu}|_{VT}}{K^p}.$$ □

Note that the presence of the $|\mu - \tilde{\mu}|_{TV}$ factor will be crucial in the sequel, but would not be present in the limit case $p = \infty$, which is therefore very different.

Lemma 3.4 (metric stability)**.** *Assume D is a compact domain of M, endowed with two different Riemannian metrics g and g' (defined on a open neighborhood of D). Denote by $|g' - g|$ the minimum number r such that*

$$e^{-2r} g_x(v, v) \leqslant g'_x(v, v) \leqslant e^{2r} g_x(v, v)$$

for all $x \in D$ and all $v \in T_x M$.

Then, denoting by W_p the Wasserstein metric computed using the distance d induced by g, and by W'_p the one obtained from the distance d' induced by g', one has for all measures μ, ν supported on D and of equal mass:

$$e^{-|g'-g|} W_p(\mu, \nu) \leqslant W'_p(\mu, \nu) \leqslant e^{|g'-g|} W_p(\mu, \nu).$$

Proof. For all $x, y \in D$ one has $d'(x, y) \leqslant e^r d(x, y)$ by computing the g'-length of a g-minimizing (or almost minimizing to avoid regularity issues on the boundary) curve connecting x to y. The same reasoning applies to transport plans: if Π is optimal from μ to ν according to d, then the d' cost of Π is at most e^{pr} times the d-cost of Π, so that $W'_p(\mu, \nu) \leqslant e^r W_p(\mu, \nu)$. The other inequality follows by symmetry. $\square$

Let us end with a result showing that no mass is moved very far away by an optimal transport plan to a N-supported measure if N is large enough.

Lemma 3.5 (localization)**.** *Let μ be a compactly supported finite measure. If μ_N is a closest N-supported measure to μ in L^p Wasserstein distance and Π_N is a L^p optimal transport plan between μ and μ_N, then when N goes to ∞,*

$$\ell_\infty(\Pi_N) \to 0.$$

Proof. Assume on the contrary that there are sequences $N_k \to \infty$, $x_k \in \operatorname{supp} \mu$ and a number $\varepsilon > 0$ such that Π_{N_k} moves x_k by a distance at least ε. There is a covering of $\operatorname{supp} \mu$ by a finite number of balls of radius $\varepsilon/3$. Up to extracting a subsequence, we can assume that all x_k lie in one of this balls, denoted by B. Since B is a neighborhood of x_k and $x_k \in \operatorname{supp} \mu$, we have $\mu(B) > 0$. Since Π_{N_k} is optimal, it moves x_k to a closest point in $\operatorname{supp} \mu_{N_k}$, which must be at distance at least ε from x_k. Therefore, every point in B is at distance at least $\varepsilon/3$ from $\operatorname{supp} \mu_{N_k}$, so that $c_p(\Pi_{N_k}) \geqslant \mu(B)(\varepsilon/3)^p > 0$, in contradiction with $W_p(\mu, \Delta_N) \to 0$. $\square$

4. Approximation rate and dimension

Theorem 1.1 is the union of the two following propositions. Note that the estimates given do not depend much on p, so that in fact Theorem 1.1 stays true when $p = \infty$.

Proposition 4.1. *If μ is a compactly supported probability measure on M and if for some $C > 0$ and for all $r \leqslant \operatorname{diam}(\operatorname{supp} \mu)$, one has*

$$C^{-1} r^s \leqslant \mu(B(x, r))$$

then for all N

$$W_p(\mu, \Delta_N) \leqslant \frac{5C^{1/s}}{N^{1/s}}.$$

Proof. The 5δ covering proposition above implies that given any $\delta > 0$, there is a subset $\mathcal{G}$ of $\operatorname{supp} \mu$ such that

- $\operatorname{supp} \mu \subset \bigcup_{x \in \mathcal{G}} B(x, 5\delta)$;
- $B(x, \delta) \cap B(x', \delta) = \varnothing$ whenever $x \neq x' \in \mathcal{G}$.

In particular, as soon as $\delta < \operatorname{diam}(\operatorname{supp}\mu)$ one has

$$1 \geqslant \sum_{x\in\mathcal{G}} \mu(B(x,\delta)) \geqslant |\mathcal{G}|C^{-1}\delta^s$$

so that $\mathcal{G}$ is finite, with $|\mathcal{G}| \leqslant C\delta^{-s}$.

Let $\tilde\mu$ be a measure supported on $\mathcal{G}$, that minimizes the L^p distance to μ among those. A way to construct $\tilde\mu$ is to assign to a point $x \in \mathcal{G}$ a mass equal to the μ-measure of its Voronoi cell, that is of the set of points nearest to x than to any other points in $\mathcal{G}$. The mass at a point at equal distance from several elements of $\mathcal{G}$ can be split indifferently between those. The previous discussion also gives a transport plan from μ to $\tilde\mu$, where each bit of mass moves a distance at most 5δ, so that $W_p(\mu,\tilde\mu) \leqslant 5\delta$ (whatever p).

