Information and Inference: A Journal of the IMA (2022) **00**, 1–66
https://doi.org/DOI HERE

Lossy Compression of General Random Variables

ERWIN RIEGLER*
Chair for Mathematical Information Science, ETH Zurich, Sternwartstrasse 7, 8092, Zurich, Switzerland

GÜNTHER KOLIANDER
Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, 1040, Vienna, Austria

AND

HELMUT BÖLCSKEI
Chair for Mathematical Information Science, ETH Zurich, Sternwartstrasse 7, 8092, Zurich, Switzerland
*Corresponding author: eriegler@mins.ee.ethz.ch

[Received on 14 March 2022; accepted on 06 September 2022]

This paper is concerned with the lossy compression of general random variables, specifically with rate-distortion theory and quantization of random variables taking values in general measurable spaces such as, e.g., manifolds and fractal sets. Manifold structures are prevalent in data science, e.g., in compressed sensing, machine learning, image processing, and handwritten digit recognition. Fractal sets find application in image compression and in the modeling of Ethernet traffic. Our main contributions are bounds on the rate-distortion function and the quantization error. These bounds are very general and essentially only require the existence of reference measures satisfying certain regularity conditions in terms of small ball probabilities. To illustrate the wide applicability of our results, we particularize them to random variables taking values in i) manifolds, namely, hyperspheres and Grassmannians, and ii) self-similar sets characterized by iterated function systems satisfying the weak separation property.

Keywords: Lossy compression; rate-distortion theory; quantization; manifolds; fractal sets; information theory; directional statistics.
2020 Math Subject Classification: Primary 94A34, 94A29; secondary 94A15, 60B05, 62H11.

1. Introduction

This paper is concerned with the lossy compression of general random variables, specifically with rate-distortion (R-D) theory and quantization of random variables taking values in general measurable spaces such as, e.g., manifolds and fractal sets. Manifold structures are prevalent in data science, e.g., in compressed sensing [1, 4, 8, 9, 11, 55], machine learning [21, 46], image processing [45, 58], directional statistics [48], and handwritten digit recognition [31]. Fractal sets find application in image compression and in modeling of Ethernet traffic [42].

In the following, we formally introduce the mathematical setup underlying R-D theory and quantization, review the corresponding relevant results in the literature, and summarize our main contributions. Quantities not defined directly are introduced in the notation paragraph at the end of this section.

1.1. *Rate-Distortion Theory*

In R-D theory [6, 15, 28, 29, 57], one is interested in the characterization of the ultimate limits on the compression of sequences of random variables under a distortion constraint, here expressed in terms of expected average distortion. Specifically, let $(\mathcal{A},\mathscr{A})$ and $(\mathcal{B},\mathscr{B})$ be measurable spaces equipped with a measurable function $\sigma\colon \mathcal{A}\times\mathcal{B}\to[0,\infty]$, henceforth referred to as distortion function,[1] and let $(X_i)_{i\in\mathbb{N}}$ be a sequence of random variables taking values in $(\mathcal{A},\mathscr{A})$. For every $\ell\in\mathbb{N}$, measurable mappings $g_{(\ell)}\colon \mathcal{A}^\ell\to\mathcal{B}^\ell$ with $|g_{(\ell)}(\mathcal{A}^\ell)|<\infty$ are referred to as source codes of length ℓ. For a given rate $R\in[0,\infty)$ and distortion $D\in[0,\infty)$, the pair (R,D) is said to be ℓ-achievable if there exists a source code $g_{(\ell)}$ of length ℓ with $\left|g_{(\ell)}(\mathcal{A}^\ell)\right|\le\lfloor e^{\ell R}\rfloor$ and expected average distortion

$$\mathbb{E}\left[\sigma_{(\ell)}\Big((X_1,\dots,X_\ell),g_{(\ell)}(X_1,\dots,X_\ell)\Big)\right]\le D, \tag{1.1}$$

where

$$\sigma_{(\ell)}((x_1,\dots,x_\ell),(y_1,\dots,y_\ell))=\frac{1}{\ell}\sum_{i=1}^{\ell}\sigma(x_i,y_i) \tag{1.2}$$

is the average distortion function of length ℓ. Moreover, the pair (R,D) is said to be achievable if there exists an $\ell\in\mathbb{N}$ such that (R,D) is ℓ-achievable. For every $\ell\in\mathbb{N}$ and $D\in[0,\infty)$, we set[2]

$$R_{(\ell)}(D)=\inf\{R\in[0,\infty):(R,D)\text{ is }\ell\text{-achievable}\} \tag{1.3}$$

and

$$R(D)=\inf\{R\in[0,\infty):(R,D)\text{ is achievable}\}. \tag{1.4}$$

The set of achievable (R,D)-pairs can be characterized as follows. For general measurable spaces $(\mathcal{X},\mathscr{X})$ and $(\mathcal{Y},\mathscr{Y})$ equipped with a general distortion function $\rho\colon\mathcal{X}\times\mathcal{Y}\to[0,\infty]$ and the random variable X taking values in $(\mathcal{X},\mathscr{X})$, the R-D function is defined as

$$R_X(D)=\inf_{\mu_{Y|X}:\mathbb{E}[\rho(X,Y)]\le D} I(X;Y), \tag{1.5}$$

where $I(\cdot\,;\cdot)$ denotes the mutual information (defined in (1.28)) between X and Y with Y taking values in $(\mathcal{Y},\mathscr{Y})$ and the infimum is over all $\mu_{Y|X}\colon\mathscr{Y}\times\mathcal{X}\to[0,1]$ with (i) $\mu_{Y|X}(\cdot,x)$ a probability measure on $(\mathcal{Y},\mathscr{Y})$ for all $x\in\mathcal{X}$, (ii) $\mu_{Y|X}(\mathcal{C},\cdot)$ a measurable function for all $\mathcal{C}\in\mathscr{Y}$, and (iii)

$$\mathbb{E}[\rho(X,Y)]=\mathbb{E}\left[\int\rho(X,y)\,\mathrm{d}\mu_{Y|X}(y,X)\right]. \tag{1.6}$$

Now, let $(\mathcal{A},\mathscr{A})$, $(\mathcal{B},\mathscr{B})$, and $\sigma\colon\mathcal{A}\times\mathcal{B}\to[0,\infty]$ be as defined above and consider a stationary ergodic process $(X_i)_{i\in\mathbb{N}}$, where X_i takes values in $(\mathcal{A},\mathscr{A})$ for all $i\in\mathbb{N}$. For every $\ell\in\mathbb{N}$, define

---

[1] For $\mathcal{A}=\mathcal{B}=\mathbb{R}^d$, a distortion function $\sigma\colon\mathbb{R}^d\times\mathbb{R}^d\to[0,\infty]$ is called a difference distortion function if there exists a measurable function $f\colon\mathbb{R}^d\to[0,\infty]$ such that $\sigma(x,y)=f(x-y)$ for all $x,y\in\mathbb{R}^d$.

[2] We use the convention that the infimum of the empty set is ∞.

$R_{(X_1,\dots,X_\ell)}(D)$ as in (1.5) for $X=(X_1,\dots,X_\ell)$, $Y=(Y_1,\dots,Y_\ell)$, $(\mathcal{X},\mathscr{X})=(\mathcal{A}^\ell,\mathscr{A}^\ell)$, $(\mathcal{Y},\mathscr{Y})=(\mathcal{B}^\ell,\mathscr{B}^\ell)$, and $\rho=\sigma_{(\ell)}$. Finally, set

$$R_{(X_i)_{i\in\mathbb{N}}}(D)=\lim_{\ell\to\infty}\frac{1}{\ell}R_{(X_1,\dots,X_\ell)}(D). \tag{1.7}$$

The limit in (1.7) is well-defined by [6, Problem 4.17]. Then, we have the following achievability result with matching converse:

(i) Suppose that there exists a $b^*\in\mathcal{B}$ with $\mathbb{E}\big[\sigma(X_1,b^*)\big]<\infty$. Then, for every $D\in[0,\infty)$ with $R_{(X_i)_{i\in\mathbb{N}}}(D)<\infty$, (R,D) is achievable for all $R>R_{(X_i)_{i\in\mathbb{N}}}(D)$ [6, Theorem 7.2.4].
(ii) For every $D\in[0,\infty)$ and $R<R_{(X_i)_{i\in\mathbb{N}}}(D)$, (R,D) is not achievable [6, Theorem 7.2.5].

Characterizing $R_{(X_i)_{i\in\mathbb{N}}}(D)$ analytically is, in general, very difficult, even in the case of i.i.d. sequences $(X_i)_{i\in\mathbb{N}}$, where (1.7) reduces to the single-letter expression $R_{(X_i)_{i\in\mathbb{N}}}(D)=R_{X_1}(D)$. For general stationary ergodic processes $(X_i)_{i\in\mathbb{N}}$, computation of $R_{(X_i)_{i\in\mathbb{N}}}(D)$ even requires knowledge of $R_{(X_1,\dots,X_\ell)}(D)$ for all $\ell\in\mathbb{N}$ with possibly the exception of finitely many ℓ. One therefore resorts to bounds on $R_X(D)$ for general X (e.g., $X=X_1$ for i.i.d. sequences or $X=(X_1,\dots,X_\ell)$ for general stationary ergodic processes). While upper bounds can be obtained by evaluating $I(X;Y)$ for a specific Y satisfying $\mathbb{E}[\rho(X,Y)]\le D$, lower bounds are notoriously hard to obtain. The best-known lower bounds are the Shannon lower bounds for i) discrete random variables with ρ such that, for every $s\in(0,\infty)$, $\sum_{x\in\mathcal{X}}e^{-s\rho(x,y)}$ does not depend on y [28, Lemma 4.3.1] and ii) continuous random variables with a difference distortion function [28, Equation (4.6.1)]. For continuous X of finite differential entropy under the difference distortion function $\rho(x,y)=\|x-y\|^k$, where $\|\cdot\|$ is a semi-norm and $k\in(0,\infty)$, the Shannon lower bound is known explicitly [64, Section VI] and, provided that X additionally satisfies a certain moment constraint, is tight as $D\to 0$ [39, 44]. For the class of m-rectifiable random variables [40, Definition 11], a Shannon lower bound was reported recently in [40, Theorem 55]. This bound is, however, not in explicit form and depends on a parametrized integral.

Asymptotic results on the R-D function for $(\mathcal{X},\omega)$ a metric space, $\mathcal{Y}=\mathcal{X}$, and $\rho(x,y)=\omega^k(x,y)$ with $k\in(0,\infty)$ were reported in [36] in terms of the lower and upper R-D dimensions of order k defined as

$$\underline{\dim}_{\mathrm{R}}(X)=\liminf_{D\to 0}\frac{R_X(D^k)}{\log(1/D)} \tag{1.8}$$

and

$$\overline{\dim}_{\mathrm{R}}(X)=\limsup_{D\to 0}\frac{R_X(D^k)}{\log(1/D)}, \tag{1.9}$$

respectively. If $\underline{\dim}_{\mathrm{R}}(X)=\overline{\dim}_{\mathrm{R}}(X)$, then this common value, denoted by $\dim_{\mathrm{R}}(X)$, is referred to as the R-D dimension of order k. Specifically, it is shown in [36] (see also [27, Remark 14.19]), that for a random variable X of self-similar distribution μ_X satisfying the assumptions of [27, Theorem 14.17], for every $k\in(0,\infty)$, the R-D dimension of order k exists and equals the Hausdorff dimension of μ_X given by

$$\dim_{\mathrm{H}}(\mu_X)=\inf\{\dim_{\mathrm{H}}(\mathcal{C}):\mathcal{C}\subseteq\mathbb{R}^d \text{ is Borel with } \mu_X(\mathcal{C})=1\}. \tag{1.10}$$

1.2. *Quantization*

In quantization [24, 27], one is concerned with characterizing the ultimate limits on the discretization of random variables. Specifically, let $(\mathcal{X},\mathscr{X})$ and $(\mathcal{Y},\mathscr{Y})$ be measurable spaces equipped with the distortion function $\rho\colon \mathcal{X}\times\mathcal{Y}\to[0,\infty]$, and let X be a random variable taking values in $(\mathcal{X},\mathscr{X})$. For every $n\in\mathbb{N}$, consider the set $\mathcal{F}_n(\mathcal{X},\mathcal{Y})$ of all measurable mappings $f\colon \mathcal{X}\to\mathcal{Y}$ with $|f(\mathcal{X})|\le n$. The elements of $\mathcal{F}_n$ are referred to as n-quantizers and the n-th quantization error is defined as

$$V_n(X)=\inf_{f\in\mathcal{F}_n(\mathcal{X},\mathcal{Y})}\mathbb{E}[\rho(X,f(X))]. \tag{1.11}$$

The typical setup is $\mathcal{X}\subseteq\mathcal{Y}=\mathbb{R}^d$ with difference distortion function $\rho(x,y)=\|x-y\|^k$, where $\|\cdot\|$ is a norm on $\mathbb{R}^d$ and $k\in[1,\infty)$ [29, Section III]. The case $d>1$ is commonly referred to as vector quantization (see [63] and references therein). While we restrict our attention to the expected average error as defined in (1.11), results in terms of the worst case error

$$\widetilde{V}_n(X)=\inf_{f\in\mathcal{F}_n(\mathcal{X},\mathcal{Y})}\operatorname{ess\,sup}\|X-f(X)\| \tag{1.12}$$

can be found in [27].

The n-th quantization error $V_n(X)$ is difficult to characterize in general. Asymptotic results are, however, available, specifically in terms of the lower and upper k-th quantization dimensions, defined as

$$\underline{D}_k(X)=\liminf_{n\to\infty}\frac{k\log(n)}{\log(1/V_n(X))} \tag{1.13}$$

and

$$\overline{D}_k(X)=\limsup_{n\to\infty}\frac{k\log(n)}{\log(1/V_n(X))}, \tag{1.14}$$

respectively. If $\underline{D}_k(X)=\overline{D}_k(X)$, then this common value, denoted by $D_k(X)$, is referred to as the k-th quantization dimension. For the special case $\mathcal{X}\subseteq\mathcal{Y}=\mathbb{R}^d$, $\rho(x,y)=\|x-y\|^k$ with $\|\cdot\|$ a norm on $\mathbb{R}^d$, and $k\in[1,\infty)$, it is known that $\underline{D}_k(X)\ge\max\{\dim_{\mathrm{H}}(\mu_X),\underline{\dim}_{\mathrm{R}}(X)\}$ [27, Theorems 11.6 and 11.10] with $\dim_{\mathrm{H}}(\mu_X)$ as defined in (1.10). If, in addition, μ_X is $\|\cdot\|^k$-regular of dimension m (see Definition 2.1), then $D_k(X)=\dim_H(\mu_X)=m$ [27, Theorem 12.18], i.e., the k-th quantization dimension exists and equals the Hausdorff dimension of μ_X.

If the k-th quantization dimension $D_k(X)$ exists in $(0,\infty)$, a more accurate characterization of the asymptotic behavior of $V_n(X)$ can be obtained in terms of the lower and upper k-th quantization coefficient defined as

$$\underline{C}_k(X)=\liminf_{n\to\infty}n^{\frac{k}{D_k(X)}}V_n(X) \tag{1.15}$$

and

$$\overline{C}_k(X)=\limsup_{n\to\infty}n^{\frac{k}{D_k(X)}}V_n(X), \tag{1.16}$$

respectively. If $\underline{C}_k(X) = \overline{C}_k(X)$, then this common value, denoted by $C_k(X)$, is referred to as the k-th quantization coefficient. Results on the k-th quantization coefficient are scarce and available in very specific cases only. More concretely, if $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ with difference distortion function $\rho(x,y) = \|x-y\|^k$, where $\|\cdot\|$ is a norm on $\mathbb{R}^d$ and $k \in [1,\infty)$, then the following results are available:

- If the continuous part of the distribution of X is nonzero then, for every $k \in [1,\infty)$, provided that there is a $\delta \in (0,\infty)$ such that $\mathbb{E}[\|X\|^{\delta+k}] < \infty$, $C_k(X)$ exists in $(0,\infty)$ and an analytic expression for $C_k(X)$ is available [27, Theorem 6.2, Remark 6.3].
- If the distribution of X is $\rho^{1/k}$-regular of dimension m (see Definition 2.1), then $0 < \underline{C}_k(X) \leq \overline{C}_k(X) < \infty$ [27, Theorem 12.18]. A condition for the existence of $C_k(X)$ does not seem to be available [27, Remark 12.19].
- For X of self-similar distribution, $C_k(X)$ exists in $(0,\infty)$ provided that the defining vector [53, Equation (1.1)] of the self-similar distribution is nonarithmetic[3] [53, Theorem 1]. Analytic characterizations for the lower and upper k-th quantization coefficient do not seem to be available in general, with the exception of X distributed uniformly on the middle-third Cantor set [26, Theorem 5.2].

1.3. *Contributions*

Most of our results, both for R-D theory and quantization, hold for general measurable spaces $(\mathcal{X},\mathscr{X})$ and $(\mathcal{Y},\mathscr{Y})$ equipped only with a distortion function $\rho\colon \mathcal{X}\times\mathcal{Y} \to [0,\infty]$ and do not need a common ambient space or metric on $\mathcal{X}$ or $\mathcal{Y}$. Our main contribution to R-D theory is a lower bound $R_X^{\mathrm{L}}(D)$ on the R-D function $R_X(D)$ in (1.5) for random variables X taking values in the measurable space $(\mathcal{X},\mathscr{X})$. The only requirement for this bound to apply is that the distribution μ_X of X is absolutely continuous with respect to a (σ-)finite measure μ on $(\mathcal{X},\mathscr{X})$ of finite generalized entropy $h_\mu(X)$ (as defined in (1.27)) and satisfying a certain subregularity condition. Specifically, this subregularity condition guarantees that the measure μ is not too concentrated on ρ-balls of small radii in the following sense: There exist constants $m \in [0,\infty)$, $c \in (0,\infty)$, and $\delta_0 \in (0,\infty]$ such that

$$\mu\big(\mathcal{B}_\rho(y,\delta)\big) \leq c\delta^m \quad \text{for all } y \in \mathcal{Y} \text{ and } \delta \in (0,\delta_0) \tag{1.17}$$

with the ρ-ball of radius δ centered at y defined as $\mathcal{B}_\rho(y,\delta) = \{x \in \mathcal{X} : \rho(x,y) < \delta\}$. We refer to m as subregularity dimension and to c, δ_0 as subregularity constants. The subregularity condition (1.17) is satisfied, e.g., by the m-dimensional Hausdorff measure restricted to an arbitrary regular set $\mathcal{K}$ of dimension m in $\mathbb{R}^d$ and with $\mathcal{Y} = \mathcal{K}$ (see [27, Section 12]). Specific examples of regular sets of dimension m are compact convex sets $\mathcal{K} \subseteq \mathbb{R}^m$ with $\mathrm{span}(\mathcal{K}) = \mathbb{R}^m$ [27, Example 12.7], surfaces of compact convex sets $\mathcal{K} \subseteq \mathbb{R}^{m+1}$ with $\mathrm{span}(\mathcal{K}) = \mathbb{R}^{m+1}$ [27, Example 12.8], m-dimensional compact C^1-submanifolds of $\mathbb{R}^d$ [27, Example 12.9], self-similar sets of similarity dimension m satisfying the weak separation property [22, Theorem 2.1], and finite unions of regular sets of dimension m [27, Lemma 12.4]. The lower bound $R_X^{\mathrm{L}}(D)$ we obtain allows us to conclude that the lower R-D dimension $\underline{\dim}_{\mathrm{R}}(X)$ of order $k \in (0,\infty)$ is lower-bounded by the subregularity dimension m, i.e., $\underline{\dim}_{\mathrm{R}}(X) \geq m$. For continuous X of finite differential entropy and distortion function $\rho(x-y) = \|x-y\|^k$, where $\|\cdot\|$ is a semi-norm on $\mathbb{R}^d$ and $k \in (0,\infty)$, our lower bound reduces to the classical Shannon lower bound in [64].

---

[3] A vector $(x_1,\dots,x_d) \in \mathbb{R}^d$ with $x_i > 0$ for $i = 1,\dots,d$ is said to be arithmetic if there exists a $\theta \in (0,\infty)$ such that $x_i/\theta \in \mathbb{N}$ for $i = 1,\dots,d$.

Our first main contribution to quantization is a lower bound on the n-th quantization error for random variables X taking values in the measurable space $(\mathcal{X}, \mathscr{X})$. The only requirement for this bound to apply is that the distribution μ_X of X is absolutely continuous with respect to a σ-finite measure μ on $(\mathcal{X}, \mathscr{X})$ satisfying the subregularity condition (1.17) and $\|\mathrm{d}\mu_X/\mathrm{d}\mu\|_{p/(p-1)}^{(\mu)} < \infty$ with $p \in [1,\infty)$. Further, the lower bound we obtain allows us to conclude that

$$\underline{D}_k(X) \geq m/p \quad \text{for all } k \in (0,\infty). \tag{1.18}$$

Moreover, we show that, if $D_k(X)$ exists and satisfies $D_k(X) = m/p$, then

$$\underline{C}_k(X) \geq \frac{m}{m+pk} c^{-\frac{k}{m}} \left(\|\mathrm{d}\mu_X/\mathrm{d}\mu\|_{p/(p-1)}^{(\mu)} \right)^{-\frac{pk}{m}}. \tag{1.19}$$

Our second main contribution to quantization is an upper bound on the n-th quantization error for random variables X taking values in the measurable space $(\mathcal{X}, \mathscr{X})$. This result requires the existence of a finite measure[4] ν on $(\mathcal{Y}, \mathscr{Y})$ satisfying a certain superregularity condition. Specifically, this superregularity condition guarantees that the measure ν is not too small on ρ-balls of small radii in the following sense: There exist constants $m \in [0,\infty)$, $b \in (0,\infty)$, and $\delta_0 \in (0,\infty]$ such that

$$\nu\left(\widetilde{\mathcal{B}}_\rho(x,\delta)\right) \geq b\delta^m \quad \text{for all } x \in \mathcal{X} \text{ and } \delta \in (0,\delta_0) \tag{1.20}$$

with the ρ-ball of radius δ centered at x defined as $\widetilde{\mathcal{B}}_\rho(x,\delta) = \{y \in \mathcal{Y} : \rho(x,y) < \delta\}$. We refer to m as the superregularity dimension and to b, δ_0 as superregularity constants. As in the case of subregularity, the m-dimensional Hausdorff measure restricted to an arbitrary regular set $\mathcal{K}$ of dimension m in $\mathbb{R}^d$ satisfies the superregularity condition (1.20) with $\mathcal{X} = \mathcal{K}$ (see [27, Section 12]). The upper bound we obtain allows us to conclude that the k-th upper quantization dimension $\overline{D}_k(X)$ is upper-bounded by the superregularity dimension m, i.e.,

$$\overline{D}_k(X) \leq m \quad \text{for all } k \in (0,\infty). \tag{1.21}$$

Moreover, we show that, if $D_k(X)$ exists and satisfies $D_k(X) = m$, then

$$\overline{C}_k(X) \leq \Gamma\left(1 + \frac{m}{k}\right) b^{-\frac{k}{m}}. \tag{1.22}$$

If, in addition to the assumptions pertaining to the upper bound, $D_k(X) = m$, $\mathcal{X} = \mathcal{Y} \subseteq \mathbb{R}^d$ is Borel, $\rho(x,y) = \omega^k(x,y)$ for $k \in (0,m)$ with ω a metric on $\mathbb{R}^d$ such that

$$\sup_{x,y \in \mathcal{X}} \omega(x,y) < \infty, \tag{1.23}$$

and $\nu \ll \mu_X$ with $\|\mathrm{d}\nu/\mathrm{d}\mu_X\|_\infty^{(\mu_X)} < \infty$, then we get the improved upper bound

$$\overline{C}_k(X) \leq \mathbb{E}\left[\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu_X} \right)^{\frac{k}{m}} (X) \right] \Gamma\left(1 + \frac{m}{k}\right) b^{-\frac{k}{m}}. \tag{1.24}$$

---

[4] We assume, without loss of generality, that $\nu(\mathcal{Y}) = 1$.

To illustrate the wide applicability of our results, we particularize them to i) compact manifolds, specifically hyperspheres and Grassmannians, and ii) self-similar sets generated by iterated function systems satisfying the weak separation property, with the middle third Cantor set as a specific (and prominent) example.

Notation. Sets are designated by calligraphic letters, e.g., $\mathcal{A}$, with $|\mathcal{A}|$ denoting cardinality, $\overline{\mathcal{A}}$ closure, and $\mathcal{A}^\ell$ the ℓ-fold cartesian product. σ-algebras are indicated by script letters, e.g., $\mathscr{X}$, and are assumed to contain all singleton sets. Unless stated otherwise, (subsets of) topological spaces, e.g., (subsets of) $\mathbb{R}^d$, hyperspheres, or Grassmannians are understood to be equipped with the Borel σ-algebra corresponding to the (induced) topology. Measures defined on Borel sets are assumed to be Borel measures. We let $\mathscr{X} \otimes \mathscr{Y}$ be the product σ-algebra formed by $\mathscr{X}$ and $\mathscr{Y}$ and write $\mathscr{X}^\ell$ for the ℓ-fold product σ-algebra corresponding to $\mathscr{X}$. For $k \in [1,\infty)$, $\|\cdot\|_k$ stands for the k-norm on $\mathbb{R}^d$. Measures are assumed to be positive. For a real-valued and μ-measurable function f and $p \in (0,\infty]$, we let

$$\|f\|_p^{(\mu)} = \left(\int \big|f(x)\big|^p \,\mathrm{d}\mu(x)\right)^{\frac{1}{p}} \quad \text{if } p < \infty, \tag{1.25}$$

and set $\|f\|_\infty^{(\mu)} = \operatorname{ess\,sup}|f|$.

For a measure space $(\mathcal{X}, \mathscr{X}, \mu)$, we write[5] $\mu|_\mathcal{A}$ for the restriction of μ to $\mathcal{A} \in \mathscr{X}$ and $f_*(\mu)$ for the pushforward measure corresponding to the measurable mapping f from $(\mathcal{X}, \mathscr{X}, \mu)$ into any measurable space $(\mathcal{Y}, \mathscr{Y})$. Throughout, random variables are indicated by capital letters, e.g., X, and we write μ_X for the corresponding distribution. The support of the Borel measure μ is denoted by $\operatorname{supp}(\mu)$ [2, Definition 1.64]. For measures μ and ν, $\mu \otimes \nu$ designates the corresponding product measure. For μ and σ-finite ν defined on the same measurable space with μ absolutely continuous with respect to ν, expressed by $\mu \ll \nu$, we write $\mathrm{d}\mu/\mathrm{d}\nu$ for the Radon–Nikodým derivative of μ with respect to ν. We denote the m-dimensional Hausdorff measure by $\mathscr{H}^m$ [2, Definition 2.46], write $\dim_\mathrm{H}$ for the Hausdorff dimension [2, Definition 2.51], and let $\mathscr{L}^m$ be the m-dimensional Lebesgue measure. The $\mathscr{H}^m$-measurable set $\mathcal{A} \subseteq \mathbb{R}^n$ is said to be $\mathscr{H}^m$-rectifiable [2, Definition 2.57] if $\mathscr{H}^m(\mathcal{A}) < \infty$ and there exist Lipschitz mappings $f_i \colon \mathbb{R}^m \to \mathbb{R}^n$, $i \in \mathbb{N}$, such that

$$\mathscr{H}^m\left(\mathcal{A} \setminus \bigcup_{i\in\mathbb{N}} f_i(\mathbb{R}^m)\right) = 0. \tag{1.26}$$

$\mathbb{E}[\cdot]$ denotes the expectation operator. For X taking values in the σ-finite measure space $(\mathcal{X}, \mathscr{X}, \mu)$ with $\mu_X \ll \mu$, we define the generalized entropy as

$$h_\mu(X) = -\mathbb{E}\left[\log\left(\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu}(X)\right)\right]. \tag{1.27}$$

For X taking values in $(\mathcal{X}, \mathscr{X})$ and Y taking values in $(\mathcal{Y}, \mathscr{Y})$, the mutual information between X and Y is

$$I(X;Y) = \mathbb{E}\left[\log\left(\frac{\mathrm{d}\mu_{(X,Y)}}{\mathrm{d}(\mu_X \otimes \mu_Y)}(X,Y)\right)\right] \tag{1.28}$$

[5] The restricted measure $\mu|_\mathcal{A}$ is defined on $\mathcal{X}$ equipped with the trace σ-algebra $\mathscr{X}_\mathcal{A} := \{\mathcal{Z} \cap \mathcal{A} : \mathcal{Z} \in \mathscr{X}\}$. It induces a measure on $(\mathcal{A}, \mathscr{X}_\mathcal{A})$ as well, for which we shall also use the symbol $\mu|_\mathcal{A}$.

if $\mu_{(X,Y)} \ll \mu_X \otimes \mu_Y$, and $I(X;Y) = \infty$ else. For given distortion function $\rho\colon \mathcal{X} \times \mathcal{Y} \to [0,\infty]$, we define the open ρ-balls of radius $\delta \in (0,\infty)$ centered at $y \in \mathcal{Y}$ and $x \in \mathcal{X}$ according to

$$\mathcal{B}_\rho(y,\delta) = \big\{x \in \mathcal{X} : \rho(x,y) < \delta\big\} \tag{1.29}$$

and

$$\widetilde{\mathcal{B}}_\rho(x,\delta) = \big\{y \in \mathcal{Y} : \rho(x,y) < \delta\big\}, \tag{1.30}$$

respectively. For $a \in (0,\infty)$, the gamma function is given by $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t}\,\mathrm{d}t$. For $a \in (0,\infty)$ and $s \in [0,\infty)$, the lower and upper incomplete gamma functions are $\gamma(a,s) = \int_0^s t^{a-1} e^{-t}\,\mathrm{d}t$ and $\Gamma(a,s) = \int_s^\infty t^{a-1} e^{-t}\,\mathrm{d}t$, respectively. For $a,b \in (0,\infty)$, the beta function is given by $B(a,b) = \int_0^1 u^{a-1}(1-u)^{b-1}\,\mathrm{d}u$ and satisfies the identity [3, Theorem 1.1.4]

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \tag{1.31}$$

For $a,b \in (0,\infty)$ and $s \in (0,1]$, the incomplete beta function is defined as $B_{a,b}(s) = \int_0^s u^{a-1}(1-u)^{b-1}\,\mathrm{d}u$ and

$$I_{a,b}(s) = \frac{B_{a,b}(s)}{B(a,b)} \tag{1.32}$$

denotes the normalized incomplete beta function.

