CONVERGENCE OF THE EMPIRICAL MEASURE IN EXPECTED WASSERSTEIN DISTANCE: NON ASYMPTOTIC EXPLICIT BOUNDS IN $\mathbb{R}^d$

NICOLAS FOURNIER

Abstract. We provide some non asymptotic bounds, with explicit constants, that measure the rate of convergence, in expected Wasserstein distance, of the empirical measure associated to an i.i.d. N-sample of a given probability distribution on $\mathbb{R}^d$. We consider the cases where $\mathbb{R}^d$ is endowed with the maximum and Euclidean norms.

1. Introduction

Let $d \geq 1$. We consider $\mu \in \mathcal{P}(\mathbb{R}^d)$, the set of probability measures on $\mathbb{R}^d$, and an i.i.d. sequence $(X_k)_{k\geq 1}$ of μ-distributed random variables. For $N \geq 1$, we introduce the empirical measure

$$\mu_N = \frac{1}{N}\sum_{k=1}^{N} \delta_{X_k}. \tag{1}$$

Estimating the rate of convergence of μ_N to μ is of course a fundamental problem, and it seems that measuring this convergence in Wasserstein distance is nowadays a widely adopted choice. Some seminal works on the subject are those by Dudley [6], Ajtai-Komlós-Tusnády [1] and Dobrić-Yukich [5]. More recently, some results have been established by Bolley-Guillin-Villani [3], Boissard-Le Gouic [2], Dereich-Scheutzow-Schottstedt [4] and Fournier-Guillin [7]. In particular, we can find in [4, 7] the following result.

Fix some norm $|\cdot|$ on $\mathbb{R}^d$ and consider, for $p > 0$ and $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, the transport cost

$$\mathcal{T}_p(\mu,\nu) = \inf\Big\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \xi(\mathrm{d}x,\mathrm{d}y) : \xi \in \mathcal{H}(\mu,\nu)\Big\},$$

where $\mathcal{H}(\mu,\nu)$ stands for the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν. It holds that $\mathcal{T}_p = \mathcal{W}_p^p$, with the usual notation, if $p \geq 1$. For $q > 0$, we also define

$$\mathcal{M}_q(\mu) = \int_{\mathbb{R}^d} |x|^q \mu(\mathrm{d}x).$$

There exists a constant $C_{d,p,q}$ such that for all $\mu \in \mathcal{P}(\mathbb{R}^d)$, for all $N \geq 1$, with μ_N defined in (1),

$$\mathbb{E}[\mathcal{T}_p(\mu_N,\mu)] \leq C_{d,p,q}[\mathcal{M}_q(\mu)]^{p/q} \times \begin{cases} N^{-1/2} & \text{if } p > d/2 \text{ and } q > 2p, \\ N^{-1/2}\log(1+N) & \text{if } p = d/2 \text{ and } q > 2p, \\ N^{-p/d} & \text{if } p \in (0, d/2) \text{ and } q > dp/(d-p). \end{cases}$$

These bounds are sharp, but the constants in [4, 7] are either non explicit or large. The number of required moments is also sharp. The case with less finite moments is also studied in [7].

2010 *Mathematics Subject Classification.* 60F25, 65C05.

Key words and phrases. Empirical measure, Sequence of i.i.d. random variables, Optimal transportation.

We have been asked a high number of times, in particular by applied scientists, if one can compute the constant $C_{d,p,q}$. If following the proofs in [4, 7], one finds some huge constants. But revisiting these proofs and optimizing as often as possible the computations, which is the purpose of the present paper, we obtain some rather reasonable constants, when using the maximum norm $|\cdot|_\infty$ on $\mathbb{R}^d$. In particular, they remain finite as the dimension tends to infinity (but of course, the rate of convergence in $N^{-p/d}$ is worse and worse).

The reason why the maximum norm $|\cdot|_\infty$ is used in [4, 7] it that the proofs rely on a partitioning of the unit ball, and that a cube is very easy to cut into smaller cubes. We of course deduce some bounds for the more natural Euclidean norm $|\cdot|_2$, multiplying the constant by $d^{p/2}$. This leads to a constant (for the Euclidean norm) that explodes as $d \to \infty$.

Using similar arguments, together with some ideas found in Weed-Bach [12], Lei [8, Theorem 3.1] proves that one can also find, in the case of the Euclidean norm, some constants (that he does not make explicit) that remain finite as $d \to \infty$. We also produce, in the present paper, some explicit constants in this context.

We present a unified proof for $|\cdot|_\infty$ and $|\cdot|_2$, specifying only at the end the choice of the norm. Since the situation is more intricate for the Euclidean norm, the constants we obtain are less good. As we will see, it is actually preferable to use the bound concerning $|\cdot|_\infty$ multiplied by $d^{p/2}$ for low dimensions, see e.g. Tables 3 and 4 below.

Finally, let us mention that Dudley [6] and more recently Weed-Bach [12] study the very interesting problem of obtaining some rates of convergence depending on the *true* dimension of the problem: if e.g. μ is a measure on $\mathbb{R}^d$ but is actually carried by a manifold of lower dimension, what about the rate of convergence ? They introduce some notion of dimension of the measure μ and get some bounds of $\mathbb{E}[\mathcal{T}_p(\mu_N, \mu)]$ in terms of this dimension.

We refer to the introduction of [12], as well as those of [8], [4] and [7], for some much more detailed presentations of the subject and its numerous applications.

2. Main results

2.1. **Basic notation.** We consider the Euclidean and maximum norms: for $x = (x_1, \dots, x_d) \in \mathbb{R}^d$,

$$|x|_2 = \Big(\sum_{i=1}^d x_i^2\Big)^{1/2} \qquad \text{and} \qquad |x|_\infty = \max\{|x_1|, \dots, |x_d|\}.$$

For $p > 0$, for μ, ν in $\mathcal{P}(\mathbb{R}^d)$ and for $i \in \{2, \infty\}$, we set

$$\mathcal{T}_p^{(i)}(\mu, \nu) = \inf \Big\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|_i^p \xi(\mathrm{d}x, \mathrm{d}y) : \xi \in \mathcal{H}(\mu, \nu) \Big\},$$

where we recall that $\mathcal{H}(\mu, \nu)$ is the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν. For $q > 0$, for $\mu \in \mathcal{P}(\mathbb{R}^d)$ and for $i \in \{2, \infty\}$, we define

$$\mathcal{M}_q^{(i)}(\mu) = \int_{\mathbb{R}^d} |x|_i^q \mu(\mathrm{d}x).$$

We of course have, since $|\cdot|_\infty \le |\cdot|_2 \le \sqrt{d}|\cdot|_\infty$, that

$$(2) \qquad \mathcal{T}_p^{(\infty)}(\mu, \nu) \le \mathcal{T}_p^{(2)}(\mu, \nu) \le d^{p/2} \mathcal{T}_p^{(\infty)}(\mu, \nu) \quad \text{and} \quad \mathcal{M}_q^{(\infty)}(\mu) \le \mathcal{M}_q^{(2)}(\mu) \le d^{q/2} \mathcal{M}_q^{(\infty)}(\mu).$$

2.2. **Covering number.** Our proofs are based on a suitable partitioning of the unit ball. The case of the maximum norm is not hard, because it is easy to cut a cube into smaller cubes. The case of the Euclidean case is more intricate. For $r \in (0,1)$, we define

$$N_r = \min\Big\{k \in \mathbb{N} : \exists\, x_1, \dots, x_k \in B_2(0,1) \text{ such that } B_2(0,1) \subset \cup_{i=1}^k B_2(x_i, r)\Big\}, \tag{3}$$

where $B_2(x,r) = \{y \in \mathbb{R}^d : |x-y|_2 < r\}$, as well as

$$K_d = \sup_{r\in(0,1)} r^d N_r, \quad \text{so that for all } r \in (0,1), \quad N_r \le K_d\, r^{-d}. \tag{4}$$

An estimate of K_d can be found in the work of Verger-Gaugry [11], partly based on the results of Rogers [10], see (7) below.

2.3. **Main result.** For $x > 0$ and $q > s > 0$, we set

$$H(x,s,q) = \Big(x\frac{q-s}{s} + (1+x)\Big(\frac{q}{s}\Big)^{q/(q-s)}\Big)^{s/q}\frac{q}{q-s}. \tag{5}$$

The goal of this paper is to prove the following result.

Theorem 1. *We fix $p > 0$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$, we set $\varepsilon_p = \max\{2^{-1}, 2^{-p}\}$ and we recall* (1).

(i) If $p > d/2$ and $q > 2p$, then for $i \in \{2, \infty\}$, for all $N \ge 1$,

$$\mathbb{E}[\mathcal{T}_p^{(i)}(\mu_N, \mu)] \le 2^p \frac{\kappa_{d,p}^{(i)}}{\sqrt{N}}[\mathcal{M}_q^{(i)}(\mu)]^{p/q}\theta_{d,p,q}^{(i)}, \quad \textit{where} \quad \theta_{d,p,q}^{(i)} = H\Big(\frac{\varepsilon_p}{\kappa_{d,p}^{(i)}}, 2p, q\Big),$$

with

$$\kappa_{d,p}^{(\infty)} = \frac{2^{d/2-1}}{1-2^{d/2-p}},$$

and with

$$\kappa_{d,p}^{(2)} = \min\{\kappa_{d,p,r}^{(2)} : r > 2\}, \qquad \textit{where} \qquad \kappa_{d,p,r}^{(2)} = \frac{\sqrt{K_d}}{2}\frac{r^d}{(r-2)^{d/2}(1-r^{d/2-p})}.$$

(ii) If $p = d/2$ and $q > 2p$, then for $i \in \{2, \infty\}$, for all $N \ge 1$,

$$\mathbb{E}[\mathcal{T}_p^{(i)}(\mu_N, \mu)] \le 2^p \frac{\kappa_{d,p,N}^{(i)}}{\sqrt{N}}[\mathcal{M}_q^{(i)}(\mu)]^{p/q}\theta_{d,p,q,N}^{(i)}, \quad \textit{where} \quad \theta_{d,p,q,N}^{(i)} = H\Big(\frac{\varepsilon_p}{\kappa_{d,p,N}^{(i)}}, 2p, q\Big),$$

with (here $\log_+(x) = \max\{\log x, 0\}$)

$$\kappa_{d,p,N}^{(\infty)} = \frac{2^{p-1}}{p\log 2}\log_+\Big((2^{1-p} - 2^{1-2p})\sqrt{N}\Big) + \frac{2^{p-1}}{1-2^{-p}},$$

and with

$$\kappa_{d,p,N}^{(2)} = \min\{\kappa_{d,p,N,r}^{(2)} : r > 2\},$$

where

$$\kappa_{d,p,N,r}^{(2)} = \frac{\sqrt{K_d}}{2}\frac{r^{2p}}{(r-2)^p p\log r}\log_+\Big(2(r-2)^p(r^{-2p} - r^{-3p})\sqrt{\frac{N}{K_d}}\Big) + \frac{\sqrt{K_d}}{2}\frac{r^{3p}}{(r-2)^p(r^p-1)}.$$

(iii) If $p \in (0, d/2)$ and $q > dp/(d-p)$, then for $i \in \{2, \infty\}$, for all $N \ge 1$,

$$\mathbb{E}[\mathcal{T}_p^{(i)}(\mu_N, \mu)] \le 2^p \frac{\kappa_{d,p}^{(i)}}{N^{p/d}}[\mathcal{M}_q^{(i)}(\mu)]^{p/q}\theta_{d,p,q}^{(i)}, \quad \textit{where} \quad \theta_{d,p,q}^{(i)} = H\Big(\frac{2^{1-2p/d}\varepsilon_p}{\kappa_{d,p}^{(i)}}, \frac{dp}{d-p}, q\Big),$$

with

$$\kappa^{(\infty)}_{d,p} = \frac{2^{p-2p/d}(1-2^{-d/2})^{1-2p/d}}{1-2^{p-d/2}},$$

and with

$$\kappa^{(2)}_{d,p} = \min\{\kappa^{(2)}_{d,p,r} : r > 2\}, \qquad \textit{where} \qquad \kappa^{(2)}_{d,p,r} = \Big(\frac{K_d}{4}\Big)^{p/d} \frac{r^{2p}(1-r^{-d/2})^{1-2p/d}}{(r-2)^p(1-r^{p-d/2})}.$$

2.4. **Comments.** By invariance by translation, we can replace $\mathcal{M}^{(i)}_q(\mu)$, in all the formulas, by

$$\inf\Big\{\int_{\mathbb{R}^d} |x-x_0|_i^q \mu(\mathrm{d}x) : x_0 \in \mathbb{R}^d\Big\}.$$

Let us next observe, and we will see that this is often advantageous, that concerning the bound of $\mathbb{E}[\mathcal{T}^{(2)}_p(\mu_N,\mu)]$, by (2), we can replace

- $\kappa^{(2)}_{d,p}$ by $d^{p/2}\kappa^{(\infty)}_{d,p}$ and $\theta^{(2)}_{d,p,q}$ by $\theta^{(\infty)}_{d,p,q}$ in items (i) and (iii);
- $\kappa^{(2)}_{d,p,N}$ by $d^{p/2}\kappa^{(\infty)}_{d,p,N}$ and $\theta^{(2)}_{d,p,q,N}$ by $\theta^{(\infty)}_{d,p,q,N}$ in item (ii).

Set $B_i(x,r) = \{y \in \mathbb{R}^d : |x-y|_i < r\}$. In each case, we present the bound under the form

$$\Big(\texttt{diameter of } B_i(0,1)\Big)^p \times \Big(\texttt{bound in the compact case}\Big) \times [\mathcal{M}^{(i)}_q(\mu)]^{p/q} \times \Big(\theta^{(i)}_{d,p,q} \texttt{ or } \theta^{(i)}_{d,p,q,N}\Big),$$

where *diameter of* $B_i(0,1)$ equals 2 and where by *compact case* we mean the case where μ is supported by the ball $B_i(0,1/2)$ with diameter 1.

One can tediously check that in each case, $q \to \theta^{(i)}_{d,p,q}$ (or $q \to \theta^{(i)}_{d,p,q,N}$) is decreasing and tends to 1 as $q \to \infty$. Hence if we are in the compact case, i.e. if μ is supported in $B_i(0,1/2)$, then $\mathcal{M}^{(i)}_q(\mu) \le 2^{-q}$ for all $q > 0$, and we find that

$$\lim_{q\to\infty} \Big(\texttt{diameter of } B_i(0,1)\Big)^p \times [\mathcal{M}^{(i)}_q(\mu)]^{p/q} \times \Big(\theta^{(i)}_{d,p,q} \texttt{ or } \theta^{(i)}_{d,p,q,N}\Big) = 1,$$

which justifies the denomination `bound in the compact case`. Let us mention that the explicit constant $\theta'_{d,p,q}$ provided in [4, Theorem 3], that concerns the case $p \in (0,d/2)$ and the maximum norm, tends to infinity as $q \to \infty$.

