Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2013, Vol. 49, No. 4, 1183–1203
DOI: 10.1214/12-AIHP489

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Constructive quantization: Approximation by empirical measures

Steffen Dereich[a], Michael Scheutzow[b] and Reik Schottstedt[c]

[a]*Westfälische Wilhelms-Universität Münster, Institut für Mathematische Statistik, Einsteinstr. 62, D-48149 Münster, Deutschland. E-mail: steffen.dereich@uni-muenster.de*
[b]*Technische Universität Berlin, Institut für Mathematik, Straße des 17. Juni 136, D-10623 Berlin, Deutschland. E-mail: ms@math.tu-berlin.de*
[c]*Philipps-Universität Marburg, Fachbereich Mathematik und Informatik, Hans-Meerwein-Str., D-35032 Marburg, Deutschland. E-mail: schottstedt@mathematik.uni-marburg.de*

Received 23 October 2011; revised 16 February 2012; accepted 18 March 2012

Abstract. In this article, we study the approximation of a probability measure μ on $\mathbb{R}^d$ by its empirical measure $\hat{\mu}_N$ interpreted as a random quantization. As error criterion we consider an averaged pth moment Wasserstein metric. In the case where $2p < d$, we establish fine upper and lower bounds for the error, a *high resolution formula*. Moreover, we provide a universal estimate based on moments, a *Pierce type estimate*. In particular, we show that quantization by empirical measures is of optimal order under weak assumptions.

Résumé. Dans cet article, nous étudions l'approximation d'une mesure de probabilité μ sur $\mathbb{R}^d$ par sa mesure empirique $\hat{\mu}_N$, interprétée comme quantification aléatoire. Comme critère d'erreur, nous considérons une moyenne de métrique de Wasserstein d'ordre p. Dans le cas $2p < d$, nous établissons des bornes supérieures et inférieures améliorées pour l'erreur, une *formule haute résolution*. De plus, nous donnons une estimation universelle à base de moments, nommée *estimation du type Pierce*. En particulier, nous prouvons que, sous de faibles hypothèses, la quantification par des mesures empiriques est d'ordre optimal.

MSC: Primary 60F25; secondary 65D32

Keywords: Constructive quantization; Wasserstein metric; Transportation problem; Zador's theorem; Pierce's lemma; Random quantization

1. Introduction

Constructive quantization is concerned with the efficient computation of discrete approximations to probability distributions. The need for such approximations mainly stems from two applications: firstly from information theory, where the approximation is a discretized version of an original signal which is to be stored on a data storage medium or transmitted via a channel (see e.g. [5,10,30]); secondly, from numerical integration, where integrals with respect to the original measure are replaced by the integral with respect to the discrete approximation (see e.g. [19]).

In both applications the objective is to find an optimal discrete subset of a metric space (E, d) of cardinality N say, a so-called *codebook*, depending on the given probability measure μ on E. In the first application one further needs fast coding and decoding schemes that find for a signal a digital representation of a close element of the codebook or, resp., translate the digital representation back. Clearly, the best coding scheme would map a signal to a digital representation of a closest neighbor in the codebook. The *quantization number* measures the smallest possible averaged distance of a μ-distributed point to the codebook and hence the performance of the best possible approximate coding of μ using N approximating points which corresponds to using $\log_2 N$ bits.

During the last decade, quantization attracted much interest mainly due to the second application, see for instance [22] for a recent review on financial applications. Here one aims at finding a codebook together with probability weights and the objective is to determine these in such a way that the distance between μ and the discrete probability measure is minimal with respect to some metric (e.g. a Wasserstein metric). Typically, the optimal solution of both problems are closely related. The optimal codebook of the first problem is also optimal for the second one and the optimal probability weights are the μ-weights of the corresponding Voronoi cells. In particular, the optimal approximation errors are again the *quantization numbers*. A regularly updated list of articles dealing with quantization can be found at http://www.quantize.maths-fi.com/.

From a constructive point of view, the two applications differ significantly and our research is mainly motivated by the second application. For moderate codebook sizes and particular probability measures it is feasible to run optimization algorithms and find approximations that are arbitrarily close to the optimum (see e.g. [18,20]). See also [17] for a recent constructive approach toward discrete approximation of marginals of stochastic differential equations. For large codebook sizes and probability measures that are defined implicitly, it is often not feasible to find close to optimal quantizations in reasonable time. Large codebooks can be used for approximate sampling of the distribution μ: if sampling of μ is costly (since it may be given only implicitly), one might prefer to sample from its quantization instead. This approach is analyzed in work in progress [25], where μ is the distribution of multiple Itô integrals.

As an alternative approach we analyze the use of the *empirical measure* $\hat{\mu}_N$ generated by N independent random variables distributed according to the original measure μ. As error criterion we consider an averaged L^p-Wasserstein metric. We stress that in our case the codebook is generated by i.i.d. samples and that the weights all have *equal* mass so that once the codebook is generated no further processing is needed. The advantage of using the empirical measure as a discrete approximation of μ is that it is usually easy to generate efficiently even for large N. The disadvantage is, of course, that for given N, the averaged Wasserstein distance between μ and $\hat{\mu}_N$ is larger than that between μ and the optimal probability measure supported on N points.

The estimation of the approximation error of $\hat{\mu}_N$ in the Wasserstein metric is the concern of various articles. Asymptotic results are derived for the uniform distribution in Ajtai et al. [1] for $d = 2$ and in Talagrand [27] for $d \geq 3$. In particular, these results indicate that the approach is not order optimal in dimensions one and two when compared with optimal quantization and we will restrict attention to dimensions greater or equal to three in this article. An upper bound for more general distributions can be found in [12]. Closely related problems are the bipartite matching problem [9] and the traveling salesman problem [3]. Interestingly, there has been progress on this class of problems in several aspects [2,4] parallel to our research.

We will show that in the case $E = \mathbb{R}^d$ equipped with some norm (which is the only case we consider in this article), the loss of performance is essentially a constant times an explicit term depending on the absolutely continuous part of μ.

A full treatment of quantization typically includes the derivation of asymptotic formulas in terms of the density of the absolutely continuous part of μ, a *high resolution formula*. Such a formula has been established for optimal quantization under norm-based distortions [6], for general Orlicz-norm distortions [8], and, very recently, also in the dual quantization problem [21]. In this article, we prove a high resolution formula for the empirical measure under an averaged L^p-Wasserstein metric. Further, a Pierce type result is derived. In particular, we obtain order optimality of the new approach under weak assumptions.

The article is organized as follows. Section 1 introduces the basic notation and summarizes the main results. Section 2 is devoted to the Pierce type result, see Theorem 1 below. Section 3 treats the particular case where μ is the uniform distribution on $[0, 1)^d$. It includes a proof of part (i) of Theorem 2 below. Finally, the high resolution formula provided by Theorem 2 is proved in Section 4.

1.1. *Notation*

We introduce the relevant notation along an example. Consider the following problem arising from logistics. There is a demand for a certain economic good on $\mathbb{R}^d$ modeled by a finite measure μ. Typically one would expect to have $d = 2$ in this example. The demand shall be accommodated by N service centers that are placed at positions $x_1, \dots, x_N \in \mathbb{R}^d$ and that have nonnegative capacities $p_1, \dots, p_N$ summing up to $\|\mu\| := \mu(\mathbb{R}^d)$. We associate a given choice of *supporting points* $x_1, \dots, x_N$ and *weights* $p_1, \dots, p_N$ with a measure $\hat{\mu} = \sum_{i=1}^N p_i \delta_{x_i}$, where δ_x denotes the Dirac measure in x. In order to cover the demand, goods have to be transported from the centers to the customers and we describe a transport schedule by a measure ξ on $\mathbb{R}^d \times \mathbb{R}^d$ such that its first, respectively second, marginal measure is equal to μ, respectively $\hat{\mu}$. The set of admissible transport schedules (*transports*) is denoted by

$\mathcal{M}(\mu, \hat{\mu})$ and supposing that transporting a unit mass from y to x causes cost $c(x, y)$, a transport $\xi \in \mathcal{M}(\mu, \hat{\mu})$ causes overall cost

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \,\mathrm{d}\xi(x, y).$$

In this article, we focus on norm based cost functions. In general, we assume that the demand is a finite measure on $\mathbb{R}^d$ and that the cost is of the form

$$c(x, y) = \|x - y\|^p,$$

where $p \geq 1$ and $\|\cdot\|$ is a fixed norm on $\mathbb{R}^d$. Given μ and $\hat{\mu}$, the minimal cost is the pth Wasserstein metric. In contrast to the above example, we will restrict attention to the case $d \geq 3$.

Definition 1 (pth Wasserstein metric). *The pth* Wasserstein metric *of two finite measures μ and ν on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, *which have equal mass*, *is given by*

$$\rho_p(\mu, \nu) = \inf_{\xi \in \mathcal{M}(\mu, \nu)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p \xi(\mathrm{d}x, \mathrm{d}y) \right)^{1/p},$$

where $\mathcal{M}(\mu, \nu)$ is the set of all finite measures ξ on $\mathbb{R}^d \times \mathbb{R}^d$ having marginal distributions μ in the first component and ν in the second component.

The Wasserstein metric originates from the *Monge–Kantorovich mass transportation problem*, which was introduced by G. Monge in 1781 [16]. Important results about the Wasserstein metric were achieved within the scope of *transportation theory*, for instance by Kantorovich [14], Kantorovich and Rubinstein [15], Wasserstein [29], Rachev and Rüschendorf [23,24], Villani [28] and others.

Note that the Wasserstein metric is homogeneous in (μ, ν) so that one can restrict attention to probability measures. In this article, we analyze for a given probability measure μ on $\mathbb{R}^d$ the quality of the empirical measure as approximation. More explicitly, we denote by $\hat{\mu}_N$ the (random) empirical measure of N independent μ-distributed random variables $X_1, \ldots, X_N$, that is

$$\hat{\mu}_N = \frac{1}{N} \sum_{j=1}^{N} \delta_{X_j},$$

and, for fixed $p \geq 1$, we analyze the asymptotic behavior of the so-called *random quantization error*

$$V_{N,p}^{\mathrm{rand}}(\mu) := \mathbb{E}\big[\rho_p^p(\mu, \hat{\mu}_N)\big]^{1/p}$$

as $N \in \mathbb{N}$ tends to infinity.

This quantity should be compared with the optimal approximation in the L^p-Wasserstein metric supported by N points, that is

$$V_{N,p}^{\mathrm{opt}}(\mu) := \inf_{\nu} \rho_p(\mu, \nu), \tag{1}$$

where the infimum is taken over all probability measures ν on $\mathbb{R}^d$ that are supported on N points. The quantity $V_{N,p}^{\mathrm{opt}}(\mu)$ is local in the sense that for a given set $\mathcal{C} \subset \mathbb{R}^d$ of supporting points used in an approximation ν, the optimal choice for ν is $\mu \circ \pi_{\mathcal{C}}^{-1}$, where $\pi_{\mathcal{C}}$ denotes a projection from $\mathbb{R}^d$ to $\mathcal{C}$. Hence, the minimization of the latter quantity reduces to a minimization over all sets $\mathcal{C} \subset \mathbb{R}^d$ of at most N elements. The minimal error is the so-called Nth *quantization number*

$$V_{N,p}^{\mathrm{opt}}(\mu) = \inf_{\mathcal{C}} \left(\int \min_{y \in \mathcal{C}} \|x - y\|^p \mu(\mathrm{d}x) \right)^{1/p}.$$

For a measure μ on $\mathbb{R}^d$ we denote by $\mu = \mu_a + \mu_s$ its Lebesgue decomposition with μ_a denoting the absolutely continuous part with respect to Lebesgue measure λ^d and μ_s the singular part. Further, we denote the uniform distribution on $[0,1)^d$ by $\mathcal{U}$ and define

$$\underline{V}_{N,p}^{\mathrm{rand}} := \mathbb{E}\Big[\inf_{\substack{\mathcal{U}' \in \Lambda \\ \xi \in \mathcal{M}(\mathcal{U}', \hat{\mathcal{U}}_N)}} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|^p \xi(\mathrm{d}x, \mathrm{d}y)\Big]^{1/p},$$

where Λ denotes the set of all probability measures $\mathcal{U}'$ on $[0,1]^d$ which satisfy $\mathcal{U}'(A) \le \mathcal{U}(A)$ for each Borel set $A \subset (0,1)^d$. Note that the latter quantity allows to have leakage in the boundaries of the support of the uniform measure $\mathcal{U}$. Therefore, $\underline{V}_{N,p}^{\mathrm{rand}} \le V_{N,p}^{\mathrm{rand}}(\mathcal{U})$. It seems plausible that the ratio of $\underline{V}_{N,p}^{\mathrm{rand}}$ and $V_{N,p}^{\mathrm{rand}}(\mathcal{U})$ converges to one as $N \to \infty$. However, this has not been proved yet.

