Optimal vector quantization in terms of Wasserstein distance

Wolfgang Kreitmeier

Wolfgang Kreitmeier, University of Passau, Department of Informatics and Mathematics, D-94032 Passau, Germany

Abstract

The optimal quantizer in memory-size constrained vector quantization induces a quantization error which is equal to a Wasserstein distortion. However, for the optimal (Shannon-)entropy constrained quantization error a proof for a similar identity is still missing. Relying on principal results of the optimal mass transportation theory, we will prove that the optimal quantization error is equal to a Wasserstein distance. Since we will state the quantization problem in a very general setting, our approach includes the Rényi-α-entropy as a complexity constraint, which includes the special case of (Shannon-)entropy constrained ($\alpha = 1$) and memory-size constrained ($\alpha = 0$) quantization. Additionally, we will derive for certain distance functions codecell convexity for quantizers with a finite codebook. Using other methods, this regularity in codecell geometry has already been proved earlier by György and Linder [12, 13].

Key words: Wasserstein distance, optimal quantization error, codecell convexity, Rényi-α-entropy
2000 MSC: 60B10, 60E05, 62E17, 68P30, 94A17, 94A29

1. Introduction

Optimal quantization often arises in electrical engineering in connection with signal processing and data compression. The survey article of Gray and Neuhoff [9] provides a comprehensive overview of this subject. In mathematical terms, quantization is concerned with the approximation of a given proba-

Email address: `wolfgang.kreitmeier@uni-passau.de` (Wolfgang Kreitmeier)

bility by another probability which is induced as an image under a quantizer. In doing so, the complexity of the quantizer must not exceed a certain bound. Optimal quantization is achieved if the quantization error between the original distribution and the approximation is, subject to the given bound, minimal. A first rigorous treatment of this problem in a higher dimensional space apparently goes back to Steinhaus [30]. The complexity of a quantizer can be measured by different mappings. Standard choices are the support cardinality of the approximating distribution [8] or its (Shannon-)entropy [12, 13]. In the first case, we are talking about memory-size (or fixed-rate) quantization, the second one is called (Shannon-)entropy constrained quantization. Recently, Rényi-α-entropy has been suggested as complexity mapping [16, 17], which contains the special case of (Shannon-)entropy constrained ($\alpha = 1$) and memory-size constrained ($\alpha = 0$) quantization. Alternatively, the quantization error between the original probability and its approximation can be interpreted as the costs arising out of the mass transport between these two distributions. Indeed, in case of $\alpha = 0$ and for distance mapping $l(x) = x^r$ with $r \geq 1$, it is well-known (see e.g. [8, Lemma 3.4], [24]) that the optimal quantization error is equal to the Wasserstein distance between original and (optimal) approximation, which reflects these costs and is a key term in the theory of optimal mass transportation. Our main goal is to prove that this identity remains valid also for general complexity and distance mappings.

By using principal results in optimal mass transportation theory, we will show in this paper for a fairly large class of distance mappings l that the optimal quantization error is equivalent to the minimization of a Wasserstein distance (cf. Thm. 3.2). Because we make only very few assumptions regarding the complexity mapping (cf. Definition 2.1), the case of Rényi-α-entropy and, therefore, the special cases of memory-size and (Shannon-) entropy constrained optimal quantization are included. Results from mass transportation theory yield that the codecells, i.e. the preimages of the quantizer, have the shape of convex polytopes if the distance mapping is quadratic. Moreover, the codecells are intervals for a large class of distance mappings in the one-dimensional setting. Using other methods, this regularity in codecell geometry has already been proved earlier by György and Linder [12, 13].

The rest of this paper is organized as follows: The second section contains the setup of optimal quantization and mass transportation theory. In the third section, we introduce different types of quantization errors and Wasserstein distances. Our main result (Theorem 3.2) shows that these different notions coincide under very general assumptions on distance and complexity mapping. In particular, we obtain codecell regularity for special distance mappings. Additionally, for distance mapping $l(x) = x^r$ with $r \geq 1$, we will generalize a consistency result for the optimal quantization error (cf. Corollary 3.9). The last section introduces Rényi-α-entropy as complexity mapping and compares the results of this paper with known results for the cases $\alpha \in \{0, 1\}$. Most of our proofs are given in the appendix.

2. Setup and notation

2.1. Optimal quantization

We begin with a very general definition of optimal quantization. Let $d \in \mathbb{N} = \{1, 2, ..\}$ and μ be a Borel probability measure on $\mathbb{R}^d$. Let $I \subset \mathbb{N}$ and $\mathcal{S} = \{S_i : i \in I\}$ be a countable and measurable partition of $\mathbb{R}^d$. Moreover, let $\mathcal{C} = \{c_i : i \in I\}$ be a countable set of points in $\mathbb{R}^d$. Now $(S_i, c_i)_{i \in I}$ defines a **quantizer** $q : \mathbb{R}^d \to \mathcal{C}$ with

$$q(x) = c_i \qquad \text{if and only if} \qquad x \in S_i.$$

We call $\mathcal{C}$ a **codebook** consisting of codepoints c_i. Every $S_i \in \mathcal{S}$ is called a **codecell**. Clearly, $\mathcal{C} = q(\mathbb{R}^d)$. Moreover, if we assume w.l.o.g. that $c_i \neq c_j$ for every $i, j \in I, i \neq j$, then

$$\mathcal{S} = \{q^{-1}(z) : z \in q(\mathbb{R}^d)\}.$$

Denote by $\mathcal{Q}_d$ the set of all quantizers and by δ_a the Dirac measure in $a \in \mathbb{R}^d$. For every $q \in \mathcal{Q}_d$, the image measure

$$\mu \circ q^{-1}(\cdot) = \sum_{i \in I} \mu(S_i) \delta_{c_i}(\cdot)$$

has a countable support and defines an approximation of μ, the so-called quantization of μ by q. Now let $\mathcal{P}$ be the space of all probability vectors on $[0, 1]^{\mathbb{N}}$, i.e., for every $p = (p_i)_{i \in \mathbb{N}} \in \mathcal{P}$ we have $p_i \in [0, 1]$ and $\sum_{i=1}^{\infty} p_i = 1$.

Definition 2.1. *We call a mapping $\mathcal{P} \ni p \to H(p) \in [0, \infty]$ a* ***complexity mapping*** *if*

(a) for any bijection $\tau : \mathbb{N} \to \mathbb{N}$ and $(p_i)_{i \in \mathbb{N}} \in \mathcal{P}$ we have $H((p_i)_{i \in \mathbb{N}}) = H((p_{\tau(i)})_{i \in \mathbb{N}})$, and

(b) for every $p \in \mathcal{P}$ and $k \geq 2$ with $p^k = (p_1, .., p_{k-1}, \sum_{i \geq k} p_i, 0, ..) \in \mathcal{P}$ we have $H(p^k) \leq H(p)$.

With any enumeration $\{i_1, i_2, ...\}$ of I we define

$$H_\mu(q) = H((\mu(S_{i_1}), \mu(S_{i_2}), ...))$$

as the $H-$complexity of q w.r.t μ. Now we intend to quantify the distance between μ and its approximation under q. To this end, let $\| \cdot \|$ be the Euclidean norm on $\mathbb{R}^d$ and $l : [0, \infty) \to [0, \infty)$ a strictly increasing (and therefore Borel-measurable) distance mapping with $l(0) = 0$. For $q \in \mathcal{Q}_d$ we define as the distance between μ and $\mu \circ q^{-1}$ the quantization error

$$D_\mu(q) = \int l(\| x - q(x) \|) d\mu(x). \tag{1}$$

For any $R \geq 0$ we denote

$$D_\mu^H(R) = \inf\{D_\mu(q) : q \in \mathcal{Q}_d, H_\mu(q) \leq R\} \tag{2}$$

as the optimal quantization error of μ under H-complexity bound R. We call a quantizer q optimal for μ under H-complexity bound R if $D_\mu(q) = D_\mu^H(R)$.

Denote by $\mathcal{Q}_d^*$ the set of all quantizers whose range is finite. It is essential for this paper and also of principal interest that we can replace $\mathcal{Q}_d$ with $\mathcal{Q}_d^*$ in relation (2) under a moment condition on μ. To this end, let $\mathcal{M}(\mathbb{R}^d)$ be the set of all Borel probability measures on $\mathbb{R}^d$ with finite $l-$moment, i.e. $\int l(\|x\|) d\mu(x) < \infty$ for every $\mu \in \mathcal{M}(\mathbb{R}^d)$. The following statement is proved in the appendix.

Proposition 2.2. *Let $\mu \in \mathcal{M}(\mathbb{R}^d)$. For every complexity mapping H and bound $R \geq 0$ we have*

$$D_\mu^H(R) = \inf\{D_\mu(q) : q \in \mathcal{Q}_d^*, H_\mu(q) \leq R\}.$$

As already stated in the introduction, two choices for the complexity mapping are of great practical importance. Quantization with the complexity mapping

$$\mathcal{P} \ni p \to H(p) = \log(\sum_{i \in \mathbb{N}} 1_{(0,1]}(p_i)) \tag{3}$$

is called memory-size constrained quantization if we denote by 1_A the characteristic function on a set $A \subset \mathbb{R}^d$. If

$$\mathcal{P} \ni p \to H(p) = -\sum_{i\in\mathbb{N}} p_i \log(p_i), \tag{4}$$

then we are talking of (Shannon-)entropy constrained quantization. Rényi-α-entropy as a more general complexity constraint is discussed in section 4. For memory-size constrained quantization, optimal quantizers exist under weak assumptions on μ (see e.g. [8, Thm. 4.12], [24]). If μ is non-atomic, György and Linder [12, Thm. 3] have shown for (Shannon-)entropy constrained quantization in the one-dimensional case that always optimal quantizers exist. If μ is absolutely continuous with respect to the Lebesgue measure and $l(x) = x^2$, then this existence results holds also for higher dimensions (cf. [13, Thm. 3]). Unfortunately, optimal quantizers do not exist in general. There are complexity mappings H which lead to the non-existence of optimal quantizers (cf. [16, Thm. 3.1]).