One can tediously check from (iii) that $p \to (\kappa^{(i)}_{d,p})^{1/p}$ is increasing for $p \in (0,d/2)$, which is natural by monotony in p of $[\mathcal{T}_p(\mu,\nu)]^{1/p}$. Actually, in the case of the Euclidean norm, it holds that $p \to (\kappa^{(2)}_{d,p,r})^{1/p}$ is increasing for $p \in (0,d/2)$ for each $r > 2$. Still concerning (iii), an explicit constant $\kappa'_{d,p}$ is provided in [4, Proposition 1] for the maximum norm, but $(\kappa'_{d,p})^{1/p}$ tends to infinity as p decreases to 0.

However, in the non compact case, we did not manage to guarantee such a property: it does not hold true that in item (iii), $p \to [\kappa^{(i)}_{d,p}\theta^{(i)}_{d,p,q}]^{1/p}$ is increasing for $p \in (0,d/2)$, as it should. Hence it may be sometimes be preferable to use the bound: for $p \in (0,d/2)$ and $q > dp/(d-p)$,

$$\mathbb{E}[\mathcal{T}^{(i)}_p(\mu_N,\mu)] \le \inf_{p'\in[p,d/2)} \mathbb{E}[\mathcal{T}^{(i)}_{p'}(\mu_N,\mu)]^{p/p'} \le \inf_{p'\in[p,d/2)} 2^p \frac{[\kappa^{(i)}_{d,p'}]^{p/p'}}{N^{p/d}} [\mathcal{M}^{(i)}_q(\mu)]^{p/q} [\theta^{(i)}_{d,p',q}]^{p/p'}, \tag{6}$$

with the convention that $\theta^{(i)}_{d,p',q} = \infty$ if $q \le dp'/(d-p')$. This is the major default of this work. We identified some computations that might be done more carefully, but this led to awful complications, without producing some marked improvements.

Finally, one can easily check from (iii) that for any $p > 0$,

$$\lim_{d\to\infty} \kappa^{(\infty)}_{d,p} = 2^p \qquad \text{and} \qquad \lim_{d\to\infty} \kappa^{(2)}_{d,p} = 8^p.$$

Concerning the Euclidean case, we use (7) below which implies that $\lim_{d\to\infty} K_d^{1/d} = 1$, we find that $\lim_{d\to\infty} \kappa^{(2)}_{d,p,r} = (r-2)^{-p} r^{2p}$ for each $r > 2$, and this (optimally) equals 8^p with $r = 4$. The constant $\kappa'_{d,p}$ of [4, Proposition 1], for the maximum norm, increases to infinity as $d \to \infty$.

2.5. **Numerical values in the compact case.** We start with the maximum norm.

$d=1$	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$	$d=8$	$d=9$	$d=100$	$d=500$
$\frac{2.42}{N^{1/2}}$	$\frac{0.73\log N+1}{N^{1/2}}$	$\frac{3.72}{N^{1/3}}$	$\frac{2.45}{N^{1/4}}$	$\frac{2.09}{N^{1/5}}$	$\frac{1.94}{N^{1/6}}$	$\frac{1.87}{N^{1/7}}$	$\frac{1.84}{N^{1/8}}$	$\frac{1.82}{N^{1/9}}$	$\frac{1.98}{N^{1/100}}$	$\frac{2.00}{N^{1/500}}$

TABLE 1. Bound of $\mathbb{E}[\mathcal{T}_1^{(\infty)}(\mu_N,\mu)]$ for $N \geq 1$ (actually $N \geq 4$ when $d = 2$), if $\mu \in \mathcal{P}(B_\infty(0,1/2))$.

$d=1$	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$	$d=8$	$d=9$	$d=100$	$d=500$
$\frac{1.05}{N^{1/4}}$	$\frac{1.42}{N^{1/4}}$	$\frac{2.20}{N^{1/4}}$	$\frac{\sqrt{0.73\log N+1.26}}{N^{1/4}}$	$\frac{2.75}{N^{1/5}}$	$\frac{2.20}{N^{1/6}}$	$\frac{2.01}{N^{1/7}}$	$\frac{1.92}{N^{1/8}}$	$\frac{1.87}{N^{1/9}}$	$\frac{1.98}{N^{1/100}}$	$\frac{2.00}{N^{1/500}}$

TABLE 2. Bound of $\sqrt{\mathbb{E}[\mathcal{T}_2^{(\infty)}(\mu_N,\mu)]}$ for $N \geq 1$ (actually $N \geq 8$ when $d = 4$), if $\mu \in \mathcal{P}(B_\infty(0,1/2))$.

We carry on with the Euclidean norm. By Verger-Gaugry [11, (1.1)-(1.3)-(1.4)], where (1.1) is due to Rogers [10], we have the following bound of K_d if $d \geq 8$ (we know from the author that there is a typo in [11] and $d \geq 8$ is the correct condition, instead of $d \geq 2$):

$$K_d \leq \max\{K_{d,1}, K_{d,2}, K_{d,3}\}, \tag{7}$$

where (recall (3)-(4); we have $N_r = \nu_{T,n}$ with $2T = 1/r$ and $n = d$ in the notation of [11])

$$K_{d,1} = d^2[\log d + \log\log d + 5],$$

$$K_{d,2} = \frac{7^{4(\log 7)/7}}{4}\sqrt{\frac{\pi}{2}}d^{3/2}\frac{2(d-1)\log d + \frac{1}{2}\log d + \log(\frac{\pi\sqrt{2d}}{\sqrt{\pi d}-2})}{(1-2/\log d)(1-2/\sqrt{\pi d})(\log d)^2},$$

$$K_{d,3} = \sqrt{2\pi d}\frac{(d-1)\log(2d) + (d-1)\log\log d + \frac{1}{2}\log d + \log(\frac{\pi\sqrt{2d}}{\sqrt{\pi d}-2})}{(1-2/\log d)(1-2/\sqrt{\pi d})}.$$

$d=8$	$d=15$	$d=20$	$d=25$	$d=30$	$d=35$	$d=50$	$d=75$	$d=100$	$d=500$
$\frac{23.44}{N^{1/8}}$	$\frac{13.40}{N^{1/15}}$	$\frac{11.98}{N^{1/20}}$	$\frac{11.20}{N^{1/25}}$	$\frac{10.69}{N^{1/30}}$	$\boldsymbol{\frac{10.34}{N^{1/35}}}$	$\boldsymbol{\frac{9.70}{N^{1/50}}}$	$\boldsymbol{\frac{9.19}{N^{1/75}}}$	$\boldsymbol{\frac{8.93}{N^{1/100}}}$	$\boldsymbol{\frac{8.24}{N^{1/500}}}$
$\boldsymbol{\frac{5.18}{N^{1/8}}}$	$\boldsymbol{\frac{7.11}{N^{1/15}}}$	$\boldsymbol{\frac{8.36}{N^{1/20}}}$	$\boldsymbol{\frac{9.47}{N^{1/25}}}$	$\boldsymbol{\frac{10.47}{N^{1/30}}}$	$\frac{11.38}{N^{1/35}}$	$\frac{13.76}{N^{1/50}}$	$\frac{17.01}{N^{1/75}}$	$\frac{19.73}{N^{1/100}}$	$\frac{44.60}{N^{1/500}}$

TABLE 3. Bound of $\mathbb{E}[\mathcal{T}_1^{(2)}(\mu_N,\mu)]$ for $N \geq 1$ if $\mu \in \mathcal{P}(B_2(0,1/2))$, using the bound proposed for the Euclidean norm in Theorem 1-(iii) (second line) and using $\sqrt{d}$ times the bound proposed for the maximum norm (third line). In bold the one to be used.

$d=8$	$d=15$	$d=20$	$d=25$	$d=30$	$d=35$	$d=50$	$d=75$	$d=100$	$d=500$
$\frac{23.86}{N^{1/8}}$	$\frac{13.41}{N^{1/15}}$	$\frac{11.98}{N^{1/20}}$	$\frac{11.20}{N^{1/25}}$	$\frac{10.69}{N^{1/30}}$	$\mathbf{\frac{10.34}{N^{1/35}}}$	$\mathbf{\frac{9.70}{N^{1/50}}}$	$\mathbf{\frac{9.19}{N^{1/75}}}$	$\mathbf{\frac{8.93}{N^{1/100}}}$	$\mathbf{\frac{8.24}{N^{1/500}}}$
$\mathbf{\frac{5.41}{N^{1/8}}}$	$\mathbf{\frac{7.13}{N^{1/15}}}$	$\mathbf{\frac{8.36}{N^{1/20}}}$	$\mathbf{\frac{9.47}{N^{1/25}}}$	$\mathbf{\frac{10.47}{N^{1/30}}}$	$\frac{11.38}{N^{1/35}}$	$\frac{13.76}{N^{1/50}}$	$\frac{17.01}{N^{1/75}}$	$\frac{19.73}{N^{1/100}}$	$\frac{44.60}{N^{1/500}}$

TABLE 4. Bound of $\sqrt{\mathbb{E}[\mathcal{T}_2^{(2)}(\mu_N,\mu)]}$ for $N \geq 1$ if $\mu \in \mathcal{P}(B_2(0,1/2))$, using the bound proposed for the Euclidean norm in Theorem 1-(iii) (second line) and using $\sqrt{d}$ times the bound proposed for the maximum norm (third line). In bold the one to be used.

As we can see, in large dimension, the bounds concerning $p=1$ and $p=2$ are very similar. Also, we see that, for $p=1$ and $p=2$, it is better to use $\sqrt{d}$ times the bound proposed for the maximum norm when $d \in \{8,\dots,30\}$. Hence it does not seem crucial to find a formula similar to (7) when $d \in \{2,\dots,7\}$.

2.6. **Numerical values in the non-compact case.** Here we study how $\theta^{(i)}_{d,p,q}$ is far from 1. When $p=d/2$, we observe that $N \mapsto \theta^{(i)}_{d,p,q,N}$ is decreasing, and we e.g. study $\theta^{(i)}_{d,p,q,100}$ (which controls $\theta^{(i)}_{d,p,q,N}$ for all $N \geq 100$). We start with the maximum norm.

$d=1$	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$	$d=8$	$d=9$	$d=100$	$d=500$
4.4	4.2	3.3	3.0	2.9	2.8	2.8	2.7	2.7	2.5	2.4
9.8	9.4	7.3	6.8	6.5	6.4	6.3	6.2	6.1	5.5	5.5
40.4	39.0	29.9	27.7	26.6	25.9	25.3	25.0	24.6	22.3	22.1

TABLE 5. Here $p=1$. Minimum value of q so that $\theta^{(\infty)}_{d,1,q} \leq c$ if $d \neq 2$ or $\theta^{(\infty)}_{d,1,q,100} \leq c$ if $d=2$, with $c=4$ (second line), $c=2$ (third line), $c=1.25$ (fourth line).

$d=1$	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$	$d=8$	$d=9$	$d=100$	$d=500$
5.1	5.0	4.9	4.9	4.1	3.7	3.5	3.3	3.2	2.6	2.5
9.5	8.9	8.4	8.4	6.9	6.4	6.0	5.8	5.7	4.6	4.5
37.0	34.5	32.4	32.5	26.7	24.6	23.4	22.5	21.9	17.7	17.4

TABLE 6. Here $p=2$. Minimum value of q so that $\sqrt{\theta^{(\infty)}_{d,2,q}} \leq c$ if $d \neq 4$ or $\sqrt{\theta^{(\infty)}_{d,2,q,100}} \leq c$ if $d=4$, with $c=4$ (second line), $c=2$ (third line), $c=1.25$ (fourth line).

Comparing Tables 5 and 6, it seems clear that, at least for large values of d, it is preferable to use the bound (6).

We do the same job concerning the Euclidean norm. We only deal with $\theta^{(2)}_{d,1,q}$ as defined in Theorem 1 for simplicity, even if we recall that it is preferable to use (2) and the bound concerning the maximum norm for low dimensions.

$d=8$	$d=15$	$d=20$	$d=25$	$d=30$	$d=35$	$d=50$	$d=75$	$d=100$	$d=500$
2.4	2.3	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2
5.2	5.0	4.9	4.9	4.8	4.8	4.8	4.8	4.8	4.8
21.5	20.5	20.2	20.1	20.0	19.9	19.8	19.7	19.7	19.6

TABLE 7. Here $p=1$. Minimum value of q so that $\theta^{(2)}_{d,1,q} \le c$ with $c=4$ (second line), $c=2$ (third line), $c=1.25$ (fourth line).

$d=8$	$d=15$	$d=20$	$d=25$	$d=30$	$d=35$	$d=50$	$d=75$	$d=100$	$d=500$
3.2	2.8	2.7	2.6	2.6	2.6	2.5	2.5	2.5	2.4
5.4	4.7	4.5	4.4	4.4	4.3	4.2	4.2	4.2	4.1
20.5	17.7	17.1	16.7	16.5	16.4	16.1	15.9	15.8	15.5

TABLE 8. Here $p=2$. Minimum value of q so that $\sqrt{\theta^{(2)}_{d,2,q}} \le c$ with $c=4$ (second line), $c=2$ (third line), $c=1.25$ (fourth line).

Comparing Tables 7 and 8, it seems again clear that it is not vain to use the bound (6).

2.7. **On a possible lowerbound.** As mentioned to us by Pagès, we have the following lower-bound, holding for $i \in \{2,\infty\}$ and actually for any other norm. Consider $X_1,\dots,X_N$ independent and μ-distributed. It holds that $\mathcal{T}_p^{(i)}(\mu_N,\mu) \ge \mathcal{S}_p^{(i)}(\mu;X_1,\dots,X_N)$, where

$$\mathcal{S}_p^{(i)}(\mu;x_1,\dots,x_d) = \inf\Big\{\mathcal{T}_p^{(i)}\Big(\mu,\sum_{i=1}^N \alpha_i\delta_{x_i}\Big) : (\alpha_i)_{i=1,\dots,N}\in[0,1]^N,\ \sum_{i=1}^N \alpha_i = 1\Big\},$$

and Luschgy-Pagès show in [9] that, under some technical conditions on μ,

$$\lim_{N\to\infty} N^{-p/d}\mathbb{E}[\mathcal{S}_p^{(i)}(\mu;X_1,\dots,X_n)] = \frac{\Gamma(1+p/d)}{[\lambda_d(B_{(i)}(0,1))]^{p/d}}\int_{\mathbb{R}^d}[f(x)]^{1-p/d}\mathrm{d}x,$$

where f is the density of μ, where λ_d is the Lebesgue measure on $\mathbb{R}^d$ and where Γ is the classical Γ function. Choosing for μ the uniform law on $B_i(0,1/2)$, we find that for $i\in\{2,\infty\}$,

$$\liminf_{N\to\infty} N^{-p/d}\mathbb{E}[\mathcal{T}_p^{(i)}(\mu_N,\mu)] \ge \lim_{N\to\infty} N^{-p/d}\mathbb{E}[\mathcal{S}_p^{(i)}(\mu;X_1,\dots,X_n)] = \frac{\Gamma(1+p/d)}{2^p} =: \gamma_{d,p},$$

to be compared with $\kappa^{(i)}_{d,p}$ when $p\in(0,d/2)$. This may be a rough lowerbound, because this is an asymptotic bound as $N\to\infty$, and because $\mathcal{S}_p^{(i)}(\mu;X_1,\dots,X_N)$ is likely to be really smaller than $\mathcal{T}_p^{(i)}(\mu_N,\mu)$ (in particular it converges in $N^{-p/d}$ when $p>d/2$ while $\mathcal{T}_p^{(i)}(\mu_N,\mu)$ only converges in $N^{-1/2}$ in general when $p>d/2$).