For $a \in \mathbb{R}$, we let $\lfloor a \rfloor$ be the largest integer less than or equal to a. log denotes the logarithm to base e and we set $\operatorname{sinc}(x) = \sin(\pi x)/(\pi x)$ for all $x \in (-\infty,\infty)\setminus\{0\}$ with $\operatorname{sinc}(0) := 1$ by continuous extension. We write $\chi_{\mathcal{A}}(\cdot)$ for the indicator function on the set $\mathcal{A}$. For $r \in (0,\infty)$ and $d \in \mathbb{N}$ with $d > 1$, we set $\mathcal{S}^{d-1}(r) = \{x \in \mathbb{R}^d : \|x\|_2 = r\}$ with corresponding surface area

$$a^{(d-1)}(r) := \mathscr{H}^{d-1}(\mathcal{S}^{d-1}(r)) = \frac{2\pi^{\frac{d}{2}}}{\Gamma\big(\frac{d}{2}\big)} r^{d-1}. \tag{1.33}$$

For $d \in \mathbb{N}$ and $\|\cdot\|$ a norm on $\mathbb{R}^d$, we write $\kappa_d(\|\cdot\|)$ for the volume of the corresponding unit ball in $\mathbb{R}^d$ and we set

$$v^{(d)}(r) = \kappa_d(\|\cdot\|_2) r^d = \frac{\pi^{\frac{d}{2}} r^d}{\Gamma\big(1+\frac{d}{2}\big)}. \tag{1.34}$$

We use the convention $0 \cdot \infty = 0$.

2. The Regularity Conditions

We start by stating the formal definitions of subregularity, superregularity, and regularity of measures.

Definition 2.1. Let $(\mathcal{X},\mathscr{X})$ and $(\mathcal{Y},\mathscr{Y})$ be measurable spaces equipped with the distortion function $\rho\colon \mathcal{X} \times \mathcal{Y} \to [0,\infty]$.

(i) A measure μ on $(\mathcal{X},\mathscr{X})$ is said to be ρ-subregular of dimension $m \in [0,\infty)$ if there exist subregularity constants $c \in (0,\infty)$ and $\delta_0 \in (0,\infty]$ such that

$$\mu\big(\mathcal{B}_\rho(y,\delta)\big) \le c\delta^m \quad \text{for all } y \in \mathcal{Y} \text{ and } \delta \in (0,\delta_0). \tag{2.1}$$

(ii) A measure ν on $(\mathcal{Y},\mathscr{Y})$ is said to be ρ-superregular of dimension $m \in [0,\infty)$ if there exist superregularity constants $b \in (0,\infty)$ and $\delta_0 \in (0,\infty]$ such that

$$\nu\Big(\widetilde{\mathcal{B}}_\rho(x,\delta)\Big) \ge b\delta^m \quad \text{for all } x \in \mathcal{X} \text{ and } \delta \in (0,\delta_0). \tag{2.2}$$

(iii) A measure ν on $(\mathcal{Y},\mathscr{Y})$ is said to be ρ-regular of dimension $m \in [0,\infty)$ if there exist regularity constants $b,c \in (0,\infty)$ and $\delta_0 \in (0,\infty]$ such that

$$b\delta^m \le \nu\Big(\widetilde{\mathcal{B}}_\rho(x,\delta)\Big) \le c\delta^m \quad \text{for all } x \in \mathcal{X} \text{ and } \delta \in (0,\delta_0). \tag{2.3}$$

If ν is ρ-regular, then it is also ρ-superregular with the same constants b and δ_0. For $\mathcal{X} = \mathcal{Y}$ and ρ symmetric, i.e., $\rho(x,y) = \rho(y,x)$ for all $x,y \in \mathcal{X}$, ν is ρ-regular if and only if it is both ρ-subregular and ρ-superregular. For ν a Borel measure on $\mathbb{R}^d$ with $\mathcal{X} = \operatorname{supp}(\nu)$ compact and $\rho\colon \mathcal{X} \times \mathbb{R}^d \to [0,\infty)$, $\rho(x,y) = \|x-y\|$, where $\|\cdot\|$ is a norm on $\mathbb{R}^d$, ρ-regularity of dimension m agrees with regularity of dimension m as defined in [27, Definition 12.1]. In this case, ρ-regularity also implies ρ-subregularity provided that ν is a finite measure [27, Lemma 12.3].

We next give some examples of ρ-regular measures for $\mathcal{X} \subseteq \mathbb{R}^d$ and $\rho\colon \mathcal{X} \times \mathbb{R}^d \to [0,\infty)$, $\rho(x,y) = \|x-y\|$, where $\|\cdot\|$ is a norm on $\mathbb{R}^d$:

(i) Lebesgue measure on $\mathcal{X} = \mathbb{R}^d$ satisfies (2.3) with $m = d$, $b = c$, and $\delta_0 = \infty$.
(ii) Measures supported on discrete sets $\mathcal{X} \subseteq \mathbb{R}^d$ satisfy (2.3) for $m = 0$.
(iii) The restricted Hausdorff measure $\mu = \mathscr{H}^d|_{\mathcal{X}}$ satisfies $0 < \mathscr{H}^d(\mathcal{X}) < \infty$ and (2.3) with $m = d$ for the following classes of sets $\mathcal{X}$: compact convex sets $\mathcal{X} \subseteq \mathbb{R}^d$ with $\operatorname{span}(\mathcal{X}) = \mathbb{R}^d$ [27, Example 12.7], surfaces of compact convex sets $\mathcal{X} \subseteq \mathbb{R}^{d+1}$ with $\operatorname{span}(\mathcal{X}) = \mathbb{R}^{d+1}$ [27, Example 12.8], d-dimensional compact C^1-submanifolds of $\mathbb{R}^n$ [27, Example 12.9], self-similar sets of similarity dimension d satisfying the weak separation property [22, Theorem 2.1], and finite unions of any such sets [27, Lemma 12.4].

In all these examples $\mathcal{X} \subseteq \mathcal{Y} = \mathbb{R}^d$, $\rho(x,y) = \|x-y\|$, where $\|\cdot\|$ is a norm on $\mathbb{R}^d$, and the regularity dimension of the measure equals the Hausdorff dimension of $\mathcal{X}$ (for Items (i) and (ii), this follows from arguments in [19, Section 3.2]; for (iii) it is a consequence of the assumption $0 < \mathscr{H}^d(\mathcal{X}) < \infty$). Note that Definition 2.1 is much more general as it does not require $\mathcal{X} \subseteq \mathcal{Y}$ and the distortion function need not be symmetric. The following simple example illustrates this aspect for $\mathcal{X}$ a Grassmannian and $\mathcal{Y}$ a hypersphere (the more complicated case where $\mathcal{X}$ and $\mathcal{Y}$ are both Grassmannians is investigated in Section 6.2).

Example 2.1. Let $p,d \in \mathbb{N}$ with $p < d$ and denote by $\mathcal{G}^{\mathbb{R}}(p,d)$ the $p(d-p)$-dimensional Grassmannian consisting of all p-dimensional subspaces of $\mathbb{R}^d$ [10, Section 1.3.2]. For every $x \in \mathcal{G}^{\mathbb{R}}(p,d)$, let $\mathrm{P}_x^\perp$ denote the orthogonal projection onto the orthogonal complement of x in $\mathbb{R}^d$. Let $\gamma_{p,d}$ be the unique uniformly distributed Borel regular measure on $\mathcal{G}^{\mathbb{R}}(p,d)$ with $\gamma_{p,d}(\mathcal{G}^{\mathbb{R}}(p,d)) = 1$ [50, p. 49]. Fix $r \in (0,\infty)$ and consider the distortion function $\rho\colon \mathcal{G}^{\mathbb{R}}(p,d) \times \mathcal{S}^{d-1}(r) \to [0,\infty)$ defined according to

$$\rho(x,y) = \|\mathrm{P}_x^\perp(y)\|_2 \quad \text{for } x \in \mathcal{G}^{\mathbb{R}}(p,d) \text{ and } y \in \mathcal{S}^{d-1}(r). \tag{2.4}$$

It follows that [50, Proof of Lemma 3.11]

$$\gamma_{p,d}\big(\mathcal{B}_\rho(y,\delta)\big) = g_{p,d,r}(\delta) \quad \text{for all } y \in \mathcal{S}^{d-1}(r) \text{ and } \delta \in (0,\infty), \tag{2.5}$$

where

$$g_{p,d,r}(\delta) = \frac{\mathscr{H}^{d-1}\Big\{y \in \mathcal{S}^{d-1}(1) : \sum_{i=p+1}^{d} y_i^2 < (\delta/r)^2\Big\}}{\mathscr{H}^{d-1}(\mathcal{S}^{d-1}(1))} \tag{2.6}$$

with y_i denoting the i-th entry of $y = (y_1,\dots,y_d)$. The function $g_{p,d,r}(\delta)$ can be upper-bounded as follows:

$$g_{p,d,r}(\delta) = \mathscr{L}^d\left\{y \in \mathbb{R}^d : \|y\|_2 \le 1, \sum_{i=p+1}^{d} y_i^2 < (\delta/r)^2\right\}\frac{\Gamma\Big(1+\frac{d}{2}\Big)}{\pi^{\frac{d}{2}}} \tag{2.7}$$

$$\le \mathscr{L}^d\left\{y \in \mathbb{R}^d : \sum_{i=1}^{p} y_i^2 \le 1, \sum_{i=p+1}^{d} y_i^2 < (\delta/r)^2\right\}\frac{\Gamma\Big(1+\frac{d}{2}\Big)}{\pi^{\frac{d}{2}}} \tag{2.8}$$

$$= \frac{\Gamma\Big(1+\frac{d}{2}\Big)}{r^{d-p}\Gamma\big(1+\frac{p}{2}\big)\Gamma\Big(1+\frac{d-p}{2}\Big)}\delta^{d-p} \tag{2.9}$$

$$= \frac{2}{p\,r^{d-p}B(p/2, 1+(d-p)/2)}\delta^{d-p} \quad \text{for all } \delta \in (0,\infty), \tag{2.10}$$

where (2.7) is by [50, Equation (3.6)], in (2.9) we used (1.34), and (2.10) follows from (1.31) and $\Gamma(1+p/2) = (p/2)\Gamma(p/2)$. We can hence conclude that $\gamma_{p,d}$ is ρ-subregular of dimension $d-p$. The corresponding subregularity constants are

$$c_{p,d,r} = \frac{2}{p\,r^{d-p}B(p/2, 1+(d-p)/2)} \tag{2.11}$$

and $\delta_0 = \infty$.

We continue by listing properties of subregularity and superregularity needed throughout the paper. The first result constitutes an extension of (2.1)–(2.3) to the half-closed interval $(0,\delta_0]$ applicable whenever $\delta_0 < \infty$.

Lemma 2.1 *Let $(\mathcal{X},\mathscr{X})$ and $(\mathcal{Y},\mathscr{Y})$ be measurable spaces equipped with the distortion function $\rho\colon \mathcal{X}\times\mathcal{Y} \to [0,\infty]$. Then, the following properties hold:*

(i) If μ satisfies (2.1) for $\delta_0 < \infty$, then

$$\mu\big(\mathcal{B}_\rho(y,\delta)\big) \le c\delta^m \quad \textit{for all } y \in \mathcal{Y} \textit{ and } \delta \in (0,\delta_0]. \tag{2.12}$$

(ii) If ν satisfies (2.2) for $\delta_0 < \infty$, then

$$\nu\Big(\widetilde{\mathcal{B}}_\rho(x,\delta)\Big) \ge b\delta^m \quad \textit{for all } x \in \mathcal{X} \textit{ and } \delta \in (0,\delta_0]. \tag{2.13}$$

(iii) *If μ satisfies (2.3) for $\delta_0 < \infty$, then*

$$b\delta^m \leq \mu\left(\widetilde{\mathcal{B}}_\rho(x,\delta)\right) \leq c\delta^m \quad \text{for all } x \in \mathcal{X} \text{ and } \delta \in (0,\delta_0]. \tag{2.14}$$

Proof To establish Item (i), suppose that (2.1) holds with $\rho_0 < \infty$. Then, we have

$$\mu\left(\mathcal{B}_\rho(y,\delta_0)\right) = \mu\left(\bigcup_{n=\lceil 1/\rho_0\rceil+1}^{\infty} \mathcal{B}_\rho\left(y,\delta_0 - \frac{1}{n}\right)\right) \tag{2.15}$$

$$= \lim_{n\to\infty} \mu\left(\mathcal{B}_\rho\left(y,\delta_0 - \frac{1}{n}\right)\right) \tag{2.16}$$

$$\leq c \lim_{n\to\infty}\left(\delta_0 - \frac{1}{n}\right)^m \tag{2.17}$$

$$= c\delta_0^m, \tag{2.18}$$

where (2.16) follows from [5, Lemma 3.4, Item (a)] and (2.17) is by (2.1) with $0 < \delta = \delta_0 - 1/n < \delta_0$ for all $n \geq \lceil 1/\rho_0\rceil + 1$. The proofs of Items (ii) and (iii) follow along the same lines. □

Next, we establish scaling properties of subregular and superregular measures.

Lemma 2.2 *Let $(\mathcal{X},\mathscr{X})$ and $(\mathcal{Y},\mathscr{Y})$ be measurable spaces equipped with the distortion function $\rho\colon \mathcal{X}\times\mathcal{Y} \to [0,\infty]$ and fix $k,\alpha,\beta \in (0,\infty)$. Then, for a measure μ on $(\mathcal{X},\mathscr{X})$, the following statements are equivalent:*

- *(i)* *μ is $\rho^{1/k}$-subregular of dimension m with subregularity constants $c \in (0,\infty)$ and $\delta_0 \in (0,\infty]$.*
- *(ii)* *$\beta\mu$ is $(\alpha\rho)$-subregular of dimension m/k with subregularity constants $\tilde{c} = \beta c/\alpha^{m/k}$ and $\tilde{\delta}_0 = \alpha\delta_0^k$.*

Similarly, for a measure ν on $(\mathcal{Y},\mathscr{Y})$, the following statements are equivalent:

- *(a)* *ν is $\rho^{1/k}$-superregular of dimension m with superregularity constants $b \in (0,\infty)$ and $\delta_0 \in (0,\infty]$.*
- *(b)* *$\beta\nu$ is $(\alpha\rho)$-superregular of dimension m/k with superregularity constants $\tilde{b} = \beta b/\alpha^{m/k}$ and $\tilde{\delta}_0 = \alpha\delta_0^k$.*

Proof Follows directly from Definition 2.1. □

By Lemma 2.2, applied with $k = \alpha = 1$, we can hence assume, without loss of generality, that finite subregular measures μ on $(\mathcal{X},\mathscr{X})$ and finite superregular measures ν on $(\mathcal{Y},\mathscr{Y})$ are normalized according to $\mu(\mathcal{X}) = 1$ and $\nu(\mathcal{Y}) = 1$, respectively.

Next, we show that if μ is subregular with subregularity constants $c,\delta_0 \in (0,\infty)$ and $\mu(\mathcal{X}) = 1$, then c can be modified to make the subregularity condition (2.1) hold for $\delta_0 = \infty$. The corresponding formal statement is as follows.

Lemma 2.3 *Let $(\mathcal{X},\mathscr{X},\mu)$ be a finite measure space with $\mu(\mathcal{X}) = 1$, let $(\mathcal{Y},\mathscr{Y})$ be a measurable space, and consider the distortion function $\rho\colon \mathcal{X}\times\mathcal{Y} \to [0,\infty]$. If μ is ρ-subregular of dimension m with* *subregularity constants $c,\delta_0 \in (0,\infty)$, then μ is also ρ-subregular of dimension m with subregularity constants $\tilde{c} = \max(c, \delta_0^{-m})$ and $\tilde{\delta}_0 = \infty$, i.e.,*

$$\mu\big(\mathcal{B}_\rho(y,\delta)\big) \le \tilde{c}\delta^m \quad \textit{for all } y \in \mathcal{Y} \textit{ and } \delta \in (0,\infty). \tag{2.19}$$

Proof Let $y \in \mathcal{Y}$ and $\delta \in (0,\infty)$ be arbitrary but fixed. If $\delta < \delta_0$, then (2.19) follows directly from (2.1). For $\delta \ge \delta_0$, (2.19) holds by $\mu\Big(\mathcal{B}_\rho(y,\delta)\Big) \le \mu(\mathcal{X}) = 1 \le \delta_0^{-m}\delta^m$. □

In light of Lemma 2.3, if μ is ρ-subregular of dimension m with subregularity constants $c, \delta_0 \in (0,\infty)$ satisfying $c \ge \delta_0^{-m}$ and if $\mu(\mathcal{X}) = 1$, then the subregularity condition (2.1) holds for $\delta_0 = \infty$ with c unchanged.

The following result allows to infer subregularity of $\tilde{\mu} \ll \mu$ from subregularity of μ.

Lemma 2.4 *Let $(\mathcal{X}, \mathscr{X}, \mu)$ be a measure space, let $(\mathcal{Y}, \mathscr{Y})$ be a measurable space, and consider the distortion function $\rho\colon \mathcal{X} \times \mathcal{Y} \to [0,\infty]$. Suppose that μ is ρ-subregular of dimension m with subregularity constants $c \in (0,\infty)$ and $\delta_0 \in (0,\infty]$. If $\tilde{\mu} \ll \mu$ and $\|\mathrm{d}\tilde{\mu}/\mathrm{d}\mu\|_{p/(p-1)}^{(\mu)} < \infty$ with $p \in [1,\infty)$, then $\tilde{\mu}$ is ρ-subregular of dimension m/p with subregularity constants $\|\mathrm{d}\tilde{\mu}/\mathrm{d}\mu\|_{p/(p-1)}^{(\mu)} c^{1/p}$ and δ_0.*

Proof We have

$$\tilde{\mu}\big(\mathcal{B}_\rho(y,\delta)\big) = \left\| \frac{\mathrm{d}\tilde{\mu}}{\mathrm{d}\mu} \chi_{\mathcal{B}_\rho(y,\delta)} \right\|_1^{(\mu)} \tag{2.20}$$

$$\le \left\| \chi_{\mathcal{B}_\rho(y,\delta)} \right\|_p^{(\mu)} \left\| \frac{\mathrm{d}\tilde{\mu}}{\mathrm{d}\mu} \right\|_{p/(p-1)}^{(\mu)} \tag{2.21}$$

$$= \Big(\mu\big(\mathcal{B}_\rho(y,\delta)\big)\Big)^{\frac{1}{p}} \left\| \frac{\mathrm{d}\tilde{\mu}}{\mathrm{d}\mu} \right\|_{p/(p-1)}^{(\mu)} \tag{2.22}$$

$$\le c^{\frac{1}{p}} \left\| \frac{\mathrm{d}\tilde{\mu}}{\mathrm{d}\mu} \right\|_{p/(p-1)}^{(\mu)} \delta^{\frac{m}{p}} \quad \text{for all } y \in \mathcal{Y} \text{ and } \delta \in (0,\delta_0), \tag{2.23}$$

where in (2.21) we applied Hölder's inequality [35, Theorem 1, p. 372] and (2.23) is by subregularity of μ. □

If $\|\mathrm{d}\tilde{\mu}/\mathrm{d}\mu\|_\infty^{(\mu)} < \infty$, then Lemma 2.4 can be applied with $p = 1$ to conclude that not only does $\tilde{\mu}$ inherit subregularity from μ, but does so while preserving the subregularity dimension as a consequence of $m/p = m$. If $\|\mathrm{d}\tilde{\mu}/\mathrm{d}\mu\|_\infty^{(\mu)} = \infty$, the subregularity dimension need not be preserved, even when $\|\mathrm{d}\tilde{\mu}/\mathrm{d}\mu\|_q^{(\mu)} < \infty$ for all $q \in [1,\infty)$, as is illustrated by the following example.

Example 2.2. Take $\mathcal{X} = \mathcal{Y} = [0,1]$ and $\tilde{\mu} \ll \mathscr{L}^1|_{\mathcal{X}}$ with

$$\frac{\mathrm{d}\tilde{\mu}}{\mathrm{d}\mathscr{L}^1|_{\mathcal{X}}}(x) = -\log(x). \tag{2.24}$$

It follows that

$$\|\mathrm{d}\tilde{\mu}/\mathrm{d}\mathscr{L}^1|_{\mathcal{X}}\|_q^{(\mathscr{L}^1|_{\mathcal{X}})} = \left(\int_0^1 (-\log(x))^q \,\mathrm{d}x\right)^{\frac{1}{q}} \tag{2.25}$$

$$= \Gamma^{\frac{1}{q}}(1+q) \tag{2.26}$$

$$< \infty \quad \text{for all } q \in [1,\infty) \tag{2.27}$$

and $\|\mathrm{d}\tilde{\mu}/\mathrm{d}\mathscr{L}^1|_{\mathcal{X}}\|_\infty^{(\mathscr{L}^1|_{\mathcal{X}})} = \infty$. Lemma 2.4 with $p = q/(q-1)$ together with (2.25)–(2.27) therefore implies that $\tilde{\mu}$ is $|\cdot|$-subregular of dimension m/p for all $p \in (1,\infty)$. But

$$\tilde{\mu}\{x \in [0,1] : x < \delta\} = -\int_0^\delta \log(x)\,\mathrm{d}x \tag{2.28}$$

$$= \Gamma(2, \log(1/\delta)) \tag{2.29}$$

$$\geq \log(1/\delta)\delta \quad \text{for all } \delta \in (0,1), \tag{2.30}$$

where (2.30) follows from [33, Equation (6)], which implies that $\tilde{\mu}$ is not $|\cdot|$-subregular of dimension 1.

The following result allows to deduce subregularity/superregularity of product measures from subregularity/superregularity of their constituent measures.

Proposition 2.1 *Fix $k \in (0,\infty)$. For $i = 1,\dots,\ell$, let $(\mathcal{X}_i, \mathscr{X}_i)$ and $(\mathcal{Y}_i, \mathscr{Y}_i)$ be measurable spaces equipped with the distortion function $\rho_i \colon \mathcal{X}_i \times \mathcal{Y}_i \to [0,\infty]$. Consider the weighted distortion function*

$$\rho_{(\ell)}((x_1,\dots,x_\ell),(y_1,\dots,y_\ell)) = \sum_{i=1}^{\ell} \alpha_i \rho_i(x_i,y_i) \tag{2.31}$$

on $(\mathcal{X}_1 \times \cdots \times \mathcal{X}_\ell, \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_\ell)$, where $\alpha_i \in (0,\infty)$ for $i = 1,\dots,\ell$. Then, the following holds:

(i) Suppose that, for $i = 1,\dots,\ell$, μ_i is a σ-finite $\rho_i^{1/k}$-subregular measure on $(\mathcal{X}_i, \mathscr{X}_i)$ of dimension $m_i \in (0,\infty)$ with subregularity constants $c_i \in (0,\infty)$ and $\delta_i \in (0,\infty]$. Then, $\mu^{(\ell)} := \mu_1 \otimes \cdots \otimes \mu_\ell$ satisfies the subregularity condition

$$\mu^{(\ell)}\left(\mathcal{B}_{\rho_{(\ell)}^{1/k}}\left(y^{(\ell)},\delta\right)\right) \leq c_{(\ell)}\delta^{m_{(\ell)}} \tag{2.32}$$

for all $y^{(\ell)} \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_\ell$ and $\delta \in (0,\delta_{(\ell)})$ with subregularity dimension $m_{(\ell)} = \sum_{i=1}^{\ell} m_i$ and subregularity constants

$$c_{(\ell)} = \frac{\prod_{i=1}^{\ell}\Gamma\left(1+\frac{m_i}{k}\right)}{\Gamma\left(1+\sum_{i=1}^{\ell}\frac{m_i}{k}\right)} \times \prod_{i=1}^{\ell}\frac{c_i}{\alpha_i^{\frac{m_i}{k}}} \tag{2.33}$$

and

$$\delta_{(\ell)} = \min\left\{\alpha_1^{1/k}\delta_1,\dots,\alpha_\ell^{1/k}\delta_\ell\right\}. \tag{2.34}$$

(ii) *Suppose that, for $i=1,\dots,\ell$, ν_i is a σ-finite $\rho_i^{1/k}$-superregular measure on $(\mathcal{Y}_i,\mathscr{Y}_i)$ of dimension $m_i\in(0,\infty)$ with superregularity constants $b_i\in(0,\infty)$ and $\delta_i\in(0,\infty]$. Then, $\nu^{(\ell)}:=\nu_1\otimes\cdots\otimes\nu_\ell$ satisfies the superregularity condition*

$$\nu^{(\ell)}\left(\widetilde{\mathcal{B}}_{\rho_{(\ell)}^{1/k}}\left(x^{(\ell)},\delta\right)\right)\geq b_{(\ell)}\delta^{m_{(\ell)}} \tag{2.35}$$

for all $x^{(\ell)}\in\mathcal{X}_1\times\cdots\times\mathcal{X}_\ell$ and $\delta\in(0,\delta_{(\ell)})$ with superregularity dimension $m_{(\ell)}=\sum_{i=1}^{\ell}m_i$ and superregularity constants

$$b_{(\ell)}=\frac{\prod_{i=1}^{\ell}\Gamma\left(1+\frac{m_i}{k}\right)}{\Gamma\left(1+\sum_{i=1}^{\ell}\frac{m_i}{k}\right)}\times\prod_{i=1}^{\ell}\frac{b_i}{\alpha_i^{\frac{m_i}{k}}} \tag{2.36}$$

and

$$\delta_{(\ell)}=\min\left\{\alpha_1^{1/k}\delta_1,\dots,\alpha_\ell^{1/k}\delta_\ell\right\}. \tag{2.37}$$

Proof See Appendix A. □

The weighted distortion function $\rho_{(\ell)}$ in (2.31) covers the following important special cases:

- If $\alpha_i=1/\ell$ and $\rho_i=\sigma$ for $i=1,\dots,\ell$, then $\rho_{(\ell)}$ equals the average distortion function of length ℓ in (1.2).
- If $\alpha_i=1$ and $\rho_i(x,y)=\|x-y\|_k^k$ for $i=1,\dots,\ell$, then

$$\rho_{(\ell)}((x_1,\dots,x_\ell),(y_1,\dots,y_\ell))=\|(x_1-y_1,\dots,x_\ell-y_\ell)\|_k^k. \tag{2.38}$$

For illustration purposes, we show that the bounds in Proposition 2.1 are tight enough to recover (1.34) using regularity of Lebesgue measure $\mathscr{L}^\ell$ on $\mathbb{R}^\ell$. To this end, note that [38, Corollary 6.7]

$$\mathscr{L}^\ell=\underbrace{\mathscr{L}^1\otimes\cdots\otimes\mathscr{L}^1}_{\ell\text{ times}}. \tag{2.39}$$

Moreover, $\mathscr{L}^1(-\delta,\delta)=2\delta$ for all $\delta\in(0,\infty)$ implies that $\mathscr{L}^1$ is $|\cdot|$-regular of dimension 1 with regularity constants $b=c=2$ and $\delta_0=\infty$. Application of Items (i) and (ii) of Proposition 2.1 with $\alpha_i=1$, $\rho_i=|\cdot|^k$, $m_i=1$, $b_i=c_i=2$, and $\delta_i=\infty$, all for $i=1,\dots,\ell$, yields

$$\mathscr{L}^\ell(\mathcal{B}_{\|\cdot\|_k}(x,\delta))=\frac{\Gamma^\ell\left(1+\frac{1}{k}\right)}{\Gamma\left(1+\frac{\ell}{k}\right)}2^\ell\delta^\ell\quad\text{for all }x\in\mathbb{R}^\ell\text{ and }k,\delta\in(0,\infty), \tag{2.40}$$

which recovers the well-known volume formula for $\|\cdot\|_k$-balls in $\mathbb{R}^\ell$ (see, e.g., [41, Theorem 5]). Particularized to $k=2$ and using $\Gamma(3/2)=\sqrt{\pi}/2$, this yields (1.34).

Note that for the Hausdorff measures, a factorization akin to that for Lebesgue measure in (2.39) is not possible in general. There is, however, an important exception. Concretely, consider an $\mathscr{H}^s$-rectifiable set $\mathcal{E} \subseteq \mathbb{R}^m$ and let $\mathcal{F} \subseteq \mathbb{R}^n$ be such that $\mathcal{F} = f(\mathcal{A})$ with $f\colon \mathbb{R}^p \to \mathbb{R}^n$ a Lipschitz mapping and $\mathcal{A} \subseteq \mathbb{R}^p$ compact. Then, $\mathcal{F}$ is $\mathscr{H}^p$-rectifiable by construction and we have [20, Theorem 3.2.23]

$$\mathscr{H}^{s+p}|_{\mathcal{E}\times\mathcal{F}} = \mathscr{H}^s|_{\mathcal{E}} \otimes \mathscr{H}^p|_{\mathcal{F}}. \tag{2.41}$$

The specific form of the $\mathscr{H}^p$-rectifiable set $\mathcal{F}$ is required as $\mathscr{H}^p$-rectifiability alone is not enough for (2.41) to hold [20, Remark 3.2.24]. Applying (2.41) inductively for a set $\mathcal{F}$ as above, we find that

$$\mathscr{H}^{\ell p}|_{\mathcal{F}^\ell} = \underbrace{\mathscr{H}^p|_{\mathcal{F}} \otimes \cdots \otimes \mathscr{H}^p|_{\mathcal{F}}}_{\ell \text{ times}} \quad \text{for all } \ell \in \mathbb{N}. \tag{2.42}$$

This factorization now allows the application of Proposition 2.1 (with $\alpha_i = 1$ and $\rho_i = \|\cdot\|^k$ for $i = 1,\dots,\ell$) to get upper/lower bounds on $\mathscr{H}^{\ell p}|_{\mathcal{F}^\ell}(\mathcal{B}_{\|\cdot\|_k}(x,\delta))$ for all $x \in \mathbb{R}^{\ell p}$ and $\delta \in (0,\infty)$ in terms of upper/lower bounds on $\mathscr{H}^p|_{\mathcal{F}}(\mathcal{B}_{\|\cdot\|_k}(x,\delta))$ for all $x \in \mathbb{R}^p$ and $\delta \in (0,\infty)$ upon noting that σ-finiteness of $\mathscr{H}^p|_{\mathcal{F}}$ is guaranteed by [50, Lemma 15.5].

3. A Lower Bound on the Rate-Distortion Function

The best-known lower bounds on $R_X(D)$ in (1.5) are the Shannon lower bounds for i) discrete X of finite entropy and with distortion function $\rho\colon \mathcal{X} \times \mathcal{Y} \to [0,\infty]$ such that $\sum_{x\in\mathcal{X}} e^{-s\rho(x,y)}$ is independent of y for all $s \in (0,\infty)$ [28, Lemma 4.3.1] and ii) continuous X of finite differential entropy and with a difference distortion function [28, Equation (4.6.1)]. We next present a general Shannon-type lower bound that applies to all random variables X taking values in a σ-finite measure space $(\mathcal{X}, \mathscr{X}, \mu)$, where μ is such that $\mu_X \ll \mu$ and $|h_\mu(X)| < \infty$, but arbitrary otherwise.