Now, let N be a positive integer and choose $\delta = (C/N)^{1/s}$. The family $\mathcal{G}$ obtained from that δ has less than N elements, so that $W_p(\mu,\Delta_N) \leqslant 5(C/N)^{1/s}$. □

Proposition 4.2. *If μ is a probability measure on M and if for some $C > 0$ and for all r, one has*

$$\mu(B(x,r)) \leqslant Cr^s$$

then for all N,

$$\left(\frac{s}{s+p}\right)^{1/p} C^{-1/s}\ \frac{1}{N^{1/s}} \leqslant W_p(\mu,\Delta_N).$$

Proof. Consider a measure $\mu_N \in \Delta_N$ that minimizes the distance to μ. For all $\delta > 0$, the union of the balls centered at $\operatorname{supp}\mu_N$ and of radius δ has μ-measure at most $NC\delta^s$. In any transport plan from μ to μ_N, a quantity of mass at least $1 - NC\delta^s$ travels a distance at least δ, so that in the best case the quantity of mass traveling a distance between $\delta < (NC)^{-1/s}$ and $\delta + \mathrm{d}\delta$ is $NCs\delta^{s-1}\mathrm{d}\delta$. It follows that

$$W_p^p(\mu,\mu_N) \geqslant \int_0^{(NC)^{-1/s}} sNC\delta^{s-1}\delta^p \mathrm{d}\delta$$

so that $W_p(\mu,\mu_N) \geqslant (s/(s+p))^{1/p}(NC)^{-1/s}$. □

In fact, Theorem 1.1 applies to more general measures, for example combination of Ahlfors regular ones, thanks to the following.

Lemma 4.3. *If $\mu = a_1\mu^1 + a_2\mu^2$ where $a_i > 0$ and μ^i are probability measures such that $W_p(\mu^2,\Delta_N) \lesssim W_p(\mu^1,\Delta_N)$ and $W_p(\mu^1,\Delta_N) \lesssim W_p(\mu^1,\Delta_{2N})$ then $W_p(\mu,\Delta_N) \approx W_p(\mu^1,\Delta_N)$.*

Proof. By the monotony lemma, $W_p(a_1\mu^1,\Delta_N) \leqslant W_p(\mu,\Delta_N)$ so that $W_p(\mu^1,\Delta_N) \leqslant a_1^{-1/p}W_p(\mu,\Delta_N)$.

The summing lemma gives

$$W_p(\mu,\Delta_{2N}) \leqslant \big(W_p(a_1\mu^1,\Delta_N)^p + W_p(a_2\mu^2,\Delta_N)^p\big)^{1/p}$$

so that

$$W_p(\mu,\Delta_{2N}) \lesssim W_p(a_1\mu^1,\Delta_N) \lesssim W_p(\mu^1,\Delta_{2N}).$$

Since $W_p(\mu,\Delta_{2N+1}) \leqslant W_p(\mu,\Delta_{2N})$ we also get

$$W_p(\mu,\Delta_{2N+1}) \lesssim W_p(\mu^1,\Delta_{2N}) \lesssim W_p(\mu^1,\Delta_{4N}) \leqslant W_p(\mu^1,\Delta_{2N+1}).$$ □

The following is now an easy consequence of this lemma.

Corollary 4.4. *Assume that $\mu = \sum_{i=1}^{k} a_i \mu^i$ where $a_i > 0$ and μ^i are probability measures that are compactly supported and Ahlfors regular of dimension $s_i > 0$. Let $s = \max_i(s_i)$. Then*

$$W_p(\mu, \Delta_N) \approx \frac{1}{N^{1/s}}.$$

5. Absolutely continuous measures

In this section we prove Theorem 1.2. To prove the Euclidean case, the idea is to approximate (in the L^1 sense) a measure with density by a combination of uniform measures in squares. Then a measure on a manifold can be decomposed as a combination of measures supported in charts, and by metric stability the problem reduces to the Euclidean case.

The following key lemma shall be used several times to extend the class of measures for which we have precise approximation estimates.

Lemma 5.1 (combination)**.** *Let μ be an absolutely continuous measure on M. Let D_i ($1 \leqslant i \leqslant I$) be domains of M whose interiors do not overlap, and assume we can decompose $\mu = \sum_{i=1}^{I} \mu^i$ where μ^i is non-zero and supported on D_i. Assume moreover that there are numbers $(\alpha_1, \ldots, \alpha_I) = \alpha$ such that*

$$W_p(\mu^i, \Delta_N) \sim \frac{\alpha_i}{N^{1/d}}.$$

Let $\mu_N \in \Delta_N$ be a sequence minimizing the W_p distance to μ and define N_i (with implicit dependency on N) as the number of points of $\operatorname{supp} \mu_N$ that lie on D_i, the points lying on a common boundary of two or more domains being attributed arbitrarily to one of them.