2.2. Transportation theory and its relation to quantization and the Wasserstein distance

The problem of optimal transportation in the sense of Kantorovich [15] can be stated on finite dimensional spaces as follows: Let X and Y be closed and non-empty subsets of $\mathbb{R}^d$. Let μ be a Borel probability measure on X and ν be a Borel probability measure on Y. Consider the set $\Gamma(\mu, \nu)$ of all Borel probability measures on $X \times Y$ with first marginal μ and second marginal ν. Kantorovich's problem was to determine the minimal transport cost

$$w(\mu, \nu) = \inf\{\int_{X\times Y} c(x, y) d\gamma(x, y) : \gamma \in \Gamma(\mu, \nu)\} \tag{5}$$

with the measurable cost function $c(\cdot, \cdot) : X \times Y \to [0, \infty)$. We call $\gamma \in \Gamma(\mu, \nu)$ an optimal transport plan if

$$w(\mu, \nu) = \int_{X\times Y} c(x, y) d\gamma(x, y).$$

Because the cost function has only non-negative values in our setting, an optimal solution always exists (cf. [31, Thm. 4.1]). Now we specify for the rest of this paper

$$c(x, y) = l(\|x - y\|) \qquad \text{for every } (x, y) \in X \times Y.$$

Definition 2.3. *A transport plan $\gamma \in \Gamma(\mu, \nu)$ is said to be* ***deterministic*** *if a measurable mapping $q : X \to Y$ exists, such that $\gamma = \mu \circ \phi^{-1}$ with the mapping*

$$X \ni x \to \phi(x) = (x, q(x)).$$

The mapping q is called Monge mapping or transport mapping.

Consequently, every deterministic transport plan is induced by a Monge mapping q which is μ-almost surely uniquely defined (when the transport plan has been fixed). Moreover, $\nu = \mu \circ q^{-1}$. Roughly speaking, one could say that the Monge mapping q transports the mass represented by the measure μ to the mass represented by the measure ν.

If we restrict the transport plans in (5) to be deterministic, the problem of optimal transport turns into the so-called Monge transportation problem, where we have to determine the optimal Monge mapping q, such that

$$\begin{aligned} & \int l(\|x - q(x)\|) d\mu(x) \\ = \;& \inf\{\int l(\|x - t(x)\|) d\mu(x) : t \text{ measurable }, \mu \circ t^{-1} = \nu\} \\ = \;& \inf\{\int_{X \times Y} c(x, y) d\gamma(x, y) : \gamma \in \Gamma(\mu, \nu), \gamma \text{ deterministic }\}. \end{aligned} \tag{6}$$

Now, if we compare (5), (6) and (1), it turns out that the optimal total cost of transportation would be a quantization error if the optimal transport plan would be deterministic and the related optimal Monge map would be a quantizer. The following Theorem 2.6 states that this is the case if the target distribution ν is discrete and the source distribution μ satisfies a certain continuity assumption. Our proof of this fundamental statement relies on principal results in the theory of optimal mass transportation. We need the following definition:

Definition 2.4. *[31, Definition 5.2 and Remark 5.6] A mapping $f : X \to \mathbb{R}$ is said to be $c-$convex if there exists a mapping $g : Y \to \mathbb{R}$ such that*

$$f(x) = \sup\{g(y) - c(x, y) : y \in Y\} \quad \textit{for every } x \in X.$$

Moreover, the $c-$subdifferential or $c-$subgradient $\partial_c f(x)$ of the mapping f at the point $x \in X$ is defined by

$$\partial_c f(x) = \{y \in Y : f(x) + c(x, y) \leq f(z) + c(z, y) \textit{ for every } z \in X\}.$$

We rely in this paper on the following fundamental result. Recall definition (5) of $w(\mu, \nu)$. Denote by card the cardinality of a set.

Theorem 2.5. *[31, Theorem 5.30] Assume that $w(\mu,\nu) < \infty$. If for any $c-$convex mapping $f : X \to \mathbb{R}$ the set*

$$\{x \in X : \mathrm{card}(\partial_c f(x)) > 1\}$$

is contained in a $\mu-$measure-zero set, then there exists a unique optimal transport plan γ which is deterministic. The Monge map q which induces γ is characterized by the existence of a $c-$convex function $\tilde{f}$ such that

$$q(x) \in \partial_c \tilde{f}(x) \text{ for } \mu - a.e. \quad x \in X.$$

Now we intend to apply Theorem 2.5 to a discrete target distribution ν on $\mathbb{R}^d$. To be precise, let $m \in \mathbb{N}$ and $a_1, .., a_m \in \mathbb{R}^d$ be m different points in $\mathbb{R}^d$. Let $p_1, .., p_m \in (0,1]$ be such that $\sum_{i=1}^m p_i = 1$ and let

$$\nu = \sum_{i=1}^{m} p_i \delta_{a_i} \tag{7}$$

For $(\lambda_1, .., \lambda_m) \in \mathbb{R}^m$ and $i \in \{1, .., m\}$, we define the set

$$\begin{aligned} A_i &= A_i(\lambda_1, .., \lambda_m) \quad (8) \\ &= \{x \in \mathbb{R}^d : l(\|x - a_i\|) - \lambda_i = \min\{l(\|x - a_j\|) - \lambda_j : j \in \{1, .., m\}\}\}. \end{aligned}$$

Moreover, let

$$\begin{aligned} \mathcal{Q}(\nu) &= \{q \in \mathcal{Q}_d^* : q(\mathbb{R}^d) = \{a_1, .., a_m\} \\ &\qquad \text{with } \mu(q^{-1}(a_i)) = p_i \text{ for every } i \in \{1, .., m\}\} \end{aligned}$$

Theorem 2.6. *Let ν be a discrete probability on $\mathbb{R}^d$ as defined in (7). Assume that μ vanishes on the boundary of $A_i(\beta_1, .., \beta_m)$ for every $(\beta_1, .., \beta_m) \in \mathbb{R}^m$ and $i \in \{1, .., m\}$. If $w(\mu, \nu) < \infty$, then a quantizer $q \in \mathcal{Q}(\nu)$ and $(\lambda_1, .., \lambda_m) \in \mathbb{R}^m$ exist, such that*

$$\begin{aligned} w(\mu, \nu) &= \int l(\|x - q(x)\|) d\mu(x) \\ &= \min\{\int l(\|x - \tilde{q}(x)\|) d\mu(x) : \tilde{q} \in \mathcal{Q}(\nu)\} \end{aligned}$$

and

$$q^{-1}(a_i) = A_i(\lambda_1, .., \lambda_m) \quad \mu - \textit{almost surely for every } i \in \{1, .., m\}.$$

The quantizer q is $\mu-$almost surely uniquely defined, i.e. we have $\mu \circ q^{-1} = \mu \circ \tilde{q}^{-1}$ for every $\tilde{q} \in \mathcal{Q}(\nu)$ which attains the above minimum .

Theorem 2.6 is proved in the appendix by applying Theorem 2.5. If $r \geq 1$ and $l(x) = x^r$, then the mapping

$$\rho_r(\cdot, \cdot) = l^{-1} \circ w(\cdot, \cdot) = w(\cdot, \cdot)^{1/r} \tag{9}$$

is a metric on $\mathcal{M}(\mathbb{R}^d) \times \mathcal{M}(\mathbb{R}^d)$ (see e.g. [31, Definition 6.1]) and often called **Wasserstein distance**. In this paper, we are generally interested in such continuous distance mappings l, where $l^{-1} \circ w$ satisfies the triangle inequality, i.e. if for every $\mu_1, \mu_2, \mu_3 \in \mathcal{M}(\mathbb{R}^d)$ with $w(\mu_1, \mu_3) < \infty$ the relation

$$l^{-1}(w(\mu_1, \mu_3)) \leq l^{-1}(w(\mu_1, \mu_2)) + l^{-1}(w(\mu_2, \mu_3))$$

holds. Even if the triangle inequality is satisfied, it is not clear if the mapping $l^{-1} \circ w$ has always finite values. Thus $l^{-1} \circ w$ does not satisfy a priori all axioms of a metric, even if the triangle inequality is in force. For a historical overview of Kantorovich's problem and further aspects of transportation theory, the reader is referred to Rüschendorf [28] and the references therein. Ambrosio et.al. [4] is also a good source of information for Wasserstein distances.

3. The optimal quantization error in terms of a Wasserstein distortion

Recall $\mathcal{M}(\mathbb{R}^d)$ as the set of all Borel probability measures on $\mathbb{R}^d$ with finite $l-$moment. We denote by $\text{supp}(\mu)$ the support of $\mu \in \mathcal{M}(\mathbb{R}^d)$ and define

$$\mathcal{M}^*(\mathbb{R}^d) = \{\mu \in \mathcal{M}(\mathbb{R}^d) : \text{card}(\text{supp}(\mu)) < \infty\},$$
$$\mathcal{M}^\infty(\mathbb{R}^d) = \{\mu \in \mathcal{M}(\mathbb{R}^d) : \text{card}(\text{supp}(\mu)) \leq \text{card}(\mathbb{N})\}.$$

Let $\nu \in \mathcal{M}^\infty(\mathbb{R}^d)$ and denote $\text{supp}(\nu) = \{a_i : i \in I\}$ with $I \subset \mathbb{N}$. By adding zeros - if necessary - the distribution ν induces a probability vector $p^\nu \in \mathcal{P}$, where $\sum_{i=1}^\infty p_i^\nu = \sum_{i \in I} \nu(a_i) = 1$. According to property (a) of H the mapping $\nu \to H(p^\nu)$ is well-defined. For $\mu \in \mathcal{M}(\mathbb{R}^d)$ and $R \geq 0$ we define

$$V_\mu^H(R) = \inf\{w(\mu, \nu) : \nu \in \mathcal{M}^\infty(\mathbb{R}^d), H(p^\nu) \leq R\}$$

as the optimal total cost of transportation for source distribution μ, where the target distributions ν have countable support and induce a $H-$complexity which is lower or equal than the bound R. If $l \circ w^{-1}$ satisfies the triangle inequality or l is bounded, then we can replace $\mathcal{M}^\infty(\mathbb{R}^d)$ by $\mathcal{M}^*(\mathbb{R}^d)$ in the definition of $V_\mu^H(R)$. The following statement is proved in the appendix.

