Sapienza University of Rome

Department of Mathematics Guido Castelnuovo
PhD in Mathematics

Thesis For The Degree Of Doctor Of Philosophy

Random Euclidean Matching for exponentially and polynomially decaying densities

Thesis Advisor
Prof. Emanuele Caglioti

Candidate
Francesca Pieroni
1801532

Academic Year 2024-2025 (XXXVIII cycle)

Abstract.

This thesis is devoted to the study of the Random Euclidean Matching problem in the case of exponentially and polynomially decaying densities. Such problem consists in finding the optimal matching between a sequence of independent random variables with common probability distribution μ taking values in a domain $\Omega \subseteq \mathbb{R}^d$ and their reference measure. Such points are denoted by $\{X_i\}_{i=1}^n$. Here *optimal* refers to the cost function determined by the power p of the Euclidean distance. More precisely, the problem is to estimate for very large n the expectation $\mathbb{E}$ of the cost function determined by the p-Wasserstein distance between two measures: the first one is the empirical measure on the points X_i and the second one is $n\mu$. Briefly speaking, one has to match each point X_i to a region $\Omega_i \subseteq \Omega$ such that $\mu(\Omega_i) = \frac{1}{n}$ in such a way to minimize the Euclidean distance.

There are two reasons why one could be interested in studying such quantity. The first one is that it is the semi-discrete version of a problem in which one has to match a sequence of iid points X_i to another sequence of iid points Y_j. But it creates much less technical difficulties since in our case one of the two measures is absolutely continuous with respect to the Lebesgue measure. Furthermore, the problem is interesting in its own right because it is related to the estimation of the averaged Wasserstein distance W_p between n points sampled from the distribution μ (the empirical measure) and the distribution itself.

Such problem has been deeply studied in the case of variables distributed on bounded connected domains with uniformly positive probability densities, but the results concerning the case of points distributed according to a probability density supported in $\mathbb{R}^d$ were much less. There were partial results and conjectures for the Gaussian density in the case $p \leq d$ and no conjectures at all for $p > d$. This thesis investigates that direction: we focus on radial probability densities ρ supported in $\mathbb{R}^d$ with an exponential and polynomial tail. The Gaussian density is one of these.

Our purpose is to prove asymptotic estimates for the expectation $\mathbb{E}$ of the cost function we described above in the case of iid points distributed with a probability measure whose density has an exponential and polynomial tail. The strategy of the proof is to adapt the results in the case of bounded connected domains Ω.

In the very first part, we focus on the rigorous definition of the problem and on the state of the art.

Then, we recall some basic notions and we prove the result we mostly need. In particular, since the Random Matching problem finally turns out to be an Optimal Transport problem, in the following we focus on the couplings between two measures, the Optimal Transport maps and the Benamou-Brenier formula, which leads back the Wasserstein distance to a PDEs problem.

Another part is devoted to the concentration properties of binomial random variables, since, given a sequence of iid random variables, the number of them in a fixed region is finally a binomial random variable.

Then, since one of the main difficulties in the strategy we will use is the concept of *geometric decomposition of the domain*, we will need also some extra tools, which we will focus on later. Indeed the geometric decomposition in our case is performed exploiting the symmetries of the density. In particular, to partition the space in much smaller regions, we use the Exponential Map, that projects the tangent space to a manifold (in our case, the sphere) on the manifold itself, and the Haar measure, i.e., the left and right-invariant measure, on the orthogonal group. Therefore some Chapters are dedicated to this.

In the very last Chapter we prove our main result: we find a critical exponent p^{cric} at which the rate of the expectation of the cost function for exponentially and polynomially decaying densities changes abruptly. Such p^{cric} depends on the density itself and on the dimension. As we said before, the way we prove it is inspired by the results in the case of bounded and connected domains, but, in order to adapt the proof to an unbounded domain and to a probability density vanishing for large $|x|$, we need to exploit the radial symmetry of the density. Indeed all the estimates used in the compact case fail in our setting. Our result is the following: we determine the rate of the cost function for any p and d and, for $p \leq p^{crit}$, we prove some more precise estimates about the limit of the cost function divided by its rate.

Contents

Acknowledgments, Ringraziamenti.

I really would like to express my gratitude to some people without whom this thesis, and generally speaking these three years of PhD, would have not existed.

First, I would like to thank my advisor, Emanuele Caglioti. Thank you for having been my advisor, also for the Bachelor and the Master theses, for having introduced me this problem and having shared with me your physical intuitions about the topic, for the willingness to speak about the subject and my doubts at any time, the advices within the drafting of this thesis, the (answered and not) questions, the patience and the support within these three years. Thank you also for having taught me that a sequence of vain attempts in studying a problem is not necessarily a failure.

I also would like to thank some other people I've been honored to work with during the development of this thesis: Michael Goldman and Dario Trevisan. Thank you Michael for having shared with me your knowledge of the subject and offered me a different perspective, for your advices (both strictly mathematical and not), your availability and your encouragements in a lot of occasions. Thank you Dario for having spoken to me about the subject through your probabilistic point of view (almost completely unknown to me), for having involved me in your mathematical problems, for your patience and your attention to the possible developments of my research. Also thank both of you for having included me in the Galileo project and for the useful discussions we shared about the random matching for polynomial densities.

I am really grateful to other people I have been introduced to and worked with in these three years. Thanks to Chiara Boccato, for having shared with me your knowledge, especially in the quantum setting, and having found the quantum optimal transport as a contact point with me (thanks also to Dario for this). Also thanks to Sergio Simonella for having introduced me the Lorentz Gas (thanks Emanuele too), and for your availability to speak about related problems at any time. I am also very grateful to François Golse for the time you spent with me, and also with Emanuele and Sergio, speaking about Boltzmann equation and other related topics.

I also would like to thank the other members of the GEP (Gender Equality Plan) Commission for the hard work we made together in the last year and for the mutual support. Another special thanks is for all the participants in the tutoring projects in penitentiary institutions: sharing this opportunity with all of you has been a very enriching experience.

The other people I am grateful to are my family and my friends, so I am going to write the leading part of these acknowledgments in Italian.

Un grazie sincero va alla mia famiglia per tutto il supporto, non solo in questi tre anni: a mamma, un poc(hin)o apprensiva, grazie per sopportare sempre i miei nervosismi, perché, anche se "le mamme hanno i loro diritti", alla fine accetti sempre serenamente le mie idee stravaganti, ma anche per essere l'esempio vivente che con la forza di volontà si può fare tutto; a papi, per i continui tentativi di accompagnarmi all'aeroporto, o dove che sia, per le scorte infinite in tema di spesa e per avermi ascoltato in momenti in cui forse avevi più bisogno tu di ascolto; a Eli, perché per colpa tua sono quasi vegana (anche se nel mio frigo aleggia ancora del formaggio), per aver sopportato sveglie a qualsiasi ora del giorno e della notte, e perché i rapporti stretti a volte hanno i litigi come effetto collaterale, ma ne vale la pena; e a Fede per la tua presenza silenziosa, i film d'azione che con te diventano improvvisamente guardabili e le tue uscite esplosive. Ma soprattutto, dopo tre ore spese a cercare di capire come fare un frontespizio, grazie per l'idea a me completamente non ovvia di comporlo a parte e poi incollarlo alla tesi di forza. Grazie a nonna, per i cappuccini del venerdì pomeriggio e per essere il porto sicuro di ogni più vario momento di crisi. Grazie a Giulia per le confidenze e i consigli reciproci sui saponi anti-acne, a Luisa, e al resto della tribù, che qui non ho spazio per inserire.

Grazie alle amicizie preziose trovate all'università: a tutto il gruppo del τ_{84}(Capodanno), ma anche a Maria Chiara, Sergio, Federica e Lisa.

Grazie a Isabelle, Paola, Veronica e Jeannette per le parole giuste al momento giusto.

Grazie a Manuela, per avermi insegnato che a volte cambiare non vuol dire perdersi ma ritrovarsi e per avermi detto, quando potevo coglierlo, che siamo noi il nostro posto sicuro.

Grazie all'auletta dottorandi, fonte inesauribile di rumore e di conseguente voglia di non lavorare: nel dire questo, mi assumo tutte le mie responsabilità. Non mi dilungo con l'elenco dei nomi perché a questo punto diventerebbero veramente troppi, ma ci tengo a precisare che in quell'auletta mi sono stati chiariti svariati dubbi sulla mappa esponenziale (e ringrazio chi di dovere per questo).

Chapter 1

Introduction.

The main topic of this thesis is the the Random Matching problem for distributions with unbounded support. The base example is the Gaussian distribution, but we consider the general case of radial distributions that decay exponentially and polynomially. We first define the random matching problem, then we focus both on the case of uniform and not uniform densities on bounded and connected domains. Finally we focus on the case of unbounded domains, and we come to the exponential and polynomial densities.

1.1 The random matching problem: the state of the art.

The random matching problem is the problem of finding the best coupling between two independent samples of iid random variables. More precisely, the problem is stated as follows: one has two sequences of independent random variables $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ with a probability distribution μ on $\mathbb{R}^d$ (let us denote them by variables of type X and of type Y), and one defines a cost function as follows. If one denotes

$$\mathcal{S}_n := \{\text{permutations of } \{1, \dots, n\}\},$$

the cost function $C_n^{p,d}$, with $p \geq 1$, is defined as

$$C_n^{p,d} := \min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^n |X_i - Y_{\sigma_i}|^p. \tag{1.1.0.1}$$

In other words, one has to match each random variable of type X with only one random variable of type Y, and then take the sum of the power p of the distances. Of course, the previous quantity depends on the realization of the variables X and Y, and therefore one is interested in the average of the cost function. Therefore, if $\mathbb{E}$ is the expected value with respect to the law of the random variables, the quantity one investigates is

$$\mathbb{E}[C_n^{p,d}], \qquad \text{for } n \gg 1.$$

One can reformulate the previous problem in terms of Wasserstein distance between two measures μ and ν, which we will define rigorously in Section 2.1. Briefly speaking, such distance is the infimum of the quantities $\|x - y\|_{L^p(\Omega\times\Omega, d\pi)}$ among all the possible couplings π of μ and ν.

It is possible to show (see for instance [10]) that

$$C_n^{p,d} = W_p^p\left(\sum_{i=1}^n \delta_{X_i}, \sum_{i=1}^n \delta_{Y_i}\right). \tag{1.1.0.2}$$

Let us spend a few words about the equality in (1.1.0.2): the upper bound is immediate to prove, indeed it is enough to notice that for any $\sigma \in \mathcal{S}_n$ the measure

$$\pi_\sigma := \sum_{i=1}^n \delta_{X_i, Y_{\sigma_i}},$$

is a coupling of $\sum_{i=1}^n \delta_{X_i}$ and $\sum_{i=1}^n \delta_{Y_i}$, and then to minimize among $\sigma \in \mathcal{S}_n$. The validity of the lower bound reflects the fact that the best coupling one can find between such measures is, in fact, a one to one matching between the points.

This problem has been studied by several authors, and we are going to focus on the state of the art starting with the uniform case (we are going to generalize later).

1.1.1 The uniform matching.

By this expression we mean that the problem taken into account is the case in which the points are independent and uniformly distributed on $[0,1]^d$.

The scaling hypothesis. An heuristic argument in [29] explains why one would expect that, at least in the case of iid random variables uniformly distributed on the hypercube $[0,1]^d$, the quantity above should have order $n^{1-\frac{p}{d}}$, that is, for a constant $C>0$ one should expect that

$$\mathbb{E}[C^{p,d}_{n,[0,1]^d,dx}] \approx n^{1-\frac{p}{d}}, \tag{1.1.1.1}$$

where here and later $\approx$ means that for any $n \in \mathbb{N}$ the left hand side is bounded from above and below by the right hand side (for some positive constants). We added to the notation of the cost the domain where the points are distributed in (in this case $[0,1]^d$) and their distribution (in this case dx, the uniform measure).

The argument why one should expect the behavior in (1.1.1.1) is the following: the averaged distance between a point of type X and the associated point of type Y should have order $n^{-\frac{1}{d}}$, since one has n points in a space of measure 1 and in dimension d. Hence, taking the power p and then summing over i one gets $n^{1-\frac{p}{d}}$. In fact, this does not happen in dimensions $d=1$ and $d=2$, but it holds if $d \geq 3$. That is, in the case of uniform iid random variables, the rate is the following:

$$\mathbb{E}[C^{p,d}_{n,[0,1]^d,dx}] \approx f_n := n^{1-\frac{p}{d}} \begin{cases} n^{\frac{p}{2}} & d=1, \\ (\log n)^{\frac{p}{2}} & d=2, \\ 1 & d \geq 3. \end{cases} \tag{1.1.1.2}$$

The case $d=1$ has been deeply studied in [14] (see also [3, 4, 8]), and the scaling of the expectation of the cost can be computed by direct inspection (the optimal matching is the monotone one in dimension 1).

We will mostly focus on dimension $d \geq 2$.

As for the case $d=2$, the presence of the extra logarithmic term has been proved in [3].

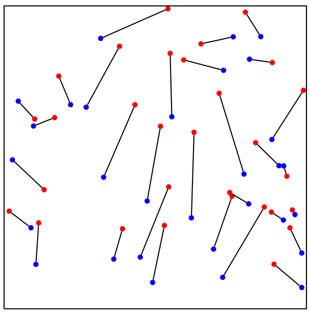

Figure 1.1.1.1: Optimal matching on $[0,1]^2$ with $n=29$.

The case $d \geq 3$ has been studied in [5, 16, 17].

In dimension $d=1$ and $d=2$ the averaged cost is much larger than expected, and this fact can be understood as the proof that the typical distance between a point of type X and the associated point of type Y has order $n^{-\frac{1}{2}}$ in dimension $d=1$ (instead of n^{-1}) and $n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}$ in dimension $d=2$ (instead of $n^{-\frac{1}{2}}$). Briefly speaking, this difference between the small dimensions and the higher ones is due to the variance of the number of points in each subset of the hypercube $[0,1]^d$.

Let us point out that the validity of (1.1.1.2) has been proved in [25, Section 4] even in the case of iid points on a compact Riemannian manifold such as, for instance, the $d-1$-dimensional sphere.

The existence of the limit. The case $p=2$ has a special role in optimal transport, hence we begin with focusing on that.

If $d=1$ the cost can be computed explicitly and on the interval one has

$$\lim_{n\to\infty} \mathbb{E}[C^{2,1}_{n,[0,1],dx}] = \frac{1}{6}. \tag{1.1.1.3}$$

The same can be done replacing $[0,1]$ by $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$ and $\frac{1}{6}$ by $\frac{1}{12}$ (see [14]). The reason why the cost on the the one dimensional torus is asymptotically smaller than the cost on the segment is that the segment can be closed to get a torus, hence the distance between close points decreases.

As for the case $p=d=2$, we mentioned in equation (1.1.1.2) that, if the points are independent and uniformly distributed on $[0,1]^2$, then

$$\mathbb{E}[C^{2,2}_{n,[0,1]^2,dx}] \approx \log n,$$

and later it has been conjectured in [13] that

$$\lim_{n\to\infty} \frac{\mathbb{E}[C^{2,2}_{n,[0,1]^2,dx}]}{\log n} = \frac{1}{2\pi}, \tag{1.1.1.4}$$

even if one substitutes $[0,1]^2$ with $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$. This conjecture is based on optimal transport tools and it has been proved but, before stating the complete result, we have to mention another problem that has been introduced. That is, one can consider

$$\tilde{C}^{p,d}_n := W^p_p\left(\sum_{i=1}^n \delta_{X_i}, n\mu\right), \tag{1.1.1.5}$$

where μ is the uniform measure on $[0,1]^2$. This can be done in general, that is, for any sequence of iid random variables $\{X_i\}_{i=1}^\infty$ with common distribution μ, by the law of large numbers the quantity $\frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ converges in probability to the reference measure. Hence one can substitute one of the two sums of delta functions with

$$\sum_{i=1}^n \delta_{Y_i} = n\frac{1}{n}\sum_{i=1}^n \delta_{Y_i} \simeq n\mu, \qquad \text{for } n \gg 1.$$

The point of doing this is that from a technical point of view the estimate of the distance between two measures one of whom is absolutely continuous with respect to the Lebesgue measure is easier than dealing with two singular measures. This problem is sometimes referred as the semi-discrete matching problem, and it corresponds to associate to each point X_i a region of $[0,1]^2$ (or of $\Omega \subseteq \mathbb{R}^d$ if $\{X_i\}_{i=1}^\infty$ take values in Ω), instead of another point Y_j.

The rate (in n) is exactly the same as in equation (1.1.1.2), only the constant changes. Hence, when we refer to one or the other the estimates about the rate estimates are the same and we omit to say it.

Theorem 1.1.1 ([4, 1]). *Let $\{X_i, Y_i\}_{i=1}^\infty$ be a sequence of iid random variables uniformly distributed on a compact Riemannian manifold Σ with no boundary and measure 1, or on $[0,1]^2$. Then, if $C^{p,d}_n$ and $\tilde{C}^{p,d}_n$ are the cost functions defined respectively in (1.1.0.1) and (1.1.1.5) and m is the uniform measure on Σ, it holds*

$$\lim_{n\to\infty} \frac{\mathbb{E}[C^{2,2}_{n,\Sigma,\mathsf{m}}]}{\log n} = \frac{1}{2\pi} \quad \text{and} \quad \lim_{n\to\infty} \frac{\mathbb{E}[\tilde{C}^{2,2}_{n,\Sigma,\mathsf{m}}]}{\log n} = \frac{1}{4\pi}. \tag{1.1.1.6}$$

Concerning the previous Theorem, in [1] the authors have proved the same results as [4] with different methods. Therefore some estimates have been improved: indeed the authors proved that

$$\left|\mathbb{E}[\tilde{C}^{2,2}_{n,\Sigma,\mathsf{m}}] - \frac{\log n}{4\pi}\right| \lesssim \sqrt{\log n \log\log n},$$

where in the previous expression and in the following the symbol $\lesssim$ means that the left hand side is bounded by the right-hand side for a positive constant. Such estimate has been further improved in [21], as follows:

$$\left|\mathbb{E}[\tilde{C}^{2,2}_{n,\Sigma,\mathsf{m}}] - \frac{\log n}{4\pi}\right| \lesssim \log\log n.$$

Moreover, for different exponents $p \geq 1$ and dimensions $d \geq 3$ there are results which exploit subadditivity arguments and state that the limit of the expectation of the cost function in the uniform case divided by the correct rate does exist (see [16, 9, 5, 23]), but, up to now, it is not known its value.

We shall denote by c^p_d and $\tilde{c}^p_d$ the limits of the expectations of the cost divided by the correct rate respectively for the bipartite matching and for semi-discrete one, that is

$$c^p_d := \lim_{n\to\infty} \frac{\mathbb{E}[C^{p,d}_n]}{f_n} \quad \text{and} \quad \tilde{c}^p_d := \lim_{n\to\infty} \frac{\mathbb{E}[\tilde{C}^{p,d}_n]}{f_n}, \tag{1.1.1.7}$$

with f_n defined by (1.1.1.2).

Let us point out that, using this notation, thanks to (1.1.1.3), (1.1.1.4) and (1.1.1.6) one has

$$c^2_1 = \frac{1}{6}, \quad c^2_2 = \frac{1}{2\pi} \quad \text{and} \quad \tilde{c}^2_2 = \frac{1}{4\pi}.$$

1.1.2 The not-uniform case.

Another case has been later considered, that is the case in which the random variables are iid on $\Omega \subseteq \mathbb{R}^d$ with a probability distribution $d\mu = \rho dx$ absolutely continuous with respect to the Lebesgue measure. We begin with the case $p = d = 2$, which is the only one in which the limiting constant is explicit.

The case $p = d = 2$. It has been proved that if the points of type X and Y are iid distributed in a bounded and connected domain $\Omega \subseteq \mathbb{R}^2$ with Lipschitz boundary and according to a probability measure $d\mu = \rho dx$, with ρ α-Hölder and uniformly bounded from above and below by strictly positive constants, then the expectation of the cost function is asymptotically the same as the uniform case except for a correction constant due to the volume of the domain. That is, the following result holds.

Theorem 1.1.2 ([7, 2]). *Let $\{X_i, Y_i\}_{i=1}^{\infty}$ be a sequence of iid random variables on $\Omega \subseteq \mathbb{R}^2$ with probability distribution $d\mu = \rho dx$, where Ω is a bounded and connected domain with Lipschitz boundary and ρ is α-Hölder and such that $\frac{1}{C} \leq \rho \leq C$, for $C > 0$. Then*

$$\frac{\mathbb{E}[C^{2,2}_{n,\Omega,\mu}]}{\log n} \xrightarrow[n\to\infty]{} \frac{|\Omega|}{2\pi}, \qquad \textit{and} \qquad \frac{\mathbb{E}[\tilde{C}^{2,2}_{n,\Omega,\mu}]}{\log n} \xrightarrow[n\to\infty]{} \frac{|\Omega|}{4\pi}. \tag{1.1.2.1}$$

Let us spend a few words about this result. The idea used by the authors is the following (we use the case of bipartite matching but the idea is the same): if one decomposes Ω in a disjoint union of squares $\{Q_j\}_j$ (in fact, one has to use a Whitney partition), denoting by $\sigma \in \mathcal{S}_n$ the optimal assignment, one has

$$\mathbb{E}[C^{2,2}_{n,\Omega,\mu}] = \mathbb{E}\left[\sum_{i=1}^{n} |X_i - Y_{\sigma_i}|^2\right] \simeq \sum_j \mathbb{E}\left[\sum_{i:X_i,Y_{\sigma_i}\in Q_j} |X_i - Y_{\sigma_i}|^2\right], \tag{1.1.2.2}$$

where in the approximation one pretends that the contribution of the points matched with other points outside the same square is negligible.

Here and later the symbol $\simeq$ is used to mean that the left hand side divided by the right hand side is asymptotic to 1 for large n.

If one chooses the squares small enough to transport the points distribution $\mu|_{Q_j}$ into the uniform one (and this can be done if ρ is regular enough), in the right hand side of the expression above one finds the cost of the bipartite matching of iid random variables uniformly distributed on Q_j, which is a square of side L_j. Hence, if

$$Q_j = q_j + [0, L_j]^2,$$

one has

$$|X_i - Y_{\sigma_i}|^2 = L_j^2 \left| \frac{X_i - q_j}{L_j} - \frac{Y_{\sigma_i} - q_j}{L_j} \right|^2 = |Q_j| |\tilde{X}_i - \tilde{Y}_{\sigma_i}|^2, \tag{1.1.2.3}$$

where the points $\tilde{X}$ and $\tilde{Y}$ are iid and (almost) uniformly distributed on $[0,1]^2$ if X and Y are (almost) uniformly distributed on Q_j.

The number of points of type X and Y in Q_j is a random variable too (let us denote by $n(Q_j)$ the number of points of type X in Q_j) and its expectation is

$$\mathbb{E}[n(Q_j)] = n\mu(Q_j) = n \int_{Q_j} dx \rho(x). \tag{1.1.2.4}$$

Hence, combining (1.1.2.3) and (1.1.2.4), one can approximate

$$\mathbb{E}\left[\sum_{i:X_i,Y_{\sigma_i}\in Q_j} |X_i - Y_{\sigma_i}|^2\right] \simeq |Q_j| \mathbb{E}\left[C^{2,2}_{n(Q_j),[0,1]^2,dx}\right] \simeq |Q_j| \mathbb{E}\left[\frac{1}{2\pi}\log(n(Q_j))\right] \simeq \frac{|Q_j|}{2\pi}\log(n\mu(Q_j)), \tag{1.1.2.5}$$

in such a way to get by (1.1.2.2)

$$\mathbb{E}[C^{2,2}_{n,\Omega,\mu}] \simeq \sum_j \frac{|Q_j|}{2\pi}\log(n\mu(Q_j)) = \sum_j \frac{|Q_j|}{2\pi}\log n + \sum_j \frac{|Q_j|}{2\pi}\log(\mu(Q_j)) = \frac{|\Omega|}{2\pi}\log n + C(\Omega),$$

where $C(\Omega)$ is a constant depending on Ω and on μ. This estimate is the reason why, up to some approximations, one should expect the behavior in (1.1.2.1).

The case $d \geq 3$ and $p \geq 1$. The same argument has been used to prove an analogue estimate for $d \geq 3$ and any exponent $p \geq 1$, where one has the following result.

Theorem 1.1.3 ([5, 24]). *Let $\{X_i, Y_i\}_{i=1}^{\infty}$ be iid random variables distributed on $\Omega \subseteq \mathbb{R}^d$ $(d \geq 3)$ with probability distribution $d\mu = \rho dx$, where Ω is a bounded and connected domain with Lipschitz boundary and ρ is α-Hölder and such that $\frac{1}{C} \leq \rho \leq C$, for $C > 0$. Then, for any $p \geq 1$, it holds*

$$\limsup_{n\to\infty} \frac{\mathbb{E}[C^{p,d}_{n,\Omega,\mu}]}{n^{1-\frac{p}{d}}} \leq c^p_d \int_\Omega \rho^{1-\frac{p}{d}} \qquad \textit{and} \qquad \limsup_{n\to\infty} \frac{\mathbb{E}[\tilde{C}^{p,d}_{n,\Omega,\mu}]}{n^{1-\frac{p}{d}}} \leq \tilde{c}^p_d \int_\Omega \rho^{1-\frac{p}{d}}, \tag{1.1.2.6}$$

with c^p_d and $\tilde{c}^p_d$ in (1.1.1.7).

The argument is the same as before: as in (1.1.2.2), one gets

$$\mathbb{E}[C^{p,d}_{n,\Omega,\mu}] \simeq \sum_j \mathbb{E}\left[\sum_{i:X_i,Y_{\sigma_i}\in Q_j} |X_i - Y_{\sigma_i}|^p\right], \tag{1.1.2.7}$$

and assuming again that

$$Q_j = q_j + [0, L_j]^d,$$

as in (1.1.2.3) one has

$$|X_i - Y_{\sigma_i}|^p = L_j^p \left|\frac{X_i - q_j}{L_j} - \frac{Y_{\sigma_i} - q_j}{L_j}\right|^p = |Q_j|^{\frac{p}{d}} |\tilde{X}_i - \tilde{Y}_{\sigma_i}|^p, \tag{1.1.2.8}$$

where the points $\tilde{X}$ and $\tilde{Y}$ are iid and (almost) uniformly distributed on $[0,1]^d$ if X and Y are (almost) uniformly distributed on Q_j (if the size of Q_j is small enough). Hence, combining (1.1.2.8) with (1.1.2.7), one gets

$$\begin{aligned}\mathbb{E}[C^{p,d}_{n,\Omega,\mu}] &\simeq \sum_j \mathbb{E}\left[\sum_{i:X_i,Y_{\sigma_i}\in Q_j} |X_i - Y_{\sigma_i}|^p\right] = \sum_j |Q_j|^{\frac{p}{d}} \mathbb{E}\left[C^{p,d}_{n(Q_j),[0,1]^d,dx}\right] \simeq \sum_j |Q_j|^{\frac{p}{d}} \mathbb{E}[c_d^p n(Q_j)^{1-\frac{p}{d}}] \\ &\simeq c_d^p \sum_j |Q_j|^{\frac{p}{d}} (n\mu(Q_j))^{1-\frac{p}{d}} = c_d^p n^{1-\frac{p}{d}} \sum_j |Q_j| \left(\frac{\mu(Q_j)}{|Q_j|}\right)^{1-\frac{p}{d}} \simeq c_d^p n^{1-\frac{p}{d}} \int_\Omega dx \rho(x)^{1-\frac{p}{d}}, \end{aligned}\tag{1.1.2.9}$$

and this explains why one should expect (1.1.2.6).

Let us point out that this argument can be made rigorous for the upper bound, thanks to the property of subadditivity of the Wasserstein distance (see (2.1.0.2)). We now explain how does it work at least in the case of the semi-discrete matching, since the bipartite one is very similar but a little bit more technical. If $\{Q_j\}_j$ is the decomposition of Ω in hypercubes, by triangle inequality one gets

$$\begin{aligned}\mathbb{E}[C^{p,d}_{n,\Omega,\mu}] &= \mathbb{E}\left[W_p^p\left(\sum_{i=1}^n \delta_{X_i}, n\mu\right)\right] \le \mathbb{E}\left[\left(W_p\left(\sum_{i=1}^n \delta_{X_i}, \sum_j \frac{n(Q_j)}{\mu(Q_j)}\mu \mathbb{1}_{Q_j}\right) + W_p\left(\sum_j \frac{n(Q_j)}{\mu(Q_j)}\mu \mathbb{1}_{Q_j}, n\mu\right)\right)^p\right] \\ &\le (1+\eta)\mathbb{E}\left[W_p^p\left(\sum_{i=1}^n \delta_{X_i}, \sum_j \frac{n(Q_j)}{\mu(Q_j)}\mu \mathbb{1}_{Q_j}\right)\right] + \frac{c}{\eta^{p-1}}\mathbb{E}\left[W_p^p\left(\sum_j \frac{n(Q_j)}{\mu(Q_j)}\mu \mathbb{1}_{Q_j}, n\mu\right)\right]. \end{aligned}\tag{1.1.2.10}$$

In the previous expression, η is finally to be sent to 0.

As we said before, the first summand in (1.1.2.10) has been estimated by the authors thanks to the subadditivity of the Wasserstein distance, indeed both the measures involved in it can be written as the sum over j of measures concentrated on Q_j (and with the same mass on Q_j, that is, $n(Q_j)$), as follows:

$$\mathbb{E}\left[W_p^p\left(\sum_{i=1}^n \delta_{X_i}, \sum_j \frac{n(Q_j)}{\mu(Q_j)}\mu \mathbb{1}_{Q_j}\right)\right] \le \sum_j \mathbb{E}\left[W_p^p\left(\sum_{i:X_i\in Q_j} \delta_{X_i}, \frac{n(Q_j)}{\mu(Q_j)}\mu \mathbb{1}_{Q_j}\right)\right].$$

Such term is usually referred to as local term and, up to choosing the squares Q_j small enough, we can approximate each term in the previous sum with the analogue quantity for uniformly distributed variables on Q_j. This produces the pre-factor $\int_\Omega \rho^{1-\frac{p}{d}}$ and the logarithmic correction in dimension $d = 2$. Briefly speaking, this is the rigorous way to get (1.1.2.5) and (1.1.2.9).

The second summand in (1.1.2.10) is a little bit more difficult to estimate, and usually one has to prove that it is much smaller than the first one. That one is the so called global term.

For the lower bound the authors have used a boundary version of the Wasserstein distance, that is, the quantity defined in Subsection 2.1.1. The power p of such distance is superadditive, hence it works well for the lower bound, and by using that one gets the boundary matching problem.

The boundary matching problem.

This problem has been introduced by [5, 2] to prove a lower bound similar to (1.1.2.6). The point is estimating, for large n, the following boundary cost functions

$$Cb^{p,d}_{n,\Omega,\mu} := Wb_p^p\left(\sum_{i=1}^n \delta_{X_i}, \sum_{i=1}^n \delta_{Y_i}\right) \quad \text{and} \quad \tilde{C}b^{p,d}_{n,\Omega,\mu} := Wb_p^p\left(\sum_{i=1}^n \delta_{X_i}, n\mu\right), \tag{1.1.2.11}$$

where $\{X_i, Y_i\}_{i=1}^{\infty}$ are iid random variables on Ω with probability distribution $d\mu = \rho dx$ with the same properties as before. The difference from the previous cost functions is that in this case one considers the boundary Wasserstein distance Wb (see Subsection 2.1.1) instead of the ordinary one. Roughly speaking, one can add points on $\partial\Omega$ (not independent on X and Y) to build the optimal matching between the random variables.

Of course, this produces two limiting constants as in (1.1.1.7), a priori different from those ones. That is, in the case of $\{X_i, Y_i\}_{i=1}^{\infty}$ iid random variables uniformly distributed on $[0,1]^d$, one gets two other constants cb_d^p and $\tilde{c}b_d^p$, defined as

$$cb_d^p := \lim_{n\to\infty} \frac{\mathbb{E}[Cb^{p,d}_{n,[0,1]^d,dx}]}{f_n} \quad \text{and} \quad \tilde{c}b_d^p := \lim_{n\to\infty} \frac{\mathbb{E}[\tilde{C}b^{p,d}_{n,[0,1]^d,dx}]}{f_n} \tag{1.1.2.12}$$

where the existence of such limits for $d \geq 3$ has been proved in [23, 2] using the superadditivity of the boundary Wasserstein distance (see Lemma 2.1.1 in Subsection 2.1.1).

Abusing a little bit with the notation, for any p and d, hereafter we will denote

$$\tilde{c}b_d^p := \liminf_{n\to\infty} \frac{\mathbb{E}[Cb^{p,d}_{n,[0,1]^d,dx}]}{f_n} \quad \text{and} \quad \tilde{c}_d^p := \limsup_{n\to\infty} \frac{\mathbb{E}[\tilde{C}^{p,d}_{n,[0,1]^d,dx}]}{f_n}, \tag{1.1.2.13}$$

where the only case in which it is not sure that the previous lim sup and lim inf coincide with the limit of the respective quantities is the case $d = 2, p \neq 2$.

As for the not-uniform case, the following result holds.

Theorem 1.1.4 ([5, 24])**.** *Let $\{X_i, Y_i\}_{i=1}^{\infty}$ be iid random variables on $\Omega \subseteq \mathbb{R}^d (d \geq 3)$ with distribution $d\mu = \rho dx$, where Ω is a bounded and connected domain with Lipschitz boundary and ρ is an α-Hölder probability density such that $\frac{1}{C} \leq \rho \leq C$, for $C > 0$. Then for any $p \geq 1$ the cost of the boundary matching in (1.1.2.11) verifies*

$$\liminf_{n\to\infty} \frac{\mathbb{E}[Cb^{p,d}_{n,\Omega,\mu}]}{n^{1-\frac{p}{d}}} \geq cb_d^p \int_\Omega \rho^{1-\frac{p}{d}} \quad \textit{and} \quad \liminf_{n\to\infty} \frac{\mathbb{E}[\tilde{C}b^{p,d}_{n,\Omega,\mu}]}{n^{1-\frac{p}{d}}} \geq \tilde{c}b_d^p \int_\Omega \rho^{1-\frac{p}{d}}, \tag{1.1.2.14}$$

with cb_d^p and $\tilde{c}b_d^p$ the constants defined in (1.1.2.12).

Notice that comparing (1.1.2.14) and (1.1.2.6), one could be able to prove that the lim sup and the lim inf are actually limits and do coincide if

$$c_d^p = cb_d^p \quad \text{and} \quad \tilde{c}_d^p = \tilde{c}b_d^p. \tag{1.1.2.15}$$

Of course, since $Wb \leq W$ (see Subsection 2.1.1) it holds

$$cb_d^p \leq c_d^p \quad \text{and} \quad \tilde{c}b_d^p \leq c_d^p,$$

and moreover, (1.1.2.15) holds for both the constants c and $\tilde{c}$ ([2]) for $p = d = 2$, and in general for $p = 2$ and any $d \geq 3$ in the case of the semidiscrete matching (see [22]), that is

$$\tilde{c}_d^2 = \tilde{c}b_d^p, \quad \forall d \geq 3.$$

The pre-factor $\int_\Omega \rho^{1-\frac{p}{d}}$ in (1.1.2.14) is obtained as in Theorems 1.1.2 and 1.1.3. The difference is that since the boundary Wasserstein distance can be computed between measures with different masses, one does not have to pass through the global term as in the upper bound. Indeed, by superadditivity on disjoint sets (see Lemma 2.1.1), if $\{Q_j\}_j$ is the decomposition in disjoint hypercubes of Ω, one has

$$Wb_p^p\left(\sum_{i=1}^n \delta_{X_i}, n\mu\right) \geq \sum_j Wb_p^p\left(\sum_{i:X_i\in Q_j} \delta_{X_i}, n\mu \mathbb{1}_{Q_j}\right).$$

Taking the expectation of such quantities, one can argue as for the local term in the upper bound.

Summarizing the results mentioned in Theorems 1.1.2, 1.1.3 and 1.1.4 one gets the following result.

Theorem 1.1.5. *Let $\{X_i, Y_i\}_{i=1}^{\infty}$ be iid random variables on $\Omega \subseteq \mathbb{R}^d$ with distribution $d\mu = \rho dx$, where Ω is a bounded and connected domain with Lipschitz boundary and ρ is an α-Hölder probability density such that $\frac{1}{C} \leq \rho \leq C$, for $C > 0$. Then for any $p \geq 1$ and $d \geq 2$ it holds*

$$\limsup_{n\to\infty} \frac{\mathbb{E}[C^{p,d}_{n,\Omega,\mu}]}{n^{1-\frac{p}{d}}(\chi_{d\geq 3} + (\log n)^{\frac{p}{2}}\chi_{d=2})} \leq c_d^p \int_\Omega \rho^{1-\frac{p}{d}}, \qquad \limsup_{n\to\infty} \frac{\mathbb{E}[\tilde{C}^{p,d}_{n,\Omega,\mu}]}{n^{1-\frac{p}{d}}(\chi_{d\geq 3} + (\log n)^{\frac{p}{2}}\chi_{d=2})} \leq \tilde{c}_d^p \int_\Omega \rho^{1-\frac{p}{d}}, \tag{1.1.2.16}$$

and

$$\liminf_{n\to\infty} \frac{\mathbb{E}[Cb^{p,d}_{n,\Omega,\mu}]}{n^{1-\frac{p}{d}}(\chi_{d\geq 3} + (\log n)^{\frac{p}{2}}\chi_{d=2})} \geq cb^p_d \int_\Omega \rho^{1-\frac{p}{d}}, \qquad \liminf_{n\to\infty} \frac{\mathbb{E}[\tilde{C}b^{p,d}_{n,\Omega,\mu}]}{n^{1-\frac{p}{d}}(\chi_{d\geq 3} + (\log n)^{\frac{p}{2}}\chi_{d=2})} \geq \tilde{c}b^p_d \int_\Omega \rho^{1-\frac{p}{d}}. \tag{1.1.2.17}$$

The proof of the previous result in the case $p = 2, d \neq 2$ in completely analogue to the proof of the result in the other cases, but it has been rigorously stated in [11].

1.1.3 The Gaussian matching.

A natural question is what happens in the case in which the points X and Y are distributed in an unbounded domain $\Omega \subset \mathbb{R}^d$, that is, if the scaling and/or the constant change.

A classical example is the case in which the distribution of the points in $\mathbb{R}^d$ is $d\mu = \rho dx$, with

$$\rho(x) = \frac{e^{-\frac{|x|^2}{2}}}{(2\pi)^{\frac{d}{2}}},$$

and in this case the state of the art is the following.

Theorem 1.1.6. *Let $\{X_i, Y_i\}_{i=1}^\infty$ be a sequence of iid random variables distributed in $\mathbb{R}^d$ according to the gaussian probability measure μ. Then*

- *in dimension $d = 1$ and with exponent $p = 2$ (see [8]) it holds*

$$\mathbb{E}[\tilde{C}^{2,1}_{n,\mathbb{R},\mu}] \approx \log\log n,$$

- *in dimension $d = 2$ and with exponent $p = 2$ (see [25, 32, 26]) it holds*

$$\mathbb{E}[C^{2,2}_{n,\mathbb{R}^2,\mu}] \approx (\log n)^2,$$

- *in dimension $d \geq 3$ and for any $1 \leq p < d$ (see [27]) it holds*

$$\mathbb{E}[\tilde{C}^{p,d}_{n,\mathbb{R}^d,\mu}] \approx n^{1-\frac{p}{d}}.$$

Let us spend a few words about the previous theorem. As for the dimension $d = 1$, the result in [8] extends to the cost $C^{2,1}$ (the bipartite case) and holds for general log-concave distributions (i.e., for distributions whose density is the exponential of a strictly concave function).

As for the dimension $d = 2$, the rate $(\log n)^2$ instead of $\log n$ is due to the fact that the most of the n points of type X and Y are contained in the disk of radius $\sqrt{2\log n}$, and therefore, if one pretends that in (1.1.2.1) one can substitute $\Omega = B_{\sqrt{2\log n}}$, one gets

$$\mathbb{E}[C^{p,d}_{n,\mathbb{R}^2,\mu}] \simeq \frac{|\Omega|}{2\pi}\log n = (\log n)^2.$$

Of course, such an estimate is not precise since one can not approximate the gaussian density on $B_{\sqrt{2\log n}}$ with a density bounded from below from a positive constant, as required for the validity of (1.1.2.1). Nevertheless, that is the reason why one gets such a rate.

As for $d \geq 3$, recall that in the case of bounded and connected domain, for $1 \leq p < d$ by (1.1.2.6) and (1.1.2.14) one has the estimate

$$\mathbb{E}[C^{p,d}_{n,\Omega,\mu}] \simeq c^p_d n^{1-\frac{p}{d}} \int_\Omega \rho^{1-\frac{p}{d}}, \tag{1.1.3.1}$$

where the constant c^p_d depends on wherever one is considering the upper or the lower bound. However, since for $1 \leq p < d$ the $(1 - \frac{p}{d})^{th}$ power of the gaussian density is integrable, if one substitutes in the previous expression in (1.1.3.1) $\Omega = \sqrt{2\log n}$, such a constant does not generate an extra term in the scaling rate.

Finally, let us stress that for $p = d$ in [32] it was conjectured that

$$\mathbb{E}[C^{p,d}_{n,\mathbb{R}^d,\mu}] \approx (\log n)^{\frac{d}{2}},$$

and this complies with the estimate in (1.1.3.1) since, if one substitutes $\Omega = \sqrt{2\log n}$ in the case $p = d$, one gets exactly the measure of the ball of radius $\sqrt{2\log n}$, which increases with n as $(\log n)^{\frac{d}{2}}$.

1.2 The exponential and polynomial matching.

The purpose of this thesis is to complete the classification of the behavior of the cost function in the gaussian case in Theorem 1.1.6, to study whenever one can get some estimates similar to (1.1.2.6) and (1.1.2.14), and to extend those results to the case of polynomially decaying densities.

The results we are showing have been proved in [12, 11].

Here we state the main result and we explain the strategy we use to prove it. We are considering the following class of densities:

$$\rho(x) = c_d \frac{e^{-\frac{|x|^q}{q}}}{1+|x|^\gamma}, \quad \text{with} \begin{cases} \gamma \in \mathbb{R}, & q > 0, \\ \gamma > d+1, & q = 0. \end{cases} \tag{1.2.0.1}$$

with c_d the normalizing constant. By $q = 0$ we mean the case in which there is only the polynomial part (and the exponential one disappears).

The cost function changes it behavior for large n when p reaches the following critical threshold:

$$p^{crit} := \begin{cases} d, & q > 0, \\ p^* := \frac{d(\gamma-d)}{\gamma}, & q = 0, \end{cases} \tag{1.2.0.2}$$

Notice that p^{crit} is the critical threshold for p such that the integral in (1.1.3.1) is not finite.

For f_n as in (1.1.1.2), the cost function $\mathbb{E}[C^{p,d}_{n,\mathbb{R}^d,\rho dx}]$ (respectively $\mathbb{E}[Cb^{p,d}_{n,\mathbb{R}^d,\rho dx}]$) is bounded from above (respectively from below) by the following rate (for some positive constants):

$$\tau^{p,d}_n := \begin{cases} f_n, & 1 \le p < d \\ (\log n)^{1+\frac{2}{q}}, & p = d = 2 \\ (\log n)^{\frac{d}{q}}, & p = d \ge 3 \\ (\log n)^{d-1-p(1-\frac{1}{q})}, & p > d \end{cases} \text{for } q > 0, \quad \tau^{p,d}_n := \begin{cases} f_n, & 1 \le p < p^* \\ (\log n)^{2-\frac{2}{\gamma}} n^{\frac{2}{\gamma}}, & p = p^*, d = 2 \\ \log n n^{\frac{d}{\gamma}}, & p = p^*, d \ge 3 \\ n^{\frac{p}{\gamma-d}}, & p > p^* \end{cases} \text{for } q = 0, \quad . \tag{1.2.0.3}$$

Moreover, if $p \le p^{crit}$, we can estimate the limiting constants by multiplying c^p_d and $\tilde{c}^p_d$ by the following positive constants

$$\zeta^p_d := \begin{cases} \int_{\mathbb{R}^d} \rho^{1-\frac{p}{d}}, & 1 \le p < d, \\ \int_{B_{\sqrt[q]{q}}} \left(1 - \frac{|x|^q}{q}\right), & p = d = 2, \\ |B_{\sqrt[q]{q}}|, & p = d \ge 3, \end{cases} \quad \text{for } q > 0, \tag{1.2.0.4}$$

$$\zeta^p_d := \begin{cases} \int_{\mathbb{R}^d} \rho^{1-\frac{p}{d}}, & 1 \le p < p^*, \\ (\gamma-2)^{-1} \lim_{R\to\infty} \frac{\int_{B_R \setminus B_1} \rho^{\frac{2}{\gamma}} [\log(R/|x|)]^{1-\frac{2}{\gamma}}}{(\log R)^{2-\frac{2}{\gamma}}}, & p = d = 2, \\ (\gamma-d)^{-1} \lim_{R\to\infty} \frac{\int_{B_R} \rho^{\frac{d}{\gamma}}}{\log R}, & p = p^* \ge 3, \end{cases} \quad \text{for } q = 0.$$

Therefore the results we shall prove are the following, stated separating the upper and the lower bound.

Theorem 1.2.1 ([12, 11], Upper bound)**.** *Let $\{X_i\}_{i=1}^\infty$ be a sequence of iid random variables distributed in $\mathbb{R}^d$ with probability distribution $d\mu = \rho dx$, with ρ in (1.2.0.1). Then, for p^{crit} defined in (1.2.0.2), $\tau^{p,d}_n$ defined in (1.2.0.3) and ζ^p_d defined in (1.2.0.4), it holds*

$$\limsup_{n\to\infty} \frac{\mathbb{E}[\tilde{C}^{p,d}_{n,\mathbb{R}^d,\mu}]}{\tau^{p,d}_n} \le \tilde{c}^p_d \zeta^p_d, \quad \forall p \in [1, p^{crit}],$$
$$\mathbb{E}[\tilde{C}^{p,d}_{n,\mathbb{R}^d,\mu}] \le C\tau^{p,d}_n, \quad \forall p > p^{crit},$$

where the second inequality holds for a positive constant $C > 0$.

Theorem 1.2.2 ([12, 11], Lower bound)**.** *Let $\{X_i\}_{i=1}^\infty$ be a sequence of iid random variables distributed in $\mathbb{R}^d$ with probability distribution $d\mu = \rho dx$, with ρ in (1.2.0.1). Then, for p^{crit} defined in (1.2.0.2), $\tau^{p,d}_n$ defined in (1.2.0.3) and ζ^p_d defined in (1.2.0.4), it holds*

$$\liminf_{n\to\infty} \frac{\mathbb{E}[\tilde{C}b^{p,d}_{n,\mathbb{R}^d,\mu}]}{\tau^{p,d}_n} \ge \tilde{c}b^p_d \zeta^p_d, \quad \forall p \in [1, p^{crit}],$$
$$\mathbb{E}[\tilde{C}b^{p,d}_{n,\mathbb{R}^d,\mu}] \ge C\tau^{p,d}_n, \quad \forall p > p^{crit},$$

where the second inequality holds for a positive constant $C > 0$.

To explain the results, let us denote, for this moment, by S_n the radius

$$S_n = \begin{cases} \sqrt[q]{q \log n}, & q > 0, \\ n^{\frac{1}{\gamma - d}}, & q = 0. \end{cases} \tag{1.2.0.5}$$

We begin with the case $d \geq 3$: by (1.1.2.6) and (1.1.2.14), in the case of uniformly positive densities on bounded connected domains, one has

$$\mathbb{E}[\tilde{C}^{p,d}_{n,\Omega,\mu}] \simeq \tilde{c}^p_d n^{1-\frac{p}{d}} \int_\Omega \rho^{1-\frac{p}{d}} = \tilde{c}^p_d n^{1-\frac{p}{d}} \int_{B_{S_n}} \rho^{1-\frac{p}{d}}, \tag{1.2.0.6}$$

where the constant depends again on whenever one considers the lower bound or the upper bound.

The reason why one should substitute $\Omega = B_{S_n}$, with S_n in (1.2.0.5), is explained in Section 6.1, but, briefly speaking, the idea is that the most of the points are constrained in such ball (or disk).

However, by (1.2.0.6) it is clear that if $\rho^{1-\frac{p}{d}}$ is integrable, nothing should change (and, in fact, nothing changes), while the critical threshold $p = p^{crit}$ in (1.2.0.2) corresponds to the infimum of the exponents p such that $\rho^{1-\frac{p}{d}}$ is not integrable. The idea behind the case $p > p^{crit}$ is the same but it is a little bit more delicate since, while for $p \leq p^{crit}$ there is a small degree of freedom in how to choose S_n, for $p > p^{crit}$ there is not (see Section 6.1 for more details). Hence if one uses (1.2.0.6) to estimate the cost for large n the technique does not work for S_n in (1.2.0.5). Moreover, for $p > p^{crit}$, in the estimate there are a lot of not negligible contributions which do not appear for $p \leq p^{crit}$, and this is the reason why we do not get precise estimates on the constant.

1.3 Structure of the thesis.

The structure of the thesis is the following.

In Chapter 2 we recall some Optimal Transport notions and other related properties. In particular, in Section 2.1 we define the marginals of a measure on a product space, the notion of coupling of two measures, the Wasserstein distance (Definition 2.1.2) and its subadditivity. Then, in Subsection 2.1.1 we focus on the boundary Wasserstein distance, which is instead superadditive (Lemma 2.1.1), and we spend some words about the way it is related to the classical one. In Section 2.2 we then define the transport maps, the push-forward of a measure and we recall its properties: the density of a pushforwarded measure, the push-forward of a push-forward, the way it changes the (classical and boundary) Wasserstein distance between two measures (Lemmas 2.2.2 and 2.2.3) and the necessary conditions to get a measure as the push-forward on another one in an hypercube (Lemma 2.2.6). We finally focus on the Benamou-Brenier formula (Section 2.3) and on some properties of the decomposition of measures (Section 2.4).

In Chapter 3 we recall some concentration properties of the Binomial random variables: in particular, we need an estimate on the centered moments of binomial Random variables (Lemma 3.1.1) and the Chernoff bound (Proposition 3.1.1).

In Chapter 4 we focus on the exponential map. Indeed, since the decomposition we use exploits the radial symmetry of the density we are considering, to transport each subset into a cube we take advantage on the exponential map, which projects the tangent space to a manifold on the manifold itself. The case we are interest in is the $d-1$-dimensional sphere. We start with the definitions of geodesics and geodesic distance between points (Section 4.1), and then we focus on some properties about the first and second derivatives of the exponential map (Section 4.2).

In Chapter 5 we recall some notions of Measure Theory: in Section 5.1 we recall the definition of the Hausdorff measure, while in Section 5.2 we briefly recall some properties of the Haar measure on $O(d)$, which we need later for some technical reasons in Chapter 6.

In Chapter 6 we prove the main result following the scheme we mentioned in the Introduction. To do that, we first focus on the exact domain we have to consider to approximate the cost as we mentioned before, that is, a ball with a radius increasing with n, and we estimate the error in substituting the distribution ρ with another one supported in such ball (Section 6.1). Then, in Section 6.2, we explain how the decomposition in hypercubes is to be made both in the upper bound, where the regions are not properly hypercubes (see Subsection 6.2.1) and where one has to estimate the global term as in (1.1.2.10), and in the lower bound (Subsection 6.2.3). Then we focus on the local term and we approximate the cost of the matching in each subset of the decomposition with the cost of the matching for uniformly iid random variables (see Section 6.3). Finally, we prove Theorems 1.2.1 and 1.2.2, respectively in Sections 6.5 and 6.6.

Chapter 2

Some generalities about the Optimal Transport.

In this Chapter we focus on some general facts about Optimal Transport.

First we define the Wasserstein distance between two measures.

2.1 The Wasserstein distance.

The most of the content of this Section is taken from [20, 5, 31, 33].

To define the Wasserstein distance, let us denote with $\mathcal{M}$ the set of positive measures on a fixed region ($\mathcal{M}(\Omega)$ is the set of positive Borel Radon measures on Ω). In particular, if one has a measure $\pi \in \mathcal{M}(\Omega \times \Omega)$, with $\Omega \subseteq \mathbb{R}^d$, its marginals $\Gamma^1_\sharp \pi$ and $\Gamma^2_\sharp \pi$ are the two measures on Ω satisfying

$$\int_\Omega d\Gamma^1_\sharp \pi(x) f(x) = \int_{\Omega\times\Omega} d\pi(x,y) f(x), \quad \text{and} \quad \int_\Omega d\Gamma^2_\sharp \pi(y) f(y) = \int_{\Omega\times\Omega} d\pi(x,y) f(y), \tag{2.1.0.1}$$

for any bounded continuous function $f : \Omega \to \mathbb{R}$. Using the marginals, we get the notion of coupling of two measures with the same mass, as follows.

Definition 2.1.1 ([31], Problem 1.2)**.** *Let $\Omega \subseteq \mathbb{R}^d$ be a Borel space and $\mu, \nu \in \mathcal{M}(\Omega)$ be two measures such that $\mu(\Omega) = \nu(\Omega)$. The set of couplings of μ and ν is*

$$\Gamma(\mu,\nu) := \left\{ \pi \in \mathcal{M}(\Omega \times \Omega) : \Gamma^1_\sharp \pi = \mu, \Gamma^2_\sharp \pi = \nu \right\},$$

where the marginals $\Gamma^1_\sharp \pi$ and $\Gamma^2_\sharp \pi$ of π are defined in (2.1.0.1)*.*

The condition of having the same mass is sufficient and necessary for the existence of a coupling, indeed by substituting $f = 1$ in (2.1.0.1) one gets

$$\pi(\Omega \times \Omega) = \Gamma^1_\sharp \mu(\Omega) = \Gamma^2_\sharp \nu(\Omega),$$

hence there does not exist a coupling between two measures with different masses.

On the contrary, the set of possible couplings between two measures with the same mass is for sure a non void set, indeed one can always find the product coupling (which is the worst one for our purposes), that is,

$$\frac{1}{\mu(\Omega)} \mu \otimes \nu \in \Gamma(\mu, \nu).$$

If the two measures coincide the most natural coupling is

$$d\pi(x,y) = d\mu(x) \delta(y - x),$$

that is, the coupling π such that for any bounded and continuous function $f : \Omega \times \Omega \to \mathbb{R}$

$$\int_{\Omega\times\Omega} d\pi(x,y) f(x,y) = \int_\Omega d\mu(x) f(x,x).$$

Now we can define the Wasserstein distance between two measures with the same mass.

Definition 2.1.2 ([31], Problem 1.2)**.** *Let $\Omega \subseteq \mathbb{R}^d$, and let μ and ν be two positive measures on Ω and $p \geq 1$. The Wasserstein distance of order p between μ and ν is*

$$W_p(\mu,\nu) := \inf\left\{ \left(\int_{\Omega\times\Omega} d\pi(x,y)|x-y|^p \right)^{\frac{1}{p}} : \pi \in \Gamma(\mu,\nu) \right\}.$$

It is possible to prove (see for example [31, Proposition 5.1]) that the Wasserstein distance has the three properties of a distance (strictly positive unless the two measures coincide, symmetric and satisfying the triangle inequality), and also that under particular conditions the infimum is achieved.

Before that, we state the following properties of the Wasserstein distance.

The subadditivity. A crucial property of the Wasserstein distance is the subadditivity. For further details, see [33, Theorem 4.8], where it is also referred as convexity.

The property is the following: if $\Omega \subseteq \mathbb{R}^d$ and $(E, \mathcal{E}, \lambda)$ is a measure space such that, for any $z \in E$, μ_z and ν_z are measures on Ω with the same mass, one has

$$W_p^p \left(\int_E d\lambda(z)\mu_z, \int_E d\lambda(z)\nu_z \right) \leq \int_E d\lambda(z) W_p^p(\mu_z,\nu_z). \tag{2.1.0.2}$$

The proof of that is immediate: if, for any $z \in E$, one chooses a coupling π_z of μ_z and ν_z, the measure

$$\int_E d\lambda(z)\pi_z,$$

is a coupling of $\int_E d\lambda(z)\mu_z$ and $\int_E d\lambda(z)\nu_z$, and therefore

$$W_p^p \left(\int_E d\lambda(z)\mu_z, \int_E d\lambda(z)\nu_z \right) \leq \int_E d\lambda(z) \int_{\Omega\times\Omega} d\pi_z(x,y)|x-y|^p,$$

and (2.1.0.2) follows from a measurable selection of the optimal couplings $\pi_z \in \Gamma(\mu_z,\nu_z)$.

Notice that if one chooses $E = \mathbb{N}$ and λ is the counting measure, one gets

$$W_p^p \left(\sum_{i=0}^{\infty} \mu_i, \sum_{i=0}^{\infty} \nu_i \right) \leq \sum_{i=0}^{\infty} W_p^p(\mu_i,\nu_i),$$

and notice also that the previous result (2.1.0.2) does not require the measures to be concentrated on disjoint domains.

The homogeneity. The homogeneity property is the following: if μ and ν are two positive measures with the same mass on a metric space Ω and α a positive constant, then

$$W_p^p(\alpha\mu,\alpha\nu) = \alpha W_p^p(\mu,\nu). \tag{2.1.0.3}$$

this is a consequence of the fact that, π is a coupling of μ and ν if and only if $\alpha\pi$ is a coupling of $\alpha\mu$ and $\alpha\nu$.

The monotonicity. We now briefly focus on the monotonicity with respect to p: if μ and ν are two positive measures on a metric space Ω such that $\mu(\Omega) = \nu(\Omega) = 1$, then, for any $1 \leq p \leq q \leq \infty$, it holds

$$W_p(\mu,\nu) \leq W_q(\mu,\nu). \tag{2.1.0.4}$$

This is a consequence of the fact that, for any π coupling of μ and ν, one has $\pi(\Omega\times\Omega) = \mu(\Omega) = \nu(\Omega) = 1$, and therefore, by Young inequality,

$$W_p(\mu,\nu) \leq \left(\int_{\Omega\times\Omega} d\pi(x,y)|x-y|^p \right)^{\frac{1}{p}} \leq \left(\int_{\Omega\times\Omega} d\pi(x,y)|x-y|^q \right)^{\frac{1}{q}},$$

and by taking the infimum of the right-hand side with respect to π we get (2.1.0.4)

2.1.1 The boundary Wasserstein distance.

In this subsection we focus on a variant of the Wasserstein distance, that is, the boundary Wasserstein distance, introduced in [20] and also in [5]. The point of doing that is that while the Wasserstein distance has the subadditivity property (2.1.0.2), which is very useful in the proof of the upper bounds one needs, the lower bound is very similar but one needs a version of the distance satisfying a superadditivity inequality. In other words, one defines a new set of couplings between two measures μ and ν on an open domain $\Omega \subseteq \mathbb{R}^d$, as follows.