If the vector $(N_i/N)_i$ has a cluster point $x = (x_i)$, then x minimizes

$$F(\alpha; x) = \left(\sum_i \frac{\alpha_i^p}{x_i^{p/d}} \right)^{1/p}$$

and if $(N_i/N)_i \to x$ when $N \to \infty$, then

$$W_p(\mu, \mu_N) \sim \frac{F(\alpha; x)}{N^{1/d}}.$$

Note that the assumption that none of the μ^i vanish is obviously unnecessary (but convenient). If some of the μ^i vanish, one only has to dismiss them.

Proof. For simplicity we denote $c_p(N) = W_p^p(\mu, \Delta_N)$. Let ε be any positive, small enough number.

We can find a $\delta > 0$ and domains $D_i' \subset D_i$ such that: each point of D_i' is at distance at least δ from the complement of D_i; if μ'^i is the restriction of μ^i to D_i', $|\mu'^i - \mu^i|_{VT} \leqslant \varepsilon^{1+p/d}$.

Assume x is the limit of $(N_i/N)_i$ when $N \to \infty$. Let us first prove that none of the x_i vanishes. Assume the contrary for some index i: then $N_i = o(N)$. For each N choose an optimal transport plan Π_N from μ to μ_N. Let $\nu_N \leqslant \mu^i$ be the part of μ^i that is sent by Π_N to the N_i points of $\operatorname{supp} \mu_N$ that lie in D_i, constructed as in the summing lemma, and let $m_N = \mu^i(D_i') - \nu_N(D_i')$ be the mass that moves from D_i' to the exterior of D_i under Π_N. Then the cost of Π_N is bounded from below by $m_N \delta^p + W_p^p(\nu_N, \Delta_{N_i})$. Since it goes to zero, we have $m_N \to 0$ and up to extracting a subsequence $\nu_N \to \nu$ where $\mu'^i \leqslant \nu \leqslant \mu^i$. The cost of Π_N is therefore bounded from below by all number less than $W_p^p(\nu_N, \Delta_{N_i}) \gtrsim N_i^{-p/d} \gg N^{-p/d}$, a contradiction.

Now, by considering optimal transport plans between μ^i and optimal N_i-supported measures of D_i, we get that

$$
\begin{aligned}
c_p(N) &\leqslant \sum_i W_p^p(\mu^i, \Delta_{N_i}) \\
&\leqslant \sum_i \frac{(\alpha_i+\varepsilon)^p}{N_i^{p/d}}
\end{aligned}
$$

when all N_i are large enough, which happens if N itself is large enough given that $x_i \neq 0$.

For N large enough, the localization lemma ensures that no mass is moved more than δ by an optimal transport plan between μ and μ_N. This implies that the cost $c_p(N)$ is bounded below by $\sum_i W_p^p(\mu'^i, \Delta_{N_i})$. By L^1-stability this gives the bound

$$
c_p(N) \geqslant \frac{\alpha_i^p(1-\varepsilon)^{p/d} + O(\varepsilon)}{(x_i N)^{p/d}}.
$$

The two inequalities above give us

$$
c_p(N) N^{p/d} \to \sum_i \frac{\alpha_i^p}{x_i^{p/d}} = F^p(\alpha; x).
$$

Now, if x is a mere cluster point of (N_i/N), this still holds up to a subsequence. If x did not minimize $F(\alpha; x)$, then by taking best approximations of μ^i supported on $x'_i N$ points where x' is a minimizer, we would get by the same computation a sequence μ'_N with better asymptotic behavior than μ_N (note that we used the optimality of μ_N only to bound from above each $W_p^p(\mu^i, \Delta_{N_i})$). □

The study of the functional F is straightforward.

Lemma 5.2. *Fix a positive vector $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_I)$ and consider the simplex $X = \{(x_i)_i \mid \sum_i x_i = 1, x_i > 0\}$. The function $F(\alpha; \cdot)$ has a unique minimizer $x^0 = (x_i^0)$ in X, which is proportional to $\left(\alpha_i^{\frac{dp}{d+p}}\right)_i$, with*

$$
F(\alpha; x^0) = \left(\sum_i \alpha_i^{\frac{dp}{d+p}}\right)^{\frac{d+p}{dp}} =: |\alpha|_{\frac{dp}{d+p}}.
$$

As a consequence, in the combination lemma the vector (N_i/N) must converge to x^0.

Proof. First $F(\alpha; \cdot)$ is continuous and goes to ∞ on the boundary of X, so that is must have a minimizer. Any minimizer must be a critical point of F^p and therefore satisfy

$$
\sum_i \alpha_i^p x_i^{-p/d-1} \eta_i = 0
$$

for all vector (η_i) such that $\sum_i \eta_i = 0$. This holds only when $\alpha_i^p x_i^{-p/d-1}$ is a constant and we get the uniqueness of x^0 and its expression:

$$
x_i^0 = \frac{\alpha_i^{\frac{dp}{d+p}}}{\sum_j \alpha_j^{\frac{dp}{d+p}}}.
$$

The value of $F(\alpha; x^0)$ follows.