Proposition 3.1. *Let $\mu \in \mathcal{M}(\mathbb{R}^d)$ and assume that l is continuous and $l^{-1} \circ w$ satisfies the triangle inequality or assume that l is bounded and continuous. For every complexity mapping H and bound $R \geq 0$ we have*

$$V_\mu^H(R) = \inf\{w(\mu, \nu) : \nu \in \mathcal{M}^*(\mathbb{R}^d), H(p^\nu) \leq R\}. \tag{10}$$

Because we are interested in the case where the target distributions ν are induced by a quantizer whose range is finite, we also define

$$W_\mu^H(R) = \inf\{w(\mu, \mu \circ q^{-1}) : q \in \mathcal{Q}_d^*, H_\mu(q) \leq R\}. \tag{11}$$

Obviously,

$$V_\mu^H(R) \leq W_\mu^H(R). \tag{12}$$

Example 4.3 will show that inequality (12) can be strict. Recall definition (2) of the optimal quantization error $D_\mu^H(R)$. In view of Proposition 2.2 and Proposition 3.1 it is natural to ask under which conditions (if any) the quantities $W_\mu^H(R)$, $V_\mu^H(R)$ and $D_\mu^H(R)$ coincide. In addition to this question (which will be answered in Theorem 3.2), we want to know more about the codecell geometry of the quantizers. Such knowledge has proved very useful in analyzing optimal scalar and vector quantizer performance [9]. Of particular interest is the question if it suffices to consider in definition (2) only quantizers whose image has finite cardinality and whose codecells are convex polytopes. To be precise, let $a \in \mathbb{R}^d$ and $b \in \mathbb{R}^d, a \neq b$. We define the closed halfspace

$$T(a, b) = \{x \in \mathbb{R}^d : \|x - a\| \leq \|x - b\|\}. \tag{13}$$

We call a set $P \subset \mathbb{R}^d$ a convex polytope, if P is a finite intersection of closed or open halfspaces. Let $\mathcal{Q}_d^c \subset \mathcal{Q}_d^*$ be the set of all quantizers where each codecell is a convex polytope and every codepoint of such codecell lies in the closure of the codecell. We define

$$D_\mu^{H,c}(R) = \inf\{D_\mu(q) : q \in \mathcal{Q}_d^c, H_\mu(q) \leq R\}.$$

From the definition we immediately obtain for any $R \geq 0$ that

$$D_\mu^H(R) \leq D_\mu^{H,c}(R). \tag{14}$$

Inequality (14) can be strict (cf. [12, Example 1]), but Theorem 3.2 below states conditions under which (14) turns into an equation. We are interested in two subclasses of distance functions, namely

(C1) l is continuous, twice continuously differentiable with $l'' \geq 0$ in $(0, \infty)$ and $l^{-1} \circ w$ satisfies the triangle inequality

(C2) l is continuous, twice continuously differentiable with $l'' \leq 0$ in $(0, \infty)$ and bounded.

Now we can state the main result of this paper. Theorem 3.2 will be proved in the appendix.

Theorem 3.2. *Let $\mu \in \mathcal{M}(\mathbb{R}^d)$ and assume that μ vanishes on continuously differentiable $(d-1)$-dimensional submanifolds of $\mathbb{R}^d$. Let the distance function l be of type (C1) or (C2). For every $R \geq 0$ we have*

$$V_\mu^H(R) = W_\mu^H(R) = D_\mu^H(R). \tag{15}$$

Additionally, if
(a) $d = 1$ and l is of type (C1), or
(b) $d > 1$ and $l(x) = x^2$ for every $x \geq 0$, then

$$D_\mu^{H,c}(R) = D_\mu^H(R). \tag{16}$$

As already stated, (cf. 9) the distance mappings $l(x) = x^r$ satisfy condition (C1). Now we state a criterion for distance mappings ensuring that they satisfy condition (C1). To this end, we need the following definition.

Definition 3.3. *A mapping $f : [0, \infty) \to [0, \infty)$ is called superadditive, if*

$$f(x + y) \geq f(x) + f(y) \text{ for every } x, y \in [0, \infty).$$

Remark 3.4. *Consider the following classes of distance functions*

(D1) l is continuous, twice continuously differentiable with $l'' > 0$ in $(0, \infty)$ and the mapping $\frac{l'}{l''}$ is superadditive in $(0, \infty)$

(D2) $l(x) = x$ for every $x \in [0, \infty)$.

Lemma A.2 states that every distance mapping which satisfies (D1) or (D2) also satisfies condition (C1). Insofar, Theorem 3.2 holds also for distance mappings which satisfy (D1) or (D2).

Example 3.5. *The function $l(x) = x \exp(-1/x)$ for $x > 0$, $l(0) = 0$ lies in class (D1) and thus according to Remark 3.4 also in class (C1). Theorem 3.2 is also applicable for the concave distance mappings $l(x) = \arctan(x)$ or $l(x) = \tanh(x)$ as elements of class (C2).*

Remark 3.6. *Common distance mappings are those who satisfy an Orlicz's condition (see e.g. [25, Example 2.2.1])*

$$\sup\{l(2x)/l(x) : x > 0\} < \infty. \tag{17}$$

It is not difficult to construct distance mappings which are continuous, twice continuously differentiable with $l'' \geq 0$ in $(0, \infty)$ and satisfying (17), but are not superadditive. It remains open whether Theorem 3.2 is still true if we replace the triangle inequality for $l^{-1} \circ w$ in (C1) by condition (17).

Example 4.3 will show that Theorem 3.2 becomes invalid in general if we drop that μ vanishes on continuously differentiable $(d-1)$-dimensional submanifolds of $\mathbb{R}^d$. We denote by λ^d the $d-$dimensional Lebesgue measure on $\mathbb{R}^d$.

Remark 3.7. *Every Borel measure μ on $\mathbb{R}^d$ which is absolutely continuous with respect to λ^d, vanishes on continuously differentiable $(d-1)$-dimensional submanifolds of $\mathbb{R}^d$. By a result of Mattila [23], also the Hausdorff measure restricted to a self-similar set whose span equals $\mathbb{R}^d$ and satisfies the open set condition, vanishes on continuously differentiable $(d-1)$-dimensional submanifolds of $\mathbb{R}^d$.*

Remark 3.8. *The boundedness of the distance mapping in the definition of class (C2) and the triangle inequality in (C1) is only needed in the proof of Proposition 3.1. These are sufficient conditions to ensure that equation (10) is true. It remains open to characterize those distance mappings l for which equation (10) is true.*

The identity (16) has already been shown by György and Linder [12, 13]. Although they investigate only the case of (Shannon-)entropy constrained quantization, their proof works also in our more general setting, because their construction of a (finite) quantizer with convex codecells always starts from an arbitrary one by redefining the codecells to convex ones but having the same probability.

For the special distance mapping $l(x) = x^r$ with fixed norm exponent $r \geq 1$, we can easily derive a consistency result for the optimal quantization error. In the special case of memory-size constrained quantization, the result is well-known (see e.g. [8, p. 57], [24, Thm. 9]). Recall the definition (9) of Wasserstein distance ρ_r.

Corollary 3.9. *Let $\mu_1, \mu_2 \in \mathcal{M}(\mathbb{R}^d)$ and assume that μ_i vanishes on continuously differentiable $(d-1)$-dimensional submanifolds of $\mathbb{R}^d$ for $i \in \{1,2\}$. Let $r \geq 1$ and assume that $l(x) = x^r$ for every $x \geq 0$. Let $R \geq 0$. Then*

$$|(D^H_{\mu_1}(R))^{1/r} - (D^H_{\mu_2}(R))^{1/r}| \leq \rho_r(\mu_1, \mu_2). \tag{18}$$

Proof. We distinguish two cases.
1. $D^H_{\mu_1}(R) \geq D^H_{\mu_2}(R)$.
Let $\varepsilon > 0$. According to Theorem 3.2, let $\nu_2 \in \mathcal{M}^\infty(\mathbb{R}^d)$, such that $H(p^{\nu_2}) \leq R$ and

$$(D^H_{\mu_2}(R))^{1/r} \geq \rho_r(\mu_2, \nu_2) - \varepsilon.$$

Again by Theorem 3.2, we obtain

$$\begin{aligned} & |(D^H_{\mu_1}(R))^{1/r} - (D^H_{\mu_2}(R))^{1/r}| \\ = \; & (D^H_{\mu_1}(R))^{1/r} - (D^H_{\mu_2}(R))^{1/r} \\ \leq \; & \inf\{\rho_r(\mu_1, \nu) - \rho_r(\mu_2, \nu_2) : \nu \in \mathcal{M}^\infty(\mathbb{R}^d), H(p^\nu) \leq R\} + \varepsilon \\ \leq \; & \inf\{\rho_r(\mu_1, \mu_2) + \rho_r(\nu, \nu_2) : \nu \in \mathcal{M}^\infty(\mathbb{R}^d), H(p^\nu) \leq R\} + \varepsilon \\ = \; & \rho_r(\mu_1, \mu_2) + \varepsilon. \end{aligned}$$

By letting $\varepsilon \to 0$, we obtain (18).
2. $D^H_{\mu_1}(R) < D^H_{\mu_2}(R)$.
This case is handled similarly to the first one. □

Corollary 3.9 becomes invalid if we drop the condition that μ_i vanishes on continuously differentiable $(d-1)$-dimensional submanifolds of $\mathbb{R}^d$ for $i \in \{1,2\}$. For a (counter-)example, the reader is referred to [12, Example 2]. Insofar we cannot apply Corollary 3.9 to the important case, where

$\mu_2 = \mu_2^{(n)}$ is the empirical ($n-$sample) version of μ_1. Although we know that consistency in this empirical case holds for memory-size constrained quantization ([8, Corollary 4.24]), it remains open if this is true in general.