Definition 2.1.3 ([20], Problem 1.1)**.** *Let $\Omega \subseteq \mathbb{R}^d$ an open domain and let μ and ν be two positive measures on Ω. The boundary couplings of μ and ν are the elements of*

$$\Gamma b(\mu, \nu) := \left\{\pi \in \mathcal{M}(\bar{\Omega} \times \bar{\Omega}) : \Gamma^1_\sharp \mu|_\Omega = \mu, \Gamma^2_\sharp \nu|_\Omega = \nu\right\}.$$

One can use this definition to introduce the boundary Wasserstein distance (and this is what we shall do), but before that let us focus a moment on the previous definition.

First notice that, while μ and ν are concentrated on Ω, π is concentrated on $\bar{\Omega} \times \bar{\Omega}$ and its marginals (concentrated on $\bar{\Omega}$) are forced to coincide with μ and ν only on Ω. This is the reason why one does not have to ask for μ and ν to have the same mass, since the mass defect can be rearranged on the boundary. In other words, $\Gamma b(\mu, \nu)$ is always a non void set: for instance, if $\mu(\Omega) > \nu(\Omega)$, one can fix $\bar{x} \in \partial\Omega$ and it is straightforward to verify that

$$\frac{1}{\mu(\Omega)} \mu \otimes (\nu + (\mu(\Omega) - \nu(\Omega))\delta_{\bar{x}}) \in \Gamma b(\mu, \nu).$$

Moreover, one can notice that if the two measures have the same mass, then

$$\Gamma(\mu, \nu) \subseteq \Gamma b(\mu, \nu), \tag{2.1.1.1}$$

since any coupling of μ and ν can be extended to a boundary coupling by not giving weight to the points of $\Omega \times \partial\Omega$, $\partial\Omega \times \Omega$ and $\partial\Omega \times \partial\Omega$. Hence, in case the two measures have the same mass, the boundary couplings are much more than the ordinary couplings.

Now we can focus on the definition of the boundary Wasserstein distance.

Definition 2.1.4 ([20], Problem 1.1)**.** *Let $\Omega \subseteq \mathbb{R}^d$ an open domain and let μ and ν be two positive measures on Ω. The boundary Wasserstein distance between μ and ν is*

$$Wb_p(\mu, \nu) := \inf \left\{ \left(\int_{\bar{\Omega} \times \bar{\Omega}} d\pi(x, y)|x - y|^p \right)^{\frac{1}{p}} : \pi \in \Gamma b(\mu, \nu) \right\}.$$

First notice that, thanks to property (2.1.1.1), one can immediately state that if μ and ν have the same mass on a bounded domain Ω, then

$$Wb_p(\mu, \nu) \leq W_p(\mu, \nu), \tag{2.1.1.2}$$

because in this case the infimum is taken on a larger set (see [20, Remark 2.10]).

The superadditivity. Contrary to the classical Wasserstein distance, the boundary Wasserstein distance has the following superadditivity property.

Lemma 2.1.1 ([2], equation (2.8): superadditivity of Wb_p^p)**.** *Let $\Omega \subseteq \mathbb{R}^d$ be an open domain, and let $\{\Omega_k\}_{k\in\mathbb{N}}$ be a sequence of open and pairwise disjoints subsets of Ω. Then, for any couple of measures μ and ν on Ω, it holds*

$$Wb_p^p(\mu, \nu) \geq \sum_{k\in\mathbb{N}} Wb_p^p(\mu|_{\Omega_k}, \nu|_\Omega).$$

Before proving the Lemma, let us notice that the statement does not require that

$$\Omega = \bigcup_k \Omega_k,$$

but only that

$$\bigcup_k \Omega_k \subseteq \Omega.$$

Proof. The proof of this fact is a little bit technical, but the idea is quite simple: given a boundary coupling π of μ and ν, one has to find a sequence of boundary couplings $\{\pi_k\}_k$ of $\mu|_{\Omega_k}$ and $\nu|_{\Omega_k}$ through some specific elements of $\partial(\cup_k \Omega_k) \cap \bar{xy}$, with $\bar{xy}$ being the segment connecting x to y. Such $\{\pi_k\}_{k\in\mathbb{N}}$ have to satisfy

$$\int_{\bar{\Omega}\times\bar{\Omega}} d\pi(x,y)|x-y|^p \geq \sum_{k\in\mathbb{N}} \int_{\bar{\Omega}_k\times\bar{\Omega}_k} d\pi_k(x,y)|x-y|^p, \tag{2.1.1.3}$$

in such a way to have

$$\int_{\bar{\Omega}\times\bar{\Omega}} d\pi(x,y)|x-y|^p \geq \sum_{k\in\mathbb{N}} Wb_p^p(\mu|_{\Omega_k}, \nu|_{\Omega_k}),$$

and the thesis will follow optimizing over π.

To do that, let us fix $\pi \in \Gamma b(\mu,\nu)$ any boundary coupling of μ and ν. We then have to decompose $\bar{\Omega}$ in a disjoint union of subsets, that is

$$\bar{\Omega} = \bigcup_k \Omega_k \cup \left(\bar{\Omega} \setminus \bigcup_k \Omega_k\right),$$

from which one has

$$\begin{aligned}
\int_{\bar{\Omega}\times\bar{\Omega}} d\pi(x,y)|x-y|^p \geq & \int_{\bigcup_k \Omega_k \times \bigcup_j \Omega_j} d\pi(x,y)|x-y|^p && (2.1.1.4)\\
&+ \int_{\bigcup_k \Omega_k \times (\bar{\Omega}\setminus \bigcup_j \Omega_j)} d\pi(x,y)|x-y|^p && (2.1.1.5)\\
&+ \int_{(\bar{\Omega}\setminus \bigcup_j \Omega_j)\times \bigcup_k \Omega_k} d\pi(x,y)|x-y|^p && (2.1.1.6)\\
&+ \int_{(\bar{\Omega}\setminus \bigcup_k \Omega_k)\times(\bar{\Omega}\setminus \bigcup_j \Omega_j)} d\pi(x,y)|x-y|^p, && (2.1.1.7)
\end{aligned}$$

and we shall estimate from below the four terms in (2.1.1.4), (2.1.1.5), (2.1.1.6) and (2.1.1.7). To estimate (2.1.1.4), for any $k \in \mathbb{N}$ we define the following function

$$F_k : \Omega_k \times (\mathbb{R}^d) \setminus \Omega_k \to \partial\Omega_k, \quad (x,y) \mapsto (1-t_k(x,y))x + t_k(x,y)y,$$

where t_k is the first time a point exits from Ω_k following the direction $x \to y$, that is

$$t_k : \Omega_k \times (\mathbb{R}^d) \setminus \Omega_k \to \partial\Omega_k, \quad (x,y) \mapsto \inf\{t \in [0,1] : (1-t)x + ty \notin \Omega_k\}.$$

Now notice that if $x \in \Omega_k$ and $y \in \Omega_j$ with $k \neq j$, we have

$$t_k(x,y) + t_j(y,x) \leq 1,$$

otherwise, by definition, one would have

$$\inf\{t : (1-t)x + ty \notin \Omega_k\} > \sup\{t : (1-t)x + ty \notin \Omega_j\} \Rightarrow \exists t \in (0,1) \text{ such that } (1-t)x + ty \in \Omega_k \cap \Omega_j,$$

but this can not happen since $\Omega_k \cap \Omega_j = \emptyset$.

Hence one has

$$\begin{aligned}
|x-y|^p &= |x-y|^p(t_k(x,y) + 1 - t_k(x,y) - t_j(y,x) + t_j(y,x))^p \\
&\geq |x-y|^p(t_k(x,y))^p + |x-y|^p(t_j(y,x))^p \\
&= |x - F_k(x,y)|^p + |F_j(y,x) - y|^p,
\end{aligned}$$

and therefore

$$\begin{aligned}
(2.1.1.4) &= \sum_{k\in\mathbb{N}} \int_{\Omega_k\times\Omega_k} d\pi(x,y)|x-y|^p + \sum_{k\neq j} \int_{\Omega_k\times\Omega_j} d\pi(x,y)|x-y|^p \\
&\geq \sum_{k\in\mathbb{N}} \int_{\Omega_k\times\Omega_k} d\pi(x,y)|x-y|^p \\
&+ \sum_{j\neq k} \int_{\Omega_k\times\Omega_j} d\pi(x,y)|x-F_k(x,y)|^p + \sum_{k\neq j} \int_{\Omega_k\times\Omega_j} d\pi(x,y)|F_j(y,x)-y|^p.
\end{aligned}$$

Moreover, (2.1.1.5) and (2.1.1.6) do not need to be estimated further, except for noticing that

$$(2.1.1.5) = \sum_{k\in\mathbb{N}} \int_{\Omega_k\times(\bar\Omega\setminus\bigcup_j\Omega_j)} d\pi(x,y)|x-y|^p \quad \text{and} \quad (2.1.1.6) = \sum_{k\in\mathbb{N}} \int_{(\bar\Omega\setminus\bigcup_j\Omega_j)\times\Omega_k} d\pi(x,y)|x-y|^p,$$

while for (2.1.1.7) we use the trivial estimate

$$(2.1.1.7) \geq 0.$$

Combining the previous bounds we get

$$\begin{aligned}\int_{\bar\Omega\times\bar\Omega} d\pi(x,y)|x-y|^p \geq & \sum_{k\in\mathbb{N}} \int_{\Omega_k\times\Omega_k} d\pi(x,y)|x-y|^p \\ & + \sum_{k\in\mathbb{N}}\sum_{j\neq k} \int_{\Omega_k\times\Omega_j} d\pi(x,y)|x-F_k(x,y)|^p + \sum_{k\in\mathbb{N}}\sum_{j\neq k} \int_{\Omega_j\times\Omega_k} d\pi(x,y)|F_k(y,x)-y|^p \\ & + \sum_{k\in\mathbb{N}} \int_{\Omega_k\times(\bar\Omega\setminus\bigcup_j\Omega_j)} d\pi(x,y)|x-y|^p + \sum_{k\in\mathbb{N}} \int_{(\bar\Omega\setminus\bigcup_j\Omega_j)\times\Omega_k} d\pi(x,y)|x-y|^p \\ & = \sum_{k\in\mathbb{N}} \int_{\bar\Omega_k\times\bar\Omega_k} d\pi_k(x,y)|x-y|^p, \end{aligned} \tag{2.1.1.8}$$

with π_k the measure implicitly defined by

$$\begin{aligned}\int_{\bar\Omega_k\times\bar\Omega_k} d\pi_k(x,y)f(x,y) := & \int_{\Omega_k\times\Omega_k} d\pi(x,y)f(x,y) \\ & + \sum_{j\neq k} \int_{\Omega_k\times\Omega_j} d\pi(x,y)f(x,F_k(x,y)) + \sum_{j\neq k} \int_{\Omega_j\times\Omega_k} d\pi(x,y)f(F_k(y,x),y) \\ & + \int_{\Omega_k\times(\bar\Omega\setminus\bigcup_j\Omega_j)} d\pi(x,y)f(x,y) + \int_{(\bar\Omega\setminus\bigcup_j\Omega_j)\times\Omega_k} d\pi(x,y)f(x,y).\end{aligned}$$

Hence one can notice that (2.1.1.8) is exactly (2.1.1.3) up to proving that $\pi_k \in \Gamma b(\mu|_{\Omega_k}, \nu|_{\Omega_k})$.

Therefore, let us show that the first marginal of π_k restricted to Ω_k is μ (restricted to Ω_k too). The analogue property for the second marginal works the same way.

Recalling that

$$F_k(y,x) \in \partial\Omega_k \Rightarrow \mathbb{1}_{\Omega_k}(F_k(y,x)) = 0,$$

and also that $\pi \in \Gamma b(\mu,\nu)$, one has

$$\begin{aligned}\int_{\bar\Omega_k\times\bar\Omega_k} d\pi_k(x,y)f(x)\mathbb{1}_{\Omega_k}(x) = & \int_{\Omega_k\times\Omega_k} d\pi(x,y)f(x)\mathbb{1}_{\Omega_k}(x) \\ & + \sum_{j\neq k} \int_{\Omega_k\times\Omega_j} d\pi(x,y)f(x)\mathbb{1}_{\Omega_k}(x) + \sum_{j\neq k} \int_{\Omega_j\times\Omega_k} d\pi(x,y)f(F_k(y,x))\mathbb{1}_{\Omega_k}(F_k(y,x)) \\ & + \int_{\Omega_k\times(\bar\Omega\setminus\bigcup_j\Omega_j)} d\pi(x,y)f(x)\mathbb{1}_{\Omega_k}(x) + \int_{(\bar\Omega\setminus\bigcup_j\Omega_j)\times\Omega_k} d\pi(x,y)f(x)\mathbb{1}_{\Omega_k}(x) \\ & = \int_{\Omega_k\times\bar\Omega} d\pi(x,y)f(x) = \int_{\bar\Omega} d\mu(x)f(x)\mathbb{1}_{\Omega_k}(x) = \int_{\Omega_k} d\mu|_{\Omega_k}(x)f(x).\end{aligned}$$

Hence (2.1.1.3) is proved and therefore the proof is concluded. □

2.2 The push-forward of a measure and its properties.

To start with, we define the push-forward $T_\sharp\mu$ of a measure μ via a map T. For our purposes, μ is again a measure (not necessarily a probability measure) on $\Omega \subseteq \mathbb{R}^d$.

Definition 2.2.1 ([33], Definition 1.2). *Let $\Omega, \Omega' \subseteq \mathbb{R}^d$ be Borel spaces, μ a measure on Ω and $T : \Omega \to \Omega'$ a μ-measurable map. The push-forward of μ via T is the measure $T_\sharp\mu$ on Ω' such that*

$$T_\sharp\mu(A) := \mu(T^{-1}(A)), \qquad \forall A \subseteq \Omega' \text{ Borel}.$$

In particular, if $X : \Sigma \to \Omega$ is a random variable with μ as probability distribution, then $T(X) : \Sigma \to \Omega'$ is a random variable with $T_\sharp\mu$ as probability distribution.

As a consequence of Definition 2.2.1, for any measurable map $f : \Omega' \to \mathbb{R}$ such that $f \circ T \in L^1(\mu)$, one has

$$\int_{\Omega'} dT_\sharp\mu(x)f(x) = \int_\Omega d\mu(x)f(T(x)).$$

It is very easy to build a coupling between a measure and its push-forward via a map T, indeed the measure

$$\pi(x, y) = \mu(x)\delta(y - T(x)),$$

meaning that for any $f : \Omega \times \Omega \to \mathbb{R}$ it holds

$$\int_{\Omega\times\Omega} d\pi(x, y)f(x, y) = \int_\Omega d\mu(x)f(x, T(x)),$$

is by definition a coupling of μ and $T_\sharp\mu$. In other words, $\pi = (\mathrm{Id}, T)_\sharp\mu$, where of course (Id, T) is a map from Ω to $\Omega \times \Omega$. In particular, this means that

$$W_p^p(\mu, T_\sharp\mu) \le \int_\Omega d\mu(x)|x - T(x)|^p. \tag{2.2.0.1}$$

Notice also that applying the map T preserves the independence of the random variables, indeed, if $\{X_i\}_{i\in E}$ is a family of independent random variables, then also $\{T(X_i)\}_{i\in E}$ is. Indeed

$$\mathbb{P}\left(\bigcap_{i\in E}\{T(X_i) \in A_i\}\right) = \mathbb{P}\left(\bigcap_{i\in E}\{X_i \in T^{-1}(A_i)\}\right) = \prod_{i\in E}\mathbb{P}(X_i \in T^{-1}(A_i)) = \prod_{i\in E}\mathbb{P}(T(X_i) \in A_i).$$

In particular, one has the following result.

Proposition 2.2.1. *Let $T : \Omega \subseteq \mathbb{R}^d \to \Omega' \subseteq \mathbb{R}^d$ be a measurable map. Then, if $\{X_i\}_{i\in E}$ are iid random variables on Ω with distribution μ, also $\{T(X_i)\}_{i\in E}$ are iid random variables on $T(\Omega)$ and their common probability distribution is $T_\sharp\mu$.*

Moreover, if $T : \Omega \to \Omega'$ is a diffeomorphism and μ is an absolutely continuous measure (denote by ρ its density), then the same holds for $T_\sharp\mu$. Indeed changing variables $y = T(x)$ one gets

$$\int_{\Omega'} dT_\sharp\mu(x)f(x) = \int_\Omega d\mu(x)f(T(x)) = \int_\Omega dx\rho(x)f(T(x)) = \int_{\Omega'} dy\rho(T^{-1}(y))\det DT^{-1}(y)f(y),$$

that is, $T_\sharp\mu$ has density

$$\frac{dT_\sharp\mu}{dx}(x) = \frac{d\mu}{dx}(T^{-1}(x))\det DT^{-1}(x).$$

In the following, we prove some properties about it: in particular, we need some informations about the push-forward of a measure through the composition of two maps and the push-forward of an empirical measure.

Lemma 2.2.1. *Let $\{X_i\}_{i=1}^n \subseteq \Omega$ and $T : \Omega \to \Omega', S : \Omega' \to \Omega''$ two measurable maps. Then*

$$(S \circ T)_\sharp\mu = S_\sharp T_\sharp\mu; \tag{2.2.0.2}$$

$$T_\sharp\left(\sum_{i=1}^n \delta_{X_i}\right) = \sum_{i=1}^n \delta_{T(X_i)}. \tag{2.2.0.3}$$

Proof. We start with the proof of (2.2.0.2): for any measurable function $f : \Omega'' \to \mathbb{R}$, we have

$$\begin{aligned}\int_{\Omega''} d(S \circ T)_\sharp\mu(x)f(x) &= \int_\Omega d\mu(x)f((S \circ T)(x)) = \int_\Omega d\mu(x)(f \circ S)(T(x))\\ &= \int_{\Omega'} dT_\sharp\mu(x)(f \circ S)(x) = \int_{\Omega''} dS_\sharp T_\sharp\mu(x)f(x),\end{aligned}$$

and therefore the statement is proved.

For the proof of (2.2.0.3) is sufficient to prove that for any function $f : \Omega \to \mathbb{R}$

$$\int_{\Omega'} dT_\sharp\left(\sum_{i=1}^n \delta_{X_i}\right)(x)f(x) = \int_{\Omega'} d\left(\sum_{i=1}^n \delta_{T(X_i)}\right)(x)f(x),$$

and this is easy to prove since

$$\int_{\Omega'} dT_\sharp\left(\sum_{i=1}^n \delta_{X_i}\right)(x)f(x) = \int_\Omega d\left(\sum_{i=1}^n \delta_{X_i}\right)(x)f(T(x)) = \sum_{i=1}^n f(T(X_i)) = \int_{\Omega'} d\left(\sum_{i=1}^n \delta_{T(X_i)}\right)(x)f(x).$$

□

Before going on, we have to mention the following result: one may ask whether the best couplings in the definition of Wasserstein distance are induced by transport maps. In the following case, which is the case interesting to us, they are.

Theorem 2.2.1 ([33], Theorem 10.38)**.** *Let μ and ν be two measures on $\mathbb{R}^d$, with μ absolutely continuous with respect to the Lebesgue measure. Then there exists an optimal coupling π for the Wasserstein distance W_p in Definition 2.1.2 and it is induced by an optimal transport map T, that is*

$$\pi = (Id, T)_\sharp \mu.$$

A further information we need about the push-forward of a measure is the relation between the Wasserstein distance of the push-forward of two measures and the Wasserstein distance of the original ones. One finds out that is quantity is related to the Lipschitz constant of T.

Lemma 2.2.2 ([7], Lemma 1 and [2], equation (2.12))**.** *Let μ and ν be two measures on $\Omega \subseteq \mathbb{R}^d$ such that $\mu(\Omega) = \nu(\Omega)$, and let $T : \Omega \to \Omega' \subseteq \mathbb{R}^d$ be a measurable map. Then*

$$W_p^p (T_\sharp\mu, T_\sharp\nu) \leq \left(\sup_{x,y\in\Omega} \frac{|T(x) - T(y)|}{|x - y|} \right)^p W_p^p(\mu, \nu). \tag{2.2.0.4}$$

Moreover, if T is bijective, it holds

$$W_p^p (T_\sharp\mu, T_\sharp\nu) \geq \left(\inf_{x,y\in\Omega} \frac{|T(x) - T(y)|}{|x - y|} \right)^p W_p^p(\mu, \nu). \tag{2.2.0.5}$$

Proof. We begin with the proof of (2.2.0.4). This statement is a direct consequence of the fact that, for any coupling π of μ and ν, one can find another coupling $\tilde{\pi}$ of $T_\sharp\mu$ and $T_\sharp\nu$ satisfying the inequality in the statement. Such a coupling is $\tilde{\pi} = (T, T)_\sharp\pi$, and for any bounded and continuous function $f : \Omega' \times \Omega' \to \mathbb{R}$ it holds

$$\int_{\Omega'\times\Omega'} d\tilde{\pi}(x, y) f(x, y) = \int_{\Omega\times\Omega} d\pi(x, y) f(T(x), T(y)).$$

To start with, we prove that $\tilde{\pi}$ is a coupling of $T_\sharp\mu$ and $T_\sharp\nu$, in fact

$$\int_{\Omega'\times\Omega'} d\tilde{\pi}(x, y) f(x) = \int_{\Omega\times\Omega} d\pi(x, y) f(T(x)) = \int_\Omega d\mu(x) f(T(x)) = \int_{\Omega'} dT_\sharp\mu(x) f(x),$$

and the same holds for $T_\sharp\nu$.

Therefore, by definition one has

$$\begin{aligned} W_p^p (T_\sharp\mu, T_\sharp\nu) &\leq \int_{\Omega\times\Omega} d\tilde{\pi}(x, y)|x - y|^p = \int_{\Omega\times\Omega} d\pi(x, y)|T(x) - T(y)|^p = \int_{\Omega\times\Omega} d\pi(x, y) \frac{|T(x) - T(y)|^p}{|x - y|^p} |x - y|^p \\ &\leq \left(\sup_{x,y\in\Omega} \frac{|T(x) - T(y)|}{|x - y|} \right)^p \int_{\Omega\times\Omega} d\pi(x, y)|x - y|^p, \end{aligned}$$

and the statement follows optimizing over π.

The proof of (2.2.0.5) is a consequence of (2.2.0.4): indeed, if one applies (2.2.0.4) substituting

$$\mu \to T_\sharp\mu, \quad \nu \to T_\sharp\nu, \quad T \to T^{-1},$$

one gets

$$W_p^p \left((T^{-1})_\sharp T_\sharp\mu, (T^{-1})_\sharp T_\sharp\nu \right) \leq \left(\sup_{x,y\in\Omega} \frac{|T^{-1}(x) - T^{-1}(y)|}{|x - y|} \right)^p W_p^p(T_\sharp\mu, T_\sharp\nu). \tag{2.2.0.6}$$

By (2.2.0.2) of Lemma 2.2.1 one has

$$(T^{-1})_\sharp T_\sharp\mu = (T^{-1} \circ T)_\sharp\mu = \mathrm{Id}_\sharp\mu = \mu,$$

and the same holds for ν. Thus (2.2.0.6) becomes

$$W_p^p(T_\sharp\mu, T_\sharp\nu) \geq \left(\sup_{x,y\in\Omega} \frac{|T^{-1}(x) - T^{-1}(y)|}{|x - y|} \right)^{-p} W_p^p(\mu, \nu) = \left(\inf_{x,y\in\Omega} \frac{|T(x) - T(y)|}{|x - y|} \right)^p W_p^p(\mu, \nu),$$

thus also (2.2.0.5) is proved. ∎

The same properties hold for the boundary Wasserstein distance.

Lemma 2.2.3 ([2], equation (2.13)). *Let μ and ν be two measures on $\Omega \subseteq \mathbb{R}^d$. Then, if $T : \Omega \to \Omega' \subseteq \mathbb{R}^d$ verifies $T(\partial\Omega) = \partial\Omega'$, it holds*

$$Wb_p^p\left(T_\sharp\mu, T_\sharp\nu\right) \le \left(\sup_{x,y\in\bar{\Omega}} \frac{|T(x)-T(y)|}{|x-y|}\right)^p Wb_p^p(\mu,\nu). \tag{2.2.0.7}$$

Moreover, if T is bijective, it holds

$$Wb_p^p\left(T_\sharp\mu, T_\sharp\nu\right) \ge \left(\inf_{x,y\in\bar{\Omega}} \frac{|T(x)-T(y)|}{|x-y|}\right)^p Wb_p^p(\mu,\nu). \tag{2.2.0.8}$$

Proof. The proof of this statement if completely analogue to the proof of Lemma 2.2.3, but since here we deal with the boundary couplings of μ and ν, we have to be careful that, given $\pi \in \Gamma b(\mu,\nu)$, defining $\tilde{\pi} = (T,T)_\sharp\pi$, that is

$$\int_{\overline{\Omega'}\times\overline{\Omega'}} d\tilde{\pi}(x,y) f(x,y) = \int_{\bar{\Omega}\times\bar{\Omega}} d\pi(x,y) f(T(x),T(y)),$$

one has $\tilde{\pi} \in \Gamma b(T_\sharp\mu, T_\sharp\nu)$. To prove this, one can notice that its first marginal coincides with $T_\sharp\mu$ on Ω' since, for any $f : \Omega' \to \mathbb{R}$, by definition

$$\int_{\overline{\Omega'}\times\overline{\Omega'}} d\tilde{\pi}(x,y) f(x)\mathbb{1}_{\Omega'}(x) = \int_{\bar{\Omega}\times\bar{\Omega}} d\pi(x,y) f(T(x))\mathbb{1}_{\Omega'}(T(x)) = \int_{\bar{\Omega}\times\bar{\Omega}} d\pi(x,y) f(T(x))\mathbb{1}_{\Omega}(x)$$

where in the second equality it is hidden the passage

$$T(\partial\Omega) = \partial\Omega' \Rightarrow \mathbb{1}_{\Omega'}(T(x)) = \mathbb{1}_{\Omega}(x).$$

Thus, since $\pi \in \Gamma b(\mu,\nu)$, one has

$$\int_{\overline{\Omega'}\times\overline{\Omega'}} d\tilde{\pi}(x,y) f(x)\mathbb{1}_{\Omega}(x) = \int_{\bar{\Omega}\times\bar{\Omega}} d\pi(x,y) f(T(x))\mathbb{1}_{\Omega}(x) = \int_{\Omega} d\mu(x) f(T(x)) = \int_{\Omega'} dT_\sharp\mu(x) f(x).$$

Thus the first marginal of $\tilde{\pi}$ is $T_\sharp\mu$, and the same argument proves that the second marginal of $\tilde{\pi}$ is $T_\sharp\nu$. Hence $\tilde{\pi} \in \Gamma b(T_\sharp\mu, T_\sharp\nu)$.

This is enough to prove both (2.2.0.7) and (2.2.0.8), as in the proof of Lemma 2.2.2. □

As a consequence of Lemmas 2.2.2 and 2.2.3, one gets that, in the case in which the quantity

$$\frac{|T(x)-T(y)|}{|x-y|}$$

does not depend on x and y, that is, for example if T is the composition of a traslation and a rescaling, the two inequalities in the Lemmas produce an equality. In other words, the following result holds.

Corollary 2.2.1. *Let μ and ν be two measures on $\Omega \subseteq \mathbb{R}^d$ such that $\mu(\Omega) = \nu(\Omega)$ and $T : \mathbb{R}^d \to \mathbb{R}^d$ the affine map*

$$T(x) = ax + b.$$

Then, it holds

$$W_p^p\left(T_\sharp\mu, T_\sharp\nu\right) = |a|^p W_p^p\left(\mu,\nu\right).$$

Moreover, the same result holds for Wb and without the hypothesis $\mu(\Omega) = \nu(\Omega)$.

Proof. Since T is invertible, thanks to Lemma 2.2.2 one has

$$\left(\inf_{x,y\in\Omega} \frac{|T(x)-T(y)|}{|x-y|}\right)^p W_p^p(\mu,\nu) \le W_p^p\left(T_\sharp\mu, T_\sharp\nu\right) \le \left(\sup_{x,y\in\Omega} \frac{|T(x)-T(y)|}{|x-y|}\right)^p W_p^p(\mu,\nu),$$

and using that the quantity

$$\frac{|T(x)-T(y)|}{|x-y|} = \frac{|ax+b-ay-b|}{|x-y|} = |a|,$$

does not depend on x or y, the statement follows.

In the case of the distance Wb, to conclude it is also necessary to notice that $T(\partial\Omega) = \partial T(\Omega)$. □

We now state a first application of the superadditivity property we proved in Proposition 2.1.1 and of the rescaling property we proved in Corollary 2.2.1. For the following result we also need the notion of Binomial random variable, which is studied in details later, in Chapter 3.

Lemma 2.2.4. *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of iid random variables uniformly distributed in $(0,1)^d$ and let N be a binomial random variable of parameters n and p independent of $\{X_i\}_{i=1}^{\infty}$. Then, if*

$$f(n,p) := \mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N}\delta_{X_i}, np\mathbb{1}_{(0,1)^d}\right)\right],$$

for any $\lambda > 1$ it holds

$$f(n,p) \geq \frac{\lfloor\lambda\rfloor^d}{\lambda^p} f\left(n, \frac{p}{\lambda^d}\right),$$

Proof. To prove the result, we argue as in [23]: we split $(0,1)^d$ in $\lfloor\lambda\rfloor^d$ hypercubes Q_j of side $\frac{1}{\lambda}$.

In this way, we have

$$\bigcup_j Q_j \subseteq (0,1)^d,$$

and the inclusion is a proper inclusion unless λ is a natural number.

Then, if

$$N(Q_j) := \sum_{i=1}^{N}\mathbb{1}(X_i \in Q_j),$$

is the number of points in Q_j, $N(Q_j)$ is a binomial random variable of parameters n and $p|Q_j| = p\lambda^{-d}$ (see (3.0.0.2) for this). Hence, we have

$$\begin{aligned} f(n,p) &= \mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N}\delta_{X_i}, np\mathbb{1}_{(0,1)^d}\right)\right] \\ &\geq \sum_{j=1}^{\lfloor\lambda\rfloor^d}\mathbb{E}\left[Wb_p^p\left(\sum_{i:X_i\in Q_j}\delta_{X_i}, np\mathbb{1}_{Q_j}\right)\right] \qquad \text{by Lemma 2.1.1} \\ &= \sum_{j=1}^{\lfloor\lambda\rfloor^d}\mathbb{E}\left[\mathbb{E}\left[Wb_p^p\left(\sum_{i:X_i\in Q_j}\delta_{X_i}, np|Q_j|\frac{\mathbb{1}_{Q_j}}{|Q_j|}\right)\middle| N(Q_j)\right]\right]. \end{aligned}$$

If now we define the map

$$T_j : Q_j = q_j + \frac{1}{\lambda}(0,1)^d \to (0,1)^d, \qquad x \mapsto \lambda(x - q_j),$$

we get

$$\begin{aligned} f(n,p) &\geq \sum_{j=1}^{\lfloor\lambda\rfloor^d}\mathbb{E}\left[\mathbb{E}\left[Wb_p^p\left((T_j^{-1})_\sharp\left(\sum_{i:X_i\in Q_j}\delta_{T_j(X_i)}\right), np|Q_j|(T_j^{-1})_\sharp(\mathbb{1}_{(0,1)^d})\right)\middle| N(Q_j)\right]\right] \\ &= \sum_{j=1}^{\lfloor\lambda\rfloor^d}\lambda^{-p}\mathbb{E}\left[\mathbb{E}\left[Wb_p^p\left(\sum_{i:X_i\in Q_j}\delta_{T_j(X_i)}, np|Q_j|\mathbb{1}_{(0,1)^d}\right)\middle| N(Q_j)\right]\right] \qquad \text{by Corollary 2.2.1} \\ &= \sum_{j=1}^{\lfloor\lambda\rfloor^d}\lambda^{-p}\mathbb{E}\left[\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N(Q_j)}\delta_{Z_i}, np|Q_j|\mathbb{1}_{(0,1)^d}\right)\middle| N(Q_j)\right]\right], \end{aligned}$$

where $\{Z_i\}_{i=1}^{\infty}$ are iid uniformly distributed on $(0,1)^d$ independent of $\{X_i\}_{i=1}^{\infty}$. In the last passage we used that $\{T_j(X_i)\}_{i:X_i\in Q_j}$ conditioned to $N(Q_j)$ are uniformly distributed in $(0,1)^d$ as well as $\{Z_i\}_{i=1}^{N(Q_j)}$.

Hence, we have

$$f(n,p) \geq \lambda^{-p}\lfloor\lambda\rfloor^d f\left(n, \frac{p}{\lambda^d}\right),$$

and the statement is proved. ∎

We also need a result concerning some sufficient conditions under which an absolutely continuous measure on an hypercube is the push-forward of the uniform measure through a Lipschitz map T whose Lipschitz constant is very close to 1.

Before proving that, we focus on a very useful property of positive and C^1 regular functions.

Lemma 2.2.5. *Let $\Omega \subseteq \mathbb{R}^d$ be a convex domain. There exists a constant $C > 0$ depending on Ω such that for any positive function $\rho : \Omega \xrightarrow{C^1} (0, +\infty)$ it holds*

$$\sup_{x,y\in\Omega}\left|\frac{\rho(x)}{\rho(y)} - 1\right| \leq \exp\left\{C \sup_{z\in\Omega}\frac{|\nabla\rho(z)|}{\rho(z)}\right\} - 1. \tag{2.2.0.9}$$

In particular, fixed $M > 0$, there exists a positive constant $C(M, \Omega)$ depending on Ω and M such that, for any ρ with the same hypothesis as before and such that $\frac{|\nabla\rho(z)|}{\rho(z)} \leq M$, it holds

$$\sup_{x,y\in\Omega}\left|\frac{\rho(x)}{\rho(y)} - 1\right| \leq C(M,\Omega) \sup_{z\in\Omega}\frac{|\nabla\rho(z)|}{\rho(z)}. \tag{2.2.0.10}$$

Proof. The proof of (2.2.0.9) is immediate: there exists a point $\xi = tx + (1-t)y$ (for a certain $t \in [0, 1]$) such that

$$\begin{aligned}
\left|\frac{\rho(x)}{\rho(y)} - 1\right| &= \left|\exp\{\log\rho(x) - \log\rho(y)\} - 1\right| \\
&= \left|\exp\left\{\frac{\nabla\rho(\xi)}{\rho(\xi)}\cdot(x-y)\right\} - 1\right| \\
&\leq \exp\left\{\left|\frac{\nabla\rho(\xi)}{\rho(\xi)}\cdot(x-y)\right|\right\} - 1 \\
&\leq \exp\left\{\operatorname{diam}(\Omega)\sup_{z\in\Omega}\frac{|\nabla\rho(z)|}{\rho(z)}\right\} - 1,
\end{aligned}$$

where $\operatorname{diam}(\Omega)$ is the diameter of Ω.

(2.2.0.10) is a consequence of (2.2.0.9) if one considers that $e^c - 1$ is asymptotic to c for small values of c. □

We can now state the result we mentioned before Lemma 2.2.5.

Lemma 2.2.6 ([31], Section 2.3)**.** *Let $d \geq 1$ and $M > 0$. There exists a constant C depending on d and on M such that for any hypercube $Q = q + [0, L]^d$ and $\rho : Q \xrightarrow{C^1} [0, +\infty)$ such that $\rho > 0$ a.s. and $\sup_{z\in Q}\frac{|\nabla\rho(z)|}{\rho(z)} \leq M$, there exists a map $T : Q \to Q$ such that*

$$\rho = T_\sharp\left(\frac{\mathbb{1}_Q}{|Q|}\right) \quad \textit{and} \quad \sup_{x,y\in Q}\left|\frac{|T(x)-T(y)|}{|x-y|} - 1\right| \leq CL \sup_{z\in Q}\frac{|\nabla\rho(z)|}{\rho(z)}.$$

Before proving this fact, let us mention that such a map is also referred as "Knothe map", and its peculiarity is the way it is constructed, that is, modifying one direction at a time.

Let us also notice that if the density ρ in the previous Lemma is uniformly bounded from below by a positive constant $c > 0$, then one gets

$$\sup_{x,y\in Q}\left|\frac{|T(x)-T(y)|}{|x-y|} - 1\right| \leq \frac{CL}{c}\sup_{z\in Q}|\nabla\rho(z)|.$$

In general, for an α-Hölder density uniformly bounded from below by a positive constant c, Lemma 2.2.5 holds with T satisfying

$$\sup_{x,y\in Q}\left|\frac{|T(x)-T(y)|}{|x-y|} - 1\right| \leq \frac{CL^\alpha}{c}C_\alpha, \tag{2.2.0.11}$$

where C_α is the Hölder constant (see [2, Proposition 2.4] for further details on this).

Proof. We prove this through a map T which modifies one direction at a time, and we can assume that $q = 0$ and $L = 1$, as follows.

Step 0: we can assume $Q=[0,1]^d$. To prove this, let us define the map

$$\hat{T}:Q=q+[0,L]^d\to[0,1]^d,\quad \hat{T}(x)=\frac{x-q}{L}.$$

Such a map is invertible, hence $(\hat{T}^{-1})_\sharp=(\hat{T}_\sharp)^{-1}$, and satisfies

$$\hat{T}_\sharp\rho(x)=\rho\left(q+Lx\right)L^d,$$

from which we have

$$\sup_{x\in[0,1]^d}\frac{|\nabla\hat{T}_\sharp\rho(x)|}{\hat{T}_\sharp\rho(x)}=L\sup_{x\in[0,1]^d}\frac{|\nabla\rho(q+Lx)|}{\rho(q+Lx)}=L\sup_{z\in Q}\frac{|\nabla\rho(z)|}{\rho(z)}. \tag{2.2.0.12}$$

Hence, if such a map T exists for $\hat{T}_\sharp\rho$, by (2.2.0.2) of Lemma 2.2.1 we have

$$\hat{T}_\sharp\rho=T_\sharp\left(\mathbb{1}_{[0,1]^d}\right)=T_\sharp\hat{T}_\sharp\left(\frac{\mathbb{1}_Q}{|Q|}\right)\quad\Rightarrow\rho=(\hat{T}_\sharp)^{-1}T_\sharp\hat{T}_\sharp\left(\frac{\mathbb{1}_Q}{|Q|}\right)=(\hat{T}^{-1}\circ T\circ\hat{T})_\sharp\left(\frac{\mathbb{1}_Q}{|Q|}\right),$$

and for any $x,y\in Q$

$$\begin{aligned}\left|\frac{|\hat{T}^{-1}\circ T\circ\hat{T}(x)-\hat{T}^{-1}\circ T\circ\hat{T}(y)|}{|x-y|}-1\right|&=\left|\frac{\left|T\left(\frac{x-q}{L}\right)-T\left(\frac{y-q}{L}\right)\right|}{\left|\frac{x-q}{L}-\frac{y-q}{L}\right|}-1\right|\\&\le C\sup_{x\in[0,1]^d}\frac{|\nabla\hat{T}_\sharp\rho(x)|}{\hat{T}_\sharp\rho(x)}=CL\sup_{z\in Q}\frac{|\nabla\rho(z)|}{\rho(z)},\end{aligned}$$

where in the last passage we have used (2.2.0.12).

Hence, $\hat{T}^{-1}\circ T\circ\hat{T}$ works for ρ as T works for $\hat{T}_\sharp\rho$, and for this reason we can assume $Q=[0,1]^d$.

To prove the thesis, we first denote

$$\rho^{(k)}(x):=\int_0^1dy_1\cdots\int_0^1dy_k\frac{\rho(y_1,\dots,y_k,x_{k+1},\dots,x_d)}{\rho([0,1]^d)},\quad k=0,1,\dots,d,\quad x\in(0,1)^d,$$

and it is straightforward to see that

$$\rho^{(0)}=\frac{\rho\mathbb{1}_{[0,1]^d}}{\rho([0,1]^d)}\text{ and }\rho^{(d)}=\mathbb{1}_{[0,1]^d}.$$

Thus, we look for a map $T^{(k)}$ such that

$$(T^{(k)})_\sharp\rho^{(k)}=\rho^{(k-1)}, \tag{2.2.0.13}$$

so that by induction

$$(T^{(1)}\circ\cdots\circ T^{(d)})_\sharp\left(\mathbb{1}_{[0,1]^d}\right)=(T^{(1)}\circ\cdots\circ T^{(d)})_\sharp\rho^{(d)}=\rho^{(0)}=\frac{\rho\mathbb{1}_{[0,1]^d}}{\rho([0,1]^d)}.$$

Therefore a suitable candidate for the map T we are looking for is

$$T:=T^{(1)}\circ\cdots\circ T^{(d)},$$

and then we claim that, up to choosing each $T^{(k)}$ in a suitable way, we get

$$T(\partial[0,1]^d)=\partial[0,1]^d\quad\text{and}\quad\sup_{x,y\in[0,1]^d}\left|\frac{|T(x)-T(y)|}{|x-y|}-1\right|\le C\sup_{x\in[0,1]^d}\frac{|\nabla\rho(x)|}{\rho(x)}. \tag{2.2.0.14}$$

To find T satisfying (2.2.0.14), we claim that it is sufficient to find $\{T^{(k)}\}_{k=1}^d$ as in (2.2.0.13) satisfying

$$T^{(k)}(\partial[0,1]^d)=\partial[0,1]^d\quad\forall k=1,\dots,d\quad\text{and}\quad\sup_{z\in[0,1]^d}\left\|D((T^{(k)})^{-1})(z)-\mathrm{Id}\right\|\le C\sup_{x\in[0,1]^d}\frac{|\nabla\rho(x)|}{\rho(x)}. \tag{2.2.0.15}$$

Of course the first statement in (2.2.0.15) implies the first statement in (2.2.0.14).

It is not completely trivial to see that the second part of (2.2.0.15) implies the second part of (2.2.0.14). To prove this, we begin with noticing that

$$\begin{aligned}\sup_{x,y\in[0,1]^d}\left|\frac{|T(x)-T(y)|}{|x-y|}-1\right| &\leq \sup_{z\in[0,1]^d}\|DT(z)-\mathrm{Id}\| \\ &= \sup_{z\in[0,1]^d}\left\|\sum_{k=1}^d D(T^{(1)}\circ\cdots\circ T^{(k)})(z)-D(T^{(1)}\circ\cdots\circ T^{(k-1)})(z)\right\| \\ &\leq \sum_{k=1}^d\prod_{j=1}^k \sup_{z\in[0,1]^d}\left\|DT^{(j)}(z)\right\| \sup_{z\in[0,1]^d}\left\|D((T^{(k)})^{-1})(z)-\mathrm{Id}\right\|,\end{aligned} \tag{2.2.0.16}$$

Thus we have to estimate both $\sup_{z\in[0,1]^d}\left\|D((T^{(k)})^{-1})(z)-\mathrm{Id}\right\|$ and $\sup_{z\in[0,1]^d}\left\|DT^{(k)}(z)\right\|$, which is a little bit more than (2.2.0.15). We begin with the estimate of the first one of these two quantities. Notice that if one proves that for any $k=1,\dots,d$

$$\max\left\{\sup_{z\in[0,1]^d}\left\|D(T^{(k)})(z)-\mathrm{Id}\right\|, \sup_{z\in[0,1]^d}\left\|D((T^{(k)})^{-1})(z)-\mathrm{Id}\right\|\right\}\leq C\sup_{x\in[0,1]^d}\frac{|\nabla\rho(x)|}{\rho(x)}, \tag{2.2.0.17}$$

one can conclude since

$$\sup_{z\in[0,1]^d}\left\|D(T^{(k)})(z)\right\|\leq 1+\sup_{z\in[0,1]^d}\left\|D(T^{(k)})(z)-\mathrm{Id}\right\|\leq 1+C\sup_{x\in[0,1]^d}\frac{|\nabla\rho(x)|}{\rho(x)},$$

and one can combine this estimate with (2.2.0.16).

Thus we only have to prove (2.2.0.17).

Step 1: prove that, for a constant C depending on d and on M, it holds

$$\sup_{z\in[0,1]^d}\left\|D((T^{(k)})^{-1})(z)-\mathrm{Id}\right\|\leq C\sup_{x\in[0,1]^d}\frac{|\nabla\rho(x)|}{\rho(x)}.$$

We claim that (2.2.0.15) holds for a certain choice of the map $T^{(k)}$.

Such map $T^{(k)}$ verifying (2.2.0.13) exists in the form

$$T^{(k)}(x)=(x_1,\dots,x_{k-1},S_{x_{k+1},\dots,x_d}(x_k),x_{k+1},\dots,x_d),$$

because none of $\rho^{(k)}$ and $\rho^{(k-1)}$ depends on $x_1,\dots,x_{k-1}$ (therefore they can be left fixed), and moreover

$$\int_0^1 dx_k\rho^{(k)}(x)=\int_0^1 dx_k\rho^{(k-1)}(x).$$

First we prove that for a suitable choice of the map S one gets

$$T^{(k)}(x)=x \qquad \forall x\in\partial Q, \tag{2.2.0.18}$$

which implies the first statement in (2.2.0.15).

To do that, notice that (2.2.0.13) can also be written as

$$\int_0^1 dy_1\cdots\int_0^1 dy_{k-1}\rho(y_1,\dots,y_{k-1},x_k,\dots,x_d)=\int_0^1 dy_1\cdots\int_0^1 dy_k\rho(y_1,\dots,y_k,x_{k+1}\dots,x_d)|(S^{-1}_{x_{k+1},\dots,x_d})'(x_k)|,$$

and therefore, by imposing the condition $S'_{x_{k+1},\dots,x_d}(x_k)>0$, one finds out that $S_{x_{k+1},\dots,x_d}(x_k)$ verifies

$$S^{-1}_{x_{k+1},\dots,x_d}(x_k)=\frac{\int_0^1 dy_1\cdots\int_0^1 dy_{k-1}\int_0^{x_k}dy_k\rho(y_1,\dots,y_k,x_{k+1},\dots,x_d)}{\int_0^1 dz_1\cdots\int_0^1 dz_k\rho(z_1,\dots,z_k,x_{k+1},\dots,x_d)}, \tag{2.2.0.19}$$

and therefore

$$S^{-1}_{x_{k+1},\dots,x_d}(0)=0,\quad S^{-1}_{x_{k+1},\dots,x_d}(1)=1,\quad S^{-1}_{x_{k+1},\dots,x_d}(x_k)\in(0,1)\forall x_k\in(0,1)\quad \text{that is,}\quad (2.2.0.18).$$

To prove that such $S_{x_{k+1},\dots,x_d}$ works well enough to satisfy the second condition in (2.2.0.15), notice that

$$D((T^{(k)})^{-1})(x)=(A_{i,j})_{i,j=1}^d,\quad \text{with } A_{i,j}=\begin{cases}\delta_{i,j} & i\neq k, j=1,\dots,d,\\ 0 & i=k, j=1,\dots,k-1,\\ \partial_j S^{-1}_{x_k,\dots,x_d} & i=k, j=k,\dots,d,\end{cases}$$

and therefore

$$\left\| D((T^{(k)})^{-1})(x) - \mathrm{Id} \right\| \leq \sqrt{\left(\partial_k S^{-1}_{x_k,\dots,x_d} - 1\right)^2 + \sum_{j=k+1}^{d} \left(\partial_j S^{-1}_{x_k,\dots,x_d}\right)^2}, \tag{2.2.0.20}$$

and to estimate this quantity we look back to (2.2.0.19). We begin with ∂_k: by (2.2.0.19), first notice that

$$\begin{aligned}
&|\partial_k S^{-1}_{x_{k+1},\dots,x_d}(x_k) - 1| \\
&= \left| \frac{\int_0^1 dy_1 \cdots \int_0^1 dy_{k-1} \rho(y_1,\dots,y_{k-1},x_k,\dots,x_d)}{\int_0^1 dz_1 \cdots \int_0^1 dz_k \rho(z_1,\dots,z_k,x_{k+1},\dots,x_d)} - 1 \right| \\
&\leq \frac{\int_0^1 dy_1 \cdots \int_0^1 dy_{k-1} \int_0^1 dy_k |\rho(y_1,\dots,y_{k-1},x_k,\dots,x_d) - \rho(y_1,\dots,y_k,x_{k+1},\dots,x_d)|}{\int_0^1 dz_1 \cdots \int_0^1 dz_k \rho(z_1,\dots,z_k,x_{k+1},\dots,x_d)} \\
&\leq C \sup_{z\in[0,1]^d} \frac{|\nabla\rho(z)|}{\rho(z)} \frac{\int_0^1 dy_1 \cdots \int_0^1 dy_{k-1} \int_0^1 dy_k \rho(y_1,\dots,y_k,x_{k+1},\dots,x_d)}{\int_0^1 dz_1 \cdots \int_0^1 dz_k \rho(z_1,\dots,z_k,x_{k+1},\dots,x_d)} \text{ with } C \text{ by (2.2.0.10) of Lemma 2.2.5} \\
&= C \sup_{z\in[0,1]^d} \frac{|\nabla\rho(z)|}{\rho(z)}.
\end{aligned} \tag{2.2.0.21}$$

Moreover, for $j \geq k+1$, looking back to (2.2.0.19) again, we have

$$\begin{aligned}
\left|\partial_j S^{-1}_{x_{k+1},\dots,x_d}(x_k)\right| &\leq \left| \frac{\int_0^1 dy_1 \cdots \int_0^1 dy_{k-1} \int_0^{x_k} dy_k \partial_j \rho(y_1,\dots,y_k,x_{k+1},\dots,x_d)}{\int_0^1 dz_1 \cdots \int_0^1 dz_k \rho(z_1,\dots,z_k,x_{k+1},\dots,x_d)} \right| \\
&+ \frac{\begin{array}{c}\int_0^1 dy_1 \cdots \int_0^1 dy_{k-1} \int_0^{x_k} dy_k \rho(y_1,\dots,y_k,x_{k+1},\dots,x_d) \\ \cdot \left|\int_0^1 dz_1 \cdots \int_0^1 dz_k \partial_j \rho(z_1,\dots,z_k,x_{k+1},\dots,x_d)\right|\end{array}}{\left(\int_0^1 dz_1 \cdots \int_0^1 dz_k \rho(z_1,\dots,z_k,x_{k+1},\dots,x_d)\right)^2} \\
&\leq 2 \frac{\int_0^1 dy_1 \cdots \int_0^1 dy_{k-1} \int_0^1 dy_k \, |\partial_j \rho(y_1,\dots,y_k,x_{k+1},\dots,x_d)|}{\int_0^1 dz_1 \cdots \int_0^1 dz_k \rho(z_1,\dots,z_k,x_{k+1},\dots,x_d)} \\
&\leq 2 \sup_{x\in[0,1]^d} \frac{|\partial_j \rho(x)|}{\rho(x)} \leq 2 \sup_{x\in[0,1]^d} \frac{|\nabla\rho(x)|}{\rho(x)}.
\end{aligned} \tag{2.2.0.22}$$

Combining (2.2.0.21) and (2.2.0.22) with (2.2.0.20), we get

$$\sup_{z\in[0,1]^d} \left\| D((T^{(k)})^{-1})(z) - \mathrm{Id} \right\| \leq C\sqrt{d-k+1} \sup_{x\in[0,1]^d} \frac{|\nabla\rho(x)|}{\rho(x)},$$

which concludes the first step.

Step 2: prove that

$$\sup_{z\in[0,1]^d} \left\| DT^{(k)}(z) - \mathrm{Id} \right\| \leq C \sup_{z\in[0,1]^d} \frac{|\nabla\rho(z)|}{\rho(z)}.$$

The estimate of this quantity is very similar to the previous one. Indeed, equation (2.2.0.19) can also be read as

$$x_k = \frac{\int_0^1 dy_1 \cdots \int_0^1 dy_{k-1} \int_0^{S_{x_{k+1},\dots,x_d}(x_k)} dy_k \rho(y_1,\dots,y_k,x_{k+1},\dots,x_d)}{\int_0^1 dz_1 \cdots \int_0^1 dz_k \rho(z_1,\dots,z_k,x_{k+1},\dots,x_d)}, \tag{2.2.0.23}$$

and since

$$\frac{\partial (T^{(k)})_i}{\partial x_j} = \begin{cases} \delta_{ij} & i \neq k, \\ 0 & i = k, j = 1,\dots,k-1, \\ \partial_j S_{x_{k+1},\dots,x_d}(x_k) & i = k, j = k,\dots,d, \end{cases}$$

we have

$$\left\| DT^{(k)}(x) - \mathrm{Id} \right\| \leq \sqrt{\left(\partial_k S_{x_{k+1},\dots,x_d}(x_k) - 1\right)^2 + \sum_{j=k+1}^{d} \left(\partial_j S_{x_{k+1},\dots,x_d}(x_k)\right)^2}, \tag{2.2.0.24}$$

and we can estimate this quantity using (2.2.0.23) as follows.

Differentiating both the sides of (2.2.0.23) with respect to x_k we get

$$\begin{aligned}
&\left|\partial_k S_{x_{k+1},\ldots,x_d}(x_k)-1\right| \\
&=\left|\frac{\int_0^1 dz_1\cdots\int_0^1 dz_{k-1}\int_0^1 dz_k[\rho(z_1,\ldots,z_k,x_{k+1},\ldots,x_d)-\rho(z_1,\ldots,z_{k-1},S_{x_{k+1},\ldots,x_d}(x_k),x_{k+1},\ldots,x_d)]}{\int_0^1 dy_1\cdots\int_0^1 dy_{k-1}\rho(y_1,\ldots,y_{k-1},S_{x_{k+1},\ldots,x_d}(x_k),x_{k+1},\ldots,x_d)}\right| \\
&\le \frac{\int_0^1 dz_1\cdots\int_0^1 dz_{k-1}\int_0^1 dz_k|\rho(z_1,\ldots,z_k,x_{k+1},\ldots,x_d)-\rho(z_1,\ldots,z_{k-1},S_{x_{k+1},\ldots,x_d}(x_k),x_{k+1},\ldots,x_d)|}{\int_0^1 dy_1\cdots\int_0^1 dy_{k-1}\rho(y_1,\ldots,y_{k-1},S_{x_{k+1},\ldots,x_d}(x_k),x_{k+1},\ldots,x_d)} \\
&\le C\sup_{z\in[0,1]^d}\frac{|\nabla\rho(z)|}{\rho(z)}\frac{\int_0^1 dz_1\cdots\int_0^1 dz_{k-1}\int_0^1 dz_k\rho(z_1,\ldots,z_{k-1},S_{x_{k+1},\ldots,x_d}(x_k),x_{k+1},\ldots,x_d)}{\int_0^1 dy_1\cdots\int_0^1 dy_{k-1}\rho(y_1,\ldots,y_{k-1},S_{x_{k+1},\ldots,x_d}(x_k),x_{k+1},\ldots,x_d)} \\
&= C\sup_{z\in[0,1]^d}\frac{|\nabla\rho(z)|}{\rho(z)},
\end{aligned}\tag{2.2.0.25}$$

where in the second inequality we used (2.2.0.10) of Lemma 2.2.5.

Moreover, differentiating with respect to x_j with $j\ge k+1$ one has

$$\begin{aligned}
&|\partial_j S_{x_{k+1},\ldots,x_d}(x_k)| \\
&=\Bigg|-\frac{\int_0^1 dy_1\cdots\int_0^1 dy_{k-1}\int_0^{S_{x_{k+1},\ldots,x_d}(x_k)} dy_k\partial_j\rho(y_1,\ldots,y_k,x_{k+1},\ldots,x_d)}{\int_0^1 dz_1\cdots\int_0^1 dz_{k-1}\rho(z_1,\ldots,z_{k-1},S_{x_{k+1},\ldots,x_d}(x_k),x_{k+1},\ldots,x_d)} \\
&+\frac{\int_0^1 dy_1\cdots\int_0^1 dy_{k-1}\int_0^{S_{x_{k+1},\ldots,x_d}(x_k)} dy_k\rho(y_1,\ldots,y_k,x_{k+1},\ldots,x_d)}{\int_0^1 dz_1\cdots\int_0^1 dz_{k-1}\rho(z_1,\ldots,z_{k-1},S_{x_{k+1},\ldots,x_d}(x_k),x_{k+1},\ldots,x_d)}\frac{\int_0^1 du_1\cdots\int_0^1 du_k\partial_j\rho(u_1,\ldots,u_k,x_{k+1},\ldots,x_d)}{\int_0^1 dw_1\cdots\int_0^1 dw_k\rho(w_1,\ldots,w_k,x_{k+1},\ldots,x_d)}\Bigg| \\
&\le \frac{\int_0^1 dy_1\cdots\int_0^1 dy_{k-1}\int_0^1 dy_k|\partial_j\rho(y_1,\ldots,y_k,x_{k+1},\ldots,x_d)|}{\int_0^1 dz_1\cdots\int_0^1 dz_{k-1}\rho(z_1,\ldots,z_{k-1},S_{x_{k+1},\ldots,x_d}(x_k),x_{k+1},\ldots,x_d)} \qquad (2.2.0.26)\\
&+\frac{\int_0^1 dy_1\cdots\int_0^1 dy_{k-1}\int_0^1 dy_k\rho(y_1,\ldots,y_k,x_{k+1},\ldots,x_d)}{\int_0^1 dz_1\cdots\int_0^1 dz_{k-1}\rho(z_1,\ldots,z_{k-1},S_{x_{k+1},\ldots,x_d}(x_k),x_{k+1},\ldots,x_d)}\frac{\int_0^1 du_1\cdots\int_0^1 du_k|\partial_j\rho(u_1,\ldots,u_k,x_{k+1},\ldots,x_d)|}{\int_0^1 dw_1\cdots\int_0^1 dw_k\rho(w_1,\ldots,w_k,x_{k+1},\ldots,x_d)}. \qquad (2.2.0.27)
\end{aligned}$$

Now in the term in the integral in the numerator of (2.2.0.26) can be estimated as follows:

$$\begin{aligned}
|\partial_j\rho(y_1,\ldots,y_k,x_{k+1},\ldots,x_d)| &\le \sup_{v\in[0,1]^d}\frac{|\partial_j\rho(v)|}{\rho(v)}\rho(y_1,\ldots,y_k,x_{k+1},\ldots,x_d) \\
&\le \sup_{v\in[0,1]^d}\frac{|\nabla\rho(v)|}{\rho(v)}\sup_{u,w\in[0,1]^d}\frac{\rho(u)}{\rho(w)}\rho(y_1,\ldots,y_{k-1},S_{x_{k+1},\ldots,x_d}(x_k),x_{k+1},\ldots,x_d) \\
&\le \sup_{v\in[0,1]^d}\frac{|\nabla\rho(v)|}{\rho(v)}\left(1+C\sup_{v\in[0,1]^d}\frac{|\nabla\rho(v)|}{\rho(v)}\right)\rho(y_1,\ldots,y_{k-1},S_{x_{k+1},\ldots,x_d}(x_k),x_{k+1},\ldots,x_d)
\end{aligned}\tag{2.2.0.28}$$

where the last inequality follows from (2.2.0.10) of Lemma 2.2.5.