In the combination lemma, we now by compacity that (N_i/N) must have cluster points, all of which must minimize $F(\alpha; \cdot)$. Since there is only one minimizer, (N_i/N) has only one cluster point and must converge to x^0. □

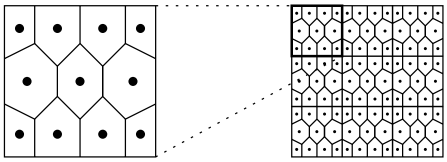

FIGURE 1. An optimal N-supported measure can be used to construct a good $k^d N$-supported measure for all k.

We are now ready to tackle more and more cases in Theorem 1.2. As a starting point, we consider the uniform measure $\square^d$ on the unit cube of $\mathbb{R}^d$ (endowed with the canonical metric).

Proposition 5.3. *There is a number $\theta(d,p) > 0$ such that*

$$W_p(\square^d, \Delta_N) \sim \frac{\theta(d,p)}{N^{1/d}}\cdot$$

The proof is obviously not new, since it is the same argument that shows that an optimal packing (or covering) of the Euclidean space must have a well-defined density (its upper and lower densities are equal).

Proof. Let $c(N) = W_p^p(\square^d, \Delta_N)$. We already know that $c(N) \approx N^{-p/d}$, so let $A = \liminf N^{p/d} c(N)$ and consider any $\varepsilon > 0$.

Let N_1 be an integer such that $c(N_1) \leqslant (A+\varepsilon) N_1^{-p/d}$ and let $\mu_1 \in \Delta_{N_1}$ be nearest to μ. For any integer ℓ, we can write $\ell = k^d + q$ where $k = \lfloor \ell^{1/d} \rfloor$ and q is an integer; then $q = O(\ell^{1-1/d}) = o(\ell)$ where the o depends only on d.

Divide the cube into k^d cubes of side length $1/k$, and consider the element μ_k of $\Delta_{k^d N_1}$ obtained by duplicating μ_1 in each of the cubes, with scaling factor k^{-1} and mass factor k^{-d} (see Fig. 1). The obvious transport plan obtained in the same way from the optimal one between $\square^d$ and μ_1 has total cost $k^{-p} c(N_1)$, so that

$$c(\ell N_1) \leqslant k^{-p} \frac{A+\varepsilon}{N_1^{p/d}} = \left(\frac{\ell}{k^d}\right)^{p/d} \frac{A+\varepsilon}{(\ell N_1)^{p/d}}\cdot$$

But since $k^d \sim \ell$, for ℓ large enough we get

$$c(\ell N_1) \leqslant \frac{A+2\varepsilon}{(\ell N_1)^{p/d}}\cdot$$

Now $N \sim \lfloor N/N_1 \rfloor N_1$ so that for N large enough $c(N) \leqslant (A+3\varepsilon) N^{-p/d}$. This proves that $\limsup N^{p/d} c(N) \leqslant A + 3\varepsilon$ for all $\varepsilon > 0$. $\square$

Note that we used the self-similarity of the cube at many different scales; the result does not hold with more general self-similar (fractal) measures, see Section 6.

Now, the combination lemma enables us to extend the validity domain of equation (1.1).

Lemma 5.4. *Let $\mu = \rho\lambda$ where λ is the Lebesgue measure on $\mathbb{R}^d$, ρ is a L^1 non-negative function supported on a union of cubes C_i with non-overlapping interiors, side length δ, and assume ρ is constant on each cube, with value ρ_i. Then equation* (1.1) *holds.*

Proof. Let μ^i the restriction of μ to C_i, removing any cube where ρ vanishes identically. Then from Proposition 5.3 we get $W_p(\mu^i, \Delta_N) \sim \alpha_i N^{-1/d}$ where

$$\alpha_i = \theta(d,p)(\rho_i \delta^d)^{1/p} \delta = \theta(d,p) \rho_i^{1/p} \delta^{\frac{d+p}{p}}$$

due to the homogeneity of W_p: μ_i is obtained from $\square^d$ by multiplication by $\rho_i\delta^d$ and dilation of a factor δ. By the combination lemma, we get $W_p(\mu,\Delta_N) \sim \min F(\alpha;\cdot)N^{-1/d}$ where

$$\begin{aligned}\min F(\alpha,\cdot) &= \theta(d,p)\left|\sum_i \rho_i^{\frac{d}{d+p}}\delta^d\right|^{\frac{d+p}{dp}}\\ &= \theta(d,p)|\rho|^{1/p}_{\frac{d}{d+p}}.\end{aligned}$$ □

Lemma 5.5. *Equation* (1.1) *holds whenever μ is an absolutely continuous measure defined on a compact domain of $\mathbb{R}^d$.*

Proof. For simplicity, we denote $\beta = d/(d+p)$. Let C be a cube containing the support of μ. Choose some $\varepsilon > 0$. Let $\tilde\mu = \tilde\rho\lambda$ be a measure such that $\tilde\rho$ is constant on each cube of a regular subdivision of C, is zero outside C, satisfies $|\rho-\tilde\rho|_1 \leqslant 2\varepsilon^{1+p/d}$ and such that $|\rho-\tilde\rho|_\beta \leqslant \varepsilon|\rho|_\beta$.

