Submitted to Bernoulli

Supplement to "Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance"

JONATHAN WEED and FRANCIS BACH

A. Omitted proofs

A.1. Proof of Proposition 1

We begin by giving an informal outline of the idea of the proof.

Consider a partition $\{Q_i\}_{i \in \mathcal{I}}$ of S, for some index set $\mathcal{I}$. The measures μ and ν both induce measures on each set in the partition. We will transport μ to ν by first moving mass *between* sets in this partition, and then moving mass *within* each set in the partition. If $\mu(Q_i) \neq \nu(Q_i)$ for one of the sets Q_i, we we need to transport an amount of mass equal to $|\mu(Q_i) - \nu(Q_i)|$ into or out of Q_i. In total, we can transport the mass that μ assigns to each set in the partition to its proper set under ν for a total cost of

$$\sum_{i \in \mathcal{I}} |\mu(Q_i) - \nu(Q_i)| \operatorname{diam}(S) \leq \sum_{i \in \mathcal{I}} |\mu(Q_i) - \nu(Q_i)| \,,$$

where we use the fact that $\operatorname{diam}(S) \leq \operatorname{diam}(X) \leq 1$ by assumption.

After the first step of the transport plan, μ has been transported so that each set in the partition contains the correct total amount of mass. It therefore suffices in the second step to properly arrange the mass *within* each set. Moving the mass within Q_i cannot cost more than $\operatorname{diam}(Q_i)$, so the total cost of arranging the mass within each set is at most

$$\sum_{i \in \mathcal{I}} \nu(Q_i) \operatorname{diam}(Q_i) \leq \max_{i \in \mathcal{I}} \operatorname{diam}(Q_i) \,.$$

We have obtained a transport of μ to ν for a total cost of approximately

$$\max_{i \in \mathcal{I}} \operatorname{diam}(Q_i) + \sum_{i \in \mathcal{I}} |\mu(Q_i) - \nu(Q_i)| \,.$$

This "single scale" bound is generally not tight, but a more refined bound can be obtained by applying the above argument recursively: instead of naïvely bounding the cost of moving the mass within Q_i by the quantity $\operatorname{diam}(Q_i)$, we can partition Q_i into smaller sets and estimate the cost of moving the mass within Q_i by first moving it between the sets of the partition before moving it within each smaller set. Iterating the argument k^* times yields the bound.

We now show how to make the above argument precise. Given two measures μ and ν on X, write $\mathcal{C}(\mu, \nu)$ for the set of couplings between μ and ν; that is, for the set of measures on $X \times X$ whose projection onto the first and second coordinate correspond to μ and ν respectively.

Fix a $k^* \geq 1$. We will define two sequences of measure π_k and ρ_k on X for $1 \leq k \leq k^*$ such that $\sum_{k=1}^{k^*} \pi_k \leq \mu$ and $\sum_{k=1}^{k^*} \rho_k \leq \nu$. Given such a sequence, we set $\mu_1 := \mu$ and $\nu_1 := \nu$ and write

$$\mu_k := \mu - \sum_{\ell=1}^{k-1} \pi_\ell$$
$$\nu_k := \nu - \sum_{\ell=1}^{k-1} \rho_\ell$$

for $k \leq k^* + 1$.

Note that if $\gamma_k \in \mathcal{C}(\pi_k, \rho_k)$ for $1 \leq k \leq k^*$ and $\gamma_{k^*+1} \in \mathcal{C}(\mu_{k^*+1}, \nu_{k^*+1})$, then

$$\sum_{k=1}^{k^*+1} \gamma_k \in \mathcal{C}\left(\sum_{k=1}^{k^*} \pi_k + \mu_{k^*+1}, \sum_{k=1}^{k^*} \rho_k + \nu_{k^*+1}\right) = \mathcal{C}(\mu, \nu)\,,$$

therefore

$$W_p^p(\mu, \nu) \leq \sum_{k=1}^{k^*} W_p^p(\pi_k, \rho_k) + W_p^p(\mu_{k^*+1}, \nu_{k^*+1})\,.$$