Proposition 3.1 *Let X be a random variable taking values in the σ-finite measure space $(\mathcal{X}, \mathscr{X}, \mu)$. Assume that $\mu_X \ll \mu$ with $|h_\mu(X)| < \infty$, let $(\mathcal{Y}, \mathscr{Y})$ be a measurable space, and consider the distortion function $\rho\colon \mathcal{X} \times \mathcal{Y} \to [0,\infty]$. Then, $R_X(D) \geq R_X^{SLB}(D)$ for all $D \in (0,\infty)$, where*

$$R_X^{SLB}(D) = h_\mu(X) - \inf_{s\geq 0}\big(sD + \log(g(s))\big) \tag{3.1}$$

with

$$g(s) = \sup_{y\in\mathcal{Y}} \int e^{-s\rho(x,y)} \mathrm{d}\mu(x) \quad \textit{for all } s \in (0,\infty). \tag{3.2}$$

Proof See Appendix B. □

Proposition 3.1 covers the following important special cases:

- For discrete X of finite entropy, μ the counting measure, and $\sum_{x\in\mathcal{X}} e^{-s\rho(x,y)}$ independent of y for all $s \in (0,\infty)$, $R_X^{\mathrm{SLB}}(D)$ in (3.1) recovers the Shannon lower bound for discrete random variables reported in [28, Lemma 4.3.1]; in this case $h_\mu(X)$ equals the Shannon entropy of X.
- For X continuous, μ the Lebesgue measure, $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$, and ρ a difference distortion function, $R_X^{\mathrm{SLB}}(D)$ in (3.1) recovers the Shannon lower bound for continuous random variables [28, Equation

(4.6.1)]; in this case $h_\mu(X)$ equals the differential entropy of X. This lower bound can be evaluated explicitly for $\rho(x,y) = \|x-y\|^k$ with $\|\cdot\|$ a semi-norm and $k \in (0,\infty)$, resulting in the classical form of the Shannon lower bound [64, Section VI]

$$R_X^{\mathrm{SLB}}(D) = h_\mu(X) + F_{d,k,\kappa_d(\|\cdot\|)}(D), \tag{3.3}$$

where, for $m,k,c,D \in (0,\infty)$, we set

$$F_{m,k,c}(D) = \log\left(\frac{\left(\frac{m}{kD}\right)^{\frac{m}{k}}}{c\Gamma\left(1+\frac{m}{k}\right)}\right) - \frac{m}{k}. \tag{3.4}$$

- For the class of m-rectifiable random variables [40, Definition 11], Proposition 3.1 recovers [40, Theorem 55]. In this case, $h_\mu(X)$ is the m-dimensional entropy [40, Definition 18] of the m-rectifiable random variable X.

The explicit expression in (3.3) is made possible by a simplification of $g(s)$ exploiting that i) ρ is a difference distortion function and ii) Lebesgue measure is translation invariant. Specifically, we have

$$g(s) = \sup_{y\in\mathcal{Y}} \int e^{-s\rho(x-y,0)}\,\mathrm{d}\mu(x) \tag{3.5}$$

$$= \int e^{-s\rho(x,0)}\,\mathrm{d}\mu(x), \tag{3.6}$$

which can be evaluated explicitly for $\rho(x,y) = \|x-y\|^k$ with $\|\cdot\|$ a semi-norm and $k \in (0,\infty)$ [64, Section II]. For general distortion functions ρ and general measures μ, a simplification of $g(s)$ and hence of the lower bound in (3.1) does not seem possible. It turns out, however, that subregular measures permit an explicit upper bound on $g(s)$, which in turn leads to the following explicit lower bound on $R_X^{\mathrm{SLB}}(D)$.

Theorem 3.1 *Let X be a random variable taking values in the σ-finite measure space $(\mathcal{X},\mathscr{X},\mu)$. Assume that $\mu_X \ll \mu$ with $\left|h_\mu(X)\right| < \infty$, let $(\mathcal{Y},\mathscr{Y})$ be a measurable space, and consider the distortion function $\rho\colon \mathcal{X}\times\mathcal{Y} \to [0,\infty]$. Further, let $k \in (0,\infty)$ and suppose that μ is $\rho^{1/k}$-subregular of dimension $m \in (0,\infty)$ with subregularity constants $c \in (0,\infty)$ and $\delta_0 \in (0,\infty]$. Finally, assume that $\mu(\mathcal{X}) = 1$ if $\delta_0 < \infty$. Then,*

$$R_X^{SLB}(D) \geq R_X^{L}(D) \quad \textit{for all } D \in (0,\infty) \tag{3.7}$$

with $R_X^L(D)$ defined as follows:

(i) If $\delta_0 = \infty$, then

$$R_X^L(D) = h_\mu(X) + F_{m,k,c}(D) \tag{3.8}$$

with $F_{m,k,c}(D)$ as defined in (3.4).

(ii) If $\delta_0 < \infty$, then

$$R_X^L(D) = h_\mu(X) - \frac{m}{k} - \log\left(c\left(\frac{m}{kD}\right)^{-\frac{m}{k}}\Gamma\left(1+\frac{m}{k}\right) + e^{-\frac{m\delta_0^k}{kD}}\right) \tag{3.9}$$

with

$$R_X^L(D) < h_\mu(X) + F_{m,k,c}(D) \quad \textit{for all } D \in (0,\infty) \tag{3.10}$$

and

$$\lim_{D\to 0}(R_X^L(D) - h_\mu(X) - F_{m,k,c}(D)) = 0. \tag{3.11}$$

Proof See Appendix C. □

The proof of Theorem 3.1 reveals that if μ satisfies the subregularity condition with equality for $\delta_0 = \infty$, then the lower bound in (3.7) holds with equality, i.e., we have

$$R_X^{\mathrm{SLB}}(D) = R_X^{\mathrm{L}}(D) = h_\mu(X) + F_{m,k,c}(D) \quad \text{for all } D \in (0,\infty) \tag{3.12}$$

with $F_{m,k,c}(D)$ as defined in (3.4). In particular, this applies to Lebesgue measure so that Theorem 3.1 generalizes the classical Shannon lower bound (3.3) to arbitrary distortion functions.

We finally note that if X satisfies the assumptions of Theorem 3.1, then the lower R-D dimension of order $k \in (0,\infty)$ is lower-bounded by the subregularity dimension. The corresponding formal statement is as follows.

Corollary 3.1 *Under the assumptions of Theorem 3.1, we have*

$$\underline{\dim}_R(X) \geq m. \tag{3.13}$$

Proof The lower bound on $\underline{\dim}_{\mathrm{R}}(X)$ follows from

$$\begin{aligned}
\underline{\dim}_{\mathrm{R}}(X) &= \liminf_{D\to 0}\frac{R_X\left(D^k\right)}{\log(1/D)} && (3.14)\\
&\geq \liminf_{D\to 0}\frac{R_X^{\mathrm{SLB}}\left(D^k\right)}{\log(1/D)} && (3.15)\\
&\geq \lim_{D\to 0}\frac{R_X^{\mathrm{L}}\left(D^k\right)}{\log(1/D)} && (3.16)\\
&= \lim_{D\to 0}\frac{F_{m,k,c}\left(D^k\right)}{\log(1/D)} && (3.17)\\
&= m, && (3.18)
\end{aligned}$$

where (3.15) is by Proposition 3.1 and in (3.16) we applied Theorem 3.1. □

4. Analysis of the n-th Quantization Error

We first derive a lower bound $L_n(X)$ on $V_n(X)$ for random variables X of distribution $\mu_X \ll \mu$, where μ is a σ-finite $\rho^{1/k}$-subregular measure satisfying $\|\mathrm{d}\mu_X/\mathrm{d}\mu\|_{p/(p-1)}^{(\mu)} < \infty$ with $p \in [1,\infty)$. The corresponding formal statement is as follows.

Theorem 4.1 *Let X be a random variable taking values in the σ-finite measure space $(\mathcal{X}, \mathscr{X}, \mu)$. Assume that $\mu_X \ll \mu$ and*

$$\Sigma_p(X) := \|\mathrm{d}\mu_X/\mathrm{d}\mu\|_{p/(p-1)}^{(\mu)} < \infty \quad \text{with } p \in [1,\infty), \tag{4.1}$$

let $(\mathcal{Y}, \mathscr{Y})$ be a measurable space, and consider the distortion function $\rho\colon \mathcal{X} \times \mathcal{Y} \to [0,\infty]$. Further, let $k \in (0,\infty)$ and suppose that μ is $\rho^{1/k}$-subregular of dimension $m \in (0,\infty)$ with subregularity constants $c \in (0,\infty)$ and $\delta_0 \in (0,\infty]$. Then, $V_n(X) \geq L_n(X)$ for all $n \in \mathbb{N}$, where

$$L_n(X) = \min\left\{ c^{-\frac{k}{m}} \Sigma_p^{-\frac{pk}{m}}(X) n^{-\frac{pk}{m}}, \delta_0^k \right\} \frac{m}{m+pk}. \tag{4.2}$$

Proof See Appendix D. □

Theorem 4.1 particularized to X taking values in a Riemannian manifold and such that μ_X itself is $\rho^{1/k}$-subregular with $\delta_0 = \infty$ recovers [37, Proposition 4.2]. We also note that the proof of [37, Proposition 4.2] uses methods from optimal transport theory [61], whereas our proof technique is based on elementary probability and measure theory.

We next use Theorem 4.1 to study the $n \to \infty$ asymptotics of $V_n(X)$ in terms of the lower k-th quantization dimension and the lower k-th quantization coefficient.

Corollary 4.1 *Under the assumptions of Theorem 4.1, the following statements hold:*

(i) $\underline{D}_k(X) \geq m/p$.
(ii) *If* $D_k(X) = m/p$, *then*

$$\underline{C}_k(X) \geq c^{-\frac{k}{m}} \Sigma_p^{-\frac{pk}{m}}(X) \frac{m}{m+pk} > 0. \tag{4.3}$$

Proof Item (i) follows from

$$\underline{D}_k(X) = \liminf_{n\to\infty} \frac{k\log(n)}{\log(1/V_n(X))} \tag{4.4}$$

$$\geq \lim_{n\to\infty} \frac{k\log(n)}{\log(1/L_n(X))} \tag{4.5}$$

$$= m/p, \tag{4.6}$$

where (4.5) is by Theorem 4.1. Item (ii) is by

$$\underline{C}_k(X) = \liminf_{n\to\infty} n^{\frac{pk}{m}} V_n(X) \tag{4.7}$$

$$\geq \lim_{n\to\infty} n^{\frac{pk}{m}} L_n(X) \tag{4.8}$$

$$= c^{-\frac{k}{m}} \Sigma_p^{-\frac{pk}{m}}(X) \frac{m}{m+pk}, \tag{4.9}$$

where (4.7) is by (1.15) combined with the assumption $D_k(X) = m/p$ and (4.8) follows from Theorem 4.1. □

Note that if $\|\mathrm{d}\mu_X/\mathrm{d}\mu\|^{(\mu)}_{p/(p-1)} < \infty$ for all $p \in (1,\infty)$, then Item (i) of Corollary 4.1 implies

$$\underline{D}_k(X) \geq \sup_{p\in(1,\infty)} \left\{\frac{m}{p}\right\} = m \tag{4.10}$$

even if $\Sigma_1(X) = \infty$, see, e.g., Example 2.2 with $\mu_X = \nu$ and $\mu = \mathscr{L}^1|_{\mathcal{X}}$.

We finally note that for the special case $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$, $k \in [1,\infty)$, and X of nonvanishing continuous part $\tilde{\mu}_X$ (in the Lebesgue decomposition of μ_X) satisfying $\|\mathrm{d}\tilde{\mu}_X/\mathrm{d}\mathscr{L}^d\|^{(\mu)}_{d/(d+k)} < \infty$, it is known that $D_k(X) = d$ and an expression for the k-th quantization coefficient is available [27, Theorem 6.2, Remark 6.3].

We next use our results in R-D theory to obtain lower bounds on the lower quantization dimension. To this end, we first establish that the k-th lower quantization dimension is lower-bounded by the lower R-D-dimension.

Lemma 4.1 *Let $(\mathcal{X},\mathscr{X})$ and $(\mathcal{Y},\mathscr{Y})$ be measurable spaces equipped with the distortion function $\rho\colon \mathcal{X}\times\mathcal{Y} \to [0,\infty]$. Further, let X be a random variable taking values in $(\mathcal{X},\mathscr{X})$. Then, $\underline{D}_k(X) \geq \underline{\dim}_R(X)$ for all $k \in (0,\infty)$.*

Proof See Appendix E. □

In the special case $\mathcal{X} \subseteq \mathcal{Y} = \mathbb{R}^d$, $\rho(x,y) = \|x-y\|^k$ with $\|\cdot\|$ a norm on $\mathbb{R}^d$, and $k \in [1,\infty)$, Lemma 4.1 particularizes to [27, Theorem 11.10].

As an immediate consequence of Lemma 4.1, we obtain the following lower bound on the k-th lower quantization dimension.

Corollary 4.2 *Let X be a random variable taking values in the σ-finite measure space $(\mathcal{X},\mathscr{X},\mu)$. Assume that $\mu_X \ll \mu$ with $\left|h_\mu(X)\right| < \infty$, let $(\mathcal{Y},\mathscr{Y})$ be a measurable space, and consider the distortion function $\rho\colon \mathcal{X}\times\mathcal{Y} \to [0,\infty]$. Further, let $k \in (0,\infty)$ and suppose that μ is $\rho^{1/k}$-subregular of dimension $m \in (0,\infty)$ with subregularity constants $c \in (0,\infty)$ and $\delta_0 \in (0,\infty]$. Finally, suppose that $\mu(\mathcal{X}) = 1$ if $\delta_0 < \infty$. Then, $\underline{D}_k(X) \geq m$.*

Proof Follows from Corollary 3.1 and Lemma 4.1. □

While the lower bound $L_n(X)$ on $V_n(X)$ in Theorem 4.1 is based on a subregular σ-finite measure μ on $(\mathcal{X},\mathscr{X})$, we next derive an upper bound on $V_n(X)$ that requires the existence of a finite superregular measure ν on $(\mathcal{Y},\mathscr{Y})$. In contrast to $L_n(X)$, which needs $\mu_X \ll \mu$ and as such depends on X, the upper bound U_n we obtain is universal in the sense of depending only on the spaces $(\mathcal{X},\mathscr{X})$ and $(\mathcal{Y},\mathscr{Y},\nu)$

and the distortion function ρ but not on the random variable X per se. To reflect this aspect, we will write U_n without the argument X.

Theorem 4.2 *Let $(\mathcal{X}, \mathscr{X})$ be a measurable space, let $(\mathcal{Y}, \mathscr{Y}, \nu)$ be a finite measure space with ν normalized according to $\nu(\mathcal{Y}) = 1$, and consider the distortion function $\rho\colon \mathcal{X} \times \mathcal{Y} \to [0,\infty]$. Further, let $k \in (0,\infty)$, suppose that ν is $\rho^{1/k}$-superregular of dimension $m \in (0,\infty)$ with superregularity constants $b \in (0,\infty)$ and $\delta_0 \in (0,\infty]$, and assume that*

$$\beta := \operatorname*{ess\,sup}_{y\in\mathcal{Y}} \sup_{x\in\mathcal{X}} \rho^{1/k}(x,y) < \infty \quad \text{if } \delta_0 < \infty. \tag{4.11}$$

Then, every X taking values in $(\mathcal{X}, \mathscr{X})$ satisfies $V_n(X) \le U_n$ for all $n \in \mathbb{N}$, where

$$U_n = \Gamma\left(1 + \frac{k}{m}\right)(bn)^{-\frac{k}{m}} \quad \text{if } \beta \le \delta_0 \tag{4.12}$$

and

$$U_n = \Gamma\left(1 + \frac{k}{m}\right)(bn)^{-\frac{k}{m}} + (\beta^k - \delta_0^k)e^{-bn\delta_0^m} \quad \text{if } \beta > \delta_0. \tag{4.13}$$

Proof See Appendix F. □

We now employ Theorem 4.2 to characterize the $n \to \infty$ asymptotics of $V_n(X)$ in terms of the upper k-th quantization dimension and the upper k-th quantization coefficient.

Corollary 4.3 *Under the assumptions of Theorem 4.2, the following statements hold:*

(i) $\overline{D}_k(X) \le m$.
(ii) *If $D_k(X) = m$, then*

$$\overline{C}_k(X) \le \Gamma\left(1 + \frac{k}{m}\right) b^{-\frac{k}{m}}. \tag{4.14}$$

Proof Item (i) follows from

$$\overline{D}_k(X) = \limsup_{n\to\infty} \frac{k\log(n)}{\log(1/V_n(X))} \tag{4.15}$$

$$\le \lim_{n\to\infty} \frac{k\log(n)}{\log(1/U_n)} \tag{4.16}$$

$$= m, \tag{4.17}$$

where (4.16) is by Theorem 4.2. Item (ii) is by

$$\overline{C}_k(X) = \limsup_{n\to\infty} n^{\frac{k}{m}} V_n(X) \tag{4.18}$$

$$\leq \lim_{n\to\infty} n^{\frac{k}{m}} U_n \tag{4.19}$$

$$= \Gamma\left(1+\frac{k}{m}\right) b^{-\frac{k}{m}}, \tag{4.20}$$

where (4.18) is by (1.16) combined with the assumption $D_k(X) = m$ and (4.19) follows from Theorem 4.2. □

Combining Corollaries 4.1–4.3, we get the following result on the k-th quantization dimension.

Corollary 4.4 *Let X be a random variable taking values in the σ-finite measure space $(\mathcal{X}, \mathscr{X}, \mu)$ with $\mu_X \ll \mu$, let $(\mathcal{Y}, \mathscr{Y}, \nu)$ be a finite measure space with ν normalized according to $\nu(\mathcal{Y}) = 1$, and consider the distortion function $\rho\colon \mathcal{X}\times\mathcal{Y} \to [0,\infty]$. Further, let $k, m, c, b \in (0,\infty)$, $\delta_0 \in (0,\infty]$, and suppose that i) μ is $\rho^{1/k}$-subregular of dimension m with subregularity constants c and δ_0 and ii) ν is $\rho^{1/k}$-superregular of dimension m with superregularity constants b and δ_0. Finally, assume that*

$$\operatorname*{ess\,sup}_{y\in\mathcal{Y}} \sup_{x\in\mathcal{X}} \rho^{1/k}(x,y) < \infty \quad \textit{if } \delta_0 < \infty. \tag{4.21}$$

Then, the following properties hold:

(i) If $\|\mathrm{d}\mu_X/\mathrm{d}\mu\|_\infty^{(\mu)} < \infty$, then

$$D_k(X) = m. \tag{4.22}$$

(ii) If i) $\left|h_\mu(X)\right| < \infty$ and ii) $\mu(\mathcal{X}) = 1$ for $\delta_0 < \infty$, then

$$\underline{\dim}_R(X) = D_k(X) = m. \tag{4.23}$$

Proof If $\|\mathrm{d}\mu_X/\mathrm{d}\mu\|_\infty^{(\mu)} < \infty$, then Item (i) of Corollary 4.1 yields $\underline{D}_k(X) \geq m$. Similarly, if i) $\left|h_\mu(X)\right| < \infty$ and ii) $\mu(\mathcal{X}) = 1$ for $\delta_0 < \infty$, then Corollary 4.2 establishes $\underline{D}_k(X) \geq m$. Finally, Item (i) of Corollary 4.3 yields $\overline{D}_k(X) \leq m$ in both cases. Taken together these results establish $D_k(X) = m$ in Items (i) and (ii) as desired. The identity $\underline{\dim}_R(X) = m$ follows from Corollary 3.1, Lemma 4.1, and $D_k(X) = m$. □

Corollary 4.4 does not hold in general when (4.21) is not satisfied. The following example, which is a slight modification of [27, Example 6.4], illustrates this by constructing a random variable with $D_k(X) = \infty$.

Example 4.1. For every $\ell \in \mathbb{N}$ with $\ell \geq 2$, set $\mathcal{I}_\ell = [2^\ell, 2^{\ell+1})$, $\mathcal{J}_\ell = [(4/3)2^\ell, (5/3)2^\ell)$, and

$$\mu_\ell = \frac{c\mathscr{L}^1|_{\mathcal{J}_\ell}}{\mathscr{L}^1(\mathcal{J}_\ell)2^{k\ell}\ell\log^2(\ell)} \tag{4.24}$$

with $c = (\sum_{\ell=2}^\infty \frac{1}{2^{k\ell}\ell\log^2(\ell)})^{-1}$. Set $\mathcal{X} = \mathcal{Y} = (0,\infty)$ and $\mu = \mathscr{L}^1|_{\mathcal{X}}$, and consider the distortion function $\rho\colon \mathcal{X}\times\mathcal{Y} \to [0,\infty)$, $\rho(x,y) = |x-y|^k$ with $k \in (0,\infty)$. Further, let X be of distribution $\mu_X = \sum_{\ell=2}^\infty \mu_\ell$.

Next, fix $n \in \mathbb{N}$ and $f \in \mathcal{F}_n$ arbitrarily and set $\mathcal{I} = \{\ell \in \mathbb{N} : \ell \geq 2 \text{ and } f([4,\infty)) \cap \mathcal{I}_\ell = \emptyset\}$. Then, we have

$$\mathbb{E}\Big[\big|f(X)-X\big|^k\Big] = \sum_{\ell=2}^{\infty} \int_{\mathcal{J}_\ell} \big|f(x)-x\big|^k \, \mathrm{d}\mu_\ell(x) \tag{4.25}$$

$$\geq \sum_{\ell \in \mathcal{I}} \int_{\mathcal{J}_\ell} \big|f(x)-x\big|^k \, \mathrm{d}\mu_\ell(x) \tag{4.26}$$

$$\geq \sum_{\ell \in \mathcal{I}} \frac{2^{k\ell}}{3^k} \mu_\ell(\mathcal{J}_\ell) \tag{4.27}$$

$$= \frac{c}{3^k} \sum_{\ell \in \mathcal{I}} \frac{1}{\ell \log^2(\ell)} \tag{4.28}$$

$$\geq \frac{c}{3^k} \sum_{\ell=n+2}^{\infty} \frac{1}{\ell \log^2(\ell)} \tag{4.29}$$

$$\geq \frac{c}{3^k} \int_{n+2}^{\infty} \frac{1}{x \log^2(x)} \, \mathrm{d}x \tag{4.30}$$

$$= \frac{c}{3^k \log(n+2)}, \tag{4.31}$$

where (4.27) follows from $\mathcal{J}_\ell \subseteq \mathcal{I}_\ell$ and $f([4,\infty)) \cap \mathcal{I}_\ell = \emptyset$ for all $\ell \in \mathcal{I}$. As $n \in \mathbb{N}$ and $f \in \mathcal{F}_n$ were arbitrary, we can conclude that

$$V_n(X) \geq \frac{c}{3^k \log(n+2)} \quad \text{for all } n \in \mathbb{N}, \tag{4.32}$$

which in turn yields $D_k(X) = \infty$.

For the particular case $\mathcal{X} \subseteq \mathcal{Y} = \mathbb{R}^d$ with $\mathcal{X}$ compact, distortion function $\rho \colon \mathcal{X} \times \mathcal{Y} \to [0,\infty]$, $\rho(x,y) = \|x-y\|^k$, where $\|\cdot\|$ is a norm on $\mathbb{R}^d$ and $k \in [1,\infty)$, and $\rho^{1/k}$-regular ν of dimension m satisfying $\operatorname{supp}(\nu) = \mathcal{X}$ and $\nu(\mathcal{X}) = 1$, combining Corollary 4.4 for this ν and $\mu = \nu|_{\mathcal{X}}$ with [27, Theorem 12.11] allows us to conclude that the k-th quantization dimension of X with $\mu_X \ll \mu$ equals the Hausdorff dimension of μ, i.e., we have

$$D_k(X) = m = \dim_{\mathrm{H}}(\mu) \leq \dim_{\mathrm{H}}(\mathcal{X}) \tag{4.33}$$

with $\dim_{\mathrm{H}}(\mu)$ as defined in (1.10), provided that $\|\mathrm{d}\mu_X/\mathrm{d}\mu\|_\infty^{(\mu)} < \infty$ or $\big|h_\mu(X)\big| < \infty$. We note that (4.33) particularized to $\mu = \mu_X$ recovers [27, Theorem 12.18]. The inequality $D_k(X) \leq \dim_{\mathrm{H}}(\mathcal{X})$ becomes particularly relevant when the ambient space dimension d is significantly larger than the Hausdorff dimension of the set $\mathcal{X}$ containing the data to be compressed (here described by a random variable taking values in $\mathcal{X}$).

While the upper bound U_n in Theorem 4.2 is sharp enough to establish $D_k(X) = m$ in Corollary 4.4, it is, in general, too weak to yield good upper bounds on $\overline{C}_k(X)$, simply as U_n does not depend on the specific random variable X taking values in $(\mathcal{X}, \mathscr{X})$. We next derive an X-dependent upper bound on $V_n(X)$ applicable to $\mathcal{X} = \mathcal{Y} \subseteq \mathbb{R}^d$, which requires, however, stronger technical assumptions than those imposed in Theorem 4.2.

Theorem 4.3 *Let $\mathcal{Y} = \mathcal{X} \subseteq \mathbb{R}^d$ be Borel and consider the distortion function $\rho = \omega^k|_{\mathcal{X}\times\mathcal{X}}$, where $k \in (0,\infty)$ and ω is a metric on $\mathbb{R}^d$. Suppose that X takes values in $\mathcal{X}$ and assume that there exists a finite measure ν on $\mathcal{X}$ normalized according to $\nu(\mathcal{X}) = 1$ with $\nu \ll \mu_X$ and*

$$\left\| \frac{\mathrm{d}\nu}{\mathrm{d}\mu_X} \right\|_\infty^{(\mu_X)} < \infty. \tag{4.34}$$

Further, let ν be $\rho^{1/k}$-superregular of dimension $m \in (k,\infty)$ with superregularity constants $b \in (0,\infty)$ and $\delta_0 \in (0,\infty]$ and assume that

$$\beta := \sup_{x,y\in\mathcal{X}} \omega(x,y) < \infty \quad \text{if } \delta_0 < \infty. \tag{4.35}$$

Then, we have

$$\limsup_{n\to\infty} n^{\frac{k}{m}} V_n(X) \le \Omega_{k/m}(X)\Gamma\left(1+\frac{k}{m}\right) b^{-\frac{k}{m}} \tag{4.36}$$

with

$$\Omega_\alpha(X) = \mathbb{E}\left[\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu_X}(X)\right)^\alpha\right] \le 1 \quad \text{for all } \alpha \in (0,1) \tag{4.37}$$

and strict inequality in (4.37) unless $\mu_X = \nu$. If, in addition, $D_k(X) = m$, then

$$\overline{C}_k(X) \le \Omega_{k/m}(X)\Gamma\left(1+\frac{k}{m}\right) b^{-\frac{k}{m}}. \tag{4.38}$$

Proof See Appendix G. □

Note that (4.38) improves upon the upper bound on $\overline{C}_k(X)$ in Item (ii) of Corollary 4.3 as the inequality in (4.37) is strict for $\alpha = k/m < 1$ unless $\mu_X = \nu$.

We close this section by stating a technical result needed later in the paper.

Lemma 4.2 *Let $(\mathcal{X},\mathscr{X})$ and $(\mathcal{Y},\mathscr{Y})$ be measurable spaces and consider the distortion function $\rho\colon \mathcal{X}\times\mathcal{Y} \to [0,\infty]$. For X taking values in $(\mathcal{X},\mathscr{X})$ and $m,k \in (0,\infty)$, the following properties hold:*

(i) *If $\liminf_{n\to\infty} n^{\frac{k}{m}} V_n(X) > 0$, then $\underline{D}_k(X) \ge m$.*
(ii) *If $\limsup_{n\to\infty} n^{\frac{k}{m}} V_n(X) < \infty$, then $\overline{D}_k(X) \le m$.*

Proof See Appendix H. □

5. Linking R-D Theory and Quantization

The purpose of this section is to establish, for i.i.d. sequences $(X_i)_{i\in\mathbb{N}}$, a relationship between R-D theory and quantization that is both conceptual and quantitative. To this end, recall the setup described in Section 1.1, where $(\mathcal{A},\mathscr{A})$ and $(\mathcal{B},\mathscr{B})$ are measurable spaces equipped with the distortion function $\sigma\colon \mathcal{A}\times\mathcal{B}\to[0,\infty]$, and let $(X_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables taking values in $(\mathcal{A},\mathscr{A})$. For a given rate $R\in[0,\infty)$ and distortion $D\in[0,\infty)$, the pair (R,D) is said to be ℓ-achievable if there exists a source code $g_{(\ell)}\colon \mathcal{A}^\ell\to\mathcal{B}^\ell$ of length ℓ with $\left|g_{(\ell)}(\mathcal{A}^\ell)\right|\le\lfloor e^{\ell R}\rfloor$ and expected average distortion

$$\mathbb{E}\left[\sigma_{(\ell)}\Big((X_1,\dots,X_\ell),g_{(\ell)}(X_1,\dots,X_\ell)\Big)\right]\le D, \tag{5.1}$$

where

$$\sigma_{(\ell)}((x_1,\dots,x_\ell),(y_1,\dots,y_\ell))=\frac{1}{\ell}\sum_{i=1}^{\ell}\sigma(x_i,y_i) \tag{5.2}$$

is the average distortion function of length ℓ, and

$$R_{(\ell)}(D)=\inf\{R\in[0,\infty):(R,D)\text{ is }\ell\text{-achievable}\}. \tag{5.3}$$

The pair (R,D) is said to be achievable if there exists an $\ell\in\mathbb{N}$ such that (R,D) is ℓ-achievable. Finally, we set

$$R(D)=\inf\{R\in[0,\infty):(R,D)\text{ is achievable}\}. \tag{5.4}$$

First, note that $R<R(D)$ in R-D theory implies $V_{\lfloor e^R\rfloor}(X_1)\ge D$ in quantization. In fact, pairs (R,D) with $R<R(D)$ are not 1-achievable. As the set of source codes of length 1 with $|f(\mathcal{X})|\le\lfloor e^R\rfloor$ is $\mathcal{F}_{\lfloor e^R\rfloor}(\mathcal{X},\mathcal{Y})$, we can therefore conclude that $R<R(D)$ implies $V_{\lfloor e^R\rfloor}(X_1)\ge D$.

Interestingly, under subregularity for $\delta_0=\infty$, reasoning in the opposite direction is also possible. Concretely, one can get a lower bound on $R(D)$ based on results from quantization. This will be accomplished by applying Theorem 4.1 to vectors $X=(X_1,\dots,X_\ell)$ and then taking the limit $\ell\to\infty$. We proceed to establish lower bounds on $R_{(\ell)}(D)$ for all $\ell\in\mathbb{N}$.

