High resolution coding of stochastic processes and small ball probabilities

vorgelegt von
Dipl.-Math. Steffen Dereich
aus Bad Kreuznach

Von der Fakultät II – Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
– Dr. rer. nat. –

genehmigte Dissertation

Promotionsausschuss

Vorsitzender:	Prof. Dr. Andreas Unterreiter
Berichter:	Prof. Dr. Jürgen Gärtner
	Prof. Dr. Michael Scheutzow

Tag der wissenschaftlichen Aussprache: 2. September 2003

Berlin 2003
D 83

Abstract

In this thesis, we study the high resolution coding problem for stochastic processes. The general coding problem concerns the finding of a good representation of a random signal, the *original*, within a class of allowed representations. The class of allowed representations is defined through restrictions on the information content of these representations. Mostly we consider three different interpretations of information. In the *quantization problem* we constrain the reconstruction corresponding to a representation, which itself is a random element, to be supported by a finite set having less than e^r elements. The *entropy coding problem* restricts the entropy of the reconstruction to be less than r. Finally, the *algorithmic coding problem* constrains the Shannon mutual information between the original and the reconstruction to be less than r.

This thesis is concerned with the asymptotic quality of the optimal coding scheme when the bound on the allowed information r tends to infinity: the *high resolution coding problem*. Our analysis considers Gaussian processes for the original and norm-based distortion measures. This means the distortion between the reconstruction and the original is measured as a power of the distance. A typical example is Wiener measure on the Banach space of continuous functions with corresponding norm. We derive asymptotic bounds for the high resolution coding problems. Our bounds are weakly and strongly tight for a broad class of originals in Banach and Hilbert spaces, respectively. Moreover, in the typical Hilbert space setting we show that the above three coding problems yield the same asymptotics.

A further result concerns the efficiency of quantization with randomly generated codebooks instead of deterministic codebooks. It is found that under certain regularity conditions, which are fulfilled in the typical Hilbert space setting, the corresponding asymptotics are the same.

A further objective is the effect of perturbations on the coding problem. These results yield a relation between the coding complexities of diffusions and Brownian motion in the entropy and algorithmic sense.

A second subject of this thesis is the study of *small ball probabilities around random centers*. We find basic properties and estimates. Moreover, we obtain that in the Hilbert space setting the random small ball probabilities are asymptotically equivalent to a deterministic function. A similar statement is proven to hold for Wiener measure in the Banach space of continuous functions. Finally, the asymptotics of random small ball functions are related to a particular coding problem.

Zusammenfassung

Diese Arbeit behandelt das asymptotische Kodierungsproblem für stochastische Prozesse. Das allgemeine Kodierungsproblem befasst sich mit dem Auffinden von guten Representationen für ein zufälliges Signal, das *Original*, in einer erlaubten Klasse von Representationen. Wir identifizieren jede Representation mit einer *Rekonstruktion* des Originals. Die erlaubten Representationen werden durch eine Bedingung an den Informationsgehalt eingeschränkt. Hierbei gibt es mehrere Möglichkeiten den Begriff des Informationsgehaltes zu definieren. Das *Quantisierungsproblem* erlaubt Rekonstruktionen, deren Träger endlich ist und weniger als e^r Elemente enthält. Das *Entropiekodierungsproblem* erlaubt Rekonstruktionen mit Entropie kleiner als r. Im *algorithmischen Kodierungsproblem* beschränkt man sich auf Rekonstruktionen, deren Shannon Information mit dem Original kleiner als r ist.

Diese Dissertation beschäftigt sich mit der asymptotischen Qualität der Rekonstruktion von optimalen Kodierungsschemata, wenn die erlaubte Information r gegen unendlich strebt; das *asymptotische Kodierungsproblem.* Unsere Analyse betrachtet Gaußsche Prozesse in Banachräumen als Original. Die Abweichung zwischen Original und Rekonstruktion wird durch eine Potenz der Distanz gemessen. Ein typisches Beispiel ist das Wienermaß auf dem Banachraum der stetigen Funktionen. Wir leiten Abschätzungen für das asymptotische Kodierungsproblem her. Unsere Schranken liefern die korrekte schwache und starke Asymptotik für eine große Klasse von Originalen in Banach- beziehungsweise Hilberträumen. Im Hilbertraumfall erhalten wir, dass alle drei Kodierungsbegriffe die gleiche Asymptotik liefern.

Ein weiteres Resultat betrachtet die Effizienz der Quantisierung mit zufällig konstruierten Codebüchern. Wir zeigen, dass unter gewissen Annahmen der neue Kodierungsbegriff die gleiche Asymptotik liefert wie der gewöhnliche Quantisierungsbegriff. Die Voraussetzungen des Resultats werden für bestimmte Maße in Hilberträumen verifiziert.

Weitere Betrachtungen befassen sich mit dem Effekt von Störungen auf das Kodierungsproblem. Diese Resultate stellen einen Zusammenhang der Kodierungskomplexität von Diffusionsprozessen und Brownscher Bewegung für das Entropie- and algorithmische Kodierungsproblem her.

Ein weiteres Thema dieser Dissertation ist das Verhalten der *Maßkonzentration in kleinen Kugeln um zufällige Zentren.* Wir leiten grundlegende Eigenschaften und Abschätzungen her. Im Hilbertraumfall wird die Äquivalenz zu einer deterministischen Funktion gezeigt. Ein ähnliches Resultat gilt für das Wienermaß im Banachraum der stetigen Funktionen. Schließlich wird die Asymptotik eines besonderen Kodierungsproblems mit der Asymptotik der Maßkonzentration in kleinen zufälligen Kugeln in Verbindung gesetzt.

Contents

Introduction

Let $(E, \|\cdot\|)$ be a separable Banach space. For a Borel measurable random element (abridged by r.e.) X in E and $s \in (0, \infty)$ we denote

$$\|X\|_{L_s(\mathbb{P})} = \big(\mathbb{E}[\|X\|^s]\big)^{1/s}.$$

The main objective of this dissertation is to study, for a given E-valued r.e. X, the *original*, and $s \in (0, \infty)$, the minimization problem

$$\inf \|X - \hat{X}\|_{L_s(\mathbb{P})}, \tag{I}$$

where the infimum is taken over a set of random elements $\hat{X}$, *reconstructions*, satisfying an *information constraint* parameterized by $r \geq 0$. We will mainly work with three constraints given by Kolmogorov in 1965 [36].

1. $\hat{X}$ is supported by a finite set with at most e^r elements (*quantization*).
2. $\hat{X}$ has entropy less than r (*entropy coding*).
3. $\hat{X}$ is such that the Shannon mutual information between X and $\hat{X}$ is less than r (*algorithmic coding*).

The terms in brackets denote the minimization problem (I) under the corresponding information constraint. The parameter $r \geq 0$, called *rate*, governs the amount of information that is allowed to be contained in $\hat{X}$.

In the quantization problem we identify a finite subset $\mathcal{C} \subset E$, the *codebook*, to a reconstruction $\hat{X}$ by demanding that $\hat{X} \in \mathcal{C}$ and

$$\min_{\hat{x} \in \mathcal{C}} \|X - \hat{x}\| = \|X - \hat{X}\|.$$

Clearly, one can restrict the minimization problem (I) to reconstructions obtained by codebooks.

We focus on the asymptotic behavior of (I) when r tends to infinity, the so called *high resolution coding problem* or *asymptotic coding problem.*

The asymptotic quantization problem was treated by Zador in 1963 ([66], [67], [68]), Bucklew and Wise in 1982 [11], and Graf and Luschgy in 2000

[27] for continuous distributions in finite dimensional spaces. It is found that for fixed $s \geq 1$ the general asymptotic quantization problem is related to the asymptotics of the quantization error of the uniform distribution on the unit cube. Although this link is known explicitly, the problem is not generally solved since the strong asymptotics of the quantization error of the uniform distribution are not known in most cases.

Analogue results were obtained for the entropy constrained quantization problem which is similar to the entropy coding problem. Gray, Linder and Li [30] related the asymptotics of the general problem to those obtained for the uniform distribution on the unit cube. As for the quantization problem, the asymptotics of the latter coding problem are known only in a few cases.

Let us now focus on the infinite dimensional setting. Let X be a centered Gaussian random element in a separable (infinite dimensional) Banach space $(E, \|\cdot\|)$ and denote by $\mu = \mathcal{L}(X)$ the law of X (e.g. X is Brownian motion in the space of continuous functions $C[0,1]$ equipped with the supremum norm). For the quantization error, upper asymptotic bounds for expression (I) were derived in Fehringer's 2001 dissertation [23] and strengthened in an article by Fehringer, Matoussi, Scheutzow and the author in 2003 [19]. The latter article contains also a lower bound for the quantization problem. These results relate the quantization problem to the asymptotics of the *small ball function*

$$\varphi(\varepsilon) = -\log \mu(B(0,\varepsilon)), \quad \varepsilon > 0.$$

Here, $B(x,r)$ denotes the closed Ball in E around the center $x \in E$ with radius $r \geq 0$. The above results are stated in Theorem 3.1.1. As we will prove in Theorem 3.1.2 the results of [19] imply that, for the quantization problem, expression (I) satisfies

$$\varphi^{-1}(r) \lesssim \inf \|X - \hat{X}\|_{L_s(\mathbb{P})} \lesssim 2\,\varphi^{-1}(r/2) \tag{II}$$

as $r \to \infty$, if

$$\varphi^{-1}(r) \approx \varphi^{-1}(2r) \tag{III}$$

as $r \to \infty$. Here, the symbols $\lesssim$ and $\approx$ denote asymptotic domination and weak asymptotic equivalence, respectively. Moreover, we use the symbol $\sim$ to indicate the strong asymptotic equivalence of two expressions. These notions are introduced rigorously at the end of Section 1.1.

Another approach for deriving upper bounds for the quantization error was presented in the dissertation by Creutzig in 2002 [14] (Theorem 4.6.5). He found that the quantization error is related to an approximation quantity called *average Kolmogorov width*. Although his result does not strengthen the results above in the Gaussian setting, it has the advantage to be applicable for non-Gaussian originals.

Suppose now that $H = E$ is a separable Hilbert space and let $s = 2$. Zador (1963) [66] observed that the quantization error of Gaussian measures on high but finite dimensional Euclidean spaces approaches the distortion rate function. This suggests a tight relation between the high resolution quantization problem and the asymptotics of the distortion rate function. Since the distortion rate function for the normal distribution is explicitly known for $s = 2$, this approach can be used for deriving the explicit asymptotics of the corresponding high resolution quantization problem. Indeed, the asymptotic equivalence of distortion rate function and quantization error is stated in Donoho [20] for certain infinite dimensional Gaussian measures. However, Donoho did not include a proof of the equivalence in his report. Luschgy and Pagès [51] (see also [50]) presented a proof under the assumption that the eigenvalues of the covariance operator of X are regularly varying. Beside the case $s = 2$ no other cases have been treated in the literature. Moreover, the high resolution coding problem for entropy coding and algorithmic coding has not appeared in the literature for infinite dimensional originals. In the sequel, we give an outline of the thesis and a brief summary of new results.

Chapters 1 and 2 have a preliminary character. Chapter 1 commences with an introduction to the basic objects. Known results on the high resolution coding problem in the finite dimensional setup are stated. Moreover, we provide the reader with basic facts about information theory. Chapter 2 is devoted to the introduction of Gaussian measures and their properties. In the appendix (Chapter A), we provide basic results for regularly varying functions. A summary of the notations and symbols used in this work is given in an extra chapter starting on page 145.

Chapter 3 is concerned with the high resolution coding problem for Gaussian originals on Banach spaces. In Theorem 3.4.1 we derive a relation between small ball probabilities (SBPs) and certain moment generating functions. This link leads to a lower bound for the algorithmic coding error (see Theorem 3.5.1). Finally, we find (Theorem 3.5.2) that formula (II) holds for all three coding problems and for any $s \geq 1$, if condition (III) is fulfilled. A second subject of this chapter is the efficiency of quantization based on ε-nets. More explicitly, we consider reconstructions associated with codebooks that are ε-nets of certain precompact subspaces of E. In Theorem 3.2.3, we find an upper bound on the corresponding coding error. In the case where condition (III) is satisfied, we obtain quantization schemes that are of the optimal weak asymptotic order. Finally, we give asymptotic bounds for several originals in Banach spaces.

Chapter 4 studies the effect of perturbations in the asymptotic coding problems under certain regularity conditions. These are satisfied, for instance, for all examples given in Section 3.6. We consider perturbations in the rate and the original. For entropy coding and the distortion rate function, it is found that perturbations $r \mapsto \Delta r$ of order $o(r)$ in the rate have no effect on the asymptotic behavior of expression (I) (see Lemmas 4.1.1 and 4.2.4). Moreover, adding less complex processes to the originals leaves the high resolution problem unchanged (see Corollaries 4.1.4 and 4.2.7). These results enable us to relate the coding problem of certain diffusion processes with that of Brownian motion (Theorem 4.4.1). For the quantization problem only weaker results are obtained.

Chapter 5 is devoted to the study of quantization with randomly generated codebooks. Theorem 5.2.1 proves that under certain assumptions random codebooks yield coding errors that are equivalent to those obtained by general quantization. One of the assumptions is that the quantization error is slowly varying in the parameter $\log r$. Unfortunately this property could not be proven for our standard examples. However, the assumptions are verified in the typical Hilbert space setting treated in Chapter 6.

Chapter 6 studies the high resolution coding problem for Gaussian measures in Hilbert spaces. The asymptotics of the quantization error with deterministic and random codebooks are found for all moments $s > 0$ if the eigenvalues $\{\lambda_n\}_{n\in\mathbb{N}}$ of the covariance operator satisfy

$$\lim_{n\to\infty} \frac{\log\log(1/\lambda_n)}{n} = 0.$$

Moreover, these asymptotics coincide with those of the entropy coding error and the distortion rate function for all moments $s \geq 2$ (Theorems 6.2.1 and 6.3.1) and we have equivalence of all moments, meaning that the asymptotics of (I) do not depend on the choice of s. Finally, examples are given.

Chapter 7 is devoted to the study of *small ball probabilities around random centers*. Let X be a Gaussian random element in the separable Banach space $(E, \|\cdot\|)$ with law μ. We consider the asymptotics of the random variables

$$-\log \mu\big(B(X,\varepsilon)\big) \tag{IV}$$

for $\varepsilon > 0$ as ε tends to 0. In Theorem 7.1.1, we find that the asymptotics of expression (IV) are related to the small ball function by

$$\varphi(\varepsilon) \lesssim -\log \mu\big(B(X,\varepsilon)\big) \lesssim 2\,\varphi(\varepsilon/2)$$

as $\varepsilon \downarrow 0$, almost surely (a.s.). In the case where E is a Hilbert space, we prove in Theorem 7.2.2 the existence of a deterministic function $\varphi_R : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying

$$-\log \mu\big(B(X,\varepsilon)\big) \sim \varphi_R(\varepsilon) \quad \text{as } \varepsilon \downarrow 0, \text{ a.s.}$$

The function φ_R can be explicitly derived from the eigenvalues of the covariance operator of μ. In Theorem 7.3.1, we obtain a similar result for Wiener measure on the Banach space of continuous functions $C[0,1]$, i.e. $C[0,1]$ equipped with the supremum norm. We find that there exists $\kappa \in \mathbb{R}_+$ such that expression (IV) satisfies

$$-\log \mu\big(B(X,\varepsilon)\big) \sim \kappa\, \varepsilon^{-2} \quad (\varepsilon \downarrow 0)$$

in probability.

A further subject of this chapter is the study of the quantization error obtained for certain randomly generated codebooks. Let $\{\tilde{X}_i\}_{i\in\mathbb{N}}$ be a sequence of independent μ-distributed random elements. We consider the quantization error corresponding to the random codebooks $\mathcal{C}_r = \{\tilde{X}_1, \dots, \tilde{X}_{\lfloor e^r \rfloor}\}$, $r \geq 0$. Here, $\lfloor x \rfloor$, $x \in \mathbb{R}$, denotes the largest integer smaller than x. We denote by $\hat{X}^{(r)}$ a reconstruction corresponding to the codebook $\mathcal{C}_r$. Our result assumes that there exists a convex function $\varphi_R : \mathbb{R}_+ \to \mathbb{R}_+$ that is one-to-one and onto, and satisfies

$$-\log \mu\big(B(X,\varepsilon)\big) \sim \varphi_R(\varepsilon) \quad (\varepsilon \downarrow 0)$$

in probability. Moreover, it is assumed that $\varphi_R^{-1}(r) \approx \varphi_R^{-1}(2r)$ as $r \to \infty$. Theorem 7.5.2 states that under these conditions, one has for all $s > 0$

$$\|X - \hat{X}^{(r)}\|_{L_s(\mathbb{P})} \sim \varphi_R^{-1}(r)$$

as $r \to \infty$.

Finally, Chapter 8 lists open problems.

Acknowledgment

Foremost, I sincerely thank my advisor Prof. Michael Scheutzow for introducing me to an intriguing area of probability theory and for his continual support throughout my work on this thesis.

I want to thank Jürgen Gärtner for accepting the task of being the co-examiner of this thesis. Thanks are also due to Profs. Jim Kuelbs, Werner Linde and Heinrich von Weizsäcker who gave me the opportunity to present my work and to Prof. Amir Dembo for fruitful discussions.

Throughout my time in Berlin, I very much appreciated the inspiring atmosphere in the stochastic community at the Technical University, Humboldt University and the Weierstrass Institute. I thank all members of the group for the pleasant time I spent here. Furthermore, I thank my colleagues and friends Stefan Ankirchner, Peter Arts, Jakob Creutzig, Georgi Dimitroff, Felix Esche, Matthias an der Heiden, Ira Matuschke, Florian Sobieczky, Michel Sortais, Christian Vormoor, Stefan Weber and Henryk Zähle for inspiring discussions. Finally, I am very grateful to my parents and Heike Traidl for their steady support throughout the years.

Financial support by the Deutsche Forschungsgemeinschaft through the Graduiertenkolleg "Stochastic Processes and Probabilistic Analysis" is gratefully acknowledged.

Chapter 1

Definitions and basic facts

The principal goal of *source coding* (or *data compression*) is to replace data by a compact representation such that the original can be reconstructed either perfectly or with high accuracy. In the compression of computer files one typically needs the reconstruction to resemble the original perfectly; so called *lossless source coding.* When the data represents audio or video signals, perfect reconstruction is usually not feasible. In these cases there will be a discrepancy between original and reconstruction, and we assume that this perceptual difference is measured by a *distortion measure.* In this case we speak of *lossy source coding.* To find good coding procedures is an important task for engineers and has attracted much attention in recent years.