For $k \geq 1$, define

$$\pi_k := \sum_{\substack{Q_i^k \in \mathcal{Q}^k \\ \mu_k(Q_i^k) > 0}} \left(1 - \frac{\nu_k(Q_i^k)}{\mu_k(Q_i^k)}\right)_+ \mu_k|_{Q_i^k}\,,$$
$$\rho_k := \sum_{\substack{Q_i^k \in \mathcal{Q}^k \\ \nu_k(Q_i^k) > 0}} \left(1 - \frac{\mu_k(Q_i^k)}{\nu_k(Q_i^k)}\right)_+ \nu_k|_{Q_i^k}\,.$$

Note that $0 \leq \pi_k \leq \mu_k$ and $0 \leq \rho_k \leq \nu_k$ for all k, hence $0 \leq \mu_k \leq \mu$ and $0 \leq \nu_k \leq \nu$ for all k as well.

Lemma A.1. *If $Q \in \mathcal{Q}^{k-1}$, then*

$$\mu_k(Q) = \nu_k(Q)$$
$$\pi_k(Q) = \rho_k(Q)\,.$$

Moreover,

$$\pi_k(S) = \rho_k(S) \le \sum_{Q_i^k \in \mathcal{Q}^k} |\mu(Q_i^k) - \nu(Q_i^k)| \,.$$

Lemma A.2. *If α and β are two measures on X such that*

$$\alpha(Q) = \beta(Q)$$

for all $Q \in \mathcal{Q}^{k-1}$, then

$$W_p^p(\alpha, \beta) \le \delta^{(k-1)p} \alpha(S) \,.$$

We can now obtain the final bound. By Lemmas A.1 and A.2,

$$W_p^p(\pi_k, \rho_k) \le \delta^{(k-1)p} \sum_{Q_i^k \in \mathcal{Q}^k} |\mu(Q_i^k) - \nu(Q_i^k)|$$

and

$$W_p^p(\mu_{k^*+1}, \nu_{k^*+1}) \le \delta^{k^* p} \mu_{k^*+1}(S) \le \delta^{k^* p} \mu(S) \le \delta^{k^* p} \,.$$

The bound follows.

A.2. Proof of Proposition 2

We prove the inequalities in order. If $d < d_H(\mu)$, then by [1, Proposition 10.3] there exists a compact set K with positive mass and a $r_0 > 0$ such that

$$\mu(B(x, r)) \le r^d$$

for all $r \le r_0$ and all $x \in K$. (See also the proof of [2, Corollary 12.16].) Let $\tau < \mu(K)/2$. If S is any set with $\mu(S) \ge 1 - \tau$, then $\mu(S \cap K) > \mu(K)/2$. If $\mathcal{N}_\varepsilon(S) = N$, then in particular there exists a covering of $S \cap K$ by at most N balls of radius ε whose centers all lie in K. Indeed, any set of diameter at most ε which intersects $S \cap K$ is contained in a ball of radius ε whose center is in K. If $\varepsilon \le r_0$, then each such ball satisfies $\mu(B(x, r)) \le \varepsilon^d$, so

$$N \ge \varepsilon^{-d} \mu(K)/2 \,.$$

We therefore have for all τ sufficiently small,

$$\liminf_{\varepsilon \to 0} \frac{\log \mathcal{N}_\varepsilon(\mu, \tau)}{-\log \varepsilon} \ge d \,.$$

Thus $d_*(\mu) \ge d$. Since $d < d_H(\mu)$ was arbitrary, we have $d_H(\mu) \le d_*(\mu)$, as desired.