Theorem 5.1 *Let $(X_i)_{i\in\mathbb{N}}$ be an i.i.d. sequence of random variables taking values in the σ-finite measure space $(\mathcal{A},\mathscr{A},\mu)$ satisfying $\mu_{X_1}\ll\mu$ and $\|\mathrm{d}\mu_{X_1}/\mathrm{d}\mu\|_{p/(p-1)}^{(\mu)}<\infty$ with $p\in[1,\infty)$, let $(\mathcal{B},\mathscr{B})$ be a measurable space, and consider the distortion function $\sigma\colon\mathcal{A}\times\mathcal{B}\to[0,\infty]$. Further, let $k\in(0,\infty)$ and suppose that μ is $\sigma^{1/k}$-subregular of dimension $m\in(0,\infty)$ with subregularity constants $c\in(0,\infty)$ and $\delta_0\in(0,\infty]$. Then, we have $R_{(\ell)}(D)\ge\widetilde{R}_{(\ell)}(D)$ for all $D\in(0,D_{(\ell)})$, where*

$$\widetilde{R}_{(\ell)}(D)=\frac{m}{pk}\log\left(\frac{\ell m}{(\ell m+pk)D}\right)-\log(d_{(\ell)}) \tag{5.5}$$

with

$$d_{(\ell)}=\frac{\Gamma\Big(1+\frac{m}{pk}\Big)\ell^{\frac{m}{pk}}}{\Gamma^{\frac{1}{\ell}}\Big(1+\frac{\ell m}{pk}\Big)}c^{\frac{1}{p}}\|\mathrm{d}\mu_{X_1}/\mathrm{d}\mu\|_{p/(p-1)}^{(\mu)} \tag{5.6}$$

and

$$D_{(\ell)} = \frac{\delta_0^k}{\ell} \frac{\ell m}{\ell m + pk}. \tag{5.7}$$

Moreover, $\widetilde{R}_{(\ell)}(D)$ *in* (5.5) *is strictly monotonically decreasing in* ℓ *with*

$$\lim_{\ell\to\infty} \widetilde{R}_{(\ell)}(D) = -\log\Big(\|\mathrm{d}\mu_{X_1}/\mathrm{d}\mu\|^{(\mu)}_{p/(p-1)}\Big) + F_{m/p,k,c^{1/p}}(D) \tag{5.8}$$

for all $D \in (0,\infty)$ *and with* $F_{m,k,c}(D)$ *as defined in* (3.4).

Proof See Appendix I. □

Now, if μ in Theorem 5.1 is $\sigma^{1/k}$-subregular with $\delta_0 = \infty$, then $D_{(\ell)} = \infty$ for all $\ell \in \mathbb{N}$ so that

$$R_{(\ell)}(D) \geq \widetilde{R}_{(\ell)}(D) \geq -\log\Big(\|\mathrm{d}\mu_{X_1}/\mathrm{d}\mu\|^{(\mu)}_{p/(p-1)}\Big) + F_{m/p,k,c^{1/p}}(D) \tag{5.9}$$

for all $D \in (0,\infty)$ and $\ell \in \mathbb{N}$. We conclude that $R(D) \geq \widetilde{R}^{\mathrm{L}}_{X_1}(D)$ for all $D \in (0,\infty)$ with

$$\widetilde{R}^{\mathrm{L}}_{X_1}(D) = -\log\Big(\|\mathrm{d}\mu_{X_1}/\mathrm{d}\mu\|^{(\mu)}_{p/(p-1)}\Big) + F_{m/p,k,c^{1/p}}(D). \tag{5.10}$$

Alternatively, suppose that the assumptions of Theorem 3.1 (applied to $X = X_1$) are satisfied for $\delta_0 = \infty$. In particular, we need $\big|h_\mu(X_1)\big| < \infty$. Then, the R-D function $R_{X_1}(D)$ is lower-bounded according to $R_{X_1}(D) \geq R^{\mathrm{L}}_{X_1}(D)$ with (see (3.8))

$$R^{\mathrm{L}}_{X_1}(D) = h_\mu(X_1) + F_{m,k,c}(D) \quad \text{for all } D \in (0,\infty). \tag{5.11}$$

Since $R_{X_1}(D) = R_{(X_i)_{i\in\mathbb{N}}}(D)$ thanks to the i.i.d. assumption, the converse [6, Theorem 7.2.5] implies that achievability of (R,D) requires $R \geq R_{X_1}(D)$, which in turn yields $R(D) \geq R^{\mathrm{L}}_{X_1}(D)$ for all $D \in (0,\infty)$.

For $p = 1$, the lower bound $\widetilde{R}^{\mathrm{L}}_{X_1}(D)$ in (5.10) differs from $R^{\mathrm{L}}_{X_1}(D)$ in (5.11) in the term $-\log\Big(\|\mathrm{d}\mu_{X_1}/\mathrm{d}\mu\|^{(\mu)}_\infty\Big)$, which is replaced by $h_\mu(X_1)$. In terms of applicability, validity of $R(D) \geq \widetilde{R}^{\mathrm{L}}_{X_1}(D)$ for $p = 1$ requires $\|\mathrm{d}\mu_{X_1}/\mathrm{d}\mu\|^{(\mu)}_\infty < \infty$, whereas $R(D) \geq R^{\mathrm{L}}_{X_1}(D)$ is based on the assumption $\big|h_\mu(X_1)\big| < \infty$. As illustrated in Examples 5.1 and 5.2 below, there are cases where $\|\mathrm{d}\mu_{X_1}/\mathrm{d}\mu\|^{(\mu)}_\infty < \infty$ and $h_\mu(X_1) = \infty$ and vice versa.

Example 5.1. Set $\mathcal{X} = \mathcal{Y} = [e,\infty)$ and let the random variable X take values in $\mathcal{X}$. Further, set $g(x) := (\mathrm{d}\mu_X/\mathrm{d}\mathscr{L}^1|_{\mathcal{X}})(x) = 1/(x\log^2(x))$. It then follows that $\|g\|^{(\mathscr{L}^1|_{\mathcal{X}})}_\infty = 1/e < \infty$, but

$$h_{\mathscr{L}^1|_{\mathcal{X}}}(X) = -\int_e^\infty g(x)\log(g(x))\,\mathrm{d}x \tag{5.12}$$

$$\geq \int_e^\infty \frac{1}{x\log(x)}\,\mathrm{d}x \tag{5.13}$$

$$= \infty. \tag{5.14}$$

Example 5.2. Set $\mathcal{X} = \mathcal{Y} = [0,1]$ and let the random variable X take values in $\mathcal{X}$. Further, set $g(x) := (\mathrm{d}\mu_X/\mathrm{d}\mathscr{L}^1|_{\mathcal{X}})(x) = 1/(2\sqrt{x})$. It then follows that $\|g\|_2^{(\mathscr{L}^1|_{\mathcal{X}})} = \infty$, which implies $\|g\|_\infty^{(\mathscr{L}^1|_{\mathcal{X}})} = \infty$, while

$$h_{\mathscr{L}^1|_{\mathcal{X}}}(X) = -\int_0^1 g(x)\log(g(x))\,\mathrm{d}x \tag{5.15}$$

$$= \log(2) + \frac{1}{2}\int_0^1 \frac{\log(\sqrt{x})}{\sqrt{x}}\,\mathrm{d}x \tag{5.16}$$

$$= \log(2) + \int_0^1 \log(u)\,\mathrm{d}u \tag{5.17}$$

$$= \log(2) - 1. \tag{5.18}$$

6. R-D Theory and Quantization for Compact Manifolds

In this section, we particularize our results on R-D theory and quantization to random variables taking values in compact manifolds, specifically hyperspheres and Grassmannians. Hyperspheres are prevalent in many areas of data science including spherical quantization [18, 49, 59], hypersphere learning [34, Section 4], and directional statistics [48]. Grassmannians find application in, e.g., code design [13, 30, 52, 66], computer vision [60], deep neural network theory [32], and the completion of low-rank matrices [7]. What we need here in order to apply the program developed above are suitable sub/super-regularity conditions, which in turn requires volume estimates of balls. For hyperspheres and Grassmannians, these estimates are obtained by direct computation. For general complete Riemannian manifolds under suitable curvature assumptions, such volume estimates can be derived using the Bishop-Günther volume bounds [23, Theorem 3.101] (see also [30, Section III.A] for an application of this method to Grassmannians and Stiefel manifolds).

6.1. *R-D Theory and Quantization for Hyperspheres*

We first consider random variables taking values in the hypersphere $\mathcal{S}^{d-1}(r)$ and derive corresponding lower bounds on the R-D function and lower and upper bounds on the n-th quantization error. The bounds on the n-th quantization error will allow us to obtain lower and upper bounds on the lower and upper k-th quantization coefficient, respectively. Finally, as an example, we evaluate our results for a random variable of von Mises-Fisher distribution.

We start by establishing subregularity and superregularity for the measure $\mathscr{H}^{d-1}|_{\mathcal{S}^{d-1}(r)}/a^{(d-1)}(r)$.

Lemma 6.1 *Fix $d \in \mathbb{N}\setminus\{1\}$ and $r \in (0,\infty)$ and consider the restricted normalized Hausdorff measure $\mu = \mathscr{H}^{d-1}|_{\mathcal{S}^{d-1}(r)}/a^{(d-1)}(r)$. Further, let $\mathcal{S}^{d-1}(r) \subseteq \mathcal{Y} \subseteq \mathbb{R}^d$ with distortion function $\rho\colon \mathcal{S}^{d-1}(r) \times \mathcal{Y} \to [0,\infty)$, $\rho(x,y) = \|x-y\|_2^2$. Then, the following holds:*

(i) The measure μ satisfies the following family of subregularity conditions:

$$\mu\left(\mathcal{B}_{\rho^{1/2}}\left(y,\delta\right)\right) \le c_{\delta_0}\delta^{d-1} \quad \textit{for all } y \in \mathcal{Y} \textit{ and } \delta \in (0,\delta_0] \tag{6.1}$$

parametrized by $\delta_0 \in (0,r]$, where $c_{\delta_0} = G(\delta_0^2/r^2)/a^{(d-1)}(r)$ with

$$G(\alpha) = \frac{a^{(d-1)}(1)}{2}\frac{I_{\frac{d-1}{2},\frac{1}{2}}(\alpha)}{\alpha^{\frac{d-1}{2}}} \tag{6.2}$$

continuous and strictly monotonically increasing on $(0,1]$ *with*

$$\lim_{\alpha\to 0} G(\alpha) = \frac{\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)} = \nu^{(d-1)}(1). \tag{6.3}$$

(ii) The measure μ *satisfies the following family of superregularity conditions:*

$$\mu\left(\widetilde{\mathcal{B}}_{\rho^{1/2}}\left(x,\delta\right)\right) \geq b_{\delta_0}\delta^{d-1} \quad \text{for all } x\in\mathcal{S}^{d-1}(r) \text{ and } \delta\in(0,\delta_0] \tag{6.4}$$

parametrized by $\delta_0 \in (0,\sqrt{2}r]$, *where*

$$b_{\delta_0} = \frac{\nu^{(d-1)}(1)}{a^{(d-1)}(r)}\left(1-\frac{\delta_0^2}{4r^2}\right)^{\frac{d-1}{2}}. \tag{6.5}$$

(iii) We have the following limits:

$$c_0 := \lim_{\delta_0\to 0} c_{\delta_0} = b_0 := \lim_{\delta_0\to 0} b_{\delta_0} = \frac{\nu^{(d-1)}(1)}{a^{(d-1)}(r)}. \tag{6.6}$$

Proof See Appendix J. □

Note that (6.6) implies that the bounds in (6.1) and (6.4) become tight for $\delta_0 \to 0$. The specific value for the limiting constant, namely $\nu^{(d-1)}(1)/a^{(d-1)}(r) = \kappa_{d-1}(\|\cdot\|_2)/a^{(d-1)}(r)$, can be explained using the following deep result on $\mathscr{H}^m$-rectifiable sets in geometric measure theory.

Theorem 6.1 *[2, Theorem 2.63] An* $\mathscr{H}^m$*-measurable set* $\mathcal{A}\subseteq\mathbb{R}^d$ *with* $0<\mathscr{H}^m(\mathcal{A})<\infty$ *is* $\mathscr{H}^m$*-rectifiable if and only if*

$$\lim_{\delta\to 0}\frac{\mathscr{H}^m|_{\mathcal{A}}\left(\mathcal{B}_{\|\cdot\|}\left(y,\delta\right)\right)}{\delta^m} = \kappa_m(\|\cdot\|) \quad \text{for } \mathscr{H}^m\text{-almost all } y\in\mathcal{A}. \tag{6.7}$$

Intuitively, Theorem 6.1 says that viewed from close up, every $\mathscr{H}^m$-rectifiable set looks almost everywhere like $\mathbb{R}^m$.

In the following, we fix $d\in\mathbb{N}\backslash\{1\}$, set

$$\mu := \frac{\mathscr{H}^{d-1}|_{\mathcal{S}^{d-1}(r)}}{a^{(d-1)}(r)}, \tag{6.8}$$

let X be a random variable taking values in $\mathcal{X}=\mathcal{S}^{d-1}(r)\subseteq\mathcal{Y}\subseteq\mathbb{R}^d$, and consider the distortion function $\rho\colon \mathcal{X}\times\mathcal{Y}\to[0,\infty)$, $\rho(x,y)=\|x-y\|_2^2$.

We first derive a lower bound on the R-D function under the assumptions $\mu_X \ll \mu$ and $h_\mu(X) > -\infty$. Before starting in earnest, we note that Jensen's inequality [51, Theorem 2.3] combined with $\mu(\mathcal{X})=1$

yields $h_\mu(X) \le \log(\mu(\mathcal{X})) = 0$, which in turn implies $|h_\mu(X)| < \infty$. Fix $\alpha \in (0,1/2)$ and set, for every $D \in (0,\infty)$, $\delta_D = \min\{D^\alpha, r\}$ and

$$c_{\delta_D} = \frac{1}{2}\frac{I_{\frac{d-1}{2},\frac{1}{2}}\left(\frac{\delta_D^2}{r^2}\right)}{\delta_D^{d-1}}. \tag{6.9}$$

Application of Theorem 3.1 with $k=2$ and $m=d-1$ combined with Item (i) of Lemma 6.1 for $\delta_0 = \delta_D$ then yields $R_X^{\mathrm{SLB}}(D) \ge R_X^{\mathrm{L}}(D)$ for all $D \in (0,\infty)$ with

$$R_X^{\mathrm{L}}(D) = h_\mu(X) - \frac{d-1}{2} \tag{6.10}$$

$$-\log\left(c_{\delta_D}\left(\frac{d-1}{2D}\right)^{-\frac{d-1}{2}}\Gamma\left(\frac{d+1}{2}\right) + e^{-\frac{(d-1)\delta_D^2}{2D}}\right). \tag{6.11}$$

Moreover, since $\delta_D \to 0$ as $D \to 0$ and $e^{-(d-1)\delta_D^2/(2D)} \to 0$ as $D \to 0$, we have

$$\lim_{D\to 0}(R_X^{\mathrm{L}}(D) - h_\mu(X) - F_{d-1,2,c_0}(D)) = 0, \tag{6.12}$$

where we used (6.6) to conclude that

$$c_0 = \frac{v^{(d-1)}(1)}{a^{(d-1)}(r)}. \tag{6.13}$$

Thus, as $D \to 0$, the lower bound $R_X^{\mathrm{L}}(D)$ approaches the classical Shannon lower bound $R_W^{\mathrm{SLB}}(D)$ in (3.3) for a continuous random variable W taking values in $\mathbb{R}^{d-1}$ and of differential entropy

$$h_{\mathscr{L}^{d-1}}(W) = \log(a^{(d-1)}(r)) + h_\mu(X) = h_{\tilde{\mu}}(X) \tag{6.14}$$

with $\tilde{\mu} = a^{(d-1)}(r)\mu = \mathscr{H}^{d-1}|_{\mathcal{S}^{d-1}(r)}$. This reflects the fact that from close up the hypersphere $\mathcal{S}^{d-1}(r)$ looks like $\mathbb{R}^{d-1}$ (see Theorem 6.1).

Next, we derive a lower bound on the n-th quantization error $V_n(X)$ under the assumptions $\mu_X \ll \mu$ and

$$\Sigma_p(X) := \left\|\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu}\right\|_{p/(p-1)}^{(\mu)} < \infty \quad \text{with } p \in [1,\infty). \tag{6.15}$$

To this end, for every $n \in \mathbb{N}$ with $n \ge n_0 := 2^{1/p}/\Sigma_p(X)$, we set

$$\delta_n = r\sqrt{I^{-1}_{\frac{d-1}{2},\frac{1}{2}}\left(\frac{2}{n^p\Sigma_p^p(X)}\right)} \tag{6.16}$$

and

$$c_{\delta_n} = \frac{1}{2}\frac{I_{\frac{d-1}{2},\frac{1}{2}}\left(\frac{\delta_n^2}{r^2}\right)}{\delta_n^{d-1}}. \tag{6.17}$$

These choices for δ_n and c_{δ_n} guarantee that $\delta_n \leq r$ and

$$\delta_n = c_{\delta_n}^{-\frac{1}{d-1}} \Sigma_p^{-\frac{p}{d-1}}(X)\, n^{-\frac{p}{d-1}} \quad \text{for all } n \geq n_0. \tag{6.18}$$

Application of Theorem 4.1 with $k=2$ and $m=d-1$ combined with Item (i) of Lemma 6.1 for $\delta_0 = \delta_n$ then yields $V_n(X) \geq L_n(X)$ for all $n \geq n_0$ with

$$L_n(X) = \frac{d-1}{d-1+2p} r^2 I^{-1}_{\frac{d-1}{2},\frac{1}{2}}\left(\frac{2}{n^p \Sigma_p^p(X)}\right). \tag{6.19}$$

This allows us to establish

$$\liminf_{n\to\infty} n^{\frac{2p}{d-1}} V_n(X) \geq \frac{d-1}{d-1+2p}\, r^2\, \Sigma_p^{-\frac{2p}{d-1}}(X)\, k_d > 0 \tag{6.20}$$

with

$$k_d = \left(\frac{2\sqrt{\pi}\,\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}\right)^{\frac{2}{d-1}}. \tag{6.21}$$

Indeed, we have

$$\begin{aligned} \liminf_{n\to\infty} n^{\frac{2p}{d-1}} V_n(X) &\geq \lim_{n\to\infty} n^{\frac{2p}{d-1}} L_n(X) &(6.22)\\ &= \frac{d-1}{d-1+2p} r^2 \lim_{n\to\infty} n^{\frac{2p}{d-1}} I^{-1}_{\frac{d-1}{2},\frac{1}{2}}\left(\frac{2}{n^p \Sigma_p^p(X)}\right) &(6.23)\\ &= \frac{d-1}{d-1+2p} r^2\, \Sigma_p^{-\frac{2p}{d-1}}(X) \left(\lim_{\alpha\to 0} \frac{2\alpha^{\frac{d-1}{2}}}{I_{\frac{d-1}{2},\frac{1}{2}}(\alpha)}\right)^{\frac{2}{d-1}}, &(6.24) \end{aligned}$$

where in (6.24) we set

$$\alpha = I^{-1}_{\frac{d-1}{2},\frac{1}{2}}\left(\frac{2}{n^p \Sigma_p^p(X)}\right). \tag{6.25}$$

Using L'Hôpital's rule, we obtain

$$\lim_{\alpha\to 0} \frac{2\alpha^{\frac{d-1}{2}}}{I_{\frac{d-1}{2},\frac{1}{2}}(\alpha)} = \frac{2\sqrt{\pi}\,\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}, \tag{6.26}$$

which when inserted in (6.24) establishes (6.20).

We next derive an upper bound U_n on the n-th quantization error $V_n(X)$. Fix $\alpha \in (0,1)$ and set, for every $n \in \mathbb{N}\backslash\{1\}$, $\delta_n = \sqrt{2}rn^{-\alpha/(d-1)}$ and

$$b_{\delta_n} = \frac{\Gamma\left(\frac{d}{2}\right)}{2\sqrt{\pi}r^{d-1}\Gamma\left(\frac{d+1}{2}\right)}\left(1-\frac{\delta_n^2}{4r^2}\right)^{\frac{d-1}{2}}. \tag{6.27}$$

As

$$\sup_{x,y\in\mathcal{S}^{d-1}(r)} \|x-y\|_2^2 = 4r^2, \tag{6.28}$$

application of Theorem 4.2 with $\mathcal{X} = \mathcal{S}^{d-1}(r)$, $\nu = \mu$, $k = 2$, and $m = d-1$ combined with Item (ii) of Lemma 6.1 for $\delta_0 = \delta_n$ then yields $V_n(X) \leq U_n$ for all $n \in \mathbb{N}$ with

$$U_n = \Gamma\left(\frac{d+1}{d-1}\right)\left(b_{\delta_n}n\right)^{-\frac{2}{d-1}} + \left(4r^2-\delta_n^2\right)e^{-b_{\delta_n}n\,\delta_n^{d-1}}. \tag{6.29}$$

Next, note that $\lim_{n\to\infty}\delta_n = 0$ and $\lim_{n\to\infty} n^{\frac{2}{d-1}}e^{-b_{\delta_n}n\,\delta_n^{d-1}} = 0$. Therefore, $V_n(X) \leq U_n$ implies

$$\limsup_{n\to\infty} n^{\frac{2}{d-1}}V_n(X) \leq \Gamma\left(\frac{d+1}{d-1}\right)r^2 k_d < \infty \tag{6.30}$$

with k_d as defined in (6.21). Combining (6.20) and (6.30) with Items (i) and (ii) of Lemma 4.2, respectively, now implies

$$\frac{d-1}{p} \leq \underline{D}_2(X) \leq \overline{D}_2(X) = d-1. \tag{6.31}$$

If $p=1$, then the 2-nd quantization dimension $D_2(X)$ exists and equals the geometric dimension $d-1$ of the hypersphere. Similarly, if $h_\mu(X) > -\infty$ (Jensen's inequality [51, Theorem 2.3] combined with $\mu(\mathcal{X}) = 1$ yields $h_\mu(X) \leq \log(\mu(\mathcal{X})) = 0$), then Item (ii) of Corollary 4.4 (with $\mathcal{X} = \mathcal{S}^{d-1}(r)$ and $\nu = \mu$) allows us to conclude that

$$\underline{\dim}_R(X) = D_2(X) = d-1. \tag{6.32}$$

In particular, we have the following results on the upper and lower 2-nd quantization coefficient. If $D_2(X) = d-1$, then by (6.30), we have

$$\overline{C}_2(X) \leq \Gamma\left(\frac{d+1}{d-1}\right)r^2 k_d < \infty, \tag{6.33}$$

and if (6.15) holds for $p=1$, then (6.20) yields

$$\underline{C}_2(X) \geq \frac{d-1}{d+1}r^2\Sigma_1^{-\frac{2}{d-1}}(X)\,k_d > 0. \tag{6.34}$$

For $\mathcal{Y} = \mathcal{S}^{d-1}(r)$, if in addition to $D_2(X) = d-1$, we have $d \geq 4$, $\mu \ll \mu_X$, and $\|\mathrm{d}\mu/\mathrm{d}\mu_X\|_\infty^{(\mu_X)} < \infty$, then we can apply Theorem 4.3 with $\nu = \mu$, $k = 2$, and $m = d-1$ to obtain the improved upper bound

$$\overline{C}_2(X) \leq \Omega_{2/(d-1)}(X)\Gamma\left(\frac{d+1}{d-1}\right)r^2 k_d, \tag{6.35}$$

where

$$\Omega_{2/(d-1)}(X) = \mathbb{E}\left[\left(\frac{\mathrm{d}\mu}{\mathrm{d}\mu_X}(X)\right)^{\frac{2}{d-1}}\right] \leq 1 \quad \text{for all } d \geq 4 \tag{6.36}$$

with strict inequality in (6.36) unless $\mu_X = \mu$.

In order to endow the results just obtained with a more specific, and, in particular, quantitative flavor, we evaluate $h_\mu(X)$, $\Sigma_1(X)$, and $\Omega_{2/(d-1)}(X)$ for X of von Mises-Fisher distribution.

Example 6.1. Fix $d \in \mathbb{N}\backslash\{1\}$ and let X be a random variable of von Mises-Fisher distribution μ_X with mean direction $y \in \mathcal{S}^{d-1}(1)$ and concentration parameter $\kappa \in (0,\infty)$, which is determined according to [48, Equation (9.3.4)]

$$\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu}(x) = c_d(\kappa)e^{\kappa y^\mathsf{T} x}, \tag{6.37}$$

where $\mu = \mathscr{H}^{d-1}|_{\mathcal{S}^{d-1}(1)}/a^{(d-1)}(1)$ and

$$c_d(\kappa) := \left(\int e^{\kappa y^\mathsf{T} x}\,\mathrm{d}\mu(x)\right)^{-1} \tag{6.38}$$

$$= \frac{\kappa^{\frac{d}{2}-1}}{\Gamma\left(\frac{d}{2}\right)2^{\frac{d}{2}-1}I_{\frac{d}{2}-1}(\kappa)} \tag{6.39}$$

with

$$I_\alpha(\kappa) = \frac{1}{2\pi}\int_0^{2\pi}\cos(\alpha t)e^{\kappa\cos(t)}\,\mathrm{d}t. \tag{6.40}$$

This distribution plays an important role in directional statistics [48, Section 9.3.2]. Specifically, it is one of the simplest parametric distributions on $\mathcal{S}^{d-1}(1)$ and has an entropy-maximizing property akin to that of the multivariate Gaussian distribution on $\mathbb{R}^d$. Concretely, among all random variables Z taking values in $\mathcal{S}^{d-1}(1)$, of distribution $\mu_Z \ll \mu = \mathscr{H}^{d-1}|_{\mathcal{S}^{d-1}(1)}/a^{(d-1)}(1)$, and with $\mathbb{E}[Z]$ fixed, the von Mises-Fisher distribution with y and κ determined by [48, Equation (9.3.7)]

$$\mathbb{E}[X] = y\frac{I_{\frac{d}{2}}(\kappa)}{I_{\frac{d}{2}-1}(\kappa)} \tag{6.41}$$

maximizes $h_\mu(Z)$ [47, Section 2.3]. The generalized entropy $h_\mu(X)$ of X with von Mises-Fisher distribution can be derived as follows. First note that

$$h_\mu(X) = -\mathbb{E}\left[\log\left(\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu}(X)\right)\right] \tag{6.42}$$

$$= -\log(c_d(\kappa)) - \kappa\mathbb{E}\left[y^\mathsf{T} X\right]. \tag{6.43}$$

Now, (6.41) together with $y \in \mathcal{S}^{d-1}(1)$ implies

$$\mathbb{E}\left[y^\mathsf{T} X\right] = \frac{I_{\frac{d}{2}}(\kappa)}{I_{\frac{d}{2}-1}(\kappa)}. \tag{6.44}$$

Using (6.44) in (6.43) results in

$$h_\mu(X) = -\log(c_d(\kappa)) - \kappa\frac{I_{\frac{d}{2}}(\kappa)}{I_{\frac{d}{2}-1}(\kappa)}. \tag{6.45}$$

Moreover,

$$\Sigma_1(X) = \left\|\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu}\right\|_\infty^{(\mu)} \tag{6.46}$$

$$= c_d(\kappa) \sup_{x\in\mathcal{S}^{d-1}(1)} e^{\kappa y^\mathsf{T} x} \tag{6.47}$$

$$= c_d(\kappa)e^\kappa, \tag{6.48}$$

which implies (see (6.34))

$$\underline{C}_2(X) \geq \frac{d-1}{d+1} r^2 \left(c_d(\kappa)e^\kappa\right)^{-\frac{2}{d-1}} k_d > 0. \tag{6.49}$$

Finally, if $d \geq 4$, then

$$\Omega_{2/(d-1)}(X) = \mathbb{E}\left[\left(\frac{\mathrm{d}\mu}{\mathrm{d}\mu_X}\right)^{\frac{2}{d-1}}(X)\right] \tag{6.50}$$

$$= \int \left(\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu}\right)^{\frac{d-3}{d-1}}(x)\, d\mu(x) \tag{6.51}$$

$$= \frac{c_d^{\frac{d-3}{d-1}}(\kappa)}{c_d\left(\kappa\frac{d-3}{d-1}\right)}. \tag{6.52}$$

We conclude this discussion by noting that the lower bound on $\underline{C}_2(X)$ in (6.34) evaluated for $d = 2$ and $\mu_X = \mu$, i.e., X uniformly distributed on the circle of radius r, is sharp enough to establish that $C_2(X) = r^2\pi^2/3$, which is shown in the following example.

Example 6.2. Let X be a random variable with uniform distribution μ_X on the circle of radius r. Using $d = 2$ and $\mu_X = \mu$ in (6.34) yields

$$\underline{C}_2(X) \geq \frac{r^2\pi^2}{3}. \tag{6.53}$$

To find a matching upper bound on $\overline{C}_2(X)$, we first derive, for every $n \in \mathbb{N}$, an upper bound $U_n(X)$ on the n-th quantization error $V_n(X)$ for $\mathcal{Y} = \mathcal{S}^1(r)$. Since more flexibility in placing the quantization points can only reduce $V_n(X)$, this upper bound applies to $\mathcal{S}^1(r) \subseteq \mathcal{Y} \subseteq \mathbb{R}^2$ as well. Concretely, for every $n \in \mathbb{N}$ and $x \in \mathcal{S}^1(r)$, we set $f_n(x) = a_i$, where $i = \operatorname{argmin}_{j\in\{1,\dots,n\}} \|x - a_j\|_2$ and $a_j = r(\cos(2\pi j/n), \sin(2\pi j/n))$ for $j = 1,\dots,n$. This yields

$$V_n(X) \leq \mathbb{E}[\|X - f_n(X)\|_2^2] \tag{6.54}$$

$$= \frac{2r^2 n}{\pi} \int_0^{\frac{\pi}{n}} (1 - \cos(\alpha))\, \mathrm{d}\alpha \tag{6.55}$$

$$= 2r^2 \left(1 - \operatorname{sinc}\left(\frac{1}{n}\right)\right) \quad \text{for all } n \in \mathbb{N}, \tag{6.56}$$

where (6.55) follows from the formula for the chord length corresponding to a circle segment of central angle α. We thus have

$$\limsup_{n\to\infty} n^2 V_n(X) \leq 2r^2 \lim_{n\to\infty} n^2 \left(1 - \operatorname{sinc}\left(\frac{1}{n}\right)\right) \tag{6.57}$$

$$= 2r^2\pi^2 \lim_{\varepsilon\to 0} \frac{\varepsilon - \sin(\varepsilon)}{\varepsilon^3} \tag{6.58}$$

$$= \frac{r^2\pi^2}{3}, \tag{6.59}$$

where in (6.57) we used (6.54)–(6.56), in (6.58) we substituted $\varepsilon = \pi/n$, and in (6.59) we applied L'Hôpital's rule three times. As $D_2(X) = 1$, see (6.31) for $d = 2$ and $p = 1$, by (6.57)–(6.59) we get

$$\overline{C}_2(X) \leq \frac{r^2\pi^2}{3}. \tag{6.60}$$

Combining (6.53) and (6.60) yields the desired result $C_2(X) = r^2\pi^2/3$.