Nevertheless, Corollary 3.9 could also be of practical relevance for algorithmic quantizer design. Algorithms for designing optimal quantizers often converge to a local error minimum which is not a global one. To avoid this effect, a perturbation approach has been proposed [1, 19] in case of memory-size constrained quantization. The original distribution μ is approximated (in a weak sense) by $\mu_n = (1 - a_n)\mu + a_n\nu$, where $(a_n) \subset (0, 1)$ is a sequence decreasing to zero. ν represents a distribution which has a unique local (and global) optimal quantizer. Now if an optimal quantizer for μ_n is used as the initial (suboptimal) quantizer for μ_{n+1}, the algorithm is likely to converge to a (global) optimal quantizer for μ_{n+1}. Corollary 3.9 ensures that the quantization errors of these (global) optimal quantizers converge towards the optimal quantization error for μ. It needs further research to determine if this approach also works for general complexity mappings.

4. Rényi-α-entropy as complexity and comparison with known results

Let us give an exact definition of Rényi-α-entropy [26, 29]. Let $\mathbb{N} := \{1, 2, ..\}$. Let $\alpha \in [-\infty, \infty]$ and $p = (p_1, p_2, ...) \in \mathcal{P}$. The Rényi-$\alpha$-entropy $H^\alpha(p) \in [0, \infty]$ is defined as (cf. [3, Definition 5.2.35], see also [14], p.1)

$$H^\alpha(p) = \begin{cases} -\sum_{i=1}^\infty p_i \log(p_i), \text{ if } \alpha = 1 \\ -\log\left(\sup\{p_i : i \in \mathbb{N}\}\right), \text{ if } \alpha = \infty \\ -\log\left(\inf\{p_i : i \in \mathbb{N}, p_i > 0\}\right), \text{ if } \alpha = -\infty \\ \frac{1}{1-\alpha} \log\left(\sum_{i=1}^\infty p_i^\alpha\right), \text{ if } \alpha \in (-\infty, \infty)\backslash\{1\}. \end{cases}$$

We use the conventions $0 \cdot \log(0) := 0$ and $0^x := 0$ for all real $x \geq 0$. Moreover, $1/0 := \infty$. The logarithm log is based on e.

Remark 4.1. *With these conventions we obtain*

$$H^0(p) = \log(\sum_{i \in \mathbb{N}} 1_{(0,1]}(p_i)).$$

Using l'Hospital's rule it is easy to see, that the case $\alpha = 1$ will be reached from $\alpha \neq 1$ by taking the limit $\alpha \to 1$. (see e.g. [3, Remark 5.2.34]).

Moreover, one has

$$\lim_{\alpha\to\infty} H^\alpha(\cdot) = H^\infty(\cdot) \text{ and } \lim_{\alpha\to-\infty} H^\alpha(\cdot) = H^{-\infty}(\cdot).$$

Now let us show that H^α is a complexity mapping in the sense of definition 2.1.

Proposition 4.2. *The mapping H^α is a complexity mapping for every $\alpha \in [-\infty, \infty]$.*

Proof. Clearly, H^α satisfies condition (a) of a complexity mapping. To show that condition (b) of a complexity mapping is also satisfied, we distinguish several cases.

1. $\alpha \in \{-\infty, 0, \infty\}$

In this case, we deduce immediately from the definition that H^α satisfies condition (b).

2. $\alpha = 1$.

Let $p \in \mathcal{P}$ and $k \geq 2$ with $p^k = (p_1, .., p_{k-1}, \sum_{i\geq k} p_i, 0, ..) \in \mathcal{P}$. If $H^1(p) = \infty$, then we have nothing to prove. So, let us assume that $H^1(p) < \infty$. From recursivity of Shannon entropy (cf. [3, relation (1.2.8)]), we obtain

$$H^1(p^k) \leq H^1(p^{k+1})$$

Due to $H^1(p^k) \to H^1(p) < \infty$ if $k \to \infty$ we obtain that condition (b) is satisfied.

3. $\alpha \in (-\infty, 1)\backslash\{0\}$.

Let $p \in \mathcal{P}$ and $k \geq 2$. Due to $\alpha < 1$ we obtain with the convention $1/0 := \infty$ that

$$\sum_{i=k}^{\infty} p_i^\alpha \geq (\sum_{i=k}^{\infty} p_i)^\alpha, \tag{19}$$

yielding $H^\alpha(p^k) \leq H^\alpha(p)$.

4. $\alpha \in (1, \infty)$.

In this case, inequality (19) holds in reversed order, yielding again $H^\alpha(p^k) \leq H^\alpha(p)$. □

In view of relation (3) memory-size constrained quantization is quantization with complexity H^0 and according to (4) Shannon-entropy constrained quantization uses H^1 as complexity mapping. Quantization with Rényi-α-entropy as complexity has been investigated in [16, 17, 18].

As already announced in section 3 we will now give an example showing that Theorem 3.2 becomes invalid in general if μ does not vanish on continuously differentiable $(d-1)$-dimensional submanifolds of $\mathbb{R}^d$.

Example 4.3. *Let $l(x) = x^2$ for $x \geq 0$ and $d = 1$. Let $z \in (1/2, 1)$ and*

$$\mu = (1/3) \cdot (\delta_0 + \delta_z + \delta_1).$$

Let $\alpha > 0$ and $p = (1/3, 2/3, 0, ..) \in \mathcal{P}$. Define $R_0 = H^\alpha(p)$. Now let

$$\nu = (1/5) \cdot \delta_0 + (4/5) \cdot \delta_{(1+z)/2}.$$

It is plain to see that $H^\alpha(p^\nu) < R_0$. Now let $R \in (0, R_0)$. Let $q \in \mathcal{Q}_1$ with $H^\alpha_\mu(q) \leq R$. From the definition of R_0 we obtain that q consists of only one codecell. As shown in [8, Example 2.3(b)], the optimal codepoint $\{c\} = q(\mathbb{R})$ equals the centre of mass, i.e.

$$c = \frac{1+z}{3}.$$

We calculate

$$\begin{aligned} D_\mu(q) &= \frac{1}{3}((0-c)^2 + (z-c)^2 + (1-c)^2) \\ &= \frac{2}{9}(1 - z + z^2) = D^{H^\alpha}_\mu(R') \end{aligned}$$

for every $R' \leq R$. Because q consists of only one codecell, we obtain

$$W^H_\mu(R) = D_\mu(q) = D^{H^\alpha,c}_\mu(R). \tag{20}$$

On the other hand we have

$$\begin{aligned} \rho_2^2(\mu, \nu) &\leq \frac{1}{5} \cdot 0^2 + (\frac{1}{3} - \frac{1}{5}) \cdot (\frac{1+z}{2})^2 + 2 \cdot \frac{1}{3}(\frac{1+z}{2} - z)^2 \\ &= \frac{1}{5}(1 - \frac{4}{3}z + z^2). \end{aligned}$$

Together with (12) and (20) we deduce

$$V^H_\mu(R) < W^H_\mu(R) = D^{H^\alpha}_\mu(R) = D^{H^\alpha,c}_\mu(R).$$

Remark 4.4. *For distance mapping $l(x) = x^2$ and Shannon-entropy as complexity ($\alpha = 1$), an identity similar to (15) has been proved by Linder [20, Lemma 1] for the so-called Lagrangian distortion. But for an entropy bound $R > 0$, this modified distortion only coincides with the quantization error (2) if the point $(R, D_\mu^{H^1}(R))$ lies on the lower convex hull of the mapping $D_\mu^{H^1}(\cdot)$, which is not the case in general (cf. [11, 16]).*

Remark 4.5. *Assume that l is of class (D2), i.e. equals the identity, and $\alpha = 1$. Moreover, assume that μ is absolutely continuous with respect to λ^d and has a compact support. In this special case, Matloub et.al. [22] have shown for every $\varepsilon > 0$ that*

$$D_\mu^{H^1}(R+\varepsilon) \leq V_\mu^H(R) \leq D_\mu^{H^1}(R). \tag{21}$$

This result follows from Theorem 3.2. If $\alpha = 0$ and $l(x) = x^r$ with $r \geq 1$, then it is well-known (see e.g. [8, Lemma 3.3, Lemma 3.4], [24]) that the equations (15) and (16) are true.