That $d_*(\mu) \le d_p^*(\mu)$ follows from the simple observation that for all positive α and τ,

$$\liminf_{\varepsilon \to 0} d_\varepsilon(\mu, \tau) \le \liminf_{\varepsilon \to 0} d_\varepsilon(\mu, \varepsilon^\alpha) \,.$$

Finally, if $d_M(\mu) \ge 2p$, then setting $s > d_M(\mu)$ yields

$$\limsup_{\varepsilon \to 0} d_\varepsilon(\mu, \varepsilon^{\frac{sp}{s-2p}}) \le \limsup_{\varepsilon \to 0} d_\varepsilon(\mu) = d_M(\mu) < s \,,$$

so $d_p^*(\mu) \le s$. Since $s > d_M(\mu)$ was arbitrary, we obtain $d_p^*(\mu) \le d_M(\mu)$. □

A.3. Proof of Proposition 3

Write $N_k := \mathcal{N}_{3^{-(k+1)}}(S_k)$, where we recall that by assumption $S_k \subseteq S$. For $1 \leq k \leq k^*$, let $C^k := \{C_1^k, \dots\}$ be a finite covering of S by balls of diameter $3^{-(k+1)}$ such that $C_1^k, \dots, C_{N_k}^k$ covers S_k. Such a covering can always be found by choosing an optimal covering of S_k and extending this covering to a covering of all of S. Since $\mathcal{N}_{3^{-(k^*+1)}}(S) < \infty$, this requires only a finite number of additional balls.

We begin by constructing $\mathcal{Q}^{k^*}$. Let $\mathcal{Q}_1^{k^*} := C_1^{k^*}$, and for $1 < \ell \leq |C^{k^*}|$ let

$$\mathcal{Q}_\ell^{k^*} := C_\ell^{k^*} \setminus \left(\bigcup_{n=1}^{\ell-1} \mathcal{Q}_n^{k^*} \right) .$$

Let $\mathcal{Q}^{k^*} := \{\mathcal{Q}_1^{k^*}, \dots\}$. Note that $\mathrm{diam}(\mathcal{Q}_\ell^{k^*}) \leq \mathrm{diam}(C_\ell^{k^*}) = 3^{-(k^*+1)} < 3^{-k^*}$, that $\mathcal{Q}^{k^*}$ forms a partition of S, and that at most N_{k^*} elements of $\mathcal{Q}^{k^*}$ intersect S_{k^*}.

We now show how to construct $\mathcal{Q}^k$ from $\mathcal{Q}^{k+1}$ and C^k. Let

$$\mathcal{Q}_1^k := \bigcup_{\substack{Q \in \mathcal{Q}^{k+1} \\ Q \cap C_1^k \neq \emptyset}} Q \,,$$

and for $1 < \ell \leq |C^{k^*}|$ let

$$\mathcal{Q}_\ell^k := \Big(\bigcup_{\substack{Q \in \mathcal{Q}^{k+1} \\ Q \cap C_\ell^k \neq \emptyset}} Q \Big) \setminus \Big(\bigcup_{n=1}^{\ell-1} \mathcal{Q}_n^k \Big) .$$

Let $\mathcal{Q}^k := \{\mathcal{Q}_1^k, \dots\}$.

The sets in $\mathcal{Q}^k$ clearly form a partition of S, and by construction at most N_k elements of $\mathcal{Q}^k$ intersect S_k Moreover, since $\mathrm{diam}(C_\ell^k) \leq 3^{-(k+1)}$ for all ℓ and $\mathrm{diam}(Q) \leq 3^{-(k+1)}$ for all $Q \in \mathcal{Q}^{k+1}$, the distance between any two points in $\mathcal{Q}_\ell^k$ is at most $3 \cdot 3^{-(k+1)} = 3^{-k}$, so each element of $\mathcal{Q}^k$ has diameter at most 3^{-k}. Finally, since each set in $\mathcal{Q}^k$ is the union of sets in $\mathcal{Q}^{k+1}$, the partition $\mathcal{Q}^{k+1}$ refines $\mathcal{Q}^k$, as desired. □

A.4. Proof of Proposition 7

The only inequality that does not follow from Proposition 2 is the first. By absolute continuity, for all $\tau > 0$ there exists a $\sigma > 0$ such that any set T for which $\mu(T) \geq 1 - \tau$ satisfies $\mathcal{H}^d(T) \geq \sigma$. If $\mathcal{H}^d(T) \geq \sigma$ then , then in particular for any covering $\{B(x_i, \varepsilon)\}$ of T by balls of radius ε for ε sufficiently small, we must have $\sum_i \varepsilon^d \geq \sigma/2$. Therefore such a covering contains at least $\sigma\varepsilon^{-d}/2$ balls, so