1.2. *Main results*

We will assume throughout the paper that $d \ge 3$. The approximation by empirical measures satisfies a so-called *Pierce type* estimate.

Theorem 1. *Let $p \in [1, \frac{d}{2})$ and $q > \frac{dp}{d-p}$. There exists a constant $\kappa_{p,q}^{\mathrm{Pierce}}$ such that for any probability measure μ on $\mathbb{R}^d$*

$$V_{N,p}^{\mathrm{rand}}(\mu) \le \kappa_{p,q}^{\mathrm{Pierce}} \Big[\int_{\mathbb{R}^d} \|x\|^q \,\mathrm{d}\mu(x)\Big]^{1/q} N^{-1/d} \tag{2}$$

for all $N \in \mathbb{N}$.

As we will see in the discussion below the assumptions on p and q cannot be relaxed. Theorem 1 improves the asymptotic estimates of [12] that focused on the case $p = 2$ and were of nonoptimal order. Interestingly, it is also possible to give estimates for $\mathbb{E}[\rho_p(\mu, \hat{\mu}_N)]$ for compactly supported measures μ in general metric spaces based on covering numbers [4].

Remark 1.

- *The constant in the statement of Theorem 1 is explicit, see Theorem 3. Its value depends on the chosen norm on $\mathbb{R}^d$.*
- *For $p > \frac{d}{2}$ and discrete measures μ, the random approach typically induces errors $V_{N,p}^{\mathrm{rand}}(\mu)$ that are not of order $\mathcal{O}(N^{-1/d})$: take, for instance, two different points $a, b \in \mathbb{R}^d$ and let $\mu = \frac{1}{2}\delta_a + \frac{1}{2}\delta_b$. Then $N\hat{\mu}_N(\{a\})$ is binomially distributed with parameters N and $\frac{1}{2}$. Consequently,*

$$V_{N,p}^{\mathrm{rand}}(\mu) = \mathbb{E}\big[\rho_p^p(\mu, \hat{\mu}_N)\big]^{1/p} = \|a-b\| \mathbb{E}\Big[\Big|\hat{\mu}_N\big(\{a\}\big) - \frac{1}{2}\Big|\Big]^{1/p}$$

  is of order $N^{-1/2p}$ and, hence, converges to zero strictly slower than $N^{-1/d}$.
- *In [1], the case where $d = 2$, $p = 1$ and $\mu = \mathcal{U}$ is treated. There it is found that the L^1-Wasserstein distance between two independent realizations of $\hat{\mathcal{U}}_N$ is typically of order $N^{-1/2}(\log N)^{1/2}$ which shows the necessity of the assumption $d \ge 3$ for Theorem 1 to hold.*
- *For the uniform distribution $\mathcal{U}$ on $[0,1)^d$, the results of Talagrand [27] imply that $V_{N,p}^{\mathrm{rand}}(\mathcal{U})$ is always of order $N^{-1/d}$ as long as $d \ge 3$.*

The following theorem is a *high resolution formula* for quantization by empirical measures.

Theorem 2. *Let* $p \in [1, \frac{d}{2})$.

(i) *Let* $\mathcal{U}$ *denote the uniform distribution on* $[0,1)^d$. *There exists a constant* $\kappa_p^{\mathrm{unif}} \in (0,\infty)$ *such that*

$$\lim_{N\to\infty} N^{1/d} V_{N,p}^{\mathrm{rand}}(\mathcal{U}) = \kappa_p^{\mathrm{unif}}.$$

Further, there exist a constant $\underline{\kappa}_p^{\mathrm{unif}} \in (0,\infty)$ *such that*

$$\lim_{N\to\infty} N^{1/d} \underline{V}_{N,p}^{\mathrm{rand}} = \underline{\kappa}_p^{\mathrm{unif}}.$$

(ii) *Let* μ *be a probability measure on* $\mathbb{R}^d$ *that has a finite* q*th moment for some* $q > \frac{dp}{d-p}$ *and suppose that* $\frac{\mathrm{d}\mu_a}{\mathrm{d}\lambda^d}$ *is Riemann integrable or* $p = 1$. *Then*

$$\limsup_{N\to\infty} N^{1/d} V_{N,p}^{\mathrm{rand}}(\mu) \le \kappa_p^{\mathrm{unif}} \left(\int_{\mathbb{R}^d} \left(\frac{\mathrm{d}\mu_a}{\mathrm{d}\lambda^d} \right)^{1-p/d} \mathrm{d}\lambda^d \right)^{1/p} \tag{3}$$

and

$$\liminf_{N\to\infty} N^{1/d} V_{N,p}^{\mathrm{rand}}(\mu) \ge \underline{\kappa}_p^{\mathrm{unif}} \left(\int_{\mathbb{R}^d} \left(\frac{\mathrm{d}\mu_a}{\mathrm{d}\lambda^d} \right)^{1-p/d} \mathrm{d}\lambda^d \right)^{1/p}. \tag{4}$$

Interestingly, very similar asymptotic formulas appear in the bipartite matching problem [9], the traveling salesman problem [3] (both for $p = 1$) and in rather general combinatorial problems [2].

Remark 2. *We conjecture that* $\underline{\kappa}_p^{\mathrm{unif}} = \kappa_p^{\mathrm{unif}}$ *in which case the inequality and* lim sup *in* (3) *are actually an equality and* lim. *Proving the equality* $\underline{\kappa}_p^{\mathrm{unif}} = \kappa_p^{\mathrm{unif}}$ *seems to be a general open problem in transport problems. Similar problems arise in* [13] *for optimal transports from Poisson point processes with Lebesgue intensity to Lebesgue measure. Furthermore, we conjecture that the high resolution formula is still valid without the assumption of Riemann integrability.*

Remark 3. *The Pierce type result is sharp with respect to the assumption on the moment* q. *We will provide an example in which inequality* (2) *is not true for* $q = \frac{dp}{d-p}$: *let* $d \ge 3$ *and* $p \in [1, \frac{d}{2})$ *be arbitrary. Choose* $\beta \in (1, \frac{d}{d-p}]$ *and consider the probability measure* μ *defined by*

$$\frac{\mathrm{d}\mu}{\mathrm{d}\lambda^d}(x) = \begin{cases} \frac{1}{Z}|x|^{-d^2/(d-p)} \big(\log|x|\big)^{-\beta}, & |x| \ge \mathrm{e}, \\ 0, & \textit{otherwise}, \end{cases}$$

where $|\cdot|$ *denotes the Euclidean norm on* $\mathbb{R}^d$ *and* Z *denotes the appropriate normalization. Using the equivalence of norms on* $\mathbb{R}^d$ *and switching to polar coordinates we obtain*

$$\int_{\mathbb{R}^d} \|x\|^q \,\mathrm{d}\mu(x) \le c_1 \int_{\mathrm{e}}^{\infty} r^{-1} \ln(r)^{-\beta} \,\mathrm{d}r$$

for an appropriate finite constant c_1. *The latter integral is finite since* $\beta > 1$. *Conversely, the integral in the high resolution formula is*

$$\int_{\mathbb{R}^d} \left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda^d} \right)^{1-p/d} \mathrm{d}\lambda^d = c_2 \int_{\mathrm{e}}^{\infty} r^{-1} \ln(r)^{-(1-p/d)\beta} \,\mathrm{d}r,$$

where c_2 *is an appropriate positive constant. By our choice of* β *this integral is infinite so that by the lower bound of the high resolution formula* (*Theorem* 2(ii))

$$\liminf_{N\to\infty} N^{1/d} V_{N,p}^{\mathrm{rand}}(\mu) = \infty$$

which contradicts the validity of the Pierce type result.

Let us compare our results with the classical high resolution formulas, see [11], Theorem 6.2. The asymptotics of $V_{N,p}^{\mathrm{opt}}$ defined in (1) satisfies

$$\lim_{N\to\infty} N^{1/d} V_{N,p}^{\mathrm{opt}}(\mu) = c_{p,d}\biggl(\int_{\mathbb{R}^d}\biggl(\frac{\mathrm{d}\mu_a}{\mathrm{d}\lambda^d}\biggr)^{d/(d+p)}\,\mathrm{d}\lambda^d\biggr)^{1/d+1/p}, \tag{5}$$

whenever μ has a finite moment of order q for some $q > p$. Here, the constant $c_{p,d}$ is the corresponding limit for the uniform distribution on the unit cube in $\mathbb{R}^d$. Its numerical value is known in a few special cases.

Note that the integral term on the right-hand side of (5) differs from the one in (3) and (4). This effect can be explained as follows: for a sequence of optimal codebooks $(\mathcal{C}(N))_{N\ge 1}$ of size N the empirical measures $\frac{1}{N}\sum_{x\in\mathcal{C}(N)}\delta_x$ tend to a measure that differs from μ. In fact optimal codebooks allocate more points in the tails of the distribution. Since our approach does not account for such a correction, it is natural to expect a loss of efficiency for heavy tailed distributions. For arbitrary codebooks whose empirical distributions tend to the measure μ, one has lower bounds which incorporate the same integral term as in our high resolution formula, see [7], Theorem 7.2.

Theorem 1 can be used to improve [11], Theorem 9.1(a): there the validity of an asymptotic formula for the random quantization error is shown to be equivalent to the uniform integrability of $(N^{p/d}\min_{1\le i\le N}\|X - X_i\|^p)_{N\ge 1}$ where $X, X_1, \dots$ are independent with law μ. Theorem 1 shows that uniform integrability holds provided that $1\le p < d/2$ and μ has a finite moment of order q for some $q > \frac{dp}{d-p}$. Hence a high resolution formula is also available for quantization with random codebooks and *optimally* chosen weights. It incorporates the same integral as in formula (3) and postprocessing the weights of a random codebook can in the limit improve the error by a constant factor, irrespective the distribution μ.

1.3. *Preliminaries*

For a finite signed measure μ on the Borel sets of $\mathbb{R}^d$, we write $\|\mu\| := |\mu|(\mathbb{R}^d)$ for its total variation norm (using the same symbol as for the norm on $\mathbb{R}^d$ should not cause any confusion). For finite (nonnegative) measures μ and ν we denote by $\mu\wedge\nu$ the largest measure that is dominated by μ and ν. Furthermore, we set $(\mu-\nu)_+ := \mu - \mu\wedge\nu$.