Using again (2.2.0.10) of Lemma 2.2.5, the term inside the first integral in the numerator of (2.2.0.27) can be estimated as

$$\begin{aligned}
\rho(y_1,\ldots,y_k,x_{k+1},\ldots,x_d) &\le \sup_{u,w\in[0,1]^d}\frac{\rho(u)}{\rho(w)}\rho(y_1,\ldots,y_{k-1},S_{x_{k+1},\ldots,x_d}(x_k),x_{k+1},\ldots,x_d) \\
&\le \left(1+C\sup_{v\in[0,1]^d}\frac{|\nabla\rho(v)|}{\rho(v)}\right)\rho(y_1,\ldots,y_{k-1},S_{x_{k+1},\ldots,x_d}(x_k),x_{k+1},\ldots,x_d),
\end{aligned}\tag{2.2.0.29}$$

while the second one as

$$\begin{aligned}
|\partial_j\rho(u_1,\ldots,u_k,x_{k+1},\ldots,x_d)| &\le \sup_{v\in[0,1]^d}\frac{|\partial_j\rho(v)|}{\rho(v)}\rho(u_1,\ldots,u_k,x_{k+1},\ldots,x_d) \\
&\le \sup_{v\in[0,1]^d}\frac{|\nabla\rho(v)|}{\rho(v)}\rho(u_1,\ldots,u_k,x_{k+1},\ldots,x_d).
\end{aligned}\tag{2.2.0.30}$$

Inserting (2.2.0.28) inside (2.2.0.26) and then (2.2.0.29) and (2.2.0.30) into (2.2.0.27), one gets

$$|\partial_j S_{x_{k+1},\ldots,x_d}(x_k)|\le 2\sup_{v\in[0,1]^d}\frac{|\nabla\rho(v)|}{\rho(v)}\left(1+C\sup_{v\in[0,1]^d}\frac{|\nabla\rho(v)|}{\rho(v)}\right). \tag{2.2.0.31}$$

Combining (2.2.0.25) and (2.2.0.31) with (2.2.0.24) one finds out that

$$\left\| DT^{(k)}(x) \right\| \le C\sqrt{d-k+1} \sup_{z\in Q} \frac{|\nabla \rho(z)|}{\rho(z)}.$$

and therefore the proof of **Step 2** is concluded. Hence, the proof of the Lemma is concluded too. □

We now prove another result concerning the boundary Wasserstein distance between measures with different masses concentrated on $[0,1]^d$.

Lemma 2.2.7 ([11], Lemma 2.5). *For any $0 \le A \le B$ positive constants and $p \ge 1$, it holds*

$$Wb_p^p \left(A\mathbb{1}_{(0,1)^d}, B\mathbb{1}_{(0,1)^d} \right) \le \frac{(B-A)^p}{B^{p-1}}.$$

Proof. The proof is quite simple, indeed we can find a one dimensional transport plan between the measures. We first define the one-dimensional map

$$T : [0,1] \to [0,1], \qquad T(x) := \begin{cases} \frac{B}{A}x, & x \in \left[0, \frac{A}{B}\right), \\ 1, & x \in \left[\frac{A}{B}, 1\right], \end{cases} \tag{2.2.0.32}$$

and then, using T, we define

$$S : [0,1]^d \mapsto [0,1]^d, \qquad S(x) := (T(x_1), x_2, \dots, x_d). \tag{2.2.0.33}$$

We first claim that the measure

$$\pi := (S, \mathrm{Id})_\sharp B\mathbb{1}_{[0,1]^d},$$

is a boundary coupling of $A\mathbb{1}_{(0,1)^d}$ and $B\mathbb{1}_{(0,1)^d}$. Indeed the second marginal of π restricted to $(0,1)^d$ is $B\mathbb{1}_{(0,1)^d}$ by definition. Moreover, since

$$\begin{aligned}
\int_{[0,1]^d\times[0,1]^d} d\pi(x,y) f(x) &= B\int_{[0,1]^d} dy f(S(y)) \quad \text{by Definition 2.2.1 of push-forward} \\
&= B\int_0^1 dy_1 \int_{[0,1]^{d-1}} d\hat{y} f(T(y_1),\hat{y}) \quad \text{by definition of } S \\
&= B\int_0^{\frac{A}{B}} dy_1 \int_{[0,1]^{d-1}} d\hat{y} f\left(\frac{B}{A}y_1, \hat{y}\right) + B\int_{\frac{A}{B}}^1 dy_1 \int_{[0,1]^{d-1}} d\hat{y} f(1,\hat{y}) \quad \text{by definition of } T \\
&= A\int_{[0,1]^d} dy f(y) + (B-A)\int_{[0,1]^{d-1}} d\hat{y} f(1,\hat{y}),
\end{aligned}$$

the first marginal of π restricted to $(0,1)^d$ is $A\mathbb{1}_{(0,1)^d}$.

Hence we have

$$\begin{aligned}
Wb_p^p \left(A\mathbb{1}_{(0,1)^d}, B\mathbb{1}_{(0,1)^d} \right) &\le \int_{[0,1]^d\times[0,1]^d} d\pi(x,y)|x-y|^p \\
&= B\int_{[0,1]^d} dy |S(y)-y|^p \\
&= B\int_0^1 dy_1 |T(y_1)-y_1|^p \\
&= B\int_0^{\frac{A}{B}} dy_1 \left|\frac{B}{A}y_1 - y_1\right|^p + B\int_{\frac{A}{B}}^1 dy_1 \, |1-y_1|^p \\
&= \frac{(B-A)^p}{B^{p-1}} \int_0^1 dz z^p = \frac{1}{p+1}\frac{(B-A)^p}{B^{p-1}}.
\end{aligned}$$

where in the last but one equality we have changed variables $y_1 = \frac{A}{B}z$ in the first integral and $1-y_1 = \left(1-\frac{A}{B}\right)z$ in the second one. Hence the proof is concluded. □

2.3 The Benamou-Brenier formula.

The Benamou-Brenier formulation of the Wasserstein distance between two measures allows to write an optimal transport problem in terms of a PDE's problem. To introduce that, we need the notion of continuity equation for measures.

The continuity equation.

It is well known that for a regular enough

$$v : [0,+\infty) \times \Omega(\subseteq \mathbb{R}^d) \to \Omega,$$

tangent to $\partial\Omega$, denoting $v_t(x) = v(t,x)$, the unique solution to the problem

$$\begin{cases} \partial_t \mu_t(x) + \mathrm{div}(\mu_t(x) v_t(x)) = 0, & x \in \Omega, t \geq 0, \\ \mu_0(x) = \mu(x), \end{cases} \tag{2.3.0.1}$$

with $\mu : \Omega \to \mathbb{R}$, is given by (see [31, Section 4.1.2])

$$\mu_t = (\varphi_t)_\sharp \mu, \tag{2.3.0.2}$$

where φ_t solves

$$\begin{cases} \partial_t \varphi_t(x) = v_t(\varphi_t(x)), \\ \varphi_0(x) = x. \end{cases} \tag{2.3.0.3}$$

Indeed, if φ_t is a solution to (2.3.0.3), using that

$$\partial_t \varphi_t(x) = v_t(\varphi_t(x)), \quad \frac{d}{dt} \det(D\varphi_t(x)) = (\mathrm{div} v_t)(\varphi_t(x)) \det(D\varphi_t(x)),$$

we have

$$\begin{aligned} \frac{d}{dt}(\varphi_t^{-1})_\sharp \mu_t(x) &= \frac{d}{dt}\left[\mu_t(\varphi_t(x)) \det(D\varphi_t(x))\right] \\ &= \left[\partial_t \mu_t(\varphi_t(x)) + \nabla \mu_t(\varphi_t(x)) \cdot \partial_t \varphi_t(x)\right] \det(D\varphi_t(x)) + \mu_t(\varphi_t(x)) \frac{d}{dt} \det(D\varphi_t(x)) \\ &= \left[\partial_t \mu_t(\varphi_t(x)) + \nabla \mu_t(\varphi_t(x)) \cdot v_t(\varphi_t(x)) + \mu_t(\varphi_t(x))(\mathrm{div} v_t)(\varphi_t(x))\right] \det(D\varphi_t(x)) \\ &= \left[\partial_t \mu_t + \mathrm{div}(\mu_t v_t)\right](\varphi_t(x)) \det(D\varphi_t(x)), \end{aligned}$$

hence equation (2.3.0.1) is satisfied if and only if the time derivative above is zero and therefore

$$(\varphi_t^{-1})_\sharp \mu_t \equiv (\varphi_0^{-1})_\sharp \mu_0 = \mu,$$

that is, (2.3.0.2).

Now we can go back to the Benamou-Brenier formulation of the transport problem.

Theorem 2.3.1 ([6], Benamou-Brenier formula)**.** *Let μ and ν be two positive measures on $\Omega \subseteq \mathbb{R}^d$. Then, one has*

$$W_p^p(\mu,\nu) = \inf \left\{ \int_0^1 dt \int_\Omega d\mu_t(x) |v_t(x)|^p : \begin{array}{l} \partial_t \mu_t + div(\mu_t v_t) = 0, \\ v_t \cdot \hat{n} = 0 \text{ on } \partial\Omega, \\ \mu_0 = \mu, \mu_1 = \nu \end{array} \right\}.$$

We are not going to prove this equality, but actually we only need the upper bound, which is quite immediate to check. For further details see also [31, Section 6.1].

If the couple (μ_t, v_t) is admissible for the infimum in the Theorem above and φ_t is the solution to (2.3.0.3) for such v_t, one has

$$\nu = \mu_{t=1} = (\varphi_{t=1})_\sharp \mu_{t=0} = (\varphi_{t=1})_\sharp \mu,$$

and therefore

$$\begin{aligned} W_p^p(\mu,\nu) = W_p^p(\mu, (\varphi_1)_\sharp \mu) &\leq \int_\Omega d\mu(x) |x - \varphi_1(x)|^p \qquad \text{by (2.2.0.1)} \\ &= \int_\Omega d\mu(x) |\varphi_0(x) - \varphi_1(x)|^p = \int_\Omega d\mu(x) \left| \int_0^1 dt \partial_t \varphi_t(x) \right|^p \\ &= \int_\Omega d\mu(x) \left| \int_0^1 dt v_t(\varphi_t(x)) \right|^p \leq \int_\Omega d\mu(x) \int_0^1 dt |v_t(\varphi_t(x))|^p \qquad \text{by convexity of the function } |\cdot|^p \\ &= \int_0^1 dt \int_\Omega d\mu(x) |v_t(\varphi_t(x))|^p = \int_0^1 dt \int_\Omega d(\varphi_t)_\sharp \mu(x) |v_t(x)|^p \qquad \text{by Definition 2.2.1 of push-forward} \\ &= \int_0^1 dt \int_\Omega d\mu_t(x) |v_t(x)|^p. \end{aligned}$$

As a consequence of Benamou-Brenier formula, one gets the following estimate, which is no more symmetric in μ and ν.

Proposition 2.3.1 ([1], Corollary 4.4 and [33], Chapter 1)**.** *Let $\Omega \subseteq \mathbb{R}^d$ be an open and bounded set, and let μ and ν be two densities on Ω such that $\mu(\Omega) = \nu(\Omega)$. Then, if $f : \mathbb{R} \to \mathbb{R}$ is the solution to*

$$\begin{cases} \Delta f = \mu - \nu \text{ on } \Omega, \\ \nabla \cdot f = 0 \text{ on } \partial\Omega, \end{cases}$$

it holds

$$W_p^p(\mu, \nu) \leq c_p \int_\Omega dx \frac{|\nabla f(x)|^p}{\mu(x)^{p-1}},$$

for a constant $c_p > 0$.

Proof. If instead for looking for v_t as in the original statement of the Benamou-Brenier formula, one looks for f_t such that $\nabla f_t = \mu_t v_t$, one gets

$$\begin{cases} \partial_t \mu_t + \Delta f_t = 0, \\ \nabla f_t \cdot \hat{n} = 0 \text{ on } \partial\Omega, \\ \mu_0 = \mu, \mu_1 = \nu. \end{cases} \tag{2.3.0.4}$$

and

$$W_p^p(\mu, \nu) \leq \int_0^1 dt \int_\Omega dx \frac{|\nabla f_t(x)|^p}{\mu_t(x)^{p-1}}. \tag{2.3.0.5}$$

The simplest thing would be to use (2.3.0.4) and (2.3.0.5) choosing for μ_t the linear interpolation between μ and ν, but doing that one gets

$$\mu_t = \mu + t(\nu - \mu) = (1-t)\mu + t\nu \geq (1-t)\mu,$$

and one would obtain $(1-t)^{-p+1}$ in the denominator which is not integrable for t close to 1 and $p \geq 2$. Hence, we make an intermediate step: by triangle inequality we have

$$W_p(\mu, \nu) \leq W_p\left(\mu, \frac{\mu+\nu}{2}\right) + W_p\left(\frac{\mu+\nu}{2}, \nu\right),$$

and substituting into the previous estimate the following one

$$W_p^p\left(\frac{\mu+\nu}{2}, \nu\right) = W_p^p\left(\frac{\mu}{2} + \frac{\nu}{2}, \frac{\nu}{2} + \frac{\nu}{2}\right) \leq W_p^p\left(\frac{\mu}{2}, \frac{\nu}{2}\right) + W_p^p\left(\frac{\nu}{2}, \frac{\nu}{2}\right) = \frac{1}{2} W_p^p(\mu, \nu),$$

one gets

$$W_p(\mu, \nu) \leq W_p\left(\mu, \frac{\mu+\nu}{2}\right) + \frac{1}{2^{\frac{1}{p}}} W_p(\mu, \nu),$$

that is

$$W_p(\mu, \nu) \leq \frac{2^{\frac{1}{p}}}{2^{\frac{1}{p}} - 1} W_p\left(\mu, \frac{\mu+\nu}{2}\right). \tag{2.3.0.6}$$

Hence in (2.3.0.4) and (2.3.0.5) we substitute ν with

$$\frac{\mu+\nu}{2}.$$

Indeed if f is chosen as in the hypothesis, one gets

$$\Delta\left(\frac{f}{2}\right) = \frac{\mu - \nu}{2} = \mu - \frac{\mu+\nu}{2},$$

and if

$$\mu_t := \mu + t\left(\frac{\mu+\nu}{2} - \mu\right) = \mu\left(1 - \frac{t}{2}\right) + \frac{t}{2}\nu \geq \frac{\mu}{2},$$

then μ_t and $f_t \equiv \frac{f}{2}$ satisfy (2.3.0.4) for μ and $\frac{\mu+\nu}{2}$ and by (2.3.0.5)

$$W_p^p\left(\mu, \frac{\mu+\nu}{2}\right) \leq \int_0^1 dt \int_\Omega dx \frac{|\nabla f_t(x)|^p}{\mu_t(x)^{p-1}} = 2^{-p} \int_0^1 dt \int_\Omega dx \frac{|\nabla f(x)|^p}{\mu_t(x)^{p-1}} \leq \frac{1}{2} \int_\Omega dx \frac{|\nabla f(x)|^p}{\mu(x)^{p-1}},$$

and then one can conclude using (2.3.0.6). □

2.4 An estimate exploiting radial ad angular symmetries.

Here we investigate the Wasserstein distance between two measures on $\mathbb{R}^d$ whose radial or angular part is the same. In particular, the following Lemma states that if the measures have the same radial marginal, then the Wasserstein distance can be estimated with the distance between the angular marginals on ∂B_1, up to rescaling with a radial factor, while if the two measures have the same angular component, then the distance can be bounded with the distance between the radial marginals.

Since the result could be confusing, in the next result we will denote by

$$W_{p,\Omega,d},$$

the Wasserstein distance of order p between two measures defined on a metric space (Ω, d).

Lemma 2.4.1. *Let Ω and Σ be two metric spaces, let μ, ν be two measures on Ω such that $\mu(\Omega) = \nu(\Omega)$, and let γ be a measure on Σ. Then, for any $\pi \in \Gamma(\mu, \nu)$, it holds*

$$W_p^p(\mu \otimes \gamma, \nu \otimes \gamma) \leq \int_{\Omega \times \Omega} d\pi(x, x') \int_\Sigma d\gamma(y) d^p((x, y), (x', y)). \tag{2.4.0.1}$$

In particular, if the two spaces are $[0, +\infty)$ and ∂B_1, with distance

$$d((r, x), (r', x')) = |rx - r'x'|,$$

one has two cases: if $\Omega = [0, +\infty)$ and $\Sigma = \partial B_1$, it holds

$$W_p^p(\mu \otimes \gamma, \nu \otimes \gamma) \leq W^p_{p,[0,+\infty),|\cdot|}(\mu, \nu)\gamma(\partial B_1), \tag{2.4.0.2}$$

while if $\Omega = \partial B_1$ and $\Sigma = [0, +\infty)$, it holds

$$W_p^p(\mu \otimes \gamma, \nu \otimes \gamma) \leq W^p_{p,\partial B_1, d_{\partial B_1}}(\mu, \nu) \int_0^\infty d\gamma(r) r^p, \tag{2.4.0.3}$$

where the distance $d_{\partial B_1}$ on ∂B_1 is the geodesic distance on the sphere.

Proof. We start with the proof of (2.4.0.1), which follows from the fact that if π is the coupling in the statement, then

$$\tilde{\pi}((x, y), (x', y')) := \pi(x, x')\gamma(y)\delta_y(y'),$$

is a coupling of $\mu \otimes \gamma$ and $\nu \otimes \gamma$. To prove this, we consider a measurable function $f : \Omega \otimes \Sigma \to \mathbb{R}$. By definition one has

$$\begin{aligned} \int_{(\Omega\times\Sigma)^2} d\tilde{\pi}((x, y), (x', y'))f(x, y) &= \int_{\Omega\times\Omega} d\pi(x, x') \int_\Sigma d\gamma(y) f(x, y) \\ &= \int_\Omega d\mu(x) \int_\Sigma d\gamma(y) f(x, y) \\ &= \int_{\Omega\times\Sigma} d(\mu \otimes \gamma)(x, y) f(x, y), \end{aligned}$$

hence the first marginal of $\tilde{\pi}$ is $\mu \otimes \gamma$, and its second marginal is $\nu \otimes \gamma$ for the same reason.

Therefore, it holds

$$W_p^p(\mu \otimes \gamma, \nu \otimes \gamma) \leq \int_{(\Omega\times\Sigma)^2} d\tilde{\pi}((x, y), (x', y'))d^p((x, y), (x', y')) = \int_{\Omega\times\Omega} d\pi(x, x') \int_\Sigma d\gamma(y) d^p((x, y), (x', y')),$$

hence (2.4.0.1) is proved. Now we shall use it to prove (2.4.0.2) and (2.4.0.3), and we begin with (2.4.0.2): if $\Omega = [0, +\infty)$ and $\Sigma = \partial B_1$, for any $x \in \partial B_1$ and $r, r' \geq 0$ one has

$$d((r, x), (r', x)) = |rx - r'x| = |r - r'||x| = |r - r'|,$$

and therefore by (2.4.0.1), for any $\pi \in \Gamma(\mu, \nu)$, one gets

$$W_p^p(\mu \otimes \gamma, \nu \otimes \gamma) \leq \int_{[0,+\infty)^2} d\pi(r, r') \int_{\partial B_1} d\gamma(y) d^p((r, x), (r', x)) = \int_{[0,+\infty)^2} d\pi(r, r')|r - r'|^p \gamma(\partial B_1),$$

and by minimizing over $\pi \in \Gamma(\mu, \nu)$ one gets the thesis.

As for (2.4.0.3), one has

$$d((r, x), (r, x')) = |rx - rx'| = r|x - x'| \leq r d_{\partial B_1}(x, x'),$$

and therefore

$$W_p^p(\mu \otimes \gamma, \nu \otimes \gamma) \leq \int_{\partial B_1\times\partial B_1} d\pi(x, x') \int_0^\infty d\gamma(r) r^p d^p_{\partial B_1}(x, x'),$$

and as before the thesis follows optimizing over $\pi \in \Gamma(\mu, \nu)$. ∎

Chapter 3

Some properties of binomial random variables.

In this Chapter we recall some notions which we need about the binomial random variables.

Let us begin with the definition of binomial random variable.

Definition 3.0.1 ([30], Section 2.2.2, Binomial random variable). *Let $(\Omega, \Sigma, \mathbb{P})$ a probability space. A random variable $N : \Omega \to \mathbb{N}$ is a Binomial random variable of parameters $n \in \mathbb{N}$ and $p \in (0, 1)$ if*

$$\mathbb{P}(N = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \forall k = 0, 1, \ldots, n.$$

Let us recall also that a Binomial random variable can be obtained the sum of n independent Bernoulli random variables, defined as follows.

Definition 3.0.2 ([30], Section 2.2.1, Bernoulli random variable). *Let $(\Omega, \Sigma, \mathbb{P})$ a probability space and $p \in (0, 1)$. A random variable $Z : \Omega \to \{0, 1\}$ is a Bernoulli random variable of parameter p if*

$$\mathbb{P}(Z = 1) = p, \quad \mathbb{P}(Z = 0) = 1 - p.$$

Let us recall that if $\{Z_i\}_{i=1}^n$ are iid Bernoulli random variables with parameter p, then

$$N := \sum_{i=1}^{n} Z_i, \tag{3.0.0.1}$$

is a Binomial random variable with parameters n and p. In particular, if $\{X_i\}_{i=1}^\infty$ are iid random variables in $\mathbb{R}^d$ with common probability distribution μ, the random variable

$$n(A) = \sum_{i=1}^{n} \mathbb{1}_A(X_i),$$

is a binomial random variable of parameters n and p with

$$p = \mathbb{P}(\mathbb{1}_A(X_1) = 1) = \mathbb{P}(X_1 \in A) = \mu(A).$$

We now state a first result concerning binomial random variables.

Lemma 3.0.1. *Let N be a binomial random variable of parameters n and p and let $\{Z_i\}_{i=1}^\infty$ be Bernoulli random variables independent of N and of parameter q. Then, the random variable*

$$M := \sum_{i=1}^{N} Z_i,$$

is a binomial random variable of parameters n and pq.

Proof. To prove that, notice that for any $k \in \{0, 1, \ldots, n\}$, one has

$$\begin{aligned}
\mathbb{P}(M = k) &= \sum_{\ell=k}^{n} \mathbb{P}(M = k | N = \ell) \mathbb{P}(N = \ell) \\
&= \sum_{\ell=k}^{n} \binom{\ell}{k} q^k (1-q)^{\ell-k} \binom{n}{\ell} p^\ell (1-p)^{n-\ell},
\end{aligned}$$

since conditioned to N, M is a binomial random variable of parameters N and q (see (3.0.0.1)). Hence

$$\begin{aligned}\mathbb{P}(M=k) &= \binom{n}{k}(pq)^k \sum_{\ell=k}^{n} \binom{n-k}{\ell-k}[p(1-q)]^{\ell-k}(1-p)^{n-\ell} \\ &= \binom{n}{k}(pq)^k \sum_{j=0}^{n-k} \binom{n-k}{j}[p(1-q)]^{j}(1-p)^{n-k-j} \\ &= \binom{n}{k}(pq)^k(1-pq)^{n-k},\end{aligned}$$

and therefore the proof is concluded. □

In particular, thanks to Lemma 3.0.1, one can conclude that if N is a binomial random variable of parameters n and p and $\{X_i\}_{i=1}^{\infty}$ is a sequence of random variables independent of N and distributed in $\mathbb{R}^d$ with common probability distribution μ, then the random variable

$$N(A) := \sum_{i=1}^{N} \mathbb{1}(X_i \in A), \tag{3.0.0.2}$$

is a binomial random variable of parameters n and $p\mu(A)$.

Finally, before going on let us briefly recall that if X is a Bernoulli random variable of parameter p, for $d \in \mathbb{N}$ one has

$$|\mathbb{E}[(X-p)^d]| = p(1-p)\left|(1-p)^{d-1} + (-1)^d p^{d-1}\right| \leq \begin{cases} 1, & \text{if } d=0, \\ 0, & \text{if } d=1, \\ p(1-p), & \text{if } d \geq 2, \end{cases} \tag{3.0.0.3}$$

where the inequality is actually an equality for $d=2$.

3.1 Concentration inequalities.

The main result we need concerns how to estimate the q^{th} moment of $n(A)$, and this is the content of the following Lemma.

Lemma 3.1.1. *Let N be a binomial random variable with parameters n and p with $np(1-p) \geq c > 0$. Then, for any $q \geq 2$ there exists a constant C depending on q and c such that*

$$\mathbb{E}[|N-np|^q] \leq C(np(1-p))^{\frac{q}{2}}.$$

Before proving it, we recall that it can be obtained by Proposition 3.1.1 as Lemma 2.2 in [23], but we prefer using the following direct proof.

Proof. Let us prove it only if $q=2k$ and $k \in \mathbb{N}\setminus\{0\}$. It is sufficient to our purposes since, if the result holds for such class of the exponents $q=2k$, then it holds for any $q \geq 2$. Indeed, if one defines

$$k(q) = \min\{k \in \mathbb{N}\setminus\{0\} : 2k \geq q\},$$

by Young inequality one gets

$$\mathbb{E}[|N-np|^q] \leq \mathbb{E}[|N-np|^{2k(q)}]^{\frac{q}{2k(q)}} \leq \left(C(np(1-p))^{\frac{2k(q)}{2}}\right)^{\frac{q}{2k(q)}} = C^{\frac{q}{2k(q)}}(np(1-p))^{\frac{q}{2}},$$

with $C>0$ depending on $2k(q)$ and therefore on q.

Hence we assume $q=2k$ and $k \in \mathbb{N}\setminus\{0\}$. In such case, if $\{X_i\}_{i=1}^{n}$ are iid Bernoulli random variables with parameter p, one has

$$\mathbb{E}[|N-np|^q] = \mathbb{E}[(N-np)^{2k}] = \mathbb{E}\left[\left(\sum_{i=1}^{n}(X_i-p)\right)^{2k}\right] = \sum_{i_1,\dots,i_{2k}=1}^{n} \mathbb{E}\left[(X_{i_1}-p)\cdots(X_{i_{2k}}-p)\right], \tag{3.1.0.1}$$

and therefore, substituting

$$E_\ell = \{j \in \{1,\dots,2k\} : i_j = \ell,\}, \quad \ell = 1,\dots,n,$$

one gets

$$
\begin{aligned}
(3.1.0.1) &= \sum_{E_1,\dots,E_n:\bigsqcup_{\ell=1}^n E_\ell=\{1,\dots,2k\}} \mathbb{E}\left[\prod_{\ell=1}^n \prod_{j\in E_\ell}(X_{i_j}-p)\right] \\
&= \sum_{E_1,\dots,E_n:\bigsqcup_{\ell=1}^n E_\ell=\{1,\dots,2k\}} \mathbb{E}\left[\prod_{\ell=1}^n (X_\ell-p)^{|E_\ell|}\right] \qquad \text{by definition of } E_\ell \\
&= \sum_{E_1,\dots,E_n:\bigsqcup_{\ell=1}^n E_\ell=\{1,\dots,2k\}} \prod_{\ell=1}^n \mathbb{E}\left[(X_\ell-p)^{|E_\ell|}\right] \qquad \text{by independence of } \{X_i\}_{i=1}^n \\
&\le \sum_{E_1,\dots,E_n:\bigsqcup_{\ell=1}^n E_\ell=\{1,\dots,2k\},|E_\ell|\neq 1\forall\ell} [p(1-p)]^{\sharp\{\ell:|E_\ell|\ge 2\}} \qquad \text{by (3.0.0.3).}
\end{aligned}
$$

Now we can perform the previous sum by counting the number of non empty sets among $\{E_1,\dots,E_n\}$. Indeed, if one denotes by m such number, one can notice that

$$
2k=\sum_{\ell=1}^n |E_\ell| \ge \sum_{\ell:|E_\ell|\ge 2} |E_\ell| \ge 2\sum_{\ell:|E_\ell|\ge 2} = 2m, \quad \Rightarrow m\in\{1,\dots,k\},
$$

hence one can write

$$
\begin{aligned}
\sum_{E_1,\dots,E_n:\bigsqcup_{\ell=1}^n E_\ell=\{1,\dots,2k\},|E_\ell|\neq 1\forall\ell} &= \sum_{E_1,\dots,E_n:\bigsqcup_{\ell=1}^n E_\ell=\{1,\dots,2k\}} \sum_{I\subseteq\{1,\dots,n\}} \mathbb{1}(|E_\ell|\ge 2\forall\ell\in I, |E_\ell|=0\forall\ell\notin I) \\
&= \sum_{E_1,\dots,E_n:\bigsqcup_{\ell=1}^n E_\ell=\{1,\dots,2k\}} \sum_{m=0}^k \sum_{I\subseteq\{1,\dots,n\},|I|=m} \mathbb{1}(|E_\ell|\ge 2\forall\ell\in I, |E_\ell|=0\forall\ell\notin I).
\end{aligned}
$$

Therefore, if $\{F_i\}_{i\in I}$ are the non empty sets among $E_1,\dots,E_n$ (which means $|F_i|\ge 2\forall i\in I$), one gets

$$
\begin{aligned}
(3.1.0.1) &\le \sum_{m=0}^k \binom{n}{m} \sum_{F_1,\dots,F_m:\bigsqcup_{\ell=1}^m F_\ell=\{1,\dots,2k\}|F_\ell|\ge 2\forall\ell} [p(1-p)]^m \\
&= \sum_{m=0}^k \binom{n}{m} [p(1-p)]^m \underbrace{\sum_{F_1,\dots,F_m:\bigsqcup_{\ell=1}^m F_\ell=\{1,\dots,2k\}|F_\ell|\ge 2\forall\ell}}_{:=C_{k,m}} \\
&= \sum_{m=0}^k \frac{1}{n^m}\frac{n!}{m!(n-m)!}[np(1-p)]^m C_{k,m} \\
&\le \underbrace{\max_{m=0,\dots,k} C_{k,m}}_{:=C_k} \sum_{m=0}^k \frac{1}{m!}[np(1-p)]^m \\
&\le C_k \sum_{m=0}^\infty \frac{1}{m!} \max_{m=0,\dots,k}[np(1-p)]^m \le C_k e C'[np(1-p)]^k,
\end{aligned}
$$

with $C'>0$ depending on k and on the fact that $np(1-p)\ge c$. □

3.1.1 Chernoff bound.

Thanks to the property of concentration of the binomial random variables, it is possible to estimate the probability of the event in which such random variable is far from its expected value. That is the aim of this Subsection.

Proposition 3.1.1 ([15], Chernoff bound). *Let N be a binomial random variables of parameters n and p. Then, for any $\delta\in[0,1]$ it holds*

$$
\mathbb{P}(N-np\ge np\delta)\le \exp\left\{-\frac{np\delta^2}{3}\right\} \quad \textit{and} \quad \mathbb{P}(N-np\le -np\delta)\le \exp\left\{-\frac{np\delta^2}{2}\right\},
$$

and, in particular

$$
\mathbb{P}(|N-np|\ge np\delta)\le 2\exp\left\{-\frac{\delta^2 np}{3}\right\}.
$$

Proof. First, we notice that the second statement is a consequence of the first one, indeed:

$$\mathbb{P}\left(|N-np|\geq np\delta\right)=\mathbb{P}\left(N\geq np(1+\delta)\right)+\mathbb{P}\left(N\leq np(1-\delta)\right)=:\mathbb{P}(A_1)+\mathbb{P}(A_2), \tag{3.1.1.1}$$

and we first estimate $\mathbb{P}(A_1)$, then we focus on $\mathbb{P}(A_2)$. By writing $N=\sum_{i=1}^n X_i$, with X_i iid Bernoulli random variables of parameter p, for any $t\geq 0$ one has

$$\begin{aligned}\mathbb{P}(A_1)&=\mathbb{P}(e^{tN}\geq e^{tnp(1+\delta)})\leq e^{-tnp(1+\delta)}\mathbb{E}[e^{tN}]=e^{-tnp(1+\delta)}\prod_{i=1}^n\mathbb{E}[e^{tX_i}]\\&=e^{-tnp(1+\delta)}(pe^t+(1-p))^n\leq\left(\frac{e^{p(e^t-1)}}{e^{tp(1+\delta)}}\right)^n,\end{aligned}$$

and since the minimum of the previous inequality is achieved for $t=\log(1+\delta)$, one gets

$$\mathbb{P}(A_1)\leq\exp\{np[\delta-(1+\delta)\log(1+\delta)]\}\leq\exp\left\{-\frac{np\delta^2}{\delta+2}\right\}\leq\exp\left\{-\frac{np\delta^2}{3}\right\}.$$

The estimate of $\mathbb{P}(A_2)$ is quite similar. One has

$$\begin{aligned}\mathbb{P}(A_2)&=\mathbb{P}(-N\geq -np(1-\delta))=\mathbb{P}(e^{-tN}\geq e^{-tnp(1-\delta)})\leq e^{tnp(1-\delta)}\mathbb{E}[e^{-tN}]=e^{tnp(1-\delta)}\prod_{i=1}^n\mathbb{E}[e^{-tX_i}]\\&=e^{tnp(1-\delta)}(pe^{-t}+(1-p))^n\leq\left(e^{-p(1-e^{-t})+tp(1-\delta)}\right)^n,\end{aligned}$$

and since here the minimum of the right hand side is achieved for $t=\log(1-\delta)$, one gets

$$\mathbb{P}(A_2)\leq\exp\left\{-np[\delta+(1-\delta)\log(1-\delta)]\right\}\leq\exp\left\{-\frac{np\delta^2}{2}\right\},$$

and using (3.1.1.1) the thesis is proved. $\square$

3.2 Some properties of polynomial functions of binomial random variables.

Now we focus on some results which will be useful in the next sections. All the result we prove can be written in a more general way, but for simplicity we only consider the case which we are interested in.

The first result is a Lemma which bounds the probability that a binomial random variable with special conditions on the parameters is smaller than a fixed value.

Lemma 3.2.1. *Let $c\in(0,1)$ and α be positive constants. Then, there exists $C<1$ depending on c and α such that for every N binomial random variable with parameters n and p with $n>\alpha$ and $np\geq c$, it holds*

$$\mathbb{P}(N\leq\alpha)\leq C.$$

Proof. First, we observe that we can assume that α is a natural number, indeed if the property holds for every natural number, we have $n>\alpha\geq\lfloor\alpha\rfloor$ and

$$\mathbb{P}(N\leq\alpha)=\mathbb{P}(N\leq\lfloor\alpha\rfloor)\leq C,$$

where $\lfloor\cdot\rfloor$ is the integer part.

We distinguish two cases: the case $np\geq 2\alpha$ and $np<2\alpha$. Let us start with the first one: by Proposition 3.1.1, if $np\geq 2\alpha$, we have

$$\mathbb{P}(N\leq\alpha)=\mathbb{P}(N-np\leq\alpha-np)=\mathbb{P}\left(N-np\leq-\frac{np}{2}\right)\leq\exp\left\{-\frac{np}{8}\right\}\leq\exp\left\{-\frac{\alpha}{4}\right\}<1.$$

We split the case $np<2\alpha$ in two more cases, as follows. We first denote by

$$\beta:=\beta(c,\alpha)=\max_{x\geq c}\sum_{k=0}^{\alpha}\frac{x^k}{k!}e^{-x}<1,$$

and then we notice that for $n > 2\alpha$ we have

$$\begin{aligned}
\mathbb{P}(N \leq \alpha) &= \sum_{k=0}^{\alpha} \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{by definition} \\
&\leq \sum_{k=0}^{\alpha} \frac{1}{k!} (np)^k (1-p)^n \left(1 - \frac{np}{n}\right)^{-k} \qquad \text{since } \frac{n!}{(n-k)!} \leq n^k \\
&\leq \left(1 - \frac{2\alpha}{n}\right)^{-\alpha} \sum_{k=0}^{\alpha} \frac{1}{k!} (np)^k e^{-np} \qquad \text{since } k \leq \alpha, np \leq 2\alpha \text{ and } \log(1-p) \leq -p \\
&\leq \left(1 - \frac{2\alpha}{n}\right)^{-\alpha} \beta. \qquad (3.2.0.1)
\end{aligned}$$

Now since $\beta < 1$ there exists $N_\alpha > 2\alpha$ (also depending on β and therefore on c) such that for $n \geq N_\alpha$ the right-hand side of the previous inequality is smaller than 1.

Hence, if $np \in [c, 2\alpha]$ and $n \geq N_\alpha$ we can conclude applying (3.2.0.1). Otherwise, in the case $np \in [c, 2\alpha]$ and $n \leq N_\alpha$, we can use the rough estimate

$$\mathbb{P}(N \leq \alpha) = \sum_{k=0}^{\alpha} \binom{n}{k} p^k (1-p)^{n-k} \leq \max_{m \in \{\alpha+1, \dots, N_\alpha\}, q \in [0,1], mq \in [c, 2\alpha]} \sum_{k=0}^{\alpha} \binom{m}{k} q^k (1-q)^{m-k} < 1.$$

$\square$

The next result is a Lemma concerning an estimate for the expectation of some function of the binomial random variables.

Lemma 3.2.2. *Let $c > 0$ be a positive constant and let $A \in \mathbb{R}$ and $B \geq 0$. Then, there exists a constant $C > 0$ depending on A, B and c such that for any N binomial random variable of parameters n and p such that $np \geq c$, it holds*

$$\mathbb{E}[(N+1)^A] \leq C(\mathbb{E}[N] + 1)^A, \qquad (3.2.0.2)$$

$$\mathbb{E}[(\log(N+1))^B] \leq C(\log(\mathbb{E}[N] + 1))^B. \qquad (3.2.0.3)$$

In particular, if

$$f(x) := (x+1)^A (\log(x+1))^B, \quad A \in \mathbb{R} \text{ and } B \geq 0,$$

for a positive constant C depending on A, B and c, it holds

$$\mathbb{E}[f(N)] \leq C f(\mathbb{E}[N]). \qquad (3.2.0.4)$$

Finally, if $M > c$, there exists a constant $C > 1$ depending on c and M such that for any N binomial random variables of parameters n and p with $np \in [c, M]$ it holds

$$\mathbb{E}[f(N)] \geq \frac{1}{C} f(\mathbb{E}[N]). \qquad (3.2.0.5)$$

Proof. We start with the proof of statement (3.2.0.2), which follows from direct computations. Notice that it is sufficient to prove the statement for integer values of the exponent A, indeed, supposing that (3.2.0.2) holds for every integer value of the exponent, for a general A by convexity one has

$$\begin{aligned}
\mathbb{E}\left[(N+1)^A\right] &= \mathbb{E}\left[\left((N+1)^{\frac{A}{|A|}}\right)^{|A|}\right] \\
&\leq \mathbb{E}\left[\left((N+1)^{\frac{A}{|A|}}\right)^{\lceil |A| \rceil}\right]^{\frac{|A|}{\lceil |A| \rceil}} \qquad \text{by convexity} \\
&\leq C((\mathbb{E}[N+1])^{\frac{A}{|A|}\lceil |A| \rceil})^{\frac{|A|}{\lceil |A| \rceil}} \qquad \text{since } \frac{A}{|A|}\lceil |A| \rceil \text{ is integer} \\
&= C(\mathbb{E}[N+1])^A,
\end{aligned}$$

where $\lceil \cdot \rceil$ is the ceiling function.

Proof of (3.2.0.2) **in the case** $A = 1, 2, \ldots$. Suppose therefore that $A \in \mathbb{N} \setminus 0$: we have to estimate

$$\mathbb{E}[(N+1)^A] = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} (k+1)^A.$$

If one defines

$$G_A(p) := \sum_{k=A}^{n} \binom{n}{k} p^{k-A} (1-p)^{n-k} \frac{k!}{(k-A)!},$$

since there exists a positive constant $C > 0$ depending on A such that for every $k \geq A$

$$(k+1)^A \leq C \frac{k!}{(k-A)!},$$

one has

$$\begin{aligned} \mathbb{E}[(N+1)^A] &= \mathbb{E}[(N+1)^A \mathbb{1}_{N<A}] + \mathbb{E}[(N+1)^A \mathbb{1}_{N \geq A}] \\ &\leq A^A + C\mathbb{E}\left[\frac{N!}{(N-A)!} \mathbb{1}_{N \geq A}\right] \\ &= A^A + Cp^A G_A(p). \end{aligned} \tag{3.2.0.6}$$

So we claim that

$$G_A(p) \leq n^A \tag{3.2.0.7}$$

If we prove (3.2.0.7) we can conclude because by (3.2.0.6) we have

$$\mathbb{E}[(N+1)^A] \leq A^A + C(np)^A,$$

that is, (3.2.0.2). Therefore we only need to prove (3.2.0.7), but this is immediate, indeed we have

$$G_A(p) = \frac{n!}{(n-A)!} \sum_{k=A}^{n} \binom{n-A}{k-A} p^{k-A} (1-p)^{n-k} = \frac{n!}{(n-A)!} \sum_{j=0}^{n-A} \binom{n-A}{j} p^j (1-p)^{n-A-j} = \frac{n!}{(n-A)!} \leq n^A.$$

Proof of (3.2.0.2) **in the case** $A = -1, -2, \ldots$. We argue as before: for $B = |A| = -A$, we define

$$F_B(p) := \sum_{k=0}^{n} \binom{n}{k} p^{k+B} (1-p)^{n-k} \frac{k!}{(k+B)!},$$

and we estimate $\mathbb{E}[(N+1)^{-B}]$ using F_B, as follows

$$\mathbb{E}[(N+1)^{-B}] = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \frac{1}{(k+1)^B} \leq C \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \frac{k!}{(k+B)!} = Cp^{-B} F_B(p). \tag{3.2.0.8}$$

We also have

$$\begin{aligned} F_B(p) &= \frac{n!}{(n+B)!} \sum_{k=0}^{n} \binom{n+B}{k+B} p^{k+B} (1-p)^{n-k} = \frac{n!}{(n+B)!} \sum_{j=B}^{n+B} \binom{n+B}{j} p^j (1-p)^{n+B-j} \\ &\leq \frac{n!}{(n+B)!} \sum_{j=0}^{n+B} \binom{n+B}{j} p^j (1-p)^{n+B-j} = \frac{n!}{(n+B)!} \leq \frac{C}{n^B}, \end{aligned} \tag{3.2.0.9}$$

and combining (3.2.0.8) and (3.2.0.9) we get

$$\mathbb{E}[(N+1)^{-B}] \leq C(np)^{-B}.$$

Hence the proof of (3.2.0.2) is concluded also for negative integers $A = -B$.

Proof of (3.2.0.3)**.** The proof of this second part of the Theorem is much simpler because the function inside the expectation is concave for large values of the argument.

If we define

$$f_B(x) := (\log(1+x))^B, \quad x \geq 0,$$

we have to prove that for any binomial random variable as in the hypothesis, it holds

$$\mathbb{E}[f_B(N)] \leq C f_B(\mathbb{E}[N]).$$

First, we observe that the function f_B is concave in the region

$$[\max\{x_B, 0\}, +\infty), \qquad \text{with } x_B := e^{B-1} - 1,$$

and convex in the region $[0, \max\{x_B, 0\})$. Hence, since the statement is valid with $C = 1$ if $B \in [0,1]$ by concavity of f_B, we can assume $B > 1$ (and therefore $x_B > 0$). In such case, we have

$$\mathbb{E}[f_B(N)] = \mathbb{E}[f_B(N)\mathbb{1}_{N<x_B}] + \mathbb{E}[f_B(N)\mathbb{1}_{N \geq x_B}], \tag{3.2.0.10}$$

and since we can roughly estimate

$$\mathbb{E}[f_B(N)\mathbb{1}_{N<x_B}] \leq \sup_{x \in [0,x_B)} f_B(x) = f_B(x_B) = (B-1)^B =: c_B, \quad \text{since } f_B \text{ is increasing,} \tag{3.2.0.11}$$

we focus on the second summand in (3.2.0.10). For that term, we can infer that

$$\begin{aligned}
\mathbb{E}[f_B(N)\mathbb{1}_{N \geq x_B}] &= \frac{\mathbb{E}[f_B(N)\mathbb{1}_{N \geq x_B}]}{\mathbb{P}(N \geq x_B)}\mathbb{P}(N \geq x_B) \\
&\leq f_B\left(\frac{\mathbb{E}[N\mathbb{1}_{N \geq x_B}]}{\mathbb{P}(N \geq x_B)}\right)\mathbb{P}(N \geq x_B) \quad \text{since } f_B \text{ is concave in the region } [x_B, +\infty) \\
&\leq f_B\left(\frac{\mathbb{E}[N]}{\mathbb{P}(N \geq x_B)}\right) \quad \text{since } f_B \text{ is increasing} \\
&= \left[\log\left(\frac{\mathbb{E}[N]}{1 - \mathbb{P}(N < x_B)} + 1\right)\right]^B \\
&\leq [\log(\mathbb{E}[N]+1) - \log(1 - \mathbb{P}(N < x_B))]^B \leq C f_B(\mathbb{E}[N]) + C \quad \text{by Lemma 3.2.1,}
\end{aligned} \tag{3.2.0.12}$$

where C is a positive constant depending on B and c.

Combining (3.2.0.12) with (3.2.0.10) and (3.2.0.11) we get (3.2.0.3).

Proof of (3.2.0.4)**.** We can now conclude: by Hölder inequality, we have

$$\begin{aligned}
\mathbb{E}[f(N)] = \mathbb{E}[(N+1)^A(\log(N+1))^B] &\leq \sqrt{\mathbb{E}[(N+1)^{2A}]}\sqrt{\mathbb{E}[(\log(N+1))^{2B}]} \\
&\leq C'(\mathbb{E}[N]+1)^A(\log(\mathbb{E}[N]+1))^B \\
&= C' f(\mathbb{E}[N]),
\end{aligned}$$

where C' is the square root of the product of the constants provided by (3.2.0.2) and (3.2.0.3) for $A' = 2A$ and $B' = 2B$.

As for the proof of (3.2.0.5), we can use the following argument: since the function f we are considering is strictly positive in $[c, +\infty)$ and continuous (hence bounded in the interval $[c, M]$), we are reduced to prove that for any binomial random variables of parameters n and p such that $np \geq c$,

$$\mathbb{E}[f(N)] \geq C', \tag{3.2.0.13}$$

for a positive constant C' depending on c.

So let us briefly prove (3.2.0.13): if $A \geq 0$ we have

$$\mathbb{E}[f(N)] \geq \mathbb{E}\left[(N+1)^A(\log(N+1))^B\mathbb{1}_{N \geq 1}\right] \geq 2^A(\log 2)^B\mathbb{P}(N \geq 1) \geq C',$$

where the constant C' comes from Lemma 3.2.1.

As for the case $A < 0$, we can argue as in the proof of (3.2.0.2): if $A = -1, -2, \dots$, we have

$$\begin{aligned}
\mathbb{E}[f(N)] &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} (k+1)^A (\log(k+1))^B \geq \frac{1}{C} \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} (k+1)^A \\
&\geq \frac{1}{C} \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} \frac{k!}{(k-A)!} = \frac{1}{C} \frac{n!}{(n-A)!} \sum_{k=1}^{n} \binom{n-A}{k-A} p^k (1-p)^{n-k} \\
&= \frac{1}{C} \frac{n! p^A}{(n-A)!} \sum_{j=1-A}^{n-A} \binom{n-A}{j} p^j (1-p)^{n-A-j} \\
&\geq \frac{1}{C^2} (np)^A \left(1 - \mathbb{P}(N' \leq -A)\right),
\end{aligned}$$

where N' is a binomial random variable with parameters $n - A$ and p (satisfying therefore $(n-A)p \geq np \geq c$). Hence, if $C'' \in (0,1)$ is the constant provided by Lemma 3.2.1, we get in this case

$$\mathbb{E}[f(N)] \geq \frac{1}{C} c^A (1 - C'').$$

To conclude the proof, we only need to prove (3.2.0.5) for $A < 0, A \notin \mathbb{Z}$, but such case is trivial since we have

$$\mathbb{E}[f(N)] = \mathbb{E}\left[(N+1)^A (\log(N+1))^B\right] \geq \mathbb{E}\left[(N+1)^{-\lceil -A \rceil} (\log(N+1))^B\right] \geq C',$$

since the property is already proved for integer values of A (in this case $-\lceil -A \rceil$). ☐

Using Lemma 3.2.2 we can prove the following concentration result.

Proposition 3.2.1. *Let $c > 0$ be a positive constant, and let $f : [0, +\infty) \to \mathbb{R}$ be the function*

$$f(x) := (x+1)^A (\log(x+1))^B, \quad A \in \mathbb{R} \text{ and } B \geq 0.$$

Then, there exist two infinitesimal functions $\omega : [c, +\infty) \to [0,1)$ and $\omega' : [c, +\infty) \to [0, +\infty)$ such that for every N binomial random variable of parameters n and p such that $np \geq c$, it holds

$$(1 - \omega(np)) f(np) \leq \mathbb{E}[f(N)] \leq f(np)(1 + \omega'(np)). \tag{3.2.0.14}$$

Proof. The proof is quite technical but it relies on the concentration properties of binomial random variables which we mentioned before. First, we fix $\alpha \in \left(\frac{1}{2}, 1\right), k \in (0, c^{1-\alpha})$ and we define the set

$$\mathcal{A} := \{|N - np| \leq k(np)^\alpha\} = \{np - k(np)^\alpha \leq N \leq np + k(np)^\alpha\}.$$

Notice that, thanks to the choice of k, one has

$$np - k(np)^\alpha \geq c - kc^\alpha > 0, \tag{3.2.0.15}$$

so that in $\mathcal{A}$ the variable N is uniformly bounded from below by a positive constant which does not depend on n or p. By Proposition 3.1.1, we have

$$\mathbb{P}(\mathcal{A}^c) \leq 2 \exp\left\{-\frac{k^2 (np)^{2\alpha - 1}}{3}\right\}. \tag{3.2.0.16}$$

Notice also that

$$f'(x) = A(x+1)^{A-1} (\log(x+1))^B + B(x+1)^{A-1} (\log(x+1))^{B-1} = \frac{f(x)}{x+1} \left(A + \frac{B}{\log(x+1)}\right),$$

so that, for a positive constant $C > 0$ depending on c, k and α, thanks to (3.2.0.15) one has

$$\frac{1}{C} \frac{f(x)}{x} \leq |f'(x)| \leq C \frac{f(x)}{x}, \quad \forall x \in [c - kc^\alpha, +\infty). \tag{3.2.0.17}$$

This estimate of course is valid even if A or B are zero (if they are both zero the Proposition is trivial).

From now on, we assume that

$$np \geq M, \tag{3.2.0.18}$$

with M to be fixed later, , otherwise for $np \in [c, M]$ the statement is obvious since by (3.2.0.4) and (3.2.0.5) of Lemma 3.2.2

$$\frac{1}{C} f(\mathbb{E}[N]) \le \mathbb{E}[f(N)] \le C f(\mathbb{E}[N]).$$

Now we can estimate

$$\frac{\mathbb{E}[|f(N) - f(np)|]}{f(np)} = \frac{\mathbb{E}[|f(N) - f(np)| \mathbb{1}_{\mathcal{A}}]}{f(np)} \tag{3.2.0.19}$$

$$+ \frac{\mathbb{E}[|f(N) - f(np)| \mathbb{1}_{\mathcal{A}^c}]}{f(np)}, \tag{3.2.0.20}$$

where $\mathcal{A}^c$ is the complement of $\mathcal{A}$. We estimate separately the summands in the right-hand side of the previous inequality, that is, (3.2.0.19) and (3.2.0.20). First we focus on (3.2.0.19): we have

$$\begin{aligned}
(3.2.0.19) &= \frac{1}{f(np)} \mathbb{E}[|f'(\xi)(N - np)| \mathbb{1}_{\mathcal{A}}] \quad \text{ with } \xi \text{ depending on } N, np \text{ and } f \\
&\le \max_{x \in [np - k(np)^\alpha, np + k(np)^\alpha]} |f'(x)| \frac{1}{f(np)} \mathbb{E}[|N - np| \mathbb{1}_{\mathcal{A}}] \\
&\le C \frac{1}{f(np)} \max_{x \in [np - k(np)^\alpha, np + k(np)^\alpha]} \frac{f(x)}{x} \mathbb{E}\left[|N - np| \mathbb{1}_{\mathcal{A}}\right] \quad \text{ thanks to (3.2.0.17)} \\
&\le C' \frac{1}{np} k(np)^\alpha \quad \text{ with } C' \text{ depending on } C, k, c \text{ and } \alpha \text{ and by definition of } \mathcal{A} \\
&= \frac{C''}{(np)^{1-\alpha}} =: \omega_1(np),
\end{aligned}$$

and so the estimate of (3.2.0.19) is completed. Notice that we can assume that the constant M in (3.2.0.18) satisfies

$$\frac{C''}{M^{1-\alpha}} \le \frac{1}{4},$$

so that $0 \le \omega_1 \le \frac{1}{4}$.

Now we focus on (3.2.0.20): by triangle inequality and Hölder inequality we have

$$(3.2.0.20) \le \frac{\mathbb{E}[f(N) \mathbb{1}_{\mathcal{A}^c}]}{f(np)} + \mathbb{P}(\mathcal{A}^c) \le \frac{\sqrt{\mathbb{E}[(f(N))^2]}}{f(np)} \sqrt{\mathbb{P}(\mathcal{A}^c)} + \mathbb{P}(\mathcal{A}^c). \tag{3.2.0.21}$$

Moreover, by (3.2.0.4), we have

$$\mathbb{E}[(f(N))^2] \le C (f(np))^2, \tag{3.2.0.22}$$

where $C > 0$ is the constant provided by (3.2.0.4) of Lemma 3.2.2 and depending on $A' = 2A, B' = 2B$ and c.

Thus, combining (3.2.0.21) and (3.2.0.22), we have

$$(3.2.0.20) \le C \left(\sqrt{\mathbb{P}(\mathcal{A}^c)} + \mathbb{P}(\mathcal{A}^c) \right) \le 2C \exp\left\{ -\frac{k^2 (np)^{2\alpha - 1}}{6} \right\} =: \omega_2(np) \quad \text{ by (3.2.0.16).}$$

As before, we can assume that the constant M in (3.2.0.18) is large enough to satisfy

$$2C \exp\left\{ -\frac{k^2 M^{2\alpha - 1}}{6} \right\} \le \frac{1}{4},$$

so that $\omega_2 \le \frac{1}{4}$.

We can finally conclude since we just proved that

$$\frac{\mathbb{E}[|f(N) - f(np)|]}{f(np)} = (3.2.0.19) + (3.2.0.20) = \omega_1(np) + \omega_2(np) =: \omega(np),$$

and ω is the function in the statement satisfying $\omega \le \frac{1}{2}$ for $np \ge M$. $\square$

Now we can use Lemma 3.2.2 and Proposition 3.2.1 to prove the following Theorem.

Theorem 3.2.1. *Let Let $c > 0$ be a positive constant, and let $f : [0, +\infty) \to \mathbb{R}$ be the function*

$$f(x) := (x + 1)^A (\log(x + 1))^B, \quad A \in \mathbb{R} \text{ and } B \ge 0.$$

Let $\omega : [0, +\infty) \to [0, +\infty)$ be a bounded and infinitesimal function. Then, there exists another infinitesimal function $\omega' : [c, +\infty) \to [0, +\infty)$ depending on A, B, c and ω such that, for any N binomial random variable of parameters n and p with $np \geq c$, it holds

$$\mathbb{E}[f(N)(1 + \omega(N))] \leq f(\mathbb{E}[N])(1 + \omega'(\mathbb{E}[N])), \tag{3.2.0.23}$$

and moreover, if $\sup \omega < 1$,

$$\mathbb{E}[f(N)(1 - \omega(N))] \geq f(\mathbb{E}[N])(1 - \omega'(\mathbb{E}[N])), \tag{3.2.0.24}$$

with ω' as before and $\sup \omega' < 1$.

Proof. To prove the statement we have to prove that the quantities

$$\alpha_\pm := \frac{\mathbb{E}[f(N)(1 \pm \omega(N))]}{f(\mathbb{E}[N])},$$

with $\omega < 1$ in the case of α_-, are bounded from above and below by positive constants and that $\alpha_\pm \xrightarrow[np\to\infty]{} 1$.

To prove that $\alpha_\pm \xrightarrow[np\to\infty]{} 1$, we first assume that $np \geq M$, for M large enough to be fixed later (since we have to consider the case $np \to \infty$). Then, we observe that

$$\begin{aligned} \mathbb{E}[f(N)(1 + \omega(N))] = \mathbb{E}[f(N)] & \qquad (3.2.0.25)\\ \pm \mathbb{E}[f(N)\omega(N)]. & \qquad (3.2.0.26) \end{aligned}$$

As for the first summand in the right-hand side of (3.2.0.25), we have

$$|\mathbb{E}[f(N)] - f(\mathbb{E}[N])| \leq f(np)\omega_1(np) = f(\mathbb{E}[N])\omega_1(\mathbb{E}[N]), \tag{3.2.0.27}$$

with $\omega_1 < 1$ the infinitesimal function provided by Proposition 3.2.1.

Moreover, the second summand in (3.2.0.26) can be estimated as follows: by Hölder inequality we have

$$\begin{aligned} |\mathbb{E}[f(N)\omega(N)]| &\leq \mathbb{E}[|f(N)\omega(N)|] \leq \sqrt{\mathbb{E}[(f(N))^2]}\sqrt{\mathbb{E}[(\omega(N))^2]} \\ &\leq \sqrt{C(f(\mathbb{E}[N]))^2}\sqrt{\mathbb{E}[(\omega(N))^2]} = C' f(\mathbb{E}[N])\sqrt{\mathbb{E}[\omega^2(N)]} \end{aligned} \tag{3.2.0.28}$$

where C is the constant provided by (3.2.0.4) of Lemma 3.2.2 with $A' = 2A$ and $B' = 2B$. So let us estimate this factor concerning ω^2: if

$$\mathcal{A} := \left\{ N \geq \frac{np}{2} \right\},$$

by Proposition 3.1.1, we have

$$\mathbb{P}(\mathcal{A}^c) = \mathbb{P}\left(N - np < -\frac{np}{2}\right) \leq \exp\left\{-\frac{np}{8}\right\}, \tag{3.2.0.29}$$

and therefore by (3.2.0.29) we have

$$\begin{aligned} \mathbb{E}[\omega^2(N)] &= \mathbb{E}[\omega^2(N)\mathbb{1}_{\mathcal{A}}] + \mathbb{E}[\omega^2(N)\mathbb{1}_{\mathcal{A}^c}] \\ &\leq \max_{x \geq \frac{np}{2}} \omega^2(x) + (\max_{x\geq 0} |\omega(x)|)^2 \mathbb{P}(\mathcal{A}^c) \\ &\leq \max_{x \geq \frac{np}{2}} \omega^2(x) + (\max_{x\geq 0} |\omega(x)|)^2 \exp\left\{-\frac{np}{8}\right\} =: \omega_2(np). \end{aligned}$$

Combining this estimate with (3.2.0.28) we get

$$|\mathbb{E}[f(N)\omega(N)]| \leq C' f(\mathbb{E}[N])\sqrt{\omega_2(np)},$$

which, if combined with (3.2.0.25), (3.2.0.26) and (3.2.0.27), gives

$$|\alpha_\pm - 1| \leq \omega_1(np) + C'\sqrt{\omega_2}(np).$$

Notice that in this case we can assume that the M such that $np \geq M$ is large enough to have $\omega_1 + C'\sqrt{\omega_2} < 1$.