$$\frac{\log \mathcal{N}_\varepsilon(\mu, \tau)}{-\log \varepsilon} \geq d + \frac{\log(\sigma/2)}{-\log \varepsilon} \,,$$

and taking limits yields that $d_*(\mu) \geq d$, as desired. □

A.5. Proof of Proposition 11

For all integers $k \geq 0$, denote by N_k the smallest positive integer such that N_k is a power of two and $\delta_{N_k} \leq 2^{-k}$. Such an integer always exists because the sequence δ_n decreases to 0. We require the following lemma, whose proof is deferred to Section B.

Lemma A.3. *The sequence N_{k+1}/N_k is bounded.*

Let m be an integer large enough that $N_{k+1}/N_k \leq 2^m$ for all n. Let $\mathcal{Q}$ be the standard dyadic partition of $[0,1]$, with $\mathcal{Q}^k$ being a partition of $[0,1]^m$ consisting of 2^{km} cubes of side length 2^{-k}.

Our measure μ will satisfy $\mathcal{N}_{2^{-k}}(\mu) = N_{k-2}$ for all $k \geq 2$. We will define a sequence of measures $\{\mu_k\}_{k=2}^{\infty}$ iteratively and construct μ as their limit in the weak topology.

Let μ_2 be the uniform distribution on $[0,1/4]^m$. For each positive integer k, the measure μ_k will be supported on N_{k-2} cubes in $\mathcal{Q}^k$, and will be uniform on its support. We will call a cube $Q_i \in \mathcal{Q}^k$ *live* if $\mu_k(Q_i) \neq 0$.

Fix an ordering $x_0, \ldots, x_{2^m-1}$ of the 2^m elements of $\{0,1\}^m$. To produce μ_{k+1} from μ_k, divide each live cube of μ_k into 2^m cubes of side length $2^{-(k+1)}$. The ordering of $\{0,1\}^m$ induces an order on these 2^m subcubes.

Given a live $Q \in \mathcal{Q}^k$, define the restriction $\mu_{k+1}|_Q$ by requiring that $\mu_{k+1}(Q) = \mu_k(Q)$ and that $\mu_{k+1}|_Q$ be uniform on the union of the first N_{k+1}/N_k subcubes of Q. Note that N_{k+1}/N_k is an integer because both N_{k+1} and N_k are powers of 2, and by assumption $N_{k+1}/N_k \leq 2^m$, the total number of subcubes of Q. Since $\mathcal{Q}^k$ forms a partition of $[0,1]^m$, combining the measures $\mu_{k+1}|_Q$ for $Q \in \mathcal{Q}^k$ yields a probability measure μ_{k+1} on $[0,1]^m$. By Prokhorov's theorem, this sequence of measures μ_k possesses a subsequence converging in distribution to some measure μ.

The following lemma collects necessary properties of μ. Its proof appears in Section B.

Lemma A.4. *If $N_k \leq n < N_{k+1}$, then*

$$\mathcal{N}_{2^{-k-4}}(\mu, 1/2) > n$$

Moreover,

$$2^{-k-2} \leq \delta_n \leq 2^{-k}$$

and

$$2^{-k-4} \leq n^{-1/d_n} \leq 2^{-k}\,.$$

We can now obtain the lower bound. Let ν be any measure supported on at most n points. If $N_k \leq n < N_{k+1}$, then by Lemma A.4, if $X \sim \mu$, then

$$\mathbb{P}[\min_{y \in \mathrm{supp}(\nu)} \|X - y\|_\infty \leq 2^{-k-5}] < 1/2\,.$$