6.2. *R-D Theory and Quantization for Grassmannians*

We now consider random variables taking values in a Grassmannian. Specifically, we derive corresponding lower bounds on the R-D function and lower and upper bounds on the n-th quantization error. We start with preparatory material on Grassmannians largely following the exposition in [13, Sections II-III]. For $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and $r, d \in \mathbb{N}$ with $1 \leq r \leq d$, let $\mathcal{G}^{\mathbb{F}}(r,d)$ denote the Grassmannian consisting of all r-dimensional subspaces of $\mathbb{F}^d$ and designate by $\gamma_{r,d}$ the unique uniformly distributed Borel regular measure on $\mathcal{G}^{\mathbb{F}}(r,d)$ with $\gamma_{r,d}(\mathcal{G}^{\mathbb{F}}(r,d)) = 1$. The dimension of $\mathcal{G}^{\mathbb{F}}(r,d)$ is given by $\beta r(d-r)$ with $\beta = 1$ if $\mathbb{F} = \mathbb{R}$ and $\beta = 2$ if $\mathbb{F} = \mathbb{C}$. For Grassmannians $\mathcal{G}^{\mathbb{F}}(r,d)$ and $\mathcal{G}^{\mathbb{F}}(s,d)$ with $r, s, d \in \mathbb{N}$ and $1 \leq r, s \leq d$, the chordal distance ρ_c is defined according to

$$\rho_c\colon \mathcal{G}^{\mathbb{F}}(r,d) \times \mathcal{G}^{\mathbb{F}}(s,d) \to [0, \sqrt{\min\{r,s\}}] \tag{6.61}$$

$$(x,y) \mapsto \sqrt{\sum_{i=1}^{\min\{r,s\}} \sin^2(\theta_i(x,y))}, \tag{6.62}$$

where $\theta_1(x,y),\dots,\theta_{\min\{r,s\}}(x,y)$ are the principal angles between the subspaces $x \in \mathcal{G}^{\mathbb{F}}(r,d)$ and $y \in \mathcal{G}^{\mathbb{F}}(s,d)$. The chordal distance can now be used to state the following sub/super-regularity conditions for the measures $\gamma_{r,d}$ on $\mathcal{G}^{\mathbb{F}}(r,d)$ and $\gamma_{s,d}$ on $\mathcal{G}^{\mathbb{F}}(s,d)$.

Lemma 6.2 *[13, Equation (6) and Corollaries 1 and 2] Consider the Grassmannians $\mathcal{X} = \mathcal{G}^{\mathbb{F}}(r,d)$ and $\mathcal{Y} = \mathcal{G}^{\mathbb{F}}(s,d)$ with $1 \le r,s \le d$. Set $a = \min\{r,s\}$, $b = \max\{r,s\}$, $m_{\mathrm{G}} = \beta a(d-b)$, and*

$$c_{a,b,d,\beta} = \begin{cases} \frac{1}{\Gamma\left(\frac{\beta}{2}a(d-b)+1\right)} \prod_{i=1}^{a} \frac{\Gamma\left(\frac{\beta}{2}(d-i+1)\right)}{\Gamma\left(\frac{\beta}{2}(b-i+1)\right)} & \text{if } a+b \le d \\ \frac{1}{\Gamma\left(\frac{\beta}{2}a(d-b)+1\right)} \prod_{i=1}^{d-b} \frac{\Gamma\left(\frac{\beta}{2}(d-i+1)\right)}{\Gamma\left(\frac{\beta}{2}(d-a-i+1)\right)} & \text{else,} \end{cases} \tag{6.63}$$

where $\beta = 1$ if $\mathbb{F} = \mathbb{R}$ and $\beta = 2$ if $\mathbb{F} = \mathbb{C}$. Then, we have

$$v_{r,s}^{(d)}(\delta) := \gamma_{r,d}\left(\mathcal{B}_{\rho_{\mathrm{c}}}(y,\delta)\right) = \gamma_{s,d}\left(\widetilde{\mathcal{B}}_{\rho_{\mathrm{c}}}(x,\delta)\right) \tag{6.64}$$

for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ and the following holds:

(i) If ($\mathbb{F} = \mathbb{R}$ and $b = a+1$) or ($\mathbb{F} = \mathbb{C}$ and $b = a$), then

$$v_{r,s}^{(d)}(\delta) = c_{a,b,d,\beta}\,\delta^{m_{\mathrm{G}}} \quad \text{for all } \delta \in (0,1]. \tag{6.65}$$

(ii) If $\mathbb{F} = \mathbb{R}$ and $a = b$, then

$$c_{a,a,d,1}\,\delta^{m_{\mathrm{G}}} \le v_{r,s}^{(d)}(\delta) \le \frac{c_{a,a,d,1}}{(1-\delta_0^2)^{\frac{a}{2}}}\,\delta^{m_{\mathrm{G}}} \tag{6.66}$$

for all $\delta \in (0,\delta_0]$ parametrized by $\delta_0 \in (0,1]$.

(iii) If ($\mathbb{F} = \mathbb{R}$ and $b \notin \{a,a+1\}$) or ($\mathbb{F} = \mathbb{C}$ and $a \ne b$), then

$$c_{a,b,d,\beta}(1-\delta_0^2)^{\frac{\beta}{2}a(b-a+1)-a}\,\delta^{m_{\mathrm{G}}} \le v_{r,s}^{(d)}(\delta) \le c_{a,b,d,\beta}\,\delta^{m_{\mathrm{G}}} \tag{6.67}$$

for all $\delta \in (0,\delta_0]$ parametrized by $\delta_0 \in (0,1]$.

In the following, fix $r,s,d \in \mathbb{N}$ with $1 \le r,s \le d$ and let $\mathcal{X} = \mathcal{G}^{\mathbb{F}}(r,d)$ and $\mathcal{Y} = \mathcal{G}^{\mathbb{F}}(s,d)$ be equipped with the distortion function $\rho(x,y) = \rho_{\mathrm{c}}^2(x-y)$. Further, set $a = \min\{r,s\}$, $b = \max\{r,s\}$, and $m_{\mathrm{G}} = \beta a(d-b)$, again with $\beta = 1$ if $\mathbb{F} = \mathbb{R}$ and $\beta = 2$ if $\mathbb{F} = \mathbb{C}$. Note that, unless $r = s$, $\mathcal{X}$ and $\mathcal{Y}$ constitute different manifolds. If $r = s$, then m_{G} equals the dimension of the Grassmannian $\mathcal{X}$. Further, let X be a random variable taking values in $\mathcal{X}$.

We proceed to derive a lower bound on the R-D function under the assumptions $\mu_X \ll \gamma_{r,d}$ and $h_{\gamma_{r,d}}(X) > -\infty$. To this end, we start by noting that Jensen's inequality [51, Theorem 2.3] combined with $\gamma_{r,d}(\mathcal{X}) = 1$ yields $h_{\gamma_{r,d}}(X) \le \log(\gamma_{r,d}(\mathcal{X})) = 0$, which implies $\left|h_{\gamma_{r,d}}(X)\right| < \infty$.

Suppose first that $\mathbb{F} = \mathbb{R}$ and $a = b$. Fix $\alpha \in (0,1/2)$ and set, for every $D \in (0,1)$, $\delta_D = D^{\alpha}$ and

$$d_{\delta_D} = \frac{c_{a,a,d,1}}{(1-\delta_D^2)^{\frac{a}{2}}}. \tag{6.68}$$

Application of Theorem 3.1 for $k=2$, $m=m_{\mathrm{G}}$, and $\mu=\gamma_{r,d}$ combined with Item (ii) of Lemma 6.2 then yields $R_X^{\mathrm{SLB}}(D)\geq R_X^{\mathrm{L}}(D)$ for all $D\in(0,1)$ with

$$R_X^{\mathrm{L}}(D)=h_\mu(X)-\frac{m_{\mathrm{G}}}{2}-\log\left(d_{\delta_D}\left(\frac{m_{\mathrm{G}}}{2D}\right)^{-\frac{m_{\mathrm{G}}}{2}}\Gamma\left(1+\frac{m_{\mathrm{G}}}{2}\right)+e^{-\frac{m_{\mathrm{G}}\delta_D^2}{2D}}\right). \tag{6.69}$$

Moreover, since $\lim_{D\to 0}\delta_D=0$ and $\lim_{D\to 0}e^{-m_{\mathrm{G}}\delta_D^2/(2D)}=0$, we have

$$\lim_{D\to 0}(R_X^{\mathrm{L}}(D)-h_\mu(X)-F_{m_{\mathrm{G}},2,c_{a,a,d,1}}(D))=0 \tag{6.70}$$

with $F_{m,k,d}(D)$ as defined in (3.4).

Next, suppose that $\mathbb{F}=\mathbb{C}$ or $a\neq b$. Application of Theorem 3.1 with $k=2$, $m=m_{\mathrm{G}}$, $\mu=\gamma_{r,d}$, and $\delta_0=1$ combined with Items (i) and (iii) of Lemma 6.2 then yields $R_X^{\mathrm{SLB}}(D)\geq R_X^{\mathrm{L}}(D)$ for all $D\in(0,1)$ with

$$R_X^{\mathrm{L}}(D)=h_{\gamma_{r,d}}(X)-\frac{m_{\mathrm{G}}}{2}-\log\left(c_{a,b,d,\beta}\left(\frac{m_{\mathrm{G}}}{2D}\right)^{-\frac{m_{\mathrm{G}}}{2}}\Gamma\left(1+\frac{m_{\mathrm{G}}}{2}\right)+e^{-\frac{m_{\mathrm{G}}}{2D}}\right). \tag{6.71}$$

Moreover, we have

$$\lim_{D\to 0}(R_X^{\mathrm{L}}(D)-h_{\gamma_{r,d}}(X)-F_{m_{\mathrm{G}},2,c_{a,b,d,\beta}}(D))=0. \tag{6.72}$$

Next, we derive a lower bound on the n-th quantization error $V_n(X)$ under the assumptions $\mu_X\ll\gamma_{r,d}$ and

$$\Sigma_p(X):=\left\|\frac{\mathrm{d}\mu_X}{\mathrm{d}\gamma_{r,d}}\right\|_{p/(p-1)}^{(\gamma_{r,d})}<\infty\quad\text{with } p\in[1,\infty). \tag{6.73}$$

Suppose first that $\mathbb{F}=\mathbb{R}$ and $a=b$ and consider the strictly monotonically increasing function

$$h\colon\ [0,1)\to[0,\infty) \tag{6.74}$$

$$u\mapsto\frac{c_{a,a,d,1}u^{m_{\mathrm{G}}}}{(1-u^2)^{\frac{a}{2}}}. \tag{6.75}$$

As $h(0)=0$ and $\lim_{u\to 1}h(u)=\infty$, h is bijective. For every $n\in\mathbb{N}$, set

$$\delta_n=h^{-1}\left(\frac{1}{n^p\Sigma_p^p(X)}\right) \tag{6.76}$$

and

$$c_{\delta_n}=\frac{c_{a,a,d,1}}{(1-\delta_n^2)^{\frac{a}{2}}}=h(\delta_n)\delta_n^{-m_{\mathrm{G}}}. \tag{6.77}$$

These choices for δ_n and c_{δ_n} ensure $\delta_n < 1$ and

$$\delta_n = c_{\delta_n}^{-\frac{1}{m_{\mathrm{G}}}} \Sigma_p(X)^{-\frac{p}{m_{\mathrm{G}}}} n^{-\frac{p}{m_{\mathrm{G}}}} \quad \text{for all } n \in \mathbb{N}. \tag{6.78}$$

Application of Theorem 4.1 with $k = 2$, $m = m_{\mathrm{G}}$, and $\mu = \gamma_{r,d}$ combined with Item (ii) of Lemma 6.2 then yields $V_n(X) \geq L_n(X)$ for all $n \in \mathbb{N}$ with

$$L_n(X) = \frac{m_{\mathrm{G}}}{m_{\mathrm{G}} + 2p} \left(h^{-1} \left(\frac{1}{n^p \Sigma_p^p(X)} \right) \right)^2. \tag{6.79}$$

This lower bound can be further simplified as follows. First, note that

$$\frac{1}{n^p \Sigma_p^p(X)} = h(\delta_n) \geq c_{a,a,d,1} \delta_n^{m_{\mathrm{G}}} \tag{6.80}$$

implies $\delta_n \leq (n^p \Sigma_p^p(X) c_{a,a,d,1})^{-1/m_{\mathrm{G}}}$ so that

$$\frac{1}{n^p \Sigma_p^p(X)} = h(\delta_n) = \frac{c_{a,a,d,1} \delta_n^{m_{\mathrm{G}}}}{(1 - \delta_n^2)^{\frac{a}{2}}} \tag{6.81}$$

$$\leq \frac{c_{a,a,d,1} \delta_n^{m_{\mathrm{G}}}}{(1 - (n^p \Sigma_p^p(X) c_{a,a,d,1})^{-\frac{2}{m_{\mathrm{G}}}})^{\frac{a}{2}}} \tag{6.82}$$

for all $n > n_0 := 1/\left(\Sigma_p(X) c_{a,a,d,1}^{1/p}\right)$. This yields

$$\delta_n^2 \geq \left(c_{a,a,d,1} n^p \Sigma_p^p(X) \right)^{-\frac{2}{m_{\mathrm{G}}}} \left(1 - (n^p \Sigma_p^p(X) c_{a,a,d,1})^{-\frac{2}{m_{\mathrm{G}}}} \right)^{\frac{1}{d-a}} \tag{6.83}$$

for all $n > n_0$. Using (6.76) and (6.83) in (6.79), we finally obtain

$$L_n(X) \geq \frac{m_{\mathrm{G}}}{m_{\mathrm{G}} + 2p} \left(c_{a,a,d,1} n^p \Sigma_p^p(X) \right)^{-\frac{2}{m_{\mathrm{G}}}} \left(1 - (n^p \Sigma_p^p(X) c_{a,a,d,1})^{-\frac{2}{m_{\mathrm{G}}}} \right)^{\frac{1}{d-a}} \tag{6.84}$$

for all $n > n_0$.

Next, suppose that $\mathbb{F} = \mathbb{C}$ or $a \neq b$. Application of Theorem 4.1 with $k = 2$, $m = m_{\mathrm{G}}$, $\mu = \gamma_{r,d}$, and $\delta_0 = 1$ combined with Items (i) and (iii) of Lemma 6.2 yields $V_n(X) \geq L_n(X)$ for all $n \geq 1/\left(\Sigma_p(X) c_{a,b,d,\beta}^{1/p}\right)$ with

$$L_n(X) = \frac{m_{\mathrm{G}}}{m_{\mathrm{G}} + 2p} \left(c_{a,b,d,\beta} n^p \Sigma_p^p(X) \right)^{-\frac{2}{m_{\mathrm{G}}}}. \tag{6.85}$$

In particular, the lower bounds in (6.84) and (6.85) yield

$$\liminf_{n \to \infty} n^{\frac{2p}{m_{\mathrm{G}}}} V_n(X) \geq \frac{m_{\mathrm{G}}}{m_{\mathrm{G}} + 2p} \Sigma_p^{-\frac{2p}{m_{\mathrm{G}}}}(X) c_{a,b,d,\beta}^{-\frac{2}{m_{\mathrm{G}}}} > 0. \tag{6.86}$$

We next derive an upper bound U_n on the n-th quantization error $V_n(X)$. Suppose first that $\mathbb{F}=\mathbb{R}$ with $b\in\{a,a+1\}$ or $\mathbb{F}=\mathbb{C}$ with $b=a$. Since (see (6.61) and recall that $\rho(x,y)=\rho_c^2(x-y)$)

$$\sup_{x\in\mathcal{X},y\in\mathcal{Y}}\rho(x,y)\le a, \tag{6.87}$$

application of Theorem 4.2 with $\nu=\gamma_{s,d}$, $m=m_G$, $k=2$, and $\delta_0=1$ combined with Items (i) and (ii) of Lemma 6.2 yields $V_n(X)\le U_n$ for all $n\in\mathbb{N}$ with

$$U_n=\Gamma\left(1+\frac{2}{m_G}\right)(c_{a,b,d,\beta}n)^{-\frac{2}{m_G}}+(a-1)e^{-nc_{a,b,d,\beta}}. \tag{6.88}$$

Next, suppose that $\mathbb{F}=\mathbb{R}$ with $b\notin\{a,a+1\}$ or $\mathbb{F}=\mathbb{C}$ with $a\neq b$. Fix $\alpha\in(0,1)$ and set, for every $n\in\mathbb{N}$, $\delta_n=n^{-\alpha/m_G}$ and

$$b_{\delta_n}=c_{a,b,d,\beta}\left(1-\delta_n^2\right)^{\frac{\beta}{2}a(b-a+1)-a}. \tag{6.89}$$

Application of Theorem 4.2 with $\nu=\gamma_{s,d}$, $m=m_G$, $k=2$, and $\delta_0=\delta_n$ combined with (6.87) and Item (iii) of Lemma 6.2 then implies $V_n(X)\le U_n$ for all $n\in\mathbb{N}$ with

$$U_n=\Gamma\left(1+\frac{2}{m_G}\right)\left(nb_{\delta_n}\right)^{-\frac{2}{m_G}}+\left(a-\delta_n^2\right)e^{-nb_{\delta_n}\delta_n^{m_G}}. \tag{6.90}$$

Next, note that $\lim_{n\to\infty}\delta_n=0$ and $\lim_{n\to\infty}n^{\frac{2}{m_G}}e^{-nb_{\delta_n}\delta_n^{m_G}}=0$. Therefore, (6.88) and (6.90) imply that

$$\limsup_{n\to\infty}n^{\frac{2}{m_G}}V_n(X)\le\Gamma\left(1+\frac{2}{m_G}\right)c_{a,b,d,\beta}^{-\frac{2}{m_G}}<\infty. \tag{6.91}$$

The bounds in (6.84), (6.85), (6.88), and (6.90) generalize [13, Theorem 4] in the sense of applying to more general, i.e., not necessarily uniformly distributed, random variables, and, in addition, do not require the condition of n being larger than an unspecified natural number.

Combining (6.86) and (6.91) with Items (i) and (ii) of Lemma 4.2, respectively, now implies

$$\frac{m_G}{p}\le\underline{D}_2(X)\le\overline{D}_2(X)=m_G. \tag{6.92}$$

If $p=1$, then the 2-nd quantization dimension $D_2(X)$ exists and equals m_G, which for $r=s$, in turn equals the geometric dimension of the Grassmannian $\mathcal{X}$. Similarly, if $h_{\gamma_{r,d}}(X)>-\infty$ (Jensen's inequality [51, Theorem 2.3] combined with $\gamma_{r,d}(\mathcal{X})=1$ yields $h_{\gamma_{r,d}}(X)\le\log(\gamma_{r,d}(\mathcal{X}))=0$), then Item (ii) of Corollary 4.4 with $\mu=\gamma_{r,d}$ and $\nu=\gamma_{s,d}$ allows us to conclude that

$$\underline{\dim}_R(X)=D_2(X)=m_G. \tag{6.93}$$

In particular, if $D_2(X)=m_G$, then we have the following results for the upper and lower 2-nd quantization coefficient. By (6.91), it follows that

$$\overline{C}_2(X)\le\Gamma\left(1+\frac{2}{m_G}\right)c_{a,b,d,\beta}^{-\frac{2}{m_G}}<\infty, \tag{6.94}$$

and if (6.73) holds for $p=1$, then (6.86) yields

$$\underline{C}_2(X)\ge\frac{m_G}{m_G+2}\Sigma_1^{-\frac{2}{m_G}}c_{a,b,d,\beta}^{-\frac{2}{m_G}}>0. \tag{6.95}$$

7. R-D Theory and Quantization for Self-Similar Sets

We now particularize our results on R-D theory and quantization to random variables taking values in self-similar sets. As in the case of compact manifolds treated in Section 6, what we need here to apply the program developed are suitable sub/super-regularity conditions, which again requires the computation of volume estimates of balls. To this end, we start with preparatory material on contracting similarities largely following the exposition in [22]. Throughout this section, we work in the ambient space $\mathbb{R}^d$ equipped with a norm $\|\cdot\|$. A bijection $s\colon \mathbb{R}^d \to \mathbb{R}^d$ is called a similarity if there exists a $\kappa \in (0,\infty)$ such that

$$\|s(u) - s(v)\| = \kappa\|u - v\| \quad \text{for all } u, v \in \mathbb{R}^d. \tag{7.1}$$

A contracting similarity is a similarity with $\kappa \in (0,1)$, which in this case is referred to as contraction parameter.

Let $\mathcal{I} = \{1,\dots,|\mathcal{I}|\}$ be a finite set of indices. An iterated function system (IFS) is a finite collection $\{s_1,\dots,s_{|\mathcal{I}|}\}$ of contracting similarities $s_i\colon \mathbb{R}^d \to \mathbb{R}^d$ with contraction parameters $\kappa_i \in (0,1)$, $i \in \mathcal{I}$. Let $\{s_1,\dots,s_{|\mathcal{I}|}\}$ be an IFS with contraction parameters $\kappa_i \in (0,1)$, $i \in \mathcal{I}$. The unique positive real number m satisfying

$$\sum_{i\in\mathcal{I}} \kappa_i^m = 1 \tag{7.2}$$

is referred to as similarity dimension. By [17, Theorem 4.1.3], there exists a unique nonempty and compact self-similar set $\mathcal{K}$ satisfying

$$\mathcal{K} = \bigcup_{i\in\mathcal{I}} s_i(\mathcal{K}) \subseteq \mathbb{R}^d. \tag{7.3}$$

Let $\mathcal{I}^* = \bigcup_{j\in\mathbb{N}} \mathcal{I}^j$. For every $j \in \mathbb{N}$ and $\alpha = (i_1,\dots,i_j) \in \mathcal{I}^*$, we set

$$\bar{\alpha} = \begin{cases} (i_1,\dots,i_{j-1}) & \text{if } j > 1 \\ \mathrm{o} & \text{if } j = 1 \end{cases} \tag{7.4}$$

with o denoting the empty sequence, which we declare to have length zero. We designate the identity mapping on $\mathbb{R}^d$ by s_{o}, set $\kappa_{\mathrm{o}} = 1$, and define

$$s_\alpha = s_{i_1} \circ s_{i_2} \circ \cdots \circ s_{i_j} \tag{7.5}$$

$$\kappa_\alpha = \kappa_{i_1}\kappa_{i_2}\dots\kappa_{i_j} \tag{7.6}$$

for all $\alpha \in \mathcal{I}^*$. As $\{s_1,\dots,s_{|\mathcal{I}|}\}$ are contracting similarities $s_i\colon \mathbb{R}^d \to \mathbb{R}^d$ with corresponding contraction parameters $\kappa_i \in (0,1)$, $i \in \mathcal{I}$, it follows from (7.5) and (7.6) that s_α is a contracting similarity with contraction parameter κ_α for all $\alpha \in \mathcal{I}^*$. For every $\delta \in (0,\infty)$ and $y \in \mathbb{R}^d$, we set

$$\mathcal{J}_\delta = \{\alpha \in \mathcal{I}^* : \kappa_\alpha \le \delta < \kappa_{\bar{\alpha}}\} \tag{7.7}$$

and

$$\mathcal{J}_\delta(y) = \Big\{\alpha \in \mathcal{J}_\delta : \mathcal{B}_{\|\cdot\|}\big(y,\delta\big) \cap s_\alpha(\mathcal{K}) \neq \emptyset\Big\}. \tag{7.8}$$

Finally, let

$$\mathcal{E} = \{s_\alpha \circ s_\beta^{-1} : \alpha, \beta \in \mathcal{I}^*, \alpha \neq \beta\} \tag{7.9}$$

and equip the group of all similarities on $\mathbb{R}^d$ with the topology induced by pointwise convergence. The IFS is said to satisfy the weak separation property if (see [65, Definition on p. 3533])

$$s_o \notin \overline{\mathcal{E} \setminus \{s_o\}}. \tag{7.10}$$

The weak separation property guarantees that the IFS does not admit infinitely many overlaps in the following sense:

Lemma 7.1 *[65, Theorem 1, Items (3a) and (5a)] Let $\mathcal{I} = \{1, \dots, |\mathcal{I}|\}$ be a finite set of indices. For every $i \in \mathcal{I}$, let $s_i\colon \mathbb{R}^d \to \mathbb{R}^d$ be a contracting similarity with contraction parameter $\kappa_i \in (0,1)$. Suppose that the self-similar set corresponding to the IFS $\{s_1, \dots, s_{|\mathcal{I}|}\}$ is not contained in any $(d-1)$-dimensional hyperplane. Then, the IFS $\{s_1, \dots, s_{|\mathcal{I}|}\}$ satisfies the weak separation property if and only if for every $y \in \mathbb{R}^d$, there exists an $\ell(y) \in \mathbb{N}$ such that*

$$\Big|\{s_\alpha \circ s_\beta : \|(s_\alpha \circ s_\beta)(y) - x\| < \delta, \alpha \in \mathcal{J}_\delta\}\Big| \leq \ell(y) \tag{7.11}$$

for all $x \in \mathbb{R}^d$, $\delta \in (0,\infty)$, and $\beta \in \mathcal{I}^*$.

The next result establishes subregularity and superregularity for the m-dimensional Hausdorff measure restricted to self-similar sets of similarity dimension m.

Lemma 7.2 *Let $\mathcal{I} = \{1, \dots, |\mathcal{I}|\}$ be a finite set of indices and let $\{s_1, \dots, s_{|\mathcal{I}|}\}$ be an IFS with corresponding contraction parameters κ_i, $i \in \mathcal{I}$. Let furthermore*

$$\mathcal{K} = \bigcup_{i \in \mathcal{I}} s_i(\mathcal{K}) \tag{7.12}$$

be the corresponding self-similar set, denote its similarity dimension by m, and set $\mu = \mathscr{H}^m|_{\mathcal{K}} / \mathscr{H}^m(\mathcal{K})$. Finally, let $k \in (0,\infty)$, $\mathcal{K} \subseteq \mathcal{Y} \subseteq \mathbb{R}^d$, and consider the distortion function $\rho\colon \mathcal{K} \times \mathcal{Y} \to [0,\infty)$, $\rho(x,y) = \|x-y\|^k$, where $\|\cdot\|$ is a norm on $\mathbb{R}^d$. Then, the following statements hold:

(i) Suppose that there exists a $c \in (0,\infty)$ such that $\big|\mathcal{J}_\delta(y)\big| \leq c$ for all $y \in \mathcal{Y}$ and $\delta \in (0,\infty)$. Then, the measure μ satisfies the following subregularity condition:

$$\mu\Big(\mathcal{B}_{\rho^{1/k}}\big(y,\delta\big)\Big) \leq c\delta^m \quad \text{for all } y \in \mathcal{Y} \text{ and } \delta \in (0,\infty). \tag{7.13}$$

(ii) The measure μ satisfies the following superregularity condition:

$$\mu\Big(\mathcal{B}_{\rho^{1/k}}\big(x,\delta\big)\Big) \geq \left(\frac{\kappa_{\min}}{\operatorname{diam}(\mathcal{K})}\right)^m \delta^m \quad \text{for all } x \in \mathcal{K} \text{ and } \delta \in (0,\operatorname{diam}(\mathcal{K})), \tag{7.14}$$

where

$$\operatorname{diam}(\mathcal{K}) = \sup_{x,y \in \mathcal{K}} \|x-y\| \tag{7.15}$$

and $\kappa_{\min} = \min\{\kappa_1, \ldots, \kappa_{|\mathcal{I}|}\}$.

(iii) If the IFS satisfies the weak separation property (7.10) *and* $\mathcal{K}$ *is not contained in any hyperplane of dimension* $d-1$, *then* $0 < \mathscr{H}^m(\mathcal{K}) < \infty$ *and there exists a* $c \in (0,\infty)$ *such that* $\left|\mathcal{J}_\delta(y)\right| \leq c$ *for all* $y \in \mathbb{R}^d$ *and* $\delta \in (0,\infty)$.

Proof The proof follows from the corresponding parts of the proof of [22, Theorem 2.1]. □

In the remainder of this section, we fix a norm $\|\cdot\|$ on $\mathbb{R}^d$ and a $k \in (0,\infty)$. Further, we consider an IFS satisfying the weak separation poperty (7.10) and such that the corresponding self-similar set $\mathcal{K} \subseteq \mathbb{R}^d$ is not contained in any hyperplane of dimension $d-1$. In particular, by Item (iii) of Lemma 7.2, we then have $0 < \mathscr{H}^m(\mathcal{K}) < \infty$, and hence $\dim_{\mathrm{H}}(\mathcal{K}) = m$ with m denoting the similarity dimension of $\mathcal{K}$. Moreover, again by Item (iii) of Lemma 7.2, there must exist a $c \in (0,\infty)$ such that

$$\left|\mathcal{J}_\delta(y)\right| \leq c \quad \text{for all } y \in \mathbb{R}^d \text{ and } \delta \in (0,\infty). \tag{7.16}$$

Finally, we consider a random variable X taking values in $\mathcal{K} \subseteq \mathcal{Y} \subseteq \mathbb{R}^d$ and the distortion function $\rho\colon \mathcal{K} \times \mathcal{Y} \to [0,\infty)$, $\rho(x,y) = \|x-y\|^k$.

We first derive a lower bound on the R-D function under the assumptions

$$\mu_X \ll \mu := \mathscr{H}^m|_{\mathcal{K}} / \mathscr{H}^m(\mathcal{K}) \tag{7.17}$$

and $h_\mu(X) > -\infty$. Jensen's inequality [51, Theorem 2.3] combined with $\mu(\mathcal{K}) = 1$ yields $h_\mu(X) \leq \log(\mu(\mathcal{K})) = 0$, which in turn implies $\left|h_\mu(X)\right| < \infty$. Application of Theorem 3.1 for $\delta_0 = \infty$ combined with (7.16) and Item (i) of Lemma 7.2 yields $R_X^{\mathrm{SLB}}(D) \geq R_X^{\mathrm{L}}(D)$ with

$$R_X^{\mathrm{L}}(D) = h_\mu(X) + F_{m,k,c}(D) \quad \text{for all } D \in (0,\infty). \tag{7.18}$$

Next, we derive a lower bound on the n-th quantization error under the assumptions $\mu_X \ll \mu$ and

$$\Sigma_p(X) := \left\| \frac{\mathrm{d}\mu_X}{\mathrm{d}\mu} \right\|_{p/(p-1)}^{(\mu)} < \infty \quad \text{with } p \in [1,\infty). \tag{7.19}$$

Application of Theorem 4.1 for $\delta_0 = \infty$ combined with (7.16) and Item (i) of Lemma 7.2 yields $V_n(X) \geq L_n(X)$ for all $n \in \mathbb{N}$ with

$$L_n(X) = \frac{m}{m+pk} c^{-\frac{k}{m}} \Sigma_p^{-\frac{pk}{m}}(X) n^{-\frac{pk}{m}}. \tag{7.20}$$

In particular, we have

$$\liminf_{n\to\infty} n^{\frac{pk}{m}} V_n(X) \geq \frac{m}{m+pk} c^{-\frac{k}{m}} \Sigma_p^{-\frac{pk}{m}}(X) > 0. \tag{7.21}$$

We now derive an upper bound U_n on the n-th quantization error $V_n(X)$. Application of Theorem 4.2 with $\beta = \operatorname{diam}(\mathcal{K})$, $\nu = \mu$, and $\delta_0 = \operatorname{diam}(\mathcal{K})$ combined with Item (ii) of Lemma 7.2 yields $V_n(X) \leq U_n$ for all $n \in \mathbb{N}$ with

$$U_n = \Gamma\left(1 + \frac{k}{m}\right)\left(\frac{\operatorname{diam}(\mathcal{K})}{\kappa_{\min}}\right)^k n^{-\frac{k}{m}}. \tag{7.22}$$

In particular, we have

$$\limsup_{n\to\infty} n^{\frac{k}{m}} V_n(X) \le \Gamma\left(1+\frac{k}{m}\right)\left(\frac{\operatorname{diam}(\mathcal{K})}{\kappa_{\min}}\right)^k < \infty. \tag{7.23}$$

Combining (7.21) and (7.23) with Items (i) and (ii) of Lemma 4.2, respectively, now yields

$$\frac{m}{p} \le \underline{D}_k(X) \le \overline{D}_k(X) \le m. \tag{7.24}$$

If $p=1$, then the k-th quantization dimension $D_k(X)$ exists and equals the similarity dimension m. Similarly, if $h_\mu(X) > -\infty$ (Jensen's inequality [51, Theorem 2.3] combined with $\mu(\mathcal{X}) = 1$ yields $h_\mu(X) \le \log(\mu(\mathcal{X})) = 0$), then Item (ii) of Corollary 4.4 (with $\mathcal{X} = \mathcal{K}$ and $\nu = \mu$) allows us to conclude that

$$\underline{\dim}_R(X) = D_k(X) = m. \tag{7.25}$$

In particular, if $D_k(X) = m$, then we obtain the following results for the upper and lower k-th quantization coefficients. By (7.23), we have

$$\overline{C}_k(X) \le \Gamma\left(1+\frac{k}{m}\right)\left(\frac{\operatorname{diam}(\mathcal{K})}{\kappa_{\min}}\right)^k < \infty, \tag{7.26}$$

and if (7.19) holds for $p=1$, then (7.21) yields

$$\underline{C}_k(X) \ge \frac{m}{m+k} c^{-\frac{k}{m}} \Sigma_1^{-\frac{k}{m}}(X) > 0. \tag{7.27}$$

For $\mathcal{Y} = \mathcal{K}$, if in addition to $D_k(X) = m$, we have $m > k$, $\mu \ll \mu_X$, and $\|\mathrm{d}\mu/\mathrm{d}\mu_X\|_\infty^{(\mu_X)} < \infty$, then we can apply Theorem 4.3 with $\nu = \mu$ to obtain the improved upper bound

$$\overline{C}_k(X) \le \Omega_{k/m}(X)\Gamma\left(1+\frac{k}{m}\right)\left(\frac{\operatorname{diam}(\mathcal{K})}{\kappa_{\min}}\right)^k, \tag{7.28}$$

where

$$\Omega_{k/m}(X) = \mathbb{E}\left[\left(\frac{\mathrm{d}\mu}{\mathrm{d}\mu_X}(X)\right)^{\frac{k}{m}}\right] \quad \text{for all } m > k \tag{7.29}$$

with strict inequality in (7.29) unless $\mu_X = \mu$.

