DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen”

Constructive quantization: Approximation by empirical measures

S. Dereich, M. Scheutzow, R. Schottstedt

Preprint 84

Edited by

AG Numerik/Optimierung
Fachbereich 12 - Mathematik und Informatik
Philipps-Universität Marburg
Hans-Meerwein-Str.
35032 Marburg

DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

Constructive quantization: Approximation by empirical measures

S. Dereich, M. Scheutzow, R. Schottstedt

Preprint 84

Constructive quantization: approximation by empirical measures

Steffen Dereich*, Michael Scheutzow†, and Reik Schottstedt*

January 27, 2011

Abstract

In this article, we study the approximation of a probability measure μ on $\mathbb{R}^d$ by its empirical measures $\hat{\mu}_N$ interpreted as random quantizations. As error criterion we consider an averaged p-th moment Wasserstein metric. In case $2p < d$, we find a sharp upper bound in terms of a high-resolution formula. In particular, we show that this approach achieves the same order of convergence as optimal quantization.

1 Introduction

Constructive quantization is concerned with the efficient computation of discrete approximations to probability distributions. The need for such approximations comes 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 computer or transmitted via a channel (see e.g. [Zad66, BW82, GG92]); 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. [GPP03, MT07]).

The two applications exhibit a number of similarities. In particular, the minimal error inferred for a given support constraint, the so called *quantization numbers*, expresses in both cases the best achievable performance. However, in the context of *constructive* quantization they are significantly different. In the first problem, the main task is to provide fast encoding and decoding algorithms which are maps mapping the original signal to a digital representation, resp. the digital representation to a reconstruction of the original signal. In the second problem, the main task is to efficiently construct the approximating measure which is formally a collection of points, called *codebook*, together with probability weights.

In this article, we investigate constructive quantization for the second problem. 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. [Pag98, PP03]). Conversely, for large

*Philipps-Universität Marburg, Fachbereich Mathematik und Informatik, Hans-Meerwein-Str., D-35032 Marburg, {dereich,schottstedt}@mathematik.uni-marburg.de

†Technische Universität Berlin, Institut für Mathematik, Straße des 17. Juni 136, D-10623 Berlin, ms@math.tu-berlin.de

codebook sizes and probability measures that are defined implicitly this approach is typically not feasible. As an alternative approach we analyze the use of empirical measures generated by independent random variables distributed according to the original measure. 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.

At first sight the approach is similar to the case with optimal weights treated in [Coh04, Yuk08]. Nonetheless the error criterion and the fixed weights make the problem nonlocal in contrast to all other approaches studied so far.

The application that motivates this approach is approximate sampling. Note that sampling according to the quantization is easy to implement since the approximating measure is the uniform distribution on the codebook.

A concise treatment of asymptotic quantization comprises the derivation of a *high resolution formula*. Such a formula has been established for optimal quantization under norm-based distortions [DGLP04] and for general Orlicz-norm distortions [DV], and, very recently, also in the dual quantization problem [PW].

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.

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 $\mathbb{R}^2$ modelled by a finite measure μ. The demand shall be accomodated by N service centers that are placed at positions $x_1, \dots, x_N \in \mathbb{R}^2$ and that have nonnegative capacities $p_1, \dots, p_N$ summing up to $\|\mu\| := \mu(\mathbb{R}^2)$. 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 $\mathbb{R}^2 \times \mathbb{R}^2$ 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

$$\int_{\mathbb{R}^2 \times \mathbb{R}^2} c(x, y) \, \mathrm{d}\xi(x, y).$$

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.

Definition 1 (pth Wasserstein metric) The pth *Wasserstein metric* of two finite measures μ and ν on $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$, 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 ρ 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 [Mon81]. Important results about the Wasserstein metric were achieved within the scope of *transportation theory*, for instance by Kantorovich [Kan42], Kantorovich and Rubinstein [KR58], Wasserstein [Was69], Rachev and Rüschendorf [RR98a],[RR98b] 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 analyse 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 $X_1, \dots, X_N$, that is

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

and, for fixed $p \geq 1$, we analyse the asymptotic behaviour of the so called *random quantization error*

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

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 minimisation of the latter quantity reduces to a minimisation over all sets $\mathcal{C} \subset \mathbb{R}^d$ of at most N elements. Furthermore, the minimal error is the so called Nth *quantization number*

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

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.

1.2 Main results

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

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

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

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

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,

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

  is of order $N^{-1/2p}$ and, hence, converges to zero more slowly than $N^{-1/d}$.
- For the uniform distribution $\mathcal{U}$ on $[0,1)^d$, the results of Talagrand [Tal94] imply that $V_{N,p}^{\mathrm{rand}}(\mathcal{U})$ is always of order $N^{-1/d}$ as long as $d \geq 3$.

The following theorem is a *high resolution formula* for the random approach.

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

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

(ii) Let μ be a probability measure on $\mathbb{R}^d$ that has a finite qth 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

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

We conjecture that the lim sup in (3) is actually a limit and that the inequality is an equality. We conjecture further that the assumption that the density is Riemann integrable is unnecessary.