In this dissertation we focus on lossy data compression of signals living in infinite dimensional spaces. The following theory is intended to be applicable for the infinite dimensional setting.

1.1 The general (lossy) coding problem

The general (lossy) coding problem assumes

- a Polish space (E, d) (i.e. complete, separable metric space), called *(source) alphabet*,
- a probability distribution μ on the Borel sets of E named *source distribution*, and
- a Borel-measurable function $\rho : E \times E \to [0, \infty)$ called *distortion measure* or *distortion.*

Here, $E \times E$ is equipped with the product topology. To avoid uninteresting technicalities we assume that E is a Polish space even though the problem can be stated in more general alphabets as well. We denote by $\mathcal{B}(E)$ the Borel σ-field on E and by $\otimes$ the standard product for σ-algebras. Furthermore, let $\mathcal{M}_1(E)$ denote the set of probability measures on the Borel sets of E.

We call the pair (μ, ρ) an *information source* on the *alphabet* E. Then μ is thought of as the underlying distribution for a random data signal. We usually denote by

$$X : \Omega \to E$$

a μ-distributed E-valued random element, which is called *original data signal* or simply *original*. Hereafter, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes the underlying probability space. It is assumed to be rich enough so as to ensure the existence of a random element with law ν for any probability distribution ν on the Borel sets of an arbitrary Polish space. This property is equivalent to either one of the following two properties:

1. $\mathbb{P}$ does not contain any atoms.

2. There exists a uniformly distributed random variable (r.v.) on $[0, 1]$.

The task of coding is to store information about the element X such that afterwards one is able to construct a signal $\hat{X}$ which is close to the original X. We assume that $\hat{X}$ is a random element in E. It is called a *reconstruction* of X. The corresponding (random) perceptual distortion between X and $\hat{X}$ is modeled by $\rho(X, \hat{X})$.

An important class of distortion measures are the *difference distortion measures*. Assume that the underlying space E is equipped with a Borel-measurable operation $-_E$,

$$E \times E \to E,\ (x, y) \mapsto x -_E y.$$

Clearly this is fulfilled if E is a topological vector space. If a distortion measure ρ on E can be represented in the form

$$\rho : E \times E \to [0, \infty),\ (x, \hat{x}) \mapsto \rho(x, \hat{x}) = \tilde{\rho}(x - \hat{x}),$$

for some measurable map $\tilde{\rho} : E \to [0, \infty)$, then ρ is called a *difference distortion measure*. With only slight abuse of notation we usually identify $\tilde{\rho}$ with ρ and write $\rho(x, y) = \rho(x - y)$.

In this dissertation we consider difference distortion measures based on Banach space norms (*norm-based distortions*), i.e. we assume that the alphabet $(E, \|\cdot\|)$ is a Banach space and consider

$$\rho(x, \hat{x}) = \|x - \hat{x}\|^s$$

for some $s > 0$. We denote the corresponding information source by $(\mu, \|\cdot\|^s)$.

The major object of this dissertation is the minimization problem

$$\inf \mathbb{E}[\rho(X, \hat{X})],$$

where the infimum is taken over all random elements $\hat{X}$ in E satisfying an information constraint parameterized by a parameter $r \geq 0$. Here, we mainly deal with three notions of information outlined, for instance, by Kolmogorov in 1965 ([36]). These are

- *combinatorial approach*: the logarithm of the cardinality of the support of $\hat{X}$
$$\log |\mathrm{supp}(\hat{X})|,$$
- *probabilistic approach*: the entropy of $\hat{X}$
$$\mathbb{H}(\hat{X}),$$
- *algorithmic approach*: the mutual information between X and $\hat{X}$
$$I(X; \hat{X}).$$

In the following sections we will introduce these notations and the corresponding coding quantities in great detail.

We need some asymptotic equality and inequality signs. Let f_1 and f_2 be two positive real-valued functions defined on a set $I \subset \mathbb{R}$ such that $[C, \infty) \subset I$ for some $C > 0$. We write

$$f_1(x) \lesssim f_2(x) \quad (x \to \infty)$$

when we mean

$$\limsup_{x \to \infty} \frac{f_1(x)}{f_2(x)} \leq 1.$$

Analogously, we write $f_1(x) \gtrsim f_2(x)$ $(x \to \infty)$ if $f_2(x) \lesssim f_1(x)$ $(x \to \infty)$. The functions f_1 and f_2 are said to be (strongly) asymptotically equivalent if $f_1(x) \lesssim f_2(x)$ and $f_1(x) \gtrsim f_2(x)$ $(x \to \infty)$. In that case we write $f_1(x) \sim f_2(x)$ $(x \to \infty)$. Moreover, f_1 and f_2 are called weakly asymptotically equivalent if there exists $C \in \mathbb{R}_+$ such that $f_1(x) \lesssim C\, f_2(x)$ and $f_2(x) \lesssim C\, f_1(x)$ as $x \to \infty$. In this case we write $f_1(x) \approx f_2(x)$ $(x \to \infty)$. Analogously, we use asymptotic relations for the limit $x \downarrow 0$.

1.2 Quantization

The notion *quantization* was first used in the late 40's in the context of pulse-code modulation (PCM). PCM is a method that translates an analog sound signal into a digital representation.

In 1924, Nyquist [53] discovered that a function f that is band-limited to W can be recovered perfectly from the sequence $\{f(n/(2W))\}_{n\in\mathbb{Z}}$. In fact, one has

$$f(x) = \sum_{n\in\mathbb{Z}} f\Big(\frac{n}{2W}\Big)\, \frac{\sin\big(2\pi W(t-\frac{n}{2W})\big)}{2\pi W(t-\frac{n}{2W})}, \quad x\in\mathbb{R}.$$

Ideally we suppose that the waveform of a sound signal is given by a curve f which is band-limited to W. In real applications, high frequencies are removed by a low pass filter. In order to translate the analog signal f into a digital representation one needs to store at each time instance $n/(2W)$, $n\in\mathbb{Z}$, the value $f(n/(2W))$. Clearly, this cannot be done perfectly and it is necessary to *discretize* the real value $f(n/(2W))$. Here one needs to find a compromise between the reconstruction quality and the coding complexity. This discretization problem was first called quantization. (see also Oliver, Pierce and Shannon [54] and Cover and Thomas [13], Theorem 10.3.1)

For an information source (μ,ρ), we define the *quantization error* by

$$\delta^{(q)}(N|\mu,\rho) = \inf\Big\{\int_E \min_{\hat{x}\in\mathcal{C}} \rho(x,\hat{x})\,\mu(dx) : \mathcal{C}\subset E, |\mathcal{C}|\le N\Big\}, \quad N\ge 1.$$

For an original X, we identify a finite set $\mathcal{C}\subset E$ (called *codebook*) with an optimal reconstruction $\hat{X}$ in $\mathcal{C}$, i.e. $\hat{X}$ satisfies

$$\rho(X,\hat{X}) = \min_{\hat{x}\in\mathcal{C}} \rho(X,\hat{x}).$$

This reconstruction does not need to be unique.

Note that each element of $\mathcal{C}$ can be uniquely assigned to a binary string of length $\lceil \log_2 |\mathcal{C}|\rceil$. Here, $\lceil x\rceil$ denotes the smallest integer greater than $x\in\mathbb{R}$. For our purposes, it is more convenient to work with natural logarithms, so we call $\log_e|\mathcal{C}|$ the *rate* of the corresponding reconstruction or the *rate* of the codebook. In the following the standard logarithm log is taken to the basis e. Additionally to $\delta^{(q)}$, we define the quantization error of a source (μ,ρ) in terms of the rate by

$$D^{(q)}(r|\mu,\rho) = \inf\Big\{\int_E \min_{y\in\mathcal{C}} \rho(x,y)\,\mu(dx) : \mathcal{C}\subset E, |\mathcal{C}|\le e^r\Big\}, \quad r\ge 0.$$

When there is no ambiguity about the source distribution μ and the distortion measure ρ these parameters are often omitted.

Known results in the finite dimensional setting

The high resolution quantization problem was first treated by Zador in 1963 ([66]; see also [67], [68]). He found the asymptotics for certain continuous probability distributions on the Euclidean space $E = \mathbb{R}^d$, $d \in \mathbb{N}$, under difference distortion measures ρ of the form

$$\rho(x, y) = |x - y|^s, \quad x, y \in E.$$

Here, $|\cdot|$ denotes Euclidean distance in $\mathbb{R}^d$. Bucklew and Wise [11] and Graf and Luschgy [27] generalized his results subsequently. We state Theorem 6.2 of Graf and Luschgy [27].

Theorem 1.2.1. *Let $E = \mathbb{R}^d$, $d \in \mathbb{N}$, be equipped with an arbitrary norm $|\cdot|_E$ and let $s \geq 1$. Let $\mu \in \mathcal{M}_1(E)$ with finite $s + \varepsilon$-th moment for some $\varepsilon > 0$, i.e.*

$$\int_E |x|_E^{s+\varepsilon} \, d\mu(x) < \infty.$$

Denote by $\mu = \mu_c + \mu_s$ the Lebesgue decomposition of μ w.r.t. λ^d, where the a.c. and the singular part of μ are denoted by μ_c and μ_s, respectively. Then

$$\lim_{N\to\infty} N^{s/d} \, \delta^{(q)}(N|\mu, |\cdot|_E^s) = \kappa(|\cdot|_E, s) \left\| \frac{d\mu_c}{d\lambda^d} \right\|_{L_{d/(d+s)}(\mathbb{R}^d)},$$

where $\kappa(|\cdot|_E, s) > 0$ depends only on the Banach space E and $s \geq 1$, but not on the distribution μ.

Let $\mathcal{U}[0,1]^d$ denote the uniform distribution on the d-dimensional unit cube $[0,1]^d$. By the previous theorem, $\kappa(|\cdot|_E, s)$ is obtained by

$$\kappa(|\cdot|_E, s) = \lim_{N\to\infty} N^{s/d} \, \delta^{(q)}(N|\mathcal{U}[0,1]^d, |\cdot|_E^s).$$

Unfortunately, the coefficient κ is known explicitly only in very few cases.

Theorem 1.2.2. *1.) Let $|\cdot|_\infty$ denote the l_∞-norm on $\mathbb{R}^d$, $d \in \mathbb{N}$. Then for $s \geq 1$,*

$$\kappa(|\cdot|_\infty, s) = \frac{d}{(d+s)2^s}.$$

2.) Let $|\cdot|_2$ denote the l_2-norm on $\mathbb{R}^2$. Then for $s \geq 1$

$$\kappa(|\cdot|_2, s) = \frac{8 \cdot 2^{s/2}}{3^{(2+s)/4}} \int_0^{1/2} \int_0^{(1-x_1)/\sqrt{3}} (x_1^2 + x_2^2)^{s/2} \, dx_2 \, dx_1.$$

In particular,

$$\kappa(|\cdot|_2, 1) = \frac{2 + 3\log\sqrt{3}}{3^{7/4}\sqrt{2}} = 0.37771\ldots \quad \textit{and}$$

$$\kappa(|\cdot|_2, 2) = \frac{5}{18\sqrt{3}} = 0.1603\ldots$$

Statement 1.) follows from the discussion in Graf and Luschgy [27], page 120. Statement 2.) was proven by Fejes Tóth ([24], [25]; see also Theorem 8.15 and Example 8.12 of [27]).

Example 1.2.3. Let us consider the standard normal distribution $\mu = \mathcal{N}(0,1)$ on $\mathbb{R}$ under squared norm distortion, i.e. the information source $(\mathcal{N}(0,1), |\cdot|^2)$. By the above results one has for any $s \geq 1$

$$\delta^{(q)}(N|\mathcal{N}(0,1), |\cdot|^s) \sim \big(\frac{\pi}{2}\big)^{s/2}(1+s)^{(s-1)/2}\, N^{-s} \quad (N \to \infty)$$

and

$$D^{(q)}(r|\mathcal{N}(0,1), |\cdot|^s) \sim \big(\frac{\pi}{2}\big)^{s/2}(1+s)^{(s-1)/2}\, e^{-sr} \quad (r \to \infty).$$

In particular, for $s = 2$, one has $D^{(q)}(r|\mathcal{N}(0,1), |\cdot|^2) \sim c\, e^{-2r}$ with $c = 2.7207\dots$

For more informations on the finite dimensional quantization problem, one may consult the recent monograph by Graf and Luschgy [27].

1.3 Variable rate compression and entropy coding

To define variable rate compression, we first need some more definitions. We denote by $\{0,1\}^*$ the set of *strings* constituted by 0's and 1's,

$$\{0,1\}^* = \bigcup_{n \in \mathbb{N}} \{0,1\}^n \cup \{\epsilon\},$$

where ϵ is the empty word. Denote by $l(x)$ the length of a string $x \in \{0,1\}^*$. Here, we define $l(\epsilon) = 0$. Let A be a countable set. An injective function

$$\Psi : A \to \{0,1\}^*$$

is called a *prefix free representation (prefix code)* for A if for any two different elements x and y of A the string $\Psi(x)$ is not a prefix of $\Psi(y)$.

Let (μ, ρ) be an information source on an alphabet E and let X be a μ-distributed random element. A *variable rate code* consists of

- a discrete random element $\hat{X}$ on E (reconstruction) and
- a prefix free representation

$$\Psi : \mathrm{supp}(\hat{X}) \to \{0,1\}^*$$

  for the support of $\hat{X}$.

Here, the support of $\hat{X}$ (denoted by $\mathrm{supp}(\hat{X})$) is the smallest set $A \subset E$ such that $\mathbb{P}(\hat{X} \in A) = 1$. For non-discrete random variables $\hat{X}$, we denote by $\mathrm{supp}(\hat{X})$ the smallest closed set such that $\mathbb{P}(\hat{X} \in A) = 1$.

The aim of variable rate compression is to minimize

$$\mathbb{E}[\rho(X, \hat{X})]$$

over all discrete random elements $\hat{X}$ such that there exists a prefix-free code Ψ of $\mathrm{supp}(\hat{X})$ with

$$\mathbb{E}[l(\Psi(\hat{X}))] \leq r.$$

Again, $r \geq 0$ is a constraint on the information contained in $\hat{X}$. By applying *Huffman coding* to $\mathrm{supp}(\hat{X})$, there exists a prefix free representation Ψ for $\mathrm{supp}(\hat{X})$ with

$$\mathbb{E}[l(\Psi(\hat{X}))] < \mathbb{H}_2(\hat{X}) + 1,$$

where $\mathbb{H}_2(\hat{X}) = -\sum_{x \in \mathrm{supp}(\hat{X})} \mathbb{P}(\hat{X} = x) \log_2 \mathbb{P}(\hat{X} = x)$ denotes the entropy of $\hat{X}$ with basis 2. The entropy of non-discrete random elements is defined to be ∞. On the other hand, for any prefix code Ψ for $\mathrm{supp}(\hat{X})$, one has

$$\mathbb{E}[l(\Psi(\hat{X}))] \geq \mathbb{H}_2(\hat{X})$$

as a consequence of the *Kraft inequality.* We define the *entropy coding error* of rate $r \geq 0$ by

$$\begin{aligned} D^{(e)}(r) &= D^{(e)}(r|\mu, \rho) \\ &= \inf\bigl\{\mathbb{E}[\rho(X, \hat{X})] : (X, \hat{X}) \text{ r.e. in } E^2,\ \mathcal{L}(X) = \mu,\ \mathbb{H}(\hat{X}) \leq r\bigr\}, \end{aligned}$$

where $\mathbb{H}(\cdot)$ denotes the entropy with basis e. Strictly speaking the definition of $D^{(e)}$ depends on the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. However, the quantity is unique under the assumption that the probability space is sufficiently rich. Recall that we have assumed the existence of a r.e. $(X, \hat{X})$ with law ν for any distribution $\nu \in \mathcal{M}_1(E^2)$.

By the results presented above, entropy coding is closely related to variable rate compression. In the literature, one often finds a slightly different definition. Instead of allowing general discrete r.e.'s $\hat{X}$ as reconstruction, one confines oneself to r.e.'s $\hat{X}$ of the form $\hat{X} = q(X)$ where $q : E \to E$ is a Borel-measurable function. Let us denote

$$\begin{aligned} D^{(E)}(r) &= D^{(E)}(r|\mu, \rho) \\ &= \inf\bigl\{\mathbb{E}[\rho(X, q(X))] \quad : X \text{ r.e. in } E,\ \mathcal{L}(X) = \mu, \\ &\qquad\qquad q : E \to E \text{ measurable with } \mathbb{H}(q(X)) \leq r\bigr\} \end{aligned}$$

for $r \geq 0$. This coding quantity will be referred to as *entropy constrained quantization*. Due to the fact that for any discrete reconstruction $\hat{X}$

$$\mathbb{H}(\hat{X}) \leq \log |\mathrm{supp}(\hat{X})|,$$

one has

$$D^{(e)}(r) \leq D^{(E)}(r) \leq D^{(q)}(r), \quad r \geq 0.$$

For more details about the construction of efficient codes, Huffman coding and the Kraft inequality, we refer the reader to Chapter 5 of Cover and Thomas [13].

Known results in the finite dimensional setting

We consider the finite dimensional setting. Let μ be a probability measure on $\mathbb{R}^d$, $d \in \mathbb{N}$, and let X be a μ-distributed r.v. We consider entropy constrained quantization under squared norm distortion on the Euclidean space. The quantity $D^{(E)}(r|\mu, |\cdot|^2)$ was first analyzed by Zador ([66], [67]). Unfortunately, his discussion contained some mistakes. Nonetheless, his main formula was justified in an article by Gray, Linder and Li [30]. A consequence of their Theorem 1 and Lemma 1 is

Theorem 1.3.1. *Let $\pi : \mathbb{R}^d \to \mathbb{Z}^d, (x_1, \dots, x_d) \mapsto (\lfloor x_1 \rfloor, \dots, \lfloor x_d \rfloor)$ and let $\mu \in \mathcal{M}_1(\mathbb{R}^d)$ be an absolutely continuous measure (w.r.t. Lebesgue measure) with Radon-Nikodym derivative $f := \frac{d\mu}{d\lambda^d}$. Under the assumptions that μ has well-defined differential entropy*

$$h(f) := -\int_{\mathbb{R}^d} f(x) \log f(x)\, dx$$

and that the r.v. $\pi(X)$ has finite entropy, one has

$$\lim_{r \to \infty} e^{\frac{2}{d} r}\, D^{(E)}(r|\mu, |\cdot|^2) = \kappa(d)\, e^{\frac{2}{d} h(f)},$$

where

$$\kappa(d) := \lim_{r \to \infty} e^{\frac{2}{d} r}\, D^{(E)}(r|\mathcal{U}[0,1]^d, |\cdot|^2).$$

In particular, $\kappa(1) = 1/12$.