The stability lemma shows that

$$W_p^p(\mu,\Delta_N) \leqslant W_p^p(\tilde\mu,\Delta_{(1-\varepsilon)N}) + O\left(\frac{|\rho-\tilde\rho|_1}{2(\varepsilon N)^{p/d}}\right)$$

so that, using the hypotheses on $\tilde\rho$ and the previous lemma,

$$W_p^p(\mu,\Delta_N) \leqslant \frac{(\theta(d,p)+\varepsilon)^p|\rho|_\beta(1+\varepsilon)(1-\varepsilon)^{-p/d}+O(\varepsilon)}{N^{p/d}}$$

for N large enough.

Symmetrically, we get (again for N large enough)

$$\begin{aligned}W_p^p(\mu,\Delta_N) &\geqslant W_p^p(\tilde\mu,\Delta_{N/(1-\varepsilon)}) - O\left(\frac{|\rho-\tilde\rho|_1}{2(\varepsilon N)^{p/d}}\right)\\ &\geqslant \frac{(\theta(d,p)-\varepsilon)^p|\rho|_\beta(1-\varepsilon)^{1+p/d}-O(\varepsilon)}{N^{p/d}}.\end{aligned}$$

Letting $\varepsilon \to 0$, the claimed equivalent follows. □

Lemma 5.6. *Equation* (1.1) *holds whenever μ is an absolutely continuous measure defined on a compact domain of $\mathbb{R}^d$, endowed with any Riemannian metric.*

Proof. Denote by g the Riemannian metric, and let C be a Euclidean cube containing the support of μ. Let ε be any positive number, and choose a regular subdivision of C into cubes C_i of center p_i such that for all i, the restriction g_i of g to C_i is almost constant: $|g(p)-g(p_i)| \leqslant \varepsilon/2$ for all $p \in C_i$. Denote by $\tilde g$ the piecewise constant metric with value $g(p_i)$ on C_i. Note that even if $\tilde g$ is not continuous, at each discontinuity point x the possible choices for the metric are within a factor $\mathrm{e}^{2\varepsilon}$ one from another, and one defines that $\tilde g(x)(v,v)$ is the least of the $g(p_i)(v,v)$ over all i such that $x \in C_i$. In this way, $\tilde g$ defines a distance function close to the distance induced by g and the metric stability lemma holds with the same proof.

If one prefers not using discontinuous metrics, then it is also possible to consider slightly smaller cubes $C_i' \subset C_i$, endow C_i' with a constant metric, and interpolate the metric between the various cubes. Then one uses the L^1 stability in addition to the metric stability in the sequel.

Denote by ρ the density of μ with respect to the volume form defined by g, by μ^i the restriction of μ to C_i and by ρ_i the density of μ^i. A domain of $\mathbb{R}^d$ endowed with a constant metric is isometric to a domain of $\mathbb{R}^d$

with the Euclidean metric so that we can apply the preceding lemma to each μ^i: denoting by W'_p the Wasserstein distance computed from the metric $\tilde{g}$,

$$W'_p(\mu^i, \Delta_N) \sim \frac{\delta(d,p)|\rho_i|^{1/p}_{\frac{d}{d+p}}}{N^{1/d}}.$$

The combination lemma then ensures that $W'_p(\mu, \Delta_N) \sim \min F(\alpha; \cdot) N^{-1/d}$ where

$$\begin{aligned}\min F(\alpha, \cdot) &= \theta(d,p)\left(\sum_i \int_{C_i} \rho_i^{\frac{d}{d+p}}\right)^{\frac{d+p}{dp}} \\ &= \theta(d,p)|\rho|^{1/p}_{\frac{d}{d+p}}.\end{aligned}$$

The metric stability lemma gives

$$\mathrm{e}^{-\varepsilon} W'_p(\mu, \Delta_N) \leqslant W_p(\mu, \Delta_N) \leqslant \mathrm{e}^{\varepsilon} W'_p(\mu, \Delta_N)$$

and we only have left to let $\varepsilon \to 0$. □

We can finally end the proof of the main theorem.

Proof of Theorem 1.2. Here μ is an absolutely continuous measure defined on a compact domain D of M. Divide the domain into a finite number of subdomains D_i, each of which is contained in a chart. Using this chart, each D_i is identified with a domain of $\mathbb{R}^d$ (endowed with the pulled-back metric of M). By combination, the previous lemma shows that equation (1.1) holds.