Let us compare our results with the classical high resolution formulas. The asymptotics of $V_{N,p}^{\mathrm{opt}}$ defined in (1) is given by

$$\lim_{N\to\infty} N^{1/d} \, V_{N,p}^{\mathrm{opt}}(\mu) = c_{p,d} \left(\int_{\mathbb{R}^d} \left(\frac{\mathrm{d}\mu_a}{\mathrm{d}\lambda^d}\right)^{\frac{d}{d+p}} \mathrm{d}\lambda^d\right)^{1/d+1/p}, \tag{4}$$

whenever μ has a finite moment of order $p + \delta$ for some $\delta > 0$, see [GL00, p. 96]. Here, the constant $c_{p,d}$ (which is explicitly known in a few special cases) is the corresponding limit for the uniform distribution on the unit cube in $\mathbb{R}^d$. By definition, the right hand side of (4) is at most equal to the right hand side of (3). Recall however that computing close to optimal quantizers is technically only feasible up to moderate codebook sizes and under complete knowledge of μ.

In the case where the codebook is generated by i.i.d. samples and the weights are chosen optimally, high resolution formulas have been derived in [GL00, p. 127ff]. One finds the same estimates as in 2, with a different constant κ_p^{unif}. Moreover, one can replace lim sup by lim and the inequality by an equality in (3).

2 Preliminaries

For a finite signed measure μ on $\mathbb{R}^d$ 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

$$\xi(\mathrm{d}x, \mathrm{d}y) = \nu(\mathrm{d}x) K(x, \mathrm{d}y), \tag{5}$$

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

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

Now we can iterate transport schedules: Let $\nu_0, \ldots, \nu_n$ 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 $\xi_1, \ldots, \xi_n$ is formally the transport described by the source ν_0 and the probability kernel $K = K_1 * \cdots * K_n$, where $K_1, \ldots, K_n$ are the kernels associated to $\xi_1, \ldots, \xi_n$. Note that the relation (5) 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 elementary properties of the Wasserstein metric in a lemma.

Lemma 1 *Let $\xi, \mu, \mu_1, \ldots$ and $\nu, \nu_1, \ldots$ be finite measures on $\mathbb{R}^d$ such that $\|\xi\| = \|\mu\| = \|\nu\|$.*

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

$$\rho_p^p(\mu, \nu) \leq \sum_{k=1}^{\infty} \rho_p^p(\mu_k, \nu_k). \tag{6}$$

(ii) ***Triangle-inequality:*** *One has*

$$\rho_p(\mu,\nu) \le \rho_p(\mu,\xi) + \rho_p(\xi,\nu). \tag{7}$$

(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*

$$\rho_p\left(\mu \circ T^{-1}, \nu \circ T^{-1}\right) = a\,\rho_p(\mu,\nu). \tag{8}$$

3 Proof of the main results

3.1 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 $N \in \mathbb{N}$*

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

Remark 2 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

$$\kappa_p^{\mathrm{cube}} = \mathfrak{d}\, 2^{\frac{d-2}{2p}} \left[\frac{1}{1-2^{p-\frac{d}{2}}} + \frac{1}{1-2^{-p}}\right]^{\frac{1}{p}}.$$

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^dary 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 $C \in \mathcal{P}$*

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

Then

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

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 masses, we call the measure ν defined by

$$\nu\big|_{A_k} = \frac{\mu_2(A_k)}{\mu_1(A_k)}\,\mu_1\big|_{A_k}$$

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 ν is associated to a transport from μ_1 to ν. Note that

$$(\mu_1 \wedge \nu)\big|_{A_k} = \frac{\mu_1(A_k) \wedge \mu_2(A_k)}{\mu_1(A_k)}\, \mu_1\big|_{A_k}$$

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

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

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

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

Proof of Lemma 2. For $l \in \mathbb{N}_0$, we set

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

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

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

so that $\mu_{l+1}|_F$ is the $\{C \in \mathcal{P}_{l+1} : C \subset A\}$-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

$$\xi_F(\{(x,y) : x \neq y\}) = \frac{1}{2} \sum_{C \text{ child of } F} \Big|\nu(C) - \nu(F)\frac{\mu(C)}{\mu(F)}\Big|. \tag{9}$$

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

$$\rho_p(\mu, \nu) \le \sup_{l \in \mathbb{N}} \rho_p(\mu, \mu_l). \tag{10}$$

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, there exists a sequence $(Z_l)_{l \in \mathbb{N}_0}$ of $[0,1)^d$-valued random variables with

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

for every $l \in \mathbb{N}$. Then the joint distribution of (Z_0, Z_l) is ξ^l. Let

$$L = \inf\{l \in N_0 : Z_{l+1} \neq Z_l\}$$

and note that 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 $k \in \mathbb{N}$

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

where we used (9) in the last step, so the assertion follows by (10). □

Proof of Proposition 1. We apply the above lemma with $\nu = \hat{\mu}_N$. Hence,

$$\rho_p^p(\mu, \hat{\mu}_N) \leq \frac{1}{2}\mathfrak{d}^p \sum_{l=0}^{\infty} 2^{-pl} \sum_{F \in \mathcal{P}_l} \sum_{C \text{ child of } F} \Big|\hat{\mu}_N(C) - \hat{\mu}_N(F)\frac{\mu(C)}{\mu(F)}\Big|. \tag{11}$$