Example 1.3.2. We consider the standard normal distribution $\mu = \mathcal{N}(0,1)$ on $\mathbb{R}$. Since then $h(\frac{d\mu}{d\lambda}) = \log(\sqrt{2\pi e})$, one has

$$D^{(E)}(r|\mathcal{N}(0,1), |\cdot|^2) \sim \frac{\pi e}{6}\, e^{-2r} \quad (r \to \infty)$$

where the coefficient $\pi e/6$ is approximately equal to 1.42329. Recall that by Example 1.2.3, one obtains the coefficient 2.7207 for the quantization analog.

For absolutely continuous measures it seems highly likely that the quantities $D^{(e)}$ and $D^{(E)}$ either coincide or are closely related. We do not concern this problem in this thesis. Our main focus is on the quantity $D^{(e)}$. The results above merely give an intuition of the asymptotic behavior of $D^{(e)}$ and $D^{(E)}$ in the finite dimensional setting.

1.4 The distortion rate function

Denote by $\mathbb{H}(\,\cdot\,\|\,\cdot\,)$ the *relative entropy*, i.e. for any Borel probability measures ξ and ν on a Polish space, let

$$\mathbb{H}(\xi\|\nu) = \begin{cases} \int \log\big(\frac{d\xi}{d\nu}\big)d\xi & \text{if } \xi \ll \nu \\ \infty & \text{else.} \end{cases}$$

For two Borel random elements A and B taking values in possibly different Polish spaces, define

$$I(A;B) = \mathbb{H}(\mathbb{P}_{AB}\,\|\,\mathbb{P}_A \otimes \mathbb{P}_B),$$

where $\mathbb{P}_{AB}$, $\mathbb{P}_A$ and $\mathbb{P}_B$ denote the distributions of (A,B), A and B, respectively. $I(A;B)$ is called the *mutual information of A and B*.

For an information source (ξ,ρ), the *Shannon distortion rate function (DRF)* is defined by

$$D(r|\mu,\rho) = \inf\big\{\mathbb{E}[\rho(X,\hat{X})] \,:\, (X,\hat{X}) \text{ r.e. in } E^2,\, \mathcal{L}(X)=\mu,\, I(X;\hat{X}) \le r\big\}$$

for $r \ge 0$. We will make use of the fact that the DRF is convex. A proof of this property is contained in Ihara [32] (Theorem 1.7.1), for instance.

Lemma 1.4.1. *For an arbitrary information source (ξ,ρ), the distortion rate function $D(\cdot|\xi,\rho)$ is convex.*

The DRF is one of the main objects used by Shannon in his works from 1948 and 1959 ([60], [62]). He considered the problem of reconstructing an original X on the basis of the information received via a channel with restricted capacity. One of his main results is that the DRF gives the asymptotically best achievable accuracy of the reconstruction for a given capacity. Much research has been devoted to the study of the distortion rate function and information transmission. For more information, we refer the reader to standard textbooks on information theory (cf. Cover and Thomas [13], Berger [4] and Gray [28]).

Example 1.4.2. Shannon [62] derived the distortion rate function for the standard normal distribution under mean squared error distortion measure. He found

$$D(r|\mathcal{N}(0,1), |\cdot|^2) = e^{-2r}, \quad r \geq 0.$$

Recall that by Examples 1.2.3 and 1.3.2, one obtains coefficients 1.42329 and 2.7207 in the entropy constraint quantization and quantization analog to the above formula, respectively. In particular, we see that the asymptotic behavior of $D^{(q)}$, $D^{(E)}$ and D differ in the finite dimensional setting.

Basic properties of mutual information and the DRF

Let A, B and C be Borel random elements in some Polish spaces E_1, E_2 and E_3, respectively. Note that there exist regular conditional probabilities $\mathbb{P}_{AB|C=\cdot}$, $\mathbb{P}_{A|C=\cdot}$ and $\mathbb{P}_{B|C=\cdot}$. For $c \in E_3$, set

$$I(A;B|C=c) = \mathbb{H}\big(\mathbb{P}_{AB|C=c} \,\|\, \mathbb{P}_{A|C=c} \otimes \mathbb{P}_{B|C=c}\big).$$

Definition 1.4.3. We call

$$I(A;B|C) = \int_{E_3} I(A;B|C=c)\, d\mathbb{P}_C(c)$$

the *conditional mutual information between A and B given C.*

Lemma 1.4.4. *Let A, B and C as above. Mutual information satisfies the following properties:*

- *(Symmetry)*

$$I(A;B) = I(B;A) \text{ and } I(A;B|C) = I(B;A|C)$$

- *(Positivity)*

$$I(A;B) \geq 0$$

- *For any Borel measurable function $f : E_2 \to E_3$,*

$$I(A;B) \geq I(A; f(B))$$

-

$$I(A;B|C) \geq 0$$

In particular, $I(A;B|C) = 0$ if and only if A and B are independent given C.

- $$I(A;(B,C)) = I(A;B) + I(A;C|B).$$
- *If A is a discrete random element, then*
$$I(A;B) \le \mathbb{H}(A).$$

All but the last property are taken from Ihara [32] (Theorem 1.6.3). The last property is a consequence of the first properties:

$$I(A;B) \le I(A;(A,B)) = I(A;A) + I(A;B|A) = I(A;A) = \mathbb{H}(A).$$

Consequently, the coding quantities are ordered as follows

$$D(r) \le D^{(e)}(r) \le D^{(q)}(r), \quad r \ge 0.$$

1.5 The source coding theorem

This section is devoted to Shannon's source coding theorem (SCT). The theory originates from Shannon's works in the years 1948 and 1959 ([60], [62]) and was subsequently extended by many authors. For a history of the source coding theorem and its generalizations, we refer the reader to the review article by Berger and Gibson [5] (end of section C).

For stating the theorem we need more definitions. First we extend the distortion measure ρ to product spaces E^m. For $m \in \mathbb{N}$ and $x, y \in E^m$, let

$$\rho_m(x,y) := \frac{1}{m}\sum_{i=1}^{m} \rho(x_i, y_i).$$

ρ_m is called the *single letter distortion measure* associated with ρ. We consider the quantization error of the product measure $\mu^{\otimes m}$ under the distortion ρ_m. For $m \in \mathbb{N}$ and $r \ge 0$, let

$$\begin{aligned} D_m(r|\mu,\rho) &= D^{(q)}(mr|\mu^{\otimes m}, \rho_m) \\ &= \inf\Big\{\int_{E^m} \min_{y\in\mathcal{C}} \rho_m(x,y)\, d\mu^{\otimes m}(x) : \mathcal{C} \subset E^m, |\mathcal{C}| \le e^{mr}\Big\}. \end{aligned} \tag{1.1}$$

$D_m(r|\mu,\rho)$ is called *quantization error of block codes of length m and rate r.*

Theorem 1.5.1. *(The Source Coding Theorem (SCT)). Let (μ,ρ) be an information source and suppose that there exists a so called* reference letter*, i.e. $\exists y^* \in E$ with $\int_E \rho(x,y^*)\, d\mu(x) < \infty$. Then*

$$\lim_{m\to\infty} D_m(r|\mu,\rho) = \inf_{m\in\mathbb{N}} D_m(r|\mu,\rho) = D(r|\mu,\rho)$$

for all $r \ge 0$.

The theorem as stated above is, for instance, an immediate consequence of the much stronger SCT proved by Gray [28].

The proof of the SCT is typically based on the so called *asymptotic equipartition property (AEP)*. Following the discussion in Dembo and Kontoyiannis [15], we give a short introduction to the AEP and a sketch of the proof of the SCT. The interested reader is referred to the more detailed article of Dembo and Kontoyiannis [15].

First let us state the *asymptotic equipartition property (AEP)*. For $\nu \in \mathcal{M}_1(E)$ and $d > 0$, define

$$R_0(\mu, \nu, d) = \inf_{\xi} \mathbb{H}(\xi \| \mu \otimes \nu),$$

where the infimum is taken over all distributions ξ on the Borel sets of E^2 with first marginal μ and

$$\int_{E\times E} \rho(x, y)\, d\xi(x, y) \le d.$$

Set

$$d_{\mathrm{av}} = \int_{E\times E} \rho(x, y)\, d\mu \otimes \nu(x, y)$$

and

$$d_{\min} = \int_{E} \operatorname{essinf}_{Y\sim\nu} \rho(x, Y)\, d\mu(x),$$

where

$$\operatorname{essinf}_{Y\sim\nu} \rho(x, Y) = \sup\bigl\{t \in \mathbb{R} : \mathbb{P}(\rho(x, Y) \ge t) = 1, Y \text{ r.e. with } \mathcal{L}(Y) = \nu\bigr\}.$$

Let $\{X_j\}_{j\in\mathbb{N}}$ be a sequence of independent μ-distributed random elements and denote $X^{(m)} = (X_1, \dots, X_m)$ for $m \in \mathbb{N}$. The *asymptotic equipartition property* states that for $d \in (d_{\min}, d_{\mathrm{av}})$ one has

$$-\frac{1}{m} \log \nu^{\otimes m}(B_{\rho_m}(X^{(m)}, d)) \to R_0(\mu, \nu, d) \quad \text{as } m \to \infty, \text{ a.s.}$$

Here,

$$B_{\rho_m}(x, r) = \bigl\{y \in E^m : \rho_m(x, y) \le r\bigr\}$$

for $x \in E^m$. Let us sketch how this result can be used to prove the upper bound in the SCT.

Let $\mu \in \mathcal{M}_1(E)$ and $\nu \in \mathcal{M}_1(E)$ denote the distributions of the original signal and the distribution used for generating the codebook, respectively. Fix $d > 0$ and assume for simplicity that $d \in (d_{\min}, d_{\mathrm{av}})$, where $d_{\min}$ and d_{av} are as above.

Fix $\varepsilon > 0$ arbitrarily and let $r = R_0(\mu, \nu, d)$. We still need to define the codebooks that will be used for coding $X^{(m)}$. This is done by using the distribution ν: For $m \in \mathbb{N}$, let $\{Y^{(m)}(i)\}_{i\in\mathbb{N}}$ be a sequence of independent (also independent of X) $\nu^{\otimes m}$-distributed random elements and let $y^* \in E$ be a reference letter as in the SCT. Consider, for $m \in \mathbb{N}$,

$$\mathcal{C}_m = \big\{Y^{(m)}(i) : i = 1, \ldots, \lfloor e^{(1+\varepsilon)mr} \rfloor\big\} \cup \big\{(y^*, \ldots, y^*)\big\}$$

and

$$\mathcal{T}_m = \big\{x^{(m)} \in E^m : \frac{1}{m} \log \nu^{\otimes m}(B_{\rho_m}(X^{(m)}, d)) \geq -(1 + \varepsilon/2)r\big\}.$$

Our aim is to bound the expected coding error $\mathbb{E}[\rho_m(X^{(m)}, \mathcal{C}_m)]$. We split the expectation into three parts:

$$\begin{aligned}\mathbb{E}[\rho_m(X^{(m)}, \mathcal{C}_m)] &= \mathbb{E}[1_{\mathcal{T}_m}(X^{(m)})\, 1_{\{\rho_m(X^{(m)},\mathcal{C}_m)\leq d\}}\, \rho_m(X^{(m)}, \mathcal{C}_m)] \\ &\quad + \mathbb{E}[1_{\mathcal{T}_m}(X^{(m)})\, 1_{\{\rho_m(X^{(m)},\mathcal{C}_m)> d\}}\, \rho_m(X^{(m)}, \mathcal{C}_m)] \\ &\quad + \mathbb{E}[1_{\mathcal{T}_m^c}(X^{(m)})\, \rho_m(X^{(m)}, \mathcal{C}_m)] \\ &=: I_1(m) + I_2(m) + I_3(m).\end{aligned}$$

In the following we provide estimates for $I_1(m), I_2(m)$ and $I_3(m)$:
1.) Clearly, $I_1(m) \leq d$.
2.) One has,

$$\begin{aligned}I_2(m) &\leq \mathbb{E}[1_{\mathcal{T}_m}(X^{(m)})\, 1_{\{\rho_m(X^{(m)},\mathcal{C}_m)> d\}}\, \rho_m(X^{(m)}, (y^*, \ldots, y^*))] \\ &= \mathbb{E}[1_{\mathcal{T}_m}(X^{(m)})\, 1_{\{\rho_m(X^{(m)},\mathcal{C}_m)> d\}}\, \rho(X_1, y^*)].\end{aligned}$$

Note that

$$\mathbb{P}(X^{(m)} \in \mathcal{T}_m, \rho_m(X^{(m)}, \mathcal{C}_m) > d) \leq (1 - e^{-(1+\varepsilon/2)mr})^{\lfloor \exp\{(1+\varepsilon)mr\}\rfloor}$$

converges to 0 as $m \to \infty$. Hence, one has $\lim_{m\to\infty} I_2(m) = 0$.
3.) Analogously to 2.) we estimate

$$I_3(m) \leq \mathbb{E}[1_{\{X^{(m)}\notin\mathcal{T}_m\}} \rho(X_1, y^*)]$$

and it follows by the AEP that $\lim_{m\to\infty} I_3(m) = 0$.

Summarizing the results above yields that for arbitrary $\varepsilon > 0$

$$\limsup_{m\to\infty} \mathbb{E}[\rho_m(X^{(m)}, \mathcal{C}_m)] \leq d.$$

Consequently, there exist codebooks $\mathcal{C}_m \subset E^m$ with $e^{m(r+o(1))}$ elements yielding ρ_m-average distortion asymptotically less than d. In order to reduce the size of the random codebooks we optimize over the distributions $\nu \in \mathcal{M}_1(E)$. Hence, one can asymptotically reduce the rate per symbol to

$$R(d|\mu,\rho) := \inf_{\nu \in \mathcal{M}_1(E)} R_0(\mu,\nu,d).$$

By a result of Yang and Kieffer [65], we have

$$R_0(\mu,\nu,d) = \inf_{(X,\hat{X})} [I(X;\hat{X}) + \mathbb{H}(\mathbb{P}_{\hat{X}} \| \nu)],$$

where the infimum is taken over all random elements $(X,\hat{X})$ on E^2 with $\mathcal{L}(X) = \mu$ and

$$\mathbb{E}[\rho(X,\hat{X})] \le d.$$

It follows immediately that

$$R(d|\mu,\rho) = \inf_{(X,\hat{X})} I(X;\hat{X}),$$

where the infimum is taken over the same r.e.'s as before. The function $R(\cdot|\mu,\rho)$ is called *rate distortion function*. It is essentially the inverse of $D(\cdot|\mu,\rho)$. One can conclude that r nats per symbol suffice to achieve asymptotically ρ_m-distortion $D(r|\mu,\rho)$.

The ideas of the above sketch are useful in the results presented in this dissertation.

1.) The realizations of $X^{(m)}$ lie in $\mathcal{T}_m$ asymptotically with probability 1. $\mathcal{T}_m$ is a so called *typical set* for the random element $X^{(m)}$ and outcomes outside of $\mathcal{T}_m$ have no influence on the asymptotic coding error. We shall see that the concept of a typical set provides a useful tool in our considerations.

2.) In the sketch of the proof, we saw that certain randomly generated codebooks are asymptotically optimal. Keeping this in mind, we shall try to apply random codebooks in order to infer results for the high resolution coding problem.

Chapter 2

Gaussian measures

Let $(E, \|\cdot\|)$ be a separable Banach space and let E' denote its topological dual equipped with the norm

$$\|f\|_{E'} := \sup_{x \in B_E(0,1)} |f(x)|, \quad f \in E'.$$

Definition 2.0.2. A measure μ on the Borel sets of E is called *Gaussian measure on* E if

- for any $f \in E'$
  $$\mu \circ f^{-1}$$
  is a normal distribution on $\mathbb{R}$, and
- μ is not a Dirac measure on E.

Analogously, we call a random element X on E with Gaussian law, a *Gaussian random element* in E.

In this section, let μ be a Gaussian measure on E and let X denote a μ-distributed random element in E. Let us first state some basic properties of Gaussian measures. There exist an element $a_\mu \in E$ and a linear operator

$$C_\mu : E' \to E$$

such that

$$\mathbb{E}[f(X)] = f(a_\mu)$$

and

$$\mathbb{E}[f(X - a_\mu)g(X - a_\mu)] = f(C_\mu g)$$

for all $f, g \in E'$. The element a_μ and the operator C_μ are called *barycenter* and *covariance operator* of μ. A Gaussian measure is uniquely determined by its barycenter and covariance operator. The measure μ is called *centered* if $a_\mu = 0$. Note, that we can transform any Gaussian vector X into a centered Gaussian vector with the same covariance operator by shifting X by $-a_\mu$. We will restrict our attention to centered Gaussian measures μ. Results for the general case can be inferred by shifting the measures appropriately. Thereafter let μ be centered.

An important quantity will be the *small ball function*

$$\begin{aligned} \varphi : \mathbb{R}_+ &\to \mathbb{R}_+ \\ \varepsilon &\mapsto -\log \mu(B(0,\varepsilon)). \end{aligned}$$

A proof for the existence of a_μ and C_μ is contained in Lifshits [48], Section 8 (see also Bogachev [7], Theorem 3.2.3). The continuity and compactness of C_μ can be deduced from Theorem 3.2.3 and Corollary 3.2.4 in [7].

2.1 The reproducing kernel Hilbert space

Denote for a Borel measurable function $f : E \to \mathbb{R}$,

$$\|f\|_{L_2(\mu)} = \Big(\int_E f(x)^2 \, d\mu(x)\Big)^{1/2}$$

and let

$$L_2(\mu) = \{f : E \to \mathbb{R} \, : \, \|f\|_{L_2(\mu)} < \infty\}$$

be the Hilbert space of square integrable functions. For $x \in E$, we set

$$\|x\|_{H_\mu} = \sup\{f(x) : f \in E', \|f\|_{L_2(\mu)} \le 1\}.$$

and let

$$H_\mu = \{x \in E : \|x\|_{H_\mu} < \infty\}.$$

Note that one has $C_\mu(E') \subset H_\mu$, since for $x = C_\mu g \in C_\mu(E')$ and $f \in E'$

$$f(x) = f(C_\mu g) = \langle f, g\rangle_{L_2(\mu)} \le \|f\|_{L_2(\mu)} \, \|g\|_{L_2(\mu)}.$$

Here, $\langle \cdot, \cdot \rangle_{L_2(\mu)}$ denotes the scalar product in the Hilbert space $L_2(\mu)$.