The previous argument also proves the boundedness from above of $\alpha_\pm$.

As for the boundedness from below, if $\mathbb{E}[N] = np \in [c, M]$, one has

$$\alpha_+ \geq \alpha_- \geq \frac{\mathbb{E}[f(N)] \inf(1 - \omega)}{f(\mathbb{E}[N])} \geq \frac{\inf(1 - \omega)}{C},$$

with C the constant provided by (3.2.0.5) of Lemma 3.2.2. □

3.3 The convergence to a Poisson random variable.

In this Section, we first recall the Definition of Poisson random variable.

Definition 3.3.1 ([30], Section 2.2.4, Poisson random variable). *Let $(\Omega, \Sigma, \mathbb{P})$ a probability space. A random variable $M : \Omega \to \mathbb{R}$ is a Poisson random variable of parameter $c > 0$ if*

$$\mathbb{P}(M = k) = \frac{e^{-c}c^k}{k!} \quad \forall k \in \mathbb{N}.$$

We recall here that, if N is a Binomial random variable of parameters n and $\frac{c}{n}$ and M a Poisson random variable of parameter $c > 0$, for fixed $k \in \mathbb{N}$ it holds (see for example [30], Section 2.2.4):

$$\mathbb{P}(N = k) = \binom{n}{k}\left(\frac{c}{n}\right)^k\left(1 - \frac{c}{n}\right)^{n-k} = \underbrace{\frac{n!}{(n-k)!n^k}}_{\xrightarrow[n\to\infty]{}1}\frac{c^k}{k!}\underbrace{\left(1 - \frac{c}{n}\right)^{n-k}}_{\xrightarrow[n\to\infty]{}e^{-c}} \xrightarrow[n\to\infty]{} \frac{e^{-c}c^k}{k!} = \mathbb{P}(M = k).$$

Hence we now focus on a result concerning the convergence of a sequence of binomial random variables of parameters n and $\frac{c}{n}$ to a Poisson random variable of parameter c.

Proposition 3.3.1 ([30], Section 2.2.4). *Fix $c > 0$ and let: $\{N_{n,c}\}_{n=1}^{\infty}$ be a sequence of binomial random variables of parameters n and $\frac{c}{n}$, M a Poisson random variable of parameter c and $f : \mathbb{N} \to \mathbb{R}$ a real function. Then*

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|f(N_{n,c})|] < +\infty \iff \mathbb{E}[|f(M)|] < +\infty, \tag{3.3.0.1}$$

and, in this case, it holds

$$\mathbb{E}[f(N_{n,c})] \xrightarrow[n\to\infty]{} \mathbb{E}[f(M)]. \tag{3.3.0.2}$$

Proof. First notice that, by splitting $f = f_+ - f_-$, where $f_\pm$ are respectively the positive and negative part of f, we can assume that $f \geq 0$. In such case $|f| = f$.

Therefore we assume $f : \mathbb{N} \to [0, +\infty)$.

We start with the proof of (3.3.0.1), and, to do that, we prove the following intermediate estimate:

$$\mathbb{E}[f(M)] \leq \liminf_{n\to\infty} \mathbb{E}[f(N_{n,c})] \leq \limsup_{n\to\infty} \mathbb{E}[f(N_{n,c})] \leq e^c\mathbb{E}[f(M)]. \tag{3.3.0.3}$$

Since one has:

$$\mathbb{E}[f(N_{n,c})] = \sum_{k=0}^{n}\binom{n}{k}\left(\frac{c}{n}\right)^k\left(1 - \frac{c}{n}\right)^{n-k} f(k) = \sum_{k=0}^{n}\frac{n!}{(n-k)!n^k}\frac{c^k}{k!}\left(1 - \frac{c}{n}\right)^{n-k} f(k), \tag{3.3.0.4}$$

by summing only over $k \leq K$ for fixed $K \in \mathbb{N}$ and then taking the liminf, one gets

$$\begin{aligned}
\liminf_{n\to\infty} \mathbb{E}[f(N_{n,c})] &\geq \liminf_{n\to\infty} \sum_{k=0}^{K}\frac{n!}{(n-k)!n^k}\frac{c^k}{k!}\left(1 - \frac{c}{n}\right)^{n-k} f(k) \\
&\geq \sum_{k=0}^{K}\liminf_{n\to\infty}\frac{n!}{(n-k)!n^k}\frac{c^k}{k!}\left(1 - \frac{c}{n}\right)^{n-k} f(k) \\
&= \sum_{k=0}^{K}\frac{c^k}{k!}e^{-c}f(k),
\end{aligned}$$

and then taking the limit as $K \to \infty$, one gets

$$\liminf_{n\to\infty} \mathbb{E}[f(N_{n,c})] \geq \mathbb{E}[f(M)],$$

that is, the first inequality in (3.3.0.3). To prove the second one, using again (3.3.0.4), since

$$\frac{n!}{(n-k)!n^k} \leq 1, \quad \text{and} \quad \left(1 - \frac{c}{n}\right)^{n-k} \leq 1, \tag{3.3.0.5}$$

one gets

$$\limsup_{n\to\infty}\mathbb{E}[f(N_{n,c})]\le\limsup_{n\to\infty}\sum_{k=0}^{n}\frac{c^k}{k!}f(k)=e^c\mathbb{E}[f(M)],$$

hence also the second inequality in (3.3.0.3) is proved. Therefore also (3.3.0.1) holds as a consequence of (3.3.0.3). So we now focus on (3.3.0.2): if we prove that

$$\limsup_{n\to\infty}\mathbb{E}[f(N_{n,c})]\le\mathbb{E}[f(M)], \tag{3.3.0.6}$$

combining (3.3.0.6) with (3.3.0.3) and the assumption that $\mathbb{E}[f(M)]$ is finite, also (3.3.0.2) is proved. To prove (3.3.0.6), we argue as follows: let $k(n)\in\mathbb{N}$ be a natural number such that

$$k(n)\xrightarrow[n\to\infty]{}\infty,\qquad \frac{k(n)}{n}\xrightarrow[n\to\infty]{}0.$$

Then, using again (3.3.0.4) and (3.3.0.5), one gets

$$\begin{aligned}\mathbb{E}[f(N_{n,c})]&\le\sum_{k=0}^{n}\frac{c^k}{k!}\left(1-\frac{c}{n}\right)^{n-k}f(k)\\&\le\left(1-\frac{c}{n}\right)^{n(1-\frac{k(n)}{n})}\sum_{k=0}^{k(n)}\frac{c^k}{k!}f(k) &(3.3.0.7)\\&+\sum_{k=k(n)+1}^{\infty}\frac{c^k}{k!}f(k), &(3.3.0.8)\end{aligned}$$

and since thanks to the choice of $k(n)$

$$(3.3.0.7)\xrightarrow[n\to\infty]{}\mathbb{E}[f(M)],$$

and

$$(3.3.0.8)=e^c\mathbb{E}[f(M)\mathbb{1}_{M>k(n)}]\xrightarrow[n\to\infty]{}0,$$

one gets (3.3.0.6). □

Proposition 3.3.1 can be used to prove the following result.

Corollary 3.3.1. *Let $c>0$ be a positive constant, $N_{n,c}$ a sequence of binomial random variables of parameters n and $\frac{c}{n}$, M a Poisson random variable of parameter c and $X=\{X_i\}_{i=1}^{\infty}$ a sequence of iid random variables uniformly distributed in $(0,1)^d$. Then, if $f_{X,p}^{(j)}:\mathbb{N}\to[0,+\infty)$, $j=1,2,3$ are the following functions*

$$f_{X,p}^{(1)}(k):=W_p^p\left(\sum_{i=1}^{k}\delta_{X_i},k\mathbb{1}_{(0,1)^d}\right),f_{X,p}^{(2)}(k):=Wb_p^p\left(\sum_{i=1}^{k}\delta_{X_i},k\mathbb{1}_{(0,1)^d}\right),f_{X,p}^{(3)}(k):=Wb_p^p\left(\sum_{i=1}^{k}\delta_{X_i},c\mathbb{1}_{(0,1)^d}\right),$$

it holds

$$\mathbb{E}\left[f_{X,p}^{(j)}(N_{n,c})\right]\xrightarrow[n\to\infty]{}\mathbb{E}\left[f_X^{(j)}(M)\right].$$

Proof. To prove the statement, by Proposition 3.3.1, it is sufficient to prove that

$$\sup_{n\in\mathbb{N}}\mathbb{E}[f_{X,p}^{(j)}(N_{n,c})]<+\infty, \tag{3.3.0.9}$$

for any $j=1,2,3$.

We start with $j=1$: for any $n\in\mathbb{N}$ one has

$$\begin{aligned}\mathbb{E}\left[f_{X,p}^{(1)}(N_{n,c})\right]&=\mathbb{E}\left[\mathbb{E}\left[f_{X,pq}^{(1)}(N_{n,c})|N_{n,c}\right]\mathbb{1}_{N_{n,c}\ge1}\right]\\&\le C\mathbb{E}\left[N_{n,c}^{1-\frac{p}{d}}(\chi_{d\ge3}+(\log(N_{n,c}+1))^{\frac{p}{2}}\chi_{d=2})\mathbb{1}_{N_{n,c}\ge1}\right]\qquad\text{by (1.1.1.7)}\\&\le C'\mathbb{E}\left[(N_{n,c}+1)^{1-\frac{p}{d}}(\chi_{d\ge3}+(\log(N_{n,c}+1))^{\frac{p}{2}}\chi_{d=2})\right]\\&\le C''(\mathbb{E}[N_{n,c}]+1)^{1-\frac{p}{d}}(\log(\mathbb{E}[N_{n,c}]+1))^{\frac{p}{2}}\qquad\text{by (3.2.0.4) of Lemma 3.2.2}\\&=C''(c+1)^{1-\frac{p}{d}}(\log(c+1))^{\frac{p}{2}}=:C'''.\end{aligned}$$

Hence we only have to prove that the same holds for $f^{(2)}$ and $f^{(3)}$.

By (2.1.1.2), for any $k \in \mathbb{N}$ we have

$$f^{(2)}_{X,p}(k) \le f^{(1)}_{X,p}(k),$$

hence also

$$\mathbb{E}\left[f^{(2)}_{X,p}(N_{n,c})\right] \le \mathbb{E}\left[f^{(1)}_{X,p}(N_{n,c})\right] \le C''',$$

for any $n \in \mathbb{N}$.

As for $f^{(3)}_{X,p}$, by triangle inequality we have

$$\begin{aligned} f^{(3)}_{X,p}(k) = Wb_p^p\left(\sum_{i=1}^k \delta_{X_i}, c\mathbb{1}_{(0,1)^d}\right) &\le 2^p Wb_p^p\left(\sum_{i=1}^k \delta_{X_i}, k\mathbb{1}_{(0,1)^d}\right) + 2^p Wb_p^p\left(k\mathbb{1}_{(0,1)^d}, c\mathbb{1}_{(0,1)^d}\right) \\ &\le 2^p f^{(2)}_{X,p}(k) + 2^p \frac{|k-c|^p}{\max\{k,c\}^{p-1}} \qquad \text{by Lemma 2.2.7} \\ &\le 2^p f^{(2)}_{X,p}(k) + 2^p c^{1-p}|k-c|^p, \end{aligned}$$

hence, for any $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}\left[f^{(3)}_{X,p}(N_{n,c})\right] &\le 2^p\mathbb{E}\left[f^{(2)}_{X,p}(N_{n,c})\right] + 2^p c^{1-p}\mathbb{E}\left[|N_{n,c}-c|^p\right] \\ &\le 2^p C''' + 2^p c^{1-p} C'''' c^{\frac{p}{2}}, \qquad \text{by Lemma 3.1.1.} \end{aligned}$$

Hence, we can apply Proposition 3.3.1 to infer that

$$\mathbb{E}\left[f^{(j)}_{X,p}(N_{n,c})\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[f^{(j)}_{X,p}(M)\right],$$

for any $j = 1, 2, 3$. $\square$

Chapter 4

The geodesics on the sphere and the exponential map.

This Chapter is dedicated to the study of the exponential map, which we need in the geometric decomposition for the upper bound, that is, in Subsection 6.2.1. To define it, we begin with the definition of geodesics on the sphere.

The most of the notions and properties we need are taken from [28].

4.1 The geodesics.

To introduce the concept of geodesic, let us begin with the notion of distance between two points on ∂B_1.

Definition 4.1.1 ([28], page 58)**.** *Given $P, Q \in \partial B_1$, the geodesic distance between P and Q is*

$$d(P,Q) := \min \left\{ \int_0^1 dt |\dot{\gamma}(t)| : \begin{array}{l} \gamma : [0,1] \xrightarrow{C^1} \partial B_1 \\ \gamma(0) = P, \gamma(1) = Q \end{array} \right\}.$$

The minimizing curves γ on the spheres are the arcs of circles, that is, the following property holds.

Lemma 4.1.1 ([28], Proposition 5.13)**.** *For $P \neq Q \in \partial B_1$ such that $Q \neq -P$, the minimizing curve in Definition 4.1.1 is the following:*

$$\gamma(t) = P \cos(\arccos(P \cdot Q)t) + \frac{Q - (P \cdot Q)P}{\sqrt{1 - (P \cdot Q)^2}} \sin(\arccos(P \cdot Q)t), \quad t \in [0,1].$$

Otherwise, if $Q = -P$, for any v such that $v \cdot P = 0$ the curve

$$\gamma(t) = P \cos(\pi t) + v \sin(\pi t), \quad t \in [0,1],$$

is a minimizer.

In particular, one has

$$d(P,Q) = \arccos(P \cdot Q).$$

With these notions, a geodesic on ∂B_1 is defined as follows.

Definition 4.1.2 ([28], page 58)**.** *A curve $\gamma : [0,1] \to \partial B_1$ is a geodesic if there exists a constant $v > 0$ such that for any $t \in (0,1)$ there exists $\varepsilon > 0$ such that for any $t - \varepsilon < t_1 < t_2 < t + \varepsilon$ it holds*

$$d(\gamma(t_1), \gamma(t_2)) = v(t_2 - t_1).$$

By the previous Lemma 4.1.1 it is easy to derive the following property.

Proposition 4.1.1 ([28], Proposition 5.13)**.** *If $\gamma : [0,1] \to \partial B_1$ is a geodesic curve, there exist $P, Q \in \partial B_1$ such that*

$$P \cdot Q = 0,$$

and $\omega > 0$ such that

$$\gamma(t) := P \cos(\omega t) + Q \sin(\omega t).$$

4.1.1 A particular class of rotations on the sphere.

What we need in the following is a rotation $R_P^Q : \mathbb{R}^d \to \mathbb{R}^d$ which moves ∂B_1 along the geodesic from P to Q. To build such a map, we proceed as follows: we let the vectorial space $\{P, Q\}^\perp$ be fixed, and in the plane generated by the vectors P and Q we apply the rotation by angle $\theta = \arccos(P \cdot Q)$. In other words, if we fix the orthonormal basis

$$v_1 := P, \quad v_2 := \frac{Q - (P \cdot Q)P}{\sqrt{1 - (P \cdot Q)^2}}, \quad v_3, \dots, v_d \text{ such that } v_i \cdot v_j = \delta_{i,j},$$

we get

$$P = v_1, \quad Q = (P \cdot Q)v_1 + \sqrt{1 - (P \cdot Q)^2}v_2,$$

so that the coordinates of P and Q with respect to the basis $\{v_i\}_{i=1}^d$ are respectively

$$(1, 0, 0, \dots, 0) \quad \text{and} \quad (P \cdot Q, \sqrt{1 - (P \cdot Q)^2}, 0, \dots, 0),$$

and the matrix representing the rotation R_P^Q we mentioned above with respect to the orthonormal basis $v_1, \dots, v_d$ is

$$M_P^Q = \begin{pmatrix} P \cdot Q & -\sqrt{1 - (P \cdot Q)^2} & 0 & \cdots & 0 \\ \sqrt{1 - (P \cdot Q)^2} & P \cdot Q & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

and this is the map we were looking for.

The rotation R_P^Q has the property move each point in ∂B_1 to another point whose distance from the first one is not larger than the distance between P and Q, as the following Proposition states.

Proposition 4.1.2. *Let $P, Q, S \in \partial B_1$. Then, if R_P^Q is the rotation defined before, it holds*

$$d_{\partial B_1}(S, R_P^Q(S)) \le d_{\partial B_1}(P, Q).$$

Proof. By Lemma 4.1.1, if we denote by $\theta = \arccos(P \cdot Q)$, we have to prove that

$$\arccos(S \cdot R_P^Q(S)) \le \theta.$$

If we now use the basis $v_1, \dots, v_d$ mentioned above and denote by

$$s_k := S \cdot v_k, \quad k = 1, \dots, d,$$

the coordinates of S with respect to that basis, we can write

$$\begin{aligned} S \cdot R_P^Q(S) &= (s_1, s_2, s_3 \dots, s_d) \cdot (s_1 \cos\theta - s_2 \sin\theta, s_1 \sin\theta + s_2 \cos\theta, s_3, \dots, s_d) \\ &= (s_1^2 + s_2^2) \cos\theta + \sum_{k=3}^d s_k^2 \\ &= 1 - (s_1^2 + s_2^2)(1 - \cos\theta), \end{aligned}$$

where in the last inequality we have used that

$$|S|^2 = \sum_{k=1}^d s_k^2 = 1.$$

Therefore we have

$$\arccos(S \cdot R_P^Q(S)) = \arccos(1 - (s_1^2 + s_2^2)(1 - \cos\theta)) \le \max_{\alpha \in [0,1]} \arccos(1 - \alpha(1 - \cos\theta)) = \theta,$$

since the above function is increasing with respect to α and therefore the maximum is achieved for $\alpha = 1$.

Hence the thesis is proved. □

Before defining the exponential map, here we state another result concerning the particular class of rotations R_P^Q we just mentioned.

Lemma 4.1.2. *Let $P, Q \in \partial B_1$, R_P^Q the rotation defined before, and $V \in O(d)$ be an orthonormal operator. Then, the operator*

$$L := (R_P^{VQ})^T V R_P^Q,$$

is an orthonormal operator and satisfies

$$Lv \cdot P = v \cdot P \quad \forall v \in \mathbb{R}^d, \tag{4.1.1.1}$$

in particular

$$LP = P \quad \text{and} \quad L\{P\}^\perp = \{P\}^\perp. \tag{4.1.1.2}$$

Proof. To start with, let us stress that L is orthonormal since it is a composition of orthonormal operators.

Then, we prove (4.1.1.1): for $v \in \mathbb{R}^d$ we have

$$Lv \cdot P = (R_P^{VQ})^T V R_P^Q v \cdot P = V R_P^Q v \cdot R_P^{VQ} P = V R_P^Q v \cdot VQ = R_P^Q v \cdot Q = v \cdot (R_P^Q)^T Q = v \cdot (R_P^Q)^{-1} Q = v \cdot P,$$

that is, the thesis. Finally, to prove (4.1.1.2) we can use (4.1.1.1) as follows: since we have

$$LP \cdot P = P \cdot P = 1 = |LP||P|,$$

hence by Cauchy-Schwarz inequality one has $LP = \lambda P$ for a certain $\lambda \in \mathbb{R}$, but applying (4.1.1.1) again we have

$$\lambda = \lambda P \cdot P = LP \cdot P = P \cdot P = 1,$$

and therefore $LP = P$. Moreover, as for the second property in (4.1.1.2), we can observe that for any $v \in \{P\}^\perp$, by (4.1.1.1) one has

$$Lv \cdot P = v \cdot P = 0 \Rightarrow Lv \in \{P\}^\perp \Rightarrow L\{P\}^\perp \subseteq \{P\}^\perp,$$

where the last inequality is actually an equality since L is orthonormal and therefore invertible. □

Now we define and prove some properties of the exponential map, and this is the content of the next section.

4.2 The exponential map.

The exponential map is a way to project the tangent space to a manifold on the manifold itself. Since the densities we are considering are radially symmetric, we need some properties about the exponential map on the $d-1$ dimensional sphere.

We begin with the definition of exponential map, but, before that, recall that given a point $P \in \partial B_1$, the tangent space to ∂B_1 in P is the orthogonal to P, that is,

$$T_P(\partial B_1) := \{v \in \mathbb{R}^d : v \cdot P = 0\}.$$

Definition 4.2.1 ([28], page 72, the Exponential map)**.** *Let $P \in \partial B_1$ and $v \in T_P(\partial B_1)$. The exponential map is defined as*

$$\exp_P : T_P(\partial B_1) \to \partial B_1, \qquad v \mapsto \gamma_v^P(1),$$

with γ_v^P the geodesic on ∂B_1 such that

$$\gamma_v^P(0) = P, \qquad \frac{\partial \gamma_v^P(t)}{\partial t}\Big|_{t=0} = v.$$

The geodesic on the sphere are the circles from Proposition 4.1.1, thus one can write explicitly

$$\gamma_v^P(t) := P\cos(|v|t) + \frac{v}{|v|}\sin(|v|t),$$

and therefore one has

$$\exp_P(v) = \gamma_v^P(1) = P\cos(|v|) + \frac{v}{|v|}\sin(|v|).$$

Through this map one can build a diffeomorphism (up to restricting to a suitable subset of $T_P(\partial B_1)$) between ∂B_1 and $T_P(\partial B_1)$. We want to study the differential of this map.

To do that, to avoid considering the local coordinates of a boundary of the point P, we write $D\exp_P(v)$ through its formal definition: for any $v \in T_P(\partial B_1)$, the differential of $\exp_P$ in v is the linear map

$$D\exp_P(v) : T_P(\partial B_1) \to T_P(\partial B_1), \qquad D\exp_P(v)h := \lim_{\varepsilon \to 0} \frac{\exp_P(v + \varepsilon h) - \exp_P(v)}{\varepsilon}. \tag{4.2.0.1}$$

Using the rescaling property (see [28, Lemma 5.8])

$$\gamma_{cv}^P(t) = \gamma_v^P(ct), \tag{4.2.0.2}$$

it is immediate to see that

$$D\exp_P(v)|_{v=0} = \mathrm{Id}.$$

This fact is proven within the proof of [28, Lemma 5.10], and it is quite immediate to show indeed for any $h \in T_P(\partial B_1)$ one has

$$\begin{aligned}
D\exp_P(v)|_{v=0}h &= \lim_{\varepsilon\to 0}\frac{\exp_P(v+\varepsilon h)-\exp_P(v)}{\varepsilon}|_{v=0} \quad \text{by definition}\\
&= \lim_{\varepsilon\to 0}\frac{\gamma_{v+\varepsilon h}^P(1)-\gamma_v^P(1)}{\varepsilon}|_{v=0} \quad \text{by definition of exponential map}\\
&= \lim_{\varepsilon\to 0}\frac{\gamma_{\varepsilon h}^P(1)-\gamma_0^P(1)}{\varepsilon} \quad \text{by definition}\\
&= \lim_{\varepsilon\to 0}\frac{\gamma_{h}^P(\varepsilon)-\gamma_h^P(0)}{\varepsilon} \quad \text{by rescaling property (4.2.0.2)}\\
&= h \quad \text{by definition of } \gamma_h^P.
\end{aligned}$$

But we need to know that also the second derivatives of $\exp_P$ are bounded, indeed one can do the same for the second derivatives: for any $v, h \in T_P(\partial B_1)$, $D^2\exp_P(v)h$ is the linear map

$$D_P^2(v)h : T_P(\partial B_1) \to T_P(\partial B_1), \qquad D^2\exp_P(v)hk := \lim_{\varepsilon\to 0}\frac{D\exp_P(v+\varepsilon k)h - D\exp_P(v)h}{\varepsilon} \tag{4.2.0.3}$$

Thus we shall prove the following result about the first and the second derivatives of the exponential map on the sphere. Hereafter, we denote by $\hat{v}$ the normalization of a vector v, as follows.

Definition 4.2.2. *For $v \in \mathbb{R}^d \setminus \{0\}$, $\hat{v}$ is defined as*

$$\hat{v} := \frac{v}{|v|}.$$

Proposition 4.2.1. *There exists a constant $C > 0$ such that for any $P \in \partial B_1$, $v, h, k \in T_P(\partial B_1)$, it holds*

$$|D\exp_P(v)h - h| \le C|h||v|; \tag{4.2.0.4}$$
$$|D^2\exp_P(v)hk + Ph\cdot k| \le C|v||h||k|; \tag{4.2.0.5}$$
$$|\exp_P(v) - P| \le C|v|. \tag{4.2.0.6}$$

This is a consequence of the following Lemma.

Lemma 4.2.1. *Let $P \in \partial B_1$, $v, h, k \in T_P(\partial B_1)$ and $\hat{v}$ as in Definition 4.2.2. Then the following hold:*

$$D\exp_P(v)h = \left[-P\sin(|v|) + v\left(\frac{\cos(|v|)}{|v|} - \frac{\sin(|v|)}{|v|^2}\right)\right]\hat{v}\cdot h + h\frac{\sin(|v|)}{|v|}; \tag{4.2.0.7}$$

$$\begin{aligned}
D^2\exp_P(v)hk &= \left[-P\frac{\sin(|v|)}{|v|} + \hat{v}\left(\frac{\cos(|v|)}{|v|} - \frac{\sin(|v|)}{|v|^2}\right)\right]k\cdot h\\
&\quad + (k\hat{v}\cdot h + h\hat{v}\cdot k)\left(\frac{\cos(|v|)}{|v|} - \frac{\sin(|v|)}{|v|^2}\right)\\
&\quad - \hat{v}\cdot h\hat{v}\cdot k\left[\hat{v}\sin(|v|) + (P|v| + 3\hat{v})\left(\frac{\cos(|v|)}{|v|} - \frac{\sin(|v|)}{|v|^2}\right)\right].
\end{aligned} \tag{4.2.0.8}$$

Before proving Lemma 4.2.1 we exploit it to prove Proposition 4.2.1.

Proof of Proposition 4.2.1. As for the proof of (4.2.0.4), by (4.2.0.7) we have

$$\begin{aligned}
|D\exp_P(v)h - h| &= \left|\left[-P\sin(|v|) + v\left(\frac{\cos(|v|)}{|v|} - \frac{\sin(|v|)}{|v|^2}\right)\right]\hat{v}\cdot h + h\left(\frac{\sin(|v|)}{|v|} - 1\right)\right|\\
&\le |\sin(|v|)| + |v|\left|\frac{\cos(|v|)}{|v|} - \frac{\sin(|v|)}{|v|^2}\right||h| + |h|\left|\frac{\sin(|v|)}{|v|} - 1\right|,
\end{aligned}$$

where in the inequality we have used that $|P| = |\hat{v}| = 1$ and Cauchy-Schwarz inequality.

Hence, one can prove (4.2.0.4) recalling that for any $t \in [0,1]$

$$|\sin t| \leq t, \qquad |\cos t - 1| \leq \frac{t^2}{2}, \qquad \left|\frac{\cos t}{t} - \frac{\sin t}{t^2}\right| \leq Ct, \qquad \left|\frac{\sin t}{t} - 1\right| \leq Ct^2. \tag{4.2.0.9}$$

As for (4.2.0.5), by (4.2.0.8) one has

$$\begin{aligned}
|D^2 \exp_P(v)hv + Ph \cdot k| \leq \Bigg| &\left[P\left(1 - \frac{\sin(|v|)}{|v|}\right) + \hat{v}\left(\frac{\cos(|v|)}{|v|} - \frac{\sin(|v|)}{|v|^2}\right)\right] k \cdot h \\
&+ (k\hat{v} \cdot h + h\hat{v} \cdot k)\left(\frac{\cos(|v|)}{|v|} - \frac{\sin(|v|)}{|v|^2}\right) \\
&- \hat{v} \cdot h\hat{v} \cdot k \left[\hat{v}\sin(|v|) + (P|v| + 3\hat{v})\left(\frac{\cos(|v|)}{|v|} - \frac{\sin(|v|)}{|v|^2}\right)\right]\Bigg| \\
\leq &\left|1 - \frac{\sin(|v|)}{|v|}\right| |h||k| + (6 + |v|)\left|\frac{\cos(|v|)}{|v|} - \frac{\sin(|v|)}{|v|^2}\right| |h||k| \\
&+ |h||k||\sin(|v|)|,
\end{aligned}$$

where we have used again that $|P| = |\hat{v}| = 1$ and Cauchy-Schwarz inequality. Applying again (4.2.0.9) one gets the thesis.

As for (4.2.0.6), by definition of exponential map on the sphere we have

$$\begin{aligned}
|\exp_P(v) - P| &= \left|P(\cos(|v|) - 1) + \frac{v}{|v|}\sin(|v|)\right| \\
&\leq |P||\cos(|v|) - 1| + |\hat{v}||\sin(|v|)| \\
&\leq \min\left\{2, \frac{|v|^2}{2}\right\} + |v| \qquad \text{since } |P| = |\hat{v}| = 1 \text{ and using again (4.2.0.9)} \\
&\leq C|v|.
\end{aligned}$$

$\square$

Now that we saw how to use it, we can prove Lemma 4.2.1.

Proof of Lemma 4.2.1. First we prove (4.2.0.7). By definition we have

$$\begin{aligned}
D\exp_P(v)h &= \lim_{\varepsilon \to 0} \frac{\gamma^P_{v+\varepsilon h}(1) - \gamma^P_v(1)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{P\cos(|v + \varepsilon h|) + \frac{v+\varepsilon h}{|v+\varepsilon h|}\sin(|v + \varepsilon h|) - P\cos(|v|) + \hat{v}\sin(|v|)}{\varepsilon} \\
&= P\lim_{\varepsilon \to 0} \frac{\cos(|v + \varepsilon h|) - \cos(|v|)}{\varepsilon} + h\lim_{\varepsilon \to 0} \frac{\sin(|v + \varepsilon h|)}{|v + \varepsilon h|} + v\frac{\frac{\sin(|v+\varepsilon h|)}{|v+\varepsilon h|} - \frac{\sin(|v|)}{|v|}}{\varepsilon} \\
&= -P\frac{\sin(|v|)}{|v|}v \cdot h + h\frac{\sin(|v|)}{|v|} + v\left(\frac{\cos(|v|)}{|v|} - \frac{\sin(|v|)}{|v|^2}\right)\hat{v} \cdot h,
\end{aligned}$$

which is what we needed.

As for (4.2.0.8), by the previous equation we have

$$
\begin{aligned}
D^2\exp_P(v)hk &= \lim_{\varepsilon\to 0}\frac{D\exp_P(v+\varepsilon k)h - D\exp_P(v)h}{\varepsilon}\\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Bigg[-P\frac{\sin(|v+\varepsilon k|)}{|v+\varepsilon k|}(v+\varepsilon k)\cdot h + h\frac{\sin(|v+\varepsilon k|)}{|v+\varepsilon k|}\\
&+ (v+\varepsilon k)\left(\frac{\cos(|v+\varepsilon k|}{|v+\varepsilon k|^2} - \frac{\sin(|v+\varepsilon k|)}{|v+\varepsilon k|^3}\right)(v+\varepsilon k)\cdot h\\
&+ P\frac{\sin(|v|)}{|v|}v\cdot h - h\frac{\sin(|v|)}{|v|} - v\left(\frac{\cos(|v|)}{|v|^2} - \frac{\sin(|v|)}{|v|^3}\right)v\cdot h\Bigg]\\
&= \lim_{\varepsilon\to 0}\frac{-P\left(\frac{\sin(|v+\varepsilon k|)}{|v+\varepsilon k|} - \frac{\sin(|v|)}{|v|}\right)}{\varepsilon}(v+\varepsilon k)\cdot h + h\lim_{\varepsilon\to 0}\frac{\frac{\sin(|v+\varepsilon k|)}{|v+\varepsilon k|} - \frac{\sin(|v|)}{|v|}}{\varepsilon}\\
&+ (v+\varepsilon k)(v+\varepsilon k)\cdot h\lim_{\varepsilon\to 0}\frac{\frac{\cos(|v+\varepsilon k|)}{|v+\varepsilon k|^2} - \frac{\sin(|v+\varepsilon k|)}{|v+\varepsilon k|^3} - \frac{\cos(|v|)}{|v|^2} + \frac{\sin(|v|)}{|v|^3}}{\varepsilon}\\
&- P\frac{\sin(|v|)}{|v|}k\cdot h + \left(\frac{\cos(|v|)}{|v|^2} - \frac{\sin(|v|)}{|v|^3}\right)\lim_{\varepsilon\to 0}\frac{(v+\varepsilon k)(v+\varepsilon k)\cdot h - vv\cdot h}{\varepsilon}\\
&= -P\left(\frac{\cos(|v|)}{|v|} - \frac{\sin(|v|)}{|v|^2}\right)\hat{v}\cdot kv\cdot h + h\left(\frac{\cos(|v|)}{|v|} - \frac{\sin(|v|)}{|v|^2}\right)\hat{v}\cdot k\\
&+ vv\cdot h\left(-\frac{\sin(|v|)}{|v|^2} - 3\frac{\cos(|v|)}{|v|^3} + 3\frac{\sin(|v|)}{|v|^4}\right)\hat{v}\cdot k\\
&- P\frac{\sin(|v|)}{|v|}k\cdot h + \left(\frac{\cos(|v|)}{|v|^2} - \frac{\sin(|v|)}{|v|^3}\right)(kv\cdot h + vk\cdot h)\,,
\end{aligned}
$$

that is, (4.2.0.8). □

4.2.1 Some properties of the exponential map.

In this Section we state some results about the composition of the exponential map with an orthogonal operator R. In particular, in the proof of Lemma 6.2.4, we need the following two properties.

Lemma 4.2.2. *Let $R \in O(d)$ be an orthonormal operator. Then, if $P \in \partial B_1$ and $v \in T_P(\partial B_1)$, the following conditions hold:*

$$T_{RP}(\partial B_1) = RT_P(\partial B_1), \tag{4.2.1.1}$$

$$R\exp_P(v) = \exp_{RP}(Rv). \tag{4.2.1.2}$$

Proof. We begin with the proof of (4.2.1.1), which follows by definition

$$T_{RP}(\partial B_1) = \{x \in \mathbb{R}^d : RP\cdot x = 0\} = \{x \in \mathbb{R}^d : P\cdot R^{-1}x = 0\} = R\{y : P\cdot y = 0\} = T_P(\partial B_1),$$

and therefore the statement is proved.

As for (4.2.1.2) we have

$$R\exp_P(v) = R\gamma_P^v(1) = \gamma_{RP}^{Rv}(1) = \exp_{RP}(Rv),$$

and therefore the thesis is proved and there is no need to involve the explicit form of the exponential map. □

Chapter 5

Some other notions of Measure Theory.

5.1 The Hausdorff Measure.

In this Section we briefly recall the definition of Hausdorff measure of dimension $s > 0$ in $\mathbb{R}^d$.

Definition 5.1.1 ([19], Definition 2.1: Hausdorff measure). *Let $A \subseteq \mathbb{R}^d$ and $s, \delta \in (0, +\infty)$. $\mathcal{H}^s_\delta(A)$ is defined as*

$$\mathcal{H}^s_\delta(A) := \inf \left\{ \sum_{j=1}^{\infty} \frac{\pi^{\frac{s}{2}}}{\Gamma(\frac{s}{2}+1)} \left(\frac{diam\ C_j}{2} \right)^s : A \subseteq \cup_{j=1}^{\infty} C_j, diam\ C_j \leq \delta \right\},$$

and the s-dimensional Hausdorff measure of A is

$$\mathcal{H}^s(A) := \sup_{\delta > 0} \mathcal{H}^s_\delta(A).$$

Briefly speaking, the Hausdorff measure is a way to perform the measure of a set embedded in $\mathbb{R}^d$, such as, for instance, a k-dimensional surface with $k < d$.

For further details on the Hausdorff measure see [19], but to our purposes it is sufficient to know that the Hausdorff measure $\mathcal{H}^k$ restricted to a k-dimensional surface sufficiently regular, say C^1, is the uniform measure on the surface itself.

5.2 The Haar measure on $O(d)$.

In this Section we briefly recall some notions about the existence of the Haar measure on a compact Lie group, and in particular on $O(d)$. We need it in order to prepare the geometric decomposition for the upper bound in Subsection 6.2.1, since it is necessary to use a multiplication invariant probability measure.

Since all the results we will take from this Chapter do not need a specific formulation, we only state the results we need, that we take from [18]. To start with, we define what is a Lie group.

Definition 5.2.1 ([18], Definition 2.1). *A Lie group $(G, \cdot)$ is a manifold with a group structure such that the multiplication map $\cdot$ is regular.*

As we mentioned before, the Lie group we are interested into is $O(d)$, and it is straightforward to see that it is included in the previous definition.

Remark 5.2.1 ([18], Section 4.4). *For $d \in \mathbb{N} \setminus \{0\}$, $O(d)$ is a compact Lie group.*

Let us spend a few lines about the previous expression.

$O(d)$ is described by

$$O(d) := \left\{ M \in \mathcal{M}^{d \times d}(\mathbb{R}) : MM^T = \mathrm{Id} \right\},$$

and therefore, since the previous condition $MM^T = \mathrm{Id}$ is a system of $\frac{d(d+1)}{2}$ independent equations in $\mathbb{R}^{d^2}$, by the implicit function theorem $O(d)$ is a $\frac{d(d-1)}{2}$ dimensional manifold.

The multiplication map is the multiplication between matrices and $O(d)$ is compact since it is closed and bounded in $\mathbb{R}^{d^2}$. The closedness is trivial to verify by definition while the boundedness is a consequence of the fact that, if $U \in O(d)$, then

$$|U_{i,j}| = |\langle e_i, Ue_j \rangle| \leq |e_i||Ue_j| \leq \|U\|_\infty = 1,$$

since

$$\|Uv\|^2 = \langle Uv, Uv \rangle = \langle v, U^T Uv \rangle = \|v\|^2.$$

We mentioned before that we need a multiplication invariant probability measure on the compact Lie group $O(d)$, that is, a measure μ such that for any $U \in O(d)$ and $\mathcal{E} \subseteq O(d)$, it holds

$$\mu(U \cdot \mathcal{E}) = \mu(\mathcal{E}),$$

and that is the content of the following Section.

5.2.1 Left-invariant and right-invariant probability measures.

To start with, we define rigorously the left and right-invariant probability measures.

Definition 5.2.2 (Multiplication invariant measures.)**.** *A measure μ on a Lie group $(G, \cdot)$ is said to be left-invariant (respectively right-invariant) if for any $U \in G$ and $\mathcal{E} \subseteq G$, it holds*

$$\mu(U \cdot \mathcal{E}) = \mu(\mathcal{E}),$$

and respectively

$$\mu(\mathcal{E} \cdot U) = \mu(\mathcal{E}).$$

As examples of left and right-invariant measures on Lie groups one can consider: the Lebesgue measure dx on $(\mathbb{R}^d, +)$ or on the torus $(\mathbb{T}^d, +)$, or the measure $\frac{dx}{x}$ on $(\mathbb{R}^+, \cdot)$ where $\cdot$ is the canonical multiplication on $\mathbb{R}$.

We shall see that there exists also a unique left-invariant and right-invariant measure on $(O(d), \cdot)$, where $\cdot$ is the canonical multiplication map between matrices.

The case $d = 1$ is trivial since one has $O(1) = \{\pm 1\}$, and therefore the only left (which is also right) invariant probability measure is μ such that $\mu(\pm 1) = \frac{1}{2}$.

The left and right invariant measures on $O(2)$. The case $d = 2$ is a little bit less obvious than the case $d = 1$ but it is explicit: one has

$$O(2) := \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} : \theta \in [0, 2\pi) \right\} =: \{R^+(\theta), R^-(\theta), \theta \in [0, 2\pi)\},$$

and therefore any element of $O(2)$ is described by its determinant (± 1) and by the angle θ.

Since one has

$$R^-(\theta) = R^+(\theta) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad R^+(-\theta) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} R^+(\theta) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{and} \quad R^+(\theta)R^+(\theta') = R^+(\theta + \theta'),$$

it is straightforward to verify that

$$R^+(\theta)R^+(\theta') = R^+(\theta + \theta'), \quad R^+(\theta)R^-(\theta') = R^-(\theta + \theta'), \tag{5.2.1.1}$$

$$R^-(\theta)R^+(\theta') = R^-(\theta - \theta'), \quad R^-(\theta)R^-(\theta') = R^+(\theta - \theta'). \tag{5.2.1.2}$$

By (5.2.1.1) and (5.2.1.2) it follows that

$$(R^+(\theta))^{-1} = R^+(-\theta) \quad \text{and} \quad (R^-(\theta))^{-1} = R^-(\theta). \tag{5.2.1.3}$$

Therefore the only left-invariant (which is also right-invariant) probability measure on $O(2)$ is μ defined via

$$\int_{O(2)} d\mu(U) f(U) = \frac{1}{4\pi} \int_0^{2\pi} d\theta [f(R^+(\theta)) + f(R^-(\theta))],$$

where $d\theta$ is the Lebesgue measure on $\mathbb{R}$.

By setting $f = 1$ one finds that μ is a probability measure on $O(2)$, and it is straightforward to see that μ is both left and right-invariant since for any $U \in O(2)$ and $\mathcal{E} \subseteq O(2)$, assuming that $U = R^+(\bar{\theta})$ (the same holds if $U = R^-(\bar\theta)$), one has

$$
\begin{aligned}
\mu(U\mathcal{E}) &= \frac{1}{4\pi}\int_0^{2\pi} d\theta[\mathbb{1}_{R^+(\bar\theta)\mathcal{E}}(R^+(\theta)) + \mathbb{1}_{R^+(\bar\theta)\mathcal{E}}(R^-(\theta))] \\
&= \frac{1}{4\pi}\int_0^{2\pi} d\theta[\mathbb{1}_{\mathcal{E}}((R^+(\bar\theta))^{-1}R^+(\theta)) + \mathbb{1}_{\mathcal{E}}((R^+(\bar\theta))^{-1}R^-(\theta))] \\
&= \frac{1}{4\pi}\int_0^{2\pi} d\theta[\mathbb{1}_{\mathcal{E}}(R^+(-\bar\theta))R^+(\theta)) + \mathbb{1}_{\mathcal{E}}(R^+(-\bar\theta)R^-(\theta))] \quad \text{by (5.2.1.3)} \\
&= \frac{1}{4\pi}\int_0^{2\pi} d\theta[\mathbb{1}_{\mathcal{E}}(R^+(-\bar\theta+\theta)) + \mathbb{1}_{\mathcal{E}}(R^-(-\bar\theta+\theta))] \quad \text{by (5.2.1.1)} \\
&= \frac{1}{4\pi}\int_{-\bar\theta}^{-\bar\theta+2\pi} d\theta'[\mathbb{1}_{\mathcal{E}}(R^+(\theta')) + \mathbb{1}_{\mathcal{E}}(R^-(\theta'))] \quad \text{changing variables } \theta' = \theta - \bar\theta \\
&= \frac{1}{4\pi}\int_0^{2\pi} d\theta'[\mathbb{1}_{\mathcal{E}}(R^+(\theta')) + \mathbb{1}_{\mathcal{E}}(R^-(\theta'))] \quad \text{by periodicity of the integrand function} \\
&= \mu(\mathcal{E}) \quad \text{by definition of } \mu.
\end{aligned}
$$

Using (5.2.1.1), (5.2.1.2) and (5.2.1.3) one can prove that μ is also right-invariant.

The question is what happens in the case $d \geq 3$, where the structure of $O(d)$ is much less obvious. In this case one can use a more general existence and uniqueness result.

5.2.2 Existence and uniqueness of the Haar measure.

Theorem 5.2.1 ([18], Theorem 37.1, Existence and uniqueness of the left Haar measure.)**.** *Let $(G, \cdot)$ be a locally compact Lie group, and let Σ be the Borel σ-algebra on G. Then, up to normalization, there exists a unique measure μ on (G, Σ) such that:*

- *μ is left-invariant,*
- *μ is finite on every compact subset $K \in \Sigma$,*
- *μ is outer regular, that is*

$$\mu(S) = \inf\{\mu(U) : S \subseteq U, U open\}, \quad \forall S \in \Sigma,$$

- *μ is inner regular, that is*

$$\mu(U) = \sup\{\mu(K) : K \subseteq U, K compact\}, \quad \forall U \in \Sigma \textit{ open set.}$$

Such measure μ is said a left-invariant Haar measure.

The previous Theorem holds of course in the case of right-invariant measures too, since there is nothing more specific in the concept of left-invariance than of right-invariance, but a simple way to get a right-invariant measure from a left-invariant one is the following: if μ is left-invariant, then $\tilde\mu$ defined via

$$\tilde\mu(A) := \mu\left\{g \in G : g^{-1} \in A\right\},$$

is right-invariant. Moreover, if μ is finite on every compact subset and it is inner and outer regular, then also $\tilde\mu$ is.

One may ask whether the left and right Haar measures do coincide: in such case, the group G is said to be *unimodular*. The following Theorem provides a sufficient condition for a group to be unimodular.

Theorem 5.2.2 ([18], Proposition 35.4)**.** *If $(G, \cdot)$ is a compact Lie group, then it is unimodular.*

As a consequence of Theorems 5.2.1 and 5.2.2 and of Remark 5.2.1, one gets the following result.

Proposition 5.2.1. *For any dimension $d \geq 1$, $O(d)$ admits a unique left and right Haar measure.*

Chapter 6

Random matching for exponential and polynomial densities.

The density we are going to consider as distribution of the variables $\{X_i\}_{i=1}^n$ here have radial symmetries, thus

$$\rho(x) = \rho(|x|),$$

with

$$\rho(r) := c_d \frac{e^{-\frac{r^q}{q}}}{1+r^\gamma}, \text{ with } \begin{cases} \gamma \in \mathbb{R}, & q > 0, \\ \gamma > d, & q = 0, \end{cases} \tag{6.0.0.1}$$

where the case $q = 0$ is a slight abuse of notation to mean that one has only the polynomial contribution and not the exponential one. Moreover, c_d chosen in such a way to have

$$\int_{\mathbb{R}^d} dx \rho(x) = 1.$$

Since we have to deal with the p^{th} moment, with $p \geq 1$, we have to be sure that

$$\int_{\mathbb{R}^d} dx \rho(x)|x|^p < \infty, \text{ that is } \int_0^\infty dr r^{d+p-1}\rho(r) < \infty,$$

and the previous property holds for

$$\begin{cases} \text{any } p \geq 1, & q > 0, \\ 1 \leq p < \gamma - d, & q = 0, \end{cases}$$

and of course this means that we need $\gamma > d + 1$.

This class of densities has the property that for any $a \in \mathbb{R}$, $a < \gamma - 1$ whenever $q = 0$, there exists a constant $C > 0$ (depending on γ and a) such that for any $R \geq 1$

$$\frac{1}{C} R^{a-q+1}\rho(R) \leq \int_R^\infty dr \rho(r) r^a \leq C R^{a-q+1}\rho(R). \tag{6.0.0.2}$$

This property is a consequence of the fact that there exists a constant $C > 0$ depending on q and γ such that for any $R \geq 1$

$$\frac{1}{C} R^{q-1} \leq \frac{|\rho'(R)|}{\rho(R)} \leq C R^{q-1}.$$

This is the crucial property we are going to use, and it also implies that

$$\int_0^R dr \rho(r)^a r^b \lesssim \begin{cases} 1 & q > 0, a > 0 \text{ or } q > 0, a = 0, b < -1 \text{ or } q = 0, -\gamma a + b < -1, \\ \log R & q > 0, a = 0, b = -1 \text{ or } q = 0, -\gamma a + b = -1, \\ \rho(R)^a R^{b-q+1} & q > 0, a < 0 \text{ or } q = 0, -\gamma a + b > -1. \end{cases}$$

The critical exponent. As we mentioned in the Introduction, and in particular in (1.1.3.1), one expects that the behavior of the cost function changes whenever the exponent p becomes such that $\rho^{1-\frac{p}{d}}$ is no more an integrable function. The class of densities we are considering is vanishing for large $|x|$, hence the more p is large, the less the $(1-\frac{p}{d})^{th}$ power of the density is integrable.

Hence the critical exponent p is defined as

$$p^{crit} := \sup\left\{p \geq 1 : \int_{\mathbb{R}^d} dx \rho(x)^{1-\frac{p}{d}} < \infty\right\},$$

and it is straightforward to see that since for $q = 0$

$$\begin{aligned}\int_{\mathbb{R}^d} dx \rho(x)^{1-\frac{p}{d}} &= \mathcal{H}^{d-1}(\partial B_1) c_d^{1-\frac{p}{d}} \int_0^\infty dr r^{d-1}(1+r^{-\gamma})^{1-\frac{p}{d}} \\ &= C\left(1 + \int_1^\infty dr r^{-(\gamma-d)+\frac{\gamma p}{d}-1}\right),\end{aligned}$$

one has

$$p^{crit} = \begin{cases} d, & q > 0, \\ p^* = \frac{d(\gamma-d)}{\gamma} < \min\{d, \gamma - d\}, & q = 0, \end{cases}$$

and notice that while if $q > 0$ one has $p^{crit} = d \geq 1$, if $q = 0$, $p^{crit} \geq 1$ if and only if

$$\gamma \geq \frac{d^2}{d-1} > d + 1 = \text{ initial lower bound for } \gamma,$$

and therefore the critical exponent p which changes significantly the behavior does not necessarily exist, that is, if $q = 0$ and $\gamma \in \left(d+1, \frac{d^2}{d-1}\right)$, one has

$$p^{crit} < 1 \text{ and therefore } p \in [1, \gamma - d) \Rightarrow p > p^{crit}.$$

6.1 Restriction to a ball of radius R.

Since our purpose is to apply the method we mentioned in Subsection 1.1.2 to the probability distribution with density ρ, we want to divide all the domain in which the random variables are distributed on, that is, $\mathbb{R}^d$, in a set of squares. We will focus later on the way this partition is to be made, but what we want to stress now is that there is a region of $\mathbb{R}^d$ where the averaged number of variables is too small. Indeed, if we consider a ball (or a disk, in dimension $d = 2$) of radius R, we get

$$n - \mathbb{E}[n(B_R)] = n \int_{B_R^c} dx \rho(x) = n\mathcal{H}^{d-1}(\partial B_1) \int_R^\infty dr r^{d-1} \rho(r) \simeq n R^{d-q} \rho(R).$$

Hence, the radius R inside which we want to constrain the variables is to be chosen in such a way to have $n - \mathbb{E}[n(B_R)] \ll n$, in such a way to have a very small fraction of the variables outside the region we are excluding, that is

$$R = \begin{cases} (q \log n)^{\frac{1}{q}} & q > 0, \\ n^{\frac{1}{\gamma-d}} & q = 0. \end{cases}.$$

This choice of R is however not the best one: indeed to compare the cost of the matching with the cost of the uniform matching, one has to be sure that in each hypercube we divide the ball (or the disk) into, two conditions are satisfied:

- the density has to be almost constant (that is, it has to be transported in the uniform one through a quasi 1-Lipschitz map),
- there is a large number of variables $\{X_i\}$ to be matched with the reference measure.

Before we deal with these two requests, let us substitute the previous R with a slightly smaller radius:

$$R = \begin{cases} \sqrt[q]{q \log\left(\frac{n}{(\log n)^\alpha}\right)} & q > 0, \\ \left(\frac{n}{(\log n)^\beta}\right)^{\frac{1}{\gamma-d}} & q = 0. \end{cases}, \tag{6.1.0.1}$$

with α and β to be fixed below according to the conditions we need.

Let use begin with the first one: it means that ρ restricted to the hypercube must be transported into the constant one. By Lemma 2.2.6, it is possible if for any x in the hypercube, denoting by L its side, it holds

$$L\left|\frac{\nabla\rho(x)}{\rho(x)}\right| = L\left|\frac{\rho'(r)}{\rho(r)}\right| \leq C\varepsilon, \text{ with } r = |x| \text{ and } \varepsilon \text{ to be finally sent to } 0. \tag{6.1.0.2}$$

We recall here that by Lemma 2.2.5 the previous condition ensures that if x and y are in the same hypercube Q, one has

$$\sup_{x,y\in Q}\left|\frac{\rho(x)}{\rho(y)}-1\right|\leq C'\varepsilon. \tag{6.1.0.3}$$

For x close to the boundary of the ball, that is, $|x|\approx R$, property (6.1.0.2) writes as

$$L\left|\frac{\rho'(R)}{\rho(R)}\right|\approx LR^{q-1}\leq C\varepsilon,\ \text{and therefore } L\approx\varepsilon R^{1-q}. \tag{6.1.0.4}$$

Here comes the second condition: using (6.1.0.4) one finds out that the expectation of the number of points inside the hypercube Q_L of side L and at distance R from the origin is

$$\mathbb{E}\left[n(Q_L)\right]=n\int_{Q_L}dx\rho(x)\approx n\rho(R)L^d\approx n\varepsilon\rho(R)R^{-d(q-1)}, \tag{6.1.0.5}$$

where the first approximation is a consequence of the fact that the density is almost constant inside the hypercube, that is, (6.1.0.3), and the second one comes from (6.1.0.4). By substituting definition (6.1.0.1) of the radius R in (6.1.0.5), for $q>0$ we get

$$\mathbb{E}\left[n(Q_L)\right]\approx n\varepsilon\frac{(\log n)^\alpha}{n}\sqrt[q]{q\log\left(\frac{n}{(\log n)^\alpha}\right)}^{-\gamma-d(q-1)}\approx\varepsilon(\log n)^{\alpha-\frac{\gamma}{q}-d(1-\frac{1}{q})},$$

while for $q=0$

$$\mathbb{E}\left[n(Q_L)\right]\approx n\varepsilon\left(\frac{n}{(\log n)^\beta}\right)^{\frac{-\gamma+d}{\gamma-d}}=\varepsilon(\log n)^\beta.$$

To satisfy the second condition one has to choose $\alpha>\frac{\gamma}{q}+d(1-\frac{1}{q})$ if $q>0$ and $\beta>0$ if $q=0$. As we are going to prove later such a choice is possible for $p\leq p^{crit}$. To explain why it happens, we first notice that large values of α and β worsen the estimates of such a cut off, in Proposition 6.1.2, but improve the estimates of rearrange the masses inside the hypercubes, in Propositions 6.2.1 and 6.2.2. This happens since large values of α and β ensure that the averaged number of points inside the hypercube is very large, and therefore their variance is very small (in comparison). While if $p\leq p^{crit}$ one can choose α and β large enough as we said before but small enough to not affect the cost of the cut off in Proposition 6.1.2, for $p>p^{crit}$ it is straightforward to see that the only choice of α and β which balances the estimates in Proposition 6.1.2, 6.2.1 and 6.2.2 is

$$\alpha=\frac{\gamma}{q}+d(1-\frac{1}{q}),\ \text{if } q>0,\ \text{and } \beta=0 \text{ if } q=0.$$

Thus the critical distance is defined as follows.

Definition 6.1.1. *The radius S_n is defined as*

$$S_n:=\begin{cases}\sqrt[q]{q\log\left(\frac{n}{(\log n)^\alpha}\right)} & q>0,\\ \left(\frac{n}{(\log n)^\beta}\right)^{\frac{1}{\gamma-d}} & q=0,\end{cases}$$

with

$$\alpha\begin{cases}\in\left(\frac{\gamma}{q}+d(1-\frac{1}{q}),\frac{\gamma}{q}+d(1-\frac{1}{q})+1\right) & p\leq d,\\ =\frac{\gamma}{q}+d(1-\frac{1}{q}) & p>d,\end{cases}$$

and

$$\beta\begin{cases}>0 & p<p^*,\\ \in\left(0,\frac{1}{1-\frac{p}{\gamma-d}}\right) & p=p^*,\\ =0 & p>p^*.\end{cases}$$

What we want to prove is that the cost is basically determined by the cost of the variables which are inside the ball of radius S_n. S_n is not exactly the final radius R_n where we are going to apply the cut off, since a further step will be to divide B_{S_n} in a sequence of annuli through a corresponding sequence of radii $\{r_k\}_{k\geq1}$, and therefore we have to be careful that in this decomposition the last annulus is not too much small. In this lines we do not want to deal with the exact decomposition in the sequence of the annuli (Subsection 6.2.1) or with the exact choice of the radii ((6.2.2.8) and Definition 6.2.3). To give the definition of R_n, it is sufficient to say that, if the sequence $\{r_k\}_{k\geq 1}$ is chosen as in (6.2.2.8), $\bar{k}$ is chosen as

$$\bar{k} := \max\{k \geq 1 : r_k \leq S_n < r_{k+1}\}, \text{ and } R_n := r_{\bar{k}+1}. \tag{6.1.0.6}$$

Such R_n is the critical radius, and by (6.1.0.2) it is straightforward to see that

$$\frac{1}{C}\rho(S_n) \leq \rho(R_n) \leq C\rho(S_n), \tag{6.1.0.7}$$

where the constant $C > 0$ does not depend on n. This holds since the distance between r_k and r_{k+1} is chosen in such a way to satisfy (6.1.0.4), that is

$$(r_{k+1} - r_k)\frac{|\rho'(r_k)|}{\rho(r_k)} \leq C\varepsilon \quad \text{and therefore} \quad 1 - c\varepsilon \leq \frac{\rho(r)}{\rho(s)} \leq 1 + c\varepsilon \quad \text{for any } r_k \leq r, s \leq r_{k+1}.$$

Now we focus on the cut off at radius R_n. This means that one needs to replace all the n variables $\{X_i\}_{i=1}^n$ with other n iid random variables $\{\tilde{X}_i\}_{i=1}^n$ constrained inside a ball of radius R (or a disk, in dimension $d = 2$). To make this, we first define the following new density

Definition 6.1.2. *For $R > 0$, ρ the density in (6.0.0.1),*

$$\delta = \varepsilon R^{-q},$$

and $A_{R,R(1+\delta)} := \{x \in \mathbb{R}^d : |x| \in [R, R(1+\delta))\}$, ρ_R^ε is the probability density defined as

$$\rho_R^\varepsilon := \rho(\mathbb{1}_{B_R} + \frac{1 - \rho(B_R)}{\rho(A_{R,R(1+\delta)})}\mathbb{1}_{A_{R,R(1+\delta)}}).$$

Let us stress that the previous measure coincide with ρ in the ball of radius R but moves all the mass outside B_R in the annulus $A_{R,R(1+\delta)}$. Notice that the previous density is obtained as the sum of two terms: the first one, i.e., the coefficient multiplying $\mathbb{1}_{B_R}$, has order 1, while the second one can be approximated for large R and small $\delta = \varepsilon R^{-q}$ by

$$\begin{aligned}\frac{1 - \rho(B_R)}{\rho(A_{R,R(1+\delta)})} &= \frac{\int_R^\infty dr r^{d-1}\rho(r)}{\int_R^{R(1+\delta)} dr r^{d-1}\rho(r)} \simeq \frac{R^{d-q}\rho(R)}{R^{d-1}\delta R\rho(R)} \\ &= \frac{1}{\delta R^q} = \frac{1}{\varepsilon} \text{ in both the cases } q > 0 \text{ and } q = 0 \text{ by the choice of } \delta,\end{aligned} \tag{6.1.0.8}$$

and therefore

$$\rho(x)(\mathbb{1}_{B_R} + \frac{1}{C\varepsilon}\mathbb{1}_{A_{R,R(1+\delta)}}) \leq \rho_R^\varepsilon(x) \leq \rho(x)(\mathbb{1}_{B_R} + \frac{C}{\varepsilon}\mathbb{1}_{A_{R,R(1+\delta)}}) \tag{6.1.0.9}$$

Also notice that this new density ρ_R^ε has the same properties as ρ, that is (6.0.0.2), indeed

$$\int_S^\infty dr \rho_R^\varepsilon(r) r^a \leq C\rho(S)S^{a-q+1}. \tag{6.1.0.10}$$

Let us spend a few lines about (6.1.0.10). One has to distinguish the three cases $S > R(1+\delta)$, $S \in (R, R(1+\delta)]$ and $S \leq R$. Of course if $S > R(1+\delta)$

$$\int_S^\infty dr \rho_R^\varepsilon(r) r^a = 0,$$

since $\rho_R^\varepsilon(r) = 0$ for any $r > R(1+\delta)$. Then, if $S \in (R, R(1+\delta)]$, by property (6.1.0.9) one has

$$\int_S^\infty dr \rho_R^\varepsilon(r) r^a \leq \frac{C}{\varepsilon}\int_S^{R(1+\delta)} dr \rho(r) r^a \leq \frac{C}{\varepsilon}\rho(S)S^a R\delta = \frac{C}{\varepsilon}\rho(S)S^{a-q+1}\varepsilon\left(\frac{R}{S}\right)^{1-q} \leq C\rho(S)S^{a-q+1}, \tag{6.1.0.11}$$

and finally, if $S \leq R$, by the previous property (6.1.0.11) one has

$$\begin{aligned}\int_S^\infty dr \rho_R^\varepsilon(r) r^a &= \int_S^R dr \rho_R^\varepsilon(r) r^a + \int_R^\infty dr \rho_R^\varepsilon(r) r^a \leq \int_S^\infty dr \rho(r) r^a + \int_R^\infty dr \rho_R^\varepsilon(r) r^a \\ &\leq C\rho(S)S^{a-q+1} + C\rho(R)R^{a-q+1} \leq C'\rho(S)S^{a-q+1},\end{aligned}$$

which conclդes.