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 $(\mu(C)/\mu(F))_{C \text{ child of } F}$. Hence,

$$\begin{aligned}&\mathbb{E}\Big[\sum_{C \text{ child of } F} \Big|\hat{\mu}_N(C) - \hat{\mu}_N(F)\frac{\mu(C)}{\mu(F)}\Big|\Big|N\hat{\mu}_N(F) = k\Big] \\ &\qquad = \frac{1}{N}\, \mathbb{E}\Big[\sum_{C \text{ child of } F} \Big|N\, \hat{\mu}_N(C) - k\, \frac{\mu(C)}{\mu(F)}\Big|\Big|N\hat{\mu}_N(F) = k\Big] \\ &\qquad \leq \frac{1}{N} \sum_{C \text{ child of } F} \operatorname{var}\big(N\, \hat{\mu}_N(C)\big|\big|N\hat{\mu}_N(F) = k\big)^{1/2} \\ &\qquad \leq \frac{\sqrt{k}}{N} \sum_{C \text{ child of } F} \sqrt{\frac{\mu(C)}{\mu(F)}} \leq 2^{\frac{d}{2}}\frac{\sqrt{k}}{N},\end{aligned}$$

where we used Jensen's inequality in the last step. We set $\zeta(t) = \sqrt{t} \wedge t$ ($t \geq 0$) and observe that

$$\mathbb{E}\Big[\sum_{C \text{ child of } F} \Big|\hat{\mu}_N(C) - \hat{\mu}_N(F)\frac{\mu(C)}{\mu(F)}\Big|\Big] \leq \frac{2^{\frac{d}{2}}}{N}\, \zeta(N\mu(F)).$$

Consequently, it follows from (11) and Jensen's inequality that

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

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

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

so the assertion follows. □

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 $\|.\|_{\max}$.

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*

$$V_{N,p}^{\mathrm{rand}}(\mu) \leq \kappa_{p,q}^{\mathrm{Pierce}} \left[\int_{\mathbb{R}^d} \|x\|_{\max}^q \, \mathrm{d}\mu(x) \right]^{1/q} N^{-1/d}, \tag{12}$$

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

Proof. By the scaling invariance of inequality (12), 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

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

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

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

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 $\mathcal{M}(\mu, \nu)$, such that

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

Note that $N\hat{\mu}_N(B_n) \sim \mathrm{Bin}(N, \mu(B_n))$ and that by the Markov inequality

$$\mu(B_n) \leq 2^{-q(n-1)} \int \|x\|_{\max}^q \, \mathrm{d}\mu(x) = 2^{-q(n-1)}. \tag{13}$$

The inequality remains true for $n = 0$. Thus

$$\begin{aligned}
\mathbb{E}[\rho_p^p(\mu,\nu)] &\le \sum_{n=0}^{\infty} 2^{p-1}2^{np}\mathfrak{d}^p\,\mathbb{E}\left[|\mu(B_n) - \hat{\mu}_N(B_n)|\right] \\
&\le \sum_{n=0}^{\infty} 2^{p-1}2^{np}\mathfrak{d}^p\, N^{-\frac{1}{2}}\mu(B_n)^{\frac{1}{2}} \qquad (14)\\
&\le 2^{p+\frac{q}{2}-1}\mathfrak{d}^p N^{-\frac{1}{2}} \sum_{n=0}^{\infty} 2^{n(p-\frac{1}{2}q)} = \frac{2^{p+\frac{q}{2}-1}}{1-2^{p-\frac{1}{2}q}}\mathfrak{d}^p N^{-\frac{1}{2}}.
\end{aligned}$$

It remains to analyse $\mathbb{E}[\rho_p^p(\nu,\hat{\mu}_N)]$. 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 (i) and Proposition 1,

$$\begin{aligned}
\mathbb{E}[\rho_p^p(\nu,\hat{\mu}_N)] &\le \sum_{n=0}^{\infty} \mathbb{E}[\rho_p^p(\nu|_{B_n},\hat{\mu}_N|_{B_n})] \\
&\le \sum_{n=0}^{\infty}\sum_{k=1}^{\infty} \mathbb{P}(N\hat{\mu}_N(B_n) = k)\, 2^{(n+1)p}\,\frac{k}{N}(\kappa_p^{\text{cube}})^p k^{-p/d}.
\end{aligned}$$

Using that $\mathbb{E}\left[\hat{\mu}_N(B_n)\right] = \mu(B_n)$, we conclude with Jensen's inequality that

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

We use again inequality (13) to derive

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

Note that $\frac{p}{d} \le \frac{1}{2}$ and altogether, we finish the proof by applying the triangle inequality (property (ii) of Lemma 1) and equation (14) to deduce that

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

□

3.2 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 limit

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

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 an enumeration of the entries of v in A.

For a Borel set A 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 $\lambda^d(A) < \infty$ 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

$$\mathbb{E}\left[\sum_{j=1}^{N} \mathbb{1}_{\{X_j \neq Y_j\}}\right] \leq \frac{\sqrt{K}\sqrt{N}}{2}. \tag{15}$$

Proof. For $k = 1, \dots, K$, denote by $X^{(k)}$ the A_k-subvector of X. For each k with $\xi_k \leq \text{length}(X^{(k)})$, 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 $\xi_k - \text{length}(X^{(k)})$ 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. Since the length of the A_k-subvector is binomially distributed with parameters N and $q_k := \frac{\lambda^d(A_i \cap A)}{\lambda^d(A)}$, we get

$$\mathbb{E}\left[\sum_{j=1}^{N} \mathbb{1}_{\{X_j \neq Y_j\}}\right] = \frac{1}{2}\mathbb{E}\left[\sum_{k=1}^{K} \left|\text{length}(X^{(k)}) - \xi_k\right|\right] \leq \frac{1}{2}\sum_{i=1}^{K} \text{var}\left(\text{length}(X^{(k)})\right)^{1/2}$$

$$\leq \frac{1}{2}\sqrt{N}\sum_{k=1}^{K}\sqrt{q_k} \leq \frac{1}{2}\sqrt{K}\sqrt{N}.$$

□

Proof of Statement (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 $B^{\text{main}} := [0, \lfloor 1/a \rfloor a)^d$ and the remainder $B^{\text{rem}} := [0,1)^d \setminus B^{\text{main}}$. Note that $\lambda^d(B^{\text{rem}}) \to 0$ as $N \to \infty$. We represent B^{main} as the union of $n = \lfloor a^{-1} \rfloor^d$ pairwise disjoint translates of B_0:

$$B^{\text{main}} = \cup_{k=1}^{n} B_k.$$

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

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

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 $A_{\mathrm{child}} \in \mathcal{T}_l$ is the child of a box $A_{\mathrm{parent}} \in \mathcal{T}_{l-1}$ if $A_{\mathrm{child}} \subset A_{\mathrm{parent}}$.