The set H_μ equipped with the norm $\|\cdot\|_{H_\mu}$ constitutes a Hilbert space; the so called *Cameron-Martin space* or *reproducing kernel Hilbert space* associated with μ ([7], Lemma 2.4.1 and Theorem 3.2.3). The space H_μ is compactly embedded in E ([7], Corollary 3.2.4). We denote

$$\sigma = \sigma(\mu) = \sup_{x \in E \setminus \{0\}} \frac{\|x\|_E}{\|x\|_{H_\mu}} < \infty. \tag{2.1}$$

Example 2.1.1. Let μ denote the Wiener measure on the Banach space of continuous functions $C([0,1],\mathbb{R}^d)$ equipped with the supremum norm $\|\cdot\|_{[0,1]}$. In that case the reproducing kernel Hilbert space is the set

$$H_\mu = \Big\{ \int_0^{\cdot} f(s)\,ds : f \in L_2([0,1],\mathbb{R}^d) \Big\}$$

equipped with the Hilbert space norm $\| \int_0^{\cdot} f(s)\,ds \|_{H_\mu} = \|f\|_{L_2[0,1]}$.

2.2 Basic properties of Gaussian measures

Lemma 2.2.1. *The support of μ ($\mathrm{supp}(\mu)$) is a closed linear subspace of E and one has*

$$\mathrm{supp}(\mu) = \overline{H_\mu},$$

where the closure is taken in the Banach space norm.

Lemma 2.2.2. *(Anderson's inequality). Let A be a symmetric convex Borel set in E. For any $x \in E$, one has*

$$\mu(x + A) \le \mu(A),$$

where $x + A = \{x + z : z \in A\}$.

Gaussian measures satisfy a zero-one law:

Lemma 2.2.3. *(Zero-one law). Suppose that A is a Borel set on E such that for any $h \in C_\mu(E')$,*

$$\mu(A + h) = \mu(A).$$

Then, $\mu(A) \in \{0, 1\}$.

We will need an estimate for the measure of shifted balls. We denote for $x \in E$ and $\varepsilon \ge 0$

$$I(x,\varepsilon) = \inf\Big\{ \frac{\|h\|_{H_\mu}^2}{2} \;:\; h \in H_\mu \cap B_E(x,\varepsilon) \Big\}.$$

Here the infimum of the empty set is assumed to be ∞.

Lemma 2.2.4. *(Estimate of shifted balls). For any $x \in E$, $\varepsilon > 0$ and $a \in [0,1]$, it holds*

$$\mu(B(x,\varepsilon)) \ge \exp\{-I(x,a\varepsilon) - \varphi((1-a)\varepsilon)\}.$$

Lemma 2.2.5. *(Cameron-Martin formula). Let $f \in E'$ and $h = C_\mu f \in H_\mu$. The measure $\mu_h \in \mathcal{M}_1(E)$ defined by*

$$\mu_h(A) = \mu(A - h), \quad A \in \mathcal{B}(E),$$

is absolutely continuous w.r.t. μ and the corresponding Radon-Nikodym derivative is

$$\frac{d\mu_h}{d\mu}(x) = \exp\{f(x) - \frac{1}{2}\|h\|_{H_\mu}^2\}, \quad x \in E.$$

Lemma 2.2.6. *Assume that H_μ is infinite dimensional and denote by $\{e_j\}_{j\in\mathbb{N}}$ a complete orthonormal system in H_μ. Let $\{X_j\}_{j\in\mathbb{N}}$ be a sequence of i.i.d. standard normals. Then the limit*

$$X := \lim_{n\to\infty} \sum_{j=1}^{n} X_j e_j$$

exists a.s. in the Banach space norm and $\mathcal{L}(X) = \mu$.

A proof of Anderson's inequality is contained in [7], Theorem 2.8.10. The zero-one law and the series representation (Lemma 2.2.6) are proven in [7] Theorem 2.5.2 and Theorem 3.5.1. The estimate of shifted balls is taken from Li and Shao [47], Theorem 3.2. The Cameron-Martin formula is proven in [7], Corollary 2.4.3.

2.3 Concentration inequalities for Gaussian measures

Set for $t \in \mathbb{R}$

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2}\, dx.$$

Moreover, we let $\Phi(\infty) = 1$ and $\Phi(-\infty) = 0$. Then,

$$\Phi : \mathbb{R} \cup \{-\infty, \infty\} \to [0, 1]$$

is bijective. Moreover, we let $\Psi = 1 - \Phi$ and denote by

$$\mathcal{K} = \{x \in H_\mu : \|x\|_{H_\mu} \le 1\}$$

the closed unit ball of H_μ.

Lemma 2.3.1. *(Isoperimetric or Borell's inequality). Let $A \in \mathcal{B}(E)$ and $t \ge 0$. One has*

$$\mu(A + t\mathcal{K}) \ge \Phi(t + \Phi^{-1}(\mu(A))).$$

The isoperimetric inequality shows that the measure is concentrated around 0. Let M denote a median of the random variable $\|X\|$ and let $t > 0$. Note that

$$B(0, M+t) \supset B(0, M) + \frac{t}{\sigma}\,\mathcal{K},$$

where $\sigma = \sigma(\mu)$ is as in (2.1). Consequently, with the isoperimetric inequality,

$$\mu(B(0, M+t)) \geq \Phi\Big(\frac{t}{\sigma} + \Phi^{-1}(1/2)\Big) = \Phi\big(\frac{t}{\sigma}\big).$$

Therefore,

$$\mu(B(0, M+t)^c) \leq \Psi(t/\sigma) \leq \frac{1}{2}\exp\{-t^2/2\sigma^2\}. \tag{2.2}$$

In particular, the random variable $\|X\|$ has Gaussian tails and all moments of $\|X\|$ are finite. A consequence of this strong concentration property is

Lemma 2.3.2. *(Equivalence of moments). For any $p, q > 0$ there exist universal constants $K_{p,q}$ such that*

$$\mathbb{E}[\|X\|^p]^{1/p} \leq K_{p,q}\,\mathbb{E}[\|X\|^q]^{1/q}$$

for any Gaussian vector X on an arbitrary separable Banach space E.

Lemma 2.3.3. *(Ehrhard's inequality). Let A_1 and A_2 be two nonempty convex Borel sets in E. Assume that $\mu(A_1) < 1$ and $\mu(A_2) < 1$. Then, for all $\gamma \in [0,1]$,*

$$\Phi^{-1}\big(\mu_*\{\gamma A_1 + (1-\gamma)A_2\}\big) \geq \gamma\,\Phi^{-1}(\mu(A_1)) + (1-\gamma)\,\Phi^{-1}(\mu(A_2)),$$

where μ_* denotes the inner measure of μ.

Corollary 2.3.4. *For any $x \in E$, the function*

$$\begin{aligned} \mathbb{R}_+ &\to [-\infty, \infty) \\ t &\mapsto \Phi^{-1}(\mathbb{P}(\|X - x\| \leq t)) \end{aligned}$$

is concave.

Proof. Fix $x \in E$ and let $t_1, t_2 > 0$ and $\gamma \in (0,1)$. Choose $A_1 = B(x, t_1)$ and $A_2 = B(x, t_2)$. Then $B(x, \gamma t_1 + (1-\gamma)t_2) = \gamma A_1 + (1-\gamma)A_2$ and the result is obtained via Lemma 2.3.3. □

A consequence of the previous corollary is

Lemma 2.3.5. *The small ball function*

$$\begin{aligned} \varphi : \mathbb{R}_+ &\to \mathbb{R}_+ \\ \varepsilon &\mapsto -\log \mu(B(0,\varepsilon)) \end{aligned}$$

is monotonically decreasing, one-to-one, onto and convex.

The isoperimetric inequality and the Ehrhard inequality are taken out of Bogachev [7] (Theorem 4.3.3, Theorem 4.2.2). The equivalence of moments is proven by Ledoux and Talagrand [44], Corollary 3.2. See also chapter 3.1 on integrability and tail behavior in [44].

Chapter 3

Coding Gaussian measures on Banach spaces

3.1 An upper bound for the quantization error

Let μ be a centered Gaussian measure on a separable Banach space $(E, \|\cdot\|)$. First we consider the high resolution quantization problem for $(\mu, \|\cdot\|^s)$ for some $s > 0$. This problem was first treated in the dissertation by Fehringer [23]. In a proceeding article by Dereich et al. [19], Fehringer's results were extended.

Fehringer considered *quantization with random codebooks generated by the distribution of the original.* We denote, for an arbitrary information source (ξ, ρ) and $r \geq 0$,

$$D^{(R)}(r|\xi, \rho) = \int \int \min_{j=1,\ldots,N} \rho(x, y_j)\, d\mu^{\otimes N}(y_1, \ldots, y_N)\, d\mu(x),$$

where $N = \lfloor e^r \rfloor$.

As before let φ be the small ball function of μ, i.e.

$$\varphi(\varepsilon) = -\log \mu(B(0, \varepsilon)), \quad \varepsilon > 0.$$

It follows Theorem 2.1 and Theorem 3.1 of [19]:

Theorem 3.1.1. *1.)* Upper bound. *Assume that*

$$\lim_{\varepsilon \downarrow 0} \frac{\varphi(\varepsilon)}{(\log \frac{1}{\varepsilon})^{1/a}} = \infty$$

for some $a \in (0, 1)$. Then for any $\varsigma \in (0, 1)$

$$D^{(R)}(r|\mu, \|\cdot\|^s)^{1/s} \lesssim 2\, \varphi^{-1}\Big(\frac{(1-\varsigma)r}{2}\Big) \quad (r \to \infty).$$

2.) Lower bound. *For every $s, \kappa > 0$, one has*

$$D^{(q)}(r|\mu, \|\cdot\|^s)^{1/s} \gtrsim (1 - e^{-\kappa})^{1/s} \varphi^{-1}(r + \kappa) \quad (r \to \infty).$$

A consequence of Theorem 3.1.1 is

Theorem 3.1.2. *Suppose φ fulfills*

$$\varphi^{-1}(2r) \approx \varphi^{-1}(r) \quad (r \to \infty). \tag{3.1}$$

Then one has for arbitrary $s > 0$

$$\begin{aligned}\varphi^{-1}(r) &\lesssim D^{(q)}(r|\mu, \|\cdot\|^s)^{1/s} \\ &\leq D^{(R)}(r|\mu, \|\cdot\|^s)^{1/s} \lesssim 2\,\varphi^{-1}(r/2)\end{aligned}$$

as $r \to \infty$.

In particular, Theorem 3.1.2 yields the weak asymptotics of the quantization error if condition (3.1) is satisfied.

Remark 3.1.3. 1.) For most infinite dimensional Gaussian measures of interest, the small ball function $\varphi(\varepsilon)$ is regularly varying at 0 with some index $-\alpha < 0$, i.e. $\varphi(\varepsilon) = \varepsilon^{-\alpha} l(1/\varepsilon)$ for some slowly varying function l. Then, by the theory of regularly varying functions (see Remark A.5), the inverse φ^{-1} is regularly varying at ∞. In particular, it holds condition (3.1) and the previous theorem is applicable. Examples for which the asymptotics of φ are known are provided in Chapter 3.6.

2.) Note that it is not in the scope of Theorem 3.1.1 to give tight bounds for the weak asymptotics of $D^{(q)}$ and $D^{(R)}$ if condition (3.1) is not satisfied. The problem of finding the weak asymptotics of the quantization error is still open when the small ball function tends to infinity (in 0) slower than every polynomial $\varepsilon^{-\alpha}$, $\alpha > 0$. For instance, this is the case when the underlying process X is a smooth Gaussian process in $C[0, 1]$.

Lemma 3.1.4. *Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be a monotonically decreasing, convex function satisfying*

$$f(2r) \approx f(r) \quad (r \to \infty).$$

Then, for $\Delta : \mathbb{R}_+ \to \mathbb{R}$, $r \mapsto \Delta r$ with $\Delta r = o(r)$ $(r \to \infty)$, one has

$$f(r + \Delta r) \sim f(r) \text{ as } r \to \infty. \tag{3.2}$$

Proof. Statement (3.2) is equivalent to

$$\lim_{r \to \infty} \frac{f(r + \Delta r) - f(r)}{f(r)} = 0$$

By the convexity and monotonicity of f, it follows for every $r > 0$ with $\Delta r > -r$ that

$$\Big|\frac{f(r+\Delta r)-f(r)}{f(r)}\Big| \le \frac{f(r-|\Delta r|)-f(r)}{f(r)}. \tag{3.3}$$

Again by the convexity of f, one has

$$f(r-|\Delta r|)-f(r) \le \frac{1}{t}[f(r-t|\Delta r|)-f(r)] \tag{3.4}$$

for any $t \ge 1$ with $t|\Delta r| < r$. Let $r_0 \ge 0$ be sufficiently large such that $|\Delta r|/r \le 1/2$ for every $r \ge r_0$. For $r \ge r_0$, set

$$t(r) = \frac{r}{2|\Delta r|}.$$

Combining (3.3) and (3.4) yields, for any $r \ge r_0$,

$$\begin{aligned}\Big|\frac{f(r+\Delta r)-f(r)}{f(r)}\Big| &\le \frac{1}{t(r)}\Big[\frac{f(r-t(r)|\Delta r|)}{f(r)}-1\Big]\\ &= \frac{1}{t(r)}\Big[\frac{f(r/2)}{f(r)}-1\Big].\end{aligned}$$

By assumption, $\limsup_{r\to\infty} f(r/2)/f(r) < \infty$. Since $t(r)$ tends to infinity, the previous expression converges to 0 as $r \to \infty$. □

Proof of Theorem 3.1.2. We start with verifying the assumptions of Theorem 3.1.1. We fix $r_0 \ge 0$ and $\eta \in (0,1)$ such that

$$\varphi^{-1}(2r) \ge \eta\varphi^{-1}(r)$$

for all $r \ge r_0$. Hence, for any $n \in \mathbb{N}$, one has

$$\varphi^{-1}(2^n r_0) \ge \eta^n\varphi^{-1}(r_0).$$

For $r \ge r_0$ we choose $n = \lfloor \log\frac{r}{r_0}/\log 2\rfloor + 1$. Then

$$\begin{aligned}\varphi^{-1}(r) &\ge \varphi^{-1}(2^n r_0) \ge \eta^n \varphi^{-1}(r_0)\\ &\ge \eta^{1+\log\frac{r}{r_0}/\log 2}\varphi^{-1}(r_0) = \eta\Big(\frac{r}{r_0}\Big)^{\log\eta/\log 2}\varphi^{-1}(r_0).\end{aligned}$$

Denoting $a := -\log\eta/\log 2 > 0$ and $c := \eta r_0^a\varphi^{-1}(r_0) > 0$, one has

$$\varphi^{-1}(r) \ge cr^{-a}$$

for $r \ge r_0$. Set $\varepsilon_0 := \varphi^{-1}(r_0)$ and consider arbitrary $\varepsilon \in (0,\varepsilon_0]$ and $r := \varphi(\varepsilon) \le r_0$. Using the previous inequality one has

$$\varepsilon = \varphi^{-1}(r) \ge cr^{-a}$$

and, hence,

$$\varphi(\varepsilon) = r \geq \left(\frac{\varepsilon}{c}\right)^{-1/a}.$$

The assumption of part 1.) of Theorem 3.1.1 follows. By an application of the previous lemma we obtain the correct upper bound. The lower bound follows immediately by Lemma 3.1.4, since $\varphi^{-1}(r+\kappa) \sim \varphi^{-1}(r)$. □

Theorem 3.1.2 implies that random coding yields coding errors of asymptotically optimal order in many cases. Motivated by this fact, one may implement a quantization scheme by generating randomly a codebook $\mathcal{C} = \{Y_1, \dots, Y_{\lfloor e^r \rfloor}\}$ of some rate $r \geq 0$ and keeping it fixed afterwards. Unfortunately, this method has a non-negligible disadvantage in its practicability. The elements of $\mathcal{C}$ are arbitrarily placed in E and do not have nice geometric properties. Hence, one encounters the problem of constructing "fast" algorithms that find good representations of X in $\mathcal{C}$. Since $r > 0$ is typically thought to be large, it seems not feasible to go through the codebook sequentially to find the best representation.

Alternatively, we consider quantization with codebooks that are based on *ε-nets*.

3.2 Using ε-nets as codebooks

Let $A \subset E$. A set $A_0 \subset E$ is called ε-net of the set A if

$$A_0 + B(0,\varepsilon) \supset A.$$

Here, + denotes the Minkowski sum of sets. Let

$$N_e(\varepsilon, A) = \inf\{|A_0| : A_0 \text{ is an } \varepsilon\text{-net of the set } A\}, \quad \varepsilon > 0,$$

be the *covering number* of A, where $|A_0|$ denotes the cardinality of the set A_0. For precompact subsets $A \subset E$, the function

$$\mathbb{R}_+ \to \mathbb{N}_0,\ \varepsilon \mapsto \log N_e(\varepsilon, A)$$

is called *metric entropy of* A. Note that the precompactness property implies that $\log N_e(\varepsilon, A)$ is finite for any $\varepsilon > 0$. Metric entropy represents a measure of the complexity of a set. High metric entropy means that much information is needed to describe an element with high accuracy.

First studies of metric entropy concerned its asymptotic behavior for certain regular (e.g. Hölder-continuous) precompact subsets of $C[0,1]$ equipped with the supremum norm (see for instance Kolmogorov's review on the theory of information transmission [35] and [38]). The research on entropy numbers continues and became a part of the more general approximation theory. For an exposition on this topic, one may consult the book by Carl and Stephani [12].

A first relation between entropy numbers and small ball probabilities was discovered by Kuelbs and Li in 1993 [40] and sharpened by Li and Linde [46]. They found

Theorem 3.2.1. *Let $l : \mathbb{R}_+ \to \mathbb{R}_+$ be slowly varying such that for all $s > 0$,*

$$l(x) \approx l(x^s) \quad (x \to \infty).$$

Then, for $a > 0$,

$$\varphi(\varepsilon) \approx \varepsilon^{-a}\, l(1/\varepsilon)$$

if and only if

$$\log N_e(\varepsilon, \mathcal{K}) \approx \varepsilon^{-2a/(2+a)}\, l(1/\varepsilon)^{2/(2+a)}.$$

Here $\mathcal{K}$ denotes the closed unit ball of the Cameron-Martin space of μ.