In order to further quantify our results, the following example considers a random variable X taking values in the middle third Cantor set.

Example 7.1. Consider the middle third Cantor set $\mathcal{C} \subseteq [0,1]$, i.e., the self-similar set corresponding to the choice $\mathcal{I} = \{1,2\}$, $\kappa_1 = \kappa_2 = 1/3$, $s_1(x) = x/3$, and $s_2(x) = x/3 + 2/3$. This self-similar set has similarity dimension $m_{\mathcal{C}} = \log(2)/\log(3)$ and satisfies $0 < \mathscr{H}^{m_{\mathcal{C}}}(\mathcal{C}) < \infty$ [19, Example 4.5]. Let $\mu = \mathscr{H}^{m_{\mathcal{C}}}|_{\mathcal{C}}/\mathscr{H}^{m_{\mathcal{C}}}(\mathcal{C})$. We now employ Lemma 7.2 to establish the subregularity dimension and the subregularity constants of μ for $\mathcal{Y} = \mathcal{C}$ and $\mathcal{Y} = \mathbb{R}$ with

$$\rho\colon \mathcal{C} \times \mathcal{Y} \to [0,\infty), \;\; \rho(x,y) = |x-y|^2. \tag{7.30}$$

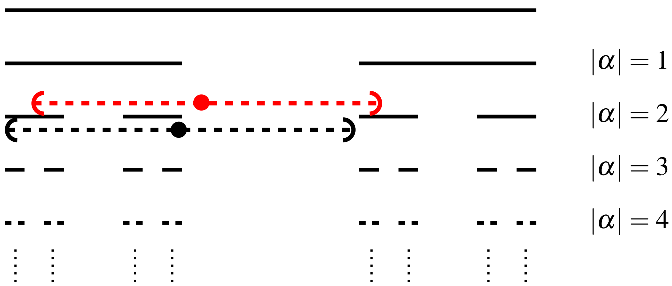

FIG. 1. The sets $s_\alpha([0,1])$ with $|\alpha| = j$ have length 3^{-j}. If $y \in \mathbb{R}$, then at most three different sets $s_\alpha([0,1])$ with $|\alpha| = j$ intersect an open interval of length $2(3^{-j+1})$ centered at y (depicted in red for $j = 2$). If $y \in \mathcal{C}$, then at most two different sets $s_\alpha([0,1])$ with $|\alpha| = j$ intersect an open interval of length $2(3^{-j+1})$ centered at y (depicted in black for $j = 2$).

Specifically, note that $\kappa_\alpha = 3^{-j}$ for all $\alpha = (i_1, \dots, i_j)$ and $j \in \mathbb{N}$. Thus,

$$\mathcal{J}_\delta = \{\alpha \in \mathcal{I}^* : \kappa_\alpha \le \delta < \kappa_{\bar{\alpha}}\} \tag{7.31}$$

$$= \{\alpha : |\alpha| = j\} \quad \text{for all } \delta \in \left[3^{-j}, 3^{-j+1}\right) \text{ and } j \in \mathbb{N}, \tag{7.32}$$

which implies (see Figure 1)

$$\left|\mathcal{J}_\delta(y)\right| \le c_{\mathcal{C}} \quad \text{for all } \delta \in (0,1) \text{ and } y \in \mathcal{Y} \tag{7.33}$$

with

$$c_{\mathcal{C}} = \begin{cases} 2 & \text{if } \mathcal{Y} = \mathcal{C} \\ 3 & \text{if } \mathcal{Y} = \mathbb{R}. \end{cases} \tag{7.34}$$

Therefore, by Item (i) in Lemma 7.2, the subregularity dimension of μ is $m_{\mathcal{C}}$ and the corresponding subregularity constants are given by $c_{\mathcal{C}}$ as defined in (7.34) and $\delta_0 = \infty$.

Now, suppose that the random variable X takes values in $\mathcal{C}$. If $\mu_X \ll \mu$ with $h_\mu(X) > -\infty$, then (7.18) evaluated for $\mathcal{K} = \mathcal{C}$, $m = m_{\mathcal{C}}$, $c = c_{\mathcal{C}}$, and $k = 2$ yields the R-D lower bound

$$R_X^{\mathrm{L}}(D) = h_\mu(X) - \frac{m_{\mathcal{C}}}{2} + \frac{m_{\mathcal{C}}}{2}\log\left(\frac{m_{\mathcal{C}}}{2D}\right) - \log\left(c_{\mathcal{C}}\Gamma\left(1 + \frac{m_{\mathcal{C}}}{2}\right)\right). \tag{7.35}$$

If $\mu_X \ll \mu$ with (7.19) satisfied for $p = 1$ or $\left|h_\mu(X)\right| < \infty$, then $D_2(X) = m_{\mathcal{C}}$ owing to (7.24). Furthermore, (7.26) and (7.27) particularized to $\mathcal{K} = \mathcal{C}$, $m = m_{\mathcal{C}}$, $c = c_{\mathcal{C}}$, $k = 2$, $\kappa_{\min} = 1/3$, and $\operatorname{diam}(\mathcal{C}) = 1$ yields the following chain of inequalities for the lower and upper 2-nd quantization coefficient

$$0 < \frac{m_{\mathcal{C}}}{m_{\mathcal{C}} + 2}\Sigma_1^{-\frac{2}{m_{\mathcal{C}}}}(X)\, c_{\mathcal{C}}^{-\frac{2}{m_{\mathcal{C}}}} \le \underline{C}_2(X) \le \overline{C}_2(X) \le 9\Gamma\left(1 + \frac{2}{m_{\mathcal{C}}}\right) < \infty. \tag{7.36}$$

We finally note that the n-th quantization error for the special case of X taking values in $\mathcal{C}$ with uniform distribution $\mu_X = \mu$ and $\mathcal{Y} = \mathbb{R}$ is known explicitly and equals [26, Theorem 5.2]

$$V_n(X) = \frac{1}{18^{l_n}}\frac{1}{8}\left(2^{l_n+1} - n + \frac{1}{9}\left(n - 2^{l_n}\right)\right), \tag{7.37}$$

where $l_n = \lfloor \log(n)/\log(2) \rfloor$ for all $n \in \mathbb{N}$. The set of accumulation points of the sequence $V_n(X)n^{2/m_C}$ is given by the interval $[1/8, f(17/(8+4m_C))]$, where $f(s) = (1/72)s^{2/m_C}(17-8s)$ [26, Theorem 6.3], which implies $\underline{C}_2(X) = 1/8$ and $\overline{C}_2(X) = f(17/(8+4m_C))$. In particular, $C_2(X)$ does not exist.

8. Future Work and Open Problems

An interesting research direction is the identification of concrete examples of non-i.i.d. stationary ergodic processes $(X_i)_{i\in\mathbb{N}}$ which allow explicit expressions for the corresponding Shannon lower bounds according to Theorem 3.1. Specifically, suppose that there exists a subregular measure μ on $(\mathcal{X}, \mathscr{X})$ such that $\mu_{X_1} \times \cdots \times \mu_{X_\ell} \ll \mu^{(\ell)}$ with

$$\mu^{(\ell)} = \underbrace{\mu \times \cdots \times \mu}_{\ell \text{ times}} \quad \text{for all } \ell \in \mathbb{N}. \tag{8.1}$$

Then, Proposition 2.1 can be used to infer subregularity of $\mu^{(\ell)}$ from subregularity of μ for all $\ell \in \mathbb{N}$, and the lower bound $R^{\mathrm{L}}_{X^{(\ell)}}(D)$ in (3.7) can be evaluated for all sections $X^{(\ell)} = (X_1, \dots, X_\ell)$ of $(X_i)_{i\in\mathbb{N}}$. The resulting Shannon lower bound is then explicit and depends on the subregularity dimension and the subregularity parameters of μ. To the best of our knowledge, results on such Shannon lower bounds for non-i.i.d. stationary ergodic processes have not been reported in the literature.

Another open problem is related to the assumptions in Theorem 4.1. Specifically, in (D.4) in the proof of Theorem 4.1, subregularity of μ_X is obtained from subregularity of μ by application of Hölder's inequality [35, Theorem 1, p. 372]. A corollary to Theorem 4.1, namely Corollary 4.1, then allows us to conclude that, if there exists a $p \in [1, \infty)$ such that

$$\|\mathrm{d}\mu_X/\mathrm{d}\mu\|^{(\mu)}_{p/(p-1)} < \infty, \tag{8.2}$$

we get the lower bound

$$\underline{D}_k(X) \geq m/p. \tag{8.3}$$

For $p > 1$, this bound is strictly smaller than the subregularity dimension m. Note that if

$$\|\mathrm{d}\mu_X/\mathrm{d}\mu\|^{(\mu)}_{q/(q-1)} = \infty \quad \text{for all } q \in [1, p), \tag{8.4}$$

then (8.3) is also the largest possible lower bound on $\underline{D}_k(X)$ that can be obtained by application of Theorem 4.1. It would now be interesting to understand if there is a concrete example saturating (8.3) for $p \in (1, \infty)$ or whether the somewhat crude step of applying Hölder's inequality in the proof of Theorem 4.1 renders the lower bound in Item (i) of Corollary 4.1 structurally loose. Two broad classes of random variables can, however, already be excluded from consideration as candidates by the following arguments:

- If $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$, $\mu = \mathscr{L}^d$, and $k \in [1, \infty)$, then [27, Theorem 6.2, Remark 6.3] imply $D_k(X) = d$ provided that

$$\|\mathrm{d}\mu_X/\mathrm{d}\mu\|^{(\mu)}_{d/(d+k)} < \infty. \tag{8.5}$$

  This hence rules out all continuous random variables satisfying (8.5).

- If μ is subregular of subregularity dimension m with $\mu(\mathcal{X}) = 1$ for $\delta_0 < \infty$, then Corollary 4.2 implies $\underline{D}_k(X) \geq m$ provided that $\left|h_\mu(X)\right| < \infty$. This hence rules out every X of finite generalized entropy.

Finally, it would be interesting to understand whether the link between quantization and R-D theory established in Section 5 extends to the case where the subregularity parameter δ_0 is finite, which is, e.g., the case for the compact manifolds described in Section 6. The problem with $\delta_0 < \infty$ here is that the lower bound $R_{(\ell)}(D) \geq \widetilde{R}_{(\ell)}(D)$ in Theorem 5.1 is valid for $D \in (0, D_{(\ell)})$, but unfortunately $D_{(\ell)} \to 0$ as $\ell \to \infty$ unless $\delta_0 = \infty$.

A. Proof of Proposition 2.1

We prove Item (i) only as the proof of Item (ii) follows along the exact same lines using superregularity of the measures ν_i instead of subregularity of the μ_i. First, we consider the special case $\ell = 2$, $\alpha_1 = \alpha_2 = k = 1$, which constitutes the core piece of the proof. The general case then follows by induction over ℓ and subsequent application of Lemma 2.2 to extend the result to arbitrary $k, \alpha_1, \ldots, \alpha_\ell \in (0,\infty)$.

Lemma A.1 *For $i = 1,2$, let $(\mathcal{X}_i, \mathscr{X}_i)$ and $(\mathcal{Y}_i, \mathscr{Y}_i)$ be measurable spaces and consider the distortion function $\rho_i\colon \mathcal{X}_i \times \mathcal{Y}_i \to [0,\infty]$. Suppose that, for $i = 1,2$, μ_i is a σ-finite ρ_i-subregular measure on $(\mathcal{X}_i, \mathscr{X}_i)$ of dimension $m_i \in (0,\infty)$ with subregularity constants $c_i \in (0,\infty)$ and $\delta_i \in (0,\infty]$, and set*

$$\bar{\rho}((x_1,x_2),(y_1,y_2)) = \rho_1(x_1,y_1) + \rho_2(x_2,y_2). \tag{A.1}$$

Then, $\bar{\mu} = \mu_1 \otimes \mu_2$ is $\bar{\rho}$-subregular of dimension $m_1 + m_2$ with subregularity constants $c_1c_2\Gamma(1+m_1)\Gamma(1+m_2)/\Gamma(1+m_1+m_2)$ and $\min(\delta_1,\delta_2)$, i.e.,

$$\bar{\mu}\left(\mathcal{B}_{\bar{\rho}}(y,\delta)\right) \leq c_1c_2\frac{\Gamma(1+m_1)\Gamma(1+m_2)}{\Gamma(1+m_1+m_2)}\delta^{m_1+m_2} \tag{A.2}$$

for all $y \in \mathcal{Y}_1 \times \mathcal{Y}_2$ and $\delta \in (0, \min(\delta_1,\delta_2))$.

Proof Fix $y = (y_1,y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2$ and $\delta \in (0,\min(\delta_1,\delta_2))$ arbitrarily. We have

$$\bar{\mu}\left(\mathcal{B}_{\bar{\rho}}(y,\delta)\right) = \bar{\mu}\left((x_1,x_2) \in \mathcal{X}_1 \times \mathcal{X}_2 : \rho_1(x_1,y_1) + \rho_2(x_2,y_2) < \delta\right) \tag{A.3}$$

$$= \int I(x_1)\,\mathrm{d}\mu_1(x_1), \tag{A.4}$$

where (A.4) is by Tonelli's theorem for characteristic functions [5, Theorem 10.9] with

$$I(x_1) = \mu_2\left(\mathcal{B}_{\rho_2}\left(y_2, \left(\delta - \min\left(\rho_1(x_1,y_1),\delta\right)\right)\right)\right). \tag{A.5}$$

Now, ρ_2-subregularity of μ_2 implies

$$I(x_1) \leq c_2 g(x_1) \tag{A.6}$$

with

$$g(x_1) = \Big(\delta - \min\big(\rho_1(x_1,y_1),\delta\big)\Big)^{m_2}, \tag{A.7}$$

which, when used in (A.4), yields

$$\bar{\mu}\big(\mathcal{B}_{\bar{\rho}}(y,\delta)\big) \le c_2 \int g(x_1)\,\mathrm{d}\mu_1(x_1). \tag{A.8}$$

Next, note that

$$\int g(x_1)\,\mathrm{d}\mu_1(x_1) \tag{A.9}$$

$$= \int_0^{\delta^{m_2}} \mu_1(\{x_1 : g(x_1) > t\})\,\mathrm{d}t \tag{A.10}$$

$$= \int_0^{\delta^{m_2}} \mu_1\left(\mathcal{B}_{\rho_1}\left(y_1, \left(\delta - t^{\frac{1}{m_2}}\right)\right)\right)\mathrm{d}t \tag{A.11}$$

$$\le c_1 \int_0^{\delta^{m_2}} \left(\delta - t^{\frac{1}{m_2}}\right)^{m_1} \mathrm{d}t \tag{A.12}$$

$$= c_1\delta^{m_1} \int_0^{\delta^{m_2}} \left(1 - \delta^{-1} t^{\frac{1}{m_2}}\right)^{m_1} \mathrm{d}t \tag{A.13}$$

$$= c_1\delta^{m_1+m_2} m_2 \int_0^1 (1-s)^{m_1} s^{m_2-1}\,\mathrm{d}s \tag{A.14}$$

$$= c_1\delta^{m_1+m_2} m_2 B_{1+m_1,m_2}(1), \tag{A.15}$$

where (A.10) follows from Lemma K.1 upon noting that $0 \le g(x_1) \le \delta^{m_2}$ for all $x_1 \in \mathcal{X}_1$, in (A.12) we employed ρ_1-subregularity of μ_1, and in (A.14) we changed variables according to $s = \delta^{-1} t^{\frac{1}{m_2}}$. Using (A.9)–(A.15) in (A.8), we obtain

$$\bar{\mu}\big(\mathcal{B}_{\bar{\rho}}(y,\delta)\big) \le c_1 c_2 \delta^{m_1+m_2} m_2 B_{1+m_1,m_2}(1) \tag{A.16}$$

$$= c_1 c_2 \frac{\Gamma(1+m_1)\Gamma(1+m_2)}{\Gamma(1+m_1+m_2)} \delta^{m_1+m_2}, \tag{A.17}$$

where the last step follows from (1.31) and $m_2\Gamma(m_2) = \Gamma(1+m_2)$. $\square$

The generalization of Lemma A.1 to arbitrary ℓ is effected by induction as follows.

Lemma A.2 *For $i = 1,\dots,\ell$, let $(\mathcal{X}_i,\mathscr{X}_i)$ and $(\mathcal{Y}_i,\mathscr{Y}_i)$ be measurable spaces and consider the distortion function $\rho_i\colon \mathcal{X}_i \times \mathcal{Y}_i \to [0,\infty]$. Suppose that, for $i = 1,\dots,\ell$, μ_i is a σ-finite ρ_i-subregular measure of dimension $m_i \in (0,\infty)$ with subregularity constants $c_i \in (0,\infty)$ and $\delta_i \in (0,\infty]$, i.e.,*

$$\mu_i\big(\mathcal{B}_{\rho_i}(y_i,\delta)\big) \le c_i \delta^{m_i} \quad \text{for all } y_i \in \mathcal{Y}_i \text{ and } \delta \in (0,\delta_i). \tag{A.18}$$

Then, $\mu^{(\ell)} := \mu_1 \otimes \cdots \otimes \mu_\ell$ *satisfies*

$$\mu^{(\ell)}\left(\mathcal{B}_{\tilde{\rho}_{(\ell)}}\left(y^{(\ell)}, \delta\right)\right) \leq \tilde{c}_{(\ell)} \delta^{\sum_{i=1}^{\ell} m_i} \tag{A.19}$$

for all $y^{(\ell)} \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_\ell$ *and* $\delta \in (0, \min(\delta_1, \ldots, \delta_\ell))$ *with*

$$\tilde{\rho}_{(\ell)}((x_1, \ldots, x_\ell), (y_1, \ldots, y_\ell)) = \sum_{i=1}^{\ell} \rho_i(x_i, y_i), \tag{A.20}$$

and

$$\tilde{c}_{(\ell)} = \frac{\prod_{i=1}^{\ell} \Gamma(1+m_i)}{\Gamma\left(1+\sum_{i=1}^{\ell} m_i\right)} \times \prod_{i=1}^{\ell} c_i. \tag{A.21}$$

Proof The proof is by induction on ℓ. The base case $\ell = 2$ was established in Lemma A.1. For the induction step, assume that (A.19) holds for $\ell - 1$, i.e.,

$$\mu^{(\ell-1)}\left(\mathcal{B}_{\tilde{\rho}_{(\ell-1)}}\left(y^{(\ell-1)}, \delta\right)\right) \leq \tilde{c}_{(\ell-1)} \delta^{\sum_{i=1}^{\ell-1} m_i} \tag{A.22}$$

for all $y^{(\ell-1)} \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_{\ell-1}$ and $\delta \in (0, \min(\delta_1, \ldots, \delta_{\ell-1}))$. Since

$$\tilde{\rho}_{(\ell)}((x_1, \ldots, x_\ell), (y_1, \ldots, y_\ell)) \tag{A.23}$$

$$= \tilde{\rho}_{(\ell-1)}((x_1, \ldots, x_{\ell-1}), (y_1, \ldots, y_{\ell-1})) + \rho_\ell(x_\ell, y_\ell) \tag{A.24}$$

and $\mu^{(\ell)} = \mu^{(\ell-1)} \otimes \mu_\ell$ with $\mu^{(\ell-1)}$ satisfying (A.22) and μ_ℓ obeying (A.18), Lemma A.1 applied to $\mu^{(\ell-1)}$, μ_ℓ, $\tilde{\rho}_{(\ell-1)}$, and ρ_ℓ yields

$$\mu^{(\ell)}\left(\mathcal{B}_{\tilde{\rho}_{(\ell)}}\left(y^{(\ell)}, \delta\right)\right) \leq \tilde{c}_{(\ell-1)} c_\ell \frac{\Gamma\left(1+\sum_{i=1}^{\ell-1} m_i\right)\Gamma(1+m_\ell)}{\Gamma\left(1+\left(\sum_{i=1}^{\ell-1} m_i\right)+m_\ell\right)} \delta^{(\sum_{i=1}^{\ell-1} m_i)+m_\ell} \tag{A.25}$$

$$= \tilde{c}_{(\ell)} \delta^{\sum_{i=1}^{\ell} m_i} \tag{A.26}$$

for all $y^{(\ell)} \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_\ell$ and $\delta \in (0, \min(\delta_1, \ldots, \delta_\ell))$. □

To complete the proof of Proposition 2.1, we need to generalize Lemma A.2 to arbitrary $k \in (0, \infty)$ and arbitrary $\alpha_1, \ldots, \alpha_\ell \in (0, \infty)$. To this end, note that $\rho_i^{1/k}$-subregularity of μ_i, by Lemma 2.2 implies that μ_i is $(\alpha_i \rho_i)$-subregular of dimension m_i/k with subregularity constants $\tilde{c}_i = c_i \alpha_i^{-m_i/k}$ and $\tilde{\delta}_i = \alpha_i \delta_i^k$. Lemma A.2 applied to the $(\alpha_i \rho_i)$-subregular measures μ_i and the distortion functions $\alpha_i \rho_i$ therefore yields that $\mu^{(\ell)}$ is $\rho_{(\ell)}$-subregular of dimension $m_{(\ell)}/k$ with subregularity constants $c_{(\ell)}$ (as defined in (2.33)) and $\min(\alpha_1 \delta_1^k, \ldots, \alpha_\ell \delta_\ell^k)$. Another application of Lemma 2.2 then establishes that $\mu^{(\ell)} = \mu_1 \otimes \cdots \otimes \mu_\ell$ is $\rho_{(\ell)}^{1/k}$-subregular of dimension $m_{(\ell)}$ with subregularity constants $c_{(\ell)}$ and $\delta_{(\ell)}$ as defined in (2.34). □

B. Proof of Proposition 3.1

By [36, Lemma A.1] we can lower-bound $R_X(D)$ according to[6]

$$R_X(D) \geq \sup_{\alpha,s}\big(\mathbb{E}[\log(\alpha(X))] - sD\big) \quad \text{for all } D \in (0,\infty), \tag{B.1}$$

where the supremization is over all measurable functions $\alpha\colon \mathcal{X} \to (0,\infty)$ and $s \in [0,\infty)$ such that

$$\mathbb{E}\Big[\alpha(X)e^{-s\rho(X,y)}\Big] \leq 1 \quad \text{for all } y \in \mathcal{Y}. \tag{B.2}$$

The main idea of the proof is to lower-bound the supremum in (B.1) using a judiciously parametrized family $(\alpha_s)_{s\geq 0}$ of measurable functions $\alpha_s\colon \mathcal{X} \to (0,\infty)$ satisfying

$$\mathbb{E}\Big[\alpha_s(X)e^{-s\rho(X,y)}\Big] \leq 1 \quad \text{for all } s \in [0,\infty) \text{ and } y \in \mathcal{Y}. \tag{B.3}$$

Specifically, for every $s \in [0,\infty)$, we set

$$\alpha_s(x) = \begin{cases} \left(\dfrac{\mathrm{d}\mu_X}{\mathrm{d}\mu}(x)g(s)\right)^{-1} & \text{if } \dfrac{\mathrm{d}\mu_X}{\mathrm{d}\mu}(x) > 0 \\ 1 & \text{else,} \end{cases} \tag{B.4}$$

where $(\mathrm{d}\mu_X/\mathrm{d}\mu)(x)$ exists and is nonnegative and measurable by the Radon–Nikodým theorem [5, Theorem 8.9]. Now (B.3) is established by noting that

$$\begin{aligned} \mathbb{E}\Big[\alpha_s(X)e^{-s\rho(X,y)}\Big] &= \int \alpha_s(x)e^{-s\rho(x,y)}\,\mathrm{d}\mu_X(x) && \text{(B.5)} \\ &= \int_{\mathcal{A}} \alpha_s(x)e^{-s\rho(x,y)}\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu}(x)\,\mathrm{d}\mu(x) && \text{(B.6)} \\ &= g(s)^{-1}\int_{\mathcal{A}} e^{-s\rho(x,y)}\,\mathrm{d}\mu(x) && \text{(B.7)} \\ &\leq 1 \quad \text{for all } s \in [0,\infty) \text{ and } y \in \mathcal{Y}, && \text{(B.8)} \end{aligned}$$

where in (B.6) we set

$$\mathcal{A} = \left\{x \in \mathcal{X} : \frac{\mathrm{d}\mu_X}{\mathrm{d}\mu}(x) > 0\right\}, \tag{B.9}$$

(B.7) follows from (B.4), and (B.8) is by (3.2). It therefore follows from (B.1) that

$$R_X(D) \geq \sup_{s\geq 0}\big(\mathbb{E}[\log(\alpha_s(X))] - sD\big) \quad \text{for all } D \in (0,\infty). \tag{B.10}$$

Finally,

$$\mathbb{E}[\log(\alpha_s(X))] \tag{B.11}$$

[6] Imposing slightly stronger constraints on the distortion function ρ than just measurability, (B.1) can actually be made to hold with equality [12, Theorem 2.3].

$$= \int \log(\alpha_s(x))\, \mathrm{d}\mu_X(x) \tag{B.12}$$

$$= \int \log(\alpha_s(x)) \frac{\mathrm{d}\mu_X}{\mathrm{d}\mu}(x)\, \mathrm{d}\mu(x) \tag{B.13}$$

$$= -\int \log\left(\frac{\mathrm{d}\mu_X}{\mathrm{d}\mu}(x)\right) \frac{\mathrm{d}\mu_X}{\mathrm{d}\mu}(x)\, \mathrm{d}\mu(x) - \log(g(s)) \int \frac{\mathrm{d}\mu_X}{\mathrm{d}\mu}(x)\, \mathrm{d}\mu(x) \tag{B.14}$$

$$= h_\mu(X) - \log(g(s)) \quad \text{for all } s \in [0,\infty), \tag{B.15}$$

where (B.14) follows from (B.4) and (B.15) holds owing to $\int \frac{\mathrm{d}\mu_X}{\mathrm{d}\mu}(x)\, \mathrm{d}\mu(x) = \mu_X(\mathcal{X}) = 1$. Inserting (B.11)–(B.15) into (B.10) concludes the proof. □

C. Proof of Theorem 3.1

We start by upper-bounding $g(s)$ in Proposition 3.1 as follows:

$$g(s) = \sup_{y\in\mathcal{Y}} \int e^{-s\rho(x,y)}\, \mathrm{d}\mu(x) \tag{C.1}$$

$$= \sup_{y\in\mathcal{Y}} \int_0^1 \mu\left(\left\{x \in \mathcal{X} : e^{-s\rho(x,y)} > t\right\}\right) \mathrm{d}t \tag{C.2}$$

$$= \sup_{y\in\mathcal{Y}} \int_0^\infty e^{-u} \mu\left(\left\{x \in \mathcal{X} : u > s\rho(x,y)\right\}\right) \mathrm{d}u \tag{C.3}$$

$$= \sup_{y\in\mathcal{Y}} \int_0^\infty e^{-u} \mu\left(\mathcal{B}_\rho\left(y, \frac{u}{s}\right)\right) \mathrm{d}u \tag{C.4}$$

$$= \sup_{y\in\mathcal{Y}} \left(\int_0^{s\delta_0^k} e^{-u} \mu\left(\mathcal{B}_\rho\left(y, \frac{u}{s}\right)\right) \mathrm{d}u + \int_{s\delta_0^k}^\infty e^{-u} \mu\left(\mathcal{B}_\rho\left(y, \frac{u}{s}\right)\right) \mathrm{d}u\right) \tag{C.5}$$

$$\le c s^{-\frac{m}{k}} \int_0^{s\delta_0^k} e^{-u} u^{\frac{m}{k}}\, \mathrm{d}u + e^{-s\delta_0^k} \tag{C.6}$$

$$= c s^{-\frac{m}{k}} \gamma\left(1 + \frac{m}{k}, s\delta_0^k\right) + e^{-s\delta_0^k} \quad \text{for all } s \in (0,\infty), \tag{C.7}$$

where (C.2) follows from Lemma K.1 upon noting that $e^{-s\rho(x,y)} \in [0,1]$, in (C.3) we applied the change of variables $t = e^{-u}$, and in (C.6) we used that i) thanks to Lemma 2.2, μ is ρ-subregular of dimension m/k with subregularity constants c and δ_0^k and ii) the assumption $\mu(\mathcal{X}) = 1$ if $\delta_0 < \infty$.