Before estimating the distance between ρ and ρ_R^ε, let us see how this new density ρ_R can substitute the previous one, that is, ρ. The main result we are going to prove is the following.

Proposition 6.1.1. *Let $\{X_i\}_{i=1}^n$ be iid random variables in $\mathbb{R}^d$ with common distribution ρ. Then there exist $\{\tilde{X}_i\}_{i=1}^n$ iid random variables in $\mathbb{R}^d$ with common distribution ρ_R^ε such that*

$$\mathbb{E}\left[W_p^p\left(\sum_{i=1}^n \delta_{X_i}, \sum_{i=1}^n \delta_{\tilde{X}_i}\right)\right] \le nW_p^p(\rho, \rho_R^\varepsilon).$$

Proof. Since ρ is absolutely continuous with respect to Lebesgue measure, thanks to Theorem 2.2.1 there exists a map $T : \mathbb{R}^d \to \mathbb{R}^d$ (depending on R and ε) such that

$$\rho_R^\varepsilon = T_\sharp \rho, \text{ and } W_p^p(\rho, \rho_R^\varepsilon) = \int_{\mathbb{R}^d} dx \rho(x)|x - T(x)|^p. \tag{6.1.0.12}$$

Therefore if $\{X_i\}_{i=1}^n$ are iid random variables with common distribution ρ, $\{T(X_i)\}_{i=1}^n$ are iid random variables with common distribution ρ_R^ε. In fact, by Definition 2.2.1 of push-forward, the law of $T(X)$ is $T_\sharp$(law of X) and the independence is preserved by the fact that $T(X_i)$ depends only on X_i.

Therefore, by setting

$$\tilde{X}_i := T(X_i),$$

and noticing that the measure (in $\mathbb{R}^d \times \mathbb{R}^d$)

$$\sum_{i=1}^n \delta_{X_i, \tilde{X}_i}$$

is a coupling of $\sum_{i=1}^n \delta_{X_i}$ and $\sum_{i=1}^n \delta_{\tilde{X}_i}$, one has

$$\begin{aligned}
\mathbb{E}\left[W_p^p\left(\sum_{i=1}^n \delta_{X_i}, \sum_{i=1}^n \delta_{\tilde{X}_i}\right)\right] &\le \mathbb{E}\left[\sum_{i=1}^n |X_i - \tilde{X}_i|^p\right] = \mathbb{E}\left[\sum_{i=1}^n |X_i - T(X_i)|^p\right] \\
&= \sum_{i=1}^n \mathbb{E}\left[|X_i - T(X_i)|^p\right] = \sum_{i=1}^n \int_{\mathbb{R}^d} dx \rho(x)|x - T(x)|^p \\
&= nW_p^p(\rho, \rho_R^\varepsilon) \text{ by (6.1.0.12)},
\end{aligned}$$

and this proves the statement. □

Therefore now the question is how much the density ρ_R^ε is far from ρ (in the Wasserstein metric). The answer is in the following Lemma.

Lemma 6.1.1. *If ρ is probability density defined in (6.0.0.1) and ρ_R^ε is the probability density in Definition 6.1.2, it holds*

$$W_p^p(\rho, \rho_R^\varepsilon) \le CR^{d-q-p(q-1)}\rho(R). \tag{6.1.0.13}$$

Proof. A possible way to prove the statement could be by finding first an explicit map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $T_\sharp \rho = \rho_R^\varepsilon$, and then estimating $|T(x) - x|$. It is possible, but we prefer to prove this by first using the subadditivity of the Wasserstein distance (see (2.1.0.2)) to get, in both the cases, the inequality

$$\begin{aligned}
W_p^p(\rho, \rho_R^\varepsilon) &= W_p^p\left(\rho \mathbb{1}_{B_R} + \rho \mathbb{1}_{B_R^c}, \rho \mathbb{1}_{B_R} + \frac{1 - \rho(B_R)}{\rho(A_{R,R(1+\delta)})} \mathbb{1}_{A_{R,R(1+\delta)}}\right) \\
&\le W_p^p(\rho \mathbb{1}_{B_R}, \rho \mathbb{1}_{B_R}) + W_p^p\left(\rho \mathbb{1}_{B_R^c} + \frac{1 - \rho(B_R)}{\rho(A_{R,R(1+\delta)})} \mathbb{1}_{A_{R,R(1+\delta)}}\right) \\
&= W_p^p\left(\rho \mathbb{1}_{B_R^c}, \frac{1 - \rho(B_R)}{\rho(A_{R,R(1+\delta)})} \mathbb{1}_{A_{R,R(1+\delta)}}\right),
\end{aligned} \tag{6.1.0.14}$$

and then by applying the triangle inequality and involving the third measure

$$\hat{\rho} := \frac{1 - \rho(B_R)}{\mathcal{H}^{d-1}(\partial B_R)} \mathcal{H}^{d-1}|_{\partial B_R},$$

where $\mathcal{H}^{d-1}$ is the $d-1$-dimensional Hausdorff measure in Definition 5.1.1. $\hat{\rho}$ moves all the mass outside B_R on the boundary of the ball (and not on the annulus).

Therefore, to prove (6.1.0.13) in both the cases, we apply the triangle inequality to the three measures involved in such a way to get

$$\begin{aligned}
(6.1.0.14) &\le W_p^p\left(\rho\mathbb{1}_{B_R^c}, \frac{1-\rho(B_R)}{\rho(A_{R,R(1+\delta)})}\mathbb{1}_{A_{R,R(1+\delta)}}\right)\\
&\le \left(W_p\left(\rho\mathbb{1}_{B_R^c},\hat\rho\right) + W_p(\hat\rho, \frac{1-\rho(B_R)}{\rho(A_{R,R(1+\delta)})}\mathbb{1}_{A_{R,R(1+\delta)}})\right)^p\\
&\le 2^pW_p^p\left(\rho\mathbb{1}_{B_R^c},\hat\rho\right) + 2^pW_p^p\left(\hat\rho, \frac{1-\rho(B_R)}{\rho(A_{R,R(1+\delta)})}\mathbb{1}_{A_{R,R(1+\delta)}}\right). \qquad (6.1.0.15)
\end{aligned}$$

The bound for the second summand in (6.1.0.15) can be performed as follows: one can observe that the map $T: A_{R,R(1+\delta)} \to \partial B_R$ defined as

$$T(x) := R\hat{x},$$

verifies

$$T_\sharp\left(\frac{1-\rho(B_R)}{\rho(A_{R,R(1+\delta)})}\mathbb{1}_{A_{R,R(1+\delta)}}\right) = \frac{1-\rho(B_R)}{\mathcal{H}^{d-1}(\partial B_R)}\mathcal{H}^{d-1}|_{\partial B_R} = \hat\rho,$$

and also

$$|T(x) - x| \le R\delta. \qquad (6.1.0.16)$$

Therefore

$$\begin{aligned}
W_p^p\left(\hat\rho, \frac{1-\rho(B_R)}{\rho(A_{R,R(1+\delta)})}\mathbb{1}_{A_{R,R(1+\delta)}}\right) &= W_p^p\left(\frac{1-\rho(B_R)}{\mathcal{H}^{d-1}(\partial B_R)}\mathcal{H}^{d-1}|_{\partial B_R}, \frac{1-\rho(B_R)}{\rho(A_{R,R(1+\delta)})}\rho\mathbb{1}_{A_{R,R(1+\delta)}}\right)\\
&\le \int_{A_{R,R(1+\delta)}} dx \frac{1-\rho(B_R)}{\rho(A_{R,R(1+\delta)})}\rho(x)|T(x)-x|^p \quad \text{by (2.2.0.1)}\\
&\le \frac{1-\rho(B_R)}{\rho(A_{R,R(1+\delta)})}\rho(A_{R,R(1+\delta)})(R\delta)^p \quad \text{by (6.1.0.16)}\\
&= (R\delta)^p\mathcal{H}^{d-1}(\partial B_1)\int_R^\infty dr r^{d-1}\rho(r)\\
&\le C(R\delta)^p \quad \rho(R)R^{d-q} \text{ by (6.0.0.2)}\\
&= \varepsilon^p C\rho(R)R^{d-q-p(q-1)} \quad \text{by Definition 6.1.2,} \qquad (6.1.0.17)
\end{aligned}$$

with $q = 0$ in the case of polynomial density.

The estimate of the other summand in (6.1.0.15) is slightly different.

In the following chain of bounds the first one is obtained as in the previous estimates.

One has

$$\begin{aligned}
W_p^p(\rho\mathbb{1}_{B_R^c},\hat\rho) &\le \mathcal{H}^{d-1}(\partial B_1)c_d^q\int_R^\infty dr\rho(r)r^{d-1}|r-R|^p\\
&= \mathcal{H}^{d-1}(\partial B_1)c_d^q\int_R^\infty dr\rho(r)r^{d-1}\frac{|r-R|^p}{|r^q-R^q|^p}|r^q-R^q|^p\\
&\le C\int_R^\infty dr r^{d-1+p(1-q)}|r^q-R^q|^p\rho(r)\\
&= C\rho(R)\int_R^\infty dr r^{d-1+p(1-q)}|r^q-R^q|^p\frac{\rho(r)}{\rho(R)}\\
&= CR^{d+p(1-q)}\rho(R)\int_1^\infty ds s^{d-1+p(1-q)-\gamma}e^{-R^q\frac{s^q-1}{q}}(R^q(s^q-1))^p \text{ with } r = Rs\\
&\le C'R^{d+p(1-q)}\rho(R)\int_1^\infty ds s^{d-1+p(1-q)-\gamma}e^{-C''R^q(s^q-1)}\\
&= C'R^{d+p(1-q)}\rho(R)\int_R^\infty dr r^{d-1+p(1-q)-\gamma}e^{-C''r^q}R^{-(d-1+p(1-q)-\gamma)-1}e^{C''R^q} \text{ with } s = r/R\\
&\le C''R^{d+p(1-q)-q}\rho(R). \qquad (6.1.0.18)
\end{aligned}$$

Then, combining all the estimates, one gets

$$W_p^p(\rho,\rho_R^\varepsilon) \le (6.1.0.15) \le (6.1.0.17) + (6.1.0.18) \le (\varepsilon^pC + C')\rho(R)R^{d-q-p(q-1)},$$

that is what we wanted to prove. $\square$

Now that we have proved the previous Lemma, we can estimate the cost of the cut off. To shorten the notation, from now on we will use the following one.

Definition 6.1.3. *For $\bar{k}$ as in (6.1.0.6), the density ρ_n is defined as*

$$\rho_n := \rho^{\varepsilon}_{r_{\bar{k}}}.$$

Proposition 6.1.2. *Let R_n be the critical radius. Then, it holds*

$$nW_p^p(\rho,\rho_n) \le C \begin{cases} (\log n)^{\frac{d}{q}-1-p(1-\frac{1}{q})+\alpha-\frac{\gamma}{q}} & q>0, \\ n^{\frac{p}{\gamma-d}}(\log n)^{\beta(1-\frac{p}{\gamma-d})} & q=0. \end{cases}$$

Proof. First we notice that, by Lemma 6.1.1, we have

$$W_p^p(\rho,\rho_n) \le CR_n^{d-q-p(q-1)}\rho(R_n),$$

and therefore, recalling property (6.1.0.7), we have

$$nW_p^p(\rho,\rho_n) \le CnS_n^{d-q-p(q-1)}\rho(S_n) \le CnS_n^{d-q-p(q-1)}e^{-\frac{S_n^q}{q}}S_n^{-\gamma}. \tag{6.1.0.19}$$

Therefore, to prove the inequality in the case $q>0$, it is sufficient to recall that in this case, by Definition 6.1.1, we have

$$S_n = \sqrt[q]{\log\left(\frac{n}{(\log n)^{\alpha}}\right)},$$

and then notice that by (6.1.0.19) we get

$$nW_p^p(\rho,\rho_n) \le C(\log n)^{\frac{d}{q}-1-p(1-\frac{1}{q})+\alpha-\frac{\gamma}{q}}.$$

Instead, in the case of the polynomial density ($q=0$) by Definition 6.1.1, we have

$$S_n = \left(\frac{n}{(\log n)^{\beta}}\right)^{\frac{1}{\gamma-d}},$$

and then, by Lemma 6.1.1,

$$nW_p^p(\rho,\rho_n) \le CnR_n^{d+p-\gamma} \le Cn^{\frac{p}{\gamma-d}}(\log n)^{\beta(1-\frac{p}{\gamma-d})}.$$

□

6.2 The geometric decomposition.

Our purpose is to divide the space in a lot of (not necessarily disjoint) subsets. The hypercubes would be the most eligible choices for this decomposition since we know the behavior (for large n) of cost of the uniform matching on an hypercube, and in fact in the lower bound (where there is no need to estimate a global term) we will use a set of disjoint hypercubes to estimate the cost. But the probability measures we are considering are radially symmetric, therefore we have to exploit this symmetry in the upper bound, where we have to estimate the distance between the real density and the density in which the masses have been rearranged.

6.2.1 Geometric decomposition for the upper bound.

To explain the partition which we are going to use we begin with the standard definition of an annulus:

$$A_{r,s} = \{x \in \mathbb{R}^d : |x| \in [r, s)\}.$$

We need to divide an annulus in a non countable union of deformed hypercubes, which we define as follows.

Definition 6.2.1. *Let $r, \delta > 0$, $y \in \partial B_1$, $U \in O(d-1)$ and R^y the rotation of $\mathbb{R}^d$ which follows the geodesic from e_d to y, defined in Subsection 4.1.1. $Q^{y,U}_{r,\delta}$ is defined as*

$$Q^{y,U}_{r,\delta} := \{x \in \mathbb{R}^d : |x| \in [r, r(1+\delta)), \hat{x} \in \Delta^{y,U}_{\delta}\}, \tag{6.2.1.1}$$

where

$$\Delta^{y,U}_{\delta} := \exp_y(R^y(U((-\delta/2, \delta/2)^{d-1}) \times \{0\})) \subseteq \partial B_1. \tag{6.2.1.2}$$

Roughly speaking, $Q^{y,U}_{r,\delta}$ is obtained as follows: we have $(-\delta/2, \delta/2)^{d-1}$ which is a $d-1$ dimensional hypercube (of side δ) on $T_{e_d}(\partial B_1)$. We rotate it by the orthogonal operator $U \in O(d-1)$ and then we embed it into $\mathbb{R}^d$ setting the last component to 0. Then we roteate this embedded hypercube by $R^y \in SO(d)$, that is, the rotation which follows the geodesic from e_d to y. This way, we get a $d-1$ dimensional hypercube of side δ on $T_y(\partial B_1)$, which we project it on ∂B_1 through the exponential map, as in Figure 6.2.1.1.

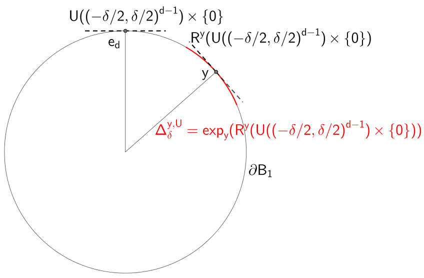

Figure 6.2.1.1: The angular component of $Q^{y,U}_{r,\delta}$ is $\Delta^{y,U}_{\delta}$, that is, the red arc.

We finally rescale it by r and add the radial component, as in Figure 6.2.1.2.

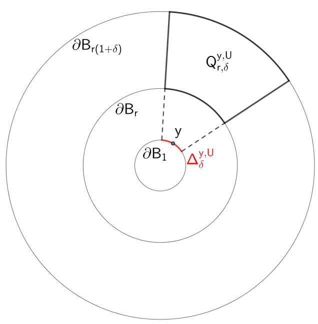

Figure 6.2.1.2: The region delimited by the thick boundary is $Q^{y,U}_{r,\delta}$.

It is fairly straightforward to see that the regions defined above closely resemble hypercubes, up to a suitable transformation. But what we need is instead a more precise estimate on this transformation: in particular, the result we need about this $Q^{y,U}_{r,\delta}$ is the following.

Lemma 6.2.1. *There exist a diffeomorphim $S : [0,1]^d \to Q^{y,U}_{r,\delta}$ and a constant $C > 0$ such that*

$$\sup_{x\in[0,1]^d} \left\| \frac{DS}{r\delta}(x) - R^y U \right\| \le C\delta, \tag{6.2.1.3}$$

$$\left| \frac{\det DS(x)}{(r\delta)^d} - \det U \right| \le C\delta, \tag{6.2.1.4}$$

$$\sup_{x\in[0,1]^d} \frac{|\nabla \det DS(x)|}{|\det DS(x)|} \le C\delta. \tag{6.2.1.5}$$

We shall prove it in Subsection 6.2.2. Notice also that the union of this deformed hypercubes over y and U, for fixed r and δ, is exactly the annulus with radii r and $r(1+\delta)$, indeed

$$\bigcup_{y\in\partial B_1, U\in O(d-1)} Q^{y,U}_{r,\delta} = A_{r,r(1+\delta)}.$$

In a lot of cases, we will not use the partition we mentioned before in that order. That is, one can partition the space in annuli $A_{r_k,r_{k+1}}$ (the sequence $\{r_k\}_k$ of the radii to be defined later) and then divide each annulus in the non countable and not disjoint union of $Q^{y,U}_{r_k,\delta_k}$. Nevertheless, sometimes doing the partition following such order produces an error term too much large. Therefore, in the most of the cases the partition we are going to use is made as follows: first we divide the ball B_{R_n} in a disk (with radius r_1) and an annulus A_{r_1,R_n}, in such a way to exclude the origin. Then, we divide this large annulus A_{r_1,R_n} following the radial partition given by $\Delta^{y,U}_\delta$ in (6.2.1.2) in Definition 6.2.1. This means that we want to consider regions like

$$\{x \in \mathbb{R}^d : |x| \in [r_1, R_n), \hat{x} \in \Delta^{y,U}_\delta\}, \tag{6.2.1.6}$$

and therefore we use the following more general notation.

Definition 6.2.2. *For $\Sigma \subseteq \partial B_1$ and $r < s$ the truncated pyramid $A^\Sigma_{r,s}$ is*

$$A^\Sigma_{r,s} := \left\{x \in \mathbb{R}^d : r \le |x| < s, \hat{x} \in \Sigma\right\}, \qquad \textit{with } r < s.$$

An example of region $A^\Sigma_{r,s}$ is in the following Figure 6.2.1.3. Notice that the region we defined in (6.2.1.6)

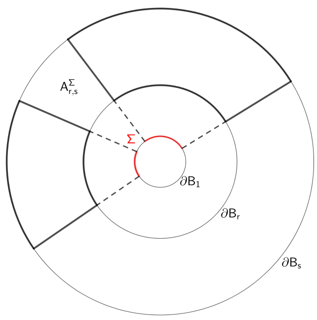

Figure 6.2.1.3: If Σ is the union of the two red arcs, $A^\Sigma_{r,s}$ is the region inside the continue thick lines.

is a truncated pyramid as in Definition 6.2.2, that is,

$$(6.2.1.6) = A^{\Delta^{y,U}_\delta}_{r_1,R_n},$$

and also that, for $\Sigma = \partial B_1$ $A^\Sigma_{r,s}$, coincides with the annulus $A_{r,s}$.

Moreover the union over $\Sigma = \Delta^{y,U}_\delta$ of $A^\Sigma_{r,s}$ is the annulus $A_{r,s}$, that is

$$\bigcup_{y\in\partial B_1, U\in O(d-1)} A^{\Delta^{y,U}_\delta}_{r,s} = A_{r,s}.$$

Once made this radial decomposition of A_{r_1,R_n} in the previous pyramidal regions, the next one is to split each pyramid in $Q^{y,U}_{r,\delta}$ as in (6.2.1.1) in Definition 6.2.1, and this partition only involves the radial component.

Now that we have explained the decomposition we are going to use in the upper bound, we can focus on the proof of Lemma 6.2.1, which makes rigorous the idea of moving each $Q^{y,U}_{r_k,\delta_k}$ in an hypercube.

Before proving such result, we state the following estimates about the matrix norms. They are quite immediate but we shall prove it later. For $A \in \mathcal{M}^{d\times d}(\mathbb{R})$, we will denote

$$\|A\|_\infty := \max_{i,j} |A_{i,j}|, \quad \|A\| := \max_{|v|=1} |Av|.$$

Lemma 6.2.2. *Let $A, B \in \mathcal{M}^{d\times d}(\mathbb{R})$. Then, if we denote by A_j the j^{th} column of A, we have*

$$|A_{i,j}| \le \|A_j\| \le \|A\| \le \sqrt{\sum_{k=1}^d \|A_k\|^2} \le d\|A\|_\infty; \tag{6.2.1.7}$$

$$|\det(A+B) - \det B| \le d!d\|A\|_\infty(\|A\|_\infty + \|B\|_\infty)^{d-1}. \tag{6.2.1.8}$$

Moreover, if $A = A(x)$ depends on $x \in \mathbb{R}^d$, it holds

$$|\nabla \det A(x)| \le d!\sqrt{d} \max_j \left\|\frac{\partial}{\partial x_j} A(x)\right\|_\infty \|A(x)\|_\infty^{d-1}. \tag{6.2.1.9}$$

6.2.2 Proof of Lemmas 6.2.1 and 6.2.2.

Proof of Lemma 6.2.1. To build such a map S we exploit the exponential map which we defined in the previous Section: we want to move the angular component $\Delta^{y,U}_{r,\delta}$ in $[0,1]^{d-1}$ and then the radial component $[r_k, r_k(1+\delta_k)]$ into $[0,1]$. To make this, for $U \in O(d-1)$, first we move $[0,1]^d$ in $U([-\frac{1}{2},\frac{1}{2}]^{d-1}) \times [0,1]$ via the map

$$\hat{S} : [0,1]^d \to U\left(\left[-\frac{1}{2},\frac{1}{2}\right]^{d-1}\right) \times [0,1], \quad \hat{S}(x) = \tilde{U}\begin{pmatrix} x_1 - \frac{1}{2} \\ \vdots \\ x_{d-1} - \frac{1}{2} \\ x_d \end{pmatrix}, \quad \tilde{U} = \begin{pmatrix} U & 0 \\ 0 & 1 \end{pmatrix}.$$

The first $d-1$ components have to be moved into the angular component and the last one into the radial one. To do that, we denote again by $R^y = R^y_{e_d} \in M(\mathbb{R}^{d\times d})$ the rotation from $e_d = (0,\dots,0,1)$ to y following the geodesic between the two points (see Subsection 4.1.1). Then, we consider

$$x \in U\left(\left[-\frac{1}{2},\frac{1}{2}\right]^{d-1}\right), \quad t \in [0,1], \quad \tilde{x} := (x,0) \in \mathbb{R}^d,$$

and with this notation we define the map

$$F : U\left(\left[-\frac{1}{2},\frac{1}{2}\right]^{d-1}\right) \times [0,1] \subseteq \mathbb{R}^d, \qquad F(x,t) = \exp_y(\delta R^y \tilde{x}) r(1+t\delta).$$

Roughly speaking, the map F takes (x,t) as an argument, rescales all by δ, sends the first $d-1$ components (that is, x) into the corresponding vector in $T_{ry}(\partial B_r)$, sends the point obtained via this procedure in ∂B_r through the exponential map, and finally sends the last component (that is, t) into the radial component (rescaled by r).

The map S we were looking for is

$$S := F \circ \hat{S},$$

and now we shall prove that such a map has all the properties we claimed before.

We claim that the following three properties hold:

$$\sup_{(x,t)} \left\|\frac{DF(x,t)}{r\delta} - R^y\right\| \le c\delta, \tag{6.2.2.1}$$

$$\sup_{(x,t)} \left|\frac{\det DF(x,t)}{(r\delta)^d} - 1\right| \le c\delta \tag{6.2.2.2}$$

$$\sup_{(x,t)} \frac{|\nabla(\det DF)(x,t)|}{\det DF(x,t)} \le c\delta. \tag{6.2.2.3}$$

Before proving them, notice that these two estimates conclude the proof of this first step, indeed one has

$$\begin{aligned}\sup_{x\in[0,1]^d}\left\|\frac{DS(x)}{r\delta}-R^y\tilde{U}\right\| &= \sup_{x\in[0,1]^d}\left\|\frac{DF(\hat{S}(x))}{r\delta}\tilde{U}-R^y\tilde{U}\right\| \le \sup_{(x,t)}\left\|\frac{DF(x,t)}{r\delta}-R^y\right\|\|\tilde{U}\| \\ &= \sup_{(x,t)}\left\|\frac{DF(x,t)}{r\delta}-R^y\right\| \le c\delta, \quad \text{by (6.2.2.1)},\end{aligned}$$

which proves (6.2.1.3).

As for (6.2.1.4), one has

$$\left|\frac{\det DS(x)}{(r\delta)^d}-\det U\right| = \left|\frac{\det DS(x)}{(r\delta)^d}-\det\tilde{U}\right| = \left|\frac{\det DF(\hat{S}(x))\det\tilde{U}}{(r\delta)^d}-\det\tilde{U}\right| = \left|\frac{\det DF(\hat{S}(x))}{(r\delta)^d}-1\right| \le c\delta,$$

by (6.2.2.2).

Moreover, using again that $D\hat{S}=\tilde{U}$, we also have

$$\begin{aligned}\det DS(x) &= \det(DF(\hat{S}(x))\tilde{U}) = \det(DF(\hat{S}(x)))\det\tilde{U} \\ &\Rightarrow \nabla(\det DS)(x) = \tilde{U}^T\nabla(\det DF)(\hat{S}(x))\det\tilde{U} \\ &\Rightarrow \frac{|\nabla(\det DS)(x)|}{|\det DS(x)|} = \frac{|\tilde{U}^T\nabla(\det DF)(\hat{S}(x))|}{\det(DF(\hat{S}(x)))}|\det\tilde{U}| = \frac{|\nabla(\det DF)(\hat{S}(x))|}{\det(DF(\hat{S}(x)))} \le c\delta, \quad \text{by (6.2.2.3)},\end{aligned}$$

and this proves (6.2.1.5).

Thus, we only have to prove (6.2.2.1), (6.2.2.2) and (6.2.2.3).

Proof of (6.2.2.1). Differentiating with respect to x_j for some $j=1,\dots,d-1$, we get

$$\begin{aligned}\frac{\partial}{\partial x_j}F(x,t) &= \lim_{\xi\to 0}\frac{F(x+\xi e_j,t)-F(x,t)}{\xi} = r(1+t\delta)\lim_{\xi\to 0}\frac{\exp_y(\delta R^y\tilde{x}+\xi\delta R^y e_j)-\exp_y(\delta R^y\tilde{x})}{\xi} \\ &= r\delta(1+t\delta)D\exp_y(\delta R^y\tilde{x})R^y e_j, \quad \text{by definition of } D\exp_P \text{ in (4.2.0.1)},\end{aligned} \tag{6.2.2.4}$$

while, differentiating with respect to the last variable t, we get

$$\frac{\partial F(x,t)}{\partial t} = \frac{\partial}{\partial t}\exp_y(\delta R^y\tilde{x})r(1+t\delta) = r\delta\exp_y(\delta R^y\tilde{x}). \tag{6.2.2.5}$$

Thus we can apply (6.2.1.7) of Lemma 6.2.2 with

$$A = \frac{DF(x,t)}{r\delta}-R^y, \quad \text{so that } A_j = \begin{cases}(1+t\delta)D\exp_y(\delta R^y\tilde{x})R^y e_j - R^y e_j & j=1,\dots,d-1 \\ \exp_y(\delta R^y\tilde{x}) - R^y e_d = \exp_y(\delta R^y\tilde{x}) - y & j=d\end{cases}$$

in order to get

$$\begin{aligned}\left\|\frac{DF(x,t)}{r\delta}-R^y\right\|^2 &\le \sum_{j=1}^{d-1}\left|(1+t\delta)(D\exp_y(\delta R^y\tilde{x})R^y e_j - R^y e_j)+t\delta R^y e_j\right|^2 + |\exp_y(\delta R^y\tilde{x})-y|^2 \\ &\le 2(1+t\delta)^2\sum_{j=1}^{d-1}\left|D\exp_y(\delta R^y\tilde{x})R^y e_j - R^y e_j\right|^2 + 2(t\delta)^2\sum_{j=1}^{d-1}|R^y e_j|^2 + |\exp_y(\delta R^y\tilde{x})-y|^2 \\ &\le 2(1+t\delta)^2\sum_{j=1}^{d-1}C|\delta R^y\tilde{x}|^2|R^y e_j|^2 + 2(d-1)(t\delta)^2 + |\delta R^y\tilde{x}|^2 \text{ by (4.2.0.4) and (4.2.0.6) of Lemma 4.2.1} \\ &= (1+t\delta)^2\sum_{j=1}^{d-1}C\delta^2|\tilde{x}|^2 + \delta^2|\tilde{x}| \le Cd\delta^2, \quad \text{that is, (6.2.2.1).}\end{aligned}$$

Proof of (6.2.2.2). By (6.2.1.8) of Lemma 6.2.2 we also have

$$\begin{aligned}\left|\frac{\det(DF(x,t))}{(r\delta)^d}-1\right| &= \left|\det\left(\frac{DF(x,t)}{r\delta}\right)-\det R^y\right| \\ &\le d!d\left\|\frac{DF(x,t)}{r\delta}-R^y\right\|\left(\left\|\frac{DF(x,t)}{r\delta}-R^y\right\|+\|R^y\|\right)^{d-1} \\ &\le C\delta(1+C\delta)^{d-1} \quad \text{by (6.2.2.1)},\end{aligned}$$

hence (6.2.2.2) is proved.

Proof of (6.2.2.3)**.** Let us start with the estimate of the numerator, for which we have to deal with the second derivatives of F. We claim that, for a constant $C > 0$, for any $(x,t) \in [0,1]^d$ we have

$$\left\| \frac{\partial}{\partial x_k}(DF)(x,t) \right\|_\infty \leq Cr\delta^2, \quad \left\| \frac{\partial}{\partial t} DF(x,t) \right\|_\infty \leq Cr\delta^2, \quad \|DF(x,t)\|_\infty \leq r\delta. \tag{6.2.2.6}$$

Before proving the inequalities in (6.2.2.6), we see how to conclude the proof using them: by (6.2.1.9) of Lemma 6.2.2 we have

$$\begin{aligned} |\nabla(\det(DF(x,t)))| &\leq \sqrt{d}d! \max\left\{ \max_j \left\| \frac{\partial DF(x,t)}{\partial x_j} \right\|_\infty, \left\| \frac{\partial DF(x,t)}{\partial t} \right\|_\infty \right\} \|DF(x,t)\|_\infty^{d-1} \\ &\leq \sqrt{d}d! r\delta^2 (r\delta)^{d-1} = \sqrt{d}d!(r\delta)^d \delta, \quad \text{by (6.2.2.6)}. \end{aligned} \tag{6.2.2.7}$$

Therefore, thanks to (6.2.2.7) and to (6.2.2.2), which implies

$$\det(DF(x,t)) \geq (r\delta)^d (1 - c\delta),$$

we have estimated both the numerator and the denominator in (6.2.2.3), and therefore

$$\frac{|\nabla(\det DF)(x,t)|}{\det DF(x,t)} \leq C\sqrt{d}d!d \frac{\delta}{1 - c\delta},$$

hence (6.2.2.3) is proved.

Therefore we only have to prove (6.2.2.6), which is straightforward by (6.2.2.4) and (6.2.2.5): for $j,k = 1,\dots,d-1$ we have

$$\begin{aligned} \frac{\partial^2}{\partial x_j \partial x_k} F(x,t) &= \lim_{\xi \to 0} \frac{\frac{\partial}{\partial x_j} F(x + \xi e_k, t) - \frac{\partial}{\partial x_j} F(x,t)}{\xi} \\ &= r\delta(1 + t\delta) \lim_{\xi \to 0} \frac{D\exp_y(\delta R^y \tilde{x} + \xi \delta R^y e_k) R^y e_j - D\exp_y(\delta R^y \tilde{x}) R^y e_j}{\xi} \\ &= (1 + t\delta) r\delta^2 D^2 \exp_y(\delta R^y \tilde{x}) R^y e_j R^y e_k, \quad \text{by definition of } D^2 \exp_P \text{ in (4.2.0.3)}. \end{aligned}$$

Hence we have

$$\begin{aligned} \left| \frac{\partial^2}{\partial x_j \partial x_k} F_i(x,t) \right| &\leq \left| \frac{\partial^2}{\partial x_j \partial x_k} F(x,t) \right| \\ &= (1 + t\delta) r\delta^2 \left| D^2 \exp_y(\delta R^y \tilde{x}) R^y e_j R^y e_k \right| \\ &\leq Cr\delta^2 |R^y e_j| |R^y e_k| \quad \text{with } C \text{ by (4.2.0.8) of Lemma 4.2.1} \\ &= Cr\delta^2. \end{aligned}$$

Moreover by (6.2.2.4) we have

$$\left| \frac{\partial^2}{\partial x_k \partial t} F(x,t) \right| = \left| r\delta^2 D\exp_y(\delta R^y \tilde{x}) R^y e_k \right| \leq Cr\delta^2 |R^y e_k| = Cr\delta^2, \quad \text{by (4.2.0.7) of Lemma 4.2.1},$$

and by (6.2.2.5)

$$\frac{\partial^2}{\partial t^2} F(x,t) = 0,$$

therefore the first two estimates in (6.2.2.6) are proved. The third one is a consequence of (6.2.2.1), indeed

$$\|DF(x,t)\| = r\delta \left\| \frac{DF(x,t)}{r\delta} - R^y + R^y \right\| \leq r\delta \left(\left\| \frac{DF(x,t)}{r\delta} - R^y \right\| + \|R^y\| \right) \leq r\delta(C\delta + 1),$$

and therefore the proof is concluded. □

Now we focus on the estimates in Lemma 6.2.2 which we left aside before.

Proof of Lemma 6.2.2. We begin with the proof of (6.2.1.7). The first inequality is trivial since

$$|A_{i,j}| \leq \sqrt{\sum_{k=1}^d |A_{k,j}|^2} = \|A_j\|,$$

and the second one is a consequence of the fact that

$$\|A_j\| = \|Ae_j\| \leq \|A\|\|e_j\| = \|A\|.$$

Then, for the third one we can notice that for any $v \in \mathbb{R}^d$

$$\|Av\|^2 = \sum_{i=1}^d |(Av)_i|^2 = \sum_{i=1}^d \left(\sum_{j=1}^d A_{i,j} v_j\right)^2 \leq \sum_{i=1}^d \left(\sum_{j=1}^d A_{i,j}^2\right) \sum_{k=1}^d v_k^2 = \sum_{k=1}^d v_k^2 \sum_{j=1}^d \|A_j\|^2,$$

and therefore one can conclude optimizing over $v \in \mathbb{R}^d$.

As for the fourth inequality in (6.2.1.7), by the third inequality we have

$$\|A\| \leq \sqrt{\sum_{j=1}^d \|A_j\|^2} = \sqrt{\sum_{i,j=1}^d A_{i,j}^2} \leq \sqrt{d^2\|A\|_\infty^2} = d\|A\|_\infty.$$

As for (6.2.1.8) we can use the formula for the determinant, that is

$$\begin{aligned}
\det(A+B) &= \sum_{\sigma \in \mathcal{S}^d} (-1)^\sigma \prod_{i=1}^d (A+B)_{i,\sigma(i)} = \sum_{\sigma \in \mathcal{S}^d} (-1)^\sigma \sum_{I \subseteq \{1,\dots,d\}} \prod_{i \in I} A_{i,\sigma(i)} \prod_{i \notin I} B_{i,\sigma(i)} \\
&= \sum_{\sigma \in \mathcal{S}^d} (-1)^\sigma \prod_{i=1}^d B_{i,\sigma(i)} + \sum_{\sigma \in \mathcal{S}^d} (-1)^\sigma \sum_{I \subseteq \{1,\dots,d\}, |I| \geq 1} \prod_{i \in I} A_{i,\sigma(i)} \prod_{i \notin I} B_{i,\sigma(i)} \\
&= \det B + \sum_{\sigma \in \mathcal{S}^d} (-1)^\sigma \sum_{I \subseteq \{1,\dots,d\}, |I| \geq 1} \prod_{i \in I} A_{i,\sigma(i)} \prod_{i \notin I} B_{i,\sigma(i)}.
\end{aligned}$$

Therefore, since by definition $|A_{i,j}| \leq \|A\|_\infty$, we have

$$\begin{aligned}
|\det(A+B) - \det B| &\leq \sum_{\sigma \in \mathcal{S}^d} \sum_{I \subseteq \{1,\dots,d\}, |I| \geq 1} \prod_{i \in I} |A_{i,\sigma(i)}| \prod_{i \notin I} |B_{i,\sigma(i)}| \leq d! \sum_{I \subseteq \{1,\dots,d\}, |I| \geq 1} \|A\|_\infty^{|I|} \|B\|_\infty^{d-|I|} \\
&= d! \sum_{k=1}^d \binom{d}{k} \|A\|_\infty^k \|B\|_\infty^{d-k} = d!\|A\|_\infty \sum_{\ell=0}^{d-1} \binom{d}{\ell+1} \|A\|_\infty^\ell \|B\|_\infty^{d-1-\ell} \\
&\leq d! d \|A\|_\infty \sum_{\ell=0}^{d-1} \binom{d}{\ell} \|A\|_\infty^\ell \|B\|_\infty^{d-1-\ell} = d! d \|A\|_\infty (\|A\|_\infty + \|B\|_\infty)^{d-1},
\end{aligned}$$

and therefore also (6.2.1.8) is proved.

Then, (6.2.1.9) follows by direct computations: one has

$$\frac{\partial \det A(x)}{\partial x_j} = \sum_{\sigma \in \mathcal{S}^d} (-1)^\sigma \frac{\partial}{\partial x_j} \prod_{i=1}^d A_{i,\sigma(i)}(x) = \sum_{\sigma \in \mathcal{S}^d} (-1)^\sigma \sum_{k=1}^d \frac{\partial}{\partial x_j} A(x) \prod_{i \neq k} A_{i,\sigma(i)},$$

hence

$$\left|\frac{\partial \det A(x)}{\partial x_j}\right| \leq d! \left\|\frac{\partial A(x)}{\partial x_j}\right\|_\infty \|A(x)\|_\infty^{d-1},$$

and therefore

$$|\nabla \det A(x)| = \sqrt{\sum_{j=1}^d \left(\frac{\partial}{\partial x_j} A(x)\right)^2} \leq \sqrt{d} d! \max_j \left\|\frac{\partial A(x)}{\partial x_j}\right\|_\infty \|A(x)\|_\infty^{d-1},$$

thus also (6.2.1.9) is proved. □

The global estimate (rearrange the mass inside the squares).

To make the decomposition one defines the sequence of radii

$$r_1 \text{ fixed}, \qquad r_{k+1} = r_k(1+\delta_k), \text{ for } k \geq 0, \qquad \text{with } \delta_k \text{ to be defined below.} \tag{6.2.2.8}$$

Since it is a bit heavy to write $\sum_{k=1}^{\bar{k}}$ and $\bigcup_{k=1}^{\bar{k}}$, with $\bar{k}$ as in (6.1.0.6), we will use the following notation.

Notation 6.2.1. *For $\bar{k}$ chosen as in* (6.1.0.6)*, we will denote*

$$\sum_{k\geq 1} := \sum_{k=1}^{\bar{k}}, \qquad \textit{and} \qquad \bigcup_{k\geq 1} := \bigcup_{k=1}^{\bar{k}}.$$

The decomposition we are going to use is the following:

$$B_{R_n} = B_{r_1} \cup \bigcup_{y\in\partial B_1} \bigcup_{U\in O(d-1)} \bigcup_{k\geq 1} Q^{y,U}_{r_k,\delta_k},$$

and of course it depends on the choice of δ_k. We will use the following notation.

Definition 6.2.3. *Given the sequence of radii in* (6.2.2.8)*, the decomposition of B_{R_n}*

$$B_{R_n} = B_{r_1} \cup \bigcup_{y\in\partial B_1} \bigcup_{U\in O(d-1)} \bigcup_{k\geq 1} Q^{y,U}_{r_k,\delta_k},$$

is said to be of type radial-angular if

$$\delta_k = \varepsilon r_k^{-q}, \tag{6.2.2.9}$$

whereas it is said to be of type angular-radial if

$$\delta_k = \delta = \varepsilon r_{\bar{k}}^{-q}, \tag{6.2.2.10}$$

with $\bar{k}$ as in (6.1.0.6).

In the previous Definition, the case of the polynomial densities has been included with $q = 0$. To clarify the way these two decompositions are made one can look at the following Figure 6.2.2.1.

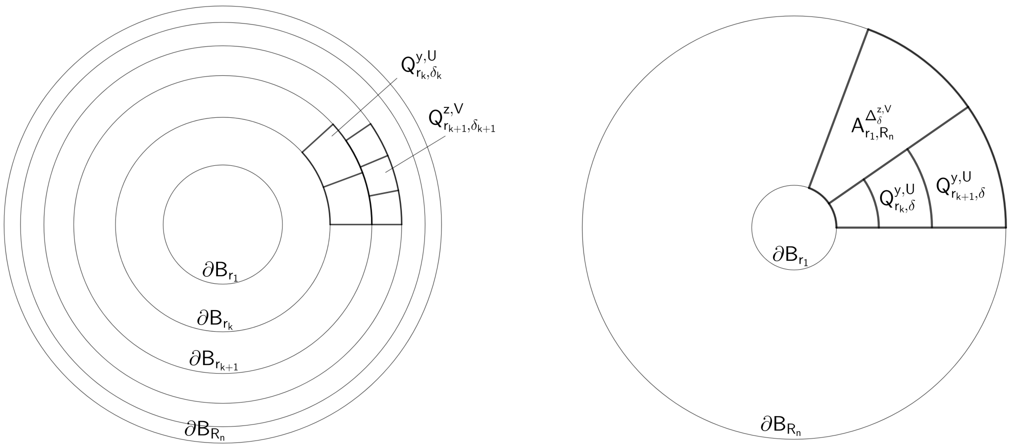

Figure 6.2.2.1: On the left: the radial-angular decomposition. On the right: the angular-radial decomposition.

Now in the upper bound we have to rearrange the mass inside each $Q^{y,U}_{r_k,\delta_k}$. As a general notation, from now on we will denote as follows the number of random variables inside domain A.

Notation 6.2.2. *Given $\{\tilde{X}_i\}_{i=1}^n$ be iid random variables with common probability density ρ_n and $A \subseteq \mathbb{R}^d$ a measurable set, we will use the notation*

$$n(A) := \sum_{i=1}^{n} \mathbb{1}_A(\tilde{X}_i).$$

In particular, using Notation 6.2.2, we have to estimate the distance between the two measures

$$W_p^p\left(n\rho_n\mathbb{1}_{B_{R_n}}, \frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}} + \sum_{k\geq 1}\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \frac{n(Q^{y,U}_{r_k,\delta_k})}{\mathcal{M}_{\delta_k}\rho_n(Q^{y,U}_{r_k,\delta_k})}\rho_n\mathbb{1}_{Q^{y,U}_{r_k,\delta_k}}\right). \tag{6.2.2.11}$$

To specify the role of the constant $\mathcal{M}_\delta$ appearing before, let us pretend that, instead of such decomposition, we had one indexed on a countable set (the current one is indexed on k, which is the only discrete index, but also on $y \in \partial B_1$ and $U \in O(d-1)$). If it was, $\mathcal{M}_\delta$ would represent the number of sets in (6.2.1.2) in Definition 6.2.1 which any x in the annulus $A_{r_k,r_{k+1}}$ belongs to. Since it is not, $\mathcal{M}_\delta$ is the measure of the (y, U) such that x belongs to $Q^{y,U}_{r_k,\delta_k}$, for x in the annulus $A_{r_k,r_{k+1}}$, that is

$$\mathcal{M}_\delta := \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\mathbb{1}_{Q^{y,U}_{r,\delta}}(x) = \mathbb{1}[r, r(1+\delta))(|x|)\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\mathbb{1}_{\Delta^{y,U}_\delta}(\hat{x}),$$

which does not depend on x by rotation invariance.

Let us spend a few lines about this: it is quite intuitive to understand why it is, since if one projects an hypercube on the surface of a sphere, by radial symmetry there is no reason why the measure of such hypercube should change if one changes the starting point of the projection. But let us prove that rigorously: if $x, z \in \mathbb{R}^d$, let us denote by V any orthonormal operator such that $\hat{z} = V\hat{x}$. Our purpose is to prove that

$$\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\mathbb{1}_{\Delta^{y,U}_\delta}(V\hat{x}) = \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\mathbb{1}_{\Delta^{y,U}_\delta}(\hat{x}), \tag{6.2.2.12}$$

and it is true since

$$\begin{aligned}\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\mathbb{1}_{\Delta^{y,U}_\delta}(V\hat{x}) &= \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\mathbb{1}_{V^T\Delta^{y,U}_\delta}(\hat{x})\\ &= \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\mathbb{1}_{\Delta^{V^Ty,L(V,y)U}_\delta}(\hat{x}) \quad \text{same argument as (6.2.2.32)}\\ &= \int_{\partial B_1} d\mathcal{H}^{d-1}(y')\int_{O(d-1)} dU'\mathbb{1}_{\Delta^{y',U'}_\delta}(\hat{x}), \quad \text{with } y' = V^Ty, U' = L(V,y)U,\end{aligned}$$

and where the orthonormal operator $L(V, y)$ is obtained as in (6.2.2.34). Therefore (6.2.2.12) is proved.

Now we can go back to $\mathcal{M}_\delta$. Even if we did not emphasized it in the notation, $\mathcal{M}_\delta$ also depends on r, but it is constant inside each annulus. A final clarification about $\mathcal{M}_{\delta_k}$: for any measure μ and function f one has

$$\sum_{k\geq 1}\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\frac{1}{\mathcal{M}_{\delta_k}}\int_{Q^{y,U}_{r_k,\delta_k}} d\mu(x)f(x) = \int_{B_{R_n}\setminus B_{r_1}} d\mu(x)f(x), \tag{6.2.2.13}$$

indeed

$$\begin{aligned}&\sum_{k\geq 1}\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\frac{1}{\mathcal{M}_{\delta_k}}\int_{Q^{y,U}_{r_k,\delta_k}} d\mu(x)f(x)\\ &= \sum_{k\geq 1}\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\frac{1}{\mathcal{M}_{\delta_k}}\int_{\mathbb{R}^d} d\mu(x)f(x)\mathbb{1}_{[r_k,r_k(1+\delta_k)]}(|x|)\mathbb{1}_{\Delta^{y,U}_{r_k,\delta_k}}(\hat{x})\\ &= \int_{\mathbb{R}^d} d\mu(x)f(x)\sum_{k\geq 1}\mathbb{1}_{[r_k,r_k(1+\delta_k)]}(|x|) \quad \text{by definition of } \mathcal{M}_{\delta_k}\\ &= \int_{B_{R_n}\setminus B_{r_1}} d\mu(x)f(x),\end{aligned}$$

and this concludes.

Our purpose is to prove the following result about (6.2.2.11).

Theorem 6.2.1. *Let ρ_n be the density in Definition 6.1.3, $n(\cdot)$ as in Notation 6.2.2 and $g(n)$ be the quantity*

$$g(n) := \mathbb{E}\left[W_p^p\left(n\rho_n\mathbb{1}_{B_{R_n}}, \frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}} + \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\sum_{k\geq 1}\frac{n(Q^{y,U}_{r_k,\delta})}{\mathcal{M}_\delta\rho_n(Q^{y,U}_{r_k,\delta})}\rho_n\mathbb{1}_{Q^{y,U}_{r_k,\delta}}\right)\right],$$

(depending also on p, d, q, γ and r_1). Then, if the decomposition is chosen

- *as in* (6.2.2.9) *for* $1 \le p < d = 2, q > 0$,
- *as in* (6.2.2.10) *otherwise,*

it holds

$$g(n) \begin{cases} \ll \tau_n^{p,d} & p \le p^{crit}, \\ \lesssim \tau_n^{p,d} & p > p^{crit}, \end{cases}$$

where the implicit constant depends on ε.

To achieve this result, we have to prove it step by step, exploiting the radial symmetry of the distribution ρ_n. Thus we begin with two preliminary Lemmas, the first one concerning the radial estimate and the second one the angular one, and then we see how we can combine them to estimate the expectation of (6.2.2.11).

Lemma 6.2.3 (The radial estimate). *Let* Σ *be any subset of* ∂B_1, $\{r_k\}_{k\ge1}$ *be the sequence of radii defined in* (6.2.2.8) *and* $\{\tilde X_i\}_{i=1}^n$ *iid random variables with common probability density* ρ_n *as in Definition 6.1.3. Then, using* $n(\cdot)$ *as in Notation 6.2.2, there exists a constant* $C(\varepsilon) > 0$ *depending on* p, q, d, γ, r_1 *and* ε *such that:*

$$\mathbb{E}\left[W_p^p\left(n\rho_n \mathbb{1}_{B_{R_n}}, \frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}} + \frac{n(A_{r_1,R_n})}{\rho_n(A_{r_1,R_n})}\rho_n\mathbb{1}_{A_{r_1,R_n}}\right)\right] \le C(\varepsilon)n^{1-\frac{p}{2}}, \tag{6.2.2.14}$$

$$\mathbb{E}\left[W_p^p\left(\frac{n(A^\Sigma_{r_1,R_n})}{\rho_n(A^\Sigma_{r_1,R_n})}\rho_n\mathbb{1}_{A^\Sigma_{r_1,R_n}}, \sum_{k\ge1}\frac{n(A^\Sigma_{r_k,r_{k+1}})}{\rho_n(A^\Sigma_{r_k,r_{k+1}})}\rho_n\mathbb{1}_{A^\Sigma_{r_k,r_{k+1}}}\right)\right] \le C(\varepsilon)(n\mathcal{H}^{d-1}(\Sigma))^{1-\frac{p}{2}}\int_{r_1}^{R_n} dr \rho(r)^{1-\frac{p}{2}} r^{d-1-p\frac{d+q-2}{2}} \tag{6.2.2.15}$$

Proof. The proofs of the first two statements are quite similar, and we begin with the first one.

Proof of (6.2.2.14). Notice that the two measures involved in the Wasserstein distance are the product of a radial part and of an angular part, and that the angular parts coincide. That is

$$n\rho_n\mathbb{1}_{B_{R_n}} = n\mathbb{1}_{[0,R_n]}(r)\rho_n(r)r^{d-1}dr \otimes \mathbb{1}_{\partial B_1}(y)d\mathcal{H}^{d-1}(y),$$

and, since $\mathbb{E}[n(B_{r_1})] = n\rho_n(B_{r_1})$ and $\mathbb{E}[n(A_{r_1,R_n})] = n\rho_n(A_{r_1,R_n})$, one has

$$\frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}} + \frac{n(A_{r_1,R_n})}{\rho_n(A_{r_1,R_n})}\rho_n\mathbb{1}_{A_{r_1,R_n}} = n\left[\frac{n(B_{r_1})}{\mathbb{E}[n(B_{r_1})]}\mathbb{1}_{[0,r_1]}(r) + \frac{n(A_{r_1,R_n})}{\mathbb{E}[n(A_{r_1,R_n})]}\mathbb{1}_{(r_1,R_n]}(r)\right]\rho_n(r)r^{d-1}dr$$
$$\otimes \mathbb{1}_{\partial B_1}(y)d\mathcal{H}^{d-1}(y),$$

thus, by (2.4.0.2) of Lemma 2.4.1

$$\mathbb{E}\left[W_p^p\left(n\rho_n\mathbb{1}_{B_{R_n}}, \frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}} + \frac{n(A_{r_1,R_n})}{\rho_n(A_{r_1,R_n})}\rho_n\mathbb{1}_{A_{r_1,R_n}}\right)\right]$$
$$\le n\mathcal{H}^{d-1}(\partial B_1)\mathbb{E}\left[W_p^p\left(\mathbb{1}_{[0,R_n]}(r)\rho_n(r)r^{d-1}, \left[\frac{n(B_{r_1})}{\mathbb{E}[n(B_{r_1})]}\mathbb{1}_{[0,r_1]}(r) + \frac{n(A_{r_1,R_n})}{\mathbb{E}[n(A_{r_1,R_n})]}\mathbb{1}_{(r_1,R_n]}(r)\right]\rho_n(r)r^{d-1}\right)\right]. \tag{6.2.2.16}$$

To estimate (6.2.2.16) we use the Benamou-Brenier formula by Proposition 2.3.1. That is, if we find a function $f : [0, R_n] \to \mathbb{R}$ such that its second derivative f'' is equal to the difference between the two measures which we consider, we can estimate

$$(6.2.2.16) \le n\mathcal{H}^{d-1}(\partial B_1)\mathbb{E}\left[\int_0^{R_n} dr \frac{|f'(r)|^p}{(\rho_n(r)r^{d-1})^{p-1}}\right] = n\mathcal{H}^{d-1}(\partial B_1)\int_0^{R_n} dr\frac{\mathbb{E}[|f'(r)|^p]}{(\rho_n(r)r^{d-1})^{p-1}}. \tag{6.2.2.17}$$

Hence we look for f such that

$$f''(r) = \rho_n(r)r^{d-1}\left[\frac{n(B_{r_1})}{\mathbb{E}[n(B_{r_1})]}\mathbb{1}_{[0,r_1]}(r) + \frac{n(A_{r_1,R_n})}{\mathbb{E}[n(A_{r_1,R_n})]}\mathbb{1}_{(r_1,R_n]}(r) - 1\right], \quad f'(0) = f'(R_n) = 0,$$

that is, since

$$n(A_{r_1,R_n}) - \mathbb{E}[n(A_{r_1,R_n})] = -n(B_{r_1}) + \mathbb{E}[n(B_{r_1})],$$

we have

$$f'(r) = \frac{n(B_{r_1}) - \mathbb{E}[n(B_{r_1})]}{n\mathcal{H}^{d-1}(\partial B_1)}\begin{cases} \rho_n(B_r)(\rho_n(B_{r_1}))^{-1} \text{ if } r \in [0, r_1] \\ \rho_n(A_{r,R_n})(\rho_n(A_{r_1,R_n}))^{-1} \text{ if } r \in [r_1, R_n] \end{cases} \tag{6.2.2.18}$$

We can use the previous explicit expression of f' in (6.2.2.18) to prove that, for a constant $C > 0$ depending on q, d, and eventually on r_1 (which is fixed), it holds

$$|f'(r)| \le C\frac{|n(B_{r_1}) - \mathbb{E}[n(B_{r_1})]|}{n}\rho_n(r)r^d\begin{cases} 1 & r \le r_1, \\ r^{-q} & r \ge r_1, \end{cases} \tag{6.2.2.19}$$

where the second estimate is a consequence of (6.1.0.10).

Plugging the estimate (6.2.2.19) into (6.2.2.17), in the case of the exponential densities ($q>0$), we get

$$\begin{aligned}(6.2.2.16) &\le Cn\frac{\mathbb{E}[|n(B_{r_1})-\mathbb{E}[n(B_{r_1})]|^p]}{n^p}\left[\int_0^{r_1}dr\frac{(\rho_n(r)r^d)^p}{(\rho_n(r)r^{d-1})^{p-1}}+\int_{r_1}^{R_n}dr\frac{(\rho_n(r)r^{d-q})^p}{(\rho_n(r)r^{d-1})^{p-1}}\right]\\ &\le Cn\mathbb{E}[n(B_{r_1})]^{-\frac{p}{2}}\left[1+\int_{r_1}^{R_n}dr\rho_n(r)r^{d-1-p(q-1)}\right]\\ &\le Cn^{1-\frac{p}{2}},\end{aligned}$$

and also in the case of the γ polynomial densities ($q=0$) since it has the same properties as the exponential case with $q=0$ hence, in this case, we get

$$(6.2.2.16)\le Cn\mathbb{E}[n(B_{r_1})]^{-\frac{p}{2}}\left[1+\int_{r_1}^{R_n}dr\rho_n(r)r^{d-1+p}\right]\le Cn^{1-\frac{p}{2}},$$

where we have used that $\rho_n(r)r^{d-1+p}\simeq r^{d+p-1-\gamma}$ is integrable.