Other relationships as presented above, are known to exist between a number of approximation quantities like average Kolmogorov width and many more. Such correspondences have been the topic of the recent dissertation by Creutzig [14].

Parts of the proof of the previous theorem rely on the same properties of Gaussian measures that were applied in the dissertation by Fehringer [23] and in the article by Dereich et al. [19] to obtain the upper bound for the quantization error. In fact, we will use Lemma 1 of [40] in the following considerations:

Lemma 3.2.2. *For $\lambda, \varepsilon > 0$, one has*

$$\log N_e(2\varepsilon, \lambda\mathcal{K}) \le \varphi(\varepsilon) + \frac{\lambda^2}{2},$$

where $\mathcal{K}$ denotes the closed unit ball in H_μ.

Denote by Φ the distribution function of a standard normal random variable and let $\Psi = 1 - \Phi$. Recall that Ψ satisfies the basic inequality

$$\Psi(t) \le \frac{1}{2}\, e^{-t^2/2}$$

for $t \ge 0$ (see for instance Ledoux and Talagrand [44], p. 57).

Theorem 3.2.3. *Let μ be a Gaussian measure with infinite dimensional support. There exists a function $c : \mathbb{R}_+ \to \mathbb{R}_+$, $\varepsilon \mapsto c_\varepsilon$ satisfying*

$$\lim_{\varepsilon \downarrow 0} \frac{c_\varepsilon^2}{\varphi(\varepsilon)} = 0 \quad \textit{and} \tag{3.5}$$

$$\lim_{\varepsilon \downarrow 0} \frac{\log(1/\varepsilon)}{c_\varepsilon^2} = 0. \tag{3.6}$$

Set $\lambda_\varepsilon := c_\varepsilon - \Phi^{-1}(\mu(B(0,\varepsilon)))$, $\varepsilon > 0$. For any family of codebooks $\{\mathcal{C}_\varepsilon\}_{\varepsilon>0}$, in E satisfying

$$\mathcal{C}_\varepsilon + B(0, 2\varepsilon) \supset \lambda_\varepsilon \mathcal{K}, \tag{3.7}$$

one has, for any $s > 0$,

$$\mathbb{E}[d(X, \mathcal{C}_\varepsilon)^s]^{1/s} \lesssim 3\,\varepsilon \quad (\varepsilon \downarrow 0).$$

Furthermore, there exist codebooks $\mathcal{C}_\varepsilon$, $\varepsilon > 0$, satisfying (3.7) *with*

$$\log |\mathcal{C}_\varepsilon| \le \varphi(\varepsilon) + \lambda_\varepsilon^2/2 \lesssim 2\,\varphi(\varepsilon) \quad (\varepsilon \downarrow 0). \tag{3.8}$$

Remark 3.2.4. Suppose the small ball function φ satisfies $\varphi^{-1}(r) \approx \varphi^{-1}(2r)$ as $r \to \infty$. Then the above theorem yields the upper bound

$$D^{(q)}(r|\mu, \|\cdot\|^s)^{1/s} \lesssim 3\,\varphi^{-1}(r/2) \quad (r \to \infty).$$

Note, that the estimate is slightly weaker than the one obtained in Theorem 3.1.2,

$$D^{(q)}(r|\mu, \|\cdot\|^s)^{1/s} \lesssim 2\,\varphi^{-1}(r/2) \quad (r \to \infty).$$

However, quantization with ε-nets yields quantization errors of the optimal order whenever condition (3.1) is satisfied.

For the proof of Theorem 3.2.3 we need

Proposition 3.2.5. *If μ has infinite dimensional support, then for any $\kappa > 0$, there exists $\varepsilon_0(\kappa)$ such that*

$$\varphi(\varepsilon) \ge \kappa \log(1/\varepsilon)$$

for all $\varepsilon \in (0, \varepsilon_0(\kappa))$.

Proof. We apply Lemma 2.2.6. Let $\{X_j\}_{j\in\mathbb{N}}$ denote a sequence of independent standard normal random variables and let $\{e_j\}_{j\in\mathbb{N}}$ denote a complete orthonormal system in H_μ. Then $X := \sum_{j\in\mathbb{N}} X_j\, e_j$ converges a.s.

and one has $\mathcal{L}(X) = \mu$. Let $n \in \mathbb{N}$ and $A_n = \operatorname{span}(e_1, \dots, e_n)$ denote a n-dimensional subspace of the reproducing kernel Hilbert space H_μ. Then

$$\mathbb{P}\big(X \in B(0,\varepsilon)\big) = \mathbb{E}\Big[\mathbb{P}\Big(\sum_{j=1}^{n} X_j e_j \in B\big(-\sum_{j=n+1}^{\infty} X_j e_j, \varepsilon\big)\big|\{X_j\}_{j\geq n+1}\Big)\Big].$$

Note that $\sum_{j=1}^{n} X_j e_j$ is a centered Gaussian measure in E and Anderson's inequality implies

$$\mathbb{P}\big(X \in B(0,\varepsilon)\big) \leq \mathbb{P}\Big(\sum_{j=1}^{n} X_j e_j \in B(0,\varepsilon)\Big)$$

Denote $\pi : \mathbb{R}^n \to E$, $(x_1, \dots, x_n) \mapsto \sum_{j=1}^{n} x_j e_j$ and let f_n be the density of the n-dimensional standard normal distribution. Then, for $\varepsilon > 0$,

$$\mathbb{P}\big(\sum_{j=1}^{n} X_j e_j \in B(0,\varepsilon)\big) = \int_{\mathbb{R}^n} 1_{\pi^{-1}(B(0,\varepsilon))}(z) f_n(z) dz.$$

Since f_n is uniformly bounded by $(2\pi)^{-n/2}$ one obtains

$$\mu(B(0,\varepsilon)) \leq (2\pi)^{-n/2}\, \lambda^n(\pi^{-1}(B(0,\varepsilon))) = (2\pi)^{-n/2}\, \lambda^n(\pi^{-1}(B(0,1)))\, \varepsilon^n, \tag{3.9}$$

where λ^n denotes n-dimensional Lebesgue measure. It remains to show the finiteness of $\lambda^n(\pi^{-1}(B(0,1)))$. Since π is linear and one-to-one, the map

$$A_n \to [0,\infty),\ x \mapsto |\pi^{-1}(x)|$$

defines a norm on A_n. On A_n, this norm is equivalent to the norm $\|\cdot\|$. Therefore, there exists $c > 0$ such that

$$|\pi^{-1}(x)| \leq c\,\|x\|$$

for all $x \in A_n$. In particular, $\pi^{-1}(B(0,1)) \subset B_{\mathbb{R}^n}(0,c)$. The statement follows from equation (3.9), since $n \in \mathbb{N}$ was arbitrary. □

Proof of Theorem 3.2.3. According to Proposition 3.2.5, there exists a function $c : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying (3.5) and (3.6).

Let $\{c_\varepsilon\}_{\varepsilon>0}$ and $\{\mathcal{C}_\varepsilon\}_{\varepsilon>0}$ fulfill the assumptions of the theorem. Let $\varepsilon > 0$ and consider

$$A_\varepsilon = B(0,\varepsilon) + \lambda_\varepsilon \mathcal{K}$$

where $\lambda_\varepsilon := c_\varepsilon - \Phi^{-1}(\mu(B(0,\varepsilon)))$. We estimate

$$\mathbb{E}[d(X,\mathcal{C}_\varepsilon)^s] \leq \mathbb{E}[1_{A_\varepsilon}(X) d(X,\mathcal{C}_\varepsilon)^s] + \mathbb{E}[1_{A_\varepsilon^c}(X) d(X,\mathcal{C}_\varepsilon)^s] =: I_1(\varepsilon) + I_2(\varepsilon). \tag{3.10}$$

By assumption (3.7), it follows

$$A_\varepsilon = B(0,\varepsilon) + \lambda_\varepsilon \mathcal{K} \subset \mathcal{C}_\varepsilon + B(0,3\varepsilon).$$

Consequently, $I_1(\varepsilon) \le (3\varepsilon)^s$. Applying the Cauchy-Schwarz inequality to $I_2(\varepsilon)$ yields

$$I_2(\varepsilon) \le \mu(A_\varepsilon^c)^{1/2}\,\mathbb{E}\big[d(X,\mathcal{C}_\varepsilon)^{2s}\big]^{1/2}. \tag{3.11}$$

Since $d(x,\mathcal{C}_\varepsilon) \le \|x\| + \min_{z\in\mathcal{C}_\varepsilon}\|z\| \le \|x\| + 2\varepsilon$ for $x \in E$, one has

$$\mathbb{E}\big[d(X,\mathcal{C}_\varepsilon)^{2s}\big]^{1/2} \le \mathbb{E}\big[(\|X\|+2)^{2s}\big]^{1/2} =: \kappa_s < \infty$$

for all $\varepsilon \in (0,1]$. Hence,

$$I_2(\varepsilon) \le \kappa_s\,\mu(A_\varepsilon^c)^{1/2}, \quad \varepsilon \in (0,1].$$

On the other hand, by Borell's inequality, (see Lemma 2.3.1)

$$\mu(A_\varepsilon^c) \le \Psi\big(\Phi^{-1}(\mu(B(0,\varepsilon))) + c_\varepsilon - \Phi^{-1}(\mu(B(0,\varepsilon)))\big) = \Psi(c_\varepsilon)$$

and we obtain with (3.11)

$$I_2(\varepsilon) \le \kappa_s\,\Psi(c_\varepsilon)^{1/2} \le \kappa_s\,\exp\{-\frac{c_\varepsilon^2}{4}\},$$

since $c_\varepsilon \ge 0$. Using assumption (3.6), one has for any $\eta > 0$,

$$I_2(\varepsilon) \le \kappa_s\,e^{-c_\varepsilon^2/4} \lesssim \kappa_s\,\varepsilon^{\eta/4} \quad (\varepsilon \downarrow 0).$$

Since $\eta > 0$ is arbitrary, $I_2(\varepsilon) = o(\varepsilon^s)$ as $\varepsilon \downarrow 0$. The estimates for I_1 and I_2 combined with (3.10) finally yield

$$\mathbb{E}\big[d(X,\mathcal{C}_\varepsilon)^s\big]^{1/s} \lesssim 3\,\varepsilon \quad (\varepsilon \downarrow 0).$$

It remains to be shown that there exists a family of codebooks $\{\mathcal{C}_\varepsilon\}_{\varepsilon>0}$ that satisfies the assumptions (3.7) and (3.8). By Lemma 3.2.2, we find codebooks $\mathcal{C}_\varepsilon$, $\varepsilon > 0$, satisfying (3.7) with

$$\log|\mathcal{C}_\varepsilon| \le \varphi(\varepsilon) + \lambda_\varepsilon^2/2 \tag{3.12}$$

for $\varepsilon > 0$. Since $\Psi(t) \le \exp\{-t^2/2\}$, $t \ge 0$, one has, for $x \le 1/2$,

$$x \le \exp\{-\Psi^{-1}(x)^2/2\}.$$

Consequently,

$$\sqrt{-2\log x} \ge \Psi^{-1}(x).$$

Therefore, for $\varepsilon > 0$ with $\mu(B(0,\varepsilon)) \le 1/2$, one obtains

$$\begin{aligned}\lambda_\varepsilon^2 &= \left[c_\varepsilon + \Psi^{-1}(\mu(B(0,\varepsilon)))\right]^2\\ &\le \left[c_\varepsilon + \sqrt{-2\log\mu(B(0,\varepsilon))}\right]^2\\ &= 2\varphi(\varepsilon) + 2c_\varepsilon\sqrt{2\varphi(\varepsilon)} + c_\varepsilon^2 \sim 2\,\varphi(\varepsilon),\end{aligned}$$

where the last equivalence follows from equation (3.5). Hence, with equation (3.12),

$$\log|\mathcal{C}_\varepsilon| \lesssim 2\,\varphi(\varepsilon) \quad (\varepsilon \downarrow 0).$$

□

3.3 A lower bound for the distortion rate function

Let (μ,ρ) be an information source on a Polish space E. We will use the source coding theorem given in Section 1.5. Recall that for $r \ge 0$

$$\lim_{k\to\infty} D_k(r|\mu,\rho) = \inf_{k\in\mathbb{N}} D_k(r|\mu,\rho) = D(r|\mu,\rho),$$

if there exists $y \in E$ with $\int \rho(x,y)\,d\mu(x) < \infty$.

We estimate the distortion rate function against some term, which measures the local mass concentration. For $y \in E$ and $t \ge 0$, set

$$B_\rho(y,t) = \{x \in E : \rho(x,y) \le t\}$$

and consider

$$F(t) = \sup_{y\in E} \mu(B_\rho(y,t)).$$

It follows that $F(\cdot)$ is non-negative, monotonically increasing and converges to 1 as $t \to \infty$. We denote by F_+ the right continuous version of F and let $\nu \in \mathcal{M}_1[0,\infty)$ be the unique probability measure satisfying $\nu[0,t] = F_+(t)$, $t \ge 0$. Moreover, we denote by Z a ν-distributed random variable.

Theorem 3.3.1. *Assume that the information source (μ,ρ) admits an element $y \in E$ with $\int \rho(x,y)\,d\mu(x) < \infty$. Let Λ_Z be the logarithmic moment generating function of ν, i.e.*

$$\Lambda_Z(\theta) = \log \int e^{\theta y}\,d\nu(y), \quad \theta \in \mathbb{R},$$

and denote by $\Lambda_Z^*(t) = \sup_{\theta\le 0}[\theta t - \Lambda(\theta)]$, $t \ge 0$ its Legendre transform. Then one has for $r \ge 0$

$$D(r|\mu,\rho) \ge \sup\{t \ge 0 : \Lambda_Z^*(t) > r\}, \tag{3.13}$$

where the supremum of the empty set is assumed to be 0.

Remark 3.3.2. The definition $\Lambda_Z^*(t) = \sup_{\theta \le 0}[\theta t - \Lambda_Z(\theta)]$, $t \ge 0$, differs slightly from the standard definition where one considers $\sup_{\theta \in \mathbb{R}}[\theta t - \Lambda_Z(\theta)]$. However, both notions are closely related:

Note that Λ_Z is differentiable on $(-\infty, 0)$ with

$$\Lambda_Z'(\theta) = \frac{\mathbb{E}[Ze^{\theta Z}]}{\mathbb{E}[e^{\theta Z}]}, \quad \theta < 0.$$

By monotone convergence,

$$\lim_{\theta \nearrow 0} \Lambda_Z'(\theta) = \mathbb{E}Z.$$

Moreover, $\Lambda_Z(0) = 0$. Due to the convexity of the moment generating function (see for instance Dembo and Zeitouni, 1998, Lemma 2.2.5), one has

$$\Lambda_Z(\theta) \ge \theta\, \mathbb{E}[Z]$$

for all $\theta \in \mathbb{R}$. Hence, for any $\theta \ge 0$ and $t \le \mathbb{E}[Z]$,

$$\theta t - \Lambda_Z(\theta) \le \theta(t - \mathbb{E}[Z]) \le 0.$$

Since $\sup_{\theta \in \mathbb{R}}[\theta t - \Lambda_Z(\theta)]$ is non-negative for all $t \in \mathbb{R}$, we conclude that, for $t \le \mathbb{E}[Z]$,

$$\Lambda_Z^*(t) = \sup_{\theta \le 0}[\theta t - \Lambda_Z(\theta)] = \sup_{\theta \in \mathbb{R}}[\theta t - \Lambda_Z(\theta)].$$

On the other hand, it holds for $t \ge \mathbb{E}[Z]$ and $\theta \le 0$

$$\theta t - \Lambda_Z(\theta) \le \theta(t - \mathbb{E}Z) \le 0.$$

Therefore,

$$\Lambda_Z^*(t) = 0$$

for $t \ge \mathbb{E}[Z]$.

As in the section on the SCT (Section 1.5), we use the single letter distortion measures related to ρ. Denote

$$\rho_k(x, y) := \frac{1}{k} \sum_{i=1}^{k} \rho(x_i, y_i)$$

for $k \in \mathbb{N}$ and $x, y \in E^k$.