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

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

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

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

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

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

Together with the former estimate we get

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

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

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

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

$$\mathbb{E}\big[\frac{1}{N}\sum_{j=1}^{N}\|X_j-Y_j\|^p\big]^{1/p} \le \mathbb{E}\big[\frac{1}{N}\sum_{j=1}^{N}\|X_j-X_j^0\|^p\big]^{1/p}+\mathbb{E}\big[\frac{1}{N}\sum_{j=1}^{N}\|X_j^0-X_j^L\|^p\big]^{1/p}$$
$$\le C\left[N^{-\frac{1}{2p}}+M^{-(\frac{1}{2p}-\frac{1}{d})}\,N^{-\frac{1}{d}}\right]. \tag{16}$$

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 $\hat{\mu}_M^{(k)}$. Morever, 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 that

$$N\,\mathbb{E}[\rho_p^p(\hat{\mu}_N^Y,\mathcal{U})] \le \sum_{k=1}^{n} M\,\mathbb{E}[\rho_p^p(\hat{\mu}_M^{(k)},\mathcal{U}(B_k))]+(N-nM)\,\mathbb{E}[\rho_p^p(\hat{\mu}^{\mathrm{rem}}_{N-nM},\mathcal{U}(B^{\mathrm{rem}}))]$$
$$\le nMa^p(V^{\mathrm{rand}}_{M,p}(\mathcal{U}))^p+(\kappa_p^{\mathrm{cube}})^p(N-nM)^{1-p/d}.$$

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

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

Moreover, (16) implies that

$$\limsup_{N\to\infty} N^{1/d}\,\mathbb{E}[\rho_p^p(\hat{\mu}_N^X,\hat{\mu}_N^Y)]^{1/p}\le C\,M^{-(\frac{1}{2p}-\frac{1}{d})}.$$

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

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

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

3.3 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 any $\varepsilon>0$ a finite number of cubes $B_1,\dots,B_n$ (which are parallel to the coorinate axes) and positive reals $\alpha_1,\dots,\alpha_n$ such that $\nu:=\sum\alpha_k\,\mathcal{U}(B_k)$ satisfies

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

Remark 3 Since we can express a measure which is approachable from below as the limit of a sequence of measures with Lebesgue density, it has itself a Lebesgue density. Conversely, any finite measure which has a density which is Riemann integrable on any cube, is approachable from below.

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

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

Proof. Let $\varepsilon > 0$ and choose a finite number of pairwise disjoint cubes $B_1, \ldots, B_K$ and positive reals $\alpha_1, \ldots, \alpha_K$ such that $\mu^* := \sum_{k=1}^K \alpha_k \mathcal{U}(B_k) \le \mu$ and $\|\mu - \mu^*\| \le \varepsilon$. For $k = 1, \ldots, K$ 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 $M = (M_k)_{k=0,\ldots,K}$ with parameters N and $(\alpha_k)_{k=0,\ldots,K}$ 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

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

The first expression on the right hand side is of order $\mathcal{O}(N^{-1/2p})$, (see proof of Proposition 3). By Theorem 2 (i), 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

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

Denote by $a_1, ..., a_K$ the edge lengths of the cubes $B_1, ..., B_K$ 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 and Jensen's inequality,

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

so that

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

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

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

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 $B_1,\dots,B_K$ with

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

We set $B_0=[0,1)^d\backslash(B_1\cup\dots\cup B_K)$ and define the probability measure ν, as in Lemma 2, by $\nu:=\sum_{k=0}^K \nu|_{B_k}$ where

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

Then the vector $Z:=(N\hat\mu_N(B_k))_{k=0,\dots,K}$ is multinomially distributed with parameters N and $(\mu(B_k))_{k=0,\dots,K}$. Hence, by Lemma 2

$$\mathbb{E}[\rho_p^p(\mu,\nu)]^{1/p} \le \Big(\frac{1}{2N}\mathfrak{d}^{\mathfrak{p}}\sum_{k=0}^K \mathbb{E}|Z_k - N\mu(B_k)|\Big)^{1/p} = \mathcal{O}(N^{-1/2p}). \tag{17}$$

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

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

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

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

It follows from (17) and the triangle inequality that

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

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 has density f which is approachable from below, then*

$$\limsup_{N\to\infty} N^{1/d}\, V_{N,p}^{rand}(\mu) \le \kappa_p^{\text{unif}} \left(\int_{\mathbb{R}^d} f(x)^{1-\frac{p}{d}}\, \mathrm{d}x\right)^{1/p}. \tag{18}$$

Proof. Let $\delta > 0$ and set

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

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 $\nu = \sum_{k=1}^3 \frac{M_k}{N} \mu^{(k)}$