Remark 4.6. *If $\alpha = 1$ and l is of type (D2), then we could immediately deduce equation (15) from relation (21) if the mapping $D_\mu^{H^1}(\cdot)$ would be continuous. Although Theorem 3.2 is true for every $\alpha \in [-\infty, \infty]$, the mapping $D_\mu^{H^\alpha}(\cdot)$ is generally non-continuous for $\alpha \leq 0$ (see e.g. [18, Lemma 7.2], [8, Example 5.5]). In case of $\alpha > 0$, the mapping $D_{U([0,1])}^{H^\alpha}(\cdot)$ has been completely determined and is Lipschitz continuous if $U([0,1])$ denotes the uniform distribution on $[0,1]$ (cf. [16]). It remains open if $D_\mu^{H^\alpha}(\cdot)$ is Lipschitz continuous for $\alpha > 0$ and distributions μ which are vanishing on continuously differentiable $(d-1)$-dimensional submanifolds of $\mathbb{R}^d$.*

Remark 4.7. *In view of (8), optimal codecells for finite quantizers are no polytopes in general if $l(x) \neq x^2$. For illustrations of such codecells, the reader is referred to [27, chapter 1.2], [6, Fig. 2] and [2]. For a detailed study of the topological properties of the codecells defined by (8), the author recommends [5]. If $l(x) = x^2$, then the polytopes defined in (8) are often called Laguerre tesselations in the literature (cf. [10]).*

A. Appendix

Recall 1_A as the characteristic function on a set $A \subset \mathbb{R}^d$.

Proof of Proposition 2.2.
Obviously, we can assume w.l.o.g. that $D_\mu^H(R) < \infty$. Let $\varepsilon > 0$. According to relation (2), let $q \in \mathcal{Q}_d$ with $H_\mu(q) \le R$ and $D_\mu^H(R) \ge D_\mu(q) - \varepsilon$. Denote $q(\mathbb{R}^d) = \{a_i : i \in I\}$ with a countable set $I \subset \mathbb{N}$ and points $a_i \in \mathbb{R}^d$, such that $a_i \ne a_j$ for every $i, j \in I$ with $i \ne j$. Because $\int l(\|x\|)d\mu(x) < \infty$ and due to

$$\infty > D_\mu(q) = \sum_{i \in I} \int_{q^{-1}(a_i)} l(\|x - a_i\|)d\mu(x),$$

a finite set $J \subset I$ exists, such that

$$\int_{\cup_{i \in I \setminus J} q^{-1}(a_i)} l(\|x - a_i\|)d\mu(x) < \varepsilon \tag{22}$$

and

$$\int_{\cup_{i \in I \setminus J} q^{-1}(a_i)} l(\|x\|)d\mu(x) < \varepsilon. \tag{23}$$

Now we define the quantizer

$$q_J = \sum_{j \in J} a_j 1_{q^{-1}(a_j)} + 0 \cdot 1_{\cup_{i \in I \setminus J} q^{-1}(a_i)}.$$

Let $M < \infty$ be the cardinality of J and $\{j_1, .., j_M\}$ be an enumeration of J. Let us first assume that $0 \in \{a_j : j \in J\}$, i.e. a $k \in \{1, .., M\}$ exists, such that $a_{j_k} = 0$. Applying property (a) and (b) of H from definition 2.1, we obtain

$$\begin{aligned} H_\mu(q_J) &= H(\mu(q^{-1}(a_{j_1})), .., \mu(q^{-1}(a_{j_{k-1}})), \mu(\cup_{i \in \{j_k\} \cup I \setminus J} q^{-1}(a_i)), \\ &\qquad \mu(q^{-1}(a_{j_{k+1}})), .., \mu(q^{-1}(a_{j_M})), 0, ..) \\ &\le H(\mu(q^{-1}(a_{j_1})), .., \mu(q^{-1}(a_{j_M})), \mu(\cup_{i \in I \setminus J} q^{-1}(a_i)), 0, ..). \end{aligned} \tag{24}$$

If $0 \notin \{a_j : j \in J\}$, then (24) turns into an equation. Again, by property (a) and (b) of H from definition 2.1, we deduce in any case that

$$\begin{aligned} H_\mu(q_J) &\le H(\mu(q^{-1}(a_{j_1})), .., \mu(q^{-1}(a_{j_M})), \mu(\cup_{i \in I \setminus J} q^{-1}(a_i)), 0, ..) \\ &\le H_\mu(q) \le R. \end{aligned}$$

Now we derive from (22) and (23) that

$$\begin{aligned}
D_\mu^H(R) &\geq D_\mu(q) - \varepsilon \\
&\geq D_\mu(q_J) - |D_\mu(q) - D_\mu(q_J)| - \varepsilon \\
&\geq D_\mu(q_J) - \Big|\int_{\cup_{i\in I\setminus J} q^{-1}(a_i)} (l(\|x - q(x)\|) - l(\|x\|))d\mu(x)\Big| - \varepsilon \\
&\geq D_\mu(q_J) - 3\varepsilon.
\end{aligned}$$

By letting $\varepsilon \to 0$ we get

$$\begin{aligned}
D_\mu^H(R) &\geq \inf\{D_\mu(q) : q \in \mathcal{Q}_d^*, H_\mu(q) \leq R\} \\
&\geq \inf\{D_\mu(q) : q \in \mathcal{Q}_d, H_\mu(q) \leq R\} = D_\mu^H(R),
\end{aligned}$$

which yields the assertion. □

Proof of Theorem 2.6.
From definition (5) we obtain

$$w(\mu, \nu) = \inf\{\int_{\mathbb{R}^d\times\mathbb{R}^d} l(\|x - y\|)d\gamma(x, y) : \gamma \in \Gamma(\mu, \nu)\}.$$

For every $\gamma \in \Gamma(\mu, \nu)$, the second marginal of γ equals ν and, therefore, has support $\{a_1, .., a_m\}$. Let $\tilde{\nu}$ be the restriction of ν to $\{a_1, .., a_m\}$. Recall (cf. [31, Thm. 4.1]) that an optimal solution always exists, i.e. choose $\tilde{\gamma} \in \Gamma(\mu, \nu)$, such that $w(\mu, \nu) = \int_{\mathbb{R}^d\times\mathbb{R}^d} l(\|x - y\|)d\tilde{\gamma}(x, y)$. Because the support of $\tilde{\gamma}$ is concentrated on $\mathbb{R}^d \times \{a_1, .., a_m\}$ (see also [31, Thm. 5.19]), we deduce that

$$\begin{aligned}
w(\mu, \nu) &= \inf\{\int_{\mathbb{R}^d\times\{a_1,..,a_m\}} l(\|x - y\|)d\gamma(x, y) : \gamma \in \Gamma(\mu, \tilde{\nu})\} \\
&= \int_{\mathbb{R}^d\times\{a_1,..,a_m\}} l(\|x - y\|)d\gamma_0(x, y)
\end{aligned}$$

with γ_0 as the restriction of $\tilde{\gamma}$ to $\mathbb{R}^d \times \{a_1, .., a_m\}$. If we denote by $\tilde{c}$ the restriction of c to $\mathbb{R}^d \times \{a_1, .., a_m\}$, then γ_0 is also an optimal solution of the Kantorovich problem for source μ and target $\tilde{\nu}$ on $\mathbb{R}^d \times \{a_1, .., a_m\}$. Now let f be a $\tilde{c}-$convex mapping on $\mathbb{R}^d$. According to Definition 2.4, let $\beta_1, .., \beta_m \in \mathbb{R}$ such that

$$f(x) = \max\{\beta_i - l(\|x - a_i\|) : i \in \{1, .., m\}\} \quad \text{for every } x \in \mathbb{R}^d.$$

Next we will show that

$$A_i(\beta_1, .., \beta_m) = \{x \in \mathbb{R}^d : a_i \in \partial_{\tilde{c}} f(x)\}. \tag{25}$$

To this end, let $x \in A_i(\beta_1, .., \beta_m)$. Applying (8) we obtain for every $z \in \mathbb{R}^d$ that

$$\begin{aligned} f(x) &= \beta_i - l(\|x - a_i\|) \\ &= l(\|z - a_i\|) + \beta_i - l(\|z - a_i\|) - l(\|x - a_i\|) \\ &\leq l(\|z - a_i\|) + \max\{\beta_j - l(\|z - a_j\|) : j \in \{1, .., m\}\} - l(\|x - a_i\|) \\ &= \tilde{c}(z, a_i) + f(z) - \tilde{c}(x, a_i). \end{aligned}$$

Consequently, $a_i \in \partial_{\tilde{c}} f(x)$ according to Definition 2.4. Now let $x \in \mathbb{R}^d$, such that $a_i \in \partial_{\tilde{c}} f(x)$. Let $z \in \mathbb{R}^d$. Choose $i_x, i_z \in \{1, .., m\}$, such that

$$f(z) = \beta_{i_z} - l(\|z - a_{i_z}\|) \quad \text{and} \quad f(x) = \beta_{i_x} - l(\|x - a_{i_x}\|). \tag{26}$$

Using $a_i \in \partial_{\tilde{c}} f(x)$ we deduce

$$f(x) + l(\|x - a_i\|) \leq f(z) + l(\|z - a_i\|). \tag{27}$$

Combining (26) and (27) we obtain

$$l(\|x - a_i\|) - \beta_{i_z} \leq l(\|x - a_{i_x}\|) - \beta_{i_x} - l(\|z - a_{i_z}\|) + l(\|z - a_i\|). \tag{28}$$

Now specialize $z \in \mathbb{R}^d$, such that $i_z = i$. Because $A_i(\beta_1, .., \beta_m)$ is non-empty for every $i \in \{1, .., m\}$ such a choice for z is always possible. From (28) we get

$$\begin{aligned} l(\|x - a_i\|) - \beta_i &\leq l(\|x - a_{i_x}\|) - \beta_{i_x} \\ &= \min\{l(\|x - a_j\|) - \beta_j : j \in \{1, .., m\}\}. \end{aligned}$$