Proof of Theorem 3.3.1. Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of independent μ-distributed random elements and write $X^{(k)} := (X_1, \dots, X_k)$, $k \in \mathbb{N}$. Let

$\mathcal{C}^k$ denote a codebook in E^k of rate $kr \geq 0$, i.e. $|\mathcal{C}^k| \leq e^{kr}$ and $\mathcal{C}^k \subset E^k$. Then

$$\begin{aligned}\mathbb{E}\big[\rho_k(X^{(k)}, \mathcal{C}^k)\big] &= \int_0^\infty \mathbb{P}\big\{\rho_k(X^{(k)}, \mathcal{C}^k) > t\big\}\,dt \\ &= \int_0^\infty \Big(1 - \mathbb{P}\big\{\rho_k(X^{(k)}, \mathcal{C}^k) \leq t\big\}\Big)\,dt. \qquad (3.14)\end{aligned}$$

Estimate

$$\mathbb{P}\big(\rho_k(X^{(k)}, \mathcal{C}^k) \leq t\big) = \mu^{\otimes k}\big(\bigcup_{y\in\mathcal{C}^k} B_{\rho_k}(y,t)\big) \leq \sum_{y\in\mathcal{C}^k} \mu^{\otimes k}\big(B_{\rho_k}(y,t)\big). \qquad (3.15)$$

By the definition of F_+, it holds $\mu\big(B_\rho(y_1,t)\big) \leq F_+(t)$ for every $y_1 \in E$. Therefore, for fixed $y_1 \in E$, one can define a ν-distributed random variable Z_1, which is coupled with $\rho(X_1, y_1)$ in such a way that

$$Z_1 \leq \rho(X_1, y_1).$$

Hence, when denoting by $\{Z_n\}_{n\in\mathbb{N}}$ an independent sequence of ν-distributed random variables, one has

$$\mathbb{P}\Big(\frac{1}{k}\sum_{i=1}^k Z_i \leq t\Big) \geq \mu^{\otimes k}\big(B_{\rho_k}(y,t)\big)$$

for all $y \in E^k$. Set $\bar{Z}_k := \sum_{i=1}^k Z_i/k$, $k \in \mathbb{N}$. With equation (3.15), one obtains

$$\mathbb{P}\big(\rho_k(X^{(k)}, \mathcal{C}^k) \leq t\big) \leq |\mathcal{C}^k|\,\mathbb{P}(\bar{Z}_k \leq t) \leq \exp\Big[k\Big(r + \frac{\log \mathbb{P}(\bar{Z}_k \leq t)}{k}\Big)\Big].$$

Due to (3.14), it follows

$$\mathbb{E}\big[\rho_k(X^{(k)}, \mathcal{C}^k)\big] \geq \int_0^\infty \Big(1 - \exp\Big[k\big(r + \frac{\log \mathbb{P}(\bar{Z}_k \leq t)}{k}\big)\Big]\Big)^+ dt,$$

where we denote $(\cdot)^+ = \cdot \vee 0$. The previous estimate holds uniformly for all codebooks of rate r and, therefore,

$$D_k(r) \geq \int_0^\infty \Big(1 - \exp\Big[k\big(r + \frac{\log \mathbb{P}(\bar{Z}_k \leq t)}{k}\big)\Big]\Big)^+ dt.$$

Note that for any $\theta \leq 0$ and $k \in \mathbb{N}$

$$\Lambda_{\bar{Z}_k}(\theta k) := \log \mathbb{E} e^{\theta k \bar{Z}_k} = k\Lambda_Z(\theta)$$

and by the exponentiel Chebychev inequality it follows

$$\log \mathbb{P}(\bar{Z}_k \leq t) \leq \Lambda_{\bar{Z}_k}(\theta k) - \theta k t = k(\Lambda_Z(\theta) - \theta t).$$

Therefore

$$\log \mathbb{P}(\bar{Z}_k \leq t) \leq -k\,\Lambda_Z^*.$$

Consequently,

$$D_k(r) \geq \int_0^\infty \Big(1 - \exp\Big[k\big(r - \Lambda_Z^*(t)\big)\Big]\Big)^+ \,dt.$$

For $t_0 \geq 0$ with $\Lambda_Z^*(t_0) > r$ it follows

$$D_k(r) \geq t_0\Big(1 - \exp\Big[k\big(r - \Lambda_Z^*(t_0)\big)\Big]\Big)^+ \to t_0$$

as $r \to \infty$. By the SCT (Theorem 1.5.1), one obtains

$$D(r) = \lim_{k\to\infty} D_k(r) \geq t_0.$$

□

Remark 3.3.3. In this remark, we outline an alternative proof of the previous theorem. This alternative version uses basic properties of the AEP. The notations of Section 1.5 are adopted. Consider the rate distortion function $R(d|\mu,\rho)$. Recall that

$$R(d|\mu,\rho) = \inf_{\nu\in\mathcal{M}_1(E)} R_0(\mu,\nu,d).$$

For $\varepsilon > 0$, let $\nu \in \mathcal{M}_1(E)$ be such that $R_0(\mu,\nu,d) \leq R(d|\mu,\rho)+\varepsilon$. According to Proposition 2 of Dembo and Kontoyiannis [15], one has

$$R_0(\mu,\nu,d) = \Lambda^*(d|\nu),$$

where $\Lambda^*(d|\nu) = \sup_{\theta\leq 0}[\theta d - \Lambda(\theta|\nu)]$ and

$$\Lambda(\theta|\nu) := \int \log\Big(\int e^{\theta\rho(x,y)}\,d\nu(y)\Big)\,d\mu(x), \quad \theta \leq 0.$$

Due to Jensen's inequality, one has for $\theta \leq 0$

$$\begin{aligned}
\Lambda(\theta|\nu) &\leq \log\Big(\int\int e^{\theta\rho(x,y)}\,d\nu(y)\,d\mu(x)\Big)\\
&= \log\Big(\int\int e^{\theta\rho(x,y)}\,d\mu(x)\,d\nu(y)\Big)\\
&\leq \sup_{y\in E}\log\Big(\int e^{\theta\rho(x,y)}\,d\mu(x)\Big) \leq \Lambda_Z(\theta),
\end{aligned}$$

where $\Lambda_Z(\theta)$ is as above (see Theorem 3.3.1). Hence,

$$R_0(\mu,\nu,d) = \Lambda^*(d|\nu) \geq \Lambda_Z^*(d).$$

Since $\varepsilon > 0$ was arbitrary, one has

$$R(d|\mu,\rho) \geq \Lambda_Z^*(d).$$

Since $R(\,\cdot\,|\mu,\rho)$ is essentially the inverse of the DRF $D(\,\cdot\,|\mu,\rho)$, the statement of the theorem can be deduced.

3.4 SBPs and moment generating functions

Let throughout the rest of this section μ denote a centered Gaussian measure on a separable Banach space $(E, \|\cdot\|)$. We denote by X a μ-distributed r.e. on E.

In the previous section, we established a lower bound for the DRF that is based on an inverse of a particular Legendre transform. In order to apply the result to Gaussian random elements, we derive a relation between this Legendre transform and SBPs. The link will be useful in several parts of this thesis.

Let $x \in E$ and let $\tau : [0,\infty) \to [0,\infty)$ be a Young function, i.e. τ is convex, one-to-one and satisfies $\tau(0) = 0$. Denote, for $t \geq 0$ and $\theta \in \mathbb{R}$,

$$\begin{aligned} Z &= \tau(\|X - x\|), \\ F(t) &= \mathbb{P}(Z \leq t), \\ \Lambda_Z(\theta) &= \log \mathbb{E}[e^{\theta Z}], \\ \Lambda_Z^*(t) &= \sup_{\theta \leq 0}[\theta t - \Lambda_Z(\theta)]. \end{aligned}$$

Theorem 3.4.1. *For $0 < t < 1/2$, set*

$$h(t) = \frac{2\log(1/t)}{(\Phi^{-1}(t))^2}$$

and $h(0) = \lim_{t\downarrow 0} h(t) = 1$, where $\Phi(t) = (2\pi)^{-1/2}\int_{-\infty}^t e^{-u^2/2}du$, $t \in \mathbb{R}$. Then

$$\Lambda_Z^*(t) \leq -\log F(t) \leq h(F(t))\, \Lambda_Z^*(t)$$

for all $t > 0$ with $F(t) < 1/2$.

Proof. For every $\theta \leq 0$ and $t > 0$, one has by the Markov inequality

$$\Lambda_Z(\theta) = \log \mathbb{E}[e^{\theta Z}] \geq \theta x + \log \mathbb{P}(Z \leq t).$$

Therefore, for any $t > 0$,

$$\Lambda_Z^*(t) = \sup_{\theta \leq 0}[t\theta - \Lambda_Z(\theta)] \leq -\log \mathbb{P}(Z \leq t).$$

We proceed with the proof of the second inequality. Suppose first that $t_0 > 0$ is such that $\mathbb{P}(Z \leq t_0) = 0$ and fix $p \in (0,1)$ arbitrarily. Then there exists $\varepsilon > 0$ such that $\mathbb{P}(Z < t_0 + \varepsilon) \leq p$. Consequently, for $\theta \leq 0$,

$$\Lambda_Z(\theta) = \log \mathbb{E}e^{\theta Z} \leq \log[pe^{\theta t_0} + (1-p)e^{\theta(t_0+\varepsilon)}]$$

and

$$\Lambda^*(t_0) \geq \limsup_{\theta \to -\infty} [\theta t_0 - \Lambda_Z(\theta)] \geq -\log p.$$

Since $p \in (0,1)$ was arbitrary, it follows that $\Lambda^*(t_0) = \infty$.

Now let $t_0 > 0$ with $\mathbb{P}(Z \leq t_0) \in (0, 1/2)$. In order to show the second inequality, we let $G(t) = \mathbb{P}(\|X - x\| \leq t)$, $t > 0$, and consider the function

$$f := \Phi^{-1} \circ F = \Phi^{-1} \circ G \circ \tau^{-1} : (0, \infty) \to \mathbb{R} \cup \{-\infty\}.$$

Recall that $\Phi^{-1} \circ G$ is concave as a consequence of the Ehrhard inequality (see Corollary 2.3.4). Moreover, since τ^{-1} is concave and $\Phi^{-1} \circ G$ is monotonically increasing, it follows that f is concave.

Let now q denote a tangent of the graph of f at the point $(t_0, f(t_0))$. Represent q in the form $q(t) = (t - m)s$, where $m, s > 0$ are appropriate constants. Let N denote a standard normal r.v. and associate the random variable $Z_q = q^{-1}(N) = N/s + m$ with q. Z_q has distribution function $\Phi \circ q$ and, hence, it is a normal r.v. on $\mathbb{R}$. Note, that $F(t) = \Phi \circ f(t)$. Consequently, the distribution function of the r.v. $f^{-1}(N)$ equals F. We assume without loss of generality that $Z = f^{-1}(N)$. Since q is a tangent of the concave function f, one has $q \geq f$. Therefore, $Z_q \leq Z$ and one has for every $\theta \leq 0$

$$\Lambda_{Z_q}(\theta) = \log \mathbb{E}[e^{\theta Z_q}] \geq \log \mathbb{E}[e^{\theta Z}] = \Lambda_Z(\theta).$$

Consequently,

$$\Lambda_Z^*(t) \geq \sup_{\theta \leq 0} [t\theta - \Lambda_{Z_q}(\theta)] =: \Lambda_{Z_q}^*(t)$$

for every $t > 0$.

On the other hand, one has $\Lambda_{Z_q}(\theta) = (\theta/s)^2/2 + m\theta$ and, for $t \in (0, m]$,

$$\begin{aligned}
\Lambda_{Z_q}^*(t) &= \sup_{\theta \leq 0} [\theta t - \Lambda_{Z_q}(\theta)] \\
&= \sup_{\theta \leq 0} \Big[-\frac{1}{2s^2} (\theta + s^2(m - t))^2 + \frac{s^2(m-t)^2}{2} \Big] \\
&= \frac{s^2(m-t)^2}{2} = \frac{q(t)^2}{2}.
\end{aligned}$$

Note that $t_0 < m$ and one has

$$\Lambda_{Z_q}^*(t_0) = \frac{q(t_0)^2}{2} = \frac{f(t_0)^2}{2} = \frac{(\Phi^{-1}(F(x_0)))^2}{2}.$$

Hence,

$$\frac{-\log \mathbb{P}(Z \leq t_0)}{\Lambda_Z^*(t_0)} \leq \frac{2 \log(1/F(t_0))}{(\Phi^{-1}(F(t_0)))^2} = h(F(t_0)).$$

The convergence $\lim_{t \downarrow 0} h(t) = 1$ is established in the lemma below. □

Lemma 3.4.2.

$$h(\varepsilon) = 1 + \frac{\log\log(1/\varepsilon)}{2\log(1/\varepsilon)} + o\Big(\frac{\log\log(1/\varepsilon)}{\log(1/\varepsilon)}\Big) \quad (\varepsilon \downarrow 0),$$

where o denotes the Landau symbol.

Proof. Consider the functions

$$g : (0,\infty) \to (0,1),\ t \mapsto e^{-t/2} \quad \text{and}$$
$$\tilde g : (-\infty,0) \to (0,1),\ t \mapsto e^{-t^2/2}.$$

Both functions are one-to-one and possess inverse functions g^{-1} and $\tilde g^{-1}$. Denote $\tilde\Phi = \Phi|_{(-\infty,0)} : (-\infty,0) \to (0,1/2)$. Then

$$\big(\tilde\Phi^{-1}\big)^2 = g^{-1} \circ g \circ \big(\tilde\Phi^{-1}\big)^2 = g^{-1} \circ \tilde g \circ \tilde\Phi^{-1} = g^{-1} \circ \big(\tilde\Phi \circ \tilde g^{-1}\big)^{-1}. \tag{3.16}$$

Since $\Phi(t) \sim (2\pi)^{-1/2} e^{-t^2/2}/(-t)$ as $t \to -\infty$, one has,

$$\tilde\Phi \circ \tilde g^{-1}(\varepsilon) \sim \frac{\varepsilon}{\sqrt{2\pi \log(1/\varepsilon^2)}} \quad (\varepsilon \downarrow 0).$$

The latter function is regularly varying and its inverse is asymptotically equivalent to (see Bingham et al. [6], p. 28)

$$\big(\tilde\Phi \circ \tilde g^{-1}\big)^{-1}(\varepsilon) \sim \sqrt{2\pi\log(1/\varepsilon^2)}\,\varepsilon \quad (\varepsilon \downarrow 0).$$

Due to equation (3.16), it holds

$$\lim_{\varepsilon\downarrow 0} (\tilde\Phi^{-1}(\varepsilon))^2 + 2\log(\sqrt{2\pi\log(1/\varepsilon^2)}\,\varepsilon) = 0.$$

Denote $\eta(\varepsilon) = (\tilde\Phi^{-1}(\varepsilon))^2 - [2\log(1/\varepsilon) - \log(\log(1/\varepsilon)) - \log(4\pi)]$, $\varepsilon > 0$. Then

$$h(\varepsilon) = 1 + \frac{\log(\log(1/\varepsilon)) + \log(4\pi) - \eta(\varepsilon)}{2\log(1/\varepsilon) - \log(\log(1/\varepsilon)) - \log(4\pi) + \eta(\varepsilon)}$$

and the proof is finished. □

Basic analysis applied to the previous theorem and lemma gives

Corollary 3.4.3. *For $\eta > 0$, there exists a constant $r_0 = r_0(\eta) \ge 0$ such that the following holds: Let X be an arbitrary Gaussian random element on an arbitrary separable Banach space $(E, \|\cdot\|)$, and let $x \in E$, $\varepsilon > 0$ and $s \ge 1$. One has*

$$-\log \mathbb{P}(\|X - x\|^s \le \varepsilon) \le \Big[\Lambda_x^*(\varepsilon) + \frac{1+\eta}{2}\log\Lambda_x^*(\varepsilon)\Big] \vee r_0,$$

where $\Lambda_x(\theta) = \log\mathbb{E}[e^{\theta\|X-x\|^s}]$, $\theta \le 0$, and $\Lambda_x^*(\varepsilon) = \sup_{\theta\le 0}[\varepsilon\theta - \Lambda_x(\theta)]$, $\varepsilon > 0$.

3.5 A lower bound for the DRF for Gaussian measures

Let μ denote a centered Gaussian measure on a separable Banach space $(E, \|\cdot\|)$. For fixed $s \geq 1$, we consider the information source $(\mu, \|\cdot\|^s)$. In this section, we combine the previous results to obtain an asymptotic lower bound for the DRF $D(r|\mu, \|\cdot\|^s)$. Again the bound uses the small ball function

$$\varphi(\varepsilon) = -\log \mu(B(0, \varepsilon)), \quad \varepsilon > 0.$$

Theorem 3.5.1. *For every $\varepsilon > 0$ and $s \geq 1$, there exists $r_0 \geq 1$ such that*

$$D(r|\mu, \|\cdot\|^s)^{1/s} \geq \varphi^{-1}\Big(r + \frac{1+\varepsilon}{2} \log r\Big)$$

for all $r \geq r_0$.

Proof. Let $\varepsilon > 0$ be arbitrary and let $\rho(x, y) = \|x - y\|^s$, $x, y \in E$. By the Anderson inequality (Lemma 2.2.2), the mass of centered Gaussian measures is concentrated around zero, i.e.

$$F(t) := \sup_{y \in E} \mu(B_\rho(y, t)) = \mu(B_\rho(0, t)) = \mu(B(0, t^{1/s})), \quad t > 0.$$

For a μ-distributed r.e. X, consider $Z := \|X\|^s$. Then, $\mathbb{P}(Z \leq t) = F(t)$ for all $t > 0$. By Theorem 3.3.1, one has for $r > 0$,

$$\begin{aligned} D(r|\mu, \rho) &\geq \sup\{t \geq 0 : \Lambda_Z^*(t) > r\} \\ &= \sup\Big\{t \geq 0 : \Lambda_Z^*(t) + \frac{1+\varepsilon}{2} \log \Lambda_Z^*(t) > r + \frac{1+\varepsilon}{2} \log r\Big\}. \end{aligned}$$

By Corollary 3.4.3, there exists $r_0 \geq 1$ such that

$$-\log F(t) \leq \Big[\Lambda_Z^*(t) + \frac{1+\varepsilon}{2} \log \Lambda_Z^*(t)\Big] \vee r_0$$

for all $t > 0$. Hence, for $r \geq r_0$,

$$D(r|\mu, \rho) \geq \sup\Big\{t \geq 0 : -\log F(t) > r + \frac{1+\varepsilon}{2} \log r\Big\}.$$

Consider

$$\tilde{\varphi} := -\log \circ F : (0, \infty) \to (0, \infty).$$

The function $\tilde{\varphi}$ is related to the small ball function by $\tilde{\varphi}(t) = \varphi(t^{1/s})$, $t > 0$. Hence, it has inverse $\tilde{\varphi}^{-1}(x) = \varphi^{-1}(x)^s$, $x > 0$. Finally, we obtain that

$$D(r|\mu, \|\cdot\|^s) \geq \varphi^{-1}\Big(r + \frac{1+\varepsilon}{2} \log r\Big)^s.$$

□

Due to Lemma 3.1.4, the new lower bound is asymptotically equivalent to $\varphi^{-1}(r)$ if $\varphi^{-1}(r) \approx \varphi^{-1}(2r)$ $(r \to \infty)$. As a consequence of Theorems 3.1.2 and 3.5.1 one obtains

Theorem 3.5.2. *Suppose that $\varphi^{-1}(r) \approx \varphi^{-1}(2r)$ as $r \to \infty$. Then, for any $s \geq 1$,*

$$\varphi^{-1}(r) \lesssim D(r|\mu, \|\cdot\|^s)^{1/s} \leq D^{(R)}(r|\mu, \|\cdot\|^s)^{1/s} \lesssim 2\,\varphi^{-1}(r/2)$$

as $r \to \infty$. In particular, all coding quantities are of the same weak asymptotic order.

3.6 Known asymptotics of small ball probabilities

Considerable effort was recently put on determining the asymptotic behavior of

$$\varphi(\varepsilon) = -\log \mu(B(0, \varepsilon)), \quad \varepsilon > 0,$$

as $\varepsilon \downarrow 0$ for centered Gaussian measures μ on Banach spaces E. Beside quantization, these results can be used to derive certain types of the law of the iterated logarithm and to get hold of certain metric entropies. An overview on the topic can be found in Li and Shao [47]. We summarize some results below.