It holds that $\gamma_{d,1}\in(0.44,0.5)$ for all $d\ge 3$, to be compared with the numerators in Tables 1 and 3. Hence when $p=1$ and say $d\ge 4$, see Table 1, $\kappa^{(\infty)}_{d,1}$ is at worst 6 times too large. When $p=1$ and $d\ge 8$, see Table 3, $\min\{\kappa^{(2)}_{d,1},\sqrt{d}\kappa^{(\infty)}_{d,1}\}$ is at worst 24 times too large. We hope this is pessimistic.

It holds that $\sqrt{\gamma_{d,2}}\in(0.47,0.5)$ for all $d\ge 5$, to be compared with the numerators in Tables 2 and 4. When $p=2$ and $d\ge 5$, see Table 2, $(\kappa^{(\infty)}_{d,2})^{1/2}$ is at worst 6 times too large. When $p=2$

and $d \geq 8$, see Table 4, $\min\{(\kappa^{(2)}_{d,2})^{1/2}, \sqrt{d}(\kappa^{(\infty)}_{d,2})^{1/2}\}$ is at worst 23 times too large. Again, we hope this is pessimistic.

We do not discuss the non compact case, but the numerical results do not seem quite favorable.

2.8. **The case with a low order finite moment.** Since this last result is likely to be much less useful for applications, we only treat the case of the maximum norm.

Theorem 2. *Let $q > p > 0$ such that $q < \min\{2p, dp/(d-p)\}$, i.e. $q \in (p, 2p)$ if $p \geq d/2$ and $q \in (p, dp/(d-p))$ if $p \in (0, d/2]$. Fix $\mu \in \mathcal{P}(\mathbb{R}^d)$. For all $N \geq 1$,*

$$\mathbb{E}[\mathcal{T}_p^{(\infty)}(\mu_N, \mu)] \leq 2^p [\mathcal{M}_q^{(\infty)}(\mu)]^{p/q} \frac{\zeta^{(\infty)}_{d,p,q}}{N^{(q-p)/q}},$$

where, setting $\varepsilon_p = \max\{2^{-1}, 2^{-p}\}$,

$$\zeta^{(\infty)}_{d,p,q} = \Big[\varepsilon_p 2^{2p/q-1} + \Big(\frac{2^{d-1}}{2^{d/2}-1}\Big)^{2(q-p)/q} \frac{1-2^{-p}}{1-2^{d-p-dp/q}}\Big] \min_{a\in(1,\infty)} \Big(\frac{a^p}{a^{p-q/2}-1} + \frac{a^p}{1-a^{p-q}}\Big).$$

2.9. **Plan of the paper.** In Section 3, we provide a general estimate of the transport cost between two measures. In Section 4, we apply this general estimate to derive a bound of $\mathbb{E}[\mathcal{T}_p(\mu_N, \mu)]$, for a general norm, for all the values of $p > 0$ and in any dimension. In Section 5, we precisely study some elementary series. We obtain a bound of $\mathbb{E}[\mathcal{T}_p(\mu_N, \mu)]$, for a general norm, separating the cases $p > d/2$, $p = d/2$ and $p \in (0, d/2)$ in Section 6. We conclude the proof of Theorem 1 for the maximum norm in Section 7 and for the Euclidean norm in Section 8. Finally, we check Theorem 2 in Section 9.

3. Upperbound of the transport cost between two measures

The result we prove in this section, Proposition 4, is more or less classical, see Boissard-Le Gouic [2, Proposition 1.1], Dereich-Scheutzow-Schottstedt [4, Lemma 2 and Theorem 3], Fournier-Guillin [7, Lemma 5] and Weed-Bach [12, Proposition 1]. As noted in [12], similar ideas can already be found in Ajtai-Komlós-Tusnády [1]. However, we provide a slightly more precise version, that allows us to get some smaller constants. We consider fixed the following objects.

Setting 3. *(a) We fix a norm $|\cdot|$ on $\mathbb{R}^d$. We denote by $B(x,r) = \{y \in \mathbb{R}^d : |x-y| < r\}$ the corresponding balls, by $\mathcal{T}_p(\mu,\nu) = \inf_{\xi\in\mathcal{H}(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \xi(\mathrm{d}x,\mathrm{d}y)$ the corresponding transport cost, and by $\mathcal{M}_q(\mu) = \int_{\mathbb{R}^d} |x|^q \mu(\mathrm{d}x)$ the corresponding moments.*

(b) Let $G_0 = B(0,1)$. For $a > 1$, we set $G_0^a = G_0$ and, for all $n \geq 1$, $G_n^a = B(0,a^n) \setminus B(0,a^{n-1})$.

(c) We consider a family $(\mathcal{Q}_\ell)_{\ell=0,\dots,k}$ of nested partitions of G_0 such that $\mathcal{Q}_0 = \{G_0\}$. For each $\ell = 1,\dots,k$, each $C \in \mathcal{Q}_\ell$, there exists a unique $F \in \mathcal{Q}_{\ell-1}$ such that $C \subset F$; we then say that C is a child of F. For $\ell = 0,\dots,k$, we denote by $|\mathcal{Q}_\ell|$ the cardinal of $\mathcal{Q}_\ell$ and we set

$$\delta_\ell = \max_{C\in\mathcal{Q}_\ell} \sup_{x,y\in C} |x-y|.$$

For $C \subset \mathbb{R}^d$ and $r > 0$, we put $rC = \{rx : x \in C\}$. Recall that $\varepsilon_p = 2^{-1} \vee 2^{-p}$.

Proposition 4. *We adopt Setting 3 and consider $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$. For all $a > 1$, there are some nonnegative numbers $(r_{a,n,\ell}(\mu,\nu))_{n\geq 0,\ell=0,\dots,k-1}$ satisfying*

$$\sum_{\ell=0}^{k-1} r_{a,n,\ell}(\mu,\nu) \leq 1 \tag{8}$$

for all $n \geq 0$ and, with the convention that $0/0 = 0$,

$$r_{a,n,\ell}(\mu,\nu) \leq \frac{1}{2} \sum_{F \in \mathcal{Q}_\ell} \Big(\frac{\mu(a^n F \cap G_n^a)}{\mu(G_n^a)} \wedge \frac{\nu(a^n F \cap G_n^a)}{\nu(G_n^a)} \Big) \sum_{C \textit{ child of } F} \Big| \frac{\mu(a^n C \cap G_n^a)}{\mu(a^n F \cap G_n^a)} - \frac{\nu(a^n C \cap G_n^a)}{\nu(a^n F \cap G_n^a)} \Big| \tag{9}$$

for all $n \geq 0$, all $\ell = 0, \dots, k-1$, and such that for all $p > 0$,

$$\mathcal{T}_p(\mu,\nu) \leq \sum_{n \geq 0} a^{pn} \Big(2^p \varepsilon_p |\mu(G_n^a) - \nu(G_n^a)| + (\mu(G_n^a) \wedge \nu(G_n^a)) \Big[\delta_k^p + \sum_{\ell=0}^{k-1} \delta_\ell^p r_{a,n,\ell}(\mu,\nu) \Big] \Big).$$

The coefficients $r_{a,n,\ell}(\mu,\nu)$ are actually explicit, but it seems difficult to use more than the properties (8)-(9). We start with the compact case.

Lemma 5. *Let $\mu, \nu \in \mathcal{P}(B(0,1))$. There is $(u_\ell(\mu,\nu))_{\ell=0,\dots,k-1} \in \mathbb{R}_+^k$ satisfying*

$$\sum_{\ell=0}^{k-1} u_\ell(\mu,\nu) \leq 1 \tag{10}$$

and

$$u_\ell(\mu,\nu) \leq \frac{1}{2} \sum_{F \in \mathcal{Q}_\ell} (\mu(F) \wedge \nu(F)) \sum_{C \textit{ child of } F} \Big| \frac{\mu(C)}{\mu(F)} - \frac{\nu(C)}{\nu(F)} \Big|, \qquad \ell = 0, \dots, k-1, \tag{11}$$

and such that for all $p > 0$,

$$\mathcal{T}_p(\mu,\nu) \leq \delta_k^p + \sum_{\ell=0}^{k-1} \delta_\ell^p u_\ell(\mu,\nu).$$

Proof. For all $0 \leq i \leq \ell \leq k$ and $C \in \mathcal{Q}_\ell$, let $f_i(C)$ be the unique element of $\mathcal{Q}_i$ containing C.

Step 1: construction of the coupling. For all $F \in \mathcal{Q}_k$, we set

$$\xi_F(\mathrm{d}x, \mathrm{d}y) = \frac{\mu|_F(\mathrm{d}x)}{\mu(F)} \frac{\nu|_F(\mathrm{d}y)}{\nu(F)}. \tag{12}$$

Then by reverse induction, for $\ell \in \{0, \dots, k-1\}$ and $F \in \mathcal{Q}_\ell$, we build

$$\xi_F(\mathrm{d}x, \mathrm{d}y) = \sum_{C \text{ child of } F} \rho_C \xi_C(\mathrm{d}x, \mathrm{d}y) + q_F \alpha_F(\mathrm{d}x) \beta_F(\mathrm{d}y), \tag{13}$$

where

$$\rho_C = \frac{\mu(C)}{\mu(F)} \wedge \frac{\nu(C)}{\nu(F)},$$

which depends only on $C \in \mathcal{Q}_{\ell+1}$ since $F = f_\ell(C)$, where

$$q_F = \frac{1}{2} \sum_{C \text{ child of } F} \Big| \frac{\mu(C)}{\mu(F)} - \frac{\nu(C)}{\nu(F)} \Big|, \tag{14}$$

and where

$$\alpha_F(\mathrm{d}x) = \frac{1}{q_F} \sum_{C \text{ child of } F} \Big(\frac{\mu(C)}{\mu(F)} - \frac{\nu(C)}{\nu(F)} \Big)_+ \frac{\mu|_C(\mathrm{d}x)}{\mu(C)},$$

$$\beta_F(\mathrm{d}x) = \frac{1}{q_F} \sum_{C \text{ child of } F} \Big(\frac{\nu(C)}{\nu(F)} - \frac{\mu(C)}{\mu(F)} \Big)_+ \frac{\nu|_C(\mathrm{d}x)}{\nu(C)}.$$

It holds that α_F and β_F are two probability measures on F, because

$$q_F = \sum_{C \text{ child of } F} \Big(\frac{\mu(C)}{\mu(F)} - \frac{\nu(C)}{\nu(F)}\Big)_+ = \sum_{C \text{ child of } F} \Big(\frac{\nu(C)}{\nu(F)} - \frac{\mu(C)}{\mu(F)}\Big)_+. \tag{15}$$

Step 2. Here we show that $\xi_{G_0} \in \mathcal{H}(\mu,\nu)$, so that $\mathcal{T}_p(\mu,\nu) \le \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \xi_{G_0}(\mathrm{d}x,\mathrm{d}y)$. We recall that $G_0 = B(0,1)$ is the unique element of $\mathcal{Q}_0$.

We actually prove by reverse induction that for all $\ell \in \{0,\dots,k\}$, all $F \in \mathcal{Q}_\ell$, it holds that $\xi_F \in \mathcal{H}(\frac{\mu|_F}{\mu(F)}, \frac{\nu|_F}{\nu(F)})$. The result will then follow by choosing $\ell = 0$ and $F = G_0$.

This is obvious if $\ell = k$, see (12). Next, we assume that this holds true for some $\ell+1 \in \{1,\dots,k\}$, and we consider $F \in \mathcal{Q}_\ell$. For $A \in \mathcal{B}(\mathbb{R}^d)$, we use (13) to write

$$\xi_F(A\times\mathbb{R}^d) = \sum_{C \text{ child of } F} \rho_C \xi_C(A\times\mathbb{R}^d) + q_F\alpha_F(A) = \sum_{C \text{ child of } F} \rho_C \frac{\mu(A\cap C)}{\mu(C)} + q_F\alpha_F(A)$$

by induction assumption. Thus

$$\xi_F(A\times\mathbb{R}^d) = \sum_{C \text{ child of } F} \Big[\frac{\mu(C)}{\mu(F)} \wedge \frac{\nu(C)}{\nu(F)} + \Big(\frac{\mu(C)}{\mu(F)} - \frac{\nu(C)}{\nu(F)}\Big)_+\Big] \frac{\mu(A\cap C)}{\mu(C)} = \sum_{C \text{ child of } F} \frac{\mu(C)}{\mu(F)}\frac{\mu(A\cap C)}{\mu(C)},$$

whence $\xi_F(A\times\mathbb{R}^d) = \mu(A\cap F)/\mu(F)$. One shows similarly that $\xi_F(\mathbb{R}^d\times A) = \nu(A\cap F)/\nu(F)$.