Proof of (6.2.2.15). To estimate this term, we first claim that we can restrict to the very suitable set

$$\mathcal{A}:=\left\{n(A^\Sigma_{r_1,R_n})\ge\frac{\mathbb{E}[n(A^\Sigma_{r_1,R_n})]}{2}\right\},$$

indeed using Proposition 3.1.1 one has

$$\mathbb{P}(\mathcal{A}^c)\le\exp\left\{-c\mathbb{E}[n(A^\Sigma_{r_1,R_n})]\right\}=\exp\left\{-cn(1-\rho_n(B_{r_1}))\frac{\mathcal{H}^{d-1}(\Sigma)}{\mathcal{H}^{d-1}(\partial B_1)}\right\},$$

which, even multiplied for the worst estimate we can have for the Wasserstein distance of the measure involved, that is

$$W_p^p\left(\frac{n(A^\Sigma_{r_1,R_n})}{\rho_n(A^\Sigma_{r_1,R_n})}\rho_n\mathbb{1}_{A^\Sigma_{r_1,R_n}},\sum_{k\ge1}\frac{n(A^\Sigma_{r_k,r_{k+1}})}{\rho_n(A^\Sigma_{r_k,r_{k+1}})}\rho_n\mathbb{1}_{A^\Sigma_{r_k,r_{k+1}}}\right)\le CnR_n^p,$$

is much smaller than the right hand side in the thesis.

Then, one has reduced the proof of (6.2.2.15) to the estimate of the quantity

$$\mathbb{E}\left[\mathbb{1}_\mathcal{A}W_p^p\left(\frac{n(A^\Sigma_{r_1,R_n})}{\rho_n(A^\Sigma_{r_1,R_n})}\rho_n\mathbb{1}_{A^\Sigma_{r_1,R_n}},\sum_{k\ge1}\frac{n(A^\Sigma_{r_k,r_{k+1}})}{\rho_n(A^\Sigma_{r_k,r_{k+1}})}\rho_n\mathbb{1}_{A^\Sigma_{r_k,r_{k+1}}}\right)\right].\tag{6.2.2.20}$$

To deal with it, one can repeat the same argument as before to infer that, if $f:[r_1,R_n]\to\mathbb{R}$ solves

$$f''(r)=\rho_n(r)r^{d-1}\sum_{k\ge1}\mathbb{1}_{[r_k,r_{k+1})}(r)\left[\frac{n(A^\Sigma_{r_1,R_n})}{\mathbb{E}[n(A^\Sigma_{r_1,R_n})]}-\frac{n(A^\Sigma_{r_k,r_{k+1}})}{\mathbb{E}[n(A^\Sigma_{r_k,r_{k+1}})]}\right],\quad f'(r_1)=f'(R_n)=0,\tag{6.2.2.21}$$

then

$$\begin{aligned}(6.2.2.20)&\le n\mathcal{H}^{d-1}(\Sigma)\mathbb{E}\left[\mathbb{1}_\mathcal{A}\int_{r_1}^{R_n}dr\frac{|f'(r)|^p}{(\frac{n(A^\Sigma_{r_1,R_n})}{\mathbb{E}[n(A^\Sigma_{r_1,R_n})]}\rho_n(r)r^{d-1})^{p-1}}\right]\\ &\le 2^{p-1}n\mathcal{H}^{d-1}(\Sigma)\int_{r_1}^{R_n}dr\frac{\mathbb{E}[|f'(r)|^p]}{(\rho_n(r)r^{d-1})^{p-1}},\end{aligned}\tag{6.2.2.22}$$

where in the last inequality one has first used that in $\mathcal{A}$ it holds $\frac{n(A^\Sigma_{r_1,R_n})}{\mathbb{E}[n(A^\Sigma_{r_1,R_n})]}\ge\frac{1}{2}$, and then removed the restriction to $\mathcal{A}$. Then, to estimate (6.2.2.22), we only have to bound the quantity $\mathbb{E}[|f'(r)|^p]$. To do that, notice that by (6.2.2.21) one gets an explicit expression of f', that is

$$\begin{aligned}f'(r)=\sum_{k\ge1}\mathbb{1}_{[r_k,r_{k+1})}(r)&\left[\frac{n(A^\Sigma_{r_k,R_n})-\mathbb{E}[n(A^\Sigma_{r_k,R_n})]}{n\mathcal{H}^{d-1}(\Sigma)}\frac{\rho_n(A^\Sigma_{r_1,r})}{\rho_n(A^\Sigma_{r_1,R_n})}-\frac{n(A^\Sigma_{r_k,r_{k+1}})-\mathbb{E}[n(A^\Sigma_{r_k,r_{k+1}})]}{n\mathcal{H}^{d-1}(\Sigma)}\frac{\rho_n(A^\Sigma_{r_k,r})}{\rho_n(A^\Sigma_{r_k,r_{k+1}})}\right.\\ &\left.-\frac{n(A^\Sigma_{r_1,r_k})-\mathbb{E}[n(A^\Sigma_{r_1,r_k})]}{n\mathcal{H}^{d-1}(\Sigma)}\frac{\rho_n(A^\Sigma_{r,R_n})}{\rho_n(A^\Sigma_{r_1,R_n})}\right],\end{aligned}$$

which is enough to ensure that for $r \in [r_k, r_{k+1})$

$$\begin{aligned}
\mathbb{E}[|f'(r)|^p] &\le C\frac{\mathbb{E}[|n(A^{\Sigma}_{r_k,R_n}) - \mathbb{E}[n(A^{\Sigma}_{r_k,R_n})]|^p] + \mathbb{E}[|n(A^{\Sigma}_{r_k,r_{k+1}}) - \mathbb{E}[n(A^{\Sigma}_{r_k,r_{k+1}})]|^p]}{(n\mathcal{H}^{d-1}(\Sigma))^p} \\
&+ \frac{\mathbb{E}[|n(A^{\Sigma}_{r_1,r_k}) - \mathbb{E}[n(A^{\Sigma}_{r_1,r_k})]|^p(\rho_n(A^{\Sigma}_{r,R_n}))^p]}{(n\mathcal{H}^{d-1}(\Sigma))^p} \\
&\le C\frac{\mathbb{E}[n(A^{\Sigma}_{r_k,R_n})]^{\frac{p}{2}} + \mathbb{E}[n(A^{\Sigma}_{r_k,r_{k+1}})]^{\frac{p}{2}} + \mathbb{E}[n(A^{\Sigma}_{r_1,r_k})]^{\frac{p}{2}}\left(\rho_n(A^{\Sigma}_{r,R_n})\right)^p}{(n\mathcal{H}^{d-1}(\Sigma))^p} \\
&\le C'\frac{\mathbb{E}[n(A^{\Sigma}_{r_k,R_n})]^{\frac{p}{2}} + n^{\frac{p}{2}}\left(\rho_n(A^{\Sigma}_{r,R_n})\right)^p}{(n\mathcal{H}^{d-1}(\Sigma))^p}.
\end{aligned} \tag{6.2.2.23}$$

Notice that in the second inequality we have used Lemma 3.1.1 making sure that

$$\mathbb{E}[n(A^{\Sigma}_{r_k,R_n})] \ge c = c(\varepsilon),$$

and therefore the previous constant C actually depends on ε.

Let us focus one moment on the first term in (6.2.2.23): for $r \in [r_k, r_{k+1})$ one has

$$\begin{aligned}
\mathbb{E}[n(A^{\Sigma}_{r_k,R_n})] &= n\mathcal{H}^{d-1}(\Sigma)\int_{r_k}^{R_n} ds s^{d-1}\rho_n(s) \\
&\le Cn\mathcal{H}^{d-1}(\Sigma)\rho(r_k)r_k^{d-q} \quad \text{by property (6.1.0.10)} \\
&\le Cn\mathcal{H}^{d-1}(\Sigma)\rho(r)r^{d-q},
\end{aligned} \tag{6.2.2.24}$$

where in the last inequality we have used the property that, for any $r_k \le r \le s \le r_{k+1}$, we have

$$\frac{1}{C} \le \frac{\rho_n(r)}{\rho_n(s)} \le C, \quad \text{and} \quad \frac{1}{C} \le \frac{r}{s} \le C.$$

Again, the case of the polynomial density corresponds to $q = 0$.

Similarly, for the second summand in (6.2.2.23), one has

$$\rho_n(A^{\Sigma}_{r,R_n}) = \mathcal{H}^{d-1}(\Sigma)\int_r^{R_n} ds s^{d-1}\rho_n(s) \le C\mathcal{H}^{d-1}(\Sigma)\rho(r)r^{d-q} \quad \text{by property (6.1.0.10)}. \tag{6.2.2.25}$$

Hence, combining (6.2.2.24) and (6.2.2.25) with (6.2.2.23), one gets

$$\mathbb{E}[|f'(r)|^p] \le C\left(\frac{\rho(r)r^{d-q}}{n\mathcal{H}^{d-1}(\Sigma)}\right)^{\frac{p}{2}},$$

which, if combined with (6.2.2.22), gives

$$(6.2.2.20) \le C(\varepsilon)(n\mathcal{H}^{d-1}(\Sigma))^{1-\frac{p}{2}}\int_{r_1}^{R_n} dr\rho(r)^{1-\frac{p}{2}}r^{d-1-p\frac{d+q-2}{2}},$$

that is, the second statement (6.2.2.15). □

Now we want to focus on the angular estimate, that is, partitioning an annulus $A_{r,r(1+\delta)}$ in the subsets $Q^{y,U}_{r,\delta}$ we have mentioned before in (6.2.1.1) in Definition 6.2.1. To do this, we need to define the kernel K.

Definition 6.2.4. *The kernel $K : \partial B_1 \times \partial B_1 \to [0, +\infty)$ is defined as*

$$K(x,z) := \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \mathbb{1}_{\Delta^{y,U}_{\delta}\times\Delta^{y,U}_{\delta}}(x,z),$$

with $\Delta^{y,U}_{\delta}$ as in (6.2.1.2) in Definition 6.2.1.

Intuitively, for any couple $(x,z) \in \partial B_1 \times \partial B_1$, $K(x,z)$ counts the number of $\Delta^{y,U}_{\delta}$ which they both belong to. Of course it does not "count" in a proper sense, it integrates, but the spirit of the definition is the same. Notice that

$$\begin{aligned}
\int_{\partial B_1} d\mathcal{H}^{d-1}(x)K(x,z) &= \int_{\partial B_1} d\mathcal{H}^{d-1}(x)\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \mathbb{1}_{\Delta^{y,U}_{\delta}}(x)\mathbb{1}_{\Delta^{y,U}_{\delta}}(z) \\
&= \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \mathbb{1}_{\Delta^{y,U}_{\delta}}(z)\int_{\partial B_1} d\mathcal{H}^{d-1}(x)\mathbb{1}_{\Delta^{y,U}_{\delta}}(x) \\
&= \mathcal{M}_{\delta}\mathcal{H}^{d-1}(\Delta_{\delta}) \text{ for any } z \in \partial B_1,
\end{aligned} \tag{6.2.2.26}$$

where, in the last equality, we have omitted (y,U) from the notation since $\mathcal{H}^{d-1}(\Delta^{y,U}_{\delta})$ does not depend on it.

The definition of K extends to measures on ∂B_1 as follows.

Definition 6.2.5. *Given a measure μ on ∂B_1, $K(\mu)$ is the measure such that, for any $f : \partial B_1 \to \mathbb{R}$*

$$\int_{\partial B_1} dK(\mu)(x) f(x) = \int_{\partial B_1} d\mathcal{H}^{d-1}(x) f(x) \int_{\partial B_1} d\mu(y) K(x,y).$$

We begin the angular estimate by noticing that K has the following property on measures on ∂B_1.

Lemma 6.2.4. *Let μ and ν be two measures on ∂B_1 and $p \geq 1$. Then*

$$W_p^p\left(K(\mu), K(\nu)\right) \leq \mathcal{M}_\delta \mathcal{H}^{d-1}(\Delta_\delta) W_p^p(\mu,\nu).$$

We do not want to prove immediately the previous Lemma since we want to see how we can use it to prove the following result.

Lemma 6.2.5 (The angular estimate)**.** *Let $\{\tilde{X}_i\}_{i=1}^n$ be a sequence of iid random variables with common probability density ρ_n as in Definition 6.1.3, and $n(\cdot)$ the number of variables in a domain as in Notation 6.2.2. Then, there exists a constant $C > 0$ such that if $A_{r,s}^{\Delta_\delta^{y,U}}$ is the pyramid in (6.2.1.2) in Definition 6.2.1 and 6.2.2, it holds*

$$\mathbb{E}\left[W_p^p\left(\frac{n(A_{r,s})}{\rho_n(A_{r,s})}\rho_n \mathbb{1}_{A_{r,s}}, \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \frac{n(A_{r,s}^{\Delta_\delta^{y,U}})}{\rho_n(A_{r,s}^{\Delta_\delta^{y,U}})\mathcal{M}_\delta}\rho_n \mathbb{1}_{A_{r,s}^{\Delta_\delta^{y,U}}}\right)\right] \leq C\frac{\int_r^s du \rho_n(u) u^{p+d-1}}{\int_r^s du\rho_n(u)u^{d-1}}\tau_{d-1}^p(\mathbb{E}[n(A_{r,s})]).$$

Proof. We begin this proof in a way which is very similar to the radial estimate. Let us recall that, by (6.2.1.2) in Definition 6.2.1, we have

$$\Delta_\delta^{y,U} = \exp_y(R^y(U((-\delta/2, \delta/2)^{d-1}) \times \{0\})) \subseteq \partial B_1.$$

The two measures involved have the same radial component and only their angular part is different. In other words, the first measure can be written as

$$\frac{n(A_{r,s})}{\rho_n(A_{r,s})}\mathbb{1}_{\partial B_1}(z) d\mathcal{H}^{d-1}(z) \otimes \rho_n(u) u^{d-1}\mathbb{1}_{[r,s)}(u) du,$$

and the second one as

$$\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \frac{n(A_{r,s}^{\Delta_\delta^{y,U}})}{\rho_n(A_{r,s}^{\Delta_\delta^{y,U}})\mathcal{M}_\delta}\mathbb{1}_{\Delta_\delta^{y,U}}(z) d\mathcal{H}^{d-1}(z) \otimes \rho_n(u)u^{d-1}\mathbb{1}_{[r,s)}(u)du.$$

Therefore by (2.4.0.3) of Lemma 2.4.1, one has

$$\begin{aligned}
&\mathbb{E}\left[W_p^p\left(\frac{n(A_{r,s})}{\rho_n(A_{r,s})}\rho_n \mathbb{1}_{A_{r,s}}, \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \frac{n(A_{r,s}^{\Delta_\delta^{y,U}})}{\rho_n(A_{r,s}^{\Delta_\delta^{y,U}})\mathcal{M}_\delta}\rho_n \mathbb{1}_{A_{r,s}^{\Delta_\delta^{y,U}}}\right)\right] \\
&\leq \int_r^s du\rho_n(u)u^{p+d-1}\mathbb{E}\left[W_p^p\left(\frac{n(A_{r,s})}{\rho_n(A_{r,s})}\mathbb{1}_{\partial B_1}, \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \frac{n(A_{r,s}^{\Delta_\delta^{y,U}})}{\rho_n(A_{r,s}^{\Delta_\delta^{y,U}})\mathcal{M}_\delta}\mathbb{1}_{\Delta_\delta^{y,U}}\right)\right],
\end{aligned} \tag{6.2.2.27}$$

therefore we can conclude the proof if we estimate (6.2.2.27), which is our next purpose.

To do that, let us write better the two measures in (6.2.2.27) through the kernel K in Definition 6.2.4. In fact, we claim that for any $z \in \partial B_1$

$$\frac{n(A_{r,s})}{\rho_n(A_{r,s})}\mathbb{1}_{\partial B_1}(z) = \frac{1}{\rho_n(A_{r,s}^{\Delta_\delta^{y,U}})\mathcal{M}_\delta}\frac{n(A_{r,s})}{\mathcal{H}^{d-1}(\partial B_1)}\int_{\partial B_1} d\mathcal{H}^{d-1}(x) K(x,z), \tag{6.2.2.28}$$

and, if

$$\hat{X}_i := \frac{\tilde{X}_i}{|\tilde{X}_i|},$$

also

$$\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \frac{n(A_{r,s}^{\Delta_\delta^{y,U}})}{\rho_n(A_{r,s}^{\Delta_\delta^{y,U}})\mathcal{M}_\delta}\mathbb{1}_{\Delta_\delta^{y,U}}(z) = \frac{1}{\rho_n(A_{r,s}^{\Delta_\delta^{\bar{y},\bar{U}}})\mathcal{M}_\delta}\sum_{i:\tilde{X}_i \in A_{r,s}} K(\hat{X}_i, z). \tag{6.2.2.29}$$

Let us postpone both the proofs of (6.2.2.28) and (6.2.2.29), and let us see how we can use them to conclude: by combining them with (6.2.2.27) we get

$$\begin{aligned}
&\mathbb{E}\left[W_p^p\left(\frac{n(A_{r,s})}{\rho_n(A_{r,s})}\rho_n\mathbb{1}_{A_{r,s}}, \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\,\frac{n(A_{r,s}^{\Delta_\delta^{y,U}})}{\rho_n(A_{r,s}^{\Delta_\delta^{y,U}})\mathcal{M}_\delta}\rho_n\mathbb{1}_{A_{r,s}^{\Delta_\delta^{y,U}}}\right)\right]\\
&\leq \frac{\int_r^s du\rho_n(u)u^{p+d-1}}{\rho_n(A_{r,s}^{\Delta_\delta^{\bar y,\bar U}})\mathcal{M}_\delta}\mathbb{E}\left[W_p^p\left(\frac{n(A_{r,s})}{\mathcal{H}^{d-1}(\partial B_1)}\int_{\partial B_1} d\mathcal{H}^{d-1}(x)K(x,\cdot), \sum_{i:\tilde X_i\in A_{r,s}} K(\hat X_i,\cdot)\right)\right]\\
&= \frac{\int_r^s du\rho_n(u)u^{p+d-1}}{\rho_n(A_{r,s}^{\Delta_\delta^{\bar y,\bar U}})\mathcal{M}_\delta}\mathbb{E}\left[W_p^p\left(K\left(\frac{n(A_{r,s})}{\mathcal{H}^{d-1}(\partial B_1)}\mathbb{1}_{\partial B_1}\right), K\left(\sum_{i:\tilde X_i\in A_{r,s}}\delta_{\hat X_i}\right)\right)\right] \quad \text{by Definition 6.2.4}\\
&\leq \frac{\int_r^s du\rho_n(u)u^{p+d-1}}{\int_r^s du\rho_n(u)u^{d-1}}\mathbb{E}\left[W_p^p\left(\frac{n(A_{r,s})}{\mathcal{H}^{d-1}(\partial B_1)}\mathbb{1}_{\partial B_1}, \sum_{i:\tilde X_i\in A_{r,s}}\delta_{\hat X_i}\right)\right], \quad \text{by Lemma 6.2.4.}
\end{aligned}$$

Hence, by (1.1.1.2), we get the thesis since, conditioned to $n(A_{r,s})$, the random variables $\{\hat X_i : \tilde X_i \in A_{r,s}\}$ are uniformly distributed on ∂B_1.

Proof of (6.2.2.28) and (6.2.2.29). We begin with the proof of (6.2.2.28) (the uniform measure): by noticing that

$$\rho_n(A_{r,s}^{\Delta_\delta^{\bar y,\bar U}}) = \rho_n(A_{r,s})\frac{\mathcal{H}^{d-1}(\Delta_\delta)}{\mathcal{H}^{d-1}(\partial B_1)},$$

and then diving by $\mathcal{M}_\delta\mathcal{H}^{d-1}(\Delta_\delta)$ and multiplying by it again (see (6.2.2.26)), we have

$$\frac{n(A_{r,s})}{\rho_n(A_{r,s})}\mathbb{1}_{\partial B_1}(z) = \frac{1}{\rho_n(A_{r,s}^{\Delta_\delta^{\bar y,\bar U}})\mathcal{M}_\delta}\frac{n(A_{r,s})}{\mathcal{H}^{d-1}(\partial B_1)}\int_{\partial B_1} d\mathcal{H}^{d-1}(x)K(x,z)\mathbb{1}_{\partial B_1}(z),$$

and this proves (6.2.2.28).

Now we focus on the proof of (6.2.2.29) (the uniform measure with rearranged masses): by writing explicitly

$$n(A_{r,s}^{\Delta_\delta^{y,U}}) = \sum_{i:\tilde X_i\in A_{r,s}} \mathbb{1}_{\Delta_\delta^{y,U}}(\hat X_i),$$

we get

$$\begin{aligned}
\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\,\frac{n(A_{r,s}^{\Delta_\delta^{y,U}})}{\rho_n(A_{r,s}^{\Delta_\delta^{y,U}})\mathcal{M}_\delta}\mathbb{1}_{\Delta_\delta^{y,U}}(z) &= \frac{1}{\rho_n(A_{r,s}^{\Delta_\delta^{\bar y,\bar U}})\mathcal{M}_\delta}\sum_{i:\tilde X_i\in A_{r,s}}\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\mathbb{1}_{\Delta_\delta^{y,U}\times\Delta_\delta^{y,U}}(\hat X_i,z)\\
&= \frac{1}{\rho_n(A_{r,s}^{\Delta_\delta^{\bar y,\bar U}})\mathcal{M}_\delta}\sum_{i:\tilde X_i\in A_{r,s}} K(\hat X_i,z),
\end{aligned}$$

where the second equality comes from Definition 6.2.4 of K. This proves also (6.2.2.29). ☐

Now we focus on the proof of Lemma 6.2.4 which we left aside before.

Proof of Lemma 6.2.4. To prove this, let π be any coupling of μ ad ν. We want to find another coupling $\tilde\pi$ of $K(\mu)$ and $K(\nu)$ such that

$$\int_{\partial B_1\times\partial B_1} d\tilde\pi(x,y)d_{\partial B_1}(x,y)^p \leq \mathcal{M}_\delta\mathcal{H}^{d-1}(\Delta_\delta)\int_{\partial B_1\times\partial B_1} d\pi(x,y)d_{\partial B_1}(x,y)^p. \tag{6.2.2.30}$$

If we assume of having already found $\tilde\pi$ satisfying (6.2.2.30), we can conclude since this implies

$$W_p^p(K(\mu),K(\nu)) \leq \int_{\partial B_1\times\partial B_1} d\tilde\pi(x,y)d_{\partial B_1}(x,y)^p \leq \mathcal{M}_\delta\mathcal{H}^{d-1}(\Delta_\delta)\int_{\partial B_1\times\partial B_1} d\pi(x,y)d_{\partial B_1}(x,y)^p,$$

which leads to the thesis by minimizing among all possible couplings π of μ and ν. Therefore let us prove that there exists another coupling $\tilde\pi$ of $K(\mu)$ and $K(\nu)$ which verifies (6.2.2.30).

We claim that the coupling

$$\tilde\pi(w,z) := \int_{\partial B_1\times\partial B_1} d\pi(x,y)K(w,x)\delta(z-R_x^y(w)),$$

verifies such property, where $\delta(z-R_x^y(w))$ means that only the (w,z) such that $z = R_x^y(w)$ have weight, that is

$$\int_{\partial B_1\times\partial B_1} d\tilde\pi(w,z)f(w,z) = \int_{\partial B_1} d\mathcal{H}^{d-1}(w)\int_{\partial B_1\times\partial B_1} d\pi(x,y)K(w,x)f(w,R_x^y(w)).$$

Step 1: prove that $\tilde\pi$ is a coupling of $K(\mu)$ and $K(\nu)$. The first marginal of $\tilde\pi$ is $K(\mu)$ since

$$\begin{aligned}\int_{\partial B_1\times\partial B_1} d\tilde\pi(w,z)f(w) &= \int_{\partial B_1} d\mathcal{H}^{d-1}(w)\int_{\partial B_1\times\partial B_1} d\pi(x,y)K(w,x)f(w)\\ &= \int_{\partial B_1} d\mathcal{H}^{d-1}(w)f(w)\int_{\partial B_1} d\mu(x)K(w,x) \text{ since the first marginal of } \pi \text{ is } \mu\\ &= \int_{\partial B_1} dK(\mu)(w)f(w) \text{ by Definition 6.2.5.}\end{aligned}$$

To prove that the second marginal of $\tilde\pi$ is $K(\nu)$ is a little bit more difficult. Let us denote by R_x^y the rotation moving x in y defined in Subsection 4.1.1. Assuming that we know that for any (z,w) and (x,y) it holds

$$K(R_y^x(z),w) = K(z,R_x^y(w)), \tag{6.2.2.31}$$

which we shall prove later, it holds

$$\begin{aligned}\int_{\partial B_1\times\partial B_1} d\tilde\pi(w,z)f(z) &= \int_{\partial B_1} d\mathcal{H}^{d-1}(w)\int_{\partial B_1\times\partial B_1} d\pi(x,y)K(w,x)f(R_x^y(w))\\ &= \int_{\partial B_1} d\mathcal{H}^{d-1}(z)\int_{\partial B_1\times\partial B_1} d\pi(x,y)K(R_y^x(z),x)f(z) \text{ by changing variables } z=R_x^y(w)\\ &= \int_{\partial B_1} d\mathcal{H}^{d-1}(z)\int_{\partial B_1\times\partial B_1} d\pi(x,y)K(z,y)f(z) \text{ by (6.2.2.31) and } R_x^y(x)=y\\ &= \int_{\partial B_1} d\mathcal{H}^{d-1}(z)\int_{\partial B_1\times\partial B_1} d\nu(y)K(z,y)f(z) \text{ since the second marginal of } \pi \text{ is } \nu\\ &= \int_{\partial B_1} dK(\nu)(z)f(z) \text{ by Definition 6.2.5.}\end{aligned}$$

Therefore to conclude this first step we only have to prove (6.2.2.31).

We claim that there exists an operator $L(y')\in O(d-1)$ (depending on x, y, and y') such that

$$R_x^y(\Delta_\delta^{y',U}) = \Delta_\delta^{R_x^y(y'),L(y')U}. \tag{6.2.2.32}$$

To prove it, we use Lemmas 4.2.2 and 4.1.2 to argue that

$$\begin{aligned}R_x^y(\Delta_\delta^{y',U}) &= R_x^y(\exp_{y'}(R^{y'}(U((-\delta/2,\delta/2)^{d-1})\times\{0\}))) \text{ by definition}\\ &= \exp_{R_x^y(y')}(R_x^yR^{y'}(U((-\delta/2,\delta/2)^{d-1})\times\{0\})) \text{ by Lemma 4.2.2}\\ &= \exp_{R_x^y(y')}(R^{R_x^y(y')}(R^{R_x^y(y')})^TR_x^yR^{y'}(U((-\delta/2,\delta/2)^{d-1})\times\{0\}))\\ &= \exp_{R_x^y(y')}(R^{R_x^y(y')}L'(y')(U((-\delta/2,\delta/2)^{d-1})\times\{0\})),\end{aligned} \tag{6.2.2.33}$$

where $L'(y')$ is the orthonormal operator

$$L'(y') := (R^{R_x^y(y')})^TR_x^yR^{y'},$$

which, thanks to Lemma 4.1.2, fixes both e_d and $\{e_d\}^\perp$, and therefore it can be written as

$$L'(y') = \begin{pmatrix} L(y') & 0\\ 0 & 1\end{pmatrix}, \tag{6.2.2.34}$$

with

$$L(y')_{i,j} = e_i\cdot L(y')e_j = e_i\cdot(R^{R_x^y(y')})^TR_x^yR^{y'}e_j = R^{R_x^y(y')}e_i\cdot R_x^yR^{y'}e_j,\quad \forall i,j=1,\dots,d-1.$$

Hence, combining (6.2.2.33) and (6.2.2.34) one gets (6.2.2.32).

Now we go back to the proof of (6.2.2.31): one has

$$\begin{aligned}K(R_y^x(z),w) &= \int_{\partial B_1} d\mathcal{H}^{d-1}(y')\int_{O(d-1)} dU\,\mathbb{1}_{\Delta_\delta^{y',U}}(R_y^x(z))\mathbb{1}_{\Delta_\delta^{y',U}}(w)\\ &= \int_{\partial B_1} d\mathcal{H}^{d-1}(y')\int_{O(d-1)} dU\,\mathbb{1}_{R_x^y(\Delta_\delta^{y',U})}(z)\mathbb{1}_{\Delta_\delta^{y',U}}(w)\\ &= \int_{\partial B_1} d\mathcal{H}^{d-1}(y')\int_{O(d-1)} dU\,\mathbb{1}_{\Delta_\delta^{R_x^y(y'),L(y')U}}(z)\mathbb{1}_{\Delta_\delta^{y',U}}(w) \quad \text{by (6.2.2.32)}\\ &= \int_{\partial B_1} d\mathcal{H}^{d-1}(y'')\int_{O(d-1)} dU'\,\mathbb{1}_{\Delta_\delta^{y'',U'}}(z)\mathbb{1}_{\Delta_\delta^{R_y^x(y''),(L(R_y^x(y'')))^{-1}U'}}(w),\end{aligned}$$

where, in the last passage, we have used the unitary change of variables

$$y'' = R_x^y(y'), \qquad U' = L(y')U.$$

Now in the last equality one can write better the integrand function: indeed one has

$$\Delta_\delta^{R_y^x(y''),(L(R_y^x(y'')))^{-1}U'} = \Delta_\delta^{R_y^x(y''),M(y'')(M(y''))^{-1}(L(R_y^x(y'')))^{-1}U'} = R_y^x(\Delta_\delta^{y'',(M(y''))^{-1}(L(R_y^x(y'')))^{-1}U'}),$$

if $M(y'')$ is the submatrix of $M'(y'')$ as in (6.2.2.34), and $M'(y'')$ is the matrix obtained as in (6.2.2.32) and (6.2.2.33) with R_y^x instead of R_x^y and y'' instead of y'. Now we claim that

$$L'(R_y^x(y''))M'(y'') = \mathrm{Id}, \text{ so that } L(R_y^x(y''))M(y'') = \mathrm{Id}. \tag{6.2.2.35}$$

If we postpone the proof of (6.2.2.35) and combine it with the previous equality, we get

$$\begin{aligned} K(R_y^x(z), w) &= \int_{\partial B_1} d\mathcal{H}^{d-1}(y'') \int_{O(d-1)} dU' \mathbb{1}_{\Delta_\delta^{y'',U'}}(z) \mathbb{1}_{R_y^x(\Delta_\delta^{y'',U'})}(w) \\ &= \int_{\partial B_1} d\mathcal{H}^{d-1}(y'') \int_{O(d-1)} dU' \mathbb{1}_{\Delta_\delta^{y'',U'}}(z) \mathbb{1}_{\Delta_\delta^{y'',U'}}(R_x^y(w)) \\ &= K(z, R_x^y(w)), \end{aligned}$$

which proves (6.2.2.31).

Hence, we are left with the proof of (6.2.2.35): we have

$$\begin{aligned} L'(R_y^x(y''))M'(y'') &= (R^{R_x^y(R_y^x(y''))})^T R_x^y R^{R_y^x(y'')} (R^{R_y^x(y'')})^T R_y^x R^{y''} \\ &= (R^{R_x^y(R_y^x(y''))})^T R_x^y R_y^x R^{y''} \quad \text{ since } \quad (R^{R_y^x(y'')})^T R^{R_y^x(y'')} = \mathrm{Id} \\ &= (R^{R_x^y(R_y^x(y''))})^T R^{y''} \quad \text{ since } \quad R_x^y R_y^x = \mathrm{Id} \\ &= (R^{y''})^T R^{y''} \quad \text{ since } \quad R_x^y(R_y^x(y'')) = y'' \\ &= \mathrm{Id}, \end{aligned}$$

that is, (6.2.2.35).

Hence $\tilde{\pi}$ is a coupling of $K(\mu)$ and $K(\nu)$.

Step 2: prove that $\tilde{\pi}$ satisfies the inequality in (6.2.2.30). By definition of $\tilde{\pi}$, we have

$$\begin{aligned} \int_{\partial B_1 \times \partial B_1} d\tilde{\pi}(w,z) d(w,z)^p &= \int_{\partial B_1} d\mathcal{H}^{d-1}(w) \int_{\partial B_1 \times \partial B_1} d\pi(x,y) K(w,x) d_{\partial B_1}(w, R_x^y(w))^p \quad \text{ by Definition of } \tilde{\pi} \\ &\le \int_{\partial B_1} d\mathcal{H}^{d-1}(w) \int_{\partial B_1 \times \partial B_1} d\pi(x,y) K(w,x) d_{\partial B_1}(x,y)^p \quad \text{ by Proposition 4.1.2} \\ &= \int_{\partial B_1 \times \partial B_1} d\pi(x,y) d_{\partial B_1}(x,y)^p \int_{\partial B_1} d\mathcal{H}^{d-1}(w) K(w,x) \\ &= \mathcal{M}_\delta \mathcal{H}^{d-1}(\Delta_\delta) \int_{\partial B_1 \times \partial B_1} d\pi(x,y) d_{\partial B_1}(x,y)^p, \end{aligned}$$

and this concludes the proof. □

Now that we have proved these two Lemmas, we can estimate the expectation of (6.2.2.11).

Proposition 6.2.1 (The global estimate with decomposition radial-angular (6.2.2.9).)**.** *Let $\{\tilde{X}_i\}_{i=1}^n$ be iid random variables in $\mathbb{R}^d$ with common probability density ρ_n as in Definition 6.1.3. Denoting by $n(\cdot)$ the number of particles in a set as in Notation 6.2.2, let $g(n)$ be the quantity*

$$g(n) := \mathbb{E}\left[W_p^p \left(n\rho_n \mathbb{1}_{B_{R_n}}, \frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n \mathbb{1}_{B_{r_1}} + \sum_{k \ge 1} \int_{\partial B_1} d\mathcal{H}^{d-1}(y) \int_{O(d-1)} dU \frac{n(Q_{r_k,\delta_k}^{y,U})}{\mathcal{M}_{\delta_k}\rho_n(Q_{r_k,\delta_k}^{y,U})} \rho_n \mathbb{1}_{Q_{r_k,\delta_k}^{y,U}} \right) \right],$$

(depending also on p, d, q, γ and r_1). Then, there exist a positive constant $c > 0$ and another constant $C(\varepsilon)$ such that the following estimates hold: if $q > 0$

$$g(n) \le C(\varepsilon) \tau_n^{p,d} (\log n)^{-c(\alpha - \frac{\gamma}{q} - d(1-\frac{1}{q}))} \begin{cases} 1 & p \le d \ge 3 \text{ or } p < d = 2 \text{ or } p > d = 4, \\ (\log n)^{c(\alpha - \frac{\gamma}{q} - d(1-\frac{1}{q}))} & p = d = 2, \\ (\log \log n)^{\frac{p}{2}} & p > d = 3, \\ (\log n)^{\frac{p}{2}} & p > d = 2, \end{cases}$$

while if $q=0$

$$g(n)\le C(\varepsilon)\tau_n^{p,d}(\log n)^{-c\beta}\begin{cases}1 & p\le p^* \text{ or } p>p^*, d\neq 3,\\ (\log\log n)^{\frac{p}{2}} & p>p^*, d=3.\end{cases}$$

Before proving this estimate, let us stress that, thanks to the choices of α and β in Definition 6.1.1 the previous Proposition 6.2.1 basically states that, in a lot of cases, using the radial-angular decomposition one has

$$g(n)\begin{cases}\ll \tau_n^{p,d} & p\le p^{crit},\\ \lesssim \tau_n^{p,d} & p>p^{crit},\end{cases}, \tag{6.2.2.36}$$

where, "a lot of cases" means that the following two possibilities are excluded: $p>p^{crit}, d=3$ (for any q), and $p\ge 2=d, q>0$.

Proof. We begin with the proof of the first statement. We use the triangle inequality with the measure

$$\tilde{\nu}:=\sum_{k\ge 0}\frac{n(A_{r_k,r_{k+1}})}{\rho_n(A_{r_k,r_{k+1}})}\rho_n\mathbb{1}_{A_{r_k,r_{k+1}}},$$

and then apply the property

$$(A+B)^p\le 2^pA^p+2^pB^p,$$

to get

$$g(n)\le 2^p\mathbb{E}\left[W_p^p\left(n\rho_n\mathbb{1}_{B_{R_n}},\tilde{\nu}\right)\right] \tag{6.2.2.37}$$

$$+2^p\mathbb{E}\left[W_p^p\left(\tilde{\nu},\frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}}+\sum_{k\ge 1}\int_{\partial B_1}d\mathcal{H}^{d-1}(y)\int_{O(d-1)}dU\frac{n(Q^{y,U}_{r_k,\delta_k})}{\mathcal{M}_{\delta_k}\rho_n(Q^{y,U}_{r_k,\delta_k})}\rho_n\mathbb{1}_{Q^{y,U}_{r_k,\delta_k}}\right)\right]. \tag{6.2.2.38}$$

We estimate the first summand with Lemma 6.2.3, applying again triangle inequality with the measure

$$\frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}}+\frac{n(A_{r_1,R_n})}{\rho_n(A_{r_1,R_n})}\rho_n\mathbb{1}_{A_{r_1,R_n}},$$

to get

$$\begin{aligned}(6.2.2.37)\le{}& 2^p\mathbb{E}\left[W_p^p\left(n\rho_n\mathbb{1}_{B_{R_n}},\frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}}+\frac{n(A_{r_1,R_n})}{\rho_n(A_{r_1,R_n})}\rho_n\mathbb{1}_{A_{r_1,R_n}}\right)\right]\\ &+2^p\mathbb{E}\left[W_p^p\left(\frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}}+\frac{n(A_{r_1,R_n})}{\rho_n(A_{r_1,R_n})}\rho_n\mathbb{1}_{A_{r_1,R_n}},\sum_{k\ge 0}\frac{n(A_{r_k,r_{k+1}})}{\rho_n(A_{r_k,r_{k+1}})}\rho_n\mathbb{1}_{A_{r_k,r_{k+1}}}\right)\right],\end{aligned}$$

and therefore, using the subadditivity (2.1.0.2) in the second summand, we have

$$\begin{aligned}(6.2.2.37)\le{}& 2^p\mathbb{E}\left[W_p^p\left(n\rho_n\mathbb{1}_{B_{R_n}},\frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}}+\frac{n(A_{r_1,R_n})}{\rho_n(A_{r_1,R_n})}\rho_n\mathbb{1}_{A_{r_1,R_n}}\right)\right]\\ &+2^p\mathbb{E}\left[W_p^p\left(\frac{n(A_{r_1,R_n})}{\rho_n(A_{r_1,R_n})}\rho_n\mathbb{1}_{A_{r_1,R_n}},\sum_{k\ge 1}\frac{n(A_{r_k,r_{k+1}})}{\rho_n(A_{r_k,r_{k+1}})}\rho_n\mathbb{1}_{A_{r_k,r_{k+1}}}\right)\right].\end{aligned}$$

Now one can see that the first summand in the previous inequality is exactly (6.2.2.14) of Lemma 6.2.3 and that the second one is the term in the statement (6.2.2.15) of Lemma 6.2.3 with $\Sigma=\partial B_1$, and therefore one can also estimate

$$\begin{aligned}(6.2.2.37)&\le Cn^{1-\frac{p}{2}}+Cn^{1-\frac{p}{2}}\int_{r_1}^{R_n}dr\rho(r)^{1-\frac{p}{2}}r^{d-1-p\frac{d+q-2}{2}}\\ &\le Cn^{1-\frac{p}{2}}\int_{r_1}^{R_n}dr\rho(r)^{1-\frac{p}{2}}r^{d-1-p\frac{d+q-2}{2}}\\ &\le Cn^{1-\frac{p}{2}}\int_{r_1}^{R_n}dre^{-\frac{r^q}{q}(1-\frac{p}{2})}r^{d-\gamma+p\frac{\gamma+2-d-q}{2}-1}\\ &\le Cn^{1-\frac{p}{2}}\begin{cases}1 & q>0,p<2 \text{ or } q>2,p=2 \text{ or } q=0,p<\underline{p},\\ R_n^{2-q} & q\in(0,2),p=2,\\ \log R_n & q=p=2, \text{ or } q=0,p=\underline{p},\\ e^{-\frac{R_n^q}{q}(1-\frac{p}{2})}R_n^{d-\gamma+p\frac{\gamma+2-d-q}{2}-q} & q>0,p>2 \text{ or } q=0,p>\underline{p},\end{cases}\end{aligned}$$

with

$$\underline{p} = \frac{2(\gamma - d)}{\gamma + 2 - d}. \tag{6.2.2.39}$$

Notice that

$$p^* - \underline{p} = \frac{(\gamma - d)^2(d-2)}{\gamma(\gamma - d + 2)}, \text{ therefore } \underline{p} \le p^*, \text{ and } \underline{p} = p^* \text{ if and only if } d = 2. \tag{6.2.2.40}$$

Recalling the definition of the critical radius R_n in (6.1.0.6), thanks to the previous estimate, we can infer that if $q > 0$ one has

$$(6.2.2.37) \le C \begin{cases} n^{1-\frac{p}{2}} & p < 2 \text{ or } q > 2, p = 2, \\ (\log n)^{\frac{2-q}{q}} & q < 2, p = 2, \\ \log\log n & q = p = 2, \\ (\log n)^{(\frac{\gamma}{q} - \alpha)(\frac{p}{2} - 1) + \frac{1}{q}(d + p\frac{2-d-q}{2} - q)} & p > 2, \end{cases}$$
$$= C \begin{cases} n^{1-\frac{p}{2}} & p < 2 \text{ or } q > 2, p = 2, \\ (\log n)^{\frac{2-q}{q}} & q < 2, p = 2, \\ \log\log n & q = p = 2, \\ (\log n)^{(\frac{\gamma}{q} - \alpha + d(1-\frac{1}{q}))(\frac{p}{2} - 1) - \frac{p}{2}(d-1) + d - 1 - p(1-\frac{1}{q})} & p > 2, \end{cases}$$

while if $q = 0$

$$(6.2.2.37) \le C \begin{cases} n^{1-\frac{p}{2}} & p < \underline{p}, \\ n^{1-\frac{p}{2}} \log n & p = \underline{p}, \\ n^{\frac{p}{\gamma - d}} (\log n)^{-\beta(\frac{p}{\gamma - d} + \frac{p}{2} - 1)} & p > \underline{p}, \end{cases} \qquad \underline{p} \text{ defined in } (6.2.2.39),$$

and these bounds complies with the estimate in the statement, since the rate is $\tau_n^{p,d}$ defined in (1.2.0.3). To see this, notice that the previous thresholds in the estimates above are: $p = 2$ for $q > 0$, which is equal to $p^{crit} = d$ if $d = 2$ and strictly smaller than that if $d > 2$, and $\underline{p}$ for $q = 0$, which is equal to $p^{crit} = \frac{d(\gamma - d)}{\gamma}$ if $d = 2$ and strictly smaller than that if $d > 2$. Therefore one can divide in cases: if $d = 2$ one has $\underline{p} = p^{crit}$ (or $2 = p^{crit}$ if $q > 0$) and therefore the cases $p <, =, > p^{crit}$ coincide with the cases $p <, =, > \underline{p}$ (or $p <, =, > 2$ if $q > 0$). Otherwise, if $d \ge 3$, one can first consider the case $p < p^{crit}$, which has no void intersection with all the cases $p <, =, > \underline{p}$ (or $p <, =, > 2$ if $q > 0$) since the threshold is strictly smaller than p^{crit} by (6.2.2.40), and then one can study the cases $p \ge p^{crit}$, which are included in the case $p > \bar{p}$ (or $p > 2$ if $q > 0$) for the same reason.

Now we focus on (6.2.2.38). One can notice that

$$\sum_{k \ge 0} \frac{n(A_{r_k, r_{k+1}})}{\rho_n(A_{r_k, r_{k+1}})} \rho_n \mathbb{1}_{A_{r_k, r_{k+1}}} = \frac{n(B_{r_1})}{\rho_n(B_{r_1})} \rho_n \mathbb{1}_{B_{r_1}} + \sum_{k \ge 1} \frac{n(A_{r_k, r_{k+1}})}{\rho_n(A_{r_k, r_{k+1}})} \rho_n \mathbb{1}_{A_{r_k, r_{k+1}}},,$$

and therefore by the property (2.1.0.2) of subadditivity (the two measures in (6.2.2.38) coincide on B_{r_1}), we have

$$(6.2.2.38) \le \mathbb{E}\left[W_p^p \left(\sum_{k \ge 1} \frac{n(A_{r_k, r_{k+1}})}{\rho_n(A_{r_k, r_{k+1}})} \rho_n \mathbb{1}_{A_{r_k, r_{k+1}}}, \sum_{k \ge 1} \int_{\partial B_1} d\mathcal{H}^{d-1}(y) \int_{O(d-1)} dU \frac{n(Q^{y,U}_{r_k, \delta_k})}{\mathcal{M}_{\delta_k} \rho_n(Q^{y,U}_{r_k, \delta_k})} \rho_n \mathbb{1}_{Q^{y,U}_{r_k, \delta_k}} \right) \right]$$
$$\le \sum_{k \ge 1} \mathbb{E}\left[W_p^p \left(\frac{n(A_{r_k, r_{k+1}})}{\rho_n(A_{r_k, r_{k+1}})} \rho_n \mathbb{1}_{A_{r_k, r_{k+1}}}, \int_{\partial B_1} d\mathcal{H}^{d-1}(y) \int_{O(d-1)} dU \frac{n(Q^{y,U}_{r_k, \delta_k})}{\mathcal{M}_{\delta_k} \rho_n(Q^{y,U}_{r_k, \delta_k})} \rho_n \mathbb{1}_{Q^{y,U}_{r_k, \delta_k}} \right) \right],$$

where in the second inequality we have used again the subadditivity (over k). Then, by Lemma 6.2.5 applied with $r = r_k$ and $s = r_{k+1} = r_k(1 + \delta_k)$, recalling that

$$Q^{y,U}_{r_k, \delta_k} = A^{\Delta^{y,U}_{\delta}}_{r_k, r_{k+1}},$$

we get

$$(6.2.2.38) \le C \sum_{k \ge 1} r_k^p \tau_{d-1}^p (\mathbb{E}[n(A_{r_k, r_{k+1}})]) = \sum_{k \ge 1} r_k^p \tau_{d-1}^p (n \rho_n(A_{r_k, r_{k+1}})). \tag{6.2.2.41}$$

Thus we have to estimate $\rho_n(A_{r_k, r_{k+1}})$, that is

$$\rho_n(A_{r_k, r_{k+1}}) = \mathcal{H}^{d-1}(\partial B_1) \int_{r_k}^{r_k + 1} dr r^{d-1} \rho_n(r) \le C(r_{k+1} - r_k) r_k^{d-1} \rho_n(r_k). \tag{6.2.2.42}$$

Estimate of (6.2.2.38) in the case $d \geq 4$: in this case we see from (1.1.1.2) that $\tau^p_{d-1}(m) = m^{1-\frac{p}{d-1}}$, and therefore combining (6.2.2.42) with (6.2.2.41) we get

$$\begin{aligned}
(6.2.2.38) &\leq Cn^{1-\frac{p}{d-1}} \sum_{k\geq 1} (r_{k+1}-r_k)^{1-\frac{p}{d-1}} r_k^{d-1} \rho_n(r_k)^{1-\frac{p}{d-1}} \\
&\leq Cn^{1-\frac{p}{d-1}} \sum_{k\geq 1} (r_{k+1}-r_k) r_k^{d-1+p\frac{q-1}{d-1}} \rho_n(r_k)^{1-\frac{p}{d-1}} \qquad \text{since } r_{k+1}-r_k = \varepsilon r_k^{1-q} \text{ in this case} \\
&\leq Cn^{1-\frac{p}{d-1}} \int_{r_1}^{R_n} dr r^{d-1+p\frac{q-1}{d-1}} \rho_n(r)^{1-\frac{p}{d-1}} \\
&= Cn^{1-\frac{p}{d-1}} \int_{r_1}^{R_n} dr e^{-\frac{r^q}{q}(1-\frac{p}{d-1})} r^{d-\gamma+p\frac{q-1+\gamma}{d-1}-1} \\
&\leq Cn^{1-\frac{p}{d-1}} \begin{cases} 1 & q>0, p<d-1 \text{ or } q=0, p<\bar{p}, \\ R_n^{d+q-1} & q>0, p=d-1, \\ \log R_n & q=0, p=\bar{p}, \\ e^{-\frac{R_n^q}{q}(1-\frac{p}{d-1})} R_n^{d-\gamma+p\frac{q-1+\gamma}{d-1}-q} & q>0, p>d-1 \text{ or } q=0, p>\bar{p}. \end{cases}
\end{aligned}$$

with

$$\bar{p} = \frac{(\gamma-d)(d-1)}{\gamma-1}. \tag{6.2.2.43}$$

Notice that

$$p^* - \bar{p} = \frac{(\gamma-d)^2}{\gamma(\gamma-1)} > 0. \tag{6.2.2.44}$$

Applying the previous estimate to the definition of R_n in (6.1.1), for $q>0$ we get

$$(6.2.2.38) \leq C \begin{cases} n^{1-\frac{p}{d-1}} & p<d-1 \\ (\log n)^{\frac{d}{q}+1-\frac{1}{q}} & p=d-1, \\ (\log n)^{(\frac{\gamma}{q}-\alpha)(\frac{p}{d-1}-1)+\frac{1}{q}(d+p\frac{q-1}{d-1}-q)} = (\log n)^{(\frac{\gamma}{q}-\alpha+d(1-\frac{1}{q}))(\frac{p}{d-1}-1)+d-1-p(1-\frac{1}{q})} & p>d-1, \end{cases}$$

while for $q=0$

$$(6.2.2.38) \leq C \begin{cases} n^{1-\frac{p}{d-1}} & p<\bar{p}, \\ n^{1-\frac{p}{d-1}} \log n & p=\bar{p}, \\ n^{\frac{p}{\gamma-d}} (\log n)^{-\beta(-1+\frac{p(\gamma-1)}{(d-1)(\gamma-d)})} & p>\bar{p}. \end{cases} \qquad \bar{p} \text{ defined in (6.2.2.43)},$$

To check that this estimates are consistent with the statement it is enough to recall the definition of the rate $\tau_n^{p,d}$ in (1.2.0.3) and to divide in the two cases: since in the estimates above the thresholds are $p=d-1$ for $q>0$ and $p=\bar{p}$ for $q=0$, both strictly lower than p^{crit}. Hence, one can first analyze the case $p<p^{crit}$, which has no void intersection with $p <,=,> \underline{p}$ (or $p<,=,> d-1$ in the case $q>0$), and then $p \geq p^{crit}$, which is contained in the case $p>\bar{p}$ (or $p>d-1$ if $q>0$) by (6.2.2.44).

Now we study the other two cases ($d=2,3$).

Estimate of (6.2.2.38) in the case $d=3$: since in this case $d-1=2$, we have $\tau^p_{d-1}(m) = m^{1-\frac{p}{2}}(\log m)^{\frac{p}{2}}$: arguing as before, in this case we get

$$\begin{aligned}
(6.2.2.38) &\leq C\sum_{k\geq 1} r_k^p (n\rho_n(A_{r_k,r_{k+1}}))^{1-\frac{p}{2}} (\log(n\rho_n(A_{r_k,r_{k+1}})))^{\frac{p}{2}} \\
&\leq Cn^{1-\frac{p}{2}} \sum_{k\geq 1} (r_{k+1}-r_k) r_k^{\frac{p(q-1)}{2}+2} \rho_n(r_k)^{1-\frac{p}{2}} (\log(nr_k^{3-q}\rho_n(r_k))^{\frac{p}{2}} \\
&\leq Cn^{1-\frac{p}{2}} \int_{r_1}^{R_n} dr r^{\frac{p(q-1)}{2}+2} \rho_n(r)^{1-\frac{p}{2}} (\log(nr^{3-q}\rho_n(r))^{\frac{p}{2}} \\
&\leq Cn^{1-\frac{p}{2}} \int_{R_n'}^{R_n} dr e^{-\frac{r^q}{q}(1-\frac{p}{2})} r^{3-\gamma+p\frac{q-1+\gamma}{2}-1} (\log(nr^{3-q-\gamma}e^{-\frac{r^q}{q}}))^{\frac{p}{2}} &(6.2.2.45) \\
&+ Cn^{1-\frac{p}{2}} \int_{r_1}^{R_n'} dr e^{-\frac{r^q}{q}(1-\frac{p}{2})} r^{3-\gamma+p\frac{q-1+\gamma}{2}-1} (\log(nr^{3-q-\gamma}e^{-\frac{r^q}{q}}))^{\frac{p}{2}}, &(6.2.2.46)
\end{aligned}$$

where in (6.2.2.45) and (6.2.2.46) we have split the previous integral as the sum of two integrals, the first one in $[0, R'_n)$ and the second one in $[R'_n, R_n]$, with

$$R'_n = \begin{cases} \sqrt[q]{q \log\left(\frac{n}{(\log n)^{\alpha'}}\right)} & q > 0, \\ \sqrt[\gamma-3]{\frac{n}{(\log n)^{\beta'}}} & q = 0, \end{cases}$$

with $\alpha' \geq \alpha$ and $\beta' \geq \beta$ to be fixed later.

Let us begin with the estimate of (6.2.2.45): one can notice that if $r \geq R'_n$

$$nr^{3-q-\gamma}e^{-\frac{r^q}{q}} \leq nR'^{3-q-\gamma}_n e^{-\frac{R'^q_n}{q}} = \begin{cases} (\log n)^{\alpha' + \frac{3-q-\gamma}{q}} & q > 0, \\ (\log n)^{\beta'} & q = 0, \end{cases}$$

and therefore

$$r \geq R'_n \Rightarrow (\log(nr^{3-q-\gamma}e^{-\frac{r^q}{q}}))^{\frac{p}{2}} \leq C(\log\log n)^{\frac{p}{2}}. \tag{6.2.2.47}$$

Applying this inequality (6.2.2.47) to (6.2.2.45) we have

$$\begin{aligned} (6.2.2.45) &\leq Cn^{1-\frac{p}{2}}(\log\log n)^{\frac{p}{2}} \int_{R'_n}^{R_n} dr e^{-\frac{r^q}{q}(1-\frac{p}{2})} r^{3-\gamma+p\frac{q-1+\gamma}{2}-1} \\ &\leq C(\log\log n)^{\frac{p}{2}} n^{1-\frac{p}{2}} \begin{cases} 1 & q > 0, p < 2 \text{ or } q = 0, p < \bar{p}, \\ R_n^{2+q} & q > 0, p = 2, \\ \log R_n & q = 0, p = \bar{p}, \\ e^{-\frac{R_n^q}{q}(1-\frac{p}{2})} R_n^{3-\gamma+p\frac{q-1+\gamma}{2}-q} & q > 0, p > 2 \text{ or } q = 0, p > \bar{p}, \end{cases} \end{aligned}$$

with

$$\bar{p} = \frac{2(\gamma-3)}{\gamma-1}. \tag{6.2.2.48}$$

We used the same notations for this $\bar{p}$ in (6.2.2.48) and the $\bar{p}$ defined in (6.2.2.43) because the first definition recovers the second one for $d = 3$.

According to the definition of R_n in (6.1.0.6), for $q > 0$ the previous estimate reads as

$$(6.2.2.45) \leq C(\log\log n)^{\frac{p}{2}} \begin{cases} n^{1-\frac{p}{2}} & p < 2 \\ (\log n)^{1+\frac{2}{q}} & p = 2, \\ (\log n)^{(\frac{\gamma}{q}-\alpha)(\frac{p}{2}-1)+\frac{1}{q}(3+p\frac{q-1}{2}-q)} = (\log n)^{(\frac{\gamma}{q}-\alpha+3(1-\frac{1}{q}))(\frac{p}{2}-1)+2-p(1-\frac{1}{q})} & p > 2, \end{cases}$$

whereas for $q = 0$

$$(6.2.2.45) \leq C \begin{cases} n^{1-\frac{p}{2}}(\log\log n)^{\frac{p}{2}} & p < \bar{p}, \\ n^{1-\frac{p}{2}}(\log\log n)^{\frac{p}{2}} \log n & p = \bar{p}, \\ (\log\log n)^{\frac{p}{2}} n^{\frac{p}{\gamma-3}} (\log n)^{-\beta(-1+p\frac{\gamma-1}{2(\gamma-3)})} & p > \bar{p}. \end{cases} \qquad \bar{p} \text{ defined in (6.2.2.48)}.$$

Now we focus on (6.2.2.46): in this case, we use the rough estimate

$$(\log(nr^{3-q-\gamma}e^{-\frac{r^q}{q}}))^{\frac{p}{2}} \leq C(\log n)^{\frac{p}{2}},$$

and therefore

$$(6.2.2.46) \leq Cn^{1-\frac{p}{2}}(\log n)^{\frac{p}{2}} \int_{r_1}^{R'_n} dr e^{-\frac{r^q}{q}(1-\frac{p}{2})} r^{3-\gamma+p\frac{q-1+\gamma}{2}-1}.$$

The same steps as the ones used for $r \geq R'_n$ prove that if $q > 0$

$$(6.2.2.46) \leq Cn^{1-\frac{p}{2}}(\log n)^{\frac{p}{2}} \begin{cases} 1 & p < 2 \\ (\log n)^{1+\frac{2}{q}} & p = 2, \\ n^{\frac{p}{2}-1}(\log n)^{(\frac{\gamma}{q}-\alpha'+3(1-\frac{1}{q}))(\frac{p}{2}-1)+2-p(1-\frac{1}{q})} & p > 2, \end{cases}$$

while if $q = 0$

$$(6.2.2.46) \leq Cn^{1-\frac{p}{2}}(\log n)^{\frac{p}{2}} \begin{cases} 1 & p < \bar{p}, \\ \log n & p = \bar{p}, \\ n^{-1+\frac{p}{2}+\frac{p}{\gamma-3}}(\log n)^{-\beta'(-1+p\frac{\gamma-1}{2(\gamma-3)})} & p > \bar{p}. \end{cases} \qquad \bar{p} \text{ defined in (6.2.2.48)}.$$

Now we can see that if

$$\alpha'-\alpha=\begin{cases}\frac{p}{p-2} & p>2,\\ 0 & p\le 2,\end{cases}\quad\text{and } \beta'-\beta=\begin{cases}\frac{p}{2}(-1+p\frac{\gamma-1}{2(\gamma-3)})^{-1} & p>\bar p,\\ 0 & p\le\bar p,\end{cases}\qquad \bar p \text{ defined in (6.2.2.48)},$$

so that, in the case $p>2$ and $p>\bar p$, (6.2.2.46) is much smaller than (6.2.2.45), the previous estimates

$$(6.2.2.38)\le(6.2.2.45)+(6.2.2.46),$$

prove that if $q>0$

$$(6.2.2.38)\le\begin{cases} n^{1-\frac{p}{2}}(\log n)^{\frac{p}{2}} & p<2,\\ (\log n)^{2+\frac{2}{q}} & p=2,\\ (\log\log n)^{\frac{p}{2}}(\log n)^{(\frac{\gamma}{q}-\alpha+3(1-\frac{1}{q}))(\frac{p}{2}-1)+2-p(1-\frac{1}{q})} & p>2,\end{cases}$$

and that if $q=0$

$$(6.2.2.38)\le C\begin{cases} n^{1-\frac{p}{2}}(\log n)^{\frac{p}{2}} & p<\bar p,\\ n^{1-\frac{p}{2}}(\log n)^{1+\frac{p}{2}} & p=\bar p,\\ n^{\frac{p}{\gamma-3}}(\log\log n)^{\frac{p}{2}}(\log n)^{-\beta(-1+p\frac{\gamma-1}{2(\gamma-3)})} & p>\bar p.\end{cases}\qquad \bar p\text{ defined in (6.2.2.48)},$$

and this estimate complies with the statement too: to see this, the same argument used in the end of the previous Paragraph for $d\ge 4$ works.