Wiener measure

We consider the Wiener measure μ on various separable Banach spaces E:

- $E = C([0, 1], \mathbb{R}^d)$, equipped with a supremum norm

  $$\|f\| := \|f\|_{[0,1],G} = \sup_{t \in [0,1]} |f(t)|_G,$$

  where $|\cdot|_G$ is an arbitrary norm on $\mathbb{R}^d$. Owing to Ledoux [43],

  $$\mu(B(0, \varepsilon)) \sim e^{-\lambda_1/\varepsilon^2} f(0) \int_{-1}^{1} f(y)\, dy \quad (\varepsilon \downarrow 0),$$

  where λ_1 is the principal eigenvalue and f is the corresponding unit-norm (in $L_2(\mathbb{R}^d)$) eigenvector of the Dirichlet problem on the domain $\{x \in \mathbb{R}^d : |x|_G < 1\}$. In the case that μ is 1-dimensional Wiener measure on $E = C[0,1]$ equipped with the standard supremum norm $\|\cdot\|_{[0,1]}$, one has $\lambda_1 = \pi^2/8$. Therefore, it holds

$$\varphi(\varepsilon) \sim \frac{\pi^2}{8\varepsilon^2} \quad (\varepsilon \downarrow 0)$$

and, for $s \geq 1$,

$$\frac{\pi}{\sqrt{8r}} \lesssim D(r|\mu, \|\cdot\|^s_{[0,1]})^{1/s} \leq D^{(R)}(r|\mu, \|\cdot\|^s_{[0,1]})^{1/s} \lesssim \frac{\pi}{\sqrt{r}}.$$

as $r \to \infty$.

- $E = L_p[0,1]$, $p \geq 1$, equipped with the L_p-norm $\|\cdot\|_{L_p[0,1]}$. It is well known (see for instance Li and Shao [47]) that the small ball probabilities satisfy

$$\varphi(\varepsilon) \sim \frac{c_p}{\varepsilon^2},$$

where

$$c_p = 2^{2/p} p \Big(\frac{\lambda_1(p)}{2+p}\Big)^{(2+p)/p}$$

and

$$\lambda_1(p) = \inf\Big\{\int_{-\infty}^{\infty} |x|^p\, f(x)^2\, dx + \frac{1}{2}\int_{-\infty}^{\infty} f'(x)^2\, dx\Big\}$$

where the infimum is taken over all differentiable $f \in L_2(\mathbb{R})$ with unit-norm. Consequently, for $s \geq 1$,

$$\frac{\sqrt{c_p}}{\sqrt{r}} \lesssim D(r|\mu, \|\cdot\|^s_{L_p[0,1]})^{1/s} \leq D^{(R)}(r|\mu, \|\cdot\|^s_{L_p[0,1]})^{1/s} \lesssim \frac{\sqrt{8c_p}}{\sqrt{r}}$$

as $r \to \infty$. The small ball probabilities under the L_p-norm for general Gaussian Markov processes is treated in Li [45].

- $E = C_0^\alpha$, $\alpha \in (0,1/2)$, the space of α-Hölder continuous functions over the time $[0,1]$ starting in 0 equipped with the norm

$$\|f\|_{C^\alpha} := \sup_{0\leq s<t\leq 1} \frac{|f(t)-f(s)|}{t-s}.$$

Referring to Kuelbs and Li [41], there exists $c_\alpha > 0$ with

$$\varphi(\varepsilon) \sim \frac{c_\alpha}{\varepsilon^{2/(1-2\alpha)}}.$$

The constant c_α is not known explicitly although lower and upper bounds are derived in [41]. We obtain, for $s \geq 1$,

$$\frac{c_\alpha^{(1-2\alpha)/2}}{r^{(1-2\alpha)/2}} \lesssim D(r|\mu, \|\cdot\|^s_{C^\alpha})^{1/s}$$
$$\leq D^{(R)}(r|\mu, \|\cdot\|^s_{C^\alpha})^{1/s} \lesssim 2^{(3-2\alpha)/2}\, \frac{c_\alpha^{(1-2\alpha)/2}}{r^{(1-2\alpha)/2}}$$

as $r \to \infty$.

Gaussian sheets

Let $\gamma = (\gamma_1, \ldots, \gamma_d)$, $d \in \mathbb{N}$, $0 < \gamma_j < 2$, and denote by $X = \{X_t\}_{t\in[0,1]^d}$ the d-dimensional fractional Brownian sheet with parameter γ in $C([0,1]^d)$, i.e. X is a centered continuous Gaussian process on $[0,1]^d$ with covariance kernel

$$\mathbb{E}[X_t X_s] = \frac{1}{2^d} \prod_{j=1}^{d} \big[|t_j|^{\gamma_j} + |s_j|^{\gamma_j} - |t_j - s_j|^{\gamma_j}\big], \quad t, s \in [0,1]^d.$$

We consider X as Gaussian random element in the Banach space of continuous functions $C([0,1]^d)$ equipped with the supremum norm $\|\cdot\|_{[0,1]^d}$. The asymptotics of the small ball function

$$\varphi(\varepsilon) = -\log \mathbb{P}(\|X\|_{[0,1]^d} \leq \varepsilon), \quad \varepsilon > 0,$$

have been studied by many authors. If $d = 1$, the process is 1-dimensional fractional Brownian motion and the asymptotics of the SBPs are stated above. In the case that there is a unique minimum, say γ_1, in $\gamma = (\gamma_1, \ldots, \gamma_d)$, it was derived by Mason and Shi [52] that

$$\varphi(\varepsilon) \approx \varepsilon^{-2/\gamma_1} \quad (\varepsilon \downarrow 0).$$

Belinski and Linde [3] studied the case that there are exactly two minimal elements, say γ_1 and γ_2, in $\gamma = (\gamma_1, \ldots, \gamma_d)$. They found

$$\varphi(\varepsilon) \approx \varepsilon^{-2/\gamma_1} \big(\log(1/\varepsilon)\big)^{1+2/\gamma_1} \quad (\varepsilon \downarrow 0).$$

By this they extended a result of Talagrand [64], who had already solved the small ball problem for $\gamma = (1,1)$, i.e. in the case that X is a 2-dimensional Brownian sheet.

Chapter 4

Perturbation of the coding problem

In Chapter 3, we obtained bounds on the asymptotic coding error for certain Gaussian measures. In order to derive results for more general processes, we study the effect of small perturbations on the asymptotic coding problem.

From now on, we denote by μ a measure on the Borel sets of a Banach space $(E, \|\cdot\|)$. Moreover, we let X denote a μ distributed random element. Our aim is to study the impact on the asymptotic coding problem when "slightly" perturbing the rate and the law of the original.

Finally, we will provide an application of the perturbation results on the high resolution coding problem for diffusion processes.

4.1 Results for the DRF

Using the convexity of the DRF one obtains immediately a stability result for small perturbations in the rate. As a consequence of Lemma 3.1.4 one obtains

Lemma 4.1.1. *Let (μ, ρ) be an information source such that its DRF $D(r) = D(r|\mu, \rho)$ satisfies*

$$D(r) \approx D(2r) \quad (r \to \infty). \tag{4.1}$$

Then, for any function $\Delta : [0, \infty) \to \mathbb{R}$, $r \mapsto \Delta r$ with $\Delta r = o(r)$ as $r \to \infty$, it holds

$$D(r + \Delta r) \sim D(r) \text{ as } r \to \infty.$$

Now we consider perturbation in the distribution of the original. In the following we let X_1 and X_2 be two random elements in a Banach space $(E, \|\cdot\|)$.

Lemma 4.1.2. *Let $r_1, r_2 \geq 0$. Then, for $\kappa > 0$ and $s > 0$,*

$$D(r_1+r_2|X_1+X_2, \|\cdot\|^s) \leq (1+\kappa)^s D(r_1|X_1, \|\cdot\|^s)+(1+1/\kappa)^s D(r_2|X_2, \|\cdot\|^s).$$

The proof is based on basic properties of the mutual information and the conditional mutual information. Recall that these properties are summarized in Lemma 1.4.4.

Proof of Lemma 4.1.2. Let $\varepsilon > 0$. By the definition of $D(\cdot)$ there exist r.e.'s $\hat{X}_1, \hat{X}_2$ such that $I(X_i; \hat{X}_i) \leq r_i$ and

$$\mathbb{E}[\|X_i - \hat{X}_i\|^s] \leq D(r_i|X_i, \|\cdot\|^s) + \varepsilon \text{ for } i = 1, 2.$$

For $i = 1, 2$, denote by K_i a version of the probability kernel

$$K_i(x, A) = \mathbb{P}_{\hat{X}_i|X_i=x}(A), \quad x \in E, A \in \mathcal{B}(E),$$

and let $\mu = \mathbb{P} \circ (X_1, X_2)^{-1}$. Now let $(X_1, X_2, \tilde{X}_1, \tilde{X}_2)$ denote a random element in E^4 with

$$\begin{aligned} \mathbb{P} \circ (X_1, X_2, &\tilde{X}_1, \tilde{X}_2)^{-1}(d(x_1, x_2, \tilde{x}_1, \tilde{x}_2)) \\ &= \mu(d(x_1, x_2))K_1(x_1, d\tilde{x}_1)K_2(x_2, d\tilde{x}_2). \end{aligned} \tag{4.2}$$

We observe that $(\tilde{X}_1, X_1, X_2, \tilde{X}_2)$ forms a Markov process. By Lemma 1.4.4, one has

$$\begin{aligned} I((X_1, X_2); (\tilde{X}_1, \tilde{X}_2)) &= I((X_1, X_2); \tilde{X}_1) + I((X_1, X_2); \tilde{X}_2|\tilde{X}_1) \\ &\leq I((X_1, X_2); \tilde{X}_1) + I((X_1, X_2, \tilde{X}_1); \tilde{X}_2) \\ &= I(X_1; \tilde{X}_1) + I(X_2; \tilde{X}_1|X_1) \\ &\quad + I(X_2; \tilde{X}_2) + I((X_1, \tilde{X}_1); \tilde{X}_2|X_2) \end{aligned}$$

Since $(\tilde{X}_1, X_1, X_2, \tilde{X}_2)$ is a Markov process, it follows that the random elements X_2 and $\tilde{X}_1$ are conditionally independent given X_1. Therefore, $I(X_2; \tilde{X}_1|X_1) = 0$. Analogously, $I((X_1, \tilde{X}_1); \tilde{X}_2|X_2) = 0$. We conclude

$$I((X_1, X_2); (\tilde{X}_1, \tilde{X}_2)) \leq I(X_1; \tilde{X}_1) + I(X_2; \tilde{X}_2) \leq r_1 + r_2.$$

On the other hand, for any $\kappa > 0$,

$$\begin{aligned} \mathbb{E}[\|X_1 + X_2 - &(\tilde{X}_1 + \tilde{X}_2)\|^s] \\ &\leq \mathbb{E}[\max\{(1+\kappa)\|X_1 - \tilde{X}_1\|, (1+1/\kappa)\|X_2 - \tilde{X}_2\|\}^s] \\ &\leq (1+\kappa)^s \mathbb{E}[\|X_1 - \tilde{X}_1\|^s] + (1+1/\kappa)^s \mathbb{E}[\|X_2 - \tilde{X}_2\|^s] \\ &\leq (1+\kappa)^s D(r_1|X_1, \|\cdot\|^s) + (1+1/\kappa)^s D(r_2|X_2, \|\cdot\|^s) \\ &\quad + [(1+\kappa)^s + (1+1/\kappa)^s]\varepsilon. \end{aligned} \tag{4.3}$$

The statement follows since $\varepsilon > 0$ is arbitrary. $\square$

Remark 4.1.3. If in the previous lemma the parameter s is greater or equal to one, the calculations in (4.3) can be improved by applying the triangle inequality in the $L_s(\mathbb{P})$ space. Then

$$\mathbb{E}[\|X_1 + X_2 - (\tilde{X}_1 + \tilde{X}_2)\|^s]^{1/s} \le \mathbb{E}[\|X_1 - \tilde{X}_1\|^s]^{1/s} + \mathbb{E}[\|X_2 - \tilde{X}_2\|^s]^{1/s}$$

and one obtains

$$D(r_1 + r_2|X_1 + X_2, \|\cdot\|^s)^{1/s} \le D(r_1|X_1, \|\cdot\|^s)^{1/s} + D(r_2|X_2, \|\cdot\|^s)^{1/s}.$$

Lemma 4.1.1 and 4.1.2 imply

Corollary 4.1.4. *Fix $s > 0$ and suppose that $D(\,\cdot\,|X_1, \|\cdot\|^s)$ satisfies condition (4.1). If*

$$D(r|X_2, \|\cdot\|^s) = o(D(r|X_1, \|\cdot\|^s)) \quad (r \to \infty),$$

then

$$D(r|X_1 + X_2, \|\cdot\|^s) \sim D(r|X_1, \|\cdot\|^s) \quad (r \to \infty).$$

Proof. Set $X = X_1 + X_2$. By the previous lemma, one has, for $r \ge 0$, $\eta \in (0, 1]$ and $\kappa \in \mathbb{R}_+$,

$$\begin{aligned} D(r|X, \|\cdot\|^s) &\le (1+\kappa)^s D(r - \eta r|X_1, \|\cdot\|^s) + (1 + 1/\kappa)^s D(\eta r|X_2, \|\cdot\|^s) \\ &\sim (1+\kappa)^s D(r - \eta r|X_1, \|\cdot\|^s) \quad (r \to \infty). \end{aligned}$$

Since $\kappa > 0$ is arbitrary, it follows that

$$D(r|X, \|\cdot\|^s) \lesssim D(r - \eta r|X_1, \|\cdot\|^s).$$

Hence we can find a function $\eta : [0, \infty) \to (0, 1)$ with $\lim_{r\to\infty} \eta(r) = 0$ and

$$D(r|X, \|\cdot\|^s) \lesssim D((1 - \eta(r))r|X_1, \|\cdot\|^s).$$

With Lemma 3.1.4,

$$D(r|X, \|\cdot\|^s) \lesssim D(r|X_1, \|\cdot\|^s).$$

On the other hand, by Lemma 4.1.2, for $s > 0, \eta > 0, \kappa > 0$ and $r \ge 0$,

$$D(r + \eta r|X_1, \|\cdot\|^s) \le (1+\kappa)^s D(r|X, \|\cdot\|^s) + (1 + 1/\kappa)^s D(\eta r|X_2, \|\cdot\|^s).$$

Consequently,

$$\begin{aligned} (1+\kappa)^s D(r|X, \|\cdot\|^s) &\ge D(r + \eta r|X_1, \|\cdot\|^s) - (1 + 1/\kappa)^s D(\eta r|X_2, \|\cdot\|^s) \\ &\gtrsim D(r + \eta r|X_1, \|\cdot\|^s) \quad (r \to \infty). \end{aligned}$$

Now we proceed as in the proof of the converse inequality. $\square$

4.2 Perturbation results for entropy coding

In this section we study the perturbation problem for entropy coding. Recall that the convexity of the DRF implied immediately the result for the perturbation of rate (Lemma 4.1.1). It is not known whether the entropy coding error is a convex function. On the contrary, results of György and Linder [31] suggest that $D^{(e)}$ is not necessarily convex. They observed that the function

$$D^{(E)}(r|\mathcal{U}[0,1), |\cdot|^2) = \inf\big\{\mathbb{E}[|X-q(X)|^2] : X \text{ is a } \mathcal{U}[0,1)\text{-distributed r.v.}, \\ q:[0,1)\to[0,1), \mathbb{H}(q(X))\le r\big\}$$

is not convex in r. In order to prove an analog of Lemma 4.1.1 for entropy coding, we will relate the quantity $D^{(e)}$ to its convex hull. In the following, let (μ,ρ) be an arbitrary information source on a Polish space E and abridge $D^{(e)}(r) = D^{(e)}(r|\mu,\rho)$.

Lemma 4.2.1. *One has, for $r_1, r_2 \ge 0$ and $\gamma\in[0,1]$,*

$$D^{(e)}(\gamma r_1 + (1-\gamma) r_2 + \log 2) \le \gamma D^{(e)}(r_1) + (1-\gamma) D^{(e)}(r_2).$$

Proof. Let $\varepsilon>0$ and let $\hat X^{(i)}$ ($i\in\{1,2\}$) be such that

$$\mathbb{E}[\rho(X,\hat X^{(i)})] \le D^{(e)}(r_i)+\varepsilon$$

and

$$\mathbb{H}(\hat X^{(i)}) \le r_i.$$

Let ξ be a random variable that is independent of X and $\hat X^{(i)}$ ($i=1,2$) with

$$\mathbb{P}(\xi=1)=\gamma \text{ and } \mathbb{P}(\xi=2)=1-\gamma.$$

Denote $\hat X = \hat X^{(\xi)}$. Then

$$\begin{aligned}\mathbb{H}(\hat X) &\le \mathbb{H}(\hat X,\xi) = \mathbb{H}(\xi)+\mathbb{H}(\hat X^{(\xi)}|\xi)\\ &= \mathbb{H}(\xi) + \sum_{i=1}^2 \mathbb{P}(\xi=i)\,\mathbb{H}(\hat X^{(i)})\\ &\le \gamma r_1 + (1-\gamma) r_2 + \log 2.\end{aligned}$$

Moreover,

$$\begin{aligned}\mathbb{E}[\rho(X,\hat X)] &= \sum_{i=1}^2 \mathbb{P}(\xi=i)\,\mathbb{E}[\rho(X,\hat X^{(i)})]\\ &\le \gamma D^{(e)}(r_1) + (1-\gamma) D^{(e)}(r_2)+\varepsilon,\end{aligned}$$

where we have used the independence of ξ and $(X, \hat{X}^{(1)}, \hat{X}^{(2)})$. The proof is complete since $\varepsilon > 0$ was arbitrary. □

For $r \geq 0$, we denote

$$\begin{aligned}\bar{D}^{(e)}(r) &= \inf\Big\{\int D^{(e)}(x)\,d\nu(x) : \nu \in \mathcal{M}_1([0,\infty)), \int x\,d\nu(x) \leq r\Big\}\\ &= \inf\Big\{\int D^{(e)}(x)\,d\nu(x) : \nu \in \mathcal{M}_1([0,\infty)), \int x\,d\nu(x) = r\Big\}.\end{aligned}$$

We use results of convex analysis to study $\bar{D}^{(e)}$.

Denote by $\mathrm{conv}(D^{(e)})$ the convex hull of $D^{(e)}$, i.e. $\mathrm{conv}(D^{(e)})$ is the biggest convex function on $[0,\infty)$ that is dominated by $D^{(e)}$. Due to Carathéodory's Theorem, one can represent $\mathrm{conv}(D^{(e)})$ in the form

$$\begin{aligned}\mathrm{conv}(D^{(e)})(r) = \inf\big\{&\lambda_1\,D^{(e)}(r_1) + \lambda_2\,D^{(e)}(r_2) :\ \lambda_1\,r_1 + \lambda_2\,r_2 = r,\\ &\lambda_1+\lambda_2 = 1,\ \lambda_1,\lambda_2 \in [0,1],\ r_1, r_2 \in \mathbb{R}\big\}\end{aligned}$$

for $r \geq 0$. A proof of Caratheodory's representation theorem is contained in Rockafellar's monograph [59], Corollary 17.1.5.