Step 3. For $i \in \{0,\dots,k\}$ and $F \in \mathcal{Q}_i$, we put $m_F = \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \xi_F(\mathrm{d}x,\mathrm{d}y)$. In this step, we show by induction that for all $i \in \{0,\dots,k-1\}$,

$$m_{G_0} \le \delta_0^p q_{G_0} + \sum_{\ell=1}^{i} \delta_\ell^p \sum_{C_\ell\in\mathcal{Q}_\ell} \Big(\prod_{j=1}^{\ell} \rho_{f_j(C_\ell)}\Big) q_{C_\ell} + \sum_{C_{i+1}\in\mathcal{Q}_{i+1}} \Big(\prod_{j=1}^{i+1} \rho_{f_j(C_{i+1})}\Big) m_{C_{i+1}}. \tag{16}$$

Recalling (13), since the set of all the children of G_0 is $\mathcal{Q}_1$, and since $|x-y| \le \delta_0$ for all $x,y \in G_0$, so that $\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \alpha_{G_0}(\mathrm{d}x)\beta_{G_0}(\mathrm{d}y) \le \delta_0^p$, we see that

$$m_{G_0} \le \delta_0^p q_{G_0} + \sum_{C_1\in\mathcal{Q}_1} \rho_{C_1} m_{C_1} = \delta_0^p q_{G_0} + \sum_{C_1\in\mathcal{Q}_1} \rho_{f_1(C_1)} m_{C_1},$$

which is (16) with $i = 0$. Assume now that (16) holds true for some $i \in \{0,\dots,k-2\}$. For all $C_{i+1} \in \mathcal{Q}_{i+1}$, we use (13) and that $\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \alpha_{C_{i+1}}(\mathrm{d}x)\beta_{C_{i+1}}(\mathrm{d}y) \le \delta_{i+1}^p$ to write

$$m_{C_{i+1}} \le \delta_{i+1}^p q_{C_{i+1}} + \sum_{C_{i+2} \text{ child of } C_{i+1}} \rho_{C_{i+2}} m_{C_{i+2}}.$$

Hence, since $f_j(C_{i+2}) = f_j(C_{i+1})$ for all $j = 1,\dots,i+1$ if C_{i+2} is a child of C_{i+1},

$$\sum_{C_{i+1}\in\mathcal{Q}_{i+1}} \Big(\prod_{j=1}^{i+1} \rho_{f_j(C_{i+1})}\Big) m_{C_{i+1}} \le \delta_{i+1}^p \sum_{C_{i+1}\in\mathcal{Q}_{i+1}} \Big(\prod_{j=1}^{i+1} \rho_{f_j(C_{i+1})}\Big) q_{C_{i+1}} + \sum_{C_{i+2}\in\mathcal{Q}_{i+2}} \Big(\prod_{j=1}^{i+2} \rho_{f_j(C_{i+2})}\Big) m_{C_{i+2}}.$$

This last formula, inserted in (16), gives (16) with $i+1$ instead of i.

Step 4. For all $C_k \in \mathcal{Q}_k$, we have $m_{C_k} \le \delta_k^p$ by (12) and since $x,y \in C_k$ implies that $|x-y| \le \delta_k$. Hence, by definition of ρ_F,

$$\sum_{C_k\in\mathcal{Q}_k} \Big(\prod_{j=1}^{k} \rho_{f_j(C_k)}\Big) m_{C_k} \le \delta_k^p \sum_{C_k\in\mathcal{Q}_k} \prod_{j=1}^{k} \frac{\mu(f_j(C_k))}{\mu(f_{j-1}(C_k))} = \delta_k^p \sum_{C_k\in\mathcal{Q}_k} \mu(C_k) = \delta_k^p.$$

This, inserted in (16) with $i = k-1$, tells us that

$$m_{G_0} \le \delta_0^p q_{G_0} + \sum_{\ell=1}^{k-1} \delta_\ell^p \sum_{C_\ell \in \mathcal{Q}_\ell} \Big(\prod_{j=1}^{\ell} \rho_{f_j(C_\ell)}\Big) q_{C_\ell} + \delta_k^p.$$

Since $\mathcal{T}_p(\mu,\nu) \le m_{G_0}$ by Step 2, we conclude that

$$\mathcal{T}_p(\mu,\nu) \le \sum_{\ell=0}^{k-1} \delta_\ell^p u_\ell(\mu,\nu) + \delta_k^p,$$

where

$$u_0(\mu,\nu) = q_{G_0} \qquad \text{and} \qquad u_\ell(\mu,\nu) = \sum_{C_\ell \in \mathcal{Q}_\ell} (\prod_{j=1}^{\ell} \rho_{f_j(C_\ell)}) q_{C_\ell} \quad \text{for} \quad \ell \in \{1,\dots,k-1\}.$$

Step 5. We now check by induction that for all $n = 1,\dots,k$,

$$\sum_{\ell=0}^{n-1} u_\ell(\mu,\nu) = 1 - \sum_{C_n \in \mathcal{Q}_n} \prod_{j=1}^{n} \rho_{f_j(C_n)}, \tag{17}$$

and this will imply (10). If first $n = 1$, by (15),

$$u_0(\mu,\nu) = q_{G_0} = \sum_{C_1 \in \mathcal{Q}_1} (\mu(C_1) - \nu(C_1))_+ = 1 - \sum_{C_1 \in \mathcal{Q}_1} [\mu(C_1) \wedge \nu(C_1)] = 1 - \sum_{C_1 \in \mathcal{Q}_1} \rho_{C_1}$$

as desired. If next (17) holds with some $n \in \{1, k-1\}$, we write

$$\sum_{\ell=0}^{n} u_\ell(\mu,\nu) = 1 - \sum_{C_n \in \mathcal{Q}_n} \prod_{j=1}^{n} \rho_{f_j(C_n)} + \sum_{C_n \in \mathcal{Q}_n} \Big(\prod_{j=1}^{n} \rho_{f_j(C_n)}\Big) q_{C_n} = 1 - \sum_{C_{n+1} \in \mathcal{Q}_{n+1}} \prod_{j=1}^{n+1} \rho_{f_j(C_{n+1})},$$

because, recalling (15) and that $\rho_{C_{n+1}} = \frac{\mu(C_{n+1})}{\mu(C_n)} \wedge \frac{\nu(C_{n+1})}{\nu(C_n)}$ (if C_{n+1} is a child of C_n),

$$q_{C_n} = \sum_{C_{n+1} \text{ child of } C_n} \Big(\frac{\mu(C_{n+1})}{\mu(C_n)} - \frac{\nu(C_{n+1})}{\nu(C_n)}\Big)_+ = 1 - \sum_{C_{n+1} \text{ child of } C_n} \rho_{C_{n+1}}.$$

Step 6. It only remains to verify (11). But for $\ell = 1,\dots,k-1$, by definition (14) of q_{C_ℓ} and since $\prod_{j=1}^{\ell} \rho_{f_j(C_\ell)} \le \mu(C_\ell) \wedge \nu(C_\ell)$,

$$u_\ell(\mu,\nu) = \sum_{C_\ell \in \mathcal{Q}_\ell} \Big(\prod_{j=1}^{\ell} \rho_{f_j(C_\ell)}\Big) q_{C_\ell} \le \frac{1}{2} \sum_{C_\ell \in \mathcal{Q}_\ell} (\mu(C_\ell) \wedge \nu(C_\ell)) \sum_{C_{\ell+1} \text{ child of } C_\ell} \Big|\frac{\mu(C_{\ell+1})}{\mu(C_\ell)} - \frac{\nu(C_{\ell+1})}{\nu(C_\ell)}\Big|.$$

Hence we have (11) for any $\ell = 1,\dots,k-1$. Next, since $\mathcal{Q}_0 = \{G_0\}$ and μ,ν are carried by G_0,

$$u_0(\mu,\nu) = q_{C_0} = \frac{1}{2} \sum_{C_1 \in \mathcal{Q}_1} |\mu(C_1) - \nu(C_1)| = \frac{1}{2} \sum_{C_0 \in \mathcal{Q}_0} (\mu(C_0) \wedge \nu(C_0)) \sum_{C_1 \in \mathcal{Q}_1} \Big|\frac{\mu(C_1)}{\mu(C_0)} - \frac{\nu(C_1)}{\nu(C_0)}\Big|,$$

whence (11) with $\ell = 0$. □

We next consider the non compact case.

Lemma 6. *For any $\mu,\nu\in\mathcal{P}(\mathbb{R}^d)$, any $a>1$, any $p>0$,*

$$\mathcal{T}_p(\mu,\nu)\le\sum_{n\ge0}a^{pn}\Big(2^p\varepsilon_p|\mu(G_n^a)-\nu(G_n^a)|+(\mu(G_n^a)\wedge\nu(G_n^a))\mathcal{T}_p(\mathcal{R}_n^a\mu,\mathcal{R}_n^a\nu)\Big),\tag{18}$$

where $\mathcal{R}_n^a\mu$ is the image measure of $\frac{\mu|_{G_n^a}}{\mu(G_n^a)}$ by the map $x\mapsto a^{-n}x$.

Proof. We fix $a>1$ and $p>0$ and consider, for each $n\ge0$, the optimal coupling π_n between $\mathcal{R}_n^a\mu$ and $\mathcal{R}_n^a\nu$ for $\mathcal{T}_p$. We define ξ_n as the image of π_n by the map $(x,y)\mapsto(a^nx,a^ny)$. It holds that ξ_n belongs to $\mathcal{H}(\mu|_{G_n^a}/\mu(G_n^a),\nu|_{G_n^a}/\nu(G_n^a))$ and satisfies

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^p\xi_n(\mathrm{d}x,\mathrm{d}y)=a^{pn}\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^p\pi_n(\mathrm{d}x,\mathrm{d}y)=a^{pn}\mathcal{T}_p(\mathcal{R}_n^a\mu,\mathcal{R}_n^a\nu).\tag{19}$$

Next, we introduce $q=\frac12\sum_{n\ge0}|\mu(G_n^a)-\nu(G_n^a)|$ and we define

$$\xi(\mathrm{d}x,\mathrm{d}y)=\sum_{n\ge0}(\mu(G_n^a)\wedge\nu(G_n^a))\xi_n(\mathrm{d}x,\mathrm{d}y)+q\alpha(\mathrm{d}x)\beta(\mathrm{d}y),\tag{20}$$

where

$$\alpha(\mathrm{d}x)=\frac1q\sum_{n\ge0}(\mu(G_n^a)-\nu(G_n^a))_+\frac{\mu|_{G_n^a}(\mathrm{d}x)}{\mu(G_n^a)}\quad\text{and}\quad\beta(\mathrm{d}y)=\frac1q\sum_{n\ge0}(\nu(G_n^a)-\mu(G_n^a))_+\frac{\nu|_{G_n^a}(\mathrm{d}y)}{\nu(G_n^a)}.$$

Using that $(G_n^a)_{n\ge0}$ is a partition of $\mathbb{R}^d$, that $\xi_n\in\mathcal{H}(\mu|_{G_n^a}/\mu(G_n^a),\nu|_{G_n^a}/\nu(G_n^a))$ and that

$$q=\sum_{n\ge0}(\nu(G_n^a)-\mu(G_n^a))_+=\sum_{n\ge0}(\mu(G_n^a)-\nu(G_n^a))_+=1-\sum_{n\ge0}(\nu(G_n^a)\wedge\mu(G_n^a)),$$

it is easily checked that α and β are probability measures and that $\xi\in\mathcal{H}(\mu,\nu)$. Furthermore, setting $c_p=1\vee2^{p-1}$,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^p\alpha(\mathrm{d}x)\beta(\mathrm{d}y)\le c_p\int_{\mathbb{R}^d\times\mathbb{R}^d}(|x|^p+|y|^p)\alpha(\mathrm{d}x)\beta(\mathrm{d}y)=c_p\int_{\mathbb{R}^d}|x|^p\alpha(\mathrm{d}x)+c_p\int_{\mathbb{R}^d}|y|^p\beta(\mathrm{d}y).$$

We have $|x|<a^n$ for all $x\in G_n^a$, whence

$$\begin{aligned}q\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^p\alpha(\mathrm{d}x)\beta(\mathrm{d}y)\le&c_p\sum_{n\ge0}a^{pn}[(\mu(G_n^a)-\nu(G_n^a))_++(\nu(G_n^a)-\mu(G_n^a))_+]\\=&2^p\varepsilon_p\sum_{n\ge0}a^{pn}|\mu(G_n^a)-\nu(G_n^a)|.\end{aligned}\tag{21}$$

Using that $\mathcal{T}_p(\mu,\nu)\le\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^p\xi(\mathrm{d}x,\mathrm{d}y)$ and (20)-(19)-(21) completes the proof. □

We can now give the

Proof of Proposition 4. Fix μ and ν in $\mathcal{P}(\mathbb{R}^d)$ and $a>1$. For each $n\ge0$, the probability measures $\mathcal{R}_n^a\mu$ and $\mathcal{R}_n^a\nu$, defined in Lemma 6, are supported in $B(0,1)$, and $\mathcal{R}_n^a\mu(C)=\frac{\mu(a^nC\cap G_n^a)}{\mu(G_n^a)}$ for all $C\in\mathcal{B}(\mathbb{R}^d)$. Hence we know from Lemma 5 that there exists some numbers $r_{a,n,\ell}(\mu,\nu)=u_\ell(\mathcal{R}_n^a\mu,\mathcal{R}_n^a\nu)$ satisfying $\sum_{\ell=0}^{k-1}r_{a,n,\ell}(\mu,\nu)\le1$ and

$$r_{a,n,\ell}(\mu,\nu)\le\frac12\sum_{F\in\mathcal{Q}_\ell}\Big(\frac{\mu(a^nF\cap G_n^a)}{\mu(G_n^a)}\wedge\frac{\nu(a^nF\cap G_n^a)}{\nu(G_n^a)}\Big)\sum_{C\text{ child of }F}\Big|\frac{\mu(a^nC\cap G_n^a)}{\mu(a^nF\cap G_n^a)}-\frac{\nu(a^nC\cap G_n^a)}{\nu(a^nF\cap G_n^a)}\Big|$$

and such that

$$\mathcal{T}_p(\mathcal{R}_n^a\mu, \mathcal{R}_n^a\nu) \le \delta_k^p + \sum_{\ell=0}^{k-1} \delta_\ell^p r_{a,n,\ell}(\mu,\nu).$$

Inserting this into (18) completes the proof. □

4. A general estimate concerning the empirical measure

To go further, we need a more precise setting.

Setting 7. *Same points (a) and (b) as in Setting 3.*

(c) There are some constants $A, D > 0$ and $r > 1$ such that for each $k \ge 1$, there is a family $(\mathcal{Q}_{k,\ell})_{\ell=0,\dots,k}$ of nested partitions of G_0 such that $\mathcal{Q}_{k,0} = \{G_0\}$ and such that

$$\forall\, \ell = 1,\dots,k, \qquad |\mathcal{Q}_{k,\ell}| \le A\, r^{d\ell} \tag{22}$$

and

$$\forall\, \ell = 0,\dots,k, \qquad \delta_{k,\ell} = \max_{C\in\mathcal{Q}_{k,\ell}} \sup_{x,y\in C} |x-y| \le D r^{-\ell}. \tag{23}$$

Recall that $\varepsilon_p = 2^{-1} \vee 2^{-p}$. The goal of this section is to prove the following result.

Proposition 8. *We adopt Setting 7, consider $\mu \in \mathcal{P}(\mathbb{R}^d)$ and the associated empirical measure μ_N, see (1). Fix $p > 0$ and assume that $\mathcal{M}_q(\mu) < \infty$ for some $q > p$. For all $a > 1$, all $N \ge 1$,*

$$\mathbb{E}[\mathcal{T}_p(\mu_N,\mu)] \le K_N + \min\{L_N, M_N\},$$

where

$$\begin{aligned}
K_N =& 2^p\varepsilon_p\Big[\Big(2[1-\mu(G_0^a)]\Big) \wedge \sqrt{\frac{1-\mu(G_0^a)}{N}}\Big] + 2^p\varepsilon_p \sum_{n\ge1} a^{pn}\Big[\Big(2\mu(G_n^a)\Big) \wedge \sqrt{\frac{\mu(G_n^a)}{N}}\Big],\\
L_N =& \frac{D^p\sqrt{Ar^d}}{2\sqrt{N}} \sum_{\ell\ge0} r^{(d/2-p)\ell} \sum_{n\ge0} a^{pn}\sqrt{\mu(G_n^a)},\\
M_N =& D^p(1-r^{-p}) \sum_{n\ge0} a^{pn} \sum_{\ell\ge0} r^{-p\ell}\Big(\mu(G_n^a) \wedge \Big[\frac{\sqrt{A}r^{d\ell/2+d}}{2(r^{d/2}-1)}\sqrt{\frac{\mu(G_n^a)}{N}}\Big]\Big).
\end{aligned}$$

We simply write K_N, L_N, M_N for readability, but these quantities also depend on p, a and μ.