Inserting (C.1)–(C.7) into the right-hand side of (3.1) yields

$$R_X^{\mathrm{SLB}}(D) \ge h_\mu(X) - \inf_{s\ge 0} \left(sD + \log\left(c s^{-\frac{m}{k}} \gamma\left(1 + \frac{m}{k}, s\delta_0^k\right) + e^{-s\delta_0^k}\right)\right) \tag{C.8}$$

$$\ge h_\mu(X) - \frac{m}{k} - \log\left(c \left(\frac{m}{kD}\right)^{-\frac{m}{k}} \gamma\left(1 + \frac{m}{k}, \frac{m\delta_0^k}{kD}\right) + e^{-\frac{m\delta_0^k}{kD}}\right), \tag{C.9}$$

where in (C.9) we set $s = m/(kD)$. Since $\gamma(1+m/k, m\delta_0^k/(kD)) \leq \Gamma(1+m/k)$, this establishes the lower bounds in (3.8) and (3.9). Note that the only bounding step in the derivation of (3.8) arises from the application of ρ-subregularity of μ in (C.6) as for $\delta_0 = \infty$ we have $\gamma(1+m/k, m\delta_0^k/(kD)) = \Gamma(1+m/k)$ and the infimum in (C.8) is attained at $s = m/(kD)$. The bound in (3.10) and the limit in (3.11) follow trivially. □

D. Proof of Theorem 4.1

Let $n \in \mathbb{N}$ and $f \in \mathcal{F}_n(\mathcal{X}, \mathcal{Y})$ be arbitrary but fixed. We have

$$\mu_X\Big(\big\{x \in \mathcal{X} : \rho(x, f(x)) \geq \delta\big\}\Big) \geq \mu_X\Big(\big\{x \in \mathcal{X} : \rho(x,y) \geq \delta \text{ for all } y \in f(\mathcal{X})\big\}\Big) \tag{D.1}$$

$$= \mu_X(\mathcal{X}) - \mu_X\left(\bigcup_{y \in f(\mathcal{X})} \mathcal{B}_\rho(y,\delta)\right) \tag{D.2}$$

$$\geq 1 - \sum_{y \in f(\mathcal{X})} \mu_X\big(\mathcal{B}_\rho(y,\delta)\big) \tag{D.3}$$

$$\geq 1 - \sum_{y \in f(\mathcal{X})} c^{\frac{1}{p}} \left\| \frac{\mathrm{d}\mu_X}{\mathrm{d}\mu} \right\|_{p/(p-1)}^{(\mu)} \delta^{\frac{m}{pk}} \tag{D.4}$$

$$\geq 1 - nc^{\frac{1}{p}} \left\| \frac{\mathrm{d}\mu_X}{\mathrm{d}\mu} \right\|_{p/(p-1)}^{(\mu)} \delta^{\frac{m}{pk}} \tag{D.5}$$

$$= 1 - \left(\frac{\delta}{h(n)}\right)^{\frac{m}{pk}} \quad \text{for all } \delta \in (0, \delta_0^k), \tag{D.6}$$

where (D.3) follows from $\mu_X(\mathcal{X}) = 1$ and the union bound, in (D.4) we applied Lemmata 2.2 and 2.4, in (D.5) we used $|f(\mathcal{X})| \leq n$, and in (D.6) we set

$$h(n) = c^{-\frac{k}{m}} \left(\left\| \frac{\mathrm{d}\mu_X}{\mathrm{d}\mu} \right\|_{p/(p-1)}^{(\mu)} \right)^{-\frac{pk}{m}} n^{-\frac{pk}{m}}. \tag{D.7}$$

We thus have

$$\mathbb{E}[\rho(X, f(X))] = \int_0^\infty \mu_X\Big(\big\{x \in \mathcal{X} : \rho(x, f(x)) \geq \delta\big\}\Big)\,\mathrm{d}\delta \tag{D.8}$$

$$\geq \int_0^{\min\{h(n), \delta_0^k\}} \mu_X\Big(\big\{x \in \mathcal{X} : \rho(x, f(x)) \geq \delta\big\}\Big)\,\mathrm{d}\delta \tag{D.9}$$

$$\geq \int_0^{\min\{h(n), \delta_0^k\}} \left(1 - \left(\frac{\delta}{h(n)}\right)^{\frac{m}{pk}}\right)\mathrm{d}\delta \tag{D.10}$$

$$= \min\Big\{h(n), \delta_0^k\Big\} \left(1 - \frac{pk}{(m+pk)} \min\left\{1, \frac{\delta_0^k}{h(n)}\right\}^{\frac{m}{pk}}\right) \tag{D.11}$$

$$\geq \min\left\{h(n), \delta_0^k\right\}\left(1 - \frac{pk}{m+pk}\right) \tag{D.12}$$

$$= \min\left\{h(n), \delta_0^k\right\}\frac{m}{m+pk}, \tag{D.13}$$

where (D.8) follows from Lemma K.1 and (D.10) is by (D.1)–(D.6) upon noting that $\delta < \delta_0^k$ on the integration domain. Since $n \in \mathbb{N}$ and $f \in \mathcal{F}_n(\mathcal{X},\mathcal{Y})$ were assumed to be arbitrary, we can conclude that

$$V_n(X) = \inf_{f\in\mathcal{F}_n(\mathcal{X},\mathcal{Y})} \mathbb{E}[\rho(X, f(X))] \tag{D.14}$$

$$\geq \min\left\{h(n), \delta_0^k\right\}\frac{m}{m+pk}. \tag{D.15}$$

□

E. Proof of Lemma 4.1

If $\lim_{n\to\infty} V_n(X) > 0$, then the definition of $\underline{D}_k(X)$ in (1.13) implies $D_k(X) = \infty$, which establishes the claim. Next, consider the case $\lim_{n\to\infty} V_n(X) = 0$ and fix $n \in \mathbb{N}$ and an n-quantizer $f \in \mathcal{F}_n$ arbitrarily. Define $g\colon \mathcal{X} \to \mathcal{X}\times\mathcal{Y}$ according to $g(x) = (x, f(x))$ and consider the pushforward measures $\mu_f := f_*(\mu_X)$ and $\mu_g := g_*(\mu_X)$. We first show that $\mu_g \ll \mu_X \otimes \mu_f$ with Radon–Nikodým derivative

$$\frac{\mathrm{d}\mu_g}{\mathrm{d}(\mu_X\otimes\mu_f)}(x,y) = h(x,y) := \frac{\chi_{\{(x,y)\in\mathcal{X}\times\mathcal{Y}: f(x)=y\}}(x,y)}{\mathbb{P}[f(X)=y]}. \tag{E.1}$$

To this end, note that for every $\mathcal{A} \in \mathscr{X}\otimes\mathscr{Y}$, we have

$$\int_{\mathcal{A}} h(x,y)\,\mathrm{d}(\mu_X\otimes\mu_f)(x,y) = \int\left(\int_{\mathcal{A}_y} h(x,y)\,\mathrm{d}\mu_X(x)\right)\mathrm{d}\mu_f(y) \tag{E.2}$$

$$= \int\left(\int_{\mathcal{A}_{f(z)}} h(x,f(z))\,\mathrm{d}\mu_X(x)\right)\mathrm{d}\mu_X(z) \tag{E.3}$$

$$= \int \frac{\mathbb{P}[f(X)=f(z) \text{ and } X\in\mathcal{A}_{f(z)}]}{\mathbb{P}[f(X)=f(z)]}\mathrm{d}\mu_X(z) \tag{E.4}$$

$$= \sum_{a\in f(\mathcal{X})} \frac{\mathbb{P}[f(X)=a \text{ and } X\in\mathcal{A}_a]}{\mathbb{P}[f(X)=a]}\int_{f^{-1}(\{a\})}\mathrm{d}\mu_X(z) \tag{E.5}$$

$$= \sum_{a\in f(\mathcal{X})} \mathbb{P}[f(X)=a \text{ and } X\in\mathcal{A}_a] \tag{E.6}$$

$$= \sum_{a\in f(\mathcal{X})} \mathbb{P}[f(X)=a \text{ and } g(X)\in\mathcal{A}] \tag{E.7}$$

$$= \mathbb{P}[g(X)\in\mathcal{A}] \tag{E.8}$$

$$= \mu_g(\mathcal{A}), \tag{E.9}$$

where (E.2) follows from Tonelli's theorem [5, Theorem 10.9] and we set, for every $y\in\mathcal{Y}$, $\mathcal{A}_y = \{x \in \mathcal{X} : (x,y)\in\mathcal{A}\}$.

Next, note that

$$\mu_g(\mathcal{A}) = \int_{\mathcal{A}} h(x,y)\,\mathrm{d}(\mu_X \otimes \mu_f)(x,y) \tag{E.10}$$

$$= \int \mu_{f(X)|X}(\mathcal{A}_x|x)\mathrm{d}\mu_X(x) \quad \text{for all } \mathcal{A} \in \mathscr{X} \otimes \mathscr{Y}, \tag{E.11}$$

where (E.10) is by (E.2)–(E.9) and in (E.11) we again applied Tonelli's theorem with $\mathcal{A}_x = \{y \in \mathcal{Y} : (x,y) \in \mathcal{A}\}$ and

$$\mu_{f(X)|X}(\mathcal{B}|x) := \int_{\mathcal{B}} h(x,y)\,\mathrm{d}\mu_f(y) \quad \text{for all } x \in \mathcal{X} \text{ and } \mathcal{B} \in \mathscr{Y}. \tag{E.12}$$

Tonelli's theorem also guarantees that $\mu_{f(X)|X}(\mathcal{B}|\cdot)$ is measurable for all $\mathcal{B} \in \mathscr{Y}$. Moreover, $\mu_{f(X)|X}(\cdot|x)$ is a probability measure as

$$\mu_{f(X)|X}(\mathcal{Y}|x) = \int h(x,y)\,\mathrm{d}\mu_f(y) \tag{E.13}$$

$$= \int h(x,f(z))\,\mathrm{d}\mu_X(z) \tag{E.14}$$

$$= \sum_{a \in f(\mathcal{X})} h(x,a)\mathbb{P}[f(X)=a] \tag{E.15}$$

$$= 1 \quad \text{for all } x \in \mathcal{X}. \tag{E.16}$$

Finally, we have

$$I(X;f(X)) = \int \log(h(x,y))\mathrm{d}\mu_g(x,y) \tag{E.17}$$

$$= \int \log(h(g(x)))\mathrm{d}\mu_X(x) \tag{E.18}$$

$$\leq \log\left(\int h(g(x))\mathrm{d}\mu_X(x)\right) \tag{E.19}$$

$$= \log\left(\sum_{a \in f(\mathcal{X})} \int_{f^{-1}(a)} h(x,a)\mathrm{d}\mu_X(x)\right) \tag{E.20}$$

$$= \log\left(\sum_{a \in f(\mathcal{X})} 1\right) \tag{E.21}$$

$$\leq \log(n), \tag{E.22}$$

where (E.19) is by Jensen's inequality [51, Theorem 2.3]. We can now conclude that (see (1.5))

$$R_X(D) = \inf_{\mu_{Y|X} : \mathbb{E}[\rho(X,Y)] \leq D} I(X;Y) \tag{E.23}$$

$$\leq \log(n) \quad \text{if } \mathbb{E}[\rho(X,f(X))] \leq D. \tag{E.24}$$

Next, note that the definition of $V_n(X)$ in (1.11) implies that for every $\delta > 0$, there exists an $f_\delta \in \mathcal{F}_n$ such that $\mathbb{E}[\rho(X, f_\delta(X))] \le V_n(X) + \delta$. Combined with (E.23)–(E.24) this yields

$$R_X(V_n(X)+\delta) \le \log(n) \quad \text{for all } \delta > 0. \tag{E.25}$$

If there exists an $n_0 \in \mathbb{N}$ such that $V_{n_0}(X) = 0$, then (E.25) implies $R_X(D) \le \log(n_0)$ for all $D > 0$ so that $\dim_R(X) = 0$ and the statement to be established follows trivially. Now, suppose that $V_n(X) > 0$ for all $n \in \mathbb{N}$. Using continuity of $R_X(D)$ [12, Lemma 1.1] and the fact that n in (E.25) was assumed to be arbitrary, it follows that

$$R_X(V_n(X)) \le \log(n) \quad \text{for all } n \in \mathbb{N}. \tag{E.26}$$

We can therefore conclude that

$$\begin{aligned}
\underline{D}_k(X) &= \liminf_{n\to\infty} \frac{k\log(n)}{\log(1/V_n(X))} && \text{(E.27)}\\
&\ge \liminf_{n\to\infty} \frac{kR_X(V_n(X))}{\log(1/V_n(X))} && \text{(E.28)}\\
&\ge \liminf_{D\to 0} \frac{kR_X(D)}{\log(1/D)} && \text{(E.29)}\\
&= \liminf_{D\to 0} \frac{R_X(D^k)}{\log(1/D)} && \text{(E.30)}\\
&= \underline{\dim}_R(X), && \text{(E.31)}
\end{aligned}$$

where (E.28) follows from (E.26) and in (E.29) we used that for every function $f\colon (0,\infty) \to (0,\infty)$ and sequence $(x_n)_{n\in\mathbb{N}}$ in $(0,\infty)$ with $\lim_{n\to\infty} x_n = 0$, we have

$$\begin{aligned}
\liminf_{n\to\infty} f(x_n) &= \sup_{n\in\mathbb{N}} \inf_{k\ge n} f(x_k) \ge \sup_{n\in\mathbb{N}} \inf_{D\in(0,x_n]} f(D) && \text{(E.32)}\\
&\ge \sup_{a\in(0,\infty)} \inf_{D\in(0,a]} f(D) = \liminf_{D\to 0} f(D), && \text{(E.33)}
\end{aligned}$$

where the inequality in (E.33) follows from the observation that for every $a \in (0,\infty)$, thanks to $\lim_{n\to\infty} x_n = 0$, there exists an $n_0 \in \mathbb{N}$ such that $x_{n_0} < a$ so that $\inf_{D\in(0,x_{n_0}]} f(D) \ge \inf_{D\in(0,a]} f(D)$. □

F. Proof of Theorem 4.2

Fix $n \in \mathbb{N}$ arbitrarily. We first show that

$$V_n(X) \le \mathbb{E}\left[\min_{i=1,\dots,n} \tilde{\rho}(X, y_i)\right] \quad \text{for all } y_1,\dots,y_n \in \mathcal{Y}. \tag{F.1}$$

To this end, fix $y_1,\dots,y_n \in \mathcal{Y}$ arbitrarily and set

$$\mathcal{A}_i = \bigcap_{j=1}^{n} \{x \in \mathcal{X} : \rho(x,y_i) \le \rho(x,y_j)\} \quad \text{for } i = 1,\dots,n. \tag{F.2}$$

The sets $\mathcal{A}_i$ are measurable by virtue of being finite intersections of preimages of $[0,\beta^k]$ under measurable functions. By construction, for every $i=1,\ldots,n$, $\mathcal{A}_i$ consists of all points $x\in\mathcal{X}$ that have y_i as nearest point among $\{y_1,\ldots,y_n\}$. Next, set $\mathcal{B}_1=\mathcal{A}_1$ and

$$\mathcal{B}_j=\mathcal{A}_j\setminus\bigcup_{i=1}^{j-1}\mathcal{A}_j \quad \text{for } j=2,\ldots,n. \tag{F.3}$$

By construction, the sets $\mathcal{B}_1,\ldots,\mathcal{B}_n$ are pairwise disjoint measurable sets with $\mathcal{B}_i\subseteq\mathcal{A}_i$ for $i=1,\ldots,n$ and

$$\mathcal{X}=\bigcup_{i=1}^{n}\mathcal{B}_i. \tag{F.4}$$

Now, consider the n-quantizer $f\colon\mathcal{X}\to\mathcal{Y}$ defined according to $f(x)=y_i$ for all $x\in\mathcal{B}_i$ and $i=1,\ldots,n$. Then, we have

$$\begin{aligned}
V_n(X)&\le\mathbb{E}[\rho(X,f(X))] && \text{(F.5)}\\
&=\sum_{i=1}^{n}\int_{\mathcal{B}_i}\rho(x,f(x))\,\mathrm{d}\mu_X(x) && \text{(F.6)}\\
&=\sum_{i=1}^{n}\int_{\mathcal{B}_i}\rho(x,y_i)\,\mathrm{d}\mu_X(x) && \text{(F.7)}\\
&=\int\min_{i=1,\ldots,n}\rho(x,y_i)\,\mathrm{d}\mu_X(x) && \text{(F.8)}\\
&=\mathbb{E}\left[\min_{i=1,\ldots,n}\rho(X,y_i)\right], && \text{(F.9)}
\end{aligned}$$

where (F.8) follows from (F.2) combined with (F.3). Since $y_1,\ldots,y_n\in\mathcal{Y}$ were assumed to be arbitrary, (F.1) follows from (F.5)–(F.9).

Now, consider i.i.d. random variables $Y_1,\ldots,Y_n$ taking values in $\mathcal{Y}$, of distribution ν, and independent from X. It follows from (F.1) that

$$V_n(X)\le\mathbb{E}\left[\min_{i=1,\ldots,n}\rho(X,Y_i)\right]. \tag{F.10}$$

Fix $x\in\mathcal{X}$ arbitrarily and set $Z_n(x)=\min_{i=1,\ldots,n}\rho(x,Y_i)$. Then, we have

$$\begin{aligned}
\mathbb{P}[Z_n(x)\ge\delta]&=\mathbb{P}\big[\rho(x,Y_i)\ge\delta \quad \text{for } i=1,\ldots,n\big] && \text{(F.11)}\\
&=\mathbb{P}\big[\rho(x,Y_1)\ge\delta\big]^n && \text{(F.12)}\\
&=\Big(1-\mathbb{P}\big[\rho(x,Y_1)<\delta\big]\Big)^n && \text{(F.13)}\\
&\le e^{-n\mathbb{P}[\rho(x,Y_1)<\delta]} && \text{(F.14)}\\
&=e^{-n\nu\left(\widetilde{\mathcal{B}}_\rho(x,\delta)\right)} && \text{(F.15)}
\end{aligned}$$

$$\leq \begin{cases} e^{-bn\delta^{\frac{m}{k}}} & \text{if } \delta < \delta_0^k \\ e^{-bn\delta_0^m} & \text{if } \delta_0^k \leq \delta < \infty, \end{cases} \tag{F.16}$$

where in (F.14) we used that $1-t \leq e^{-t}$ for all $t \in [0,1]$, the case $\delta < \delta_0^k$ in (F.16) follows from the fact that ν, by Lemma 2.2, is ρ-superregular of dimension m/k with superregularity constants b and δ_0^k, and for $\delta_0^k \leq \delta < \infty$, (F.16) is by

$$\nu\left(\widetilde{\mathcal{B}}_\rho(x,\delta)\right) \geq \nu\left(\widetilde{\mathcal{B}}_\rho\left(x,\delta_0^k\right)\right) \geq b\delta_0^m, \tag{F.17}$$

which, in turn, follows from monotonicity and superregularity of ν combined with Item (ii) in Lemma 2.1. We can therefore upper-bound $\mathbb{E}[Z_n(x)]$ as follows:

$$\mathbb{E}[Z_n(x)] = \int_0^{\beta^k} \mathbb{P}\left[Z_n(x) \geq \delta\right] \mathrm{d}\delta \tag{F.18}$$

$$\leq \int_0^{\delta_0^k} e^{-bn\delta^{\frac{m}{k}}} \,\mathrm{d}\delta + e^{-bn\delta_0^m} \int_{\min\{\delta_0^k,\beta^k\}}^{\beta^k} \mathrm{d}\delta \tag{F.19}$$

$$= \frac{k}{m}(bn)^{-\frac{k}{m}} \int_0^{bn\delta_0^m} e^{-t} t^{\frac{k}{m}-1}\,\mathrm{d}t + e^{-bn\delta_0^m} \int_{\min\{\delta_0^k,\beta^k\}}^{\beta^k} \mathrm{d}\delta \tag{F.20}$$

$$= \frac{k}{m}\gamma\left(\frac{k}{m}, bn\delta_0^m\right)(bn)^{-\frac{k}{m}} + e^{-bn\delta_0^m} \int_{\min\{\delta_0^k,\beta^k\}}^{\beta^k} \mathrm{d}\delta \tag{F.21}$$

$$\leq U_n, \tag{F.22}$$

where (F.18) follows from Lemma K.1 combined with (4.11), (F.19) is by (F.11)–(F.16), and in (F.20) we changed variables according to $t = bn\delta^{m/k}$. The claim now follows from (F.10) and

$$\mathbb{E}\left[\min_{i=1,\dots,n} \rho(X,Y_i)\right] = \mathbb{E}\left[Z_n(X)\right] \tag{F.23}$$

$$= \int \mathbb{E}\left[Z_n(x)\right] \mathrm{d}\mu_X(x) \tag{F.24}$$

$$\leq \int U_n \,\mathrm{d}\mu_X(x) \tag{F.25}$$

$$= U_n, \tag{F.26}$$

where (F.25) is by (F.18)–(F.22). □

G. Proof of Theorem 4.3

We first show that

$$\nu\left(\mathcal{B}_\rho(x,\delta)\right) \geq \begin{cases} b\delta^{\frac{m}{k}} & \text{if } \delta < \delta_0^k \\ \tilde{b}\delta^{\frac{m}{k}} & \text{if } \delta_0^k \leq \delta \leq \beta^k < \infty \end{cases} \quad \text{for all } x \in \mathcal{X}, \tag{G.1}$$

where $\tilde{b} = b(\delta_0/\beta)^m$ for $\delta_0 < \infty$. The case $\delta < \delta_0^k$ follows from the fact that $\rho^{1/k}$-superregularity of ν, by Lemma 2.2 implies that ν is ρ-superregular of dimension m/k with superregularity constants b and δ_0^k. If $\delta_0^k \leq \delta \leq \beta^k < \infty$, then we have

$$\nu\big(\mathcal{B}_\rho(x,\delta)\big) \geq \nu\left(\mathcal{B}_\rho\left(x,\delta_0^k\right)\right) \tag{G.2}$$

$$\geq b\,\delta_0^m \tag{G.3}$$

$$\geq \tilde{b}\,\delta^{\frac{m}{k}} \quad \text{for all } x \in \mathcal{X}, \tag{G.4}$$

where in (G.2) we used that $\widetilde{\mathcal{B}}_\rho(x,\delta) = \mathcal{B}_\rho(x,\delta)$ as a consequence of $\mathcal{X} = \mathcal{Y}$ and symmetry of ρ, and (G.3) is by Item (ii) in Lemma 2.1. Now, let

$$g(x,\delta) = \frac{\mu_X\big(\mathcal{B}_\rho(x,\delta)\big)}{\nu\big(\mathcal{B}_\rho(x,\delta)\big)} \tag{G.5}$$

and note that there exists a $t \in (0,\infty)$ such that

$$f(x) := \frac{\mathrm{d}\nu}{\mathrm{d}\mu_X}(x) \leq 1/t \quad \text{for } \mu_X\text{-almost all } x \in \mathcal{X} \tag{G.6}$$

owing to (4.34). We therefore have

$$g(x,\delta) = \frac{\mu_X\big(\mathcal{B}_\rho(x,\delta)\big)}{\int_{\mathcal{B}_\rho(x,\delta)} f(x)\,\mathrm{d}\mu_X(x)} \tag{G.7}$$

$$\geq t \quad \text{for all } x \in \mathcal{X} \text{ and } \delta \in (0,\infty). \tag{G.8}$$

Next, note that

$$\lim_{\delta\to 0} \frac{1}{g(x,\delta)} = \lim_{\delta\to 0} \frac{1}{\mu_X\big(\mathcal{B}_\rho(x,\delta)\big)} \int_{\mathcal{B}_\rho(x,\delta)} f(x)\,\mathrm{d}\mu_X(x) \tag{G.9}$$

$$= f(x) \quad \text{for } \mu_X\text{-almost all } x \in \mathcal{X}, \tag{G.10}$$

where (G.10) follows from [50, Corollary 2.14, Item (2)] upon noting that μ_X is a Radon measure [25, Definition 6.6] thanks to [25, Theorem 6.1 and Proposition 6.7] and $f(x)$ is bounded by (G.6). The upper bound in (G.6) now implies

$$\Omega_{k/m}(X) = \mathbb{E}\left[f^{\frac{k}{m}}(X)\right] \;\leq\; t^{-\frac{k}{m}} \;<\; \infty. \tag{G.11}$$

Next, let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. copies of X and define, for every $n \in \mathbb{N}$ and $x \in \mathcal{X}$, the random variable $U_n(x) = n^{k/m}\min_{i=1,\dots,n}\rho(x,X_i)$. Then, we have

$$\mathbb{P}\big[U_n(x) \geq \delta\big] = \mathbb{P}\Big[\rho(x,X_i) \geq n^{-k/m}\delta \quad \text{for } i = 1,\dots,n\Big] \tag{G.12}$$

$$= \mathbb{P}\Big[\rho(x,X_1) \geq n^{-k/m}\delta\Big]^n \tag{G.13}$$

$$= \left(1 - \mathbb{P}\left[\rho(x, X_1) < n^{-k/m}\delta\right]\right)^n \tag{G.14}$$

$$\leq e^{-n\mathbb{P}\left[X_1 \in \mathcal{B}_\rho\left(x, n^{-\frac{k}{m}}\delta\right)\right]} \tag{G.15}$$

$$\leq \begin{cases} e^{-b\,\delta^{\frac{m}{k}} g\left(x, n^{-\frac{k}{m}}\delta\right)} & \text{if } \delta < \delta_0^k n^{\frac{k}{m}} \\ e^{-\tilde{b}\,\delta^{\frac{m}{k}} g\left(x, n^{-\frac{k}{m}}\delta\right)} & \text{if } \delta_0^k n^{\frac{k}{m}} \leq \delta \leq \beta^k n^{\frac{k}{m}} < \infty \end{cases} \tag{G.16}$$

for all $n \in \mathbb{N}$, where (G.16) follows from (G.1) and (G.5). Moreover, (4.35) implies

$$\mathbb{P}\left[U_n(x) \geq \delta\right] = 0 \quad \text{if } \delta > \beta^k n^{\frac{k}{m}}. \tag{G.17}$$

We can now upper-bound $\limsup_{n\to\infty} n^{\frac{k}{m}} V_n(X)$ as follows:

$$\limsup_{n\to\infty} n^{\frac{k}{m}} V_n(X) \leq \limsup_{n\to\infty} n^{\frac{k}{m}} \mathbb{E}\left[\min_{i=1,\dots,n} \rho(X, X_i)\right] \tag{G.18}$$

$$= \limsup_{n\to\infty} \mathbb{E}\left[U_n(X)\right] \tag{G.19}$$

$$= \limsup_{n\to\infty} \int \left(\int_0^\infty \mathbb{P}\left[U_n(x) \geq \delta\right] \mathrm{d}\delta\right) \mathrm{d}\mu_X(x) \tag{G.20}$$

$$\leq \int \left(\int_0^\infty \limsup_{n\to\infty} \mathbb{P}\left[U_n(x) \geq \delta\right] \mathrm{d}\delta\right) \mathrm{d}\mu_X(x), \tag{G.21}$$

where (G.18) is by (F.1) with $\mathcal{Y} = \mathcal{X}$, (G.20) follows from Lemma K.1, and in (G.21) we applied Fatou's lemma for lim sup [2, Theorem 1.20] twice noting that $\mathbb{P}[U_n(x) \geq \delta] \leq h(\delta)$ for all $x \in \mathcal{X}$ and $\delta \in (0, \infty)$ and

$$\left(\int_0^\infty h(\delta)\,\mathrm{d}\delta\right) \mathrm{d}\mu_X(x) = \int_0^\infty h(\delta)\,\mathrm{d}\delta < \infty \tag{G.22}$$

with

$$h(\delta) = \begin{cases} e^{-bt\,\delta^{\frac{m}{k}}} & \text{if } \delta \in (0, \delta_0^k n^{k/m}) \\ e^{-\tilde{b}t\,\delta^{\frac{m}{k}}} & \text{if } \delta_0^k n^{k/m} \leq \delta < \infty \end{cases} \tag{G.23}$$

thanks to (G.12)–(G.17) together with (G.7)–(G.8). Next, note that

$$\limsup_{n\to\infty} \mathbb{P}\left[U_n(x) \geq \delta\right] \leq \lim_{n\to\infty} e^{-b\,\delta^{\frac{m}{k}} g\left(x, n^{-\frac{k}{m}}\delta\right)} \tag{G.24}$$

$$= e^{-b\delta^{\frac{m}{k}}/f(x)} \quad \text{for all } x \in \mathcal{X} \text{ and } \delta \in (0, \infty), \tag{G.25}$$

where (G.24) is by (G.12)–(G.16) and in (G.25) we applied (G.9)–(G.10). Using (G.24)–(G.25) in (G.21) yields

$$\limsup_{n\to\infty} n^{\frac{k}{m}} V_n(X) \le \int \left(\int_0^\infty e^{-b\delta^{\frac{m}{k}}/f(x)} \, \mathrm{d}\delta \right) \mathrm{d}\mu_X(x) \tag{G.26}$$

$$= \mathbb{E}\left[(f(X)/b)^{\frac{k}{m}} \frac{k}{m} \int_0^\infty e^{-s} s^{\frac{k}{m}-1} \, \mathrm{d}s \right] \tag{G.27}$$

$$= \mathbb{E}\left[f^{\frac{k}{m}}(X) \right] \Gamma\left(1 + \frac{k}{m}\right) b^{-\frac{k}{m}}, \tag{G.28}$$

where in (G.27) we changed variables according to $s = b\delta^{m/k}/f(x)$. This establishes (4.36). Finally, we have

$$\Omega_\alpha(X) = \mathbb{E}\left[f^\alpha(X) \right] \tag{G.29}$$

$$\le \left(\mathbb{E}\left[f(X) \right] \right)^\alpha \tag{G.30}$$

$$= \left(\int f(x) \, \mathrm{d}\mu_X(x) \right)^\alpha \tag{G.31}$$

$$= (\nu(\mathcal{X}))^\alpha \tag{G.32}$$

$$= 1 \quad \text{for all } \alpha \in (0,1), \tag{G.33}$$

where (G.30) follows from Jensen's inequality [51, Theorem 2.3] and strict concavity of $t \to t^\alpha$ on $(0,\infty)$ for all $\alpha \in (0,1)$ with strict inequality in (G.30) unless $f(x) = \mathbb{E}\left[f(X)\right]$ for μ_X-almost all $x \in \mathcal{X}$ [14, Example 6]. The statement in (4.38) follows trivially from $\mathbb{E}[f(X)] = 1$ and the definition of the lower k-th quantization coefficient in (1.15). □

H. Proof of Lemma 4.2

We first prove Item (i). Toward a contradiction, assume that $\liminf_{n\to\infty} n^{\frac{k}{m}} V_n(X) > 0$ with $\underline{D}_k(X) < m$ and set $\delta = (m - \underline{D}_k(X))/2$. The definition of $\underline{D}_k(X)$ in (1.13) then implies that

$$\frac{k \log(n)}{\log(1/V_n(X))} \le m - \delta \quad \text{for } \infty\text{-many } n \in \mathbb{N}. \tag{H.1}$$

Straightforward algebraic manipulations show that (H.1) is equivalent to

$$n^{\frac{k}{m}} V_n(X) \le n^{\frac{k}{m} - \frac{k}{m-\delta}} \quad \text{for } \infty\text{-many } n \in \mathbb{N}, \tag{H.2}$$

which in turn implies the existence of an increasing sequence $(n_j)_{j\in\mathbb{N}}$ satisfying

$$\liminf_{j\to\infty} n_j^{\frac{k}{m}} V_{n_j}(X) \le \lim_{j\to\infty} n_j^{\frac{k}{m} - \frac{k}{m-\delta}} = 0 \tag{H.3}$$

and thereby contradicts the assumption $\liminf_{n\to\infty} n^{\frac{k}{m}} V_n(X) > 0$.