Thus, $x \in A_i(\beta_1, .., \beta_m)$ which proves (25). Now let $x \in \mathbb{R}^d$ with $\mathrm{card}(\partial_{\tilde{c}} f(x)) > 1$. In view of (25), we know that $i, j \in \{1, .., m\}$ exist with $i \neq j$ such that $x \in A_i(\beta_1, .., \beta_m) \cap A_j(\beta_1, .., \beta_m)$. Hence,

$$l(\|x - a_i\|) - \beta_i = l(\|x - a_j\|) - \beta_j$$

Because $a_i \neq a_j$, we will assume w.l.o.g. that $x \neq a_j$. According to Definition (8), we can find for every $\varepsilon > 0$ a point $z \in \mathbb{R}^d$ such that

$$\|z - x\| < \varepsilon, \quad \|x - a_j\| = \|z - a_j\| \text{ and } \|x - a_i\| < \|z - a_i\|.$$

Because l is strictly increasing, we obtain

$$l(\|z-a_i\|)-\beta_i > l(\|z-a_j\|)-\beta_j,$$

yielding that z lies in the complement of $A_i(\beta_1,..,\beta_m)$. This implies that x is an element of the boundary of $A_i(\beta_1,..,\beta_m)$, i.e. the set $\{x \in \mathbb{R}^d : \text{card}(\partial_{\tilde{c}} f(x)) > 1\}$ is contained in a $\mu-$measure zero set by our assumption. Now we can apply Theorem 2.5 which implies that γ_0 is deterministic. Let q_0 be the Monge map which induces γ_0 and let $\tilde{f}$ be the $\tilde{c}-$convex function, such that

$$q_0(x) \in \partial_{\tilde{c}} \tilde{f}(x) \text{ for } \mu-a.e. \quad x \in \mathbb{R}^d.$$

From above and Definition 2.4 we deduce that for $\mu-$almost every $x \in \mathbb{R}^d$ exactly one $i \in \{1,..,m\}$ exist, such that $a_i \in \partial_{\tilde{c}} \tilde{f}(x)$, i.e. $q_0(x) = a_i$. Now let $B_1,..,B_m$ be a partition of $\mathbb{R}^d$ with

$$B_i = \{x \in \mathbb{R}^d : a_i \in \partial_{\tilde{c}} \tilde{f}(x)\} \quad \mu-\text{almost surely}.$$

According to Definition 2.4 and relation (25), we obtain that $(\lambda_1,..,\lambda_m) \in \mathbb{R}^m$ exist such that for every $i \in \{1,..,m\}$ the set B_i equals $\mu-$almost surely the set $A_i(\lambda_1,..,\lambda_m)$. Consequently, we can assume w.l.o.g. that B_i is measurable for every $i \in \{1,..,m\}$. Now we define the quantizer q with $q(x) = a_i$ if $x \in B_i$. Because $q(x) = q_0(x)$ for $\mu-$almost every $x \in \mathbb{R}^d$ and q_0 is a Monge map for $\gamma_0 \in \Gamma(\mu, \tilde{\nu})$ we obtain that $\mu \circ q^{-1} = \nu$, which yields $\mu \circ q^{-1}(a_i) = \nu(\{a_i\}) = p_i$ for every $i \in \{1,..,m\}$. We get

$$w(\mu,\nu) = \int l(\|x-q_0(x)\|)d\mu(x) = \int l(\|x-q(x)\|)d\mu(x).$$

Now let $\tilde{q} \in \mathcal{Q}(\nu)$. Because $\tilde{q}$ induces a transport plan $\pi \in \Gamma(\mu,\tilde{\nu})$, we have $w(\mu,\nu) \leq \int l(\|x-\tilde{q}(x)\|)d\mu(x)$. If $w(\mu,\nu) = \int l(\|x-\tilde{q}(x)\|)d\mu(x)$, then $\pi = \gamma_0$ according to Theorem 2.5. Thus, we obtain again from above considerations that $\tilde{q}(x) = q(x)$ for μ-almost every $x \in \mathbb{R}^d$, which finally proves the assertion. $\square$

We denote with ∇F the gradient of a differentiable mapping $F : \mathbb{R}^d \to \mathbb{R}$.

Lemma A.1. *Let $\mu \in \mathcal{M}(\mathbb{R}^d)$ and assume that μ vanishes on continuously differentiable $(d-1)$-dimensional submanifolds of $\mathbb{R}^d$. Let $m \in \mathbb{N}$ and $a_1,..,a_m \in \mathbb{R}^d$ be m different points in $\mathbb{R}^d$. Let $\lambda_1,..,\lambda_m \in \mathbb{R}$ and $A_j = A_j(\lambda_1,..,\lambda_m)$ as defined in (8). If l is of type (C1) or (C2), then* *the boundary of A_j has zero $\mu-$measure. Additionally, if $d = 1$ and l is of type (C1), then A_j is an interval.*

If $d > 1$ and $l(x) = x^2$ for every $x \geq 0$, then A_j is a convex polytope.

Proof. Let $j \in \{1, .., m\}$. Obviously, we can write

$$A_j = \cap_{i=1, i\neq j}^{m} \{x \in \mathbb{R}^d : l(\|x - a_j\|) - l(\|x - a_i\|) \leq \lambda_j - \lambda_i\} \tag{29}$$

For every $i \in \{1, .., m\} \backslash \{j\}$ we define

$$G_{j,i} = \{a_j + \lambda(a_j - a_i) : \lambda \in \mathbb{R}\}$$

and for every $x \in \mathbb{R}^d$ let

$$\Psi_{j,i}(x) = l(\|x - a_j\|) - l(\|x - a_i\|).$$

Let $G_j = \cup_{i=1, i\neq j}^{m} G_{j,i}$. In order to show that the boundary of A_j has zero $\mu-$measure, we distinguish several cases.

1. $d = 1$.

1.1. l is of type (C1).

We proceed as in the proof of [12, Lemma 1]. Let $i \in \{1, .., m\} \backslash \{j\}$ and assume w.l.o.g. that $a_i > a_j$. If $x < a_j$, then

$$\Psi'_{j,i}(x) = l'(a_i - x) - l'(a_j - x).$$

Due to $l'' \geq 0$ we obtain that $\Psi_{j,i}$ is monotone increasing on $(-\infty, a_j)$. If $x > a_i$, then we deduce by similar considerations that $\Psi'_{j,i}(x) \geq 0$, i.e. $\Psi_{j,i}$ is monotone increasing on (a_i, ∞). If $a_j < x < a_i$, then we obtain that

$$\Psi'_{j,i}(x) = l'(a_i - x) + l'(x - a_j).$$

Due to $l' > 0$ we get that $\Psi'_{j,i}$ is strictly increasing on (a_j, a_i). Obviously, $\Psi_{j,i}$ is continuous on $\mathbb{R}$. Hence, $\Psi_{j,i}^{-1}((-\infty, \lambda_j - \lambda_i])$ is an (possibly empty or degenerate) interval. Due to (29), A_j is a finite intersection of intervals, yielding that the boundary of A_j consists of a finite set. Because μ is non-atomic, the assertion is proved in this case.

1.2. l is of type (C2).

Due to $l'' \leq 0$, we obtain similar to above that $\Psi_{j,i}$ is monotone decreasing on $(-\infty, a_j)$ and (a_i, ∞). Moreover, $\Psi_{j,i}$ is strictly increasing on (a_j, a_i) and continuous on $\mathbb{R}$. Note also that $\Psi_{j,i}$ is negative on $(-\infty, a_j)$ and positive on (a_i, ∞). Hence, $\Psi_{j,i}^{-1}((-\infty, \lambda_j - \lambda_i])$ consists of at most two intervals. Hence, the assertion is proved as in case 1.2.

2. $d > 1$.

Let $i \in \{1, .., m\} \backslash \{j\}$. Let $x \in \mathbb{R}^d \backslash G_j$. The mapping $\Psi_{j,i}$ is differentiable on $\mathbb{R}^d \backslash G_j$ and we obtain

$$\nabla \Psi_{j,i}(x) = l'(\|x - a_j\|) \frac{x - a_j}{\|x - a_j\|} - l'(\|x - a_i\|) \frac{x - a_i}{\|x - a_i\|}.$$

Because $x \notin G_j$, we know that $x - a_j$ and $x - a_i$ are linearly independent and $\min(\|x - a_j\|, \|x - a_i\|) > 0$. Because l' has no zeros on $(0, \infty)$, we obtain $\nabla \Psi_{j,i}(x) \neq 0$. If $\lambda_j - \lambda_i \in \Psi_{j,i}(\mathbb{R}^d \backslash G_j)$, then the submersion theorem yields that $\Psi_{j,i}^{-1}(\lambda_j - \lambda_i) \backslash G_j$ is covered by a continuously differentiable $(d-1)$-dimensional submanifold of $\mathbb{R}^d \backslash G_j$. Because G_j as a finite union of lines can be covered by a finite union of $(d-1)$-dimensional hyperplanes, we finally obtain that $\Psi_{j,i}^{-1}(\lambda_j - \lambda_i)$ is always covered by a finite union of continuously differentiable $(d-1)$-dimensional submanifolds of $\mathbb{R}^d$. Note that the boundary of A_j is contained in $\cup_{i=1, i \neq j}^m \Psi_{j,i}^{-1}(\lambda_j - \lambda_i)$, which yields the assertion also in this case.