Estimate of (6.2.2.38) in the case $d=2$: since in this case $d-1=1$, we have $\tau^p_{d-1}(m)=m^{1-\frac{p}{2}}$, and therefore by (6.2.2.42) we get

$$\begin{aligned}(6.2.2.38)&\le C\sum_{k\ge1}r_k^p(n\rho_n(A_{r_k,r_{k+1}}))^{1-\frac{p}{2}}\le Cn^{1-\frac{p}{2}}\sum_{k\ge1}r_k^p((r_{k+1}-r_k)r_k\rho_n(r_k))^{1-\frac{p}{2}}\\
&\le Cn^{1-\frac{p}{2}}\sum_{k\ge1}(r_{k+1}-r_k)r_k^{\frac{pq}{2}+1}\rho_n(r_k)^{1-\frac{p}{2}}\\
&\le Cn^{1-\frac{p}{2}}\int_{r_1}^{R_n}dr\,r^{\frac{pq}{2}+1}\rho_n(r)^{1-\frac{p}{2}}\\
&\le Cn^{1-\frac{p}{2}}\int_{r_1}^{R_n}dr\,e^{-\frac{r^q}{q}(1-\frac{p}{2})}r^{2-\gamma+p\frac{q+\gamma}{2}-1}\\
&\le Cn^{1-\frac{p}{2}}\begin{cases}1 & q>0,p<2\text{ or }q=0,p<\hat p,\\ R_n^{2+q} & q>0,p=2,\\ \log R_n & q=0,p=\hat p,\\ e^{-\frac{R_n^q}{q}(1-\frac{p}{2})}R_n^{2-\gamma+p\frac{q+\gamma}{2}-q} & q>0,p>2\text{ or }q=0,p>\hat p,\end{cases}\end{aligned}$$

with

$$\hat p=\frac{2(\gamma-2)}{\gamma}=p^*. \tag{6.2.2.49}$$

We estimate the previous quantity in the two different cases $q>0$ and $q=0$. If $q>0$ we have

$$(6.2.2.38)\le C\begin{cases} n^{1-\frac{p}{2}} & p<2,\\ (\log n)^{1+\frac{2}{q}} & p=2,\\ (\log n)^{(\frac{\gamma}{q}-\alpha)(\frac{p}{2}-1)+\frac{1}{q}(2+\frac{pq}{2}-q)}=(\log n)^{(\frac{\gamma}{q}-\alpha+2(1-\frac{1}{q}))(\frac{p}{2}-1)+\frac{p}{2}+1-p(1-\frac{1}{q})} & p>2,\end{cases}$$

while if $q=0$

$$(6.2.2.38)\le C\begin{cases} n^{1-\frac{p}{2}} & p<\hat p,\\ n^{1-\frac{p}{2}}\log n & p=\hat p,\\ n^{\frac{p}{\gamma-2}}(\log n)^{-\beta(-1+\frac{p\gamma}{2(\gamma-2)})} & p>\hat p,\end{cases}\qquad \hat p\text{ defined in (6.2.2.49)},$$

and these estimates complete the proof since the rate is $\tau_n^{p,d}$ in (1.2.0.3) and in this case the thresholds to study are $p=2$ (if $q>0$) and $p=\hat p=p^*$ (for $q=0$), both equal to p^{crit}. □

Proposition 6.2.2 (The global estimate with decomposition angular-radial (6.2.2.10).). *Given $\{\tilde X_i\}_{i=1}^n$ iid random variables in $\mathbb{R}^d$ with common probability density ρ_n and $n(\cdot)$ the number of such variables in a domain as in Notation 6.2.2. Then, if $g(n)$ is the quantity*

$$g(n) := \mathbb{E}\left[W_p^p\left(n\rho_n \mathbb{1}_{B_{R_n}}, \frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n \mathbb{1}_{B_{r_1}} + \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \sum_{k\geq 1}\frac{n(Q^{y,U}_{r_k,\delta})}{\mathcal{M}_\delta \rho_n(Q^{y,U}_{r_k,\delta})}\rho_n \mathbb{1}_{Q^{y,U}_{r_k,\delta}}\right)\right],$$

(depending also on p,d,q,γ,r_1 and ε), there exist a constant $c>0$ and a constant $C=C(\varepsilon)$ such that the following estimates hold: if $q>0$

$$g(n) \leq C(\varepsilon)\tau_n^{p,d}(\log n)^{-c(\alpha-\frac{\gamma}{d}-d(1-\frac{1}{q}))}\begin{cases} 1 & p<d\geq 3 \text{ or } p\geq d,\\ (\log n)^{c(\alpha-\frac{\gamma}{d}-d(1-\frac{1}{q}))} & p<d=2,\end{cases}$$

while if $q=0$

$$g(n) \leq C(\varepsilon)\tau_n^{p,d}(\log n)^{-c\beta}.$$

Before proving this estimate, we can notice that thanks to the choices of α and β in Definition 6.1.1, even in this Proposition 6.2.2, in a lot of cases it is ensured that

$$g(n)\begin{cases} \ll \tau_n^{p,d} & p\leq p^{crit},\\ \lesssim \tau_n^{p,d} & p> p^{crit},\end{cases} \tag{6.2.2.50}$$

where the only eventuality not included in the previous estimate is $p<d=2, q>0$. Of course the implicit constant depend on ε.

Proof. In this case, we use triangle inequality with the measure

$$\tilde\mu := \frac{n(B_{r_1})\mathbb{1}_{B_{r_1}}}{\rho_n(B_{r_1})} + \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \frac{n(A^{\Delta^{y,U}_\delta}_{r_1,R_n})}{\mathcal{M}_\delta\rho_n(A^{\Delta^{y,U}_\delta}_{r_1,R_n})}\rho_n \mathbb{1}_{A^{\Delta^{y,U}_\delta}_{r_1,R_n}},$$

obtaining

$$g(n) \leq 2^p\mathbb{E}\left[W_p^p\left(n\rho_n\mathbb{1}_{B_{R_n}},\tilde\mu\right)\right] \tag{6.2.2.51}$$

$$+2^p\mathbb{E}\left[W_p^p\left(\tilde\mu, \frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}} + \sum_{k\geq 1}\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\frac{n(Q^{y,U}_{r_k,\delta})}{\mathcal{M}_\delta\rho_n(Q^{y,U}_{r_k,\delta})}\rho_n\mathbb{1}_{Q^{y,U}_{r_k,\delta}}\right)\right]. \tag{6.2.2.52}$$

We begin with the estimate of (6.2.2.51), and to bound it we use again triangle inequality with the measure

$$\frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}} + \frac{n(A_{r_1,R_n})}{\rho_n(A_{r_1,R_n})}\rho_n\mathbb{1}_{A_{r_1,R_n}},$$

and then we use Lemmas 6.2.3 and 6.2.5, indeed

$$(6.2.2.51) \leq 2^p\mathbb{E}\left[W_p^p\left(n\rho_n\mathbb{1}_{B_{R_n}}, \frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}} + \frac{n(A_{r_1,R_n})}{\rho_n(A_{r_1,R_n})}\rho_n\mathbb{1}_{A_{r_1,R_n}}\right)\right] \tag{6.2.2.53}$$

$$+2^p\mathbb{E}\left[W_p^p\left(\frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}} + \frac{n(A_{r_1,R_n})}{\rho_n(A_{r_1,R_n})}\rho_n\mathbb{1}_{A_{r_1,R_n}},\tilde\mu\right)\right]. \tag{6.2.2.54}$$

(6.2.2.53) is exactly the term appearing in (6.2.2.14) of Lemma 6.2.3, and therefore we can estimate

$$(6.2.2.53) \leq Cn^{1-\frac{p}{2}},$$

while by subadditivity (see (2.1.0.2)) we can estimate (6.2.2.54) by noticing that the two measures involved coincide on B_{r_1}, and therefore we can estimate it with

$$\begin{aligned}(6.2.2.54) &\leq \mathbb{E}\left[W_p^p\left(\frac{n(A_{r_1,R_n})}{\rho_n(A_{r_1,R_n})}\rho_n\mathbb{1}_{A_{r_1,R_n}}, \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\frac{n(A^{\Delta^{y,U}_\delta}_{r_1,R_n})}{\mathcal{M}_\delta\rho_n(A^{\Delta^{y,U}_\delta}_{r_1,R_n})}\rho_n\mathbb{1}_{A^{\Delta^{y,U}_\delta}_{r_1,R_n}}\right)\right]\\ &\leq C\frac{\int_{r_1}^{R_n} ds\rho_n(s)s^{p+d-1}}{\int_{r_1}^{R_n} ds\rho_n(s)s^{d-1}}\tau^p_{d-1}(\mathbb{E}[n(A_{r_1,R_n})]),\end{aligned}$$

using Lemma 6.2.5 with $r = r_1$ and $s = R_n$. If we now observe that the first coefficient

$$\frac{\int_{r_1}^{R_n} ds\rho_n(s)s^{p+d-1}}{\int_{r_1}^{R_n} ds\rho_n(s)s^{d-1}} \le C,$$

because both the quantities in the numerator and in the denominator are bounded, and that

$$\mathbb{E}[n(A_{r_1,R_n})] = n\int_{r_1}^{R_n} dr\rho_n(r)r^{d-1}\mathcal{H}^{d-1}(\partial B_1) = Cn,$$

by (1.1.1.2) we get

$$(6.2.2.54) \le C\begin{cases} n^{1-\frac{p}{d-1}} & d \ge 4, \\ n^{1-\frac{p}{2}}(\log n)^{\frac{p}{2}} & d = 3, \\ n^{1-\frac{p}{2}} & d = 2, \end{cases}$$

and therefore the first part of the proof in concluded since we just proved that

$$(6.2.2.51) \le (6.2.2.53) + (6.2.2.54) \le Cn^{1-\frac{p}{2}} + C\begin{cases} n^{1-\frac{p}{d-1}} & d \ge 4, \\ n^{1-\frac{p}{2}}(\log n)^{\frac{p}{2}} & d = 3, \\ n^{1-\frac{p}{2}} & d = 2, \end{cases} \lesssim n^{-\xi}\tau_n^{p,d},$$

for a positive $\xi > 0$, and this estimate complies with the statement.

Now we estimate (6.2.2.52). We have to compare two measures, the first one is

$$\tilde{\mu} = \frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}} + \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\frac{n(A_{r_1,R_n}^{\Delta_\delta^{y,U}})}{\mathcal{M}_\delta\rho_n(A_{r_1,R_n}^{\Delta_\delta^{y,U}})}\rho_n\mathbb{1}_{A_{r_1,R_n}^{\Delta_\delta^{y,U}}},$$

while the second one is

$$\frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}} + \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\sum_{k\ge 1}\frac{n(Q_{r_k,\delta}^{y,U})}{\mathcal{M}_\delta\rho_n(Q_{r_k,\delta}^{y,U})}\rho_n\mathbb{1}_{Q_{r_k,\delta}^{y,U}},$$

and therefore by subadditivity (see (2.1.0.2)) one has

$$(6.2.2.52) \le \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\mathbb{E}\left[W_p^p\left(\frac{n(A_{r_1,R_n}^{\Delta_\delta^{y,U}})}{\mathcal{M}_\delta\rho_n(A_{r_1,R_n}^{\Delta_\delta^{y,U}})}\rho_n\mathbb{1}_{A_{r_1,R_n}^{\Delta_\delta^{y,U}}}, \sum_{k\ge 1}\frac{n(Q_{r_k,\delta}^{y,U})}{\mathcal{M}_\delta\rho_n(Q_{r_k,\delta}^{y,U})}\rho_n\mathbb{1}_{Q_{r_k,\delta}^{y,U}}\right)\right].$$

The expression we found in the previous inequality can be bounded using Lemma 6.2.3. Indeed if

$$\Sigma = \Delta_\delta^{y,U},$$

we have

$$Q_{r_k,\delta}^{y,U} = A_{r_k,r_{k+1}}^{\Sigma},$$

and therefore by (6.2.2.15) of Lemma 6.2.3 with $\Sigma = \Delta_\delta^{y,U}$, we have

$$(6.2.2.52) \le \frac{C}{\mathcal{M}_\delta}(n\mathcal{H}^{d-1}(\Delta_\delta^{\bar{y},\bar{U}}))^{1-\frac{p}{2}}\int_{r_1}^{R_n} dr\rho(r)^{1-\frac{p}{2}}r^{d-1-p\frac{d+q-2}{2}}.$$

We briefly focus on the first factor outside the integral, indeed by definition

$$\mathcal{M}_\delta = \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\mathbb{1}_{\Delta_\delta^{y,U}} = \int_{O(d-1)} dU\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\mathbb{1}_{\Delta_\delta^{y,U}} \ge \frac{1}{C}\delta^{d-1},$$

and also

$$\frac{1}{C}\delta^{d-1} \le \mathcal{H}^{d-1}(\Delta_\delta^{\bar{y},\bar{U}}) \le C\delta^{d-1},$$

hence

$$\begin{aligned}(6.2.2.52) &\le Cn^{1-\frac{p}{2}}\delta^{-\frac{p}{2}(d-1)}\int_{r_1}^{R_n} dr\rho(r)^{1-\frac{p}{2}}r^{d-1-p\frac{d+q-2}{2}} \\ &= C(\varepsilon)n^{1-\frac{p}{2}}R_n^{\frac{qp}{2}(d-1)}\int_{r_1}^{R_n} dr\rho(r)^{1-\frac{p}{2}}r^{d-1-p\frac{d+q-2}{2}},\end{aligned}$$

and we have already studied this quantity at the beginning of the proof of the previous Proposition 6.2.1, within the estimate of (6.2.2.37). Recalling the definition of R_n the same computations prove that if $q > 0$

$$
\begin{aligned}
(6.2.2.52) &\leq C n^{1-\frac{p}{2}} (\log n)^{\frac{p}{2}(d-1)} \begin{cases} 1 & p < 2 \text{ or } q > 2, p = 2, \\ (\log n)^{\frac{1}{q}(2-q)} & q < 2, p = 2, \\ \log\log n & q = p = 2, \\ n^{\frac{p}{2}-1} (\log n)^{(\frac{\gamma}{q}-\alpha)(\frac{p}{2}-1)+\frac{1}{q}(d+p\frac{2-d-q}{2}-q)} & p > 2, \end{cases} \\
&= C \begin{cases} n^{1-\frac{p}{2}} (\log n)^{\frac{p}{2}(d-1)} & p < 2 \text{ or } q > 2, p = 2, \\ (\log n)^{d-2+\frac{2}{q}} & q < 2, p = 2, \\ (\log n)^{d-1} \log\log n & q = p = 2, \\ (\log n)^{(\frac{\gamma}{q}-\alpha+d(1-\frac{1}{q}))(\frac{p}{2}-1)+d-1-p(1-\frac{1}{q})} & p > 2, \end{cases}
\end{aligned}
$$

while if $q = 0$

$$
(6.2.2.52) \leq C \begin{cases} n^{1-\frac{p}{2}} & p < \underline{p}, \\ n^{1-\frac{p}{2}} \log n & p = \underline{p}, \\ n^{\frac{p}{\gamma-d}} (\log n)^{-\beta(-1+\frac{p(\gamma-d+2)}{2(\gamma-d)})} & p > \underline{p}, \end{cases} \qquad \underline{p} = \frac{2(\gamma-d)}{\gamma+2-d} = p^* - \frac{(\gamma-d)^2(d-2)}{\gamma(\gamma-d+2)}
$$

and this completes the proof since, as in the previous Proposition, to check the estimates it is sufficient to analyze the cases:

- $d = 2$, where the previous thresholds are $p = 2$ (if $q > 0$) and $p = \underline{p}$ (if $q = 0$), both equal to p^{crit},
- $d \geq 3$, where the thresholds are $p = 2$ (if $q > 0$) and $p = \underline{p}$ (if $q = 0$), both strictly smaller than p^{crit}, where therefore $p < p^{crit}$ has no void intersection with $p <, =, > \underline{p}$ ($p <, =, > \underline{2}$ if $q > 0$), and $p \geq p^{crit}$, included in $p > \underline{p}$ ($p > 2$ if $q > 0$).

□

The consequence of Propositions 6.2.1 and 6.2.2 is that, up to choose the decomposition in the suitable way, that is, as in (6.2.2.10) except in the case $p < d = 2, q > 0$, where (6.2.2.9) is more suitable, it is possible to prove that the expectation of the global term in (6.2.2.11) is much smaller than the rate $\tau_n^{p,d}$ for $p \leq p^{crit}$ and has at most its same order for $p > p^{crit}$. In other words, as a corollary of (6.2.2.36) and (6.2.2.50), we have proved Theorem 6.2.1.

6.2.3 Geometric decomposition for the lower bound.

The geometric decomposition for the lower bound is much easier since we do not have to care about the global term. Indeed the boundary Wasserstein is superadditive, so there is no global term to estimate. The cut off at a critical radius is not needed, that is, one is going to estimate the cost only in a region which approximately coincides with the ball of radius R_n, but instead of constraining the variables inside it, one only looks at the cost given by the points which are already inside it.

Thus, since one does not have to estimate the global term, there is no need to divide B_{R_n} in such a way not to break the radial symmetry of the density: one can simply consider a set of hypercubes covering B_{R_n}.

By the same argument we used in Subsection 6.2.1, one can notice that it is still important that the hypercubes $\{Q_j\}$ are to be taken small enough iso that the density ρ restricted to Q_j can be transported to the uniform one.

Here we take the partition in a slightly different way: we fix $\ell \in \mathbb{N}$ (to be sent to ∞ eventually), and, for $x_1 > 0$, we consider the following sequence of $\{x_k\}_{k=1}^{\infty}$, that is

$$
x_{k+1} = x_k \left(1 + \frac{1}{\ell \lfloor x_k^q \rfloor}\right).
$$

The previous sequence is very similar to the sequence of radii in (6.2.2.8), with the only difference that here we need the condition

$$
\mathbb{N} \ni \frac{x_k}{x_{k+1} - x_k} = \ell \lfloor x_k^q \rfloor. \tag{6.2.3.1}
$$

Then, we consider the hypercubes

$$
Q^k := [-x_k, x_k]^d,
$$

and thanks to (6.2.3.1) we can divide $Q^{k+1}\setminus Q^k$ in a set of smaller hypercubes $\mathcal{Q}_k = \{Q_j^{(k)}\}_j$, with side $x_{k+1}-x_k$. The number of these hypercubes (that is, the cardinality of $\mathcal{Q}_k$) is

$$\frac{|Q^{k+1}\setminus Q^k|}{|Q_j^{(k)}|} = \frac{|Q^{k+1}|-|Q^k|}{|Q_j^{(k)}|} = \frac{(2x_{k+1})^d-(2x_k)^d}{(x_{k+1}-x_k)^d} = 2^d(\ell\lfloor x_k^q\rfloor)^d\left[\left(1+\frac{1}{\ell\lfloor x_k^q\rfloor}\right)^d-1\right] = 2^d\sum_{k=1}^d\binom{d}{k}(\ell\lfloor x_k^q\rfloor)^{d-k}.$$

Then we denote by

$$\mathcal{Q} = \bigcup_{k\in\mathbb{N}}\mathcal{Q}_k = \{Q_j\}_{j\in\mathbb{N}},$$

and we use the following decomposition.

Definition 6.2.6. *We define*

$$\mathcal{J} := \{j\in\mathbb{N} : Q_j\cap B_{R_n}\neq\emptyset\},$$

and the geometric decomposition for the lower bound is taken by considering

$$\bigcup_{j\in\mathcal{J}} Q_j.$$

6.3 The comparison with the uniform case.

In this section we focus on the matching problem inside an hypercube, and we use an argument very similar to [7, 2] to compare the cost of the gaussian (or polynomial) matching to the uniform one. We shall prove the following results. Just for the sake of clarity we separate the upper bound and the lower bound, indeed the geometric decomposition is different in the two cases (see Subsections 6.2.1 and 6.2.3). To start with, we state both the estimates, and then we shall prove them.

Lemma 6.3.1 (Compare to the uniform matching: upper bound.)**.** *Let $\{\tilde X_i\}_{i=1}^n$ be iid random variables with common distribution ρ_n and let $Q = Q^{y,U}_{r_k,\delta_k}$ one of the hypercubes in (6.2.1.1) in Definition 6.2.1 and r_k,δ_k by Definition 6.2.3. Then, if $\{Z_i\}_{i=1}^\infty$ is a sequence of iid random variables uniformly distributed on $[0,1]^d$, one has*

$$\mathbb{E}\left[W_p^p\left(\sum_{i:X_i\in Q}\delta_{X_i},\frac{n(Q)}{\rho_n(Q)}\rho_n\mathbb{1}_Q\right)\right]\le(1+c\varepsilon)|Q|^{\frac{p}{d}}\mathbb{E}\left[W_p^p\left(\sum_{i=1}^{n(Q)}\delta_{Z_i},n(Q)\mathbb{1}_{[0,1]^d}\right)\right],$$

with $n(Q)$ as in Notation 6.2.2.

Notice that the quantity ε appearing in the previous estimate is hidden in δ_k, in Definition 6.2.3.

The analogue estimate for the lower bound is very similar, but it involves the boundary Wasserstein distance instead of the ordinary one, and it holds on the hypercubes $\{Q_j\}$ by Definition 6.2.6 instead of 6.2.3.

Lemma 6.3.2 (Compare to the uniform matching: lower bound.)**.** *Let $\{X_i\}_{i=1}^n$ be iid random variables with common distribution ρ and let $Q = Q_j$ one of the hypercubes in Definition 6.2.6. Then, if $\{Z_i\}_{i=1}^\infty$ is a sequence of iid random variables uniformly distributed on Q, one has*

$$\mathbb{E}\left[Wb_p^p\left(\sum_{i:X_i\in Q}\delta_{X_i},\frac{\mathbb{E}[n(Q)]}{\rho(Q)}\rho\mathbb{1}_Q\right)\right]\ge\left(1-\frac{c}{\ell}\right)|Q|^{\frac{p}{d}}\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{n(Q)}\delta_{Z_i},\frac{\mathbb{E}[n(Q)]}{|Q|}\mathbb{1}_Q\right)\right],$$

with notation

$$n(Q) := \sum_{i=1}^n\mathbb{1}_Q(X_i).$$

The two proofs are quite similar, but Lemma 6.3.1 requires a little more effort. Indeed one has first to move the $Q = Q^{y,U}_{r_k,\delta_k}$ in a proper hypercube, and then to move the push-forwarded density on the hypercube into the uniform one.

Lemma 6.3.2 instead only requires to move the density over $Q = Q_j$ into the uniform one, since Q_j is already a proper hypercube.

Proof of Lemma 6.3.1. To prove this Lemma we first have to transform $Q = Q^{y,U}_{r_k,\delta_k}$ in a proper hypercube, and then move the push-forward of the obtained measure into the uniform one. As for the first passage, that is, to move Q in a proper hypercube of side 1, it is immediate to see that thanks to Lemma 6.2.1, we know that there exists a diffeomorphism

$$S:[0,1]^d\to Q = Q^{y,U}_{r_k,\delta_k},$$

such that

$$\sup_{x\in Q}\left\|\frac{DS}{r_k\delta_k}(x)-R^yU\right\|\le C\delta_k,\quad \left|\frac{\det DS(x)}{(r_k\delta_k)^d}-\det U\right|\le C\delta_k,\quad \sup_{x\in Q}\frac{|\nabla\det DS(x)|}{|\det DS(x)|}\le c\delta_k. \tag{6.3.0.1}$$

We can exploit it to conclude the proof: thanks to this estimate one gets

$$\frac{|S(x)-S(y)|}{|x-y|}\le r_k\delta_k\left(1+\sup_{x\in Q}\left\|D\left(\frac{S}{r_k\delta_k}\right)(x)-R^yU\right\|\right)\le r_k\delta_k(1+C\delta_k)\le r_k\delta_k(1+C\varepsilon). \tag{6.3.0.2}$$

Hence, using first Lemma 2.2.2 and then (6.3.0.2), one has

$$\begin{aligned}
W_p^p\left(\sum_{i:X_i\in Q}\delta_{X_i},\frac{n(Q)}{\rho_n(Q)}\rho_n\mathbb{1}_Q\right)&=W_p^p\left(\sum_{i:X_i\in Q}\delta_{X_i},\frac{n(Q)}{\rho(Q)}\rho\mathbb{1}_Q\right)\\
&=W_p^p\left(S_\sharp\left(\sum_{i:X_i\in Q}\delta_{S^{-1}(X_i)}\right),S_\sharp(S^{-1})_\sharp\left(\frac{n(Q)}{\rho(Q)}\rho\mathbb{1}_Q\right)\right) \qquad (6.3.0.3)\\
&\le\sup_{x,y\in Q}\left(\frac{|S(x)-S(y)|}{|x-y|}\right)^pW_p^p\left(\sum_{i:X_i\in Q}\delta_{S^{-1}(X_i)},(S^{-1})_\sharp\left(\frac{n(Q)}{\rho(Q)}\rho\mathbb{1}_Q\right)\right)\\
&\le(1+C\varepsilon)^p(r_k\delta_k)^pW_p^p\left(\sum_{i:X_i\in Q}\delta_{S^{-1}(X_i)},S^{-1}_\sharp\left(\frac{n(Q)}{\rho(Q)}\rho\mathbb{1}_Q\right)\right). \qquad (6.3.0.4)
\end{aligned}$$

We first claim that

$$(r_k\delta_k)^p\le(1+c\varepsilon)|Q|^{\frac{p}{d}}, \tag{6.3.0.5}$$

which, combined with (6.3.0.4), gives

$$W_p^p\left(\sum_{i:X_i\in Q}\delta_{X_i},\frac{n(Q)}{\rho_n(Q)}\rho_n\mathbb{1}_Q\right)\le(1+c\varepsilon)|Q|^{\frac{p}{d}}W_p^p\left(\sum_{i:X_i\in Q}\delta_{S^{-1}(X_i)},(S^{-1})_\sharp\left(\frac{n(Q)}{\rho(Q)}\rho\mathbb{1}_Q\right)\right). \tag{6.3.0.6}$$

Thus we have to relate the previous quantity to the analogue quantity for iid uniformly random variables, but, before that, let us prove (6.3.0.5). Since

$$\begin{aligned}
\sup_{x\in Q}\left|(r_k\delta_k)^d\det(D(S^{-1}(x))-\det U\right|&=\sup_{x\in Q}\left|(r_k\delta_k)^d\det((DS(S^{-1}(x))^{-1})-\det U\right|\\
&=\sup_{z\in[0,1]^d}\left|(r_k\delta_k)^d\det((DS(z)^{-1})-\det U\right|\\
&=\sup_{z\in[0,1]^d}\left|\frac{\det U(r_k\delta_k)^d}{\det DS(z)}-1\right|\le c\delta_k,\qquad\text{by (6.2.1.5) of Lemma 6.2.1,}
\end{aligned}$$

changing variables $x=S^{-1}(y)$, we have

$$1=\int_{[0,1]^d}dx=\int_Q dy|\det(D(S^{-1})(y))|\le\int_Q dy(r_k\delta_k)^{-d}|\det U|(1+c\delta_k)=|Q|(r_k\delta_k)^{-d}(1+c\delta_k),$$

and this proves (6.3.0.5).

Now we can go on with the proof. First we move $(S^{-1})_\sharp\rho$ into the uniform measure.

Step 1: we claim that there exists a map

$$T:[0,1]^d\to[0,1]^d$$

such that

$$(S^{-1})_\sharp\left(\frac{\rho\mathbb{1}_Q}{\rho(Q)}\right)=T_\sharp\left(\mathbb{1}_{[0,1]^d}\right)\quad\text{and}\quad\sup_{x,y\in Q}\frac{|T(x)-T(y)|}{|x-y|}\le1+C\varepsilon. \tag{6.3.0.7}$$

Let us postpone the proof and see the consequence of that. Using first Lemma 2.2.2 and then (6.3.0.7) we have

$$\begin{aligned}
W_p^p\left(\sum_{i:X_i\in Q}\delta_{S^{-1}(X_i)},(S^{-1})_\sharp\left(\frac{n(Q)}{\rho(Q)}\rho\mathbb{1}_Q\right)\right) &= W_p^p\left(T_\sharp\left(\sum_{i:X_i\in Q}\delta_{T^{-1}(S^{-1}(X_i))}\right),T_\sharp\left(n(Q)\mathbb{1}_{[0,1]^d}\right)\right)\\
&\le\left(\sup_{x,y\in Q}\frac{|T(x)-T(y)|}{|x-y|}\right)^p W_p^p\left(\sum_{i:X_i\in Q}\delta_{T^{-1}(S^{-1}(X_i))},n(Q)\mathbb{1}_{[0,1]^d}\right)\\
&\le(1+C\varepsilon)^p W_p^p\left(\sum_{i:X_i\in Q}\delta_{T^{-1}(S^{-1}(X_i))},n(Q)\mathbb{1}_{[0,1]^d}\right) \qquad (6.3.0.8)
\end{aligned}$$

Step 2: conclusion. As a result of (6.3.0.6) and (6.3.0.8), there exists a constant $C>0$ depending on p (and on ρ and d), such that

$$W_p^p\left(\sum_{i:X_i\in Q}\delta_{X_i},\frac{n(Q)}{\rho(Q)}\rho\mathbb{1}_Q\right)\le(1+C\varepsilon)|Q|^{\frac{p}{d}}W_p^p\left(\sum_{i:X_i\in Q}\delta_{T^{-1}(S^{-1}(X_i))},n(Q)\mathbb{1}_{[0,1]^d}\right).$$

Thus, taking the expectation, one gets

$$\mathbb{E}\left[W_p^p\left(\sum_{i:X_i\in Q}\delta_{X_i},\frac{n(Q)}{\rho(Q)}\rho\mathbb{1}_Q\right)\right]\le(1+C\varepsilon)|Q|^{\frac{p}{d}}\mathbb{E}\left[W_p^p\left(\sum_{i:X_i\in Q}\delta_{T^{-1}(S^{-1}(X_i))},n(Q)\mathbb{1}_{[0,1]^d}\right)\right]. \qquad (6.3.0.9)$$

Now we consider $\{Z_i\}_{i=1}^\infty$ iid uniform random variables on $[0,1]^d$. If we condition to $n(Q)$, by (2.2.0.3) of Lemma 2.2.1 $\{T^{-1}(S^{-1}(X_i))\}_{i:X_i\in Q}$ are independent and distributed on $[0,1]^d$ with probability distribution

$$(T^{-1}\circ S^{-1})_\sharp\left(\frac{\rho\mathbb{1}_Q}{\rho(Q)}\right)=\mathbb{1}_{[0,1]^d},\qquad\text{by (6.3.0.7).}$$

Thus

$$\begin{aligned}
\mathbb{E}\left[W_p^p\left(\sum_{i:X_i\in Q}\delta_{T^{-1}(S^{-1}(X_i))},n(Q)\mathbb{1}_{[0,1]^d}\right)\right]&=\mathbb{E}\left[\mathbb{E}\left[W_p^p\left(\sum_{i:X_i\in Q}\delta_{T^{-1}(S^{-1}(X_i))},n(Q)\mathbb{1}_{[0,1]^d}\right)\middle|n(Q)\right]\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[W_p^p\left(\sum_{i=1}^{n(Q)}\delta_{Z_i},n(Q)\mathbb{1}_{[0,1]^d}\right)\middle|n(Q)\right]\right]\\
&=\mathbb{E}\left[W_p^p\left(\sum_{i=1}^{n(Q)}\delta_{Z_i},n(Q)\mathbb{1}_{[0,1]^d}\right)\right]. \qquad (6.3.0.10)
\end{aligned}$$

Combining (6.3.0.9) and (6.3.0.10) one gets the thesis.
Now we go back to the step which we left aside before.

Proof of (6.3.0.7) of Step 1. Since the density is

$$\tilde\rho(x)=(S^{-1})_\sharp\left(\frac{\rho\mathbb{1}_Q}{\rho(Q)}\right)(x)=\frac{\rho\mathbb{1}_Q}{\rho(Q)}(S(x))\det DS(x),$$

by Lemma 2.2.6 such a map T exists and verifies

$$\sup_{x,y\in[0,1]^d}\frac{|T(x)-T(y)|}{|x-y|}\le1+f\left(L\sup_{z\in[0,1]^d}\frac{|\nabla\tilde\rho(z)|}{\tilde\rho(z)}\right),\qquad f(t)=C_dt(1+t)^d.$$

Thus the statement is proved if one proves that

$$\sup_{z\in[0,1]^d}\frac{|\nabla\tilde\rho(z)|}{\tilde\rho(z)}\le C\varepsilon,\quad\text{since we have already rescaled and therefore in this case } L=1, \qquad (6.3.0.11)$$

indeed in that case one would have

$$f\left(\sup_{z\in[0,1]^d}\frac{|\nabla\tilde\rho(z)|}{\tilde\rho(z)}\right)=f\left(L\sup_{z\in[0,1]^d}\frac{|\nabla\tilde\rho(z)|}{\tilde\rho(z)}\right)\le c\varepsilon\quad\text{and therefore}\quad\sup_{x,y\in[0,1]^d}\frac{|T(x)-T(y)|}{|x-y|}\le 1+c\varepsilon.$$

Thus we only have to prove (6.3.0.11). First, notice that for any $x\in[0,1]^d$

$$\frac{\nabla\tilde\rho(x)}{\tilde\rho(x)}=\frac{\nabla\left(\tilde\rho(S(x))\det DS(x)\right)}{\tilde\rho(S(x))\det DS(x)}=\frac{(DS(x))^T\nabla\rho(S(x))}{\rho(S(x))}+\frac{\nabla(\det DS(x))}{\det DS(x)},$$

hence

$$\frac{|\nabla\tilde\rho(x)|}{\tilde\rho(x)}\le\|DS(x)\|\frac{|\nabla\rho(S(x))|}{\rho(S(x))}+\frac{|\nabla(\det DS(x))|}{|\det(DS(x))|}. \tag{6.3.0.12}$$

Now, one has

$$\|DS(x)\|\le r_k\delta_k\left\|\frac{DS(x)}{r_k\delta_k}-R^yU\right\|+r_k\delta_k\|R^yU\|\le r_k\delta_k(1+C\delta_k)\le r_k\delta_k(1+\varepsilon)\le C\varepsilon r_k^{1-q}, \tag{6.3.0.13}$$

with

$$\sup_{z\in Q}\frac{|\nabla\rho(z)|}{\rho(z)}\le\sup_{z\in Q}C|z|^{q-1}\le Cr_{k+1}^{q-1}\le C'r_k^{q-1}, \tag{6.3.0.14}$$

and by (6.3.0.1)

$$\frac{|\nabla(\det DS(x))|}{|\det(DS(x))|}\le C\delta_k\le C\varepsilon. \tag{6.3.0.15}$$

Combining (6.3.0.13), (6.3.0.14) and (6.3.0.15) with (6.3.0.12) one gets

$$\frac{\nabla\tilde\rho(x)}{\tilde\rho(x)}\le C\varepsilon,\qquad\text{for all } x\in[0,1]^d,$$

hence (6.3.0.11) holds and therefore the proof is concluded. □

Proof of Lemma 6.3.2. The proof of the lower bound is quite similar to the proof of Lemma 6.3.1. The difference is that our purpose is now to apply (2.2.0.8) of Lemma 2.2.3 instead of (2.2.0.4) of Lemma 2.2.2, and therefore the same argument we used in that case works, up to finding a suitable map T such that, if $Q=Q_j$

$$\frac{1}{\rho(Q)}\rho\mathbb{1}_Q=T_\sharp\left(\frac{1}{|Q|}\mathbb{1}_Q\right),\quad T(\partial Q)=\partial Q, \tag{6.3.0.16}$$

and that

$$\frac{|T(x)-T(y)|}{|x-y|}\ge 1-\frac{c}{\ell}. \tag{6.3.0.17}$$

If we find such T, thanks to Lemma 2.2.3 we get

$$\begin{aligned}Wb_p^p\left(\sum_{i:X_i\in Q}\delta_{X_i},\frac{\mathbb{E}[n(Q)]}{\rho(Q)}\rho\mathbb{1}_Q\right)&=Wb_p^p\left(T_\sharp\left(\sum_{i:X_i\in Q}\delta_{T^{-1}(X_i)}\right),T_\sharp\left(\frac{\mathbb{E}[n(Q)]}{|Q|}\mathbb{1}_Q\right)\right)\\&\ge\left(1-\frac{c}{\ell}\right)^pWb_p^p\left(\sum_{i:X_i\in Q}\delta_{T^{-1}(X_i)},\frac{\mathbb{E}[n(Q)]}{|Q|}\mathbb{1}_Q\right),\end{aligned}$$

and taking the expectation we conclude with the same argument as the previous Lemma: conditioned to $n(Q)$, $\{T^{-1}(X_i)\}_{i:X_i\in Q}$ are uniformly iid on Q by (2.2.0.3) of Lemma 2.2.1. Then, applying the affine map which moves $Q_j=q_j+[0,L_j]^d$ to $[0,1]^d$,

$$S:Q=Q_j\to[0,1]^d, S(x)=\frac{x-q_j}{L_j},$$

by Corollary 2.2.1 we get

$$Wb_p^p\left(\sum_{i:X_i\in Q}\delta_{T^{-1}(X_i)},\frac{\mathbb{E}[n(Q)]}{|Q|}\mathbb{1}_Q\right)=Wb_p^p\left((S^{-1})_\sharp\left(\sum_{i:X_i\in Q}\delta_{S(T^{-1}(X_i))}\right),(S^{-1})_\sharp(\mathbb{E}[n(Q)]\mathbb{1}_{[0,1]^d})\right)$$
$$=L_j^pWb_p^p\left(\sum_{i:X_i\in Q}\delta_{S(T^{-1}(X_i))},\mathbb{E}[n(Q)]\mathbb{1}_{[0,1]^d}\right)$$
$$=|Q|^{\frac{p}{d}}Wb_p^p\left(\sum_{i:X_i\in Q}\delta_{S(T^{-1}(X_i))},\mathbb{E}[n(Q)]\mathbb{1}_{[0,1]^d}\right),$$

with $\{S(T^{-1}(X_i))\}_{i:X_i\in Q}$ uniformly iid on $[0,1]^d$ if conditioned to $n(Q)$.

To explain why such a map T satisfying (6.3.0.16) and (6.3.0.17) exists, let us assume that the hypercube we are considering by partition in Definition 6.2.6 is

$$Q=Q_j\subseteq Q^{k+1}\setminus Q^k.$$

In this case the side of the squares is

$$L=(x_{k+1}-x_k)=\frac{x_k}{\ell\lfloor x_k^q\rfloor}\leq\frac{C}{\ell}x_k^{1-q},$$

and for any $x\in Q_j$ one has

$$|x|\leq\sqrt{d}x_{k+1}=\sqrt{d}x_k\left(1+\frac{1}{\ell\lfloor x_k^q\rfloor}\right)\leq Cx_k,$$

hence

$$L\frac{|\nabla\rho(x)|}{\rho(x)}=L\frac{|\rho'(|x|)|}{\rho(|x|)}\leq cL|x|^{q-1}\leq\frac{C}{\ell}x_x^{1-q}x_k^{q-1}\leq\frac{C}{\ell}.$$

Hence by Lemma 2.2.6 such a map exists and verifies

$$\frac{|T(x)-T(y)|}{|x-y|}\geq 1-cL\frac{|\nabla\rho(x)|}{\rho(x)}\geq 1-\frac{c}{\ell},$$

and therefore the proof is concluded. □

6.4 Some useful formulations of previous results.

In this Section we formulate some preliminary results in a slightly different way with respect to the Introduction. To begin with, we focus on the Random Matching problem with a random number of iid points.

Lemma 6.4.1. *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of iid random variables with uniform distribution on $[0,1]^d$ and $c>0$ be a positive constant. Then, for any N binomial random variables independent of $\{X_i\}_{i=1}^{\infty}$ with parameters n and q with $nq\geq c$, it holds*

$$\mathbb{E}\left[W_p^p\left(\sum_{i=1}^N\delta_{X_i},N\mathbb{1}_{[0,1]^d}\right)\right]\leq(1+\omega(\mathbb{E}[N]))\tilde{c}_d^p(\mathbb{E}[N])^{1-\frac{p}{d}}\left(\chi_{d\geq3}+|\log(\mathbb{E}[N])|^{\frac{p}{2}}\chi_{d=2}\right),\tag{6.4.0.1}$$

$$\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^N\delta_{X_i},N\mathbb{1}_{(0,1)^d}\right)\right]\geq(1-\omega(\mathbb{E}[N]))\tilde{c}b_d^p(\mathbb{E}[N])^{1-\frac{p}{d}}\left(\chi_{d\geq3}+|\log(\mathbb{E}[N])|^{\frac{p}{2}}\chi_{d=2}\right),\tag{6.4.0.2}$$

where $\omega:[c,+\infty)\to[0,+\infty)$ is a function vanishing as the argument increases ($\omega<1$ in (6.4.0.2)) and $\tilde{c}_d^p$ and $\tilde{c}b_d^p$ are the constants defined in (1.1.1.7).

Proof. Since the proof of the two results is the same, we only prove (6.4.0.1). By conditioning with respect to N, we get $\{X_i\}_{i=1}^N$ iid random variables uniformly distributed on $[0,1]^d$. Hence, if ω,ω',ω'' and ω''' are functions with the same properties as the function ω in the statement, we have

$$\begin{aligned}
\mathbb{E}\left[W_p^p\left(\sum_{i=1}^N\delta_{X_i},N\mathbb{1}_{[0,1]^d}\right)\right]&=\mathbb{E}\left[\mathbb{E}\left[W_p^p\left(\sum_{i=1}^N\delta_{X_i},N\mathbb{1}_{[0,1]^d}\right)\middle|N\right]\right]\\
&\leq\mathbb{E}\left[(1+\omega(N))N^{1-\frac{p}{d}}\tilde{c}_d^p(\chi_{d\geq3}+(\log N)^{\frac{p}{2}}\chi_{d=2})\mathbb{1}_{N\geq1}\right]\quad\text{by (1.1.2.13)}\\
&=\mathbb{E}\left[(1+\omega'(N))(N+1)^{1-\frac{p}{d}}\tilde{c}_d^p(\chi_{d\geq3}+(\log(N+1))^{\frac{p}{2}}\chi_{d=2})\right]\\
&\leq\tilde{c}_d^p(1+\omega''(\mathbb{E}[N]))(\mathbb{E}[N]+1)^{1-\frac{p}{d}}(\chi_{d\geq3}+(\log(\mathbb{E}[N]+1))^{\frac{p}{2}}\chi_{d=2})\quad\text{by (3.2.0.23) of Theorem 3.2.1}\\
&=\tilde{c}_d^p(1+\omega'''(\mathbb{E}[N]))(\mathbb{E}[N])^{1-\frac{p}{d}}(\chi_{d\geq3}+|\log(\mathbb{E}[N])|^{\frac{p}{2}}\chi_{d=2}),
\end{aligned}$$

and therefore the proof of (6.4.0.1) is concluded. As we mentioned before, the proof of (6.4.0.2) is the same (substituting (3.2.0.23) with (3.2.0.24)), hence we omit it. $\square$

We now focus on an intermediate result about the boundary cost function.

Lemma 6.4.2. *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of iid random variables and $c > 0$ be a positive constant. Then, for any N binomial random variables independent of $\{X_i\}_{i=1}^{\infty}$ with parameters n and q with $nq \geq c$, it holds*

$$\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N}\delta_{X_i}, \mathbb{E}[N]\mathbb{1}_{(0,1)^d}\right)\right] \geq (1-\omega(\mathbb{E}[N]))\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N}\delta_{X_i}, N\mathbb{1}_{(0,1)^d}\right)\right],$$

where $\omega : [c, +\infty) \to [0, 1)$ is a function vanishing as the argument increases.

Before proving the Lemma, let us specify that also an upper bound could be proved (i.e., with $\leq (1+\omega(\mathbb{E}[N]))$, but since we only need the lower bound we leave it aside.

Proof. We have to prove that the quantity

$$\alpha = \alpha(n, q) := \frac{\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N}\delta_{X_i}, \mathbb{E}[N]\mathbb{1}_{(0,1)^d}\right)\right]}{\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N}\delta_{X_i}, N\mathbb{1}_{(0,1)^d}\right)\right]} \tag{6.4.0.3}$$

is bounded from below by a positive constant and that $\liminf_{np\to\infty} \alpha \geq 1$.

Step 1: we prove that

$$\liminf_{nq\to\infty} \alpha \geq 1.$$

To do that, we start by noticing that by triangle inequality we have

$$\left|Wb_p\left(\sum_{i=1}^{N}\delta_{X_i}, N\mathbb{1}_{(0,1)^d}\right) - Wb_p\left(\sum_{i=1}^{N}\delta_{X_i}, \mathbb{E}[N]\mathbb{1}_{(0,1)^d}\right)\right| \leq Wb_p\left(\mathbb{E}[N]\mathbb{1}_{(0,1)^d}, N\mathbb{1}_{(0,1)^d}\right).$$

Therefore, there exists a constant $c_p > 0$ depending only on p such that for any $\varepsilon \in (0, 1)$ it holds

$$\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N}\delta_{X_i}, \mathbb{E}[N]\mathbb{1}_{(0,1)^d}\right)\right] \geq (1-\varepsilon)\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N}\delta_{X_i}, N\mathbb{1}_{(0,1)^d}\right)\right] - \frac{c_p}{\varepsilon^{p-1}}\mathbb{E}\left[Wb_p^p\left(\mathbb{E}[N]\mathbb{1}_{(0,1)^d}, N\mathbb{1}_{(0,1)^d}\right)\right],$$

and moreover

$$\begin{aligned}
\mathbb{E}\left[Wb_p^p\left(\mathbb{E}[N]\mathbb{1}_{(0,1)^d}, N\mathbb{1}_{(0,1)^d}\right)\right] &\leq \mathbb{E}\left[\frac{|N-\mathbb{E}[N]|^p}{\max\{N, \mathbb{E}[N]\}^{p-1}}\right] \qquad \text{by Lemma 2.2.7} \\
&\leq \frac{\mathbb{E}[|N-\mathbb{E}[N]|^p]}{(\mathbb{E}[N])^{p-1}} \\
&\leq C(\mathbb{E}[N])^{1-\frac{p}{2}} \qquad \text{by Lemma 3.1.1.}
\end{aligned}$$

Combining these two estimates we get

$$\alpha \geq 1 - \varepsilon - \frac{c_p C}{\varepsilon^{p-1}} \frac{(\mathbb{E}[N])^{1-\frac{p}{2}}}{\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N}\delta_{X_i}, N\mathbb{1}_{(0,1)^d}\right)\right]}. \tag{6.4.0.4}$$

Now we can rearrange the term in the right-hand side of (6.4.0.4) as follows:

$$\begin{aligned}
\frac{(\mathbb{E}[N])^{1-\frac{p}{2}}}{\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N}\delta_{X_i}, N\mathbb{1}_{(0,1)^d}\right)\right]} &= \frac{(\mathbb{E}[N])^{1-\frac{p}{2}}}{(\mathbb{E}[N])^{1-\frac{p}{d}}\left(\chi_{d\geq 3} + |\log(\mathbb{E}[N])|^{\frac{p}{2}}\chi_{d=2}\right)} \frac{(\mathbb{E}[N])^{1-\frac{p}{d}}\left(\chi_{d\geq 3} + |\log(\mathbb{E}[N])|^{\frac{p}{2}}\chi_{d=2}\right)}{\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N}\delta_{X_i}, N\mathbb{1}_{(0,1)^d}\right)\right]} \\
&= (\mathbb{E}[N])^{-\frac{p(d-2)}{2d}}\left(\chi_{d\geq 3} + |\log(\mathbb{E}[N])|^{-\frac{p}{2}}\chi_{d=2}\right) \frac{(\mathbb{E}[N])^{1-\frac{p}{d}}\left(\chi_{d\geq 3} + |\log(\mathbb{E}[N])|^{\frac{p}{2}}\chi_{d=2}\right)}{\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N}\delta_{X_i}, N\mathbb{1}_{(0,1)^d}\right)\right]} \\
&\leq C(\mathbb{E}[N])^{-\frac{p(d-2)}{2d}}\left(\chi_{d\geq 3} + |\log(\mathbb{E}[N])|^{-\frac{p}{2}}\chi_{d=2}\right) \qquad \text{by (6.4.0.2) of Lemma 6.4.1} \\
&=: \omega(\mathbb{E}[N]),
\end{aligned} \tag{6.4.0.5}$$

where the function ω is vanishing as the argument increases.

Combining (6.4.0.4) and (6.4.0.5) we get

$$\alpha \geq 1 - \varepsilon - \frac{c_p C}{\varepsilon^{p-1}} \omega(\mathbb{E}[N]),$$

and we get $\liminf_{np\to\infty} \alpha \geq 1$ by taking

$$\varepsilon = \begin{cases} (\omega(\mathbb{E}[N]))^{\frac{1}{2(p-1)}}, & p > 1, \\ 0, & p = 1, \end{cases}$$

so that

$$\alpha \geq 1 - (\omega(\mathbb{E}[N]))^{\frac{1}{2(p-1)}} - c_p C\sqrt{\omega(\mathbb{E}[N])}.$$

Step 2: we prove that for any n and p such that $nq \geq c$, $\alpha(n,q) \geq \frac{1}{C}$. The constant C mentioned before depends on c. We first notice that, since we have just proved that it is asymptotic to 1 as $nq \to \infty$, we can assume that

$$np \in [c, M],$$

where M is large enough to have $1 - \omega^{\frac{1}{2(p-1)}}(M) - c_p C\sqrt{\omega(M)} \geq \frac{1}{2}$. Otherwise, if np is larger than such M, the statement has already been proved in the previous passage.

Since we already know that, up to a constant, the denominator in α is bounded from above from

$$\mathbb{E}[N]^{1-\frac{p}{d}} \left(\chi_{d\geq 3} + \chi_{d=2}(\log(\mathbb{E}[N]))^{\frac{p}{2}}\right) \leq C \quad \text{ for } \mathbb{E}[N] = np \leq M,$$

it is sufficient to prove that also the denominator is bounded from below from a positive constant if $nq \in [c, M]$. This is a consequence of the fact that, using Lemma 2.2.4 with

$$\lambda = \left(\frac{nq}{c}\right)^{\frac{1}{d}} \geq 1,$$

if $N_{n,c}$ is a binomial random variable with parameters n and $\frac{c}{n}$, we have

$$\begin{aligned}
\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N} \delta_{X_i}, np\mathbb{1}_{(0,1)^d}\right)\right] &\geq \left\lfloor \left(\frac{nq}{c}\right)^{\frac{1}{d}} \right\rfloor^d \left(\frac{c}{nq}\right)^{\frac{p}{d}} \mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N_{n,c}} \delta_{X_i}, c\mathbb{1}_{(0,1)^d}\right)\right] \\
&\geq \min_{x\in\left[1,\frac{M}{c}\right]} \left\lfloor x^{\frac{1}{d}} \right\rfloor^d x^{-\frac{p}{d}} \mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N_{n,c}} \delta_{X_i}, c\mathbb{1}_{(0,1)^d}\right)\right] \\
&=: C(c, M, p, d)\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N_{n,c}} \delta_{X_i}, c\mathbb{1}_{(0,1)^d}\right)\right], \qquad (6.4.0.6)
\end{aligned}$$

hence the numerator of α in (6.4.0.3) is bounded from below by the right-hand side of the previous inequality. Therefore we only have to prove that (6.4.0.6) is bounded from below uniformly for $n \in \mathbb{N}$. This is a consequence of the fact that, by Corollary 3.3.1, we have

$$(6.4.0.6) \xrightarrow[n\to\infty]{} \mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{M_c} \delta_{X_i}, c\mathbb{1}_{(0,1)^d}\right)\right], \qquad (6.4.0.7)$$

where M_c is a Poisson random variable with parameter c. Since the righ-hand side of (6.4.0.7) is positive, we can conclude that the numerator of α in (6.4.0.3) is uniformly bounded from below for $nq \in [c, M]$, and therefore the proof is concluded. □

We can now state the main result that we need to prove Theorems 1.2.1 and 1.2.2.

Proposition 6.4.1 ([2, 23, 11])**.** *Let $d \geq 2$, $\Omega \subseteq \mathbb{R}^d$ be a bounded connected domain with Lipschitz boundary and $\{X_i\}_{i=1}^\infty$ be a sequence of iid random variables distributed on Ω with common probability density ρ. Then, if $c > 0$ is a positive constant, for any N binomial random variables with parameters n and q such that $nq \geq c$, it holds*

$$\mathbb{E}\left[W_p^p\left(\sum_{i=1}^{N} \delta_{X_i}, N\rho\right)\right] \leq (1 + \omega(\mathbb{E}[N]))\tilde{c}_d^p(\mathbb{E}[N])^{1-\frac{p}{d}} \left(\chi_{d\geq 3} + |\log(\mathbb{E}[N])|^{\frac{p}{2}}\right) \int_\Omega \rho^{1-\frac{p}{d}}, \qquad (6.4.0.8)$$

$$\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N} \delta_{X_i}, \mathbb{E}[N]\rho\right)\right] \geq (1 - \omega(\mathbb{E}[N]))\tilde{c}b_d^p(\mathbb{E}[N])^{1-\frac{p}{d}} \left(\chi_{d\geq 3} + |\log(\mathbb{E}[N])|^{\frac{p}{2}}\right) \int_\Omega \rho^{1-\frac{p}{d}}, \qquad (6.4.0.9)$$

where $\omega : [c, +\infty) \to [0, +\infty)$ is a function vanishing as the argument increases ($\omega < 1$ in (6.4.0.9)).

Proof. The proof of (6.4.0.8) is completely analogue to the proof of (6.4.0.1) of Lemma 6.4.1, where one has to use the second inequality in (1.1.2.16) of Theorem 1.1.5 instead of (1.1.2.13).

The proof of (6.4.0.9) is a little bit more technical, since the result we proved which relate the case if deterministic and random number of points points are limited to the case of the uniform distribution. The proof uses the same techniques as [2, 23, 24], that is, the same technique of the proof of Theorem 1.2.2, so we only sketch it.

Step 0: restrict to a union of hypercupes. Let

$$\mathcal{J}_\varepsilon := \left\{k \in \mathbb{Z}^d : Q_k^\varepsilon := \varepsilon k + \varepsilon(0,1)^d \subseteq \Omega\right\}.$$

By (3.0.0.2), if

$$N(Q_k^\varepsilon) := \sum_{i=1}^N \mathbb{1}(X_i \in Q_k^\varepsilon),$$

is the number of variables in the hypercube Q_k^ε, $N(Q_k^\varepsilon)$ is a binomial random variable of parameters n and $q\rho(Q_k^\varepsilon)$, and therefore we have $\mathbb{E}[N(Q_k^\varepsilon)] = nq\rho(Q_k^\varepsilon) = \mathbb{E}[N]\rho(Q_k^\varepsilon)$. Hence, by Proposition 2.1.1

$$\begin{aligned}
\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^N \delta_{X_i}, \mathbb{E}[N]\rho\right)\right] &\geq \sum_{k\in\mathcal{J}_\varepsilon} \mathbb{E}\left[Wb_p^p\left(\sum_{i:X_i\in Q_k^\varepsilon} \delta_{X_i}, \mathbb{E}[N(Q_k^\varepsilon)]\frac{\rho\mathbb{1}_{(0,1)^d}}{\rho(Q_k^\varepsilon)}\right)\right] \\
&= \sum_{k\in\mathcal{J}_\varepsilon} \mathbb{E}\left[\mathbb{E}\left[Wb_p^p\left(\sum_{i:X_i\in Q_k^\varepsilon} \delta_{X_i}, \mathbb{E}[N(Q_k^\varepsilon)]\frac{\rho\mathbb{1}_{(0,1)^d}}{\rho(Q_k^\varepsilon)}\right)\middle| N(Q_k^\varepsilon)\right]\right].
\end{aligned}$$

Step 1: move $\{X_i\}_{i:X_i\in Q_k^\varepsilon}$ into variables uniformly distributed in $(0,1)^d$. We now apply two maps to the variables $\{X_i\}_{i:X_i\in Q_k^\varepsilon}$: the first one is T_k, the linear map moving Q_k^ε in $(0,1)^d$, that is

$$T_k : Q_k^\varepsilon \to (0,1)^d, \qquad x \mapsto \frac{x - \varepsilon k}{\varepsilon}.$$

To choose the second one we argue as follows: the variables $\{X_i\}_{i:X_i\in Q_k^\varepsilon}$, if conditioned to $N(Q_k^\varepsilon)$, have distribution $\frac{\rho\mathbb{1}_{Q_k^\varepsilon}}{\rho(Q_k^\varepsilon)}$. Hence, the variables $\{T_k(X_i)\}_{i:X_i\in Q_k^\varepsilon}$, if conditioned to $N(Q_k^\varepsilon)$, have distribution, $T_{k\sharp}\left(\frac{\rho\mathbb{1}_{Q_k^\varepsilon}}{\rho(Q_k^\varepsilon)}\right)$.

We first claim that the Hölderian constant of $T_{k\sharp}\left(\frac{\rho\mathbb{1}_{Q_k^\varepsilon}}{\rho(Q_k^\varepsilon)}\right)$ is smaller than ε^α for a constant. Indeed it holds

$$\begin{aligned}
\sup_{x,y\in(0,1)^d} \frac{\left|T_{k\sharp}\left(\frac{\rho\mathbb{1}_{Q_k^\varepsilon}}{\rho(Q_k^\varepsilon)}\right)(x) - T_{k\sharp}\left(\frac{\rho\mathbb{1}_{Q_k^\varepsilon}}{\rho(Q_k^\varepsilon)}\right)(y)\right|}{|x-y|^\alpha} &= \frac{\varepsilon^d}{\rho(Q_k^\varepsilon)} \sup_{x,y\in(0,1)^d} \frac{|\rho(\varepsilon x + \varepsilon k) - \rho(\varepsilon y + \varepsilon k)|}{|x-y|^\alpha} \\
&= \frac{\varepsilon^d}{\rho(Q_k^\varepsilon)}\varepsilon^\alpha \sup_{z,w\in Q_k^\varepsilon} \frac{|\rho(z) - \rho(w)|}{|z-w|^\alpha} \\
&\leq C\varepsilon^\alpha,
\end{aligned}$$

where in the inequality we used that

$$\rho(Q_k^\varepsilon) \geq \inf \rho |Q_k^\varepsilon| = \inf \rho\varepsilon^d, \qquad \text{and} \qquad \sup_{z,w\in Q_k^\varepsilon} \frac{|\rho(z)-\rho(w)|}{|z-w|^\alpha} \leq C,$$

since ρ is bounded from below by a positive constant and α-Hölder.