Next, we prove Item (ii). Toward a contradiction, assume that $\limsup_{n\to\infty} n^{\frac{k}{m}} V_n(X) < \infty$ with $\overline{D}_k(X) > m$ and set $\delta = (\overline{D}_k(X) - m)/2$. The definition of $\overline{D}_k(X)$ in (1.14) then implies that

$$\frac{k\log(n)}{\log(1/V_n(X))} \geq m + \delta \quad \text{for } \infty\text{-many } n \in \mathbb{N}. \tag{H.4}$$

Straightforward algebraic manipulations reveal that (H.4) is equivalent to

$$n^{\frac{k}{m}} V_n(X) \geq n^{\frac{k}{m} - \frac{k}{m+\delta}} \quad \text{for } \infty\text{-many } n \in \mathbb{N}, \tag{H.5}$$

which in turn implies the existence of an increasing sequence $(n_j)_{j\in\mathbb{N}}$ satisfying

$$\limsup_{j\to\infty} n_j^{\frac{k}{m}} V_{n_j}(X) \geq \lim_{j\to\infty} n_j^{\frac{k}{m} - \frac{k}{m+\delta}} = \infty \tag{H.6}$$

and thereby contradicts the assumption $\limsup_{n\to\infty} n^{\frac{k}{m}} V_n(X) < \infty$.

I. Proof of Theorem 5.1

Lemma 2.2 combined with Lemma 2.4 implies that every μ_{X_i} is σ-subregular of dimension $m/(pk)$ with subregularity constants $\tilde{c} = \|\mathrm{d}\mu_{X_1}/\mathrm{d}\mu\|^{(\mu)}_{p/(p-1)} c^{1/p}$ and δ_0^k. Fix $\ell \in \mathbb{N}$ arbitrarily and set $X^{(\ell)} = (X_1,\dots,X_\ell)$. It follows from Item (i) of Proposition 2.1 applied to the measures $\mu_{X_1},\dots,\mu_{X_\ell}$ with $\alpha_i = 1/\ell$ for $i = 1,\dots,\ell$ that $\mu_{X^{(\ell)}} := \mu_{X_1} \otimes \cdots \otimes \mu_{X_\ell}$ is $\sigma_{(\ell)}$-subregular of dimension $\ell m/(pk)$ with subregularity constants $d_{(\ell)}^{\ell}$ and $\delta_{(\ell)} = \delta_0^k/\ell$. Theorem 4.1 applied to $X^{(\ell)}$ (note that for $\mu = \mu_{X^{(\ell)}}$ we have $\Sigma_1(X^{(\ell)}) = 1$) therefore yields

$$V_n(X^{(\ell)}) \geq L_n\left(X^{(\ell)}\right) \quad \text{for all } n \in \mathbb{N} \tag{I.1}$$

with

$$L_n\left(X^{(\ell)}\right) = \min\left\{ d_{(\ell)}^{-\frac{pk}{m}} n^{-\frac{pk}{\ell m}}, \frac{\delta_0^k}{\ell} \right\} \frac{\ell m}{\ell m + pk}. \tag{I.2}$$

Now, set $L_n = L_n\left(X^{(\ell)}\right)$ and

$$K_{(\ell)} = \left\lceil \left(\frac{\ell}{\delta_0^k}\right)^{\frac{\ell m}{pk}} d_{(\ell)}^{-\ell} \right\rceil \tag{I.3}$$

and recall the definition of $D_{(\ell)}$ in (5.7) according to

$$D_{(\ell)} = \frac{\delta_0^k}{\ell} \frac{\ell m}{\ell m + pk}. \tag{I.4}$$

Evaluation of the minimum in (I.2) yields

$$L_n = d_{(\ell)}^{-\frac{pk}{m}} n^{-\frac{pk}{\ell m}} \frac{\ell m}{\ell m + pk} \quad \text{for all } n \geq \lceil K_{(\ell)} \rceil \tag{I.5}$$

and $L_n = D_{(\ell)}$ for all $n \leq \lfloor K_{(\ell)} \rfloor$. Next, assume that there exist $n \in \mathbb{N}$ and $R \in (0,\infty)$ such that $V_{\lfloor e^{\ell R} \rfloor}(X^{(\ell)}) < L_n$. Then, (I.1) implies that $L_{\lfloor e^{\ell R} \rfloor} < L_n$, which, as a consequence of L_n being monotonically decreasing in n, is possible for $e^{\ell R} \geq n+1$ only. We conclude that, for every $D \in [L_{n+1}, L_n)$ and $n \geq \lceil K_{(\ell)} \rceil - 1$, it holds that

$$\begin{aligned} R_{(\ell)}(D) &= \inf\{R > 0 : V_{\lfloor e^{\ell R} \rfloor}(X^{(\ell)}) \leq D\} && \text{(I.6)} \\ &\geq \inf\{R > 0 : V_{\lfloor e^{\ell R} \rfloor}(X^{(\ell)}) < L_n\} && \text{(I.7)} \\ &\geq \frac{1}{\ell}\log(n+1) && \text{(I.8)} \\ &= \frac{m}{pk}\log\left(\frac{\ell m}{(\ell m + pk)L_{n+1}}\right) - \log(d_{(\ell)}) && \text{(I.9)} \\ &\geq \frac{m}{pk}\log\left(\frac{\ell m}{(\ell m + pk)D}\right) - \log(d_{(\ell)}), && \text{(I.10)} \end{aligned}$$

where (I.9) is by (I.5). Finally, as $L_{\lfloor K_{(\ell)} \rfloor} = D_{(\ell)}$ and $\lim_{n\to\infty} L_n = 0$, we can conclude that for every $D < D_{(\ell)}$, there must exist an $n \geq \lceil K_{(\ell)} \rceil - 1$ such that $D \in [L_{n+1}, L_n)$. Hence, (I.6)–(I.10) establishes $R_{(\ell)}(D) \geq \widetilde{R}_{(\ell)}(D)$ for all $D \in (0, D_{(\ell)})$.

Next, note that $\widetilde{R}_{(\ell)}(D)$ can be written as

$$\begin{aligned} \widetilde{R}_{(\ell)}(D) &= -\log\left(\|\mathrm{d}\mu_{X_1}/\mathrm{d}\mu\|^{(\mu)}_{p/(p-1)}\right) - \frac{m}{pk}\log(D) - \log\left(c^{\frac{1}{p}}\,\Gamma\left(1+\frac{m}{pk}\right)\right) && \text{(I.11)} \\ &\quad + \frac{m}{pk}\left(\frac{pk}{\ell m}\log\left(\Gamma\left(1+\frac{\ell m}{pk}\right)\right) - \log\left(\ell + \frac{pk}{m}\right)\right) && \text{(I.12)} \\ &= -\log\left(\|\mathrm{d}\mu_{X_1}/\mathrm{d}\mu\|^{(\mu)}_{p/(p-1)}\right) - \frac{m}{pk}\log(D) - \log\left(c^{\frac{1}{p}}\,\Gamma\left(1+\frac{m}{pk}\right)\right) && \text{(I.13)} \\ &\quad + \frac{m}{pk}\log\left(\frac{m}{pk}\right) + \frac{m}{pk}\psi\left(\frac{\ell m}{pk}\right), && \text{(I.14)} \end{aligned}$$

where we set

$$\psi(x) = \frac{\log(\Gamma(1+x))}{x} - \log(1+x). \tag{I.15}$$

The properties of $\widetilde{R}_{(\ell)}(D)$ we wish to establish now follow from the fact that $\psi(x)$ is strictly monotonically decreasing on $(-1,\infty)$ with $\lim_{x\to\infty}\psi(x) = -1$ [62, Theorem 1]. □

J. Proof of Lemma 6.1

We first show that $G(\alpha)$ is continuous and strictly monotonically increasing on $(0,1]$. To this end, we write $G(\alpha)$ in the form

$$G(\alpha) = \frac{a^{(d-1)}(1)}{2B_{\frac{d-1}{2},\frac{1}{2}}(1)} g(\alpha) \tag{J.1}$$

with

$$g(\alpha) = \frac{B_{\frac{d-1}{2},\frac{1}{2}}(\alpha)}{\alpha^{\frac{d-1}{2}}}. \tag{J.2}$$

Since

$$\frac{\mathrm{d}g(\alpha)}{\mathrm{d}\alpha} = \frac{1}{\alpha}\left(\frac{1}{\sqrt{1-\alpha}} - \frac{d-1}{2}\frac{B_{\frac{d-1}{2},\frac{1}{2}}(\alpha)}{\alpha^{\frac{d-1}{2}}}\right) \tag{J.3}$$

$$= \frac{1}{\alpha}\left(\frac{1}{\sqrt{1-\alpha}} - \frac{d-1}{2}g(\alpha)\right) \quad \text{for all } \alpha \in (0,1), \tag{J.4}$$

which is strictly positive on $(0,1)$ owing to

$$g(\alpha) < \frac{1}{\alpha^{\frac{d-1}{2}}\sqrt{1-\alpha}}\int_0^{\alpha} u^{\frac{d-1}{2}-1}\,\mathrm{d}u \tag{J.5}$$

$$= \frac{2}{(d-1)\sqrt{1-\alpha}} \quad \text{for all } \alpha \in (0,1), \tag{J.6}$$

it follows that $G(\alpha)$ is continuous and strictly monotonically increasing on $(0,1)$. Since

$$\left|B_{\frac{d-1}{2},\frac{1}{2}}(1) - B_{\frac{d-1}{2},\frac{1}{2}}(1-\delta)\right| = \int_{1-\delta}^{1} u^{\frac{d-1}{2}-1}(1-u)^{-1/2}\,\mathrm{d}u \tag{J.7}$$

$$\leq (1-\delta)^{-1/2}\int_{1-\delta}^{1}(1-u)^{-1/2}\,\mathrm{d}u \tag{J.8}$$

$$= 2\sqrt{\frac{\delta}{1-\delta}} \quad \text{for all } \delta \in (0,1), \tag{J.9}$$

the function $G(\alpha)$ is continuous at $\alpha = 1$ as well so that monotonicity on $(0,1)$ implies monotonicity on $(0,1]$. The limit in (6.3) follows by application of L'Hôpital's rule together with $B_{(d-1)/2,1/2}(1) = \Gamma((d-1)/2)\Gamma(1/2)/\Gamma(d/2)$ [3, Theorem 1.1.4].

We next establish the family of subregularity conditions in Item (i). To this end, fix $\delta_0 \in (0,r]$ arbitrarily. In light of Item (i) in Lemma 2.1, it is sufficient to show that

$$\mu\left(\mathcal{B}_{\rho^{1/2}}(y,\delta)\right) \leq c_{\delta_0}\delta^{d-1} \quad \text{for all } y \in \mathcal{Y} \text{ and } \delta \in (0,\delta_0) \tag{J.10}$$

with c_{δ_0} as specified in the statement. Now, fix $y \in \mathcal{Y}$ and $\delta \in (0, \delta_0)$ arbitrarily and note that

$$\|z-y\|_2^2 = \|y\|_2^2 + r^2 - 2y^{\mathsf{T}}z \quad \text{for all } z \in \mathcal{S}^{d-1}(r) \tag{J.11}$$

yields

$$\mathcal{B}_{\rho^{1/2}}(y,\delta) = \left\{ z \in \mathcal{S}^{d-1}(r) : y^{\mathsf{T}}z > \frac{\|y\|_2^2 + r^2 - \delta^2}{2} \right\}. \tag{J.12}$$

Next, consider the one-dimensional subspace $\mathcal{A}(y) = \{ay : a \in \mathbb{R}\}$. We have

$$\begin{aligned} \operatorname{dist}(z, \mathcal{A}(y)) &:= \min\{\|z-ay\|_2 : a \in \mathbb{R}\} && \text{(J.13)} \\ &= \sqrt{\|z\|_2^2 - \frac{(y^{\mathsf{T}}z)^2}{\|y\|_2^2}} && \text{(J.14)} \\ &\le \sqrt{r^2 - \frac{(r^2-\delta^2+\|y\|_2^2)^2}{4\|y\|_2^2}} \quad \text{for all } z \in \mathcal{B}_{\rho^{1/2}}(y,\delta), && \text{(J.15)} \end{aligned}$$

where (J.14) follows from the Hilbert projection theorem [56, Theorem 4.11] on $\mathbb{R}^d$ and in (J.15) we used (J.12) together with $\delta < \delta_0 \le r$. We can now write (J.13)–(J.15) in the form

$$\operatorname{dist}(z, \mathcal{A}(y)) \le \sqrt{r^2 - f^2(\|y\|_2)} \quad \text{for all } z \in \mathcal{B}_{\rho^{1/2}}(y,\delta) \tag{J.16}$$

with

$$f(s) = \frac{r^2 - \delta^2 + s^2}{2s}. \tag{J.17}$$

As $\delta < \delta_0 \le r$, the function $f(s)$ is strictly convex on $(0,\infty)$ and, therefore, minimized for $s_0 = \sqrt{r^2-\delta^2}$. We can thus conclude that

$$\begin{aligned} \operatorname{dist}(z, \mathcal{A}(y)) &\le \sqrt{r^2 - f^2\left(\sqrt{r^2-\delta^2}\right)} && \text{(J.18)} \\ &= \delta \quad \text{for all } z \in \mathcal{B}_{\rho^{1/2}}(y,\delta). && \text{(J.19)} \end{aligned}$$

Now, (J.18)–(J.19) implies that the radius of the base of the hyperspherical cap $\mathcal{B}_{\rho^{1/2}}(y,\delta)$ is no larger than δ. We conclude that

$$\begin{aligned} \mu\left(\mathcal{B}_{\rho^{1/2}}(y,\delta)\right) &\le \frac{1}{2} I_{\frac{d-1}{2},\frac{1}{2}}\left(\frac{\delta^2}{r^2}\right) && \text{(J.20)} \\ &= \frac{G(\delta^2/r^2)}{a^{(d-1)}(r)}\delta^{d-1} && \text{(J.21)} \\ &\le \frac{G(\delta_0^2/r^2)}{a^{(d-1)}(r)}\delta^{d-1}, && \text{(J.22)} \end{aligned}$$

where (J.20) is by the hyperspherical cap formula [43, Equation (1)] with δ for the radius of the base of the cap and in (J.22) we made use of the monotonicity of $G(\alpha)$. This yields the upper bound in (J.10) as $y \in \mathcal{Y}$ and $\delta \in (0, \delta_0)$ were assumed to be arbitrary.

Next, we establish the family of superregularity conditions in Item (ii). To this end, fix $\delta_0 \in (0, \sqrt{2}r]$ arbitrarily. Again, in light of Item (ii) in Lemma 2.1, it suffices to show that

$$\mu\left(\widetilde{\mathcal{B}}_{\rho^{1/2}}(x,\delta)\right) \geq b_{\delta_0}\delta^{d-1} \quad \text{for all } x \in \mathcal{S}^{d-1}(r) \text{ and } \delta \in (0,\delta_0) \tag{J.23}$$

with b_{δ_0} as specified in the statement. To this end, fix $x \in \mathcal{S}^{d-1}(r)$ and $\delta \in (0, \delta_0)$ arbitrarily. As before, consider the one-dimensional subspace $\mathcal{A}(x) = \{ax : a \in \mathbb{R}\}$. We have

$$\operatorname{dist}(z, \mathcal{A}(x)) := \min\{\|z - ax\|_2 : a \in \mathbb{R}\} \tag{J.24}$$

$$= \sqrt{\|z\|_2^2 - \frac{(x^\mathsf{T} z)^2}{\|x\|_2^2}} \tag{J.25}$$

$$= \sqrt{r^2 - \frac{(x^\mathsf{T} z)^2}{r^2}} \quad \text{for all } z \in \mathcal{S}^{d-1}(r), \tag{J.26}$$

where again (J.25) follows from the Hilbert projection theorem on $\mathbb{R}^d$. Now, (J.24)–(J.26) is equivalent to

$$\left|x^\mathsf{T} z\right| = r\sqrt{r^2 - \operatorname{dist}^2(z, \mathcal{A}(x))} \quad \text{for all } z \in \mathcal{S}^{d-1}(r). \tag{J.27}$$

Using (J.27) and $\delta < \delta_0 \leq \sqrt{2}r$ in (J.12), we obtain

$$\mathcal{B}_{\rho^{1/2}}(x,\delta) = \left\{ z \in \mathcal{S}^{d-1}(r) : \operatorname{dist}(z, \mathcal{A}(x)) < \delta\sqrt{1 - \left(\frac{\delta}{2r}\right)^2} \text{ and } x^\mathsf{T} z > 0 \right\}. \tag{J.28}$$

Now, (J.28) implies that the radius of the base of the hyperspherical cap $\mathcal{B}_{\rho^{1/2}}(x,\delta)$ equals $\delta\sqrt{1-(\delta/(2r))^2}$. We conclude that

$$\mu\left(\widetilde{\mathcal{B}}_{\rho^{1/2}}(x,\delta)\right) = \mu\left(\mathcal{B}_{\rho^{1/2}}(x,\delta)\right) \tag{J.29}$$

$$= \frac{1}{2} I_{\frac{d-1}{2},\frac{1}{2}}\left(h\left(\frac{\delta^2}{r^2}\right)\right) \tag{J.30}$$

$$= \frac{\Gamma\left(\frac{d}{2}\right)}{2\sqrt{\pi}\,\Gamma\left(\frac{d-1}{2}\right)} B_{\frac{d-1}{2},\frac{1}{2}}\left(h\left(\frac{\delta^2}{r^2}\right)\right) \tag{J.31}$$

$$\geq \frac{\Gamma\left(\frac{d}{2}\right)}{2\sqrt{\pi}\,\Gamma\left(\frac{d+1}{2}\right)} \left(h\left(\frac{\delta^2}{r^2}\right)\right)^{\frac{d-1}{2}} \tag{J.32}$$

$$\geq \frac{\Gamma\left(\frac{d}{2}\right)}{2\sqrt{\pi}\,\Gamma\left(\frac{d+1}{2}\right)r^{d-1}}\left(1-\frac{\delta_0^2}{4r^2}\right)^{\frac{d-1}{2}}\delta^{d-1} \tag{J.33}$$

$$= \frac{\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)a^{(d-1)}(r)}\left(1-\frac{\delta_0^2}{4r^2}\right)^{\frac{d-1}{2}}\delta^{d-1} \tag{J.34}$$

where (J.29) follows from

$$\widetilde{\mathcal{B}}_{\rho^{1/2}}(x,\delta)\cap\mathcal{S}^{d-1}(r)=\{z\in\mathcal{S}^{d-1}(r):\|x-z\|_2<\delta\}=\mathcal{B}_{\rho^{1/2}}(x,\delta), \tag{J.35}$$

(J.30) is by the hyperspherical cap formula [43, Equation (1)] with $\delta\sqrt{1-(\delta/(2r))^2}$ for the radius of the base of the cap and we set $h(t)=t(1-t/4)$, and in (J.32) we used

$$B_{\frac{d-1}{2},\frac{1}{2}}(\alpha)\geq\int_0^\alpha u^{\frac{d-1}{2}-1}\,\mathrm{d}u=\frac{2}{d-1}\alpha^{\frac{d-1}{2}}\quad\text{for all }\alpha\in(0,1]. \tag{J.36}$$

The argument is concluded by noting that $x\in\mathcal{S}^{d-1}(r)$ and $\delta\in(0,\delta_0)$ were assumed to be arbitrary.

Finally, the limit for c_{δ_0} follows from (6.3) and the limit for b_{δ_0} is trivial. □

K. A Layer Cake Argument

An important tool in lower/upper-bounding integrals using sub/super-regularity of the underlying integration measure is based on the so-called layer cake argument, which is an immediate consequence of Tonelli's theorem for characteristic functions [5, Theorem 10.9]. The formal statement is as follows.

Lemma K.1 *[16, Exercise 1.7.2] Consider a σ-finite measure space $(\mathcal{X},\mathscr{X},\mu)$. If $f\colon\mathcal{X}\to\mathbb{R}$ is measurable and $f(\mathcal{X})\subseteq[0,a]$ with $a\in(0,\infty]$, then*

$$\int f(x)\,\mathrm{d}\mu(x)=\int_0^a\mu\left(\{x:f(x)\geq t\}\right)\mathrm{d}t \tag{K.1}$$

$$=\int_0^a\mu\left(\{x:f(x)>t\}\right)\mathrm{d}t. \tag{K.2}$$

Proof To establish (K.1), we note that

$$\int_0^a\mu\left(\{x:f(x)\geq t\}\right)\mathrm{d}t=\int_0^\infty\mu\left(\{x:f(x)\geq t\}\right)\mathrm{d}t \tag{K.3}$$

$$=\int\left(\int_0^{f(x)}\mathrm{d}t\right)\mathrm{d}\mu(x) \tag{K.4}$$

$$=\int f(x)\,\mathrm{d}\mu(x), \tag{K.5}$$

where (K.3) holds because $f(\mathcal{X})\subseteq[0,a]$ and (K.4) follows from Tonelli's theorem for characteristic functions [5, Theorem 10.9]. Finally, (K.2) is obtained by repeating the steps in (K.3)–(K.5) with $f(x)\geq t$ replaced by $f(x)>t$. □

Acknowledgments

The authors thank the anonymous reviewers for their valuable suggestions.

Data Availability Statement

No new data were generated or analysed in support of this review.

Statements and Declarations

- G. Koliander was supported in part by the WWTF project MA16-053.

- A weaker version of Theorem 3.1 was presented without proof at the 2018 IEEE International Symposium on Information Theory (ISIT) [54].

REFERENCES

1. G. Alberti, H. Bölcskei, C. De Lellis, G. Koliander, and E. Riegler. Lossless analog compression. *IEEE Trans. Inform. Theory*, 65(11):7480–7513, Nov. 2018.
2. L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Univ. Press, New York, NY, 2000.
3. G. E. Andrews, R. Askey, and R. Roy. *Special Functions*. Cambridge University Press, Cambridge, NY, 1999.
4. R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. *Found. Comput. Math.*, 9(1):51–77, Feb 2009.
5. R. G. Bartle. *The Elements of Integration and Lebesgue Measure*. Wiley, New York, NY, 1995.
6. T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice-Hall, Englewood Cliffs, NJ, 1971.
7. N. Boumal and P.A. Absil. Low-rank matrix completion via preconditioned optimization on the Grassmann manifold. *Linear Algebra and its Applications*, 475:200–239, 2015.
8. E. J. Candès and Y. Plan. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Trans. Inform. Theory*, 4(57):2342–2359, Apr. 2011.
9. E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009.
10. Y. Chikuse. Statistics on Special Manifolds. In *Lecture Notes in Statistics*. Springer, New York, NY, 2003.
11. R.R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.
12. I. Csiszár. On an extremum problem of information theory. *Studia Sci. Math. Hungar.*, 9:57–71, 1974.
13. W. Dai, Y. Liu, and B. Rider. Quantization bounds on Grassmann manifolds and applications to MIMO communications. *IEEE Trans. Inform. Theory*, 54(3):1108–1123, Mar. 2008.
14. G. De Barra. *Measure Theory and Integration*. E. Horwood, Chichester, UK, 1981.
15. A. Dembo and I. Kontoyiannis. Critical behavior in lossy source coding. *IEEE Trans. Inform. Theory*, 47(3):1230–1236, Mar. 2001.
16. R. Durrett. *Probability Theory and Examples*. Cambridge University Press, Cambridge, NY, 4th edition, 2010.
17. G. Edgar. *Measure, Topology, and Fractal Geometry*. Springer, New York, NY, 2nd edition, 2008.
18. S. Eghbali and T. Ladan. Deep spherical quantization for image search. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11690–11699, 2019.
19. K. Falconer. *Fractal Geometry*. Wiley, New York, NY, 3rd edition, 2014.
20. H. Federer. *Geometric Measure Theory*. Springer, New York, NY, 1969.
21. C. Fefferman, S. Mitter, and H. Narayanan. Testing the manifold hypothesis. *J. Amer. Math. Soc.*, 29:983–1049, 2016.

22. J. M. Fraser, A. M. Henderson, E. J. Olson, and J. C. Robinson. On the Assouad dimension of self-similar sets with overlaps. *Adv. Math.*, 273:188–214, 2015.
23. S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Springer, New York, NY, 2nd edition, 1993.
24. A. Gersho and R. M. Gray. *Vector Quantization and Signal Processing*. Springer, New York, NY, 1992.
25. M. Giaquinta and G. Modica. *Mathematical Analysis–Foundations and Advanced Techniques for Functions of Several Variables*. Birkhäuser, New York, NY, 2012.
26. S. Graf and H. Luschgy. The quantization of the Cantor distribution. *Math. Nachr.*, 183:113–133, 1997.
27. S. Graf and H. Luschgy. *Foundations of Quantization for Probability Distributions*. Springer, Berlin, Germany, 2000.
28. R. M. Gray. *Source Coding Theory*. Kluwer, Boston, MA, 1990.
29. R. M. Gray and D. L. Neuhoff. Quantization. *IEEE Trans. Inform. Theory*, 44(6):2325–2383, Oct. 1998.
30. O. Henkel. Sphere-packing bounds in the Grassmann and Stiefel manifolds. *IEEE Trans. Inform. Theory*, 51(10):3445–3455, Oct. 2005.
31. G. E. Hinton, P. Dayan, and M. Revow. Modeling the manifolds of images of handwritten digits. *IEEE Trans. Neural Netw.*, 8(1):65–74, Jan 1997.
32. Z. Huang, J. Wu, and L. Van Gool. Building deep networks on Grassmann manifolds. In *The Thirty-Second AAAI Conference on Artificial Intelligence*, pages 3279–32, 2018.
33. G. J. O. Jameson. The incomplete gamma functions. *Math. Gazette*, 100(548):298–306, Jun. 2016.
34. A.D. Joseph, B. Nelson, B.I.P. Rubinstein, and J.D. Tygar. *Adversarial Machine Learning*. Cambridge University Press, Cambridge, NY, 2019.
35. V. Kadets. *A Course in Functional Analysis and Measure Theory*. Springer, Cham, Switzerland, 2018.
36. T. Kawabata and A. Dembo. The rate-distortion dimension of sets and measures. *IEEE Trans. Inform. Theory*, 40(5):1564–1572, Sep. 1994.
37. B. R. Kloeckner. Approximation of finitely supported measures. *ESAIM Control Optim. Calc. Var.*, 18(2):343–359, 2012.
38. A. W. Knapp. *Basic Real Analysis*. Birkhäuser, New York, NY, 2005.
39. T. Koch. The Shannon lower bound is asymptotically tight. *IEEE Trans. Inform. Theory*, 62(11):6155–6161, Nov. 2016.
40. G. Koliander, G. Pichler, E. Riegler, and F. Hlawatsch. Entropy and source coding for integer-dimensional singular random variables. *IEEE Trans. Inform. Theory*, 62(11):6124–6154, 2016.
41. M. Kossaczká and J. Vybíral. Entropy numbers of finite-dimensional embeddings. *Expositiones Mathematicae*, 48(3):319–336, 2020.
42. W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Trans. Netw.*, 2(1):1–15, Feb 1994.
43. S. Li. Concise formulas for the area and volume of a hyperspherical cap. *Asian J. Math. Stat.*, 4(1):66–70, 2011.
44. T. Linder and R. Zamir. On the asymptotic tightness of the Shannon lower bound. *IEEE Trans. Inform. Theory*, 40(6):2026–2031, Nov. 1994.
45. H. Lu, Y. Fainman, and R. Hecht-Nielsen. Image manifolds. In *Proc. SPIE 3307, Applications of Artificial Neural Networks in Image Processing III*, pages 3307–3307, San Jose, CA, Apr 1998.
46. A. J. Lzenman. Introduction to manifold learning. *WIREs Comput. Stat.*, 4:439–446, 2012.
47. K. V. Mardia. Statistics of directional data. *J. Royal Stat. Soc.*, 37(3):349–393, 1975.
48. K. V. Mardia and P. E. Jupp. *Directional Statistics*. Wiley, New York, NY, 2000.
49. B. Matschal and J. B. Huber. Spherical logarithmic quantization. *IEEE Trans. Audio, Speech, Language Process.*, 18(1):3445–3455, Jan. 2010.
50. P. Mattila. *Geometry of Sets and Measures in Euclidean Space: Fractals and Rectifiability*. Cambridge Univ. Press, Cambridge, UK, 1999.
51. J. E. Pečarić, F. Proschan, and Y. L. Tong. *Convex Functions, Partial Ordering, and Statistical Applications*. Academic Press Ind., London, UK, 1992.

52. R.A. Pitaval, L. Wei, O. Tirkkonen, and C. Hollanti. Density of spherical embedded Stiefel and Grassmann codes. *IEEE Trans. Inform. Theory*, 64(1):225–248, Jan. 2018.
53. K. Pötzelberger. The quantization error of self-similar distributions. *Math. Proc. Camb. Phil. Soc.*, 137(3):725–740, 2004.
54. E. Riegler, G. Koliander, and H. Bölcskei. Rate-distortion theory for general sets and measures. In *Proc. IEEE Int. Symp. on Inf. Theory*, pages 101–105, Jun. 2018.
55. E. Riegler, D. Stotz, and H. Bölcskei. Information-theoretic limits of matrix completion. In *Proc. IEEE Int. Symp. on Inf. Theory*, pages 1836–1840, Jun. 2015.
56. W. Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, NY, 3rd edition, 1987.
57. C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec. Pt. 4*, 7:142–163, 1959.
58. N. Sochen and Y. Y. Zeevi. Representation of colored images by manifolds embedded in higher dimensional non-Euclidean space. In *Proc. IEEE Int. Conf. on Image Process.*, pages 166–170, Chicago, IL, Oct 1998.
59. P. Swaszek and J. Thomas. Multidimensional spherical coordinates quantization. *IEEE Trans. Inform. Theory*, 29(4):570–576, Jul. 1983.
60. P. Turaga, A. Veeraraghavan, and R. Chellappa. Statistical analysis on Stiefel and Grassmann manifolds with applications in computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.
61. C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. AMS, Providence, RI, 2nd edition, 2016.
62. H. Vogt and J. Voigt. A monotonicity property of the γ-function. *JIPAM*, 3(4):1–3, 2002.
63. Z. Wu and J. Yu. Vector quantization: A review. *Front. Inform. Technol. Electron. Eng.*, 20(4):507–524, 2019.
64. Y. Yamada, S. Tazaki, and R. M. Gray. Asymptotic performance of block quantizers with difference distortion measure. *IEEE Trans. Inform. Theory*, 26(1):6–14, Jan. 1980.
65. M. P. W. Zerner. Weak separation properties for self-similar sets. *Proc. Am. Math. Soc.*, 124(11):3529–3539, 1996.
66. L. Zheng and D.N.C. Tse. Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel. *IEEE Trans. Inform. Theory*, 48(2):359–383, Feb. 2002.