Now let $d = 1$ and assume that l is of type (C1). From 1.1. we obtain that A_j is an interval. Finally assume that $d > 1$ and $l(x) = x^2$ for every $x \geq 0$. For any $i \in \{1, .., m\} \backslash \{j\}$ let $\lambda = (\lambda_j - \lambda_i)/(2\|a_i - a_j\|^2)$ and

$$b_i = a_i + \lambda(a_i - a_j), \qquad b_j = a_j + \lambda(a_i - a_j).$$

Now recall the definition (13) of a closed halfspace. A simple calculation shows that

$$A_j = \cap_{i=1, i \neq j}^m T(b_j, b_i)$$

is a finite intersection of closed halfspaces and, therefore, a convex polytope. □

Lemma A.2. *If the distance mapping l is of type (D1) or (D2), then the triangle inequality holds.*

Proof. Let $\mu_i \in \mathcal{M}(\mathbb{R}^d)$, $i \in \{1, 2, 3\}$ and $w(\mu_1, \mu_3) < \infty$. We have to show that

$$l^{-1}(w(\mu_1, \mu_3)) \leq l^{-1}(w(\mu_1, \mu_2)) + l^{-1}(w(\mu_2, \mu_3)). \tag{30}$$

We are following the standard argumentation for ρ_r to satisfy the triangle inequality (see e.g. [31, p.94]) or [7, Proposition 2]) and apply a generalization of the Minkowski inequality (cf. [21, Thm. 3]). Obviously, we can assume w.l.o.g. that $w(\mu_1, \mu_2) < \infty$ and $w(\mu_2, \mu_3) < \infty$. If l is of type (D2), then the assertion follows immediately from the triangle inequality of the Wasserstein distance. Hence, let us assume that l is of type (D1). Because l is continuous, we deduce from general existence results in optimal transportation theory (cf. [31, Thm. 4.1]) that distributions P_1 and P_2 on $\mathbb{R}^d \times \mathbb{R}^d$ exist, such that

$$w(\mu_1, \mu_2) = \int_{\mathbb{R}^d \times \mathbb{R}^d} l(\|x - y\|) dP_1(x, y) \tag{31}$$

and

$$w(\mu_2, \mu_3) = \int_{\mathbb{R}^d \times \mathbb{R}^d} l(\|x - y\|) dP_2(x, y), \tag{32}$$

where P_1 has marginals μ_1 and μ_2 and P_2 has marginals μ_2 and μ_3. Moreover, let P be a distribution on $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$ with marginal distribution P_1 when projecting on the first two components and P_2 when projecting on the last two components. Regarding the existence of such a distribution, see e.g. [4, Remark 5.3.3] or [31, chapter 1]. Denote P_3 as the marginal of P by projecting on the first and the last component. The first marginal of P_3 equals μ_1 and the second marginal of P_3 equals μ_3. Because l and l^{-1} are increasing we deduce

$$\begin{aligned} l^{-1}(w(\mu_1, \mu_3)) &\leq l^{-1}\left(\int_{\mathbb{R}^d \times \mathbb{R}^d} l(\|x - y\|) dP_3(x, y)\right) \\ &= l^{-1}\left(\int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} l(\|x - z\|) dP(x, y, z)\right) \\ &\leq l^{-1}\left(\int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} l(\|x - y\| + \|y - z\|) dP(x, y, z)\right). \end{aligned}$$

Now let $(u_n)_{n \in \mathbb{N}}$ and $(v_n)_{n \in \mathbb{N}}$ be sequences of step functions on $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$ with $u_n \leq u_{n+1}$, $v_n \leq v_{n+1}$ and

$$\sup_{n \in \mathbb{N}} u_n(x, y, z) = \|x - y\|, \qquad \sup_{n \in \mathbb{N}} v_n(x, y, z) = \|y - z\|.$$

Note that l is continuous, monotone increasing and l^{-1} is continuous. Thus, by monotone convergence we obtain

$$l^{-1}(w(\mu_1, \mu_3)) \leq \lim_{n \to \infty} l^{-1}\left(\int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} l(u_n(x, y, z) + v_n(x, y, z)) dP(x, y, z)\right).$$

Now we distinguish two cases. Let us first assume that $\mu_i(A) \in \{0,1\}$ for every measurable $A \subset \mathbb{R}^d$ and $i \in \{1,2,3\}$. Consequently, $a_i \in \mathbb{R}^d$ exist, such that $\mu_i = \delta_{a_i}$. But then relation (30) follows immediately from the triangle inequality. Hence, we can assume that a measurable set $A \subset \mathbb{R}^d$ and $j \in \{1,2,3\}$ exists with $\mu_j(A) \in (0,1)$. Let $B = \prod_{i\in\{1,2,3\}} A_i$ with $A_j = A$ and $A_i = \mathbb{R}^d$ for every $i \neq j$. Hence, $P(B) = \mu_j(A) \in (0,1)$. Due to the assumptions on l we can apply a generalized version of the Minkowski inequality [21, Thm. 3] and deduce

$$
\begin{aligned}
& l^{-1}(w(\mu_1,\mu_3)) \\
\leq\ & \lim_{n\to\infty} l^{-1}\left(\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} l(u_n(x,y,z)+v_n(x,y,z))dP(x,y,z)\right) \\
\leq\ & \lim_{n\to\infty}\left(l^{-1}\left(\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} l(u_n(x,y,z))dP(x,y,z)\right)\right. \\
& \qquad\qquad \left. + l^{-1}\left(\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} l(v_n(x,y,z))dP(x,y,z)\right)\right) \\
=\ & l^{-1}\left(\int_{\mathbb{R}^d\times\mathbb{R}^d} l(\|x-y\|)dP_1(x,y)\right) + l^{-1}\left(\int_{\mathbb{R}^d\times\mathbb{R}^d} l(\|y-z\|)dP_2(y,z)\right)
\end{aligned}
$$

Applying the identities (31) and (32), we obtain inequality (30). □

Proof of Proposition 3.1. Obviously, we can assume w.l.o.g. that $V^H_\mu(R) < \infty$. Let $\nu \in \mathcal{M}^\infty(\mathbb{R}^d)\backslash\mathcal{M}^*(\mathbb{R}^d)$ with $H(p^\nu) \leq R$ and $w(\mu,\nu) < \infty$. Let $\{a_1, a_2, ..\} = \mathrm{supp}(\nu)$ and define for every $k \geq 2$ the distribution

$$\nu_k = \sum_{i=1}^{k-1} \nu(a_i)\delta_{a_i} + \sum_{i\geq k} \nu(a_i)\delta_0.$$

Property (b) of the mapping H implies

$$H(p^{\nu_k}) \leq H(p^\nu) \leq R.$$

1. Let us first assume that l is continuous and $l^{-1}\circ w$ satisfies the triangle inequality.
We define the transport mapping

$$q_k = \sum_{i=1}^{k-1} a_i 1_{\{a_i\}} + 0 \cdot 1_{\mathbb{R}^d\backslash\{a_1,..,a_{k-1}\}}.$$

with $\nu \circ q_k^{-1} = \nu_k$. Denote $\pi_k \in \Gamma(\nu, \nu_k)$ as the (deterministic) transport plan which is induced by q_k. Thus, we get

$$w(\nu, \nu_k) \leq \int l(\|x-y\|) d\pi_k(x,y) = \sum_{i \geq k} \nu(a_i) l(\|a_i\|). \tag{33}$$

Because $\nu \in \mathcal{M}(\mathbb{R}^d)$ we obtain $\int l(\|x\|) d\nu(x) = \sum_{i=1}^{\infty} \nu(a_i) l(\|a_i\|) < \infty$ which together with (33) yields that $w(\nu, \nu_k) \to 0$ as $k \to \infty$. Using the triangle inequality for $l^{-1} \circ w$ we conclude that

$$\infty > l^{-1}(w(\mu, \nu)) \geq l^{-1}(w(\mu, \nu_k)) - l^{-1}(w(\nu_k, \nu)).$$

Letting $k \to \infty$, we obtain from the continuity of l and l^{-1} that

$$w(\mu, \nu) \geq \liminf_{k \to \infty} w(\mu, \nu_k).$$

This implies

$$V_\mu^H(R) \geq \inf\{w(\mu, \kappa) : \kappa \in \mathcal{M}^*(\mathbb{R}^d), H(p^\kappa) \leq R\} \tag{34}$$

and yields the assertion in this first case.
2. Now we assume that l is bounded.
Note that ν_k weakly converges to ν. Moreover,

$$\liminf_{k \to \infty} w(\mu, \nu_k) \leq \sup_{x \in [0,\infty)} l(x) < \infty.$$

Denote by π_k an optimal transport plan for source μ and target ν_k, i.e.

$$w(\mu, \nu_k) = \int l(\|x-y\|) d\pi_k(x,y).$$

Applying a stability result for optimal transport plans (cf. [31, Thm. 5.20]) we obtain a subsequence of (π_k), also denote by (π_k) such that (π_k) weakly converges to an optimal transport plan π for source μ and target ν, i.e.

$$w(\mu, \nu) = \int_{\mathbb{R}^d \times \mathbb{R}^d} l(\|x-y\|) d\pi(x,y).$$

Because l is bounded, weak convergence implies

$$\begin{aligned} w(\mu, \nu_k) &= \int l(\|x-y\|) d\pi_k(x,y) \\ &\quad \to \int l(\|x-y\|) d\pi(x,y) = w(\mu, \nu), \text{ as } k \to \infty. \end{aligned}$$

As an immediate consequence we obtain

$$w(\mu,\nu) \geq \inf\{w(\mu,\kappa) : \kappa \in \mathcal{M}^*(\mathbb{R}^d), H(p^\kappa) \leq R\},$$

which yields (34) and, hence, proves the assertion also in this second case. $\square$

For any set $A \subset \mathbb{R}^d$ we denote $\overset{\circ}{A}$ as the interior of A. Recall 1_A as the characteristic function on A. Now we can prove the main result of this paper.