Let us now give the asymptotic distribution of the support of any distance minimizing μ_N. Let A be any domain in M. Let x be the limit of the proportion of $\operatorname{supp}\mu_N$ that lies inside A (x exists up to extracting a subsequence). Since the domains generates the Borel σ-algebra, we only have to prove that $x = \int_A \rho^\beta / \int_M \rho^\beta$. But this follows from the combination lemma applied to the restriction of μ to A and to its complement. □

6. The dyadic Cantor measure

In this section we study the approximation problem for the dyadic Cantor measure κ to prove Theorem 1.3. Let S^0, S^1 be the dilations of ratio $1/3$ and fixed point $0, 1$. The map defined by

$$\mathcal{S} : \mu \mapsto 1/2\, S^0_\# \mu + 1/2\, S^1_\# \mu$$

is 1/3-Lipschitz on the complete metric space of probability measures having finite p-th moment endowed with the L^p Wasserstein metric. It has therefore a unique fixed point, called the dyadic Cantor measure and denoted by κ. It can be considered as the "uniform" measure on the usual Cantor set.

By convexity of the cost function and symmetry, $c_1 := W_p(\kappa, \Delta_1)$ is realized by the Dirac measure at $1/2$. Using the contractivity of $\mathcal{S}$, we see at once that $W_p(\kappa, \Delta_{2^k}) \leqslant 3^{-k} c_1$. Denote by $s = \log 2/\log 3$ the dimension of κ. We have

$$W_p(\kappa, \Delta_{2^k})(2^k)^{1/s} \leqslant c_1$$

for all integer k.

To study the case when the number of points is not a power of 2, and to get lower bounds in all cases, we introduce a notation to code the regions of $\operatorname{supp}\kappa$. Let $I^0 = [0,1]$ and given a word $w = \epsilon_n \dots \epsilon_1$ where $\epsilon_i \in \{0,1\}$, define $I^n_w = S_{\epsilon_n} S_{\epsilon_{n-1}} \dots S_{\epsilon_1}[0,1]$. The *soul* of such an interval is the open interval of one-third length with the same center. The *sons* of I^n_w are the two intervals $I^{n+1}_{w\epsilon}$ where $\epsilon \in \{0,1\}$, and an interval is

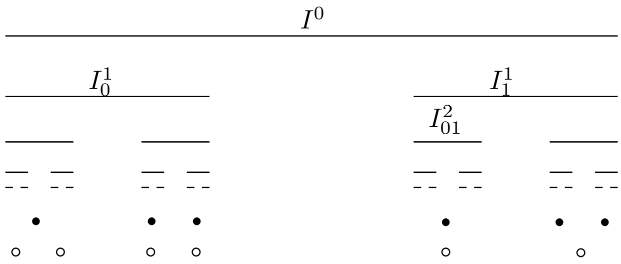

FIGURE 2. The four first steps of the construction of the Cantor set; the Cantor measure is equally divided between the intervals of a given step. The bullets show the supports of two optimal approximation of κ by 6-supported measures. We see that there is no need for the support to be equally distributed between the intervals of the first generation.

the *father* of its sons. The two sons of an interval are *brothers*. Finally, we say that n is the *generation* of the interval I^n_w.

Let N be an integer, and $\mu_N \in \Delta_N$ be a measure closest to κ, whose support is denoted by $\{x_1, \dots, x_N\}$. An interval I^n_w is said to be *terminal* if there is an x_i in its soul. A point in I^n_w is always closer to the center of I^n_w than to the center of its father. This and the optimality of μ_N implies that a terminal interval contains only one x_i, at its center.

Since the restriction of κ to I^n_w is a copy of κ with mass 2^{-n} and size 3^{-n}, it follows that

$$W_p^p(\kappa, \mu_N) = c_1^p \sum_{I^n_w} 2^{-n} 3^{-np}$$

where the sum is on terminal intervals. A simple convexity arguments shows that the terminal intervals are of at most two (successive) generations.

Consider the numbers $N_k = 3 \cdot 2^k$. The terminal intervals of an optimal μ_{N_k} must be in generations $k+1$ (for 2^k of them) and $k+2$ (for 2^{k+1} of them). Therefore

$$W_p(\kappa, \mu_{N_k})^p = c_1^p \left(3^{-(k+1)p} + 3^{-(k+2)p}\right)/2$$

and finally

$$W_p(\kappa, \Delta_{N_k}) N_k^{1/s} = c_1 \left(\frac{1+3^{-p}}{2}\right)^{1/p} 3^{\frac{\log 3}{\log 2} - 1}.$$

Note that the precise repartition of the support does not have any importance (see Fig. 2).

To see that $W_p(\kappa, \Delta_N) N^{1/s}$ has no limit, it is now sufficient to estimate the factor of c_1 in the right-hand side of the above formula. First we remark that $\left(\frac{1+3^{-p}}{2}\right)^{1/p}$ is greater than $1 - (1-3^{-p})/(2p)$ which is increasing in p and takes for $p = 1$ the value $2/3$. Finally, we compute $2/3 \cdot 3^{\frac{\log 3}{\log 2} - 1} \simeq 1.27 > 1$.