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

Further, by Lemma 1,

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

and, by Propositions 2 and 3, there exist concave functions φ_1 and φ_2 with

$$n\, V_{n,p}^{\text{rand}}(\mu^{(k)})^p \le \varphi_k(n), \qquad \text{for } n \in \mathbb{N}, \quad k = 1,2$$

and

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

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

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

Analogously,

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

and, by Proposition 3,

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

Altogether, we get

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

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

3.4 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 on $\mathbb{R}^d$ such that $\|\mu\| = \|\nu\|$. Then one has*

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

Proof. One has

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

$\square$

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

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

$$\limsup_{N\to\infty} N^{\frac{1}{d}} |V_{N,1}^{\mathrm{rand}}(\mu) - V_{N,1}^{\mathrm{rand}}(\nu)| \le 2\kappa_{1,q}^{\mathrm{Pierce}} \|\mu - \nu\|^{1-\frac{1}{d}-\frac{1}{q}} \Big(\int \|x\|_{\max}^q \, |\mu - \nu|(\mathrm{d}x)\Big)^{\frac{1}{q}}.$$

Proof. Without loss of generality, we assume that $\mu \ne \nu$. 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

$$N\,\hat{\mu}_N = M_1 \hat{\alpha}_{M_1} + M_2 \hat{\mu}^*_{M_2} \quad \text{and} \quad N\,\hat{\nu}_N = M_1 \hat{\alpha}_{M_1} + M_2 \hat{\nu}^*_{M_2}.$$

Then

$$\begin{aligned} \rho(N\mu, N\hat{\mu}_N) &\leq \rho(N\mu, M_1\alpha + M_2\mu^*) + \rho(M_1\alpha + M_2\mu^*, M_1\hat{\alpha}_{M_1} + M_2\hat{\mu}^*_{M_2}) \\ &\leq \rho(N\mu, M_1\alpha + M_2\mu^*) + \rho(M_1\alpha, M_1\hat{\alpha}_{M_1}) + \rho(M_2\mu^*, M_2\hat{\mu}^*_{M_2}). \end{aligned} \tag{19}$$

Observe that

$$\mathbb{E}[\rho(N\mu, M_1\alpha + M_2\mu^*)] = \mathcal{O}(N^{1/2}). \tag{20}$$

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

$$\mathbb{E}[\rho(M_2\mu^*, M_2\hat{\mu}^*_{M_2})] \leq \kappa^{\text{Pierce}}_{1,q} \|\mu - \nu\|^{1-\frac{1}{d}-\frac{1}{q}} N^{1-\frac{1}{d}} \Big(\int \|x\|^q_{\max}\,(\mu-\nu)_+(\mathrm{d}x)\Big)^{\frac{1}{q}} + \mathcal{O}(N^{\frac{1}{2}}), \tag{21}$$

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

$$\begin{aligned} \rho(M_1\alpha, M_1\hat{\alpha}_{M_1}) &= \rho(M_1\alpha + M_2\hat{\nu}^*_{M_2}, M_1\hat{\alpha}_{M_1} M_2\hat{\nu}^*_{M_2}) \\ &= \rho(M_1\alpha + M_2\hat{\nu}^*_{M_2}, N\hat{\nu}_N) \\ &\leq \rho(N\nu, N\hat{\nu}_N) + \rho(M_1\alpha + M_2\hat{\nu}^*_{M_2}, N\nu) \\ &= \rho(N\nu, N\hat{\nu}_N) + \rho(M_1\alpha + M_2\hat{\nu}^*_{M_2} + M_2\nu^*, N\nu + M_2\nu^*) \\ &\leq \rho(N\nu, N\hat{\nu}_N) + \rho(M_2\hat{\nu}^*_{M_2}, M_2\nu^*) + \rho(M_1\alpha + M_2\nu^*, N\nu). \end{aligned}$$

The expected values of the last two summands can be estimated like (21) and (20). Inserting the estimates into (19), the assertion of the lemma follows. □

We now prove the general upper bound 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

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

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

Acknowledgements. Reik Schottstedt acknowledge support from DFG Grant SPP-1324 DE 1423/3-1.

References

[BW82] J. A. Bucklew and G. L. Wise. Multidimensional asymptotic quantization theory with rth power distortion measures. *IEEE Trans. Inf. Theory*, 28:239–247, 1982.

[Coh04] P. Cohort. Limit theorems for random normalized distortion. *Ann. Appl. Probab.*, 14(1):118–143, 2004.

[DGLP04] S. Delattre, S. Graf, H. Luschgy, and G. Pagès. Quantization of probability distributions under norm-based distortion measures. *Stat. Decis.*, 22(4):261–282, 2004.

[DV] S. Dereich and C. Vormoor. The high resolution vector quantization problem with Orlicz norm distortion. *J. Theor. Probab.* to appear.

[GG92] A. Gersho and R. M. Gray. *Vector Quantization and Signal Processing*, volume 1 of *The Springer International Series in Engineering and Computer Science.* Springer, 1992.

[GL00] S. Graf and H. Luschgy. *Foundations of Quantization for Probability Distributions.* Lecture Notes in Mathematics. 1730. Berlin: Springer, 2000.

[GPP03] H. Pham G. Pagès and J. Printems. Optimal quantization methods and applications to numerical problems in finance. *Handbook on Numerical Methods in Finance*, 2004:253–298, 2003.

[Kan42] L. V. Kantorovich. On the translocation of masses. *Doklady Akademii Nauk SSSR*, 37(No. 7-8):227–229, 1942.

[KR58] L. V. Kantorovich and G. Rubinstein. On a space of completely additive functions. *Vestnik Leningrad Univ. Math.*, 13(7):52–59, 1958.

[Mon81] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Mémoires de l'Académie Royale des Sciences*, XVIII-XIX:666–704, 1781.