Markov's inequality therefore implies for any coupling (X, Y) of μ and ν that

$$\mathbb{E}[\|X-Y\|_\infty^p]^{1/p} \geq 2^{-k-5}\mathbb{P}[\min_{y \in \mathrm{supp}(\nu)} \|X - y\|_\infty > 2^{-k-5}]^{1/p} \geq 2^{-k-6} \geq 2^{-6} n^{-1/d_n}\,,$$

as claimed. □

A.6. Proof of Proposition 19

Both claims are standard, and details can be found in [3, Theorem 5.10]. The first follows follows from the assumption that X is a bounded Polish space. For the second, we use the fact that the supremum is achieved by an f satisfying

$$f(x) = \inf_{y \in X} f^c(y) + D(x,y)^p \quad \forall x \in X\,. \tag{1}$$

Let f be a function achieving the supremum in (4) and satisfying (1). By adding a constant to f and f^c, we can assume that $\sup_{x \in X} f(x) = 1$. Then for all $y \in X$,

$$f^c(y) = \sup_{x \in X} f(x) - D(x,y)^p \geq 0\,,$$

and (1) then implies

$$f(x) \geq 0 \quad \forall x \in X\,,$$

as claimed. □

A.7. Proof of Lemma 1

If $X \sim \nu$ is independent of $Z \sim \mathcal{N}(0, \sigma^2 I)$, then by considering the coupling $(X, X+Z)$, we obtain

$$W_p^p(\mu, \nu) \leq \mathbb{E}[\|Z\|^p] \leq \sigma^p (d+2p)^{p/2}\,,$$

where we have applied a standard bound for the moments of the χ^2 distribution.

We can couple empirical distributions $\hat{\mu}_n$ and $\hat{\nu}_n$ by letting $X_1, \dots, X_n \sim \nu$ i.i.d. and $Z_1, \dots, Z_n \sim \mathcal{N}(0, \sigma^2)$ i.i.d. and independent of $\{X_i\}$ and setting

$$\hat{\nu}_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$
$$\hat{\mu}_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i + Z_i}\,.$$

We have for this coupling

$$W_p^p(\hat{\mu}_n, \hat{\nu}_n) \leq \frac{1}{n} \sum_{i=1}^n \|Z_i\|^p\,,$$

and

$$\mathbb{E} W_p^p(\hat{\mu}_n, \hat{\nu}_n) \leq \sigma^p (d+2p)^{p/2}$$

The triangle for W_p then implies

$$\begin{aligned}\mathbb{E} W_p^p(\mu, \hat{\mu}_n) &\leq \mathbb{E}(W_p(\mu,\nu) + W_p(\nu, \hat{\nu}_n) + W_p(\hat{\mu}_n, \hat{\nu}_n))^p \\ &\leq 3^{p-1} \mathbb{E} W_p^p(\nu, \hat{\nu}_n) + 2 \cdot 3^{p-1} \sigma^p (d+2p)^{p/2}\,.\end{aligned}$$

□

B. Additional lemmas

B.1. Proof of Lemma A.1

We first show that for any $\ell < k$, if $Q \in \mathcal{Q}^\ell$, then

$$\mu_k(Q) = \nu_k(Q)\,.$$

Suppose first that $Q \in \mathcal{Q}^{k-1}$. By definition, $\mu_k = \mu_{k-1} - \pi_{k-1}$. We obtain

$$\mu_k(Q) = (\mu_{k-1} - \pi_{k-1})(Q) = \min\{\mu_{k-1}(Q), \nu_{k-1}(Q)\}\,,$$

and likewise

$$\nu_k(Q) = \min\{\mu_{k-1}(Q), \nu_{k-1}(Q)\}\,.$$

Since $\mathcal{Q}$ is a dyadic partition, any $Q \in \mathcal{Q}^\ell$ for $\ell < k$ can be written as a disjoint union of $Q_1, \dots, Q_m \in \mathcal{Q}^{k-1}$. Hence

$$\mu_k(Q) = \sum_{i=1}^m \mu_k(Q_i) = \sum_{i=1}^m \nu_k(Q_i) = \nu_k(Q)\,,$$

as claimed.