Proof of Theorem 3.2. We will proceed in several steps.
1. We will prove that $W_\mu^H(R) \leq D_\mu^H(R)$.
Let $q \in \mathcal{Q}_d^*$ with $H_\mu(q) \leq R$. With the measurable mapping

$$\mathbb{R}^d \ni x \mapsto \phi(x) = (x, q(x)) \in \mathbb{R}^d \times \mathbb{R}^d$$

and in view of (5), we obtain

$$\begin{aligned} D_\mu(q) &= \int_{\mathbb{R}^d} l(\|x - q(x)\|) d\mu(x) \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^d} l(\|x - y\|) d\mu \circ \phi^{-1}(x,y) \geq w(\mu, \mu \circ q^{-1}). \end{aligned}$$

Hence, relation (11) and Proposition 2.2 imply that $W_\mu^H(R) \leq D_\mu^H(R)$.
2. We will prove that $V_\mu^H(R) \geq D_\mu^H(R)$.
Let us assume w.l.o.g. that $V_\mu^H(R) < \infty$. Let $m \in \mathbb{N}$ and $\nu \in \mathcal{M}^*(\mathbb{R}^d)$ with $\mathrm{card}(\mathrm{supp}(\nu)) = m$, $H(p^\nu) \leq R$ and $w(\mu,\nu) < \infty$. Let us denote $\mathrm{supp}(\nu) = \{a_1, .., a_m\}$. For any shifts $\beta_1, .., \beta_m \in \mathbb{R}$ recall the definition (8) of the set $A_j(\beta_1, .., \beta_m)$. We obtain from Lemma A.1 that the boundary of $A_j(\beta_1, .., \beta_m)$ has zero $\mu-$measure. Thus, we deduce from Theorem 2.6 that a quantizer q and uniquely determined constants $\lambda_1, .., \lambda_m \in \mathbb{R}$ exist, such that $q(\mathbb{R}^d) = \{a_1, .., a_m\}$ and

$$w(\mu,\nu) = D_\mu(q) = \sum_{j=1}^m \int_{A_j(\lambda_1,..,\lambda_m)} l(\|x - a_j\|) d\mu(x), \tag{35}$$

where

$$\mu(q^{-1}(a_j)) = \mu(A_j(\lambda_1, .., \lambda_m)) = \nu(a_j) > 0 \text{ for every } j \in \{1, .., m\}.$$

Note that

$$H_\mu(q) = H(p^\nu) \le R.$$

Using (35) we obtain

$$w(\mu,\nu) = D_\mu(q) \ge D_\mu^H(R).$$

Because l satisfies (C1) or (C2), the assertion of step 2 follows from Proposition 3.1.

3. Rest of the proof.

Combining (12) with step 1 and step 2, we obtain

$$V_\mu^H(R) \le W_\mu^H(R) \le D_\mu^H(R) \le V_\mu^H(R)$$

which proves equation (15). Now, additionally, assume that condition (a) or (b) in our assumptions of this theorem is satisfied. In view of (14) and (15) it suffices to prove that $V_\mu^H(R) \ge D_\mu^{H,c}(R)$. Again, we can assume w.l.o.g. that $V_\mu^H(R) < \infty$. Let $\nu \in \mathcal{M}^*(\mathbb{R}^d)$ and q as in step 2. Due to Lemma A.1, the source distribution μ vanishes on the boundary of $A_j(\lambda_1,..,\lambda_m)$. Thus, we can assume w.l.o.g. that

$$\mathring{A}_j(\lambda_1,..,\lambda_m) \subset q^{-1}(a_j) \subset A_j(\lambda_1,..,\lambda_m), \qquad j \in \{1,..,m\},$$

and that $q^{-1}(a_j)$ is either an interval ($d=1$) or a convex polytope ($d>1$). If $d=1$, then we can clearly assume w.l.o.g. that a_j is contained in A_j. Moreover, if $d>1$ let us assume w.l.o.g. according to [8, Example 2.3(b)] that a_j is the centroid and therefore optimal for μ restricted to A_j, i.e.

$$\int_{A_j} \|x-a_j\|^2 d\mu(x) = \inf\{\int_{A_j} \|x-b\|^2 d\mu(x) : b \in \mathbb{R}^d\}\}.$$

Because we are operating with the Euclidean norm, we obtain by exactly the same argument as in the proof of [8, Lemma 2.6 (a)] that a_j is contained in A_j, which is the closure of $q^{-1}(a_j)$. By (35) we get

$$w(\mu,\nu) = D_\mu(q) \ge D_\mu^{H,c}(R).$$

Because l is of type (C1) or (C2), Proposition 3.1 yields $V_\mu^H(R) \ge D_\mu^{H,c}(R)$, which finishes the proof. □

Acknowledgements

The author would like to thank the two anonymous referees for valuable comments leading to an improved version of this paper. The research was supported by a grant from the German Research Foundation (DFG).

References

[1] E.F. Abaya and G.L. Wise, Convergence of vector quantizers with applications to optimal quantization, SIAM J. Appl. Math. **44** (1984), pp. 183-189.

[2] T. Abdellaoui, Optimal solution of a Monge-Kantorovitch transportation problem, J. Comput. Appl. Math. **96** (1998), pp. 149-161.

[3] J. Aczél and Z. Daróczy, On Measures of Information and Their Characterizations, Mathematics in Science and Engineering, Academic Press, London (1975).

[4] L. Ambrosio, N. Gigli and G. Savaré, Gradient flows in metric spaces and in the space of probability measures, 2nd ed., Lectures in Mathematics, Birkhäuser, Basel (2008).

[5] P.F. Ash and E.D. Bolker, Generalized Dirichlet tesselations, Geom. Dedicata **20** (1986), pp. 209-243.

[6] W. Gangbo and R. J. McCann, The geometry of optimal transportation, Acta Math. **177** (1996), pp. 113–161.

[7] C.R. Givens and R.M. Shortt, A class of Wasserstein metrics for probability distributions, Mich. Math. J. **31** (1984), pp. 231-240.

[8] S. Graf and H. Luschgy, Foundations of Quantization for Probability Distributions, Lecture Notes 1730 Springer, Berlin (2000).

[9] R. Gray and D. Neuhoff, Quantization, IEEE Trans. Inf. Theory **44** (1998), pp. 2325-2383.

[10] C. Lautensack and S. Zuyev, Random Laguerre tessellations, Adv. in Appl. Probab. **40** (2008), pp. 630-650.

[11] A. György and T. Linder, Optimal entropy-constrained scalar quantization of a uniform source, IEEE Trans. Inf. Theory **46** (2000), pp. 2704-2711.

[12] A. György and T. Linder, On the structure of optimal entropy-constrained scalar quantizers, IEEE Trans. Inf. Theory **48** (2002), pp. 416-427.

[13] A. György and T. Linder, Codecell convexity in optimal entropy-constrained vector quantization, IEEE Trans. Inf. Theory **49** (2003), pp. 1821-1828.

[14] P. Harremoës, Joint Range of Rényi Entropies, Kybernetika **45** (2009), pp. 901-911.

[15] L. Kantorovitch, On the translocation of masses, C. R. (Dokl.) Acad. Sci. URSS, n. Ser. **37** (1942), pp. 199-201.

[16] W. Kreitmeier, Optimal Quantization for the one-dimensional uniform distribution with Rényi-α-entropy constraints, Kybernetika **46** (2010), pp. 96-113.

[17] W. Kreitmeier, Error bounds for high-resolution quantization with Rényi-α-entropy constraints, Acta Math. Hungar. **127** (2010), pp. 34-51.

[18] W. Kreitmeier and T. Linder, High-resolution scalar quantization with Rényi entropy constraint, preprint, arXiv: 1008.1744 (2010).

[19] Y. Linde, A. Buzo and R.M. Gray, An algorithm for vector quantizer design, IEEE Trans. Comm. **28** (1980), pp. 84–95.

[20] T. Linder, Lagrangian empirical design of variable-rate vector quantizers: consistency and convergence rates, IEEE Trans. Inf. Theory **48** (2002), pp. 2998-3003.

[21] J. Matkowski, The converse of the Minkowski's inequality theorem and its generalization, Proc. Am. Math. Soc. **109** (1990), pp. 663-675.

[22] S. Matloub, D. O'Brien and R. Gray, Optimal quantizer performance and the Wasserstein distortion, Proceedings of the 2005 IEEE Data Compression Conference, 29-31 March 2005, pp. 243-250.

[23] P. Mattila, On the structure of self-similar fractals, Ann. Acad. Sci. Fenn., Ser. A **7** (1982), pp. 189-195.

[24] D. Pollard, Quantization and the method of $k-$means, IEEE Trans. Inf. Theory **28** (1982), pp. 199-205.

[25] S.T. Rachev, Probability metrics and the stability of stochastic models, Wiley Series in Probability and Statistics, Wiley, Chichester (1991).

[26] A. Rényi, On measures of entropy and information, Proc. 4th Berkeley Symp. Math. Stat. Probab. **1** (1960), pp. 547-561.

[27] L. Rüschendorf and L. Uckelmann, On optimal multivariate couplings, Proceedings of the 1996 conference, Distribution with given marginals and moment problems, Prague, Czech Republic, pp. 261–273. Kluwer (1996).

[28] L. Rüschendorf, Monge-Kantorovich transportation problem and optimal couplings, Jahresber. Dtsch. Math.-Ver. **109** (2007), pp. 113-137.

[29] M.P. Schützenberger, Contribution aux applications statistiques de la théorie de l'information, Publ. Inst. Stat. Univ. Paris **3** (1954), pp. 1-114.

[30] H. Steinhaus, Sur la division des corps matériels en parties, Bull. Acad. Polon. Sci. Cl. III **4** (1957), pp. 801-804.

[31] C. Villani, Optimal transport. Old and new, Grundlehren der Mathematischen Wissenschaften, Springer, Berlin (2009).