[MT07] I. Molchanov and N. Tontchev. Optimal approximation and quantisation. *J. Math. Anal. Appl.*, 325(2):1410–1429, 2007.

[Pag98] G. Pagès. A space quantization method for numerical integration. *J. Comput. Appl. Math.*, 89(1):1–38, 1998.

[PP03] G. Pagès and J. Printems. Optimal quadratic quantization for numerics: the Gaussian case. *Monte Carlo Methods Appl.*, 9(2):135–165, 2003.

[PW] G. Pagès and B. Wilbertz. Sharp rate for the dual quantization problem. Preprint.

[RR98a] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems*, volume I of *Probability and its applications.* Springer, 1998.

[RR98b] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems*, volume II of *Probability and its applications.* Springer, 1998.

[Tal94] M. Talagrand. The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. *The Annals of Probability*, 22(2):919–959, 1994.

[Was69] L. N. Wasserstein. Markov processes over denumerable products of spaces describing large systems of automata. *Prob. Inf. Transmission*, 5:47–52, 1969.

[Yuk08] J. Yukich. Limit theorems for multi-dimensional random quantizers. *Electron. Commun. Probab.*, 13:507–517, 2008.

[Zad66] P. L. Zador. Topics in the asymptotic quantization of continuous random variables. Bell Laboratories Technical Memorandum, 1966.

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

[1] R. Ramlau, G. Teschke, and M. Zhariy. A Compressive Landweber Iteration for Solving Ill-Posed Inverse Problems. Preprint 1, DFG-SPP 1324, September 2008.

[2] G. Plonka. The Easy Path Wavelet Transform: A New Adaptive Wavelet Transform for Sparse Representation of Two-dimensional Data. Preprint 2, DFG-SPP 1324, September 2008.

[3] E. Novak and H. Woźniakowski. Optimal Order of Convergence and (In-) Tractability of Multivariate Approximation of Smooth Functions. Preprint 3, DFG-SPP 1324, October 2008.

[4] M. Espig, L. Grasedyck, and W. Hackbusch. Black Box Low Tensor Rank Approximation Using Fibre-Crosses. Preprint 4, DFG-SPP 1324, October 2008.

[5] T. Bonesky, S. Dahlke, P. Maass, and T. Raasch. Adaptive Wavelet Methods and Sparsity Reconstruction for Inverse Heat Conduction Problems. Preprint 5, DFG-SPP 1324, January 2009.

[6] E. Novak and H. Woźniakowski. Approximation of Infinitely Differentiable Multivariate Functions Is Intractable. Preprint 6, DFG-SPP 1324, January 2009.

[7] J. Ma and G. Plonka. A Review of Curvelets and Recent Applications. Preprint 7, DFG-SPP 1324, February 2009.

[8] L. Denis, D. A. Lorenz, and D. Trede. Greedy Solution of Ill-Posed Problems: Error Bounds and Exact Inversion. Preprint 8, DFG-SPP 1324, April 2009.

[9] U. Friedrich. A Two Parameter Generalization of Lions' Nonoverlapping Domain Decomposition Method for Linear Elliptic PDEs. Preprint 9, DFG-SPP 1324, April 2009.

[10] K. Bredies and D. A. Lorenz. Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding. Preprint 10, DFG-SPP 1324, April 2009.

[11] K. Bredies and D. A. Lorenz. Regularization with Non-convex Separable Constraints. Preprint 11, DFG-SPP 1324, April 2009.

[12] M. Döhler, S. Kunis, and D. Potts. Nonequispaced Hyperbolic Cross Fast Fourier Transform. Preprint 12, DFG-SPP 1324, April 2009.

[13] C. Bender. Dual Pricing of Multi-Exercise Options under Volume Constraints. Preprint 13, DFG-SPP 1324, April 2009.

[14] T. Müller-Gronbach and K. Ritter. Variable Subspace Sampling and Multi-level Algorithms. Preprint 14, DFG-SPP 1324, May 2009.

[15] G. Plonka, S. Tenorth, and A. Iske. Optimally Sparse Image Representation by the Easy Path Wavelet Transform. Preprint 15, DFG-SPP 1324, May 2009.

[16] S. Dahlke, E. Novak, and W. Sickel. Optimal Approximation of Elliptic Problems by Linear and Nonlinear Mappings IV: Errors in L_2 and Other Norms. Preprint 16, DFG-SPP 1324, June 2009.

[17] B. Jin, T. Khan, P. Maass, and M. Pidcock. Function Spaces and Optimal Currents in Impedance Tomography. Preprint 17, DFG-SPP 1324, June 2009.

[18] G. Plonka and J. Ma. Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing. Preprint 18, DFG-SPP 1324, June 2009.

[19] G. Teschke and C. Borries. Accelerated Projected Steepest Descent Method for Nonlinear Inverse Problems with Sparsity Constraints. Preprint 19, DFG-SPP 1324, July 2009.

[20] L. Grasedyck. Hierarchical Singular Value Decomposition of Tensors. Preprint 20, DFG-SPP 1324, July 2009.

[21] D. Rudolf. Error Bounds for Computing the Expectation by Markov Chain Monte Carlo. Preprint 21, DFG-SPP 1324, July 2009.

[22] M. Hansen and W. Sickel. Best m-term Approximation and Lizorkin-Triebel Spaces. Preprint 22, DFG-SPP 1324, August 2009.

[23] F.J. Hickernell, T. Müller-Gronbach, B. Niu, and K. Ritter. Multi-level Monte Carlo Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 23, DFG-SPP 1324, August 2009.