Proposition 4.2.2. *One has*

$$\bar{D}^{(e)}(r) = \mathrm{conv}(D^{(e)})(r), \quad r \geq 0.$$

Proof. First we prove $\bar{D}^{(e)}(r) \leq \mathrm{conv}(D^{(e)})(r)$ for $r \geq 0$. Let

$$r = \lambda_1 r_1 + \lambda_2 r_2$$

be a convex combination of r_1 and r_2 in $[0,\infty)$, i.e. $\lambda_1, \lambda_2 \in [0,1]$ with $\lambda_1 + \lambda_2 = 1$. Consider the measure $\nu = \sum_{i=1}^2 \lambda_i \delta_{r_i} \in \mathcal{M}_1([0,\infty))$, with δ denoting Dirac measure. Note that $\int x\,d\nu(x) = r$ and, hence,

$$\bar{D}^{(e)}(r) \leq \int D^{(e)}(x)\,d\nu(x) = \lambda_1 D^{(e)}(r_1) + \lambda_2 D^{(e)}(r_2).$$

Consequently, $\bar{D}^{(e)}(r) \leq \mathrm{conv}(D^{(e)})(r)$ for $r \geq 0$.

It remains to prove $\bar{D}^{(e)}(r) \geq \mathrm{conv}(\bar{D}^{(e)})(r)$, $r \in \mathbb{R}$. Recall that $\mathrm{conv}(D^{(e)})$ is a convex function that is smaller than $D^{(e)}$. Hence,

$$\begin{aligned}\bar{D}^{(e)}(r) &= \inf\Big\{\int D^{(e)}(x)\,d\nu(x) : \nu \in \mathcal{M}_1([0,\infty)), \int x\,d\nu(x) = r\Big\}\\ &\geq \inf\Big\{\int \mathrm{conv}(D^{(e)})(x)\,d\nu(x) : \nu \in \mathcal{M}_1([0,\infty)), \int x\,d\nu(x) = r\Big\}\\ &\geq \mathrm{conv}(D^{(e)})(r),\end{aligned}$$

where the last estimate holds due to Jensen's inequality. □

Combining the previous proposition with Lemma 4.2.1, yields

Corollary 4.2.3. *For $r \geq 0$,*

$$D^{(e)}(r + \log 2) \leq \bar{D}^{(e)}(r).$$

In analogy to condition (4.1), we require that

$$D^{(e)}(2r) \approx D^{(e)}(r) \quad (r \to \infty) \tag{4.4}$$

in order to prove the perturbation result.

Lemma 4.2.4. *Let $\Delta : [0, \infty) \to [0, \infty)$ with $\Delta r = o(r)$ and suppose that $D^{(e)}$ satisfies condition (4.4). Then*

$$D^{(e)}(r + \Delta r) \sim D^{(e)}(r) \quad (r \to \infty).$$

Proof. By Corollary 4.2.3 and assumption (4.4), one has

$$\bar{D}^{(e)}(2r) \approx \bar{D}^{(e)}(r).$$

Due to Lemma 3.1.4 and the convexity of $\bar{D}^{(e)}$, it follows

$$\bar{D}^{(e)}(r + \Delta r) \sim \bar{D}^{(e)}(r) \quad (r \to \infty).$$

Moreover, applying Corollary 4.2.3, yields for $r \geq \log 2$

$$\bar{D}^{(e)}(r) \leq D^{(e)}(r) \leq \bar{D}^{(e)}(r - \log 2) \sim \bar{D}^{(e)}(r) \quad (r \to \infty).$$

□

Remark 4.2.5. In Section 1.3, we motivated the definition of entropy coding with the close relationship to variable rate compression. A discrete reconstruction $\hat{X}$ admits a prefix free representation Ψ of its support with

$$\frac{1}{\log 2}\,\mathbb{H}(\hat{X}) = \mathbb{H}_2(\hat{X}) \leq \mathbb{E}[l(\Psi(\hat{X}))] < \mathbb{H}_2(\hat{X}) + 1 = \frac{1}{\log 2}\,\mathbb{H}(\hat{X}) + 1.$$

Under the assumption (4.4), variable rate compression with constraint $\mathbb{E}[l(\Psi(\hat{X}))] \leq \log(2)\, r$ yields asymptotically the same coding error as entropy coding.

Now we consider perturbations in the original. Let X_1 and X_2 be two random elements in a Banach space $(E, \|\cdot\|)$.

Lemma 4.2.6. *For $s, \kappa > 0$ and $r_1, r_2 \geq 0$,*

$$\begin{aligned} &D^{(e)}(r_1 + r_2 | X_1 + X_2, \|\cdot\|^s) \\ &\qquad \leq (1+\kappa)^s D^{(e)}(r_1 | X_1, \|\cdot\|^s) + (1 + 1/\kappa)^s D^{(e)}(r_2 | X_2, \|\cdot\|^s). \end{aligned}$$

Proof. Let $\varepsilon > 0$ and let $\hat{X}_i$ $(i = 1, 2)$ be discrete Borel measurable random elements in E such that

$$\mathbb{E}[\|X_i - \hat{X}_i\|^s] \leq D^{(e)}(r_i|X_i, \|\cdot\|^s) + \varepsilon$$

and

$$\mathbb{H}(\hat{X}_i) \leq r_i$$

for $i \in \{1, 2\}$. Then $\mathbb{H}(\hat{X}_1 + \hat{X}_2) \leq \mathbb{H}(\hat{X}_1, \hat{X}_2) \leq \mathbb{H}(\hat{X}_1) + \mathbb{H}(\hat{X}_2) \leq r_1 + r_2$ by basic properties of the entropy (see Ihara [32], Theorems 1.2.1 and 1.2.2). Furthermore, analogously to the calculations in (4.3), one has

$$\begin{aligned}\mathbb{E}[\|X_1 + X_2 - (\hat{X}_1 + \hat{X}_2)\|^s] \leq & (1+\kappa)^s D^{(e)}(r_1|X_1, \|\cdot\|^s) \\ & + (1 + 1/\kappa)^s D^{(e)}(r_2|X_2, \|\cdot\|^s) \\ & + [(1+\kappa)^s + (1+1/\kappa)^s]\varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, the statement follows. □

Analogously to Corollary 4.1.4, it follows

Corollary 4.2.7. *Let $s > 0$ and suppose that $D^{(e)}(r|X_1, \|\cdot\|^s)$ fulfills condition (4.4). If*

$$D^{(e)}(r|X_2, \|\cdot\|^s) = o(D^{(e)}(r|X_1, \|\cdot\|^s)) \quad (r \to \infty),$$

then

$$D^{(e)}(r|X_1 + X_2, \|\cdot\|^s) \sim D^{(e)}(r|X_1, \|\cdot\|^s) \quad (r \to \infty).$$

4.3 Perturbation of the rate for the quantization problem

In the previous section we used convexity and "almost" convexity of the functions D and $D^{(e)}$, respectively, to prove perturbation results. Unfortunately it is not known whether such strong statements hold for the quantization error. The regularity result which we will develop is applicable in our typical setting. However, it is too weak to yield as strong conclusion as for the DRF and entropy coding.

Let again (μ, ρ) be an information source on E. Consider the quantization error

$$\delta^{(q)}(N) = \delta^{(q)}(N|\mu, \rho) = \inf_{\substack{\mathcal{C} \subset E: \\ |\mathcal{C}| \leq N}} \int \rho(x, \mathcal{C})\, \mu(dx), \quad N \in [1, \infty).$$

Lemma 4.3.1. *Let $N \in [1,\infty)$, $p \in (0,1)$ and $\kappa \geq 1$. Then*

$$\delta^{(q)}(N + \lceil p\kappa N \rceil) \leq p\delta^{(q)}(\kappa N) + (1-p)\delta^{(q)}(N).$$

Proof. We need to show that there exists a codebook with at most $\lfloor N \rfloor + \lceil p\kappa N \rceil$ elements that achieves the distortion bound of the statement. Let X denote a μ distributed random element. Without loss of generality we assume that E contains at least $\lfloor \kappa N \rfloor$ elements. Fix $\varepsilon > 0$ and let $\mathcal{C}_1 \subset E$ be a codebook with $\lfloor \kappa N \rfloor$ elements satisfying

$$\mathbb{E}[\rho(X, \mathcal{C}_1)] \leq \delta^{(q)}(\kappa N) + \varepsilon.$$

Let $\pi : E \to \mathcal{C}_1$ be a Borel measurable function such that

$$\mathbb{E}[\rho(X, \mathcal{C}_1)] = \mathbb{E}[\rho(X, \pi(X))].$$

Furthermore, let $\mathcal{C}_2 \subset E$ be a codebook with at most $\lfloor N \rfloor$ elements that satisfies

$$\mathbb{E}[\rho(X, \mathcal{C}_2)] \leq \delta^{(q)}(N) + \varepsilon.$$

We assume without loss of generality that $\lfloor N \rfloor + \lceil p\kappa N \rceil < \lfloor \kappa N \rfloor$ since otherwise the statement is trivial. We consider the following random codebook. Choose uniformly $\lceil p\kappa N \rceil$ different elements of $\mathcal{C}_1$ without preference for any combination and denote the corresponding set of points by $\mathcal{C}^{(R)}$. We assume independence of X and $\mathcal{C}^{(R)}$. Now set $\mathcal{C} = \mathcal{C}^{(R)} \cup \mathcal{C}_2$. Note that $|\mathcal{C}| \leq \lfloor N \rfloor + \lceil p\kappa N \rceil$. Moreover,

$$\begin{aligned}\mathbb{E}[\rho(X, \mathcal{C})] \leq\ & \mathbb{P}(\pi(X) \in \mathcal{C}^{(R)})(\delta^{(q)}(\kappa N) + \varepsilon) \\ &+ \mathbb{P}(\pi(X) \notin \mathcal{C}^{(R)})(\delta^{(q)}(N) + \varepsilon),\end{aligned}$$

since the event $\{\pi(X) \in \mathcal{C}^{(R)}\}$ is independent of X. Furthermore,

$$\mathbb{P}(\pi(X) \in \mathcal{C}^{(R)}) = \frac{\lceil p\kappa N \rceil}{\lfloor \kappa N \rfloor} \geq p.$$

Since $\delta^{(q)}(N) \geq \delta^{(q)}(\kappa N)$, we obtain

$$\mathbb{E}[\rho(X, \mathcal{C})] \leq p(\delta^{(q)}(\kappa N) + \varepsilon) + (1-p)(\delta^{(q)}(N) + \varepsilon).$$

Since $\varepsilon > 0$ is arbitrary the proof is complete. □

The regularity result requires the following

Assumption 4.3.2. There exist constants $c, C > 0$ and a slowly varying function $f : [1,\infty) \to [1,\infty)$ such that

$$cf(N) \lesssim \delta^{(q)}(N) \lesssim Cf(N) \quad (N \to \infty).$$

Note that the quantization error $\delta^{(q)}$ satisfies Assumption 4.3.2 in many cases. For instance, the assumption is satisfied if Theorem 3.1.2 is applicable. In particular, it is fulfilled for all examples presented in Section 3.6.

Lemma 4.3.3. *Let $\delta^{(q)}$ satisfy Assumption 4.3.2. Then, for any $r \in (0,1]$,*

$$\frac{\delta^{(q)}(rN)}{\delta^{(q)}(N)} \lesssim 1 + (1-r)(\tfrac{C}{c} - 1) \text{ as } N \to \infty.$$

Proof. Owing to Lemma 4.3.1,

$$(1-p)[\delta^{(q)}(N) - \delta^{(q)}(N + \lceil p\kappa N \rceil)] \geq p[\delta^{(q)}(N + \lceil p\kappa N \rceil) - \delta^{(q)}(\kappa N)]$$

for $N \in [1,\infty)$, $p \in (0,1)$ and $\kappa \geq 1$. Hence,

$$\begin{aligned}
&\delta^{(q)}(N) - \delta^{(q)}(\kappa N) \\
&\quad = \delta^{(q)}(N) - \delta^{(q)}(N + \lceil p\kappa N \rceil) + \delta^{(q)}(N + \lceil p\kappa N \rceil) - \delta^{(q)}(\kappa N) \\
&\quad \geq (1 + \frac{p}{1-p})[\delta^{(q)}(N + \lceil p\kappa N \rceil) - \delta^{(q)}(\kappa N)].
\end{aligned}$$

Dividing by $\delta^{(q)}(\kappa N)$ and regrouping the terms gives

$$\frac{\delta^{(q)}(N + \lceil p\kappa N \rceil)}{\delta^{(q)}(\kappa N)} \leq 1 + (1-p)\Big(\frac{\delta^{(q)}(N)}{\delta^{(q)}(\kappa N)} - 1\Big).$$

By Assumption 4.3.2, it follows

$$\frac{\delta^{(q)}(N + \lceil p\kappa N \rceil)}{\delta^{(q)}(\kappa N)} \lesssim 1 + (1-p)(\tfrac{C}{c} - 1) \text{ as } N \to \infty.$$

Now let $\kappa \geq 1$ and $p \in (0,1)$ be such that $r > (1 + p\kappa)/\kappa = p + 1/\kappa$. The previous asymptotic estimate yields

$$\frac{\delta^{(q)}(rN)}{\delta^{(q)}(N)} \lesssim 1 + (1-p)(\tfrac{C}{c} - 1) \text{ as } N \to \infty.$$

Note that for any $p \in (0,r)$ there exists $\kappa \geq 1$ with $r > p + 1/\kappa$. Therefore,

$$\frac{\delta^{(q)}(rN)}{\delta^{(q)}(N)} \lesssim 1 + (1-r)(\tfrac{C}{c} - 1) \text{ as } N \to \infty.$$

□

In particular, one obtains the following corollary.

Corollary 4.3.4. *Let $\Delta : [1,\infty) \to \mathbb{R}$, $N \mapsto \Delta N$ with $\Delta N = o(N)$ as $N \to \infty$. If Assumption 4.3.2 holds, then it holds*

$$\delta^{(q)}(N + \Delta N) \sim \delta^{(q)}(N) \text{ as } N \to \infty.$$

Translating the corollary into the notion of rates yields

$$D^{(q)}(r + \Delta r) \sim D^{(q)}(r) \quad (r \to \infty),$$

if Assumption 4.3.2 is satisfied and $\Delta : r \mapsto \Delta r$, $[0, \infty) \to \mathbb{R}$ is such that $\lim_{r\to\infty} \Delta r = 0$. This statement is much weaker than the corresponding results for the DRF (Lemma 4.1.1) and entropy coding (Lemma 4.2.4).

4.4 Coding diffusion processes

In this section, $X = \{X_t\}_{t\in[0,1]}$ denotes a stochastic process in $\mathbb{R}^d$, $d \in \mathbb{N}$, that solves the stochastic differential equation (SDE)

$$X_t = \int_0^t f(X_s, s)\, ds + B_t, \quad t \in [0, 1],$$

where $f : \mathbb{R}^d \times [0, 1] \to \mathbb{R}^d$ is a Borel measurable function and $\{B_t\}_{t\in[0,1]}$ denotes d-dimensional Brownian motion. Throughout the following considerations we impose a regularity condition on the drift function f. We suppose that there exist constants $L, c > 0$ such that

$$|f(x, t)| \le L|x| + c \tag{4.5}$$

for all $x \in \mathbb{R}^d$ and $t \in [0, 1]$. Let $G = \mathbb{R}^d$ be a Banach space equipped with a norm $|\cdot|_G$ and denote, for a function $g : [0, 1] \to \mathbb{R}^d$,

$$\|g\|_{[0,1],G} = \sup_{t\in[0,1]} |g(t)|_G$$

and

$$\|g\|_{L_p([0,1],G)} = \Big(\int_0^1 |g(t)|_G^p\, dt\Big)^{1/p}.$$

The main task of this section is to prove

Theorem 4.4.1. *Let $\|\cdot\| = \|\cdot\|_{[0,1],G}$ or $\|\cdot\| = \|\cdot\|_{L_p([0,1],G)}$ for some $p \ge 1$. Under the regularity condition (4.5), one has, for $s \in [1, \infty)$,*

$$D(r|X, \|\cdot\|^s) \sim D(r|B, \|\cdot\|^s)$$

and

$$D^{(e)}(r|X, \|\cdot\|^s) \sim D^{(e)}(r|B, \|\cdot\|^s)$$

as $r \to \infty$.

From now on, let $\|\cdot\| = \|\cdot\|_{[0,1]}$ be the standard supremum norm, i.e.

$$\|g\| = \sup_{t\in[0,1]} |g(t)|.$$

In order to prove the theorem, we decompose X into $X_t = Y_t + B_t$, $t \in [0,1]$, where $Y_t = \int_0^t f(X_t,t)\,dt$. We will show that it is easier to code the process $Y = \{Y_t\}_{t\in[0,1]}$ than coding Brownian motion.

For $g \in C([0,1],\mathbb{R}^d)$ and $\alpha \in (0,1]$, we denote

$$\|g\|_{C^\alpha} = \sup_{\substack{s,t\in[0,1],\\ s\neq t}} \frac{|g(t)-g(s)|}{|t-s|^\alpha}.$$

and let

$$C_0^\alpha = \Big\{ f \in C([0,1],\mathbb{R}^d) : f(0) = 0,\ \|f\|_{C^\alpha} < \infty \Big\}$$

be the Banach space of α-Hölder continuous functions starting in 0. Note that

$$\|Y\|_{C^1} \le \sup_{t\in[0,1]} |f(X_t,t)| \le \sup_{t\in[0,1]} [L|X_t| + c] = L\|X\| + c. \tag{4.6}$$

In order to obtain an upper bound for the coding complexity of Y we will use a result of Kolmogorov and Tikhomirov ([38], Theorem XIV).

Lemma 4.4.2. *For $\alpha \in (0,1]$, it holds*

$$\log N_e(\varepsilon, B_{C_0^\alpha}(0,1)) \approx \frac{1}{\varepsilon^{1/\alpha}} \quad (\varepsilon \downarrow 0),$$

where $N_e(\varepsilon, B_{C_0^\alpha}(0,1))$ denotes the ε-entropy of the unit ball of C_0^α in the space $C([0,1],\mathbb{R}^d)$ equipped with the supremum norm $\|\cdot\|$.

The previous lemma follows as well from Proposition 5.7.1 of Carl and Stephani [12]. We need some lemmas for the proof of Theorem 4.4.