Note that the fundamental property of κ we used is that the points in a given I^n_w are closest to its center than to that of its father. The same method can therefore be used to study the approximation of sparser Cantor measure, or to some higher-dimensional analogue like the one generated by four contractions of ratio $1/4$ on the plane, centered at the four vertices of a square.

Moreover, one could study into more details the variations in the approximations $W_p(\kappa, \Delta_N)$. As said before, here our point was only to show the limitations to Theorem 1.2.

7. Link with Centroidal Voronoi Tessellations

Here we explain the link between our optimization problem and the Centroidal Voronoi Tessellations (CVTs in short). For a complete account on CVTs, the reader can consult [6] from where all definitions below are taken. Since we use the concept of barycenter, we consider only the case $M = \mathbb{R}^d$ (with the Euclidean metric). As before, λ denotes the Lebesgue measure.

7.1. **A quick presentation**

Consider a compact convex domain Ω in $\mathbb{R}^d$ and a density (positive, L^1) function ρ on Ω.

Given a N-tuple $X = (x_1, \dots, x_N)$ of so-called *generating points*, one defines the associated *Voronoi Tessellation* as the collection of convex sets

$$V_i = \big\{x \in \Omega \,\big|\, |x - x_i| \leqslant |x - x_j| \text{ for all } j \in [\![1, N]\!]\big\}$$

and we denote it by $V(X)$. One says that V_i is the *Voronoi cell* of x_i. It is a tiling of Ω, in particular the cells cover Ω and have disjoint interiors.

Each V_i has a center of mass, equivalently defined as

$$g_i = \frac{\int_{V_i} x\rho(x)\mathrm{d}x}{\int_{V_i} \rho(x)\mathrm{d}x}$$

or as the minimizer of the energy functional

$$\mathcal{E}_{V_i}(g) = \int_{V_i} |x - g|^2 \rho(x)\mathrm{d}x.$$

One says that $(V_i)_i$ is a *Centroidal Voronoi Tessellation* or *CVT*, if for all i, $g_i = x_i$. The existence of CVTs comes easily by considering the following optimization problem: search for a N-tuple of points $X = (x_1, \dots, x_N)$ and a tiling V of Ω by N sets $V_1, \dots, V_N$ which together minimize

$$\mathcal{E}_V(X) = \sum_{i=1}^{N} \mathcal{E}_{V_i}(x_i).$$

A compacity argument shows that such a minimizer exists, so let us explain why a minimizer must be a CVT together with its generating set. First, each x_i must be the center of mass of V_i, otherwise one could reduce the total energy by moving x_i to g_i and changing nothing else. But also, V_i should be the Voronoi cell of x_i, otherwise there is a $j \neq i$ and a set of positive measure in V_i whose points are closest to x_j than to x_i. Transferring this set from V_i to V_j would reduce the total cost.

We observe that this optimization problem is exactly that of approximating the measure $\rho\lambda$ in L^2 Wasserstein distance; more precisely, finding the N-tuple x that minimizes $\inf_V \mathcal{E}_V(X)$ is equivalent to finding the support of an optimal $\mu_N \in \Delta_N$ closest to $\rho\lambda$, and then the Voronoi tessellation generated by X gives the mass of μ_N at each x_i and the optimal transport from $\rho\lambda$ to μ_N.

One says that a CVT is *optimal* when its generating set is a global minimizer of the energy functional

$$\mathcal{E}(X) = \mathcal{E}_{V(X)}(X).$$

Optimal CVTs are most important in applications, which include for example mesh generation and image analysis (see [6]).

7.2. Equidistribution of energy

The *principle of energy equidistribution* says that if X generates an optimal CVT, the energies $\mathcal{E}_{V_i}(x_i)$ of the generating points should be asymptotically independent of i when N goes to ∞.

Our goal here is to deduce a mesoscopic version of this principle from Theorem 1.2. A similar result holds for any exponent, so that we introduce the L^p energy functionals $\mathcal{E}^p_{V_i}(x_i) = \int_{V_i} |x-x_i|^p \rho(x)\mathrm{d}x$, $\mathcal{E}^p_V(X) = \sum_i \mathcal{E}^p_{V_i}(x_i)$ and $\mathcal{E}^p(X) = \inf_V \mathcal{E}^p_V(x) = \mathcal{E}^p_{V(X)}(X)$. In particular, an optimal X for this last functional is the support of an element of Δ_N minimizing the L^p Wasserstein distance to $\rho\lambda$.

Note that for $p \neq 2$ an x minimizing $\mathcal{E}^p(x)$ need not generate a CVT, since the minimizer of $\mathcal{E}^p_{V_i}$ is not always the center of mass of V_i (but it is unique as soon as $p > 1$).