[24] S. Dereich and F. Heidenreich. A Multilevel Monte Carlo Algorithm for Lévy Driven Stochastic Differential Equations. Preprint 24, DFG-SPP 1324, August 2009.

[25] S. Dahlke, M. Fornasier, and T. Raasch. Multilevel Preconditioning for Adaptive Sparse Optimization. Preprint 25, DFG-SPP 1324, August 2009.

[26] S. Dereich. Multilevel Monte Carlo Algorithms for Lévy-driven SDEs with Gaussian Correction. Preprint 26, DFG-SPP 1324, August 2009.

[27] G. Plonka, S. Tenorth, and D. Roşca. A New Hybrid Method for Image Approximation using the Easy Path Wavelet Transform. Preprint 27, DFG-SPP 1324, October 2009.

[28] O. Koch and C. Lubich. Dynamical Low-rank Approximation of Tensors. Preprint 28, DFG-SPP 1324, November 2009.

[29] E. Faou, V. Gradinaru, and C. Lubich. Computing Semi-classical Quantum Dynamics with Hagedorn Wavepackets. Preprint 29, DFG-SPP 1324, November 2009.

[30] D. Conte and C. Lubich. An Error Analysis of the Multi-configuration Time-dependent Hartree Method of Quantum Dynamics. Preprint 30, DFG-SPP 1324, November 2009.

[31] C. E. Powell and E. Ullmann. Preconditioning Stochastic Galerkin Saddle Point Problems. Preprint 31, DFG-SPP 1324, November 2009.

[32] O. G. Ernst and E. Ullmann. Stochastic Galerkin Matrices. Preprint 32, DFG-SPP 1324, November 2009.

[33] F. Lindner and R. L. Schilling. Weak Order for the Discretization of the Stochastic Heat Equation Driven by Impulsive Noise. Preprint 33, DFG-SPP 1324, November 2009.

[34] L. Kämmerer and S. Kunis. On the Stability of the Hyperbolic Cross Discrete Fourier Transform. Preprint 34, DFG-SPP 1324, December 2009.

[35] P. Cerejeiras, M. Ferreira, U. Kähler, and G. Teschke. Inversion of the noisy Radon transform on $SO(3)$ by Gabor frames and sparse recovery principles. Preprint 35, DFG-SPP 1324, January 2010.

[36] T. Jahnke and T. Udrescu. Solving Chemical Master Equations by Adaptive Wavelet Compression. Preprint 36, DFG-SPP 1324, January 2010.

[37] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Irregular Shearlet Frames: Geometry and Approximation Properties. Preprint 37, DFG-SPP 1324, February 2010.

[38] G. Kutyniok and W.-Q Lim. Compactly Supported Shearlets are Optimally Sparse. Preprint 38, DFG-SPP 1324, February 2010.

[39] M. Hansen and W. Sickel. Best m-Term Approximation and Tensor Products of Sobolev and Besov Spaces – the Case of Non-compact Embeddings. Preprint 39, DFG-SPP 1324, March 2010.

[40] B. Niu, F.J. Hickernell, T. Müller-Gronbach, and K. Ritter. Deterministic Multi-level Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 40, DFG-SPP 1324, March 2010.

[41] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Construction of Compactly Supported Shearlet Frames. Preprint 41, DFG-SPP 1324, March 2010.

[42] C. Bender and J. Steiner. Error Criteria for Numerical Solutions of Backward SDEs. Preprint 42, DFG-SPP 1324, April 2010.

[43] L. Grasedyck. Polynomial Approximation in Hierarchical Tucker Format by Vector-Tensorization. Preprint 43, DFG-SPP 1324, April 2010.

[44] M. Hansen und W. Sickel. Best m-Term Approximation and Sobolev-Besov Spaces of Dominating Mixed Smoothness - the Case of Compact Embeddings. Preprint 44, DFG-SPP 1324, April 2010.

[45] P. Binev, W. Dahmen, and P. Lamby. Fast High-Dimensional Approximation with Sparse Occupancy Trees. Preprint 45, DFG-SPP 1324, May 2010.

[46] J. Ballani and L. Grasedyck. A Projection Method to Solve Linear Systems in Tensor Format. Preprint 46, DFG-SPP 1324, May 2010.

[47] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence Rates for Greedy Algorithms in Reduced Basis Methods. Preprint 47, DFG-SPP 1324, May 2010.

[48] S. Kestler and K. Urban. Adaptive Wavelet Methods on Unbounded Domains. Preprint 48, DFG-SPP 1324, June 2010.

[49] H. Yserentant. The Mixed Regularity of Electronic Wave Functions Multiplied by Explicit Correlation Factors. Preprint 49, DFG-SPP 1324, June 2010.

[50] H. Yserentant. On the Complexity of the Electronic Schrödinger Equation. Preprint 50, DFG-SPP 1324, June 2010.

[51] M. Guillemard and A. Iske. Curvature Analysis of Frequency Modulated Manifolds in Dimensionality Reduction. Preprint 51, DFG-SPP 1324, June 2010.

[52] E. Herrholz and G. Teschke. Compressive Sensing Principles and Iterative Sparse Recovery for Inverse and Ill-Posed Problems. Preprint 52, DFG-SPP 1324, July 2010.

[53] L. Kämmerer, S. Kunis, and D. Potts. Interpolation Lattices for Hyperbolic Cross Trigonometric Polynomials. Preprint 53, DFG-SPP 1324, July 2010.

[54] G. Kutyniok and W.-Q Lim. Shearlets on Bounded Domains. Preprint 54, DFG-SPP 1324, July 2010.