Next, we introduce concatenation of transports. A transport ξ, i.e. a finite measure ξ on $\mathbb{R}^d\times\mathbb{R}^d$, will be associated to a probability kernel K and a measure ν on $\mathbb{R}^d$ via

$$\xi(\mathrm{d}x, \mathrm{d}y) = \nu(\mathrm{d}x)K(x, \mathrm{d}y), \tag{6}$$

so ν is the first marginal of ξ. We call ξ the transport with source ν and kernel K. Let $\mathcal{K}$ denote the set of probability kernels from $(\mathbb{R}^d, \mathcal{B}^d)$ into itself and consider the semigroup $(\mathcal{K}, *)$, where the operation $*$ is defined via

$$K_1 * K_2(x, A) := \int K_1(x, \mathrm{d}z)K_2(z, A) \quad \bigl(x\in\mathbb{R}^d, A\in\mathcal{B}^d\bigr).$$

Now we can iterate transport schedules: Let $\nu_0, \dots, \nu_n$ be measures on $\mathbb{R}^d$ with identical total mass and let $\xi_k \in \mathcal{M}(\nu_{k-1}, \nu_k)$. Then the concatenation of the transports $\xi_1, \dots, \xi_n$ is formally the transport described by the source ν_0 and the probability kernel $K = K_1 * \cdots * K_n$, where $K_1, \dots, K_n$ are the kernels associated to $\xi_1, \dots, \xi_n$. Note that the relation (6) defines the kernel uniquely up to ν-nullsets so that the concatenation of transport schedules is a well-defined operation on the set of transports. In analogy to the operation $*$ on $\mathcal{K}$, we write $\xi_1 * \cdots * \xi_n$ for the concatenation of the transport schedules.

We summarize some well-known properties of the Wasserstein metric in a lemma.

Lemma 1. *Let $\xi, \mu, \mu_1, \dots$ and $\nu, \nu_1, \dots$ be finite measures on $\mathbb{R}^d$ such that $\|\xi\| = \|\mu\| = \|\nu\|$ and $\|\mu_k\| = \|\nu_k\|$ for all $k = 1, \dots$. Further let $p\ge 1$.*

(i) Convexity: *Suppose that $\mu = \sum_{k\in\mathbb{N}}\mu_k$ and $\nu = \sum_{k\in\mathbb{N}}\nu_k$. Then*

$$\rho_p^p(\mu, \nu) \le \sum_{k=1}^{\infty}\rho_p^p(\mu_k, \nu_k). \tag{7}$$

(ii) Triangle-inequality: *One has*

$$\rho_p(\mu,\nu)\le\rho_p(\mu,\xi)+\rho_p(\xi,\nu). \tag{8}$$

(iii) Translation and scaling: *Let $T:\mathbb{R}^d\to\mathbb{R}^d$ be a map, which consists of a translation and a scaling by the factor $a>0$. Then*

$$\rho_p\bigl(\mu\circ T^{-1},\nu\circ T^{-1}\bigr)=a\rho_p(\mu,\nu). \tag{9}$$

(iv) Homogeneity: *For $\kappa\ge 0$ one has*

$$\kappa\rho_p^p(\mu,\nu)=\rho_p^p(\kappa\mu,\kappa\nu).$$

Proof. To see (i), use the transportation plan $\xi=\sum_{k=1}^\infty\xi_k$ where ξ_k is an optimal transportation plan from μ_k to ν_k. For (ii), see [28], p. 94. (iii) and (iv) follow directly from the definition of the Wasserstein metric. □

2. Proof of the Pierce type result

In order to prove Theorem 1, we first derive an estimate for general distributions on the unit cube $[0,1)^d$.

Proposition 1. *Let $1\le p<\frac{d}{2}$. There exists a constant $\kappa_p^{\mathrm{cube}}\in(0,\infty)$ such that for any probability measure μ on $[0,1)^d$ and $N\in\mathbb{N}$*

$$V_{N,p}^{\mathrm{rand}}(\mu)\le\kappa_p^{\mathrm{cube}}N^{-1/d}.$$

Remark 4. *The constant κ_p^{cube} is explicit. Let $\mathfrak{d}=\sup_{x,y\in[0,1)^d}\|x-y\|$ denote the diameter of $[0,1)^d$. Then*

$$\kappa_p^{\mathrm{cube}}=\mathfrak{d}2^{(d-2)/(2p)}\biggl[\frac{1}{1-2^{p-d/2}}+\frac{1}{1-2^{-p}}\biggr]^{1/p}.$$

For the proof of Proposition 1 we use a nested sequence of partitions of $B=[0,1)^d$. Note that B can be partitioned into 2^d translates $B_1,\dots,B_{2^d}$ of $2^{-1}B$. We iterate this procedure and partition each set B_k into 2^d translates $B_{k,1},\dots,B_{k,2^d}$ of $2^{-2}B$. We continue this scheme obeying the rule that each set $B_{k_1,\dots,k_l}$ is partitioned into 2^d translates $B_{k_1,\dots,k_l,1},\dots,B_{k_1,\dots,k_l,2^d}$ of $2^{-(l+1)}B$. These translates of $2^{-l}B$ form a partition of B and we denote this collection of sets by $\mathcal{P}_l$, the lth level. We now endow the sets $\mathcal{P}:=\bigcup_{l=0}^\infty\mathcal{P}_l$ with a 2^d-ary tree structure. B denotes the root of the tree and the father of a set $C\in\mathcal{P}_l$ ($l\in\mathbb{N}$) is the unique set $F\in\mathcal{P}_{l-1}$ that contains C.

Lemma 2. *Let μ and ν be two probability measures supported on B such that for all $C\in\mathcal{P}$*

$$\nu(C)>0\quad\Rightarrow\quad\mu(C)>0.$$

Then

$$\rho_p^p(\mu,\nu)\le\frac{1}{2}\mathfrak{d}^p\sum_{l=0}^\infty 2^{-pl}\sum_{F\in\mathcal{P}_l}\sum_{C\text{ child of }F}\biggl|\nu(C)-\nu(F)\frac{\mu(C)}{\mu(F)}\biggr|$$

with the convention that $\frac{0}{0}=0$.

For the proof we use couplings defined via partitions. Let (A_k) be a (finite or countably infinite) Borel partition of the Borel set $A\subset\mathbb{R}^d$. For two finite measures μ_1,μ_2 on A with equal mass, we call the measure $\bar\mu_1$ on $\mathbb{R}^d$ defined by

$$\bar\mu_1|_{A_k}=\frac{\mu_2(A_k)}{\mu_1(A_k)}\mu_1\bigg|_{A_k},$$

the (A_k)-*approximation* of μ_1 to μ_2 provided that it is well defined (i.e. that $\mu_1(A_k)=0$ implies $\mu_2(A_k)=0$).

The (A_k)-approximation $\bar\mu_1$ is associated with a transport ξ from μ_1 to $\bar\mu_1$: for each k, one has

$$(\mu_1\wedge\bar\mu_1)|_{A_k}=\frac{\mu_1(A_k)\wedge\mu_2(A_k)}{\mu_1(A_k)}\mu_1\bigg|_{A_k}$$

and we define a transport $\xi\in\mathcal{M}(\mu_1,\bar\mu_1)$ via

$$\xi=(\mu_1\wedge\bar\mu_1)\circ\psi^{-1}+\frac{1}{\delta}(\mu_1-\bar\mu_1)_+\otimes(\bar\mu_1-\mu_1)_+,$$

where $\delta:=\frac{1}{2}\sum_k|\mu_1(A_k)-\mu_2(A_k)|$ and $\psi:\mathbb{R}^d\to\mathbb{R}^d\times\mathbb{R}^d, x\mapsto(x,x)$. Then

$$\xi\big(\big\{(x,y)\in\mathbb{R}^d\times\mathbb{R}^d\colon x\neq y\big\}\big)=\delta.$$

Proof of Lemma 2. For $l\in\mathbb{N}_0:=\{0,1,2,\dots\}$, we set

$$\mu_l=\sum_{A\in\mathcal{P}_l}\frac{\nu(A)}{\mu(A)}\mu\bigg|_A$$

which is the $\mathcal{P}_l$-approximation of μ to ν. By construction, one has for each set $F\in\mathcal{P}_l$ with $l\in\mathbb{N}_0$

$$\mu_l(F)=\mu_{l+1}(F).$$

Moreover, provided that $\mu_l(F)>0$, one has for each child C of F

$$\mu_{l+1}|_C=\frac{\nu(C)}{\mu(C)}\mu\bigg|_C=\frac{\mu(F)\nu(C)}{\mu(C)\nu(F)}\mu_l\bigg|_C$$

so that $\mu_{l+1}|_F$ is the $\{C\in\mathcal{P}_{l+1}\colon C\subset F\}$-approximation of $\mu_l|_F$ to $\nu|_F$. Hence, there exists a transport $\xi_F\in\mathcal{M}(\mu_l|_F,\mu_{l+1}|_F)$ with

$$\xi_F\big(\big\{(x,y)\colon x\neq y\big\}\big)=\frac{1}{2}\sum_{C\text{ child of }F}\left|\nu(C)-\nu(F)\frac{\mu(C)}{\mu(F)}\right|. \tag{10}$$

Since each family $\mathcal{P}_l$ is a partition of the root B, we have

$$\xi_{l+1}:=\sum_{F\in\mathcal{P}_l}\xi_F\in\mathcal{M}(\mu_l,\mu_{l+1}).$$

Next, note that $\rho_p(\mu_l,\nu)\le\mathfrak{d}2^{-l}$ so that μ_l converges in the pth Wasserstein metric to ν which implies that

$$\rho_p(\mu,\nu)\le\sup_{l\in\mathbb{N}}\rho_p(\mu,\mu_l). \tag{11}$$

The concatenation of the transports $(\xi_l)_{l\in\mathbb{N}}$ leads to new transports

$$\xi^l=\xi_1*\cdots*\xi_l\in\mathcal{M}(\mu,\mu_l).$$

Each of the transports ξ_k is associated to a kernel K_k and, by Ionescu–Tulcea (see, e.g., [26]), there exists a sequence $(Z_l)_{l\in\mathbb{N}_0}$ of $[0,1)^d$-valued random variables with

$$\mathbb{P}(Z_0\in A_0,\dots,Z_l\in A_l)=\int_{A_0}\int_{A_1}\cdots\int_{A_{l-1}}K_l(x_{l-1},A_l)\cdots K_1(x_0,\mathrm{d}x_1)\mu(\mathrm{d}x_0)$$

for every $l \in \mathbb{N}$ and Borel sets $A_0, \dots, A_l \subset \mathbb{R}^d$. Then the joint distribution of (Z_0, Z_l) is ξ^l. Let

$$L = \inf\{l \in \mathbb{N}_0\colon\ Z_{l+1} \neq Z_l\}$$

and note that, almost surely, all entries $(Z_l)_{l \in \mathbb{N}_0}$ lie in one (random) set $A \in \mathcal{P}_L$, if $\{L < \infty\}$ enters, and are identical on $\{L = \infty\}$. Hence, for any $k \in \mathbb{N}$

$$\begin{aligned}\mathbb{E}\bigl[\|Z_0 - Z_k\|^p\bigr] &\le \mathfrak{d}^p \mathbb{E}\bigl[2^{-pL}\bigr] \le \mathfrak{d}^p \sum_{l=0}^{\infty} 2^{-pl} \mathbb{P}(Z_{l+1} \neq Z_l)\\ &= \mathfrak{d}^p \sum_{l=0}^{\infty} 2^{-pl} \xi_{l+1}\bigl(\bigl\{(x, y)\colon\ x \neq y\bigr\}\bigr)\\ &= \frac{1}{2}\mathfrak{d}^p \sum_{l=0}^{\infty} 2^{-pl} \sum_{F \in \mathcal{P}_l} \sum_{C \text{ child of } F} \left|\nu(C) - \nu(F)\frac{\mu(C)}{\mu(F)}\right|,\end{aligned}$$

where we used (10) in the last step, so the assertion follows by (11). □

Proof of Proposition 1. It is straight-forward to verify that the above lemma can be applied to μ and $\nu = \hat{\mu}_n(\omega)$ for almost all $\omega \in \Omega$ and we get

$$\rho_p^p(\mu, \hat{\mu}_N) \le \frac{1}{2}\mathfrak{d}^p \sum_{l=0}^{\infty} 2^{-pl} \sum_{F \in \mathcal{P}_l} \sum_{C \text{ child of } F} \left|\hat{\mu}_N(C) - \hat{\mu}_N(F)\frac{\mu(C)}{\mu(F)}\right|. \tag{12}$$