Therefore by Lemma 2.2.6 (and the remarks later) there exists another map $S_k : (0,1)^d \to (0,1)^d$ such that

$$T_{k\sharp}\left(\frac{\rho\mathbb{1}_{Q_k^\varepsilon}}{\rho(Q_k^\varepsilon)}\right) = S_{k\sharp}(\mathbb{1}_{(0,1)^d}), \qquad \text{and} \qquad \sup_{x,y\in(0,1)^d}\left|\frac{|S_k(x) - S_k(y)|}{|x-y|} - 1\right| \leq C\varepsilon^\alpha,$$

and, if conditioned to $N(Q_k^\varepsilon)$, $\{S_k^{-1}(T_k(X_i))\}_{i:X_i\in Q_k^\varepsilon}$ have distribution $(S_k^{-1}\circ T_k)_\sharp\left(\frac{\rho\mathbb{1}_{Q_k^\varepsilon}}{\rho(Q_k^\varepsilon)}\right) = \mathbb{1}_{(0,1)^d}$. Hence, if $Y_i := S_k^{-1}(T_k(X_i))$, applying first Lemma 2.2.2 and Corollary 2.2.1, and then deleting the conditioned expectation, we have

$$\begin{aligned}
\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^N \delta_{X_i}, \mathbb{E}[N]\rho\right)\right] &\geq \sum_{k\in\mathcal{J}_\varepsilon} \mathbb{E}\left[\mathbb{E}\left[Wb_p^p\left((T_k^{-1}\circ S_k)_\sharp\left(\sum_{i:X_i\in Q_k^\varepsilon}\delta_{Y_i}\right), \mathbb{E}[N(Q_k^\varepsilon)](T_k^{-1}\circ S_k)_\sharp(\mathbb{1}_{(0,1)^d})\right)\middle| N(Q_k^\varepsilon)\right]\right] \\
&\geq \sum_{k\in\mathcal{J}_\varepsilon} |Q_k^\varepsilon|^{\frac{p}{d}}(1-C\varepsilon^\alpha)\mathbb{E}\left[\mathbb{E}\left[Wb_p^p\left(\sum_{i:X_i\in Q_k^\varepsilon}\delta_{Y_i}, \mathbb{E}[N(Q_k^\varepsilon)]\mathbb{1}_{(0,1)^d}\right)\middle| N(Q_k^\varepsilon)\right]\right] \\
&= \sum_{k\in\mathcal{J}_\varepsilon} |Q_k^\varepsilon|^{\frac{p}{d}}(1-C\varepsilon^\alpha)\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{N(Q_k^\varepsilon)}\delta_{Z_i}, \mathbb{E}[N(Q_k^\varepsilon)]\mathbb{1}_{(0,1)^d}\right)\right]
\end{aligned}$$

with $\{Z_i\}_{i=1}^\infty$ iid uniformly distributed on $(0,1)^d$ and independent of $\{X_i\}_{i=1}^\infty$.

Step 2: conclusion. Now if we apply first Lemma 6.4.2 and then (6.4.0.2) of Lemma 6.4.1, since $\mathbb{E}[N(Q_k^\varepsilon)] = \mathbb{E}[N]\rho(Q_k^\varepsilon)$, we get

$$\begin{aligned}
&\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^N \delta_{X_i}, \mathbb{E}[N]\rho\right)\right] \\
&\geq (1-C\varepsilon^\alpha)\tilde{c}b_d^p \sum_{k\in\mathcal{J}_\varepsilon} |Q_k^\varepsilon|^{\frac{p}{d}}(1-\omega(\mathbb{E}[N(Q_k^\varepsilon)]))(\mathbb{E}[N(Q_k^\varepsilon)])^{1-\frac{p}{d}}\left(\chi_{d\geq 3} + |\log(\mathbb{E}[N(Q_k^\varepsilon)])|^{\frac{p}{2}}\chi_{d=2}\right) \\
&= (1-C\varepsilon^\alpha)(1-\omega(\mathbb{E}[N]\inf\rho\varepsilon^d))\tilde{c}b_d^p(\mathbb{E}[N])^{1-\frac{p}{d}} \sum_{k\in\mathcal{J}_\varepsilon}|Q_k^\varepsilon|\left(\frac{\rho(Q_k^\varepsilon)}{|Q_k^\varepsilon|}\right)^{\frac{p}{d}}\left(\chi_{d\geq 3} + |\log(\mathbb{E}[N]\rho(Q_k^\varepsilon))|^{\frac{p}{2}}\chi_{d=2}\right) \\
&\geq (1-C\varepsilon^\alpha)(1-\omega(\mathbb{E}[N]\inf\rho\varepsilon^d))(1-o(\varepsilon))\tilde{c}b_d^p(\mathbb{E}[N])^{1-\frac{p}{d}}\sum_{k\in\mathcal{J}_\varepsilon}\int_{Q_k^\varepsilon}\rho^{1-\frac{p}{d}}\left(\chi_{d\geq 3} + |\log(\mathbb{E}[N]\rho(Q_k^\varepsilon))|^{\frac{p}{2}}\chi_{d=2}\right),
\end{aligned}$$

where $o(\varepsilon)$ is a function vanishing as $\varepsilon\to 0$. Finally, dividing by $(\mathbb{E}[N])^{1-\frac{p}{d}}\left(\chi_{d\geq 3} + |\log(\mathbb{E}[N])|^{\frac{p}{2}}\chi_{d=2}\right)$, taking the liminf as $\mathbb{E}[N]\to\infty$ and then sending $\varepsilon\to 0$, we get the thesis. □

6.5 Proof of the upper bound.

Here we prove Theorem 1.2.1.

The strategy of the proof is the same as [7, 2], which we have explained in the Introduction, but we first have to constrain the points $\{X_i\}_{i=1}^n$ inside the ball of radius R_n. Then, since

$$B_{R_n} = B_{r_1} \cup \bigcup_{k\geq 1} \bigcup_{y\in\partial B_1} \bigcup_{U\in O(d-1)} Q^{y,U}_{r_k,\delta_k},$$

we have to rearrange the masses inside B_{r_1} and $Q^{y,U}_{r_k,\delta_k}$, which means that we have to estimate the global term in Theorem 6.2.1, and then we estimate the cost of the matching inside each one of these subsets and compare it with the cost of the uniform matching inside an hypercube, as in Lemma 6.3.1.

Proof. As we mentioned before, we start as follows.

Step 1: substitute $\{X_i\}_{i=1}^n \to \{\tilde{X}_i\}_{i=1}^n$. These new random variables are iid with common probability density $\rho_n = \rho^{\varepsilon}_{R_n}$ by Proposition 6.1.1. In other words, by triangle inequality we have

$$W_p\left(\sum_{i=1}^n \delta_{X_i}, n\rho\right) \leq W_p\left(\sum_{i=1}^n \delta_{X_i}, \sum_{i=1}^n \delta_{\tilde{X}_i}\right) + W_p\left(\sum_{i=1}^n \delta_{\tilde{X}_i}, n\rho_n\right) + W_p\left(n\rho_n, n\rho\right),$$

and therefore, taking the power p and then the expectation of the above expression, there exists a constant $C>0$ depending only on p such that for $\eta>0$ we have

$$\begin{aligned}
\mathbb{E}\left[W_p^p\left(\sum_{i=1}^n \delta_{X_i}, n\rho\right)\right] &\leq \frac{C}{\eta^{p-1}}\mathbb{E}\left[W_p\left(\sum_{i=1}^n \delta_{X_i}, \sum_{i=1}^n \delta_{\tilde{X}_i}\right)\right] + (1+\eta)\mathbb{E}\left[W_p^p\left(\sum_{i=1}^n \delta_{\tilde{X}_i}, n\rho_n\right)\right] + \frac{C}{\eta^{p-1}}W_p^p\left(n\rho_n, n\rho\right) \\
&\leq \frac{2C}{\eta^{p-1}} n W_p^p\left(\rho_n, \rho\right) \qquad (6.5.0.1) \\
&\quad + (1+\eta)\mathbb{E}\left[W_p^p\left(\sum_{i=1}^n \delta_{\tilde{X}_i}, n\rho_n\right)\right], \qquad (6.5.0.2)
\end{aligned}$$

where, in the second inequality, we have used first Proposition 6.1.1 and then the fact that the p^{th} power of the p-Wasserstein distance is homogeneous. By Proposition 6.1.2, we have

$$(6.5.0.1) \leq k(n) := C\begin{cases} (\log n)^{\frac{d}{q}-1-p(1-\frac{1}{q})+\alpha-\frac{\gamma}{q}} & q>0, \\ n^{\frac{p}{\gamma-d}}(\log n)^{\beta(1-\frac{p}{\gamma-d})} & q=0, \end{cases} \qquad (6.5.0.3)$$

hence to prove the thesis we focus on (6.5.0.2).

Step 2: the global estimate - rearrange the masses according to the decomposition. In this proof we use Notation 6.2.2 meaning for $n(\cdot)$ the number of variables $\tilde{X}_i$ in a domain.

To prove the thesis, we split the ball B_{R_n}, where the new $\{\tilde{X}_i\}_{i=1}^n$ are constrained and the new density ρ_n is concentrated. Using the radial-angular decomposition ((6.2.2.9) of Definition 6.2.3) in the case

$$q>0, d=2, 1\leq p<2,$$

and the angular-radial one ((6.2.2.10) of Definition 6.2.3) in all the other cases, by triangle inequality we have

$$\begin{aligned}
W_p\left(\sum_{i=1}^n \delta_{\tilde{X}_i}, n\rho_n\right) &\leq W_p\left(\sum_{i=1}^n \delta_{\tilde{X}_i}, \frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n \mathbb{1}_{B_{r_1}} + \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \sum_{k\geq 1}\frac{n(Q^{y,U}_{r_k,\delta})}{\mathcal{M}_{\delta_k}\rho_n(Q^{y,U}_{r_k,\delta})}\rho_n \mathbb{1}_{Q^{y,U}_{r_k,\delta}}\right) \\
&+ W_p\left(\frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n \mathbb{1}_{B_{r_1}} + \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \sum_{k\geq 1}\frac{n(Q^{y,U}_{r_k,\delta})}{\mathcal{M}_{\delta_k}\rho_n(Q^{y,U}_{r_k,\delta})}\rho_n \mathbb{1}_{Q^{y,U}_{r_k,\delta}}, n\rho_n\right),
\end{aligned}$$

and we make the same step as before: we take the p^{th} power and then we take the expectation, so that, if $g(n)$

is the quantity defined in Theorem 6.2.1 we get

$$\mathbb{E}\left[W_p^p\left(\sum_{i=1}^n \delta_{\tilde X_i}, n\rho_n\right)\right]$$

$$\leq (1+\eta)\mathbb{E}\left[W_p^p\left(\sum_{i=1}^n \delta_{\tilde X_i}, \frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n \mathbb{1}_{B_{r_1}} + \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \sum_{k\geq 1}\frac{n(Q^{y,U}_{r_k,\delta})}{\mathcal{M}_{\delta_k}\rho_n(Q^{y,U}_{r_k,\delta})}\rho_n \mathbb{1}_{Q^{y,U}_{r_k,\delta}}\right)\right] \tag{6.5.0.4}$$

$$+\frac{C}{\eta^{p-1}}\mathbb{E}\left[W_p^p\left(\frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n \mathbb{1}_{B_{r_1}} + \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \sum_{k\geq 1}\frac{n(Q^{y,U}_{r_k,\delta})}{\mathcal{M}_{\delta_k}\rho_n(Q^{y,U}_{r_k,\delta})}\rho_n \mathbb{1}_{Q^{y,U}_{r_k,\delta}}, n\rho_n\right)\right]. \tag{6.5.0.5}$$

This way, by Theorem 6.2.1, we have

$$(6.5.0.5) = g(n)\begin{cases} \ll \tau_n^{p,d} & p\leq p^{crit}, \\ \lesssim \tau_n^{p,d} & p > p^{crit}, \end{cases} \tag{6.5.0.6}$$

hence we focus only on (6.5.0.4): by (6.2.2.13) with $\mu=\sum_{i=1}^n \delta_{\tilde X_i}$ and $f=1$ one has

$$\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\sum_{k\geq 1}\frac{1}{\mathcal{M}_{\delta_k}}\sum_{i:\tilde X_i\in Q^{y,U}_{r_k,\delta_k}}\delta_{\tilde X_i} = \sum_{i:|\tilde X_i|\geq r_1}\delta_{\tilde X_i},$$

hence

$$\sum_{i=1}^n \delta_{\tilde X_i} = \sum_{i:\tilde X_i\in B_{r_1}}\delta_{\tilde X_i} + \sum_{i:|\tilde X_i|\geq r_1}\delta_{\tilde X_i} = \sum_{i:\tilde X_i\in B_{r_1}}\delta_{\tilde X_i} + \int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\sum_{k\geq 1}\frac{1}{\mathcal{M}_{\delta_k}}\sum_{i:\tilde X_i\in Q^{y,U}_{r_k,\delta_k}}\delta_{\tilde X_i},$$

and therefore, by subadditivity (see (2.1.0.2))

$$(6.5.0.4) \leq \mathbb{E}\left[W_p^p\left(\sum_{i:\tilde X_i\in B_{r_1}}\delta_{\tilde X_i}, \frac{n(B_{r_1})}{\rho_n(B_{r_1})}\rho_n\mathbb{1}_{B_{r_1}}\right)\right] \tag{6.5.0.7}$$

$$+\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\sum_{k\geq 1}\frac{1}{\mathcal{M}_{\delta_k}}\mathbb{E}\left[W_p^p\left(\sum_{i:\tilde X_i\in Q^{y,U}_{r_k,\delta_k}}\delta_{\tilde X_i}, \frac{n(Q^{y,U}_{r_k,\delta_k})}{\rho_n(Q^{y,U}_{r_k,\delta_k})}\rho_n\mathbb{1}_{Q^{y,U}_{r_k,\delta_k}}\right)\right], \tag{6.5.0.8}$$

and now we consider the two different cases $d\geq 3$ and $d=2$, and we begin with the first one.

Step 3A: conclusion in the case $d\geq 3$. Thanks to (1.1.2.1) one can estimate (6.5.0.7): indeed, if we condition to $n(B_{r_1})$, $\{\tilde X_i\}_{i:\tilde X_i\in B_{r_1}}$ are iid with common distribution $\frac{\rho_n \mathbb{1}_{B_{r_1}}}{\rho_n(B_{r_1})}$ (which does not depend on n), and therefore

$$(6.5.0.7) \leq (1+\omega(n))\tilde c_d^p n^{1-\frac{p}{d}}\int_{B_{r_1}} dx\rho_n(x)^{1-\frac{p}{d}} \quad \text{by (6.4.0.8) of Proposition 6.4.1},$$

and using Lemma 6.3.1 one can estimate (6.5.0.8), so that, if $\{Z_i\}_{i=1}^\infty$ are iid uniformly distributed on $[0,1]^d$, one has

$$(6.5.0.8) \leq (1+C\varepsilon)\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\sum_{k\geq 1}\frac{1}{\mathcal{M}_{\delta_k}}|Q^{y,U}_{r_k,\delta_k}|^{\frac{p}{d}}\mathbb{E}\left[W_p^p\left(\sum_{i=1}^{n(Q^{y,U}_{r_k,\delta_k})}\delta_{Z_i}, n(Q^{y,U}_{r_k,\delta_k})\mathbb{1}_{[0,1]^d}\right)\right],$$

and since

$$\mathbb{E}\left[W_p^p\left(\sum_{i=1}^{n(Q^{y,U}_{r_k,\delta_k})}\delta_{Z_i}, n(Q^{y,U}_{r_k,\delta_k})\mathbb{1}_{[0,1]^d}\right)\right] \leq \tilde c_d^p(\mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})])^{1-\frac{p}{d}}(1+\omega(\mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})])) \quad \text{by (6.4.0.1) of Lemma 6.4.1}$$

one has

$$(6.5.0.8) \leq \tilde c_d^p(1+C\varepsilon)\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\sum_{k\geq 1}\frac{1}{\mathcal{M}_{\delta_k}}|Q^{y,U}_{r_k,\delta_k}|^{\frac{p}{d}}(1+\omega(\mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})]))(\mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})])^{1-\frac{p}{d}}$$

$$= \tilde c_d^p(1+C\varepsilon)(1+\omega(\ell_n))\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU\sum_{k\geq 1}\frac{1}{\mathcal{M}_{\delta_k}}|Q^{y,U}_{r_k,\delta_k}|^{\frac{p}{d}}n^{1-\frac{p}{d}}\left(\int_{Q^{y,U}_{r_k,\delta_k}} dx\rho_n(x)\right)^{1-\frac{p}{d}},$$

with

$$\ell_n := \varepsilon^d \begin{cases} (\log n)^{\alpha - \frac{\gamma}{q} - d(1-\frac{1}{q})} & q > 0, \\ (\log n)^{\beta} & q = 0, \end{cases} \simeq n(R_n\delta_n)^d \rho_n(R_n) \leq \mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})].$$

Now, looking at the last term, since, by (6.1.0.3) for any $x, y \in Q^{y,U}_{r_k,\delta_k}$ one has

$$\frac{\rho_n(x)}{\rho_n(y)} \leq 1 + C\varepsilon,$$

if $\bar{x}$ is any point in $Q^{y,U}_{r_k,\delta_k}$, and it holds

$$\begin{aligned} \left(\int_{Q^{y,U}_{r_k,\delta_k}} dx \rho_n(x)\right)^{1-\frac{p}{d}} &\leq (1+C\varepsilon)\rho_n(\bar{x})^{1-\frac{p}{d}} |Q^{y,U}_{r_k,\delta_k}|^{1-\frac{p}{d}} = (1+C\varepsilon)|Q^{y,U}_{r_k,\delta_k}|\rho_n(\bar{x})^{1-\frac{p}{d}}|Q^{y,U}_{r_k,\delta_k}|^{-\frac{p}{d}} \\ &\leq (1+C\varepsilon)\int_{Q^{y,U}_{r_k,\delta_k}} dx \rho_n(x)^{1-\frac{p}{d}} |Q^{y,U}_{r_k,\delta_k}|^{-\frac{p}{d}}, \end{aligned}$$

and therefore

$$\begin{aligned} (6.5.0.8) &\leq \tilde{c}^p_d n^{1-\frac{p}{d}}(1+C\varepsilon)(1+\omega(\ell_n))\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(d-1)} dU \sum_{k\geq 1}\frac{1}{\mathcal{M}_{\delta_k}}\int_{Q^{y,U}_{r_k,\delta_k}} dx \rho_n(x)^{1-\frac{p}{d}} \\ &= \tilde{c}^p_d n^{1-\frac{p}{d}}(1+C\varepsilon)(1+\omega(\ell_n))\int_{B_{R_n}\setminus B_{r_1}} dx \rho_n(x)^{1-\frac{p}{d}} \quad \text{by (6.2.2.13) with } d\mu = dx \text{ and } f = \rho_n^{1-\frac{p}{d}} \\ &= \tilde{c}^p_d n^{1-\frac{p}{d}}(1+C\varepsilon)(1+\omega(\ell_n))\int_{B_{R_n}\setminus B_{r_1}} dx \rho_n(x)^{1-\frac{p}{d}}. \end{aligned}$$

Thus, combining the previous estimates on (6.5.0.7) and (6.5.0.8) with (6.5.0.3) and (6.5.0.6), we have

$$\mathbb{E}\left[W^p_p\left(\sum_{i=1}^n \delta_{\tilde{X}_i}, n\rho_n\right)\right] \leq \tilde{c}^p_d(1+C\eta+C\varepsilon+\omega(\ell_n))n^{1-\frac{p}{d}}\int_{B_{R_n}} dx\rho_n(x)^{1-\frac{p}{d}} + \frac{C}{\eta^{p-1}}(k(n)+g(n)),$$

with k and g both depending on ε.

Now since if $p > p^{crit}$, by definition of $\tau^{p,d}_n$ in (1.2.0.3) we have $k(n) = O(\tau^{p,d}_n)$, $g(n) = O(\tau^{p,d}_n)$ and $\ell_n = O(\varepsilon)$, hence we can only conclude that

$$\mathbb{E}\left[W^p_p\left(\sum_{i=1}^n \delta_{\tilde{X}_i}, n\rho_n\right)\right] \leq C\tau^{p,d}_n,$$

while if $p \leq p^{crit}$ we have $k(n) = o(\tau^{p,d}_n)$, $g(n) = o(\tau^{p,d}_n)$ and $\ell_n \xrightarrow[n\to\infty]{} \infty$, hence if we divide both the left-hand side and the right-hand side by $\tau^{p,d}_n$, we get

$$\limsup_{n\to\infty}\frac{\mathbb{E}\left[W^p_p\left(\sum_{i=1}^n \delta_{\tilde{X}_i}, n\rho_n\right)\right]}{\tau^{p,d}_n} \leq (1+C\varepsilon+C\eta)\tilde{c}^p_d \lim_{n\to\infty}\frac{n^{1-\frac{p}{d}}\int_{B_{R_n}} dx\rho_n(x)^{1-\frac{p}{d}}}{\tau^{p,d}_n},$$

and the thesis follows observing that

$$\lim_{n\to\infty}\frac{n^{1-\frac{p}{d}}\int_{B_{R_n}} dx\rho_n(x)^{1-\frac{p}{d}}}{\tau^{p,d}_n} = \lim_{n\to\infty}\frac{n^{1-\frac{p}{d}}\int_{B_{R_n}} dx\rho(x)^{1-\frac{p}{d}}}{\tau^{p,d}_n},$$

and then sending ε and η to 0.

Step 3B: conclusion in the case $d = 2$. In this case the conclusion follows by the same steps as before if one estimates (6.5.0.7) and (6.5.0.8), since $k(n)$ by (6.5.0.3) and $g(n)$ by (6.5.0.6) give a contribution which is negligible for $p \leq p^{critc}$ and has the same order as $\tau^{p,d}_n$ if $p > p^{crit}$, exactly as in the case $d \geq 3$.

Hence, let us focus on (6.5.0.7) and (6.5.0.8).

In this case we have the logarithmic term, that is, as for (6.5.0.7), by (6.4.0.8) of Proposition 6.4.1 one has

$$\begin{aligned} (6.5.0.7) &\leq \tilde{c}^p_2(1+\omega(\mathbb{E}[n(B_{r_1})]))\mathbb{E}[n(B_{r_1})]^{1-\frac{p}{2}}(\log\mathbb{E}[n(B_{r_1})])^{\frac{p}{2}}\int_{B_{r_1}} dx\left(\frac{\rho_n(x)}{\rho_n(B_{r_1})}\right)^{1-\frac{p}{2}} \\ &= \tilde{c}^p_2(1+\omega(n))n^{1-\frac{p}{2}}(\log n)^{\frac{p}{2}}\int_{B_{r_1}} dx\rho_n(x)^{1-\frac{p}{2}} \quad \text{since } \mathbb{E}[n(B_{r_1})] = O(n), \end{aligned} \tag{6.5.0.9}$$

while for (6.5.0.8), if $\{Z_i\}_{i=1}^{\infty}$ are iid uniformly distributed on $[0,1]^2$, with the same passages as before we have

$$\begin{aligned}(6.5.0.8) &\leq (1+C\varepsilon)\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(1)} dU \sum_{k\geq 1}\frac{1}{\mathcal{M}_{\delta_k}}|Q^{y,U}_{r_k,\delta_k}|^{\frac{p}{2}}\mathbb{E}\left[W_p^p\left(\sum_{i=1}^{n(Q^{y,U}_{r_k,\delta_k})}\delta_{Z_i}, n(Q^{y,U}_{r_k,\delta_k})\mathbb{1}_{[0,1]^2}\right)\right]\\ &\leq \tilde{c}_2^p n^{1-\frac{p}{2}}(1+C\varepsilon)(1+\omega(\ell_n))\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(1)} dU\sum_{k\geq 1}\frac{1}{\mathcal{M}_{\delta_k}}\int_{Q^{y,U}_{r_k,\delta_k}} dx\rho_n(x)^{1-\frac{p}{2}}\log(1+\mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})])^{\frac{p}{2}},\end{aligned} \tag{6.5.0.10}$$

with ℓ_n defined as before.

Now we study the integral in (6.5.0.10) in some different cases.

Case $p < p^{crit}$ and $d = 2$. Here one can use the trivial estimate

$$\log(1+\mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})]) \leq (1+\omega(n))\log n,$$

and therefore, putting this estimate into (6.5.0.10), the same steps as before in this case prove that

$$\begin{aligned}(6.5.0.7)+(6.5.0.8) &\leq \tilde{c}_2^p(1+\omega(n))n^{1-\frac{p}{2}}(\log n)^{\frac{p}{2}}\int_{B_{r_1}} dx\rho_n(x)^{1-\frac{p}{2}}\\ &\quad+\tilde{c}_2^p(1+C\varepsilon+\omega(\ell_n))n^{1-\frac{p}{2}}(\log n)^{\frac{p}{2}}\int_{B_{R_n}\setminus B_{r_1}} dx\rho_n(x)^{1-\frac{p}{2}}\\ &=\tilde{c}_2^p(1+C\varepsilon+\omega(\ell_n))n^{1-\frac{p}{2}}(\log n)^{\frac{p}{2}}\int_{B_{R_n}} dx\rho_n(x)^{1-\frac{p}{2}},\end{aligned}$$

and this concludes the proof in this case.

Then, if $p = p^{crit}$ we argue in the two different cases $q > 0$ and $q = 0$.

Case $q > 0$ and $p = p^{crit} = d = 2$. In this case we can write

$$\begin{aligned}\log(1+\mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})])^{\frac{p}{2}} &= (1+\omega(\mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})])\log(\mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})])\\ &=(1+\omega(n))\left(\log\left(\frac{n}{|Q^{y,U}_{r_k,\delta_k}|}\int_{Q^{y,U}_{r_k,\delta_k}} dx\rho_n(x)\right)+\log|Q^{y,U}_{r_k,\delta_k}|\right)\\ &\leq(1+\omega(n))\left(\log(n\rho_n(r_k))+\varepsilon+C|\log(\varepsilon r_k)|\right) \qquad \text{by (6.1.0.3)},\end{aligned}$$

and therefore the integral in (6.5.0.10) writes as

$$\begin{aligned}&\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(1)} dU\sum_{k\geq 1}\frac{1}{\mathcal{M}_{\delta_k}}\int_{Q^{y,U}_{r_k,\delta_k}} dx\rho_n(x)^{1-\frac{p}{2}}\log(\mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})])^{\frac{p}{2}}\\ &\leq\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(1)} dU\sum_{k\geq 1}\frac{1}{\mathcal{M}_{\delta_k}}\int_{Q^{y,U}_{r_k,\delta_k}} dx[\log(n\rho_n(r_k))+\varepsilon+C|\log(\varepsilon r_k)|]\\ &\leq(1+C\varepsilon)\int_{\partial B_1} d\mathcal{H}^{d-1}(y)\int_{O(1)} dU\sum_{k\geq 1}\frac{1}{\mathcal{M}_{\delta_k}}\int_{Q^{y,U}_{r_k,\delta_k}} dx[\log(n\rho_n(x))+\varepsilon+C|\log(\varepsilon|x|)|] \qquad \text{using (6.1.0.3) again}\\ &=(1+C\varepsilon)\int_{B_{R_n}\setminus B_{r_1}} dx\log(n\rho_n(x))+|B_{R_n}|(\varepsilon+C\log(\varepsilon R_n)) \qquad \text{by (6.2.2.13)},\end{aligned}$$

and since only the first one of these contributions has the same order as $\tau_n^{p,d} = (\log n)^{1+\frac{2}{q}}$, we get

$$\begin{aligned}(6.5.0.8) &\leq \tilde{c}_2^{p=2}(1+C\varepsilon+\omega(\ell_n))\int_{B_{R_n}\setminus B_{r_1}} dx\log(n\rho_n(x))(1+\omega(n))\\ &=\tilde{c}_2^2(1+C\varepsilon+\omega(\ell_n))\int_{B_{R_n}} dx\log(n\rho_n(x))(1+\omega(n)),\end{aligned}$$

which proves the thesis noticing that in this case (6.5.0.9) proves that

$$(6.5.0.7) \ll \tau_n^{p,d}.$$

Therefore the conclusion follows by the same steps as before.

Case $q=0$, $d=2$ and $p=p^{crit}=\frac{d(\gamma-d)}{\gamma}=\frac{2(\gamma-2)}{\gamma}$. We have to estimate (6.5.0.10): to do that, we can notice that for any $x\in Q^{y,U}_{r_k,\delta_k}$

$$\begin{aligned}\log(1+\mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})])&=(1+\omega(\mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})]))\log(\mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})])\\&\le(1+\omega(n))\log\left(n\int_{Q^{y,U}_{r_k,\delta_k}}dy\rho(y)\right)\\&\le(1+\omega(n))\log(n\rho_n(x)|Q^{y,U}_{r_k,\delta_k}|(1+C\varepsilon))\\&\le\log(n\rho(x)(\delta_kr_k)^2(1+C\varepsilon))\le\log(n\rho(x)|x|^2\varepsilon(1+C\varepsilon))\le C+\log(n\rho(x)|x|^2),\end{aligned}$$

and putting this estimate into (6.5.0.10), since $\frac{p^*}{2}<1$ one gets

$$\begin{aligned}(6.5.0.8)&\le\tilde{c}_2^{p^*}n^{1-\frac{p^*}{2}}(1+C\varepsilon)(1+\omega(\ell_n))\int_{\partial B_1}d\mathcal{H}^1(y)\int_{O(1)}dU\sum_{k\ge1}\frac{1}{\mathcal{M}_{\delta_k}}\int_{Q^{y,U}_{r_k,\delta_k}}dx\rho_n(x)^{1-\frac{p^*}{2}}(C+\log(n\rho(x)|x|^2))^{\frac{p^*}{2}}\\&\le\tilde{c}_2^{p^*}n^{1-\frac{p^*}{2}}(1+C\varepsilon+\omega(\ell_n))\int_{\partial B_1}d\mathcal{H}^1(y)\int_{O(1)}dU\sum_{k\ge1}\frac{1}{\mathcal{M}_{\delta_k}}\int_{Q^{y,U}_{r_k,\delta_k}}dx\rho_n(x)^{1-\frac{p^*}{2}}[(\log(n\rho(x)|x|^2))^{\frac{p^*}{2}}+C^{\frac{p^*}{2}}]\\&=\tilde{c}_2^{p^*}n^{1-\frac{p^*}{2}}(1+C\varepsilon+\omega(\ell_n))\int_{\partial B_1}d\mathcal{H}^1(y)\int_{O(1)}dU\sum_{k\ge1}\frac{1}{\mathcal{M}_{\delta_k}}\int_{Q^{y,U}_{r_k,\delta_k}}dx\rho_n(x)^{1-\frac{p^*}{2}}(\log(n\rho(x)|x|^2))^{\frac{p^*}{2}}\\&\quad+o(\tau_n^{p,d})\\&=c_2^{p^*}n^{1-\frac{p^*}{2}}(1+C\varepsilon+\omega(\ell_n))\int_{B_{R_n}\setminus B_{r_1}}dx\rho(x)^{1-\frac{p^*}{2}}(\log(n\rho(x)|x|^2))^{\frac{p^*}{2}}+\omega(\tau_n^{p,d})\quad\text{by (6.2.2.13)},\end{aligned}$$

and also in this case one has

$$(6.5.0.7)=o(\tau_n^{p,d}),$$

hence the thesis is proved.

Case $d=2$ and $p>p^{crit}$. Here we do not care about the constants, since one already has $k(n)$ by (6.5.0.3) and $g(n)$ by (6.5.0.6) which gives two contributions with order $\tau_n^{p,d}$. Moreover, by (6.5.0.9) it is straightforward to see that

$$(6.5.0.7)=o(\tau_n^{p,d}),$$

and therefore only (6.5.0.8) is to be bounded, starting with the estimate in (6.5.0.10). Since for any $x\in Q^{y,U}_{r_k,\delta_k}$ one has

$$\mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})]=n\int_{Q^{y,U}_{r_k,\delta_k}}dx\rho_n(x)\le Cn|Q^{y,U}_{r_k,\delta_k}|\rho(r_k)\le C\varepsilon^dnr_k^{2(1-q)}\rho_n(r_k)\le Cn|x|^{2(1-q)}\rho_n(x),$$

by (6.1.0.3), the integral in that inequality satisfies

$$\begin{aligned}&\int_{\partial B_1}d\mathcal{H}^{d-1}(y)\int_{O(1)}dU\sum_{k\ge1}\frac{1}{\mathcal{M}_{\delta_k}}\int_{Q^{y,U}_{r_k,\delta_k}}dx\rho_n(x)^{1-\frac{p}{2}}\log(1+\mathbb{E}[n(Q^{y,U}_{r_k,\delta_k})])^{\frac{p}{2}}\\&\le C\int_{\partial B_1}d\mathcal{H}^{d-1}(y)\int_{O(1)}dU\sum_{k\ge1}\frac{1}{\mathcal{M}_{\delta_k}}\int_{Q^{y,U}_{r_k,\delta_k}}dx\rho_n(x)^{1-\frac{p}{2}}\left|\log(Cn|x|^{2(1-q)}\rho_n(x))\right|^{\frac{p}{2}}\\&=C\int_{B_{R_n}\setminus B_{r_1}}dx\rho_n(x)^{1-\frac{p}{2}}\left|\log(Cn|x|^{2(1-q)}\rho_n(x))\right|^{\frac{p}{2}}\qquad\text{by (6.2.2.13)}\\&=C\mathcal{H}^1(\partial B_1)\int_{r_1}^{R_n}drr\rho_n(r)^{1-\frac{p}{2}}\left|\log\left(C'\frac{r^{2(1-q)}\rho_n(r)}{R_n^{2(1-q)}\rho_n(R_n)}\right)\right|^{\frac{p}{2}}\qquad\text{since }\frac{1}{n}=R_n^{2(1-q)}\rho_n(R_n)\\&=C\mathcal{H}^1(\partial B_1)\int_{r_1}^{R_n}drr\rho_n(r)^{1-\frac{p}{2}}\left||\log C'|-(2(1-q)-\gamma)\log\left(\frac{R_n}{r}\right)+\frac{R_n^q-r^q}{q}\right|^{\frac{p}{2}},\end{aligned}\tag{6.5.0.11}$$

and now in (6.5.0.11) we have three terms: we claim that they have all order $\tau_n^{p,d}$ and this concludes.
The first one has order $\tau_n^{p,d}$ by denfinition, and the second one too since

$$\int_{r_1}^{R_n}drr\rho_n(r)^{1-\frac{p}{2}}\left[\log\left(\frac{R_n}{r}\right)\right]^{\frac{p}{2}}=e^{\frac{R_n^q}{q}(\frac{p}{2}-1)}\int_{r_1}^{R_n}drre^{\frac{R_n^q-r^q}{q}(\frac{p}{2}-1)}r^{\gamma(\frac{p}{2}-1)}\left[\log\left(\frac{R_n}{r}\right)\right]^{\frac{p}{2}},$$

and if $q=0$ the exponential term disappears and recalling that $p^*=\frac{2(\gamma-2)}{\gamma}$, one gets

$$\begin{aligned}\int_{r_1}^{R_n} drr\rho_n(r)^{1-\frac{p}{2}}\left[\log\left(\frac{R_n}{r}\right)\right]^{\frac{p}{2}} &= \int_{r_1}^{R_n} drr^{1+\gamma(\frac{p}{2}-1)}\left[\log\left(\frac{R_n}{r}\right)\right]^{\frac{p}{2}}\\ &= R_n^{2+\gamma(\frac{p}{2}-1)}\int_1^{\frac{R_n}{r_1}} dss^{-1-2\frac{p-p^*}{2-p^*}}(\log s)^{\frac{p}{2}} \qquad \text{with } s=\frac{R_n}{r}\\ &\le CR_n^{2+\gamma(\frac{p}{2}-1)}\le C\tau_n^{p,d},\end{aligned}$$

since the logarithmic term does not affect the integral, while if $q>0$, changing variables $s=\frac{R_n^q-r^q}{q}$ one gets

$$\begin{aligned}\int_{r_1}^{R_n} drr\rho_n(r)^{1-\frac{p}{2}}\left[\log\left(\frac{R_n}{r}\right)\right]^{\frac{p}{2}} &\le Ce^{\frac{R_n^q}{q}(\frac{p}{2}-1)}\int_{r_1}^{R_n} drr(e^{\frac{R_n^q-r^q}{q}}r^{-\gamma})^{1-\frac{p}{2}}\left[\log\left(\frac{R_n}{r}\right)\right]^{\frac{p}{2}}\\ &= Ce^{\frac{R_n^q}{q}(\frac{p}{2}-1)}\int_0^{\frac{R_n^q-r_1^q}{q}} dse^{-s(\frac{p}{2}-1)}(R_n^q-sq)^{\frac{2}{q}-1+\frac{\gamma}{q}(\frac{p}{2}-1)}\left[\log\left(\frac{R_n}{(R_n^q-sq)^{\frac{1}{q}}}\right)\right]^{\frac{p}{2}},\end{aligned}$$

where one can use the trivial estimates

$$(R_n^q-sq)^{\frac{2}{q}-1+\frac{\gamma}{q}(\frac{p}{2}-1)}\le (R_n^q)^{\frac{2}{q}-1+\frac{\gamma}{q}(\frac{p}{2}-1)}=R_n^{2-q+\gamma(\frac{p}{2}-1)},$$

and

$$R_n\ge (r_1^q+sq)^{\frac{1}{q}}\Rightarrow \log\left(\frac{R_n}{(R_n^q-sq)^{\frac{1}{q}}}\right)\le \log\left(\frac{(r_1^q+sq)^{\frac{1}{q}}}{r_1}\right)\le C+C|\log s|,$$

to obtain

$$\int_{r_1}^{R_n} drr\rho_n(r)^{1-\frac{p}{2}}\left[\log\left(\frac{R_n}{r}\right)\right]^{\frac{p}{2}}\le Ce^{\frac{R_n^q}{q}(\frac{p}{2}-1)}R_n^{2-q+\gamma(\frac{p}{2}-1)}\int_0^{\frac{R_n^q-r_1^q}{q}} dse^{-s(\frac{p}{2}-1)}(1+|\log s|)^{\frac{p}{2}}\le C\tau_n^{p,d}.$$

The third term in (6.5.0.11) appears only in the case $q>0$ and can be estimated as before, indeed changing again variables $s=\frac{R_n^q-r^q}{q}$ one gets

$$\begin{aligned}\int_{r_1}^{R_n} drr\rho_n(r)^{1-\frac{p}{2}}\left[\frac{R_n^q-r^q}{q}\right]^{\frac{p}{2}} &\le Ce^{\frac{R_n^q}{q}(\frac{p}{2}-1)}\int_{r_1}^{R_n} drr(e^{\frac{R_n^q-r^q}{q}}r^{-\gamma})^{1-\frac{p}{2}}\left[\frac{R_n^q-r^q}{q}\right]^{\frac{p}{2}}\\ &= Ce^{\frac{R_n^q}{q}(\frac{p}{2}-1)}\int_0^{\frac{R_n^q-r_1^q}{q}} dse^{-s(\frac{p}{2}-1)}(R_n^q-sq)^{\frac{2}{q}-1+\frac{\gamma}{q}(\frac{p}{2}-1)}s^{\frac{p}{2}}\\ &\le Ce^{\frac{R_n^q}{q}(\frac{p}{2}-1)}R_n^{2-q+\gamma(\frac{p}{2}-1)}\int_0^{\frac{R_n^q-r_1^q}{q}} dse^{-s(\frac{p}{2}-1)}s^{\frac{p}{2}}\\ &\le Ce^{\frac{R_n^q}{q}(\frac{p}{2}-1)}R_n^{2-q+\gamma(\frac{p}{2}-1)}\le C\tau_n^{p,d},\end{aligned}$$

and therefore also the third term eventually appearing in (6.5.0.11) has order $\tau_n^{p,d}$. □

6.6 Proof of the lower bound.

Here we prove Theorem 1.2.2.

Proof. As a disclaimer, we specify that this proof we mean by $n(\cdot)$ the number of variables $\{X_i\}_{i=1}^n$ in a domain (and not $\{\tilde{X}_i\}_{i=1}^n$ as in Notation 6.2.2).

The lower bound is much easier than the upper bound, indeed we do not have to constrain the points inside B_{R_n} and we do not have to care about the global term because in the boundary Wasserstein distance the measures are allowed to have different masses. In fact, if we recall the geometric decomposition we defined for the lower bound, thanks to the superadditivity of the boundary Wasserstein distance (see Lemma 2.1.1), we can infer that

$$Wb_p^p\left(\sum_{i=1}^n \delta_{X_i}, n\rho\right) = Wb_p^p\left(\sum_{j\in\mathbb{N}}\sum_{i:X_i\in Q_j} \delta_{X_i}, n\rho\sum_{j\in\mathbb{N}} \mathbb{1}_{Q_j}\right) \geq \sum_{j\in\mathbb{N}} Wb_p^p\left(\sum_{i:X_i\in Q_j}\delta_{X_i}, n\rho\mathbb{1}_{Q_j}\right)$$

$$\geq \sum_{j\in\mathcal{J}} Wb_p^p\left(\sum_{i:X_i\in Q_j}\delta_{X_i}, n\rho\mathbb{1}_{Q_j}\right) \qquad \text{with } \mathcal{J} \text{ by Definition 6.2.6}$$

$$= \sum_{j\in\mathcal{J}} Wb_p^p\left(\sum_{i:X_i\in Q_j}\delta_{X_i}, \mathbb{E}[n(Q_j)]\frac{\rho\mathbb{1}_{Q_j}}{\rho(Q_j)}\right), \tag{6.6.0.1}$$

where in the last passage we have used that

$$\mathbb{E}[n(Q_j)] = n\rho(Q_j).$$

Taking the expectation in (6.6.0.1), if $\{Z_i\}_{i=1}^\infty$ is a sequence of iid random variables uniformly distributed on $[0,1]^d$, one gets

$$\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^n\delta_{X_i}, n\rho\right)\right] \geq \sum_{j\in\mathcal{J}}\mathbb{E}\left[Wb_p^p\left(\sum_{i:X_i\in Q_j}\delta_{X_i}, \mathbb{E}[n(Q_j)]\frac{\rho\mathbb{1}_{Q_j}}{\rho(Q_j)}\right)\right]$$

$$\geq \left(1-\frac{c}{\ell}\right)\sum_{j\in\mathcal{J}}|Q_j|^{\frac{p}{d}}\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^{n(Q_j)}\delta_{Z_i}, \mathbb{E}[n(Q_j)]\mathbb{1}_{[0,1]^d}\right)\right] \qquad \text{by Lemma 6.3.2}$$

$$\geq \tilde{c}b_d^p\left(1-\frac{c}{\ell}\right)\sum_{j\in\mathcal{J}}(1-\omega(\mathbb{E}[n(Q_j)]))|Q_j|^{\frac{p}{d}}(\mathbb{E}[n(Q_j)])^{1-\frac{p}{d}}(1+\chi_{d=2}\log(1+\mathbb{E}[n(Q_j)]))^{\frac{p}{2}}, \tag{6.6.0.2}$$

where in the last passage we used (6.4.0.9) of Proposition 6.4.1.

First we notice that, if the hypercube in the partition by Definition 6.2.6 is $Q_j \subseteq \mathbb{Q}^{k+1}\setminus Q^k$, one has

$$\mathbb{E}[n(Q_j)] = n\int_{Q_j} dx\rho(x) \geq cn\rho(x_k e_1)|Q_j| = \frac{c}{\ell^d}ne^{-\frac{x_k^q}{q}}x_k^{-\gamma}x_k^{d(1-q)} \geq n\frac{c}{\ell^d}e^{-\frac{R_n^q}{q}}R_n^{-\gamma+d(1-q)} =: \frac{c}{\ell^d}f_n,$$

where, using definition (6.1.0.6) of the critical radius R_n, one has

$$f_n \xrightarrow[n\to\infty]{} \infty \quad \text{if } p\leq p^{crit} \quad \text{and} \quad \frac{1}{C}\leq f_n\leq C \quad \text{if } p>p^{crit}. \tag{6.6.0.3}$$

Then, we can use (6.1.0.3) as in the proof of the upper bound (where ε is substituted by $\frac{1}{\ell}$) to get

$$\mathbb{E}[n(Q_j)]^{1-\frac{p}{d}} = \left(n\int_{Q_j}dx\rho(x)\right)^{1-\frac{p}{d}} \geq \left(1-\frac{c}{\ell}\right)n^{1-\frac{p}{d}}|Q_j|^{-\frac{p}{d}}\int_{Q_j}dx\rho(x)^{1-\frac{p}{d}}, \tag{6.6.0.4}$$

so that, if $d\geq 3$, combining the previous estimate in (6.6.0.4) with (6.6.0.2), one gets

$$\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^n\delta_{X_i}, n\rho\right)\right] \geq \tilde{c}b_d^p n^{1-\frac{p}{d}}\left(1-\frac{c}{\ell}-\omega(f_n)\right)\sum_{j\in\mathcal{J}}\int_{Q_j}dx\rho(x)^{1-\frac{p}{d}}$$

$$\geq \tilde{c}b_d^p n^{1-\frac{p}{d}}\left(1-\frac{c}{\ell}-\omega(f_n)\right)\int_{B_{R_n}}dx\rho(x)^{1-\frac{p}{d}},$$

where ω is a function strictly smaller that 1 and vanishing as the argument increases and where the second inequality follows from the fact that

$$B_{R_n}\subseteq\bigcup_{j\in\mathcal{J}}Q_j,$$

and this concludes passing to the limit as $n\to\infty$.

In the case $d=2$ one has to divide the proof in the three cases $p<p^{crit}$, $p=p^{crit}$ and $p>p^{crit}$.

Case $d=2$ and $p<p^{crit}$. This case is the simplest one since one can restrict the sum in (6.6.0.2) to the hypercubes Q_j with

$$j\in\mathcal{J}':=\left\{j\in\mathcal{J}:\mathbb{E}[n(Q_j)]\geq\frac{n}{\log n}\right\},$$

and notice that

$$\bigcup_{n\geq 2}\bigcup_{j\in\mathcal{J}'}Q_j=\mathbb{R}^2. \tag{6.6.0.5}$$

Before proving (6.6.0.5) we see how to conclude: by (6.6.0.3) and (6.6.0.4) one has

$$\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^n\delta_{X_i},n\rho\right)\right]\geq\tilde{c}b_2^pn^{1-\frac{p}{2}}\left(1-\frac{c}{\ell}-\omega(f_n)\right)\sum_{j\in\mathcal{J}}\int_{Q_j}dx\rho(x)^{1-\frac{p}{2}}\log(\mathbb{E}[n(Q_j)])^{\frac{p}{2}},$$

and therefore, restricting the sum to the special Q_j with $j\in\mathcal{J}'$ we mentioned before, we get

$$\begin{aligned}\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^n\delta_{X_i},n\rho\right)\right]&\geq\tilde{c}b_2^pn^{1-\frac{p}{2}}\left(1-\frac{c}{\ell}-\omega(f_n)\right)\sum_{j\in\mathcal{J}'}\int_{Q_j}dx\rho(x)^{1-\frac{p}{2}}(\log n-\log\log n)^{\frac{p}{2}}\\&=\tilde{c}b_2^pn^{1-\frac{p}{2}}(\log n)^{\frac{p}{2}}\left(1-\frac{c}{\ell}-\omega(f_n)-\omega(n)\right)\sum_{j\in\mathcal{J}'}\int_{Q_j}dx\rho(x)^{1-\frac{p}{2}}\\&=\tilde{c}b_2^pn^{1-\frac{p}{2}}(\log n)^{\frac{p}{2}}\left(1-\frac{c}{\ell}-\omega(f_n)-\omega(n)\right)\int_{\cup_{j\in\mathcal{J}'}Q_j}dx\rho(x)^{1-\frac{p}{2}}\\&=\tilde{c}b_2^pn^{1-\frac{p}{2}}(\log n)^{\frac{p}{2}}\left(1-\frac{c}{\ell}-\omega(f_n)-\omega(n)\right)\int_{\mathbb{R}^2}dx\rho(x)^{1-\frac{p}{2}}\qquad\text{by (6.6.0.5)}.\end{aligned}$$

Hence, we only have to prove (6.6.0.5), and this is quite immediate since

$$\mathbb{E}[n(Q_j)]=n\int_{Q_j}dx\rho(x)\geq cn\rho(x_ke_1)|Q_j|=\frac{c}{\ell^d}ne^{-\frac{x_k^q}{q}}x_k^{-\gamma}x_k^{d(1-q)}\geq c\frac{n}{\log n},$$

for any x_k such that

$$x_k\leq h_n:=\begin{cases}\sqrt[q]{q\log\left(\frac{\log n}{(\log\log n)^{\frac{\gamma}{q}+d(1-\frac{1}{q})}}\right)} & q>0,\\ (\log n)^{\frac{1}{\gamma-d}} & q=0,\end{cases}\quad\Rightarrow h_n\xrightarrow[n\to\infty]{}\infty,$$

and therefore

$$\bigcup_{j\in\mathcal{J}'}Q_j\supseteq\bigcup_{k:x_k\leq h_n}Q^k=[-x_k,x_k]^d\nearrow\mathbb{R}^2,$$

which proves (6.6.0.5).

Case $d=2$ and $p=p^{crit}$. We claim that for any $j\in\mathcal{J}$ and for any $x\in Q_j$

$$\log(\mathbb{E}[n(Q_j)])\geq\begin{cases}\log(n\rho(x))-c\log\log n & q>0,\\ \log(n\rho(x)|x|^2)-c & q=0,\end{cases} \tag{6.6.0.6}$$

for a constant c depending on ℓ.

Let us postpone the proof of that. Exploiting this estimate for $p=p^{crit}$ one has

$$\begin{aligned}\mathbb{E}[n(Q_j)]^{1-\frac{p}{2}}[1+\log(1+\mathbb{E}[n(Q_j)])^{\frac{p}{2}}&\geq\mathbb{E}[n(Q_j)]^{1-\frac{p^{crit}}{2}}[\log(\mathbb{E}[n(Q_j)])^{\frac{p^{crit}}{2}}\\&=|Q_j|^{-\frac{p^{crit}}{2}}\begin{cases}\int_{Q_j}dx[\log(n\rho(x))-c\log\log n] & q>0,\\ n^{1-\frac{p^*}{2}}\int_{Q_j}dx\rho(x)^{1-\frac{p^*}{2}}(\log(n\rho(x)|x|^2)-c)^{\frac{p^*}{2}} & q=0,\end{cases}\end{aligned}$$

and combining this estimate with (6.6.0.2) and (6.6.0.3) we get

$$\begin{aligned}\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^n\delta_{X_i},n\rho\right)\right]&\geq n^{1-\frac{p^{crit}}{2}}\tilde{c}b_2^{p^{crit}}\sum_{j\in\mathcal{J}}\begin{cases}\int_{Q_j}dx[\log(n\rho(x))-c\log\log n] & q>0,\\ \int_{Q_j}dx\rho(x)^{1-\frac{p^*}{2}}[(\log(n\rho(x)|x|^2))^{\frac{p^*}{2}}-c^{\frac{p^*}{2}}] & q=0,\end{cases}\\&\geq n^{1-\frac{p^{crit}}{2}}\tilde{c}b_2^{p^{crit}}\int_{B_{R_n}}dx\begin{cases}[\log(n\rho(x))-c\log\log n] & q>0,\\ \rho(x)^{1-\frac{p^*}{2}}[(\log(n\rho(x)|x|^2))^{\frac{p^*}{2}}-c^{\frac{p^*}{2}}] & q=0,\end{cases}\end{aligned}$$

and this concludes since the terms multiplied by $c \log\log n$ and c have order $o(\tau_n^{p,d})$.

Hence we only have to prove (6.6.0.6), which is quite immediate since if $q > 0$ and $x \in Q_j$ one has

$$\begin{aligned}
\log(\mathbb{E}[n(Q_j)]) &= \log\left(n\int_{Q_j} dy\rho(y)\right) = \log\left(n\frac{1}{|Q_j|}\int_{Q_j} dy\rho(y)\right) + \log|Q_j| \\
&\geq \log\left(n\left(1-\frac{c}{\ell}\right)\rho(x)\right) + \log|Q_j| \\
&\geq \log(n\rho(x)) - \frac{c}{\ell} - \left|\log\left(\frac{1}{\ell^d}|x|^{d(1-q)}\right)\right| \\
&\geq \log(n\rho(x)) - c\log\log n,
\end{aligned}$$

while if $q = 0$ and $x \in Q_j \subseteq Q^{k+1} \setminus Q^k$

$$\begin{aligned}
\log(\mathbb{E}[n(Q_j)]) &= \log\left(n\int_{Q_j} dy\rho(y)\right) \geq \log\left(n\left(1-\frac{c}{\ell}\right)\rho(x)|Q_j|\right) \\
&\geq \log\left(n\rho(x)\frac{1}{\ell^2}x_k^2\right) - \frac{c}{\ell} \\
&\geq \log\left(n\rho(x)|x|^2\right) - c,
\end{aligned}$$

and this proves (6.6.0.6).

Case $d = 2$ and $p > p^{crit}$. In this case one can conclude as in the case $d \geq 3$ noticing that

$$\log(1 + \mathbb{E}[n(Q_j)]) \geq c,$$

and combining this estimate with (6.6.0.2), (6.6.0.3) and (6.6.0.4) one gets the same conclusion as the case $d \geq 3$, that is

$$\begin{aligned}
\mathbb{E}\left[Wb_p^p\left(\sum_{i=1}^n \delta_{X_i}, n\rho\right)\right] &\geq c^{\frac{p}{2}}\tilde{c}b_2^p n^{1-\frac{p}{2}}\left(1-\frac{c}{\ell}-\omega(f_n)\right)\sum_{j\in\mathcal{J}}\int_{Q_j} dx\rho(x)^{1-\frac{p}{2}} \\
&\geq c^{\frac{p}{2}}\tilde{c}b_2^p n^{1-\frac{p}{2}}\left(1-\frac{c}{\ell}-\omega(f_n)\right)\int_{B_{R_n}} dx\rho(x)^{1-\frac{p}{2}},
\end{aligned}$$

and this concludes passing to the limit as $n \to \infty$. $\square$

Bibliography

[1] L. Ambrosio and F. Glaudo, Finer estimates on the 2-dimensional matching problem, *J. Éc. Polytech. Math.* **6** (2019) 737–765.

[2] L. Ambrosio, M. Goldman and D. Trevisan, On the quadratic random matching problem in two-dimensional domains, *Electron. J. Probab.* **27** (2022) 35. Id/No 54.

[3] M. Ajtai, J. Komlós, and G. Tusnády, On optimal matchings, *Combinatorica* **4** (1984) 259–264.

[4] L. Ambrosio, F. Stra and D. Trevisan, A PDE approach to a 2-dimensional matching problem, *Probab. Theory Relat. Fields* **173** (1-2) (2019) 433–477.

[5] F. Barthe and C. Bordenave, Combinatorial optimization over two random point sets, *Séminaire de Probabilités XLV, Springer* (2013) 483–535.

[6] J. D. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, *Numer. Math.* **84** (3) (2000) 375–393.

[7] D. Benedetto and E. Caglioti, Euclidean random matching in 2d for non-constant densities, *J. Stat. Phys.* **181** (3) (2020) 854–869.

[8] S. Bobkov and M.Ledoux, One-Dimensional Empirical Measures, Order Statistics, and Kantorovich Transport Distances, *Paperback Mem. Am. Math. Soc.* **1259** (2019).

[9] J. H. Boutet de Monvel and O. C. Martin, Almost sure convergence of the minimum bipartite matching functional in Euclidean space, *Combinatorica* **22** (2002) 523–530.

[10] H. Brezis, Remarks on the Monge-Kantorovich problem in the discrete setting, *Comptes Rendus Math.* **356** (2018) 207–213.

[11] E. Caglioti, M. Goldman, F. Pieroni and D. Trevisan, Subadditivity and optimal matching of unbounded samples, *arXiv:2407.06352.*

[12] E. Caglioti and F. Pieroni, Random Matching in 2D with Exponent 2 for Gaussian Densities, *J. Stat. Phys.*, **191** (2024) 62.

[13] S. Caracciolo, C. Lucibello, G. Parisi and G. Sicuro, Scaling hypothesis for the euclidean bipartite matching problem, *Phys. Rev. E* **90** (2014) 012118.

[14] S. Caracciolo and G. Sicuro, One dimensional Euclidean matching problem: Exact solutions, correlation functions and universality, *Phys. Rev. E* **90** (2014) 4.

[15] H. Chernoff, A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations, *Ann. Math. Statist.* **23** (1952) 493–507.

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

[17] R. M. Dudley, The Speed of Mean Glivenko-Cantelli Convergence, *Ann. Math. Statist.* **40** (1969) 40–50.

[18] P. Etnigof, Lie Groups and Lie Algebras, *arXiv:2201.09397.*

[19] L. C. Evans and R. F. Gariepy, Measure Theory and Fine Properties of Functions, Revised Edition, *Chapman and Hall/CRC*, (2015).

[20] A. Figalli and N. Gigli, A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions, *J. Math. Pures Appl. (9)* **94** (2) (2010) 107–130.

[21] M. Goldman, H. Huesmann and F. Otto, Almost Sharp Rates of Convergence for the Average Cost and Displacement in the Optimal Matching Problem, In: *Partial Differential Equations: Waves, Nonlinearities and Nonlocalities. Abel Symposia* **18** (2025) 93–109.

[22] M. Goldman, H. Huesmann and D. Trevisan, Asymptotics for Random Quadratic Transportation Costs, *arXiv:2409.08612*.

[23] M. Goldman and D. Trevisan, Convergence of asymptotic costs for random Euclidean matching problems, *Prob. and Math. Phys.* **2** (2) (2021) 341–362.

[24] M. Goldman and D. Trevisan, Optimal transport methods for combinatorial optimization over two random point sets, *Probab. Theory Relat. Fields*

[25] M. Ledoux, On optimal matching of Gaussian samples, *J. Math. Sci. NY* **238** (4) (2019) 495–522.

[26] M. Ledoux, On optimal matching of Gaussian samples II (2018).

[27] M. Ledoux and Z. Jie-Xiang, On optimal matching of Gaussian samples III, *Probab. Math. Stat.* (2021).

[28] J. M. Lee, Riemannian Manifolds: An Introduction to Curvature, Springer–Verlag, *Graduate texts in Mathematics* **176** (1997).

[29] M. Mezard and G. Parisi, The Euclidean matching problem, *J. Phys.* **49** (1988) 2019–2025.

[30] S. M. Ross, Introduction to Probability Models, 10th edition, *Academic Press* (2010).

[31] F. Santambrogio, Optimal transport for applied mathematicians, *Birkhäuser, NY.* **55** (58-63) 94 (2015).

[32] M. Talagrand, Scaling and non-standard matching theorems, *C. R. Math. Acad. Sci. Paris* **356** (6) (2018) 692–695.

[33] C. Villani, Optimal transport: old and new, *Springer Berlin Heidelberg* (2009).