Proof. We fix $\mu \in \mathcal{P}(\mathbb{R}^d)$, $a > 1$ and $p > 0$. We also fix $k \ge 1$; we will let $k \to \infty$ at the end of the proof. Applying Proposition 4, with the family $(\mathcal{Q}_{k,\ell})_{\ell=0,\dots,k}$, with μ and $\nu = \mu_N$ and taking expectations, we find

$$\mathbb{E}[\mathcal{T}_p(\mu_N,\mu)] \le U_{N,k} + V_{N,k} + W_{N,k}, \tag{24}$$

where, setting $\rho_{k,a,n,\ell} = \mathbb{E}[(\mu(G_n^a) \wedge \mu_N(G_n^a))r_{k,a,n,\ell}(\mu,\mu_N)]$ (with $r_{k,a,n,\ell}(\mu,\mu_N)$ as defined in Proposition 4 with the family $(\mathcal{Q}_{k,\ell})_{\ell=0,\dots,k}$),

$$
\begin{aligned}
U_{N,k} =& 2^p\varepsilon_p \sum_{n\geq 0} a^{pn}\mathbb{E}[|\mu(G_n^a) - \mu_N(G_n^a)|],\\
V_{N,k} =& \sum_{n\geq 0} a^{pn} \sum_{\ell=0}^{k-1} \delta_{k,\ell}^p \rho_{k,a,n,\ell} \leq D^p \sum_{n\geq 0} a^{pn} \sum_{\ell=0}^{k-1} r^{-p\ell}\rho_{k,a,n,\ell},\\
W_{N,k} =& \delta_{k,k}^p \sum_{n\geq 0} a^{pn}\mu(G_n^a) \leq D^p r^{-pk} \sum_{n\geq 0} a^{pn}\mu(G_n^a). \qquad (25)
\end{aligned}
$$

We used that $\delta_{k,\ell} \leq Dr^{-\ell}$ for all $\ell \in \{0,\dots,k\}$, see (23).

Since $N\mu_N(G_n^a)$ is Binomial$(N,\mu(G_n^a))$-distributed, it holds that $\mathbb{E}[\mu_N(G_n^a)] = \mu(G_n^a)$ and $\mathbb{V}\mathrm{ar}\,[\mu_N(G_n^a)] = N^{-1}\mu(G_n^a)(1-\mu(G_n^a))$, from which we deduce that

$$
\mathbb{E}[|\mu(G_n^a) - \mu_N(G_n^a)|] \leq \Big(2[1-\mu(G_n^a)]\Big) \wedge \Big(2\mu(G_n^a)\Big) \wedge \sqrt{\frac{\mu(G_n^a)(1-\mu(G_n^a))}{N}}.
$$

We used that $|x-y| = |(1-x)-(1-y)| \leq 1-x+1-y$ for all $x,y\in[0,1]$ for the first bound, that $|x-y| \leq x+y$ for the second one, and the Bienaymé-Tchebychev inequality for the third one. All this implies that for all $k\geq 1$,

$$
U_{N,k} \leq K_N. \qquad (26)
$$

Next, we observe that $\sum_{\ell=0}^{k-1}\rho_{k,a,n,\ell} \leq \mu(G_n^a)$ by (8) and we claim that for $\ell = 0,\dots,k-1$,

$$
\rho_{k,a,n,\ell} \leq \frac{1}{2}\sqrt{\frac{|\mathcal{Q}_{k,\ell+1}|\mu(G_n^a)}{N}}. \qquad (27)
$$

Recalling (9) and using that $(\mu(G_n^a) \wedge \mu_N(G_n^a))(\frac{\mu(a^nF\cap G_n^a)}{\mu(G_n^a)} \wedge \frac{\mu_N(a^nF\cap G_n^a)}{\mu_N(G_n^a)}) \leq \mu_N(a^nF\cap G_n^a)$,

$$
\rho_{k,a,n,\ell} \leq \frac{1}{2}\sum_{F\in\mathcal{Q}_{k,\ell}} \sum_{C \text{ child of } F} \mathbb{E}\Big[\Big|\mu_N(a^nC\cap G_n^a) - \frac{\mu_N(a^nF\cap G_n^a)\mu(a^nC\cap G_n^a)}{\mu(a^nF\cap G_n^a)}\Big|\Big]. \qquad (28)
$$

But for C a child of F, the conditional law of $N\mu_N(a^nC\cap G_n^a)$ knowing that $N\mu_N(a^nF\cap G_n^a) = i$ is Binomial$(i, \frac{\mu(a^nC\cap G_n^a)}{\mu(a^nF\cap G_n^a)})$, whence

$$
\mathbb{E}\Big[\Big|\mu_N(a^nC\cap G_n^a) - \frac{\mu_N(a^nF\cap G_n^a)\mu(a^nC\cap G_n^a)}{\mu(a^nF\cap G_n^a)}\Big| \;\Big| N\mu_N(a^nF\cap G_n^a) = i\Big] \leq \sqrt{\frac{i}{N^2}\frac{\mu(a^nC\cap G_n^a)}{\mu(a^nF\cap G_n^a)}}.
$$

Hence, since $\mathbb{E}[\sqrt{N\mu_N(A)}] \leq \sqrt{N\mu(A)}$ because $\mathbb{E}[\mu_N(A)] = \mathbb{E}[\mu(A)]$,

$$
\begin{aligned}
\mathbb{E}\Big[\Big|\mu_N(a^nC\cap G_n^a) - \frac{\mu_N(a^nF\cap G_n^a)\mu(a^nC\cap G_n^a)}{\mu(a^nF\cap G_n^a)}\Big|\Big] \leq& \sqrt{\frac{\mu(a^nC\cap G_n^a)}{N^2\mu(a^nF\cap G_n^a)}}\mathbb{E}\Big[\sqrt{N\mu_N(a^nF\cap G_n^a)}\Big]\\
\leq& \sqrt{\frac{\mu(a^nC\cap G_n^a)}{N}}.
\end{aligned}
$$

This, inserted in (28), proves the claim (27), since by the Cauchy-Schwarz inequality,

$$
\sum_{F\in\mathcal{Q}_{k,\ell}} \sum_{C \text{ child of } F} \sqrt{\frac{\mu(a^nC\cap G_n^a)}{N}} = \sum_{C\in\mathcal{Q}_{k,\ell+1}} \sqrt{\frac{\mu(a^nC\cap G_n^a)}{N}} \leq \sqrt{\frac{|\mathcal{Q}_{k,\ell+1}|\mu(G_n^a)}{N}}.
$$

We deduce from (27) and (22) that

$$\sum_{\ell=0}^{k-1} r^{-p\ell}\rho_{k,a,n,\ell} \le \frac{1}{2}\sum_{\ell=0}^{k-1} r^{-p\ell}\sqrt{\frac{|\mathcal{Q}_{k,\ell+1}|\mu(G_n^a)}{N}} \le \frac{\sqrt{A\mu(G_n^a)}}{2\sqrt{N}}\sum_{\ell=0}^{k-1} r^{-p\ell} r^{d(\ell+1)/2}.$$

Since $V_{N,k} \le D^p \sum_{n\ge 0} a^{pn} \sum_{\ell=0}^{k-1} r^{-p\ell}\rho_{k,a,n,\ell}$, we conclude that for all $k\ge 1$,

$$V_{N,k} \le D^p \sum_{n\ge 0} a^{pn}\frac{\sqrt{A\mu(G_n^a)}}{2\sqrt{N}}\sum_{\ell\ge 0} r^{-p\ell}r^{d(\ell+1)/2} = L_N \tag{29}$$

Next, we set $S_{k,a,n,\ell} = \sum_{i=0}^{\ell}\rho_{k,a,n,i}$ for $\ell = 0,\dots,k-1$ and $S_{k,a,n,-1}=0$ to write

$$\sum_{\ell=0}^{k-1} r^{-p\ell}\rho_{k,a,n,\ell} = \sum_{\ell=0}^{k-1} r^{-p\ell}(S_{k,a,n,\ell} - S_{k,a,n,\ell-1}) = (1-r^{-p})\sum_{\ell=0}^{k-1} r^{-p\ell}S_{k,a,n,\ell} + r^{-pk}S_{k,a,n,k-1}.$$

But for each $\ell = 0,\dots,k-1$, we both have $S_{k,a,n,\ell}\le \mu(G_n^a)$ (since $\sum_{\ell=0}^{k-1}\rho_{k,a,n,\ell}\le\mu(G_n^a)$, as already seen) and, by (27) and (22),

$$S_{k,a,n,\ell} \le \frac{1}{2}\sum_{i=0}^{\ell}\sqrt{\frac{|\mathcal{Q}_{k,i+1}|\mu(G_n^a)}{N}} \le \frac{\sqrt{A\mu(G_n^a)}}{2\sqrt{N}}\sum_{i=0}^{\ell} r^{d(i+1)/2} \le \frac{\sqrt{A}r^{d\ell/2+d}}{2(r^{d/2}-1)}\sqrt{\frac{\mu(G_n^a)}{N}}.$$

Hence

$$\sum_{\ell=0}^{k-1} r^{-p\ell}\rho_{k,a,n,\ell} \le (1-r^{-p})\sum_{\ell=0}^{k-1} r^{-p\ell}\Big(\mu(G_n^a)\wedge\Big[\frac{\sqrt{A}r^{d\ell/2+d}}{2(r^{d/2}-1)}\sqrt{\frac{\mu(G_n^a)}{N}}\Big]\Big) + r^{-pk}\mu(G_n^a).$$

Recalling that $V_{N,k}\le D^p\sum_{n\ge 0}a^{pn}\sum_{\ell=0}^{k-1}r^{-p\ell}\rho_{k,a,n,\ell}$, we conclude that for all $k\ge 1$, it holds that

$$\begin{aligned} V_{N,k} \le & D^p\sum_{n\ge 0}a^{pn}(1-r^{-p})\sum_{\ell\ge 0}r^{-p\ell}\Big(\mu(G_n^a)\wedge\Big[\frac{\sqrt{A}r^{d\ell/2+d}}{2(r^{d/2}-1)}\sqrt{\frac{\mu(G_n^a)}{N}}\Big]\Big) + D^p r^{-pk}\sum_{n\ge 0}a^{pn}\mu(G_n^a) \\ = & M_N + D^p r^{-pk}\sum_{n\ge 0}a^{pn}\mu(G_n^a). \end{aligned} \tag{30}$$

Gathering (24)-(25)-(26)-(29)-(30), we have proved that for all $k\ge 1$,

$$\mathbb{E}[\mathcal{T}_p(\mu_N,\mu)] \le K_N + \min\Big\{L_N, M_N + D^p r^{-pk}\sum_{n\ge 0}a^{pn}\mu(G_n^a)\Big\} + D^p r^{-pk}\sum_{n\ge 0}a^{pn}\mu(G_n^a). \tag{31}$$

Since $\mu(G_n^a)\le\mathcal{M}_q(\mu)a^{(1-n)q}$ for all $n\ge 1$ because $G_n^a\subset B(0,a^{n-1})^c$, and since $q>p$, we deduce that $\sum_{n\ge 0}a^{pn}\mu(G_n^a)<\infty$. Letting $k\to\infty$ in (31) thus completes the proof. □

Let us mention that the penultimate paragraph of this proof, where we handle a discrete integration by parts, is crucial to obtain reasonable constants when $p\in(0,d/2)$.

5. Precise study of some series

Lemma 9. *Fix $r>1$, $\beta\ge\alpha>0$ and $x\ge 0$ and put*

$$\Psi_{r,\alpha,\beta}(x) = \sum_{\ell\ge 0} r^{-\alpha\ell}[1\wedge(x\,r^{\beta\ell})].$$

With the notation $\log_+ x = (\log x) \vee 0$, *it holds that*

$$\Psi_{r,\alpha,\beta}(x) \le \Big(\frac{\log_+(1/x)}{\beta \log r} + \frac{1}{1-r^{-\alpha}}\Big)x \qquad \text{if } \beta = \alpha, \tag{32}$$

$$\Psi_{r,\alpha,\beta}(x) \le \Big(\frac{1}{r^{\beta-\alpha}-1} + \frac{1}{1-r^{-\alpha}}\Big)x^{\alpha/\beta} \qquad \text{if } \beta > \alpha. \tag{33}$$

Proof. We fix $\alpha = \beta > 0$ and prove (32). If $x > 1$, we write

$$\Psi_{r,\alpha,\beta}(x) \le \sum_{\ell \ge 0} r^{-\alpha\ell} = \frac{1}{1-r^{-\alpha}} \le \frac{x}{1-r^{-\alpha}}.$$

If $x \in [0,1]$, we set $t_x = \log(1/x)/(\beta \log r) \ge 0$, $\ell_x = \lfloor t_x \rfloor \in \mathbb{N}$ and $s_x = t_x - \ell_x \in [0,1)$ and write

$$\Psi_{r,\alpha,\beta}(x) \le \sum_{\ell=0}^{\ell_x} x + \sum_{\ell=\ell_x+1}^{\infty} r^{-\alpha\ell} = x(\ell_x+1) + \frac{r^{-\alpha(\ell_x+1)}}{1-r^{-\alpha}} = u + v,$$

where $u = xt_x + x/(1-r^{-\alpha})$ is the desired bound and where, since $x = r^{-\alpha t_x}$,

$$v = x(\ell_x + 1 - t_x) + \frac{r^{-\alpha(\ell_x+1)} - x}{1-r^{-\alpha}} = x\Big[1 - s_x + \frac{r^{-\alpha(1-s_x)}-1}{1-r^{-\alpha}}\Big].$$

To show that $v \le 0$, which will complete the proof of (32), it suffices to prove that $g(u) = u + \frac{r^{-\alpha u}-1}{1-r^{-\alpha}}$ is nonpositive for all $u \in [0,1]$. But $g''(u) = \frac{(\alpha \log r)^2}{1-r^{-\alpha}} r^{-\alpha u} \ge 0$, so that g is convex, and $g(0) = g(1) = 0$. The conclusion follows.

We fix $\beta > \alpha > 0$ and prove (33). If $x > 1$, we write

$$\Psi_{r,\alpha,\beta}(x) \le \sum_{\ell \ge 0} r^{-\alpha\ell} = \frac{1}{1-r^{-\alpha}} \le \frac{x^{\alpha/\beta}}{1-r^{-\alpha}}.$$

If $x \in [0,1]$, we set $t_x = \log(1/x)/(\beta \log r) \ge 0$, $\ell_x = \lfloor t_x \rfloor \in \mathbb{N}$ and $s_x = t_x - \ell_x \in [0,1)$ and write

$$\Psi_{r,\alpha,\beta}(x) \le \sum_{\ell=0}^{\ell_x} x r^{(\beta-\alpha)\ell} + \sum_{\ell=\ell_x+1}^{\infty} r^{-\alpha\ell} \le x\frac{r^{(\beta-\alpha)(\ell_x+1)}}{r^{\beta-\alpha}-1} + \frac{r^{-\alpha(\ell_x+1)}}{1-r^{-\alpha}} = u + v,$$

where $u = x^{\alpha/\beta}/(r^{\beta-\alpha}-1) + x^{\alpha/\beta}/(1-r^{-\alpha})$ is the desired bound and where, since $x = r^{-\beta t_x}$,

$$v = \frac{xr^{(\beta-\alpha)(\ell_x+1)} - x^{\alpha/\beta}}{r^{\beta-\alpha}-1} + \frac{r^{-\alpha(\ell_x+1)} - x^{\alpha/\beta}}{1-r^{-\alpha}} = x^{\alpha/\beta}\Big[\frac{r^{(\beta-\alpha)(1-s_x)}-1}{r^{\beta-\alpha}-1} + \frac{r^{-\alpha(1-s_x)}-1}{1-r^{-\alpha}}\Big].$$

To show that $v \le 0$, which will complete the proof of (33), it suffices to show that $g(u) = \frac{r^{(\beta-\alpha)u}-1}{r^{\beta-\alpha}-1} + \frac{r^{-\alpha u}-1}{1-r^{-\alpha}}$ is nonpositive for all $u \in [0,1]$. But $g''(u) = \frac{((\beta-\alpha)\log r)^2}{r^{\beta-\alpha}-1} r^{(\beta-\alpha)u} + \frac{(\alpha \log r)^2}{1-r^{-\alpha}} r^{-\alpha u} \ge 0$, so that g is convex, and it holds that $g(0) = g(1) = 0$. ☐

6. Theoretical result for a general norm

Recall that $\varepsilon_p = 2^{-1} \vee 2^{-p}$ and that H was defined in (5). Here we prove the following general result, to be applied to some specific norms later.

Proposition 10. *We adopt Setting 7, we fix* $\mu \in \mathcal{P}(\mathbb{R}^d)$ *and consider the associated empirical measure* μ_N, *see* (1). *We fix* $q > p > 0$ *and assume that* $\mathcal{M}_q(\mu) < \infty$.

(i) If $p > d/2$ and $q > 2p$, then for all $N \geq 1$,

$$\mathbb{E}[\mathcal{T}_p(\mu_N,\mu)] \leq 2^p \frac{\kappa_{d,p}}{\sqrt{N}} [\mathcal{M}_q(\mu)]^{p/q} H\Big(\frac{\varepsilon_p}{\kappa_{d,p}}, 2p, q\Big),$$

where

$$\kappa_{d,p} = \frac{D^p \sqrt{Ar^d}}{2^{p+1}(1-r^{d/2-p})}.$$

(ii) If $p = d/2$ and $q > 2p$, then for all $N \geq 1$,

$$\mathbb{E}[\mathcal{T}_p(\mu_N,\mu)] \leq 2^p \frac{\kappa_{d,p,N}}{\sqrt{N}} [\mathcal{M}_q(\mu)]^{p/q} H\Big(\frac{\varepsilon_p}{\kappa_{d,p,N}}, 2p, q\Big),$$

where

$$\kappa_{d,p,N} = \frac{D^p \sqrt{A} r^p}{2^{p+1} p \log r} \log_+ \Big(2(r^{-p} - r^{-2p})\sqrt{\frac{N}{A}}\Big) + \frac{D^p\sqrt{A} r^{2p}}{2^{p+1}(r^p-1)}.$$

(iii) If $p \in (0, d/2)$ and $q > dp/(d-p)$, then for all $N \geq 1$,

$$\mathbb{E}[\mathcal{T}_p(\mu_N,\mu)] \leq 2^p \frac{\kappa_{d,p}}{N^{p/d}} [\mathcal{M}_q(\mu)]^{p/q} H\Big(\frac{2^{1-2p/d}\varepsilon_p}{\kappa_{d,p}}, \frac{dp}{d-p}, q\Big),$$

where

$$\kappa_{d,p} = \frac{D^p A^{p/d} r^p (r^{d/2}-1)^{1-2p/d}}{2^{p+2p/d}(r^{d/2-p}-1)}.$$

Proof. We fix $q > p > 0$. We have $(G_0^a)^c \subset \{x \in \mathbb{R}^d : |x| \geq 1\}$ and $G_n^a \subset \{x \in \mathbb{R}^d : |x| \geq a^{n-1}\}$ for each $n \geq 1$, whence

$$1 - \mu(G_0^a) \leq \mathcal{M}_q(\mu) \quad \text{and} \quad \mu(G_n^a) \leq \mathcal{M}_q(\mu) a^{-q(n-1)} \quad \text{if } n \geq 1. \tag{34}$$

We know from Proposition 8 that $\mathbb{E}[\mathcal{T}_p(\mu_N,\mu)] \leq K_N + \min\{L_N, M_N\}$.

Case (i): $p > d/2$ and $q > 2p$. First, by (34), we have

$$K_N \leq 2^p \varepsilon_p \sqrt{\frac{\mathcal{M}_q(\mu)}{N}} + 2^p \varepsilon_p \sum_{n\geq 1} a^{pn} \sqrt{\frac{\mathcal{M}_q(\mu)}{N a^{q(n-1)}}} = 2^p \varepsilon_p \sqrt{\frac{\mathcal{M}_q(\mu)}{N}} \Big[1 + \frac{a^p}{1-a^{p-q/2}}\Big].$$

Next,

$$L_N \leq \frac{D^p\sqrt{Ar^d}}{2\sqrt{N}} \sum_{\ell \geq 0} r^{(d/2-p)\ell} \Big(1 + \sum_{n\geq 1} a^{pn} \sqrt{\frac{\mathcal{M}_q(\mu)}{a^{q(n-1)}}}\Big) = 2^p \frac{\kappa_{d,p}}{\sqrt{N}} \Big(1 + \sqrt{\mathcal{M}_q(\mu)} \frac{a^p}{1-a^{p-q/2}}\Big)$$

recall that $\kappa_{d,p} = (D/2)^p \sqrt{Ar^d}/[2(1-r^{d/2-p})]$. All in all, we have proved that

$$\mathbb{E}[\mathcal{T}_p(\mu_N,\mu)] \leq \frac{2^p}{\sqrt{N}} \Big(\kappa_{d,p} + \sqrt{\mathcal{M}_q(\mu)} \Big[\varepsilon_p + (\varepsilon_p + \kappa_{d,p}) \frac{a^p}{1-a^{p-q/2}}\Big]\Big). \tag{35}$$

This holds true for any value of $a > 1$ and we optimally choose $a = [q/(2p)]^{2/(q-2p)}$ and set

$$v_{p,q} = \frac{a^p}{1-a^{p-q/2}} = \frac{q}{q-2p}\Big(\frac{q}{2p}\Big)^{2p/(q-2p)}.$$

We thus have

$$\mathbb{E}[\mathcal{T}_p(\mu_N,\mu)] \leq \frac{2^p}{\sqrt{N}} \Big(\kappa_{d,p} + \sqrt{\mathcal{M}_q(\mu)} [\varepsilon_p + (\varepsilon_p + \kappa_{d,p}) v_{p,q}]\Big) = \frac{2^p \kappa_{d,p}}{\sqrt{N}} \Big(1 + \sqrt{\mathcal{M}_q(\mu)} \rho_{d,p,q}\Big),$$

where $\rho_{d,p,q} = \varepsilon_p/\kappa_{d,p} + (\varepsilon_p + \kappa_{d,p}) v_{p,q}/\kappa_{d,p}$.

For any $\alpha > 0$, we may apply this formula to μ^α, the image measure of μ by the map $x \mapsto \alpha x$, which satisfies $\mathbb{E}[\mathcal{T}_p(\mu_N^\alpha, \mu^\alpha)] = \alpha^p \mathbb{E}[\mathcal{T}_p(\mu_N, \mu)]$ and $\mathcal{M}_q(\mu^\alpha) = \alpha^q \mathcal{M}_q(\mu)$. We thus get

$$\mathbb{E}[\mathcal{T}_p(\mu_N, \mu)] \le \frac{2^p \kappa_{d,p}}{\sqrt{N}} \frac{1}{\alpha^p}\Big(1 + \sqrt{\alpha^q \mathcal{M}_q(\mu)}\rho_{d,p,q}\Big).$$

We optimally choose $\alpha = [(q-2p)\rho_{d,p,q}\sqrt{\mathcal{M}_q(\mu)}/(2p)]^{-2/q}$ and find

$$\begin{aligned}
\mathbb{E}[\mathcal{T}_p(\mu_N, \mu)] \le & \frac{2^p \kappa_{d,p}}{\sqrt{N}} [\mathcal{M}_q(\mu)]^{p/q} \Big(\rho_{d,p,q} \frac{q-2p}{2p}\Big)^{2p/q} \frac{q}{q-2p} \\
= & \frac{2^p \kappa_{d,p}}{\sqrt{N}} [\mathcal{M}_q(\mu)]^{p/q} \Big(\frac{\varepsilon_p}{\kappa_{d,p}} \frac{q-2p}{2p} + \frac{\varepsilon_p + \kappa_{d,p}}{\kappa_{d,p}} v_{p,q} \frac{q-2p}{2p}\Big)^{2p/q} \frac{q}{q-2p} \\
= & \frac{2^p \kappa_{d,p}}{\sqrt{N}} [\mathcal{M}_q(\mu)]^{p/q} \Big(\frac{\varepsilon_p}{\kappa_{d,p}} \frac{q-2p}{2p} + \frac{\varepsilon_p + \kappa_{d,p}}{\kappa_{d,p}} \Big(\frac{q}{2p}\Big)^{q/(q-2p)}\Big)^{2p/q} \frac{q}{q-2p} \\
= & \frac{2^p \kappa_{d,p}}{\sqrt{N}} [\mathcal{M}_q(\mu)]^{p/q} H\Big(\frac{\varepsilon_p}{\kappa_{d,p}}, 2p, q\Big).
\end{aligned}$$

Case (ii): $p = d/2$ and $q > 2p$. Exactly as in Case (i),

$$K_N \le 2^p \varepsilon_p \sqrt{\frac{\mathcal{M}_q(\mu)}{N}} \Big[1 + \frac{a^p}{1 - a^{p-q/2}}\Big].$$

We next write

$$\begin{aligned}
M_N \le & D^p(1 - r^{-p}) \sum_{n \ge 0} a^{pn} \mu(G_n^a) \sum_{\ell \ge 0} r^{-p\ell} \Big(1 \wedge \Big[\frac{\sqrt{A} r^{d\ell/2+d}}{2(r^{d/2}-1)} \frac{1}{\sqrt{N\mu(G_n^a)}}\Big]\Big) \\
= & D^p(1 - r^{-p}) \sum_{n \ge 0} a^{pn} \mu(G_n^a) \Psi_{r,p,d/2}\Big(\frac{\sqrt{A} r^d}{2(r^{d/2}-1)\sqrt{N\mu(G_n^a)}}\Big).
\end{aligned}$$

By (32) with $\alpha = \beta = p = d/2$, we can bound M_N by

$$D^p(1 - r^{-p}) \sum_{n \ge 0} a^{pn} \mu(G_n^a) \Big[\frac{\log_+(2(r^{-d/2} - r^{-d})\sqrt{N\mu(G_n^a)/A})}{p \log r} + \frac{1}{1 - r^{-p}}\Big] \frac{\sqrt{A} r^d}{2(r^{d/2}-1)\sqrt{N\mu(G_n^a)}}$$

$$\le \frac{D^p \sqrt{A} r^{2p}}{2(r^p - 1)\sqrt{N}} \Big[\frac{(1 - r^{-p}) \log_+(2(r^{-p} - r^{-2p})\sqrt{N/A})}{p \log r} + 1\Big] \sum_{n \ge 0} a^{pn} \sqrt{\mu(G_n^a)}$$

since $\mu(G_n^a) \le 1$. Observing that, by (34),

$$\sum_{n \ge 0} a^{pn} \sqrt{\mu(G_n^a)} \le 1 + \sum_{n \ge 1} a^{pn} \sqrt{\frac{\mathcal{M}_q(\mu)}{a^{q(n-1)}}} = 1 + \sqrt{\mathcal{M}_q(\mu)} \frac{a^p}{1 - a^{p-q/2}},$$

and recalling that

$$\kappa_{d,p,N} = \frac{(D/2)^p \sqrt{A} r^p \log_+(2(r^{-p} - r^{-2p})\sqrt{N/A})}{2p \log r} + \frac{(D/2)^p \sqrt{A} r^{2p}}{2(r^p - 1)},$$

we conclude that

$$M_N \le 2^p \frac{\kappa_{d,p,N}}{\sqrt{N}} \Big(1 + \sqrt{\mathcal{M}_q(\mu)} \frac{a^p}{1 - a^{p-q/2}}\Big).$$

All in all, we have proved that

$$\mathbb{E}[\mathcal{T}_p(\mu_N, \mu)] \le \frac{2^p}{\sqrt{N}} \Big(\kappa_{d,p,N} + \sqrt{\mathcal{M}_q(\mu)} \Big[\varepsilon_p + (\varepsilon_p + \kappa_{d,p,N}) \frac{a^p}{1 - a^{p-q/2}}\Big]\Big).$$

From there we conclude exactly as in Case (i) (compare the above formula to (35)) that

$$\mathbb{E}[\mathcal{T}_p(\mu_N,\mu)] \le 2^p \frac{\kappa_{d,p,N}}{\sqrt{N}} [\mathcal{M}_q(\mu)]^{p/q} H\Big(\frac{\varepsilon_p}{\kappa_{d,p,N}}, 2p, q\Big).$$

Case (iii): $p \in (0, d/2)$ and $q > dp/(d-p)$. We write, using that $2p/d \in (0,1)$ and then (34),

$$\begin{aligned}
K_N \le& 2^p \varepsilon_p \Big[2(1-\mu(G_0^a))\Big]^{1-2p/d} \Big[\sqrt{\frac{1-\mu(G_0^a)}{N}}\Big]^{2p/d} + 2^p \varepsilon_p \sum_{n\ge1} a^{pn} \Big[2\mu(G_n^a)\Big]^{1-2p/d} \Big[\sqrt{\frac{\mu(G_n^a)}{N}}\Big]^{2p/d} \\
\le& 2^p \varepsilon_p \Big[2\mathcal{M}_q(\mu)\Big]^{1-2p/d} \Big[\sqrt{\frac{\mathcal{M}_q(\mu)}{N}}\Big]^{2p/d} + 2^p \varepsilon_p \sum_{n\ge1} a^{pn} \Big[\frac{2\mathcal{M}_q(\mu)}{a^{q(n-1)}}\Big]^{1-2p/d} \Big[\sqrt{\frac{\mathcal{M}_q(\mu)}{N a^{q(n-1)}}}\Big]^{2p/d} \\
=& 2^p \varepsilon_p \frac{2^{1-2p/d}[\mathcal{M}_q(\mu)]^{1-p/d}}{N^{p/d}} + 2^p \varepsilon_p \frac{2^{1-2p/d}[\mathcal{M}_q(\mu)]^{1-p/d}}{N^{p/d}} \sum_{n\ge1} a^{pn-q(1-p/d)(n-1)} \\
=& 2^p \varepsilon_p \frac{2^{1-2p/d}[\mathcal{M}_q(\mu)]^{1-p/d}}{N^{p/d}} \Big(1 + \frac{a^p}{1-a^{p-q+pq/d}}\Big).
\end{aligned}$$

We used that $p - q + pq/d < 0$ because $q > dp/(d-p)$. Next,

$$M_N \le D^p(1-r^{-p}) \sum_{n\ge0} a^{pn} \mu(G_n^a) \Psi_{r,p,d/2}\Big(\frac{\sqrt{A} r^d}{2(r^{d/2}-1)\sqrt{N\mu(G_n^a)}}\Big)$$

as in Case (ii). Thus, using (33) with $\alpha = p$ and $\beta = d/2$,

$$\begin{aligned}
M_N \le& D^p(1-r^{-p}) \Big[\frac{1}{r^{d/2-p}-1} + \frac{1}{1-r^{-p}}\Big] \sum_{n\ge0} a^{pn} \mu(G_n^a) \Big(\frac{\sqrt{A} r^d}{2(r^{d/2}-1)\sqrt{N\mu(G_n^a)}}\Big)^{2p/d} \\
=& 2^p \frac{\kappa_{d,p}}{N^{p/d}} \sum_{n\ge0} a^{pn} [\mu(G_n^a)]^{1-p/d},
\end{aligned}$$

because

$$\Big(\frac{D}{2}\Big)^p (1-r^{-p}) \Big[\frac{1}{r^{d/2-p}-1} + \frac{1}{1-r^{-p}}\Big] \Big(\frac{\sqrt{A} r^d}{2(r^{d/2}-1)}\Big)^{2p/d} = \frac{D^p A^{p/d} r^p (r^{d/2}-1)^{1-2p/d}}{2^{p+2p/d}(r^{d/2-p}-1)} = \kappa_{d,p}.$$

But, using (34),

$$\sum_{n\ge0} a^{pn} [\mu(G_n^a)]^{1-p/d} \le 1 + \sum_{n\ge1} a^{pn} \Big[\frac{\mathcal{M}_q(\mu)}{a^{q(n-1)}}\Big]^{1-p/d} = 1 + [\mathcal{M}_q(\mu)]^{1-p/d} \frac{a^p}{1-a^{p-q+pq/d}},$$

whence

$$M_N \le 2^p \frac{\kappa_{d,p}}{N^{p/d}} \Big(1 + [\mathcal{M}_q(\mu)]^{1-p/d} \frac{a^p}{1-a^{p-q+pq/d}}\Big).$$

All in all, we have

$$\begin{aligned}
\mathbb{E}[\mathcal{T}_p(\mu_N,\mu)] \le& \frac{2^p}{N^{p/d}} \Big(\kappa_{d,p} + [\mathcal{M}_q(\mu)]^{1-p/d} \Big[2^{1-2p/d}\varepsilon_p + (2^{1-2p/d}\varepsilon_p + \kappa_{d,p}) \frac{a^p}{1-a^{p-q+pq/d}}\Big]\Big) \\
=& \frac{2^p}{N^{p/d}} \Big(\kappa_{d,p} + [\mathcal{M}_q(\mu)]^{p/\tau} \Big[2^{1-2p/d}\varepsilon_p + (2^{1-2p/d}\varepsilon_p + \kappa_{d,p}) \frac{a^p}{1-a^{p-pq/\tau}}\Big]\Big),
\end{aligned}$$

where we have set $\tau = dp/(d-p)$. We choose $a = [q/\tau]^{\tau/(p(q-\tau))} > 1$, for which

$$v_{d,p,q} = \frac{a^p}{1-a^{p-pq/\tau}} = \frac{q}{q-\tau} \Big(\frac{q}{\tau}\Big)^{\tau/(q-\tau)}.$$

Thus

$$\mathbb{E}[\mathcal{T}_p(\mu_N,\mu)] \leq \frac{2^p\kappa_{d,p}}{N^{p/d}}\Big(1+[\mathcal{M}_q(\mu)]^{p/\tau}\Big[\frac{2^{1-2p/d}\varepsilon_p}{\kappa_{d,p}}+(2^{1-2p/d}\varepsilon_p+\kappa_{d,p})\frac{\upsilon_{d,p,q}}{\kappa_{d,p}}\Big]\Big)$$
$$=\frac{2^p\kappa_{d,p}}{N^{p/d}}\Big(1+[\mathcal{M}_q(\mu)]^{p/\tau}\rho_{d,p,q}\Big),$$

where $\rho_{d,p,q}=2^{1-2p/d}\varepsilon_p/\kappa_{d,p}+(2^{1-2p/d}\varepsilon_p+\kappa_{d,p})\upsilon_{d,p,q}/\kappa_{d,p}$. As in Case (i), we deduce that

$$\mathbb{E}[\mathcal{T}_p(\mu_N,\mu)] \leq \frac{2^p\kappa_{d,p}}{N^{p/d}}\frac{1}{\alpha^p}\Big(1+[\alpha^q\mathcal{M}_q(\mu)]^{p/\tau}\rho_{d,p,q}\Big)$$

for all $\alpha>0$. With $\alpha=[\rho_{d,p,q}\mathcal{M}_q^{p/\tau}(q-\tau)/\tau]^{-\tau/(pq)}$, which is optimal, we find

$$\mathbb{E}[\mathcal{T}_p(\mu_N,\mu)] \leq \frac{2^p\kappa_{d,p}}{N^{p/d}}[\mathcal{M}_q(\mu)]^{p/q}\Big(\rho_{d,p,q}\frac{q-\tau}{\tau}\Big)^{\tau/q}\frac{q}{q-\tau}$$
$$=\frac{2^p\kappa_{d,p}}{N^{p/d}}[\mathcal{M}_q(\mu)]^{p/q}\Big(\frac{2^{1-2p/d}\varepsilon_p}{\kappa_{d,p}}\frac{q-\tau}{\tau}+\frac{2^{1-2p/d}\varepsilon_p+\kappa_{d,p}}{\kappa_{d,p}}\upsilon_{d,p,q}\frac{q-\tau}{\tau}\Big)^{\tau/q}\frac{q}{q-\tau}$$
$$=\frac{2^p\kappa_{d,p}}{N^{p/d}}[\mathcal{M}_q(\mu)]^{p/q}\Big(\frac{2^{1-2p/d}\varepsilon_p}{\kappa_{d,p}}\frac{q-\tau}{\tau}+\frac{2^{1-2p/d}\varepsilon_p+\kappa_{d,p}}{\kappa_{d,p}}\Big(\frac{q}{\tau}\Big)^{q/(q-\tau)}\Big)^{\tau/q}\frac{q}{q-\tau}$$
$$=\frac{2^p\kappa_{d,p}}{N^{p/d}}[\mathcal{M}_q(\mu)]^{p/q}H\Big(\frac{2^{1-2p/d}\varepsilon_p}{\kappa_{d,p}},\tau,q\Big)$$

as desired. □

7. Conclusion for the maximum norm

Here we consider the maximum norm $|\cdot|_\infty$. We claim that Setting 7-(c) holds true with $A=1$, $D=2$ and $r=2$. Indeed, consider, for each $\ell\geq 0$, the natural partition $\mathcal{Q}_\ell$ of $G_0=[-1,1]^d$ into $2^{d\ell}$ translations of $[-2^{-\ell},2^{-\ell}]^d$ (we actually have to remove some of the common faces, but this is of course not an issue). Then for any $k\geq 1$, $(\mathcal{Q}_\ell)_{\ell=0,\dots,k}$ is a family of nested partitions of G_0, we have $|\mathcal{Q}_\ell|=2^{d\ell}=Ar^{d\ell}$ for all $\ell=1,\dots,k$ and $\delta_\ell=\max_{C\in\mathcal{Q}_\ell}\sup_{x,y\in C}|x-y|=2\times 2^{-\ell}=Dr^{-\ell}$ for all $\ell=0,\dots,k$.

We thus may apply Proposition 10 with these values $A=1$, $D=2$ and $r=2$. This gives Theorem 1 (with the norm $|\cdot|_\infty$) with the announced formulas, which we now check.

If first $p>d/2$, we have

$$\kappa_{d,p}^{(\infty)}=\frac{D^p\sqrt{Ar^d}}{2^{p+1}(1-r^{d/2-p})}=\frac{2^{d/2-1}}{1-2^{d/2-p}}.$$

If next $p=d/2$, we have

$$\kappa_{d,p,N}^{(\infty)}=\frac{D^p\sqrt{A}r^p}{2^{p+1}p\log r}\log_+\Big(2(r^{-p}-r^{-2p})\sqrt{\frac{N}{A}}\Big)+\frac{D^p\sqrt{A}r^{2p}}{2^{p+1}(r^p-1)}$$
$$=\frac{2^{p-1}}{p\log 2}\log_+\Big((2^{1-p}-2^{1-2p})\sqrt{N}\Big)+\frac{2^{p-1}}{1-2^{-p}}.$$

If finally $p\in(0,d/2)$, we have

$$\kappa_{d,p}^{(\infty)}=\frac{D^pA^{p/d}r^p(r^{d/2}-1)^{1-2p/d}}{2^{p+2p/d}(r^{d/2-p}-1)}=\frac{2^{p-2p/d}(1-2^{-d/2})^{1-2p/d}}{1-2^{p-d/2}}.$$

8. Conclusion for the Euclidean norm

We now work with the Euclidean norm $|\cdot|_2$. The following lemma is slight modification of Boissard-Le Gouic [2, Lemma 2.1] and Weed-Bach [12, Proposition 3]. Recall that K_d was defined in (4) and that $B_2(x,r) = \{y \in \mathbb{R}^d : |y-x|_2 < r\}$.

Lemma 11. *For any $k \geq 1$, any $r > 2$, there exists a family $(\mathcal{Q}_{k,\ell})_{\ell=0,\dots,k}$ of nested partitions of $B_2(0,1)$ such that $\mathcal{Q}_{k,0} = \{B_2(0,1)\}$, with $|\mathcal{Q}_{k,\ell}| \leq K_d(r-2)^{-d}r^d r^{d\ell}$ for all $\ell = 1,\dots,k$ and $\delta_{k,\ell} = \max_{C\in\mathcal{Q}_{k,\ell}} \sup_{x,y\in C} |x-y|_2 \leq 2r^{-\ell}$ for all $\ell = 0,\dots,k$.*

Proof. We fix $k \geq 1$ and $r > 2$ and consider $\gamma \in (0,r)$ to be chosen later. By definition (3) of N_r, for each $\ell = 1,\dots,k$, there is a covering $(C_{k,\ell,i})_{i=1,\dots,N_{\gamma r^{-\ell}}}$ of $B_2(0,1)$ by balls with radius $\gamma r^{-\ell}$ (observe that for each $\ell = 1,\dots,k$, it holds that $\gamma r^{-\ell} < 1$).

We now define, by reverse induction on ℓ, a family $(\mathcal{Q}_{k,\ell})_{\ell=1,\dots,k}$ of nested partitions of $B_2(0,1)$, such that for all $\ell = 1,\dots,k$, it holds that $|\mathcal{Q}_{k,\ell}| = N_{\gamma r^{-\ell}}$ and

$$\delta_{k,\ell} := \max_{C\in\mathcal{Q}_{k,\ell}} \sup_{x,y\in C} |x-y|_2 \leq 2\gamma r^{-\ell} \sum_{i=0}^{k-\ell} (2/r)^i.$$

We first define $F_{k,k,1} = C_{k,k,1} \cap B_2(0,1)$ and by induction, for every $i = 2,\dots,N_{\gamma r^{-k}}$, $F_{k,k,i} = (C_{k,k,i} \cap B_2(0,1)) \setminus (\cup_{j=1}^{i-1} C_{k,k,j})$. Then $\mathcal{Q}_{k,k} = (F_{k,k,i})_{i=1,\dots,N_{\gamma r^{-k}}}$ is a partition of $B_2(0,1)$, of which all the elements have a $|\cdot|_2$-diameter smaller than $2\gamma r^{-k} = 2\gamma r^{-k}\sum_{i=0}^{k-k}(2/r)^i$.

We next assume that $\mathcal{Q}_{k,\ell+1}$ has been built, for some $\ell \in \{2,\cdots,k-1\}$, and we build $\mathcal{Q}_{k,\ell}$. We first set, for $i = 1,\dots,N_{\gamma r^{-\ell}}$,

$$G_{k,\ell,i} = \bigcup_{j\in\mathcal{A}_{k,\ell,i}} F_{k,\ell+1,j}, \quad \text{where} \quad \mathcal{A}_{k,\ell,i} = \Big\{ j \in \{1,\dots,N_{\gamma r^{-\ell-1}}\} : F_{k,\ell+1,j} \cap C_{k,\ell,i} \neq \emptyset \Big\}.$$

Since the $|\cdot|_2$-diameter of $C_{k,\ell,i}$ equals $2\gamma r^{-\ell}$ and the $|\cdot|_2$-diameter of $F_{k,\ell+1,j}$ is smaller, for each j, than $2\gamma r^{-\ell-1}\sum_{i=0}^{k-\ell-1}(2/r)^i$, we deduce that the $|\cdot|_2$-diameter of $G_{k,\ell,i}$ is smaller than

$$2\gamma r^{-\ell} + 4\gamma r^{-\ell-1} \sum_{i=0}^{k-\ell-1} (2/r)^i = 2\gamma r^{-\ell}\Big(1 + (2/r)\sum_{i=0}^{k-\ell-1}(2/r)^i\Big) = 2\gamma r^{-\ell}\sum_{i=0}^{k-\ell}(2/r)^i.$$

We then set $F_{k,\ell,1} = G_{k,\ell,1}$ and, by induction, for $i = 2,\dots,N_{\gamma r^{-\ell}}$, $F_{k,\ell,i} = G_{k,\ell,i} \setminus (\cup_{j=1}^{i-1} G_{k,\ell,j})$. Then $\mathcal{Q}_{k,\ell} = (F_{k,\ell,i})_{i=1,\dots,N_{\gamma r^{-\ell}}}$ is a partition of $B_2(0,1)$, composed of elements of which the $|\cdot|_2$-diameter is smaller than $2\gamma r^{-\ell}\sum_{i=0}^{k-\ell}(2/r)^i$. Finally, each element of $\mathcal{Q}_{k,\ell}$ is a union of elements of $\mathcal{Q}_{k,\ell+1}$, so that $\mathcal{Q}_{k,\ell+1}$ is a refinement of $\mathcal{Q}_{k,\ell}$.

We finally set $\mathcal{Q}_{k,0} = \{B_2(0,1)\}$. With the choice $\gamma = (r-2)/r$, the family $(\mathcal{Q}_{k,\ell})_{\ell=0,\dots,k}$ meets all the requirements of the statement. Indeed, for each $\ell = 1,\dots,k$, we have $|\mathcal{Q}_{k,\ell}| = N_{\gamma r^{-\ell}} \leq K_d(\gamma r^{-\ell})^{-d} = K_d(r-2)^{-d}r^d r^{d\ell}$ by (4) and $\max_{C\in\mathcal{Q}_{k,\ell}}\sup_{x,y\in C}|x-y|_2 \leq 2\gamma r^{-\ell}\sum_{i=0}^{k-\ell}(2/r)^i \leq 2r^{-\ell}$. This also holds true with $\ell = 0$ since $\sup_{x,y\in B_2(0,1)}|x-y|_2 = 2$. ☐

Thus Setting 7-(c) holds with any $r > 2$, with $A = K_d(r-2)^{-d}r^d$ and $B = 2$, so that we may apply Proposition 10 with these values. Optimizing in $r > 2$, this gives Theorem 1 (with the norm $|\cdot|_2$) with the announced formulas, which we now check.

If first $p > d/2$, we find $\kappa^{(2)}_{d,p} = \min\{\kappa^{(2)}_{d,p,r} : r > 2\}$, where

$$\kappa^{(2)}_{d,p,r} = \frac{D^p\sqrt{Ar^d}}{2^{p+1}(1-r^{d/2-p})} = \frac{\sqrt{K_d}}{2}\frac{r^d}{(r-2)^{d/2}(1-r^{d/2-p})}.$$

If next $p = d/2$, we have $\kappa^{(2)}_{d,p,N} = \min\{\kappa^{(2)}_{d,p,N,r} : r > 2\}$, where

$$\begin{aligned}\kappa^{(2)}_{d,p,N,r} =& \frac{D^p\sqrt{A}r^p}{2^{p+1}p\log r}\log_+\Big(2(r^{-p}-r^{-2p})\sqrt{\frac{N}{A}}\Big) + \frac{D^p\sqrt{A}r^{2p}}{2^{p+1}(r^p-1)}\\ =& \frac{\sqrt{K_d}}{2}\frac{r^{2p}}{(r-2)^p p\log r}\log_+\Big(2(r-2)^p(r^{-2p}-r^{-3p})\sqrt{\frac{N}{K_d}}\Big) + \frac{\sqrt{K_d}}{2}\frac{r^{3p}}{(r-2)^p(r^p-1)}.\end{aligned}$$

If finally $p \in (0, d/2)$, we have $\kappa^{(2)}_{d,p} = \min\{\kappa^{(2)}_{d,p,r} : r > 2\}$, where

$$\kappa^{(2)}_{d,p,r} = \frac{D^pA^{p/d}r^p(r^{d/2}-1)^{1-2p/d}}{2^{p+2p/d}(r^{d/2-p}-1)} = \Big(\frac{K_d}{4}\Big)^{p/d}\frac{r^{2p}(1-r^{-d/2})^{1-2p/d}}{(r-2)^p(1-r^{p-d/2})}.$$

9. The case of a low order finite moment

We finally handle the case where μ has a low order moment. We only treat the case of the maximum norm for simplicity. We thus may apply Proposition 8 with $A = 1$, $D = 2$ and $r = 2$, see the beginning of Section 7.

Proof of Theorem 2. We consider $p > 0$, $q \in (p, \min\{2p, dp/(d-p)\})$, $\mu \in \mathcal{P}(\mathbb{R}^d)$ and the associated empirical measure μ_N. We know that $\mathbb{E}[\mathcal{T}^{(\infty)}_p(\mu_N,\mu)] \le K_N + \min\{L_N, M_N\}$ by Proposition 8, and we have

$$\mu(G^a_n) \le \mathcal{M}^{(\infty)}_q(\mu)a^{-q(n-1)} \quad \text{if } n \ge 1. \tag{36}$$

as usual. First,

$$\begin{aligned}K_N \le& \frac{2^p\varepsilon_p}{\sqrt{N}} + 2^p\varepsilon_p\sum_{n\ge1}a^{pn}\Big[\frac{2\mathcal{M}^{(\infty)}_q(\mu)}{a^{q(n-1)}} \wedge \sqrt{\frac{\mathcal{M}^{(\infty)}_q(\mu)}{Na^{q(n-1)}}}\Big]\\ =& \frac{2^p\varepsilon_p}{\sqrt{N}} + 2^p\varepsilon_p a^p\sum_{n\ge0}a^{(p-q)n}\Big[\Big(2\mathcal{M}^{(\infty)}_q(\mu)\Big)\wedge\Big(\sqrt{\frac{\mathcal{M}^{(\infty)}_q(\mu)}{N}}a^{qn/2}\Big)\Big]\\ =& \frac{2^p\varepsilon_p}{\sqrt{N}} + 2^{p+1}\varepsilon_p\mathcal{M}^{(\infty)}_q(\mu)a^p\Psi_{a,q-p,q/2}\Big(\frac{1}{2\sqrt{N\mathcal{M}^{(\infty)}_q(\mu)}}\Big).\end{aligned}$$

Since $q/2 > q-p$ because $q < 2p$, we may apply (33) with $r = a$, with $\alpha = q-p$ and $\beta = q/2$:

$$\begin{aligned}K_N \le& \frac{2^p\varepsilon_p}{\sqrt{N}} + 2^{p+1}\varepsilon_p\mathcal{M}^{(\infty)}_q(\mu)a^p\Big[\frac{1}{a^{p-q/2}-1} + \frac{1}{1-a^{p-q}}\Big]\Big(\frac{1}{2\sqrt{N\mathcal{M}^{(\infty)}_q(\mu)}}\Big)^{2(q-p)/q}\\ =& \frac{2^p\varepsilon_p}{\sqrt{N}} + 2^p\varepsilon_p\rho_a[\mathcal{M}^{(\infty)}_q(\mu)]^{p/q}\frac{2^{2p/q-1}}{N^{(q-p)/q}},\end{aligned}$$

where $\rho_a = a^p[1/(a^{p-q/2}-1) + 1/(1-a^{p-q})]$. Next, recalling that $A = 1$, $D = 2$ and $r = 2$,

$$M_N \le 2^p(1-2^{-p})\sum_{n\ge0}a^{pn}\sum_{\ell\ge0}2^{-p\ell}\Big[\mu(G^a_n)\wedge\Big(\frac{2^{d-1}2^{d\ell/2}}{2^{d/2}-1}\sqrt{\frac{\mu(G^a_n)}{N}}\Big)\Big] \le M_{N,1} + M_{N,2},$$

where we separate the cases $n = 0$ and $n \geq 1$, i.e.

$$M_{N,1} = 2^p(1-2^{-p})\sum_{\ell\geq 0} 2^{-p\ell}\Big[1 \wedge \Big(\frac{2^{d-1}2^{d\ell/2}}{2^{d/2}-1}\sqrt{\frac{1}{N}}\Big)\Big] \leq 2^p(1-2^{-p})\sum_{\ell\geq 0} 2^{-p\ell} = 2^p,$$

and

$$\begin{aligned}
M_{N,2} =& 2^p(1-2^{-p})\sum_{n\geq 1} a^{pn}\sum_{\ell\geq 0} 2^{-p\ell}\Big[\frac{\mathcal{M}_q^{(\infty)}(\mu)}{a^{q(n-1)}} \wedge \Big(\frac{2^{d-1}2^{d\ell/2}}{2^{d/2}-1}\sqrt{\frac{\mathcal{M}_q^{(\infty)}(\mu)}{Na^{q(n-1)}}}\Big)\Big]\\
=& 2^p(1-2^{-p})a^p\sum_{\ell\geq 0} 2^{-p\ell}\sum_{n\geq 0} a^{pn}\Big[\frac{\mathcal{M}_q^{(\infty)}(\mu)}{a^{qn}} \wedge \Big(\frac{2^{d-1}2^{d\ell/2}}{2^{d/2}-1}\sqrt{\frac{\mathcal{M}_q^{(\infty)}(\mu)}{Na^{qn}}}\Big)\Big]\\
=& 2^p(1-2^{-p})a^p\sum_{\ell\geq 0} 2^{-p\ell}\sum_{n\geq 0} a^{(p-q)n}\Big[\mathcal{M}_q^{(\infty)}(\mu) \wedge \Big(\frac{2^{d-1}2^{d\ell/2}}{2^{d/2}-1}\sqrt{\frac{\mathcal{M}_q^{(\infty)}(\mu)}{N}}a^{qn/2}\Big)\Big]\\
=& 2^p(1-2^{-p})a^p\mathcal{M}_q^{(\infty)}(\mu)\sum_{\ell\geq 0} 2^{-p\ell}\Psi_{a,q-p,q/2}\Big(\frac{2^{d-1}2^{d\ell/2}}{(2^{d/2}-1)\sqrt{N\mathcal{M}_q^{(\infty)}(\mu)}}\Big).
\end{aligned}$$

By (33) with $r = a$, $\alpha = q - p$ and $\beta = q/2$, recalling that $\rho_a = a^p[1/(a^{p-q/2}-1) + 1/(1-a^{p-q})]$,

$$\begin{aligned}
M_{N,2} \leq& 2^p(1-2^{-p})\mathcal{M}_q^{(\infty)}(\mu)\rho_a\sum_{\ell\geq 0} 2^{-p\ell}\Big(\frac{2^{d-1}2^{d\ell/2}}{(2^{d/2}-1)\sqrt{N\mathcal{M}_q^{(\infty)}(\mu)}}\Big)^{2(q-p)/q}\\
=& 2^p\rho_a\frac{[\mathcal{M}_q^{(\infty)}(\mu)]^{p/q}}{N^{(q-p)/q}}\Big(\frac{2^{d-1}}{2^{d/2}-1}\Big)^{2(q-p)/q}\frac{1-2^{-p}}{1-2^{d-p-dp/q}},
\end{aligned}$$

observe that $d - p - dp/q < 0$ because $q < dp/(d-p)$. All in all, we conclude that

$$\mathbb{E}[\mathcal{T}_p^{(\infty)}(\mu_N,\mu)] \leq 2^p\Big(\frac{\varepsilon_p}{\sqrt{N}} + 1 + \rho_a\frac{[\mathcal{M}_q^{(\infty)}(\mu)]^{p/q}}{N^{(q-p)/q}}\Big[\varepsilon_p 2^{2p/q-1} + \Big(\frac{2^{d-1}}{2^{d/2}-1}\Big)^{2(q-p)/q}\frac{1-2^{-p}}{1-2^{d-p-dp/q}}\Big]\Big).$$

For any $\alpha > 0$, we may apply this above formula to μ^α, the image measure of μ by the map $x \mapsto \alpha x$, for which $\mathbb{E}[\mathcal{T}_p^{(\infty)}(\mu_N^\alpha,\mu^\alpha)] = \alpha^p\mathbb{E}[\mathcal{T}_p^{(\infty)}(\mu_N,\mu)]$ and $\mathcal{M}_q^{(\infty)}(\mu^\alpha) = \alpha^q\mathcal{M}_q^{(\infty)}(\mu)$. We get

$$\mathbb{E}[\mathcal{T}_p^{(\infty)}(\mu_N,\mu)] \leq \frac{2^p}{\alpha^p}\Big(\frac{\varepsilon_p}{\sqrt{N}} + 1 + \rho_a\frac{[\alpha^q\mathcal{M}_q^{(\infty)}(\mu)]^{p/q}}{N^{(q-p)/q}}\Big[\varepsilon_p 2^{2p/q-1} + \Big(\frac{2^{d-1}}{2^{d/2}-1}\Big)^{2(q-p)/q}\frac{1-2^{-p}}{1-2^{d-p-dp/q}}\Big]\Big).$$

Letting $\alpha \to \infty$, we find

$$\mathbb{E}[\mathcal{T}_p^{(\infty)}(\mu_N,\mu)] \leq 2^p\frac{[\mathcal{M}_q^{(\infty)}(\mu)]^{p/q}}{N^{(q-p)/q}}\rho_a\Big[\varepsilon_p 2^{2p/q-1} + \Big(\frac{2^{d-1}}{2^{d/2}-1}\Big)^{2(q-p)/q}\frac{1-2^{-p}}{1-2^{d-p-dp/q}}\Big]\Big).$$

Since $\rho_a = a^p/(a^{p-q/2}-1) + a^p/(1-a^{p-q})$ and since this result holds for any $a \in (1,\infty)$, the proof is complete. □

References

[1] M. Ajtai, J. Komlós, G. Tusnády. On optimal matchings. *Combinatorica* 4 (1984), 259–264.
[2] E. Boissard, T. Le Gouic. On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. *Ann. Inst. Henri Poincaré Probab. Stat.* 50 (2014), 539–563.
[3] F. Bolley, A. Guillin, C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields* 137 (2007), 541–593.

[4] S. Dereich, M. Scheutzow, R. Schottstedt. Constructive quantization: approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.* 49 (2013), 1183–1203.
[5] V. Dobrić, J.E. Yukich. Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.* 8 (1995), 97–118.
[6] R.M. Dudley. The speed of mean Glivenko-Cantelli convergence. *Ann. Math. Statist.* 40 (1968), 40–50.
[7] N. Fournier, A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* 162 (2015), 707–738.
[8] J. Lei. Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces. *Bernoulli* 26 (2020), 767–798.
[9] H. Luschgy, G. Pagès. Marginal and functional quantization of stochastic processes. Book in preparation.
[10] C.A. Rogers. Covering a sphere with spheres. Mathematika 10 (1963), 157–164.
[11] J.L. Verger-Gaugry. Covering a ball with smaller equal balls in $\mathbb{R}^n$. *Discrete Comput. Geom.* 33 (2005), 143–155.
[12] J. Weed, F. Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli* 25 (2019), 2620–2648.

N. Fournier, Sorbonne Université, CNRS, Laboratoire de Probabilités, Statistiques et Modélisation, F-75005 Paris, France. E-mail: nicolas.fournier@sorbonne-universite.fr