Note that conditional on the event $\{N\hat{\mu}_N(F) = k\}$ $(k \in \mathbb{N})$ the random vector $(N\hat{\mu}_N(C))_{C \text{ child of } F}$ is multinomially distributed with parameters k and success probabilities $(\mu(C)/\mu(F))_{C \text{ child of } F}$. Letting $\zeta(t) := \sqrt{t} \wedge t$ for $t \ge 0$, we obtain

$$\begin{aligned}&\mathbb{E}\biggl[\sum_{C \text{ child of } F}\left|\hat{\mu}_N(C) - \hat{\mu}_N(F)\frac{\mu(C)}{\mu(F)}\right| \Big| N\hat{\mu}_N(F) = k\biggr]\\ &\quad = \frac{1}{N}\mathbb{E}\biggl[\sum_{C \text{ child of } F}\left|N\hat{\mu}_N(C) - k\frac{\mu(C)}{\mu(F)}\right| \Big| N\hat{\mu}_N(F) = k\biggr]\\ &\quad \overset{\text{(a)}}{\le} \frac{1}{N}\sum_{C \text{ child of } F} \operatorname{var}\bigl(N\hat{\mu}_N(C) | N\hat{\mu}_N(F) = k\bigr)^{1/2}\\ &\quad \overset{\text{(b)}}{\le} \frac{\sqrt{k}}{N}\sum_{C \text{ child of } F}\sqrt{\frac{\mu(C)}{\mu(F)}} \overset{\text{(c)}}{\le} 2^{d/2}\frac{\sqrt{k}}{N} = \frac{2^{d/2}}{N}\zeta(k).\end{aligned}$$

Here, we estimate the first against the second moment in (a). In (b) we use that the conditional distribution of $N\hat{\mu}_N(C)$ is the binomial distribution with parameters k and $\mu(C)/\mu(F)$ and in (c) we apply the Cauchy–Schwarz inequality. The function ζ is concave and Jensen's inequality implies that

$$\mathbb{E}\biggl[\sum_{C \text{ child of } F}\left|\hat{\mu}_N(C) - \hat{\mu}_N(F)\frac{\mu(C)}{\mu(F)}\right|\biggr] \le \frac{2^{d/2}}{N}\zeta\bigl(N\mu(F)\bigr).$$

Consequently, it follows from (12) and Jensen's inequality that

$$\mathbb{E}\bigl[\rho_p^p(\mu, \hat{\mu}_N)\bigr] \le \frac{1}{2}\mathfrak{d}^p \sum_{l=0}^{\infty} 2^{-pl} \sum_{F \in \mathcal{P}_l} \frac{2^{d/2}}{N}\zeta\bigl(N\mu(F)\bigr) \le \mathfrak{d}^p 2^{d/2-1} N^{-1} \sum_{l=0}^{\infty} 2^{(d-p)l} \zeta\bigl(2^{-dl}N\bigr).$$

Let $l^* := \lfloor \log_2 N^{1/d} \rfloor$. Then,

$$\begin{aligned}\mathbb{E}\big[\rho_p^p(\mu,\hat\mu_N)\big] &\le \mathfrak{d}^p 2^{d/2-1} N^{-1}\Bigg[\sum_{l=0}^{l^*} 2^{((1/2)d-p)l}\sqrt{N} + \sum_{l=l^*+1}^{\infty} 2^{-pl} N\Bigg] \\ &\le \mathfrak{d}^p 2^{d/2-1} N^{-1}\Bigg[\sum_{k=0}^{\infty} 2^{(d/2-p)(l^*-k)}\sqrt{N} + 2^{-p(l^*+1)}\sum_{j=0}^{\infty} 2^{-pj} N\Bigg] \\ &\le \mathfrak{d}^p 2^{d/2-1} N^{-p/d}\bigg[\frac{1}{1-2^{p-d/2}} + \frac{1}{1-2^{-p}}\bigg],\end{aligned}$$

so the assertion follows. $\square$

We are now in the position to prove Theorem 1. Since all norms on $\mathbb{R}^d$ are equivalent, it suffices to prove the result for the maximum norm $\|\cdot\|_{\max}$.

Theorem 3. *Let $p \in [1, \frac{d}{2})$ and $q > \frac{pd}{d-p}$. One has for any probability measure μ on $\mathbb{R}^d$ that*

$$V_{N,p}^{\mathrm{rand}}(\mu) \le \kappa_{p,q}^{\mathrm{Pierce}}\bigg[\int_{\mathbb{R}^d} \|x\|_{\max}^q \,\mathrm{d}\mu(x)\bigg]^{1/q} N^{-1/d}, \tag{13}$$

where

$$\kappa_{p,q}^{\mathrm{Pierce}} = \kappa_p^{\mathrm{cube}}\bigg[\frac{2^{p-1}2^{q/2}\mathfrak{d}^p}{1-2^{p-(1/2)q}} + \frac{2^{p+q(1-p/d)}(\kappa_p^{\mathrm{cube}})^p}{1-2^{-q(1-p/d)+p}}\bigg]^{1/p}.$$

Proof. By the scaling invariance of inequality (13), we can and will assume without loss of generality that $\int \|x\|_{\max}^q \,\mathrm{d}\mu(x) = 1$. We partition $\mathbb{R}^d$ into a sequence of sets $(B_n)_{n\in\mathbb{N}_0}$ defined as

$$B_0 := B := [-1,1)^d \quad \text{and} \quad B_n := \big(2^n B\big)\backslash\big(2^{n-1}B\big) \quad \text{for } n \in \mathbb{N}.$$

We denote by ν the random (B_n)-approximation of μ to $\hat\mu_N$, that is

$$\nu|_{B_n} = \frac{\hat\mu_N(B_n)}{\mu(B_n)}\mu\bigg|_{B_n} \quad \text{for } n \in \mathbb{N}_0.$$

Then $\xi = (\mu\wedge\nu)\circ\psi^{-1} + \delta^{-1}(\mu-\nu)^+ \otimes (\nu-\mu)^+$ with $\delta := |(\mu-\nu)^+| = |(\nu-\mu)^+|$ and $\psi : \mathbb{R}^d \to \mathbb{R}^d\times\mathbb{R}^d, x \mapsto (x,x)$ defines a transport in $\mathcal{M}(\mu,\nu)$. Using that $\|x-y\|^p \le 2^{p-1}(\|x\|^p + \|y\|^p)$ for $x, y \in \mathbb{R}^d$, we get

$$\begin{aligned}\int \|x-y\|^p \xi(\mathrm{d}x,\mathrm{d}y) &= \delta^{-1}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \|x-y\|^p (\mu-\nu)^+(\mathrm{d}x)(\nu-\mu)^+(\mathrm{d}y) \\ &\le 2^{p-1}\int_{\mathbb{R}^d} \|x\|^p(\mu-\nu)^+(\mathrm{d}x) + 2^{p-1}\int_{\mathbb{R}^d}\|y\|^p(\nu-\mu)^+(\mathrm{d}y) \\ &\le 2^{p-1}\sum_{n=0}^{\infty}\int_{B_n}\|x\|^p(\mu-\nu)^+(\mathrm{d}x) + 2^{p-1}\sum_{n=0}^{\infty}\int_{B_n}\|y\|^p(\nu-\mu)^+(\mathrm{d}y) \\ &\le 2^{p-1}\sum_{n=0}^{\infty}\mathfrak{d}^p 2^{np}\cdot|\mu-\nu|(B_n).\end{aligned}$$

Note that $N\hat\mu_N(B_n)$ is binomially distributed with parameters N and $\mu(B_n)$. By the Markov inequality it follows that

$$\mu(B_n) \le 2^{-q(n-1)}\int \|x\|_{\max}^q \,\mathrm{d}\mu(x) = 2^{-q(n-1)}. \tag{14}$$

The inequality remains true for $n=0$. One has $\mathbb{E}[\hat\mu_N(B_n)]=\mu(B_n)$ and estimating the first against the second moment yields

$$\begin{aligned}\mathbb{E}\big[\rho_p^p(\mu,\nu)\big] &\le \sum_{n=0}^\infty 2^{p-1}2^{np}\mathfrak{d}^p\mathbb{E}\big[\big|\mu(B_n)-\hat\mu_N(B_n)\big|\big]\\ &\le \sum_{n=0}^\infty 2^{p-1}2^{np}\mathfrak{d}^p\operatorname{var}\big(\hat\mu_N(B_n)\big)^{1/2} \le \sum_{n=0}^\infty 2^{p-1}2^{np}\mathfrak{d}^p N^{-1/2}\mu(B_n)^{1/2}\\ &\le 2^{p+q/2-1}\mathfrak{d}^p N^{-1/2}\sum_{n=0}^\infty 2^{n(p-(1/2)q)} = \frac{2^{p+q/2-1}}{1-2^{p-(1/2)q}}\mathfrak{d}^p N^{-1/2}.\end{aligned}\tag{15}$$

It remains to analyze $\mathbb{E}[\rho_p^p(\nu,\hat\mu_N)]$. Given that $\{N\hat\mu_N(B_n)=k\}$ the random measure $\frac{N}{k}\hat\mu_N|_{B_n}$ is the empirical measure of k independent $\frac{\mu|_{B_n}}{\mu(B_n)}$-distributed random variables. By Lemma 1(iv) and Proposition 1, one has for $n\in\mathbb{N}_0$

$$\begin{aligned}&\mathbb{E}\big[\rho_p^p(\nu|_{B_n},\hat\mu_N|_{B_n})\big]\\ &\quad=\sum_{k=1}^\infty \mathbb{P}\big(N\hat\mu_N(B_n)=k\big)\frac{k}{N}\mathbb{E}\left[\rho_p^p\left(\frac{\mu|_{B_n}}{\mu(B_n)},\frac{N}{k}\hat\mu_N\bigg|_{B_n}\right)\bigg|N\hat\mu_N(B_n)=k\right]\\ &\quad\le\sum_{k=1}^\infty \mathbb{P}\big(N\hat\mu_N(B_n)=k\big)2^{(n+1)p}\frac{k}{N}\big(\kappa_p^{\mathrm{cube}}\big)^p k^{-p/d}\\ &\quad=\big(\kappa_p^{\mathrm{cube}}\big)^p N^{-p/d}2^{(n+1)p}\mathbb{E}\big[\hat\mu_N(B_n)^{1-p/d}\big].\end{aligned}$$

By Lemma 1(i) and the fact that $\mathbb{E}[\hat\mu_N(B_n)]=\mu(B_n)$, we conclude with Jensen's inequality that

$$\mathbb{E}\big[\rho_p^p(\nu,\hat\mu_N)\big]\le\sum_{n=0}^\infty\mathbb{E}\big[\rho_p^p(\nu|_{B_n},\hat\mu_N|_{B_n})\big]\le\big(\kappa_p^{\mathrm{cube}}\big)^p N^{-p/d}\sum_{n=0}^\infty 2^{(n+1)p}\mu(B_n)^{1-p/d}.$$

We use again inequality (14) to derive

$$\begin{aligned}\mathbb{E}\big[\rho_p^p(\nu,\hat\mu_N)\big]&\le\big(\kappa_p^{\mathrm{cube}}\big)^p N^{-p/d}\sum_{n=0}^\infty 2^{(n+1)p-q(n-1)(1-p/d)}\\ &=\big(\kappa_p^{\mathrm{cube}}\big)^p\frac{2^{p+q(1-p/d)}}{1-2^{-q(1-p/d)+p}}N^{-p/d}.\end{aligned}$$

Note that $\frac{p}{d}\le\frac12$ and altogether, we finish the proof by applying the triangle inequality (property (ii) of Lemma 1) and inequality (15) to deduce that

$$\mathbb{E}\big[\rho_p^p(\mu,\hat\mu_N)\big]^{1/p}\le\underbrace{\left[\frac{2^{p-1}2^{q/2}\mathfrak{d}^p}{1-2^{p-(1/2)q}}+\frac{2^{p+q(1-p/d)}(\kappa_p^{\mathrm{cube}})^p}{1-2^{-q(1-p/d)+p}}\right]^{1/p}}_{=:\kappa_{p,q}^{\mathrm{Pierce}}}N^{-1/d}.$$

□

3. Asymptotic analysis of the uniform measure

Next, we investigate the asymptotics of the random quantization of the uniform distribution $\mathcal{U}$ on the unit cube $B=[0,1)^d$. The aim of this subsection is to prove the existence of the limits

$$\kappa_p^{\mathrm{unif}}:=\lim_{N\to\infty}N^{1/d}V_{N,p}^{\mathrm{rand}}(\mathcal{U}),\qquad \underline{\kappa}_p^{\mathrm{unif}}:=\lim_{N\to\infty}N^{1/d}\underline{V}_{N,p}^{\mathrm{rand}}$$

which is the first statement of Theorem 2.