Note that this also implies for any $\ell < k$, if $Q \in \mathcal{Q}^\ell$, then

$$\pi_k(Q) = \mu_k(Q) - \mu_{k+1}(Q) = \nu_k(Q) - \nu_{k+1}(Q) = \rho_k(Q)\,.$$

We now prove the bound on $\pi_k(S)$. By definition,

$$\rho_k(S) = \sum_{Q_i^k \in \mathcal{Q}^k} (\nu_k(Q_i^k) - \mu_k(Q_i^k))_+ = \frac{1}{2} \sum_{Q_i^k \in \mathcal{Q}^k} |\nu_k(Q_i^k) - \mu_k(Q_i^k)|\,.$$

We now show that, for any $Q \in \mathcal{Q}^{k-1}$, there exist scalars $c_1, c_2 \in [0,1]$ depending on Q such that

$$\mu_k|_Q = c_1 \mu|_Q$$
$$\nu_k|_Q = c_2 \nu|_Q\,.$$

We proceed by induction on k. By symmetry, it suffices to prove the claim for μ_k and μ. Since $\mu_1 = \mu$, it holds for $k = 1$. Now assume $\mu_{k-1}|_Q = c_1 \mu|_Q$. We have

$$\mu_k|_Q = \mu_{k-1}|_Q - \pi_{k-1}|_Q = \min\left\{\frac{\nu_{k-1}(Q)}{\mu_{k-1}(Q)}, 1\right\} \mu_{k-1}|_Q = c_1' \mu|_Q\,,$$

where $c_1' = \min\left\{\frac{\nu_{k-1}(Q)}{\mu_{k-1}(Q)}, 1\right\} c_1$. This proves the claim.

Now, given such a $Q \in \mathcal{Q}^{k-1}$ and $c_1, c_2 \in [0,1]$, we have $\mu_k(Q) = \nu_k(Q)$, so

$$c_1 \mu(Q) = c_2 \nu(Q)\,.$$

Summing over the elements of $\mathcal{Q}^k$ contained in Q, we obtain

$$\begin{aligned}\sum_{Q_i^k \subset Q} |\mu_k(Q_i^k) - \nu_k(Q_i^k)| &= \sum_{Q_i^k \subset Q} |c_1 \mu(Q_i^k) - c_2 \nu(Q_i^k)| \\ &\le \sum_{Q_i^k \subset Q} c_1 |\mu(Q_i^k) - \nu(Q_i^k)| + \sum_{Q_i^k \subset Q} \nu(Q_i^k)|c_1 - c_2| \\ &= \sum_{Q_i^k \subset Q} c_1 |\mu(Q_i^k) - \nu(Q_i^k)| + c_2 |\mu(Q) - \nu(Q)| \\ &\le \sum_{Q_i^k \subset Q} (c_1 + c_2)|\mu(Q_i^k) - \nu(Q_i^k)| \\ &\le 2 \sum_{Q_i^k \subset Q} |\mu(Q_i^k) - \nu(Q_i^k)| \,.\end{aligned}$$

Finally, summing over all $Q \in \mathcal{Q}^{k-1}$ yields

$$\rho_k(S) = \frac{1}{2} \sum_{Q_i^k \in \mathcal{Q}^k} |\nu_k(Q_i^k) - \mu_k(Q_i^k)| \le \sum_{Q_i^k \in \mathcal{Q}^k} |\mu(Q_i^k) - \nu(Q_i^k)| \,,$$

as claimed. □

B.2. Proof of Lemma A.2

Let

$$\gamma := \sum_{\substack{Q_i^{k-1} \in \mathcal{Q}^{k-1} \\ \alpha(Q_i^{k-1}) > 0}} \frac{\alpha \otimes \beta}{\alpha(Q_i^{k-1})} \,.$$

Note that $\gamma \in C(\alpha, \beta)$. Indeed, for any measurable $U \subset S$, since $\mathcal{Q}^{k-1}$ is a partition of S, we have

$$\gamma(S, U) = \sum_{Q_i^{k-1} \in \mathcal{Q}^{k-1}} \frac{\alpha(Q_i^{k-1}) \beta(Q_i^{k-1} \cap U)}{\alpha(Q_i^{k-1})} = \beta(U) \,.$$

On the other hand, by assumption, $\alpha(Q_i^{k-1}) = \beta(Q_k^{k-1})$, so

$$\gamma(U, S) = \sum_{Q_i^{k-1} \in \mathcal{Q}^{k-1}} \frac{\alpha(Q_i^{k-1} \cap U) \beta(Q_i^{k-1})}{\beta(Q_k^{k-1})} = \alpha(U) \,.$$