Corollary 7.1. *Let A be a cube of Ω. Let $X^N = \{x^N_1, \dots, x^N_N\}$ be a sequence of N-sets minimizing $\mathcal{E}^p$ for the density ρ, and denote by $\bar{\mathcal{E}}^p_A(N)$ the average energy of the points of X^N that lie in A. Then*

$$\bar{\mathcal{E}}^p_A(N) N^{\frac{d+p}{d}}$$

has a limit when $N \to \infty$, and this limit does not depend on A.

The cube A could be replaced by any domain, but not by any open set. Since the union of the X_N is countable, there are indeed open sets of arbitrarily small measure containing all the points $(x^N_i)_{N,i}$.

Proof. Fix some $\varepsilon > 0$ and let $A' \subset A$ be the set of points that are at distance at least ε from $\Omega \setminus A$ and by $A'' \supset A$ the set of points at distance at most ε from A.

First, the numbers N', N'' of points of X^N in A', A'' satisfy

$$N' \sim N \frac{\int_{A'} \rho^{d/(d+p)}}{\int_\Omega \rho^{d/(d+p)}} \qquad N'' \sim N \frac{\int_{A''} \rho^{d/(d+p)}}{\int_\Omega \rho^{d/(d+p)}}.$$

The localization lemma implies that the maximal distance by which mass is moved by the optimal transport between $\rho\lambda$ and the optimal X^N-supported measure tends to 0, so that for N large enough the energy of all points in A is at least the minimal cost between $\rho_{|A'}\lambda$ and $\Delta_{N'}$ and at most the minimal cost between $\rho_{|A''}\lambda$ and $\Delta_{N''}$.

Letting $\varepsilon \to 0$ we thus get that the total energy of all points of X^N lying in A is equivalent to

$$\theta(d,p) \frac{\left(\int_A \rho^{d/(d+p)}\right)^{(d+p)/d}}{\left(N \int_A \rho^{d/(d+p)}\right)^{p/d}} = \theta(d,p) N^{-p/d} \int_A \rho^{d/(d+p)}.$$

As a consequence we have $\bar{\mathcal{E}}_A(N) \sim \left(\theta(d,p) \int_\Omega \rho^{d/(d+p)}\right) N^{-(d+p)/d}$. □

Acknowledgements. I am grateful to Romain Joly, Vincent Munnier, Hervé Pajot, Rémy Peyre and Cédric Villani for interesting discussions or comments.

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] 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.

[3] T. Champion, L. De Pascale and P. Juutinen, The ∞-Wasserstein distance: local solutions and existence of optimal transport maps. *SIAM J. Math. Anal.* **40** (2008) 1–20.

[4] V. Dobrić and J.E. Yukich, Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.* **8** (1995) 97–118.

[5] Q. Du and D. Wang, The optimal centroidal Voronoi tessellations and the Gersho's conjecture in the three-dimensional space. *Comput. Math. Appl.* **49** (2005) 1355–1373.

[6] Q. Du, V. Faber and M. Gunzburger, Centroidal Voronoi tessellations: applications and algorithms. *SIAM Rev.* **41** (1999) 637–676.
[7] K.J. Falconer, *The geometry of fractal sets*, *Cambridge Tracts in Mathematics* **85**. Cambridge University Press (1986).
[8] L. Fejes Tóth, Sur la représentation d'une population infinie par un nombre fini d'éléments. *Acta. Math. Acad. Sci. Hungar* **10** (1959) 299–304.
[9] L. Fejes Tóth, *Lagerungen in der Ebene, auf der Kugel und im Raum*, *Die Grundlehren der mathematischen Wissenschaften, Band* **65**. Zweite verbesserte und erweiterte Auflage, Springer-Verlag (1972).
[10] S. Graf and H. Luschgy, *Foundations of quantization for probability distributions*, *Lecture Notes in Mathematics* **1730**. Springer-Verlag (2000).
[11] J. Heinonen, *Lectures on analysis on metric spaces.* Universitext, Springer-Verlag (2001).
[12] J. Horowitz and R.L. Karandikar, Mean rates of convergence of empirical measures in the Wasserstein metric. *J. Comput. Appl. Math.* **55** (1994) 261–273.
[13] J.E. Hutchinson, Fractals and self-similarity. *Indiana Univ. Math. J.* **30** (1981) 713–747.
[14] F. Morgan and R. Bolton, Hexagonal economic regions solve the location problem. *Amer. Math. Monthly* **109** (2002) 165–172.
[15] S.J.N. Mosconi and P. Tilli, Γ-convergence for the irrigation problem. *J. Convex Anal.* **12** (2005) 145–158.
[16] D.J. Newman, The hexagon theorem. *IEEE Trans. Inform. Theory* **28** (1982) 137–139.
[17] C. Villani, *Topics in optimal transportation*, *Graduate Studies in Mathematics* **58**. American Mathematical Society (2003).