Notation 1. *Let A and S denote two sets with $A \subset S$ and suppose that $v = (v_j)_{j=1,\dots,N}$ is an S-valued vector. We call the vector v_A consisting of all entries of v in A the A-*subvector *of v, that is*

$$v_A := (v_{\gamma(j)}),$$

where $(\gamma(j))$ is the enumeration of the entries of v in A in canonical order.

For a Borel set $A \subset \mathbb{R}^d$ with finite nonvanishing Lebesgue measure, we denote by $\mathcal{U}(A)$ the uniform distribution on A. The proof of the existence of the limit makes use of the following lemma.

Lemma 3. *Let $K \in \mathbb{N}$ and let $A, A_1, \dots, A_K \subset \mathbb{R}^d$ be Borel sets such that $\lambda^d(A) \in (0, \infty)$ and that the sets $A_1, \dots, A_K \subset \mathbb{R}^d$ are pairwise disjoint and cover A. Fix $N \in \mathbb{N}$ and suppose that $\xi_k := N \cdot \frac{\lambda^d(A_k \cap A)}{\lambda^d(A)} \in \mathbb{N}_0$ for $k = 1, \dots, K$.*

Assume that $X = (X_1, \dots, X_N)$ is a random vector consisting of independent $\mathcal{U}(A)$-distributed entries. Then one can couple X with a random vector $Y = (Y_1, \dots, Y_N)$ which has A_k-subvectors consisting of ξ_k independent $\mathcal{U}(A_k)$-distributed entries such that the individual subvectors are independent and such that

$$\mathbb{E}\left[\sum_{j=1}^{N} \mathbb{1}_{\{X_j \neq Y_j\}}\right] \le \frac{\sqrt{K}\sqrt{N}}{2}. \tag{16}$$

Proof. For $k = 1, \dots, K$, denote by $X^{(k)}$ the A_k-subvector of X. For each k with $\xi_k \le \operatorname{length}(X^{(k)})$, we keep the first ξ_k entries of X in A_k and erase the remaining ones. For any other k's, we fill up $\xi_k - \operatorname{length}(X^{(k)})$ of the empty places by independent $\mathcal{U}(A_k)$-distributed elements. Denoting the new vector by Y, we see that Y has A_k-subvectors of length ξ_k. Clearly, Y has independent subvectors that are uniformly distributed on the respective sets. The length of the A_k-subvector of X is binomially distributed with parameters N and $q_k := \frac{\lambda^d(A_k \cap A)}{\lambda^d(A)}$, so that, in particular, $\mathbb{E}[\operatorname{length}(X^{(k)})] = N q_k = \xi_k$. Bounding the first by the second moment we get

$$\begin{aligned}\mathbb{E}\left[\sum_{j=1}^{N} \mathbb{1}_{\{X_j \neq Y_j\}}\right] &= \frac{1}{2}\mathbb{E}\left[\sum_{k=1}^{K} \left|\operatorname{length}\bigl(X^{(k)}\bigr) - \xi_k\right|\right] \le \frac{1}{2}\sum_{i=1}^{K} \operatorname{var}\bigl(\operatorname{length}\bigl(X^{(k)}\bigr)\bigr)^{1/2} \\ &\le \frac{1}{2}\sqrt{N}\sum_{k=1}^{K} \sqrt{q_k} \le \frac{1}{2}\sqrt{K}\sqrt{N},\end{aligned}$$

where we used the Cauchy–Schwarz inequality in the last step. □

Proof of the first statement of (i) of Theorem 2. Let $M \in \mathbb{N}$ be arbitrary but fixed. Further, let $N \in \mathbb{N}$, $N > 2^d M$, and denote by $B_0 = [0, a)^d$, $a^d = \frac{M}{N}$, the cube with volume $\lambda^d(B_0) = \frac{M}{N}$. We divide $[0, 1)^d$ into two parts, the main one $B^{\mathrm{main}} := [0, \lfloor 1/a \rfloor a)^d$ and the remainder $B^{\mathrm{rem}} := [0, 1)^d \backslash B^{\mathrm{main}}$. Note that $\lambda^d(B^{\mathrm{rem}}) \to 0$ as $N \to \infty$. We represent B^{main} as the union of $n = \lfloor a^{-1} \rfloor^d$ pairwise disjoint translates $B_1, \dots, B_n$ of B_0:

$$B^{\mathrm{main}} = \bigcup_{k=1}^{n} B_k.$$

Let $X = (X_1, \dots, X_N)$ denote a vector of N independent $\mathcal{U}[0, 1)^d$-distributed entries. We shall now couple X with a random vector $Y = (Y_1, \dots, Y_N)$ in such a way that most of the entries of X and Y coincide and such that the B_k-subvectors are independent and consist of M independent $\mathcal{U}(B_k)$-distributed entries. To achieve this goal we successively apply Lemma 3 to construct random vectors $X^0, \dots, X^L$ and finally set $X^L = Y$. First we apply the coupling of Lemma 3 for X with the decomposition $[0, 1)^d = B^{\mathrm{main}} \dot\cup B^{\mathrm{rem}}$ and denote by X^0 the resulting vector.

In the next step a 2^d-ary tree $\mathcal{T}$ whose leaves are the boxes $B_1, \dots, B_n$ is used to define further couplings. We let L denote the smallest integer with $2^L B_0 \supset B^{\text{main}}$, i.e. $L = \lceil -\log_2 a \rceil$, and set

$$\mathcal{T}_l := \{\gamma + 2^{L-l} B_0 \colon \gamma \in (2^{L-l} a \mathbb{Z}^d) \cap B^{\text{main}}\}$$

for $l = 0, \dots, L$. Now $\mathcal{T}$ is defined as the rooted tree which has at level l the boxes (vertices) $\mathcal{T}_l$ and a box $A_{\text{child}} \in \mathcal{T}_l$ is the child of a box $A_{\text{parent}} \in \mathcal{T}_{l-1}$ if $A_{\text{child}} \subset A_{\text{parent}}$.

We associate the vector X^0 with the 0th level of the tree. Now we define consecutively $X^1, \dots, X^L$ via the following rule. Suppose that X^l has already been defined. For each $A \in \mathcal{T}_l$ we apply the above coupling independently to the A-subvector of X^l with the representation

$$A = \dot{\bigcup_{B \text{ child of } A}} B.$$

By induction, for each $A \in \mathcal{T}_l$, the A-subvector of X^l consists of $N\lambda^d(A) \in \mathbb{N}$ independent $\mathcal{U}(A)$-distributed random variables. In particular, this is valid for the last level $Y = X^L$.

We proceed with an error analysis. Fix $\omega \in \Omega$ and $j \in \{1, \dots, N\}$ and suppose that $X_j^0(\omega), \dots, X_j^L(\omega)$ is altered in the step $l \to l+1$ for the first time and that $X_j^0(\omega) \in B \in \mathcal{T}_l$. Then it follows that $X_j^L(\omega) \in B$ so that

$$\|X_j^0(\omega) - X_j^L(\omega)\| \le \text{diameter}(B) \le a\mathfrak{d}2^{L-l},$$

where $\mathfrak{d}$ is the diameter of $[0,1)^d$. Consequently,

$$\mathbb{E}\left[\sum_{j=1}^N \|X_j^0 - X_j^L\|^p\right] \le \mathbb{E}\left[\sum_{j=1}^N \sum_{l=0}^{L-1} \mathbb{1}_{\{X_j^l \ne X_j^{l+1}\}} (a\mathfrak{d}2^{L-l})^p\right].$$

By Lemma 3 and the Cauchy–Schwarz inequality, one has, for $l = 1, \dots, L$,

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=1}^N \mathbb{1}_{\{X_j^l \ne X_j^{l-1}\}}\right] &\le \frac{1}{2}\sqrt{2^d}\sqrt{N} \sum_{A \in \mathcal{T}_{l-1}} \sqrt{\lambda^d(A)} \\
&\le \frac{1}{2}\sqrt{2^d}\sqrt{N}\left(\sum_{A \in \mathcal{T}_{l-1}} \lambda^d(A)\right)^{1/2} 2^{d(l-1)/2} \le \frac{1}{2} 2^{dl/2}\sqrt{N},
\end{aligned}$$

since $\sum_{A \in \mathcal{T}_{l-1}} \lambda^d(A) = \lambda^d(B^{\text{main}}) \le 1$ and $\mathcal{T}_{l-1}$ has at most $2^{d(l-1)}$ elements. Together with the former estimate we get

$$\mathbb{E}\left[\sum_{j=1}^N \|X_j^0 - X_j^L\|^p\right] \le \frac{1}{2}(a\mathfrak{d})^p\sqrt{N}\sum_{l=1}^L 2^{(L-l)p + dl/2} \le \frac{1}{2}\frac{(a\mathfrak{d})^p}{1 - 2^{-d/2+p}} 2^{dL/2}\sqrt{N}.$$

Next, we use that $a = (\frac{M}{N})^{1/d}$ and $2^L \le \frac{2}{a}$ to conclude that

$$\mathbb{E}\left[\sum_{j=1}^N \|X_j^0 - X_j^L\|^p\right] \le \frac{2^{d/2-1}\mathfrak{d}^p}{1 - 2^{-d/2+p}} M^{p/d-1/2} N^{1-p/d}.$$

Hence, there exists a constant C that does not depend on N and M such that

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{N}\sum_{j=1}^N \|X_j - Y_j\|^p\right]^{1/p} &\le \mathbb{E}\left[\frac{1}{N}\sum_{j=1}^N \|X_j - X_j^0\|^p\right]^{1/p} + \mathbb{E}\left[\frac{1}{N}\sum_{j=1}^N \|X_j^0 - X_j^L\|^p\right]^{1/p} \\
&\le C\left[N^{-1/(2p)} + M^{-(1/(2p)-1/d)} N^{-1/d}\right]. \qquad (17)
\end{aligned}$$

By construction, Y has for each $k=1,\dots,n$, a B_k-subvector of M independent $\mathcal{U}(B_k)$-distributed random variables and we denote the corresponding empirical measure by $\hat\mu_M^{(k)}$. Moreover, its B^{rem}-subvector contains $N-nM$ independent $\mathcal{U}(B^{\mathrm{rem}})$-distributed entries and we denote its empirical measure by $\hat\mu^{\mathrm{rem}}_{N-nM}$. Letting $\hat\mu_N^Y$ denote the empirical measure of Y, we conclude with Lemma 1 and Proposition 1 that

$$\begin{aligned} N\mathbb{E}\big[\rho_p^p\big(\hat\mu_N^Y,\mathcal{U}\big)\big] &\le \sum_{k=1}^n M\mathbb{E}\big[\rho_p^p\big(\hat\mu_M^{(k)},\mathcal{U}(B_k)\big)\big]+(N-nM)\mathbb{E}\big[\rho_p^p\big(\hat\mu^{\mathrm{rem}}_{N-nM},\mathcal{U}\big(B^{\mathrm{rem}}\big)\big)\big] \\ &\le nMa^p\big(V^{\mathrm{rand}}_{M,p}(\mathcal{U})\big)^p+\big(\kappa_p^{\mathrm{cube}}\big)^p(N-nM)^{1-p/d}. \end{aligned} \tag{18}$$

Next, we let N tend to infinity and combine the above estimates. Note that $N^{1/d}a=M^{1/d}$ and $\frac{nM}{N}\to 1$ so that

$$\limsup_{N\to\infty} N^{1/d}\mathbb{E}\big[\rho_p^p\big(\hat\mu_N^Y,\mathcal{U}\big)\big]^{1/p}\le M^{1/d}V^{\mathrm{rand}}_{M,p}(\mathcal{U}).$$

Moreover, (17) implies that

$$\limsup_{N\to\infty} N^{1/d}\mathbb{E}\big[\rho_p^p\big(\hat\mu_N^X,\hat\mu_N^Y\big)\big]^{1/p}\le CM^{-(1/(2p)-1/d)}.$$

Now fix $\varepsilon\in(0,1]$ arbitrarily and let $M\ge\frac{1}{\varepsilon}$ such that

$$M^{1/d}V^{\mathrm{rand}}_{M,p}(\mathcal{U})\le \liminf_{N\to\infty} N^{1/d}V^{\mathrm{rand}}_{N,p}(\mathcal{U})+\varepsilon.$$

Then

$$\limsup_{N\to\infty} N^{1/d}V^{\mathrm{rand}}_{N,p}(\mathcal{U})\le M^{1/d}V^{\mathrm{rand}}_{M,p}(\mathcal{U})+CM^{-(1/(2p)-1/d)}\le \liminf_{N\to\infty} N^{1/d}V^{\mathrm{rand}}_{N,p}(\mathcal{U})+\varepsilon+C\varepsilon^{1/(2p)-1/d}$$

and letting $\varepsilon\downarrow 0$ finishes the proof. □

Proof of the second statement of (i) of Theorem 2. The proof of the second statement is very similar to the proof of the first statement. The crucial difference is that the arguments are now based on superadditivity compared to the subadditivity of the Wasserstein metric (in the sense of part (i) of Lemma 1) that was used in the proof of the first statement.

We now look at a nonsymmetric modified version of the Wasserstein distance that allows leakage at the boundaries. For two probability measures ν_1 and ν_2 on $[0,1]^d$, we define

$$\underline{\rho}_p(\nu_1,\nu_2):=\inf_{\nu_1'\in\Lambda(\nu_1)}\rho_p\big(\nu_1',\nu_2\big),$$

where $\Lambda(\nu_1)$ denotes all probability measures ζ on $[0,1]^d$ which satisfy $\zeta(A)\le\nu_1(A)$ for all Borel sets A in $(0,1)^d$.

We make use of thee same notation as in the proof of the first statement. First note that similar as in (18)

$$N\mathbb{E}\big[\underline{\rho}_p^p\big(\mathcal{U},\hat\mu_N^Y\big)\big]\ge nMa^p\big(\underline{V}^{\mathrm{rand}}_{M,p}\big)^p.$$

Since, in general,

$$\underline{\rho}_p\big(\mathcal{U},\hat\mu_N^Y\big)\le\underline{\rho}_p\big(\mathcal{U},\hat\mu_N^X\big)+\rho_p\big(\hat\mu_N^X,\hat\mu_N^Y\big),$$

we conclude that

$$\begin{aligned} \liminf_{N\to\infty} N^{1/d}\mathbb{E}\big[\underline{\rho}_p^p\big(\mathcal{U},\hat\mu_N^X\big)\big]^{1/p} &\ge \liminf_{N\to\infty} N^{1/d}\mathbb{E}\big[\underline{\rho}_p^p\big(\mathcal{U},\hat\mu_N^Y\big)\big]^{1/p}-\limsup_{N\to\infty} N^{1/d}\mathbb{E}\big[\rho_p^p\big(\hat\mu_N^X,\hat\mu_N^Y\big)\big]^{1/p} \\ &\ge \liminf_{N\to\infty} N^{1/d}(nM/N)^{1/p}a\underline{V}^{\mathrm{rand}}_{M,p}(\mathcal{U})-CM^{-(1/(2p)-1/d)} \\ &\ge M^{1/d}\underline{V}^{\mathrm{rand}}_{M,p}(\mathcal{U})-CM^{-(1/(2p)-1/d)}. \end{aligned}$$

The proof is finished as above. □

4. Proof of the high resolution formula

4.1. *Proof of the high resolution formula for general p*

Definition 2. *We call a finite measure μ on $\mathbb{R}^d$* approachable from below, *if there exists for every $\varepsilon > 0$ a finite number of pairwise disjoint cubes $B_1, \dots, B_n$ (which are parallel to the coordinate axes) and positive reals $\alpha_1, \dots, \alpha_n$ such that $\nu := \sum \alpha_k \mathcal{U}(B_k)$ satisfies*

$$\nu \le \mu \quad \textit{and} \quad \|\mu - \nu\| \le \varepsilon.$$

A finite measure μ on $\mathbb{R}^d$ is called approachable from above, *if there exists for every $\delta, \varepsilon > 0$ a finite number of pairwise disjoint cubes $B_1, \dots, B_n$ and positive reals $\alpha_1, \dots, \alpha_n$ such that $\nu := \sum \alpha_k \mathcal{U}(B_k)$ satisfies*

$$\nu \ge \mu|_{B(0,\delta)} \quad \textit{and} \quad \|\nu - \mu|_{B(0,\delta)}\| \le \varepsilon.$$

Remark 5.

1. *The uniform distribution on a rectangle with positive Lebesgue measure is approachable from above and below and it is straight-forward to verify that one can allow arbitrary (nondisjoint) rectangles $B_1, \dots, B_n$ in the definition of approachability without changing the definition.*
2. *By* 1. *we can express a measure which is approachable from below or above locally as the limit of a monotone sequence of measures with Lebesgue density. Hence it has itself a Lebesgue density. Conversely, every finite measure which has a density that is Riemann integrable on any ball $B(0, \delta)$ $(\delta > 0)$, is approachable from below and above.*

Proposition 2. *Let μ denote a compactly supported probability measure that is approachable from below. Further let $p \in [1, d/2)$. Then*

$$\limsup_{N\to\infty} N^{1/d} \mathbb{E}\big[\rho_p^p(\mu, \hat{\mu}_N)\big]^{1/p} \le \kappa_p^{\mathrm{unif}} \left(\int_{\mathbb{R}^d} \left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda^d} \right)^{1-p/d} \mathrm{d}\lambda^d \right)^{1/p}.$$

Proof. Let $\varepsilon > 0$ and choose a finite number of pairwise disjoint cubes $B_1, \dots, B_K$ and positive reals $\alpha_1, \dots, \alpha_K$ such that $\mu^* := \sum_{k=1}^K \alpha_k \mathcal{U}(B_k) \le \mu$ and $\|\mu - \mu^*\| \le \varepsilon$. For $k = 1, \dots, K$ let $\mu^{(k)} = \mathcal{U}(B_k)$, set $\alpha_0 = \|\mu - \mu^*\|$ and fix a probability measure $\mu^{(0)}$ such that

$$\mu = \sum_{k=0}^K \alpha_k \mu^{(k)}.$$

For each k, we consider empirical measures $(\hat{\mu}_n^{(k)})_{n\in\mathbb{N}}$ of a sequence of independent $\mu^{(k)}$-distributed random variables. We assume independence of the individual empirical measures and observe that for an additional independent multinomial random variable $M = (M_k)_{k=0,\dots,K}$ with parameters N and $(\alpha_k)_{k=0,\dots,K}$ one has

$$N \hat{\mu}_N \stackrel{\mathcal{L}}{=} \sum_{k=0}^K M_k \hat{\mu}_{M_k}^{(k)}.$$

We assume without loss of generality strict equality in the last equation. Set $\nu = \sum_{k=0}^K \frac{M_k}{N} \mu^{(k)}$ and observe that by the triangle inequality

$$\mathbb{E}\big[\rho_p^p(\mu, \hat{\mu}_N)\big]^{1/p} \le \mathbb{E}\big[\rho_p^p(\mu, \nu)\big]^{1/p} + \mathbb{E}\big[\rho_p^p(\nu, \hat{\mu}_N)\big]^{1/p}.$$

The first expression on the right hand side is of order $\mathcal{O}(N^{-1/2p})$ (see the proof of Proposition 3). By Theorem 2(i) and Lemma 6 of the Appendix, there is a concave function $\varphi: [0, \infty) \to \mathbb{R}$ such that $\mathbb{E}[n\rho_p^p(\mathcal{U}([0,1)^d), \widehat{\mathcal{U}([0,1)^d)}_n)] \le$

$\varphi(n)$ for all $n \in \mathbb{N}_0$ and

$$\lim_{n\to\infty} \frac{1}{n^{1-p/d}} \varphi(n) = \left(\kappa_p^{\mathrm{unif}}\right)^p.$$

Denote by $a_1, \ldots, a_K$ the edge lengths of the cubes $B_1, \ldots, B_K$ and let $a_0 > 0$ be such that the support of μ is contained in a cube with side length a_0. Then, by Lemma 1, Proposition 1 and Jensen's inequality,

$$\begin{aligned} N\mathbb{E}\left[\rho_p^p(\nu, \hat{\mu}_N)\right] &\le \sum_{k=0}^{K} \mathbb{E}\left[M_k \rho_p^p\left(\mu^{(k)}, \hat{\mu}_{M_k}^{(k)}\right)\right] \\ &\le \left(\kappa_p^{\mathrm{cube}}\right)^p a_0^p \mathbb{E}\left[M_0^{1-p/d}\right] + \sum_{k=1}^{K} a_k^p \mathbb{E}\left[\varphi(M_k)\right] \\ &\le \left(\kappa_p^{\mathrm{cube}}\right)^p a_0^p (\alpha_0 N)^{1-p/d} + \sum_{k=1}^{K} a_k^p \varphi(\alpha_k N), \end{aligned}$$

so that

$$\limsup_{N\to\infty} N^{p/d}\mathbb{E}\left[\rho_p^p(\nu, \hat{\mu}_N)\right] \le \left(\kappa_p^{\mathrm{cube}}\right)^p a_0^p \varepsilon^{1-p/d} + \left(\kappa_p^{\mathrm{unif}}\right)^p \sum_{k=1}^{K} a_k^p \alpha_k^{1-p/d}.$$

Note that for $x \in B_k$, $f(x) := \frac{\mathrm{d}\mu_a}{\mathrm{d}\lambda^d} \ge \alpha_k / a_k^d$ and we get

$$a_k^p \alpha_k^{1-p/d} = \int_{B_k} a_k^{p-d} \alpha_k^{1-p/d}\,\mathrm{d}x \le \int_{B_k} f(x)^{1-p/d}\,\mathrm{d}x.$$

Finally, we arrive at

$$\limsup_{N\to\infty} N^{p/d}\mathbb{E}\left[\rho_p^p(\mu, \hat{\mu}_N)\right] \le \left(\kappa_p^{\mathrm{unif}}\right)^p \int_{\mathbb{R}^d} f(x)^{1-p/d}\,\mathrm{d}x + \left(\kappa_p^{\mathrm{cube}}\right)^p a_0^p \varepsilon^{1-p/d}.$$