We have

$$\begin{aligned}
\int D(x,y)^p \mathrm{d}\gamma(x,y) &= \sum_{Q_i^{k-1} \in \mathcal{Q}^{k-1}} \frac{1}{\alpha(Q_i^{k-1})} \int_{Q_i^{k-1}} D(x,y)^p \mathrm{d}\alpha(x)\mathrm{d}\beta(y) \\
&\le \sum_{Q_i^{k-1} \in \mathcal{Q}^{k-1}} \beta(Q_i^{k-1}) \operatorname{diam}(Q_i^{k-1})^p \\
&\le \alpha(S)\delta^{(k-1)p} .
\end{aligned}$$

□

B.3. Proof of Lemma A.3

By assumption, there exist constants c and α such that $\frac{1}{c}n^\alpha \le \delta_n \le cn^\alpha$ for all n sufficiently large. Let $M = (2c^2)^{-1/\alpha}$. Then for n sufficiently large,

$$\delta_{Mn} \le c(Mn)^\alpha = \frac{1}{2c}n^\alpha \le \frac{1}{2}\delta_n .$$

This implies that for k sufficiently large, $\delta_{N_k} \le 2^{-k}$ implies that $\delta_{MN_k} \le 2^{-k-1}$, so that $N_{k+1} \le MN_k$. Hence $N_{k+1}/N_k \le M$ for all k sufficiently large, so N_{k+1}/N_k is bounded. □

B.4. Proof of Lemma A.4

We first show the key property of μ. For any $x \in [0,1]^m$ and $r > 0$, denote by $B(x,r)$ the open ℓ_∞ ball of radius r around x. We claim that for any $x \in [0,1]^m$ and $\ell \ge 2$,

$$\mu(B(x, 2^{-\ell-1})) \le \frac{1}{N_{\ell-2}} .$$

It suffices to show this claim for all μ_k with $k \ge \ell$, and conclude via the fact that μ is the weak limit of a subsequence of the measures. The bound in question certainly holds when $B(x, 2^{-\ell-1})$ exactly coincides with one of the cubes in $\mathcal{Q}^\ell$, since each live cube in $\mathcal{Q}^\ell$ has mass exactly $1/N_{\ell-2}$ by construction.

For all other x, note that the restriction of μ_k to each live cube in $\mathcal{Q}^\ell$ is the same measure. In general, the cube $B(x, 2^{-\ell-1})$ intersects 2^m cubes cubes in $\mathcal{G}_\ell$, and we can partition $B(x, 2^{-\ell-1})$ into 2^m pieces which, via translation, exactly cover a cube of $\mathcal{Q}^\ell$. Each piece has mass at most the mass of the corresponding piece in a live cube, hence the measure is at most the measure of a live cube.

This property immediately implies a bound on the number of balls needed to cover any set S such that $\mu(S) \ge 1/2$. Since each ball of diameter $2^{-\ell}$ has mass at most $1/N_{\ell-2}$, to cover a set of mass $1/2$ requires at least $N_{\ell-2}/2$ balls. Therefore for all $\ell \ge 2$,

$$\mathcal{N}_{2^{-\ell}}(\mu, 1/2) \ge N_{\ell-2}/2 . \tag{2}$$

For all $k \geq 0$, because N_{k+1} is a power of 2 greater than one, $N_{k+1}/2$ is also a power of 2. The definition of N_{k+1} therefore implies that $\delta_{N_{k+1}/2} > 2^{-k-1}$. Because $\frac{\log n}{-\log \delta_n}$ is nondecreasing and at least 1 for all $n \geq 2$, we have for all $k \geq 0$

$$\frac{\log N_{k+1}}{-\log \delta_{N_{k+1}}} \geq \frac{\log(N_{k+1}/2)}{-\log \delta_{N_{k+1}/2}} \geq \frac{\log N_{k+1}}{-\log(\delta_{N_{k+1}/2}/2)}$$

and therefore $\delta_{N_{k+1}} \geq \frac{1}{2}\delta_{N_{k+1}/2} > 2^{-k-2}$, so that $N_{k+2} > N_{k+1}$. Since N_{k+2} is also a power of 2, in particular $N_{k+2} \geq 2N_{k+1}$. This implies $N_{k+2}/2 > n$.