[55] A. Zeiser. Wavelet Approximation in Weighted Sobolev Spaces of Mixed Order with Applications to the Electronic Schrödinger Equation. Preprint 55, DFG-SPP 1324, July 2010.

[56] G. Kutyniok, J. Lemvig, and W.-Q Lim. Compactly Supported Shearlets. Preprint 56, DFG-SPP 1324, July 2010.

[57] A. Zeiser. On the Optimality of the Inexact Inverse Iteration Coupled with Adaptive Finite Element Methods. Preprint 57, DFG-SPP 1324, July 2010.

[58] S. Jokar. Sparse Recovery and Kronecker Products. Preprint 58, DFG-SPP 1324, August 2010.

[59] T. Aboiyar, E. H. Georgoulis, and A. Iske. Adaptive ADER Methods Using Kernel-Based Polyharmonic Spline WENO Reconstruction. Preprint 59, DFG-SPP 1324, August 2010.

[60] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the Convergence of Generalized Polynomial Chaos Expansions. Preprint 60, DFG-SPP 1324, August 2010.

[61] S. Holtz, T. Rohwedder, and R. Schneider. On Manifolds of Tensors of Fixed TT-Rank. Preprint 61, DFG-SPP 1324, September 2010.

[62] J. Ballani, L. Grasedyck, and M. Kluge. Black Box Approximation of Tensors in Hierarchical Tucker Format. Preprint 62, DFG-SPP 1324, October 2010.

[63] M. Hansen. On Tensor Products of Quasi-Banach Spaces. Preprint 63, DFG-SPP 1324, October 2010.

[64] S. Dahlke, G. Steidl, and G. Teschke. Shearlet Coorbit Spaces: Compactly Supported Analyzing Shearlets, Traces and Embeddings. Preprint 64, DFG-SPP 1324, October 2010.

[65] W. Hackbusch. Tensorisation of Vectors and their Efficient Convolution. Preprint 65, DFG-SPP 1324, November 2010.

[66] P. A. Cioica, S. Dahlke, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Spatial Besov Regularity for Stochastic Partial Differential Equations on Lipschitz Domains. Preprint 66, DFG-SPP 1324, November 2010.

[67] E. Novak and H. Woźniakowski. On the Power of Function Values for the Approximation Problem in Various Settings. Preprint 67, DFG-SPP 1324, November 2010.

[68] A. Hinrichs, E. Novak, and H. Woźniakowski. The Curse of Dimensionality for Monotone and Convex Functions of Many Variables. Preprint 68, DFG-SPP 1324, November 2010.

[69] G. Kutyniok and W.-Q Lim. Image Separation Using Shearlets. Preprint 69, DFG-SPP 1324, November 2010.

[70] B. Jin and P. Maass. An Analysis of Electrical Impedance Tomography with Applications to Tikhonov Regularization. Preprint 70, DFG-SPP 1324, December 2010.

[71] S. Holtz, T. Rohwedder, and R. Schneider. The Alternating Linear Scheme for Tensor Optimisation in the TT Format. Preprint 71, DFG-SPP 1324, December 2010.

[72] T. Müller-Gronbach and K. Ritter. A Local Refinement Strategy for Constructive Quantization of Scalar SDEs. Preprint 72, DFG-SPP 1324, December 2010.

[73] T. Rohwedder and R. Schneider. An Analysis for the DIIS Acceleration Method used in Quantum Chemistry Calculations. Preprint 73, DFG-SPP 1324, December 2010.

[74] C. Bender and J. Steiner. Least-Squares Monte Carlo for Backward SDEs. Preprint 74, DFG-SPP 1324, December 2010.

[75] C. Bender. Primal and Dual Pricing of Multiple Exercise Options in Continuous Time. Preprint 75, DFG-SPP 1324, December 2010.

[76] H. Harbrecht, M. Peters, and R. Schneider. On the Low-rank Approximation by the Pivoted Cholesky Decomposition. Preprint 76, DFG-SPP 1324, December 2010.

[77] P. A. Cioica, S. Dahlke, N. Döhring, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Adaptive Wavelet Methods for Elliptic Stochastic Partial Differential Equations. Preprint 77, DFG-SPP 1324, January 2011.

[78] G. Plonka, S. Tenorth, and A. Iske. Optimal Representation of Piecewise Hölder Smooth Bivariate Functions by the Easy Path Wavelet Transform. Preprint 78, DFG-SPP 1324, January 2011.

[79] A. Mugler and H.-J. Starkloff. On Elliptic Partial Differential Equations with Random Coefficients. Preprint 79, DFG-SPP 1324, January 2011.

[80] T. Müller-Gronbach, K. Ritter, and L. Yaroslavtseva. A Derandomization of the Euler Scheme for Scalar Stochastic Differential Equations. Preprint 80, DFG-SPP 1324, January 2011.

[81] W. Dahmen, C. Huang, C. Schwab, and G. Welper. Adaptive Petrov-Galerkin methods for first order transport equations. Preprint 81, DFG-SPP 1324, January 2011.

[82] K. Grella and C. Schwab. Sparse Tensor Spherical Harmonics Approximation in Radiative Transfer. Preprint 82, DFG-SPP 1324, January 2011.

[83] D.A. Lorenz, S. Schiffler, and D. Trede. Beyond Convergence Rates: Exact Inversion With Tikhonov Regularization With Sparsity Constraints. Preprint 83, DFG-SPP 1324, January 2011.

[84] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: Approximation by empirical measures. Preprint 84, DFG-SPP 1324, January 2011.