Letting $\varepsilon \to 0$ the assertion follows. □

Proposition 3. *Let μ be a finite singular measure on the Borel sets of* $[0,1)^d$. *For* $p \in [1, d/2)$, *one has*

$$\lim_{N\to\infty} N^{1/d} V_{N,p}^{\mathrm{rand}}(\mu) = 0.$$

Proof. Without loss of generality we will assume that μ is a probability measure. Let $\varepsilon > 0$ and choose an open set $U \subset \mathbb{R}^d$ such that $\mu(U) = 1$ and $\lambda^d(U) < \varepsilon$. We fix finitely many pairwise disjoint cubes $B_1, \ldots, B_K$ with

$$U \supset B_1 \cup \cdots \cup B_K \quad \text{and} \quad \mu(B_1 \cup \cdots \cup B_K) \ge 1 - \varepsilon.$$

We set $B_0 = [0,1)^d \backslash (B_1 \cup \cdots \cup B_K)$ and consider the (B_k)-approximation of μ to $\hat{\mu}_N$. According to the discussion following Lemma 2, we consider the random probability measure ν on $[0,1)^d$ with

$$\nu|_{B_k} = \frac{\hat{\mu}_N(B_k)}{\mu(B_k)} \mu\bigg|_{B_k}.$$

Then the vector $Z := (N\hat{\mu}_N(B_k))_{k=0,\ldots,K}$ is multinomially distributed with parameters N and $(\mu(B_k))_{k=0,\ldots,K}$ and the coupling introduced below Lemma 2 achieves

$$\rho_p^p(\mu, \nu) \le \frac{1}{2}\mathfrak{d}^p \sum_{k=0}^{K} \left|\frac{Z_k}{N} - \mu(B_k)\right|.$$

Consequently,

$$\mathbb{E}\big[\rho_p^p(\mu,\nu)\big]^{1/p} \le \left(\frac{1}{2N}\mathfrak{d}^p \sum_{k=0}^{K} \mathbb{E}\big|Z_k - N\mu(B_k)\big|\right)^{1/p} = \mathcal{O}\big(N^{-1/2p}\big). \tag{19}$$

We denote by $a_1,\dots,a_K$ the edge lengths of the cubes B_k, i.e. $a_k = \lambda^d(B_k)^{1/d}$, and set $a_0 = 1$. Note that $\nu|_{B_k}$ and $\hat{\mu}_N|_{B_k}$ have the same mass for all k. We apply Lemma 1, Proposition 1 and Jensen's inequality to deduce that

$$\begin{aligned}\mathbb{E}\big[\rho_p^p(\nu,\hat{\mu}_N)\big] &\le \sum_{k=0}^{K}\mathbb{E}\big[\rho_p^p(\nu|_{B_k},\hat{\mu}_N|_{B_k})\big] \le \frac{1}{N}\big(\kappa_p^{\mathrm{cube}}\big)^p \sum_{k=0}^{K} a_k^p \mathbb{E}\big[\big(\hat{\mu}_N(B_k)N\big)^{1-p/d}\big] \\ &\le \big(\kappa_p^{\mathrm{cube}}\big)^p N^{-p/d}\sum_{k=0}^{K} a_k^p\big(\mu(B_k)\big)^{1-p/d}.\end{aligned}$$

Next, we apply Hölder's inequality with exponents d/p and $(1-p/d)^{-1}$ to get

$$\begin{aligned}\mathbb{E}\big[\rho_p^p(\nu,\hat{\mu}_N)\big] &\le \big(\kappa_p^{\mathrm{cube}}\big)^p \left(\sum_{k=1}^{K}\lambda^d(B_k)\right)^{p/d}\cdot\left(\sum_{k=1}^{K}\mu(B_k)\right)^{1-p/d} N^{-p/d} + \big(\kappa_p^{\mathrm{cube}}\big)^p \mu(B_0)^{1-p/d}N^{-p/d} \\ &\le \big(\kappa_p^{\mathrm{cube}}\big)^p\big(\varepsilon^{p/d}+\varepsilon^{1-p/d}\big)N^{-p/d}.\end{aligned}$$

It follows from (19) and the triangle inequality that

$$\limsup_{N\to\infty} N^{1/d}\mathbb{E}\big[\rho_p^p(\mu,\hat{\mu}_N)\big]^{1/p} \le \kappa_p^{\mathrm{cube}}\big(\varepsilon^{p/d}+\varepsilon^{1-p/d}\big)^{1/p}$$

which finishes the proof since $\varepsilon > 0$ is arbitrary. □

Theorem 4. *Let $p \in [1, \frac{d}{2})$ and let μ denote a probability measure on $\mathbb{R}^d$ with finite qth moment for some $q > \frac{dp}{d-p}$. If the absolutely continuous part μ_a of μ is approachable from below with density f, then*

$$\limsup_{N\to\infty} N^{1/d} V_{N,p}^{\mathrm{rand}}(\mu) \le \kappa_p^{\mathrm{unif}}\left(\int_{\mathbb{R}^d} f(x)^{1-p/d}\,\mathrm{d}x\right)^{1/p}. \tag{20}$$

If the absolutely continuous part μ_a of μ is approachable from above with density f, then

$$\liminf_{N\to\infty} N^{1/d} V_{N,p}^{\mathrm{rand}}(\mu) \ge \underline{\kappa}_p^{\mathrm{unif}}\left(\int_{\mathbb{R}^d} f(x)^{1-p/d}\,\mathrm{d}x\right)^{1/p}. \tag{21}$$

Proof. We only prove the first statement since the second one is proved analogously (first establishing a corresponding version of Proposition 2). Let $\delta > 0$ and set

$$\mu^{(1)} = \frac{\mu_a|_{B(0,\delta)}}{\mu_a(B(0,\delta))}, \qquad \mu^{(2)} = \frac{\mu_s|_{B(0,\delta)}}{\mu_s(B(0,\delta))}, \quad \text{and} \quad \mu^{(3)} = \frac{\mu|_{B(0,\delta)^c}}{\mu(B(0,\delta)^c)},$$

where we let $\mu^{(i)}$ be an arbitrary probability measure in case the denominator is zero. As in the proof of Proposition 2, we represent $\hat{\mu}_N$ with the help of independent sequences of empirical measures $(\hat{\mu}_n^{(1)})_{n\in\mathbb{N}_0},\dots,(\hat{\mu}_n^{(3)})_{n\in\mathbb{N}_0}$ and an independent multinomially distributed random variable $M = (M_k)_{k=1,2,3}$ with parameters N and $(\mu_a(B(0,\delta)), \mu_s(B(0,\delta)), \mu(B(0,\delta)^c))$ as

$$N\hat{\mu}_N = \sum_{k=1}^{3} M_k \hat{\mu}_{M_k}^{(k)}.$$

As before one observes that for the random measure $\nu = \sum_{k=1}^{3} \frac{M_k}{N} \mu^{(k)}$

$$\mathbb{E}\big[\rho_p^p(\mu, \nu)\big]^{1/p} = \mathcal{O}\big(N^{-1/2}\big).$$

Further, by Lemma 1,

$$N\mathbb{E}\big[\rho_p^p(\nu, \hat{\mu}_N)\big] \leq \sum_{k=1}^{3} \mathbb{E}\big[M_k \rho_p^p\big(\mu^{(k)}, \hat{\mu}_{M_k}^{(k)}\big)\big].$$

By Propositions 2 and 3 and Lemma 6 of the Appendix, there exist concave functions φ_1 and φ_2 with

$$nV_{n,p}^{\mathrm{rand}}\big(\mu^{(k)}\big)^p \leq \varphi_k(n) \quad \text{for } n \in \mathbb{N}, k = 1, 2$$

and

$$\varphi_1(n) \sim \big(\kappa_p^{\mathrm{unif}}\big)^p n^{1-p/d} \int_{B(0,\delta)} \frac{f(x)^{1-p/d}}{\mu_a(B(0,\delta))^{1-p/d}} \,\mathrm{d}x \quad \text{and} \quad \varphi_2(n) = \mathrm{o}\big(n^{1-p/d}\big)$$

as $n \to \infty$. By Jensen's inequality, $\mathbb{E}[M_k \rho_p^p(\mu^{(k)}, \hat{\mu}_{M_k}^{(k)})] \leq \varphi_k(\mathbb{E}[M_k])$ so that

$$\limsup_{N\to\infty} \frac{1}{N^{1-p/d}} \mathbb{E}\big[M_1 \rho_p^p\big(\mu^{(1)}, \hat{\mu}_{M_1}^{(1)}\big)\big] \leq \big(\kappa_p^{\mathrm{unif}}\big)^p \int_{B(0,\delta)} f(x)^{1-p/d} \,\mathrm{d}x.$$

Analogously, using Proposition 3,

$$\limsup_{N\to\infty} \frac{1}{N^{1-p/d}} \mathbb{E}\big[M_2 \rho_p^p\big(\mu^{(2)}, \hat{\mu}_{M_2}^{(2)}\big)\big] = 0$$

and, by Theorem 3,

$$\limsup_{N\to\infty} \frac{1}{N^{1-p/d}} \mathbb{E}\big[M_3 \rho_p^p\big(\mu^{(3)}, \hat{\mu}_{M_3}^{(3)}\big)\big] \leq \big(\kappa_{p,q}^{\mathrm{Pierce}}\big)^p \bigg[\int_{B(0,\delta)^c} \|x\|_{\max}^q \,\mathrm{d}\mu(x)\bigg]^{p/q},$$

where we used that $1 - \frac{p}{d} - \frac{p}{q} \geq 0$. Altogether, we get

$$\begin{aligned}&\limsup_{N\to\infty} N^{p/d} \mathbb{E}\big[\rho_p^p(\mu, \hat{\mu}_N)\big] \\ &\quad \leq \big(\kappa_p^{\mathrm{unif}}\big)^p \int_{B(0,\delta)} f(x)^{1-p/d} \,\mathrm{d}x + \big(\kappa_{p,q}^{\mathrm{Pierce}}\big)^p \bigg[\int_{B(0,\delta)^c} \|x\|_{\max}^q \,\mathrm{d}\mu(x)\bigg]^{p/q}\end{aligned}$$

and letting $\delta \to \infty$ finishes the proof. □

4.2. *Proof of the high resolution formula for $p = 1$*

In this section, we consider the special case $p = 1$. We will write ρ instead of ρ_1. The case $p = 1$ is special because of the following lemma.

Lemma 4. *Let μ, ν, κ be finite measures on $\mathbb{R}^d$ such that $\|\mu\| = \|\nu\|$. Then one has*

$$\rho(\mu + \kappa, \nu + \kappa) = \rho(\mu, \nu).$$

Proof. One has

$$\begin{aligned}\rho(\mu+\kappa,\nu+\kappa) &= \sup\left\{\int f\,\mathrm{d}(\mu+\kappa)-\int f\,\mathrm{d}(\nu+\kappa)\colon\ f\ \text{1-Lipschitz}\right\}\\ &= \sup\left\{\int f\,\mathrm{d}\mu-\int f\,\mathrm{d}\nu\colon\ f\ \text{1-Lipschitz}\right\}=\rho(\mu,\nu),\end{aligned}$$

where the first equality can be found in [28], p. 95. □

The following lemma shows that the map $\mu\mapsto\limsup_{N\to\infty}(N^{1/d}V_{N,1}^{\mathrm{rand}}(\mu))$ and likewise $\mu\mapsto\liminf_{N\to\infty}(N^{1/d}\times V_{N,1}^{\mathrm{rand}}(\mu))$ are continuous with respect to the total variation norm.

Lemma 5. *Let $d\geq 3$ and $q>\frac{d}{d-1}$. For probability measures μ and ν on $\mathbb{R}^d$ one has*

$$\limsup_{N\to\infty} N^{1/d}\bigl|V_{N,1}^{\mathrm{rand}}(\mu)-V_{N,1}^{\mathrm{rand}}(\nu)\bigr|\leq 2\kappa_{1,q}^{\mathrm{Pierce}}\|\mu-\nu\|^{1-1/d-1/q}\left(\int\|x\|_{\max}^q|\mu-\nu|(\mathrm{d}x)\right)^{1/q}.$$

Proof. Without loss of generality, we assume that $\mu\neq\nu$. Let $\alpha=\frac{\mu\wedge\nu}{\|\mu\wedge\nu\|}$, $\mu^*=\frac{\mu-\mu\wedge\nu}{\|\mu-\mu\wedge\nu\|}$ and $\nu^*=\frac{\nu-\mu\wedge\nu}{\|\nu-\mu\wedge\nu\|}$ (let α be an arbitrary probability measure in case $\mu\wedge\nu=0$). For fixed $N\in\mathbb{N}$ let (M_1,M_2) be multinomially distributed with parameters N and $(\|\mu\wedge\nu\|,1-\|\mu\wedge\nu\|)$. We represent $\hat\mu_N$ and $\hat\nu_N$ as combinations of independent empirical measures $(\hat\alpha_n)$, $(\hat\mu_n^*)$ and $(\hat\nu_n^*)$ as

$$N\hat\mu_N=M_1\hat\alpha_{M_1}+M_2\hat\mu_{M_2}^*\quad\text{and}\quad N\hat\nu_N=M_1\hat\alpha_{M_1}+M_2\hat\nu_{M_2}^*.$$

By Lemma 1(ii) and (i), one has

$$\begin{aligned}\rho(N\mu,N\hat\mu_N)&\leq\rho\bigl(N\mu,M_1\alpha+M_2\mu^*\bigr)+\rho\bigl(M_1\alpha+M_2\mu^*,M_1\hat\alpha_{M_1}+M_2\hat\mu_{M_2}^*\bigr)\\ &\leq\rho\bigl(N\mu,M_1\alpha+M_2\mu^*\bigr)+\rho(M_1\alpha,M_1\hat\alpha_{M_1})+\rho\bigl(M_2\mu^*,M_2\hat\mu_{M_2}^*\bigr).\end{aligned}\tag{22}$$

Observe that

$$\mathbb{E}\bigl[\rho\bigl(N\mu,M_1\alpha+M_2\mu^*\bigr)\bigr]\leq\mathbb{E}\bigl[\bigl|M_1-\mathbb{E}[M_1]\bigr|\bigr]\rho\bigl(\mu^*,\alpha\bigr)=\mathcal{O}\bigl(N^{1/2}\bigr).\tag{23}$$

Further, by Theorem 3 and Jensen's inequality, one has

$$\mathbb{E}\bigl[\rho\bigl(M_2\mu^*,M_2\hat\mu_{M_2}^*\bigr)\bigr]\leq\kappa_{1,q}^{\mathrm{Pierce}}\|\mu-\nu\|^{1-1/d-1/q}N^{1-1/d}\left(\int\|x\|_{\max}^q(\mu-\nu)_+(\mathrm{d}x)\right)^{1/q}+\mathcal{O}\bigl(N^{1/2}\bigr),\tag{24}$$

where we used that $(\mu-\nu)_+=\|\mu-\nu\|\mu^*$. Conversely, by Lemma 4 and Lemma 1,

$$\begin{aligned}\rho(M_1\alpha,M_1\hat\alpha_{M_1})&=\rho\bigl(M_1\alpha+M_2\hat\nu_{M_2}^*,M_1\hat\alpha_{M_1}+M_2\hat\nu_{M_2}^*\bigr)\\ &=\rho\bigl(M_1\alpha+M_2\hat\nu_{M_2}^*,N\hat\nu_N\bigr)\\ &\leq\rho(N\nu,N\hat\nu_N)+\rho\bigl(M_1\alpha+M_2\hat\nu_{M_2}^*,N\nu\bigr)\\ &=\rho(N\nu,N\hat\nu_N)+\rho\bigl(M_1\alpha+M_2\hat\nu_{M_2}^*+M_2\nu^*,N\nu+M_2\nu^*\bigr)\\ &\leq\rho(N\nu,N\hat\nu_N)+\rho\bigl(M_2\hat\nu_{M_2}^*,M_2\nu^*\bigr)+\rho\bigl(M_1\alpha+M_2\nu^*,N\nu\bigr).\end{aligned}$$

The expected values of the last two summands can be estimated like (24) and (23). Inserting the estimates into (22), the assertion of the lemma follows. □

We now prove the general upper and lower bounds in the case $p=1$.

Proof of Theorem 2(ii) for $p=1$. Let $\mu = \mu_a + \mu_s$ be the Lebesgue decomposition of μ and let f denote the density of μ_a. It is now straightforward to verify that $\mu^{(n)}$ with density

$$f^{(n)}(x) = 2^{-nd} \int_{S_{n,m_1,\dots,m_d}} f(y)\,\mathrm{d}y \quad \text{for } x \in S_{n,m_1,\dots,m_d},$$

where $S_{n,m_1,\dots,m_d} := 2^{-n}([m_1, m_1+1) \times \cdots \times [m_d, m_d+1))$, satisfies $\|\mu_a - \mu^{(n)}\| \to 0$ and $\int \|x\|_{\max}^q |\mu_a - \mu^{(n)}|(\mathrm{d}x) \to 0$. Since $\mu^{(n)} + \mu_s$ is approachable from below and above, Lemma 5 allows to extend the upper and lower bounds of Theorem 4 to the case with general density if $p=1$. □

Appendix

Lemma 6. *Suppose that $f, g : \mathbb{N}_0 \to [0,\infty)$ are functions with the following properties*:

- *g is nondecreasing, concave and* $\lim_{n\to\infty} g(n) = \infty$.
- $\alpha := \limsup_{n\to\infty} \frac{f(n)}{g(n)} \in [0,\infty)$.

Then there exists a concave function $\varphi : \mathbb{N}_0 \to [0,\infty)$ dominating f with

$$\lim_{n\to\infty} \frac{\varphi(n)}{g(n)} = \alpha.$$

Proof. For $\varepsilon > 0$ choose n_0 such that $f(n) \le (\alpha+\varepsilon)g(n)$ for all $n \ge n_0$. Then there exists some $C_\varepsilon \ge 0$ such that $f(n) \le (\alpha+\varepsilon)g(n) + C_\varepsilon =: \varphi_\varepsilon(n)$ for all $n \in \mathbb{N}_0$. Since all φ_ε are concave, so is $\varphi(n) := \inf_{\varepsilon>0} \varphi_\varepsilon(n)$. Then φ dominates f and

$$\limsup_{n\to\infty} \frac{\varphi(n)}{g(n)} \le \limsup_{n\to\infty} \frac{\varphi_\varepsilon(n)}{g(n)} = \alpha + \varepsilon + \limsup_{n\to\infty} \frac{C_\varepsilon}{g(n)} = \alpha + \varepsilon,$$

$$\liminf_{n\to\infty} \frac{\varphi(n)}{g(n)} = \liminf_{n\to\infty} \inf_{\varepsilon>0} \frac{\varphi_\varepsilon(n)}{g(n)} = \liminf_{n\to\infty} \inf_{\varepsilon>0} \left(\alpha + \varepsilon + \frac{C_\varepsilon}{g(n)}\right) \ge \alpha.$$

The result follows since $\varepsilon > 0$ is arbitrary. □

Acknowledgements

Reik Schottstedt acknowledges support from DFG Grant SPP-1324 DE 1423/3-1. We are grateful for the comments of two referees, in particular, for the references on further related work.

References

[1] M. Ajtai, J. Komlós and G. Tusnàdy. On optimal matchings. *Combinatorica* **4** (4) (1983) 259–264.
[2] F. Barthe and C. Bordenave. Combinatorial optimization over two random point sets. Preprint, 2011. Available at arXiv:1103.2734v2.
[3] J. Beardwood, J. H. Halton and J. M. Hammersley. The shortest path through many points. *Proc. Cambridge Philos. Soc.* **55** (1959) 299–327.
[4] E. Boissard and T. Le Gouic. On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. Preprint, 2011. Available at arXiv:1105.5263v1.
[5] J. A. Bucklew and G. L. Wise. Multidimensional asymptotic quantization theory with rth power distortion measures. *IEEE Trans. Inform. Theory* **28** (1982) 239–247.
[6] S. Delattre, S. Graf, H. Luschgy and G. Pagès. Quantization of probability distributions under norm-based distortion measures. *Statist. Decisions* **22** (4) (2004) 261–282.
[7] S. Dereich. Asymptotic formulae for coding problems and intermediate optimization problems: A review. In *Trends in Stochastic Analysis* 187–232. Cambridge Univ. Press, Cambridge, 2009.
[8] S. Dereich and C. Vormoor. The high resolution vector quantization problem with Orlicz norm distortion. *J. Theoret. Probab.* **24** (2) (2011) 517–544.

[9] V. Dobrić and J. E. Yukich. Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.* **8** (1) (1995) 97–118.
[10] A. Gersho and R. M. Gray. *Vector Quantization and Signal Processing. The Springer International Series in Engineering and Computer Science* **1**. Springer, New York, 1992.
[11] S. Graf and H. Luschgy. *Foundations of Quantization for Probability Distributions. Lecture Notes in Mathematics* **1730**. Springer, Berlin, 2000.
[12] J. Horowitz and R. L. Karandikar. Mean rates of convergence of empirical measures in the Wasserstein metric. *J. Comput. Appl. Math.* **55** (3) (1994) 261–273.
[13] M. Huesmann and T. Sturm. Optimal transport from Lebesgue to Poisson. Preprint, 2010. Available at http://arxiv.org/abs/1012.3845v1.
[14] L. V. Kantorovich. On the translocation of masses. *Dokl. Akad. Nauk* **37** (7–8) (1942) 227–229.
[15] L. V. Kantorovich and G. Rubinstein. On a space of completely additive functions. *Vestnik Leningrad Univ. Math.* **13** (7) (1958) 52–59.
[16] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Mémoires de l'Académie Royale des Sciences* **XVIII–XIX** (1781) 666–704.
[17] T. Müller-Gronbach, K. Ritter and L. Yaroslavtseva. A derandomization of the Euler scheme for scalar stochastic differential equations. Preprint 80, DFG Priority Program 1324, 2011. Available at http://www.dfg-spp1324.de/publications.php?lang=en.
[18] G. Pagès. A space quantization method for numerical integration. *J. Comput. Appl. Math.* **89** (1) (1998) 1–38.
[19] G. Pagès, H. Pham and J. Printems. Optimal quantization methods and applications to numerical problems in finance. In *Handbook on Numerical Methods in Finance* 253–298. Birkhäuser, Boston, 2004.
[20] G. Pagès and J. Printems. Optimal quadratic quantization for numerics: The Gaussian case. *Monte Carlo Methods Appl.* **9** (2) (2003) 135–165.
[21] G. Pagès and B. Wilbertz. Sharp rate for the dual quantization problem. Preprint, 2010.
[22] G. Pagès and B. Wilbertz. Optimal Delaunay and Voronoi quantization schemes for pricing American style options. Preprint, 2011.
[23] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems. Probability and Its Applications* **I**. Springer, New York, 1998.
[24] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems. Probability and Its Applications* **II**. Springer, New York, 1998. MR1619171
[25] R. Schottstedt. Constructive quantization by random sampling and numerical applications. Ph.D. thesis, Technical University Berlin, 2012.
[26] D. W. Stroock and S. R. S. Varadhan. *Multidimensional Diffusion Processes. Grundlehren der Mathematischen Wissenschaften* **233**. Springer, Berlin, 1979.
[27] M. Talagrand. The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. *Ann. Probab.* **22** (2) (1994) 919–959.
[28] C. Villani. *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften* **338**. Springer, Berlin, 2009.
[29] L. N. Wasserstein. Markov processes over denumerable products of spaces describing large systems of automata. *Probl. Inf. Transm.* **5** (1969) 47–52.
[30] P. L. Zador. Topics in the asymptotic quantization of continuous random variables. Bell Laboratories Technical Memorandum, 1966.