Choosing $\ell = k+4$ in (2) yields

$$\mathcal{N}_{2^{-k-4}}(\mu, 1/2) > n\,.$$

This proves the first claim.

We have just noted that $\delta_{N_{k+1}} > 2^{-k-2}$, and the definition of N_k implies $\delta_{N_k} \leq 2^{-k}$. If $N_k \leq n < N_{k+1}$, then the fact that δ_n is nonincreasing in n yields

$$2^{-k-2} < \delta_{N_{k+1}} \leq \delta_n \leq \delta_{N_k} \leq 2^{-k}\,.$$

This proves the second claim.

To prove the third claim, we first note that the definition of d_n implies that

$$n^{-1/d_n}$$

is nonincreasing as n increases. We can therefore prove an upper bound on n^{-1/d_n} by proving an upper bound on $N_k^{-1/d_{N_k}}$.

Recall that

$$d_{N_k} = \inf_{\varepsilon>0} \max\left\{d_{\geq\varepsilon}(\mu, \varepsilon^p), \frac{\log N_k}{-\log \varepsilon}\right\}\,.$$

Choosing $\varepsilon = 2^{-(k+2)}$ yields

$$d_{N_k} \leq \max\{d_{\geq 2^{-(k+2)}}(\mu), \frac{\log_2 N_k}{k+2}\}\,.$$

To bound the first term, note that if $\varepsilon' \in [2^{-\ell}, 2^{-\ell+1})$, then $\mathcal{N}_{\varepsilon'}(\mu) \leq \mathcal{N}_{2^{-\ell}}(\mu) = N_{\ell-2}$. We also have $\varepsilon' < 2^{-\ell+1} < \delta_{N_{\ell-2}}$. Therefore $d'_\varepsilon = \frac{\log \mathcal{N}_{\varepsilon'}(\mu)}{-\log \varepsilon'} \leq \frac{\log N_{\ell-2}}{\delta_{N_{\ell-2}}}$.

The assumption that $\frac{\log n}{-\log \delta_n}$ is nonincreasing therefore implies

$$d_{\geq 2^{-k+2}}(\mu) \leq \max_{2\leq\ell\leq k+2} \frac{\log N_{\ell-2}}{-\log \delta_{N_{\ell-2}}} \leq \frac{\log N_k}{-\log \delta_{N_k}} \leq \frac{\log_2 N_k}{k}\,.$$

We obtain

$$d_{N_k} \leq \frac{\log_2 N_k}{k}\,,$$

so $n^{-1/d_n} \le N_k^{-1/d_{N_k}} \le 2^{-k}$.

To obtain the lower bound, note that if $\varepsilon \le 2^{-(k+4)}$, then

$$d_{\ge\varepsilon}(\mu, \varepsilon^p) \ge d_{2^{-(k+4)}}(\mu, 1/2) > \frac{\log_2 n}{k+4}\,,$$

where we have used the fact proved above that $\mathcal{N}_{2^{-(k+4)}}(\mu, 1/2) > n$. If $\varepsilon > 2^{-(k+4)}$, then

$$\frac{\log n}{-\log\varepsilon} > \frac{\log_2 n}{k+4}\,.$$

Combining these bounds yields

$$d_n = \inf_{\varepsilon>0} \max\left\{d_{\ge\varepsilon}(\mu, \varepsilon^p), \frac{\log n}{-\log\varepsilon}\right\} > \frac{\log_2 n}{k+4}\,,$$

so

$$n^{-1/d_n} > 2^{-(k+4)}\,,$$

as claimed. □

References

[1] Falconer, K. (1997). *Techniques in fractal geometry* **3**. Wiley Chichester (W. Sx.).

[2] Graf, S. and Luschgy, H. (2007). *Foundations of quantization for probability distributions*. Springer.

[3] Villani, C. (2008). *Optimal transport: old and new* **338**. Springer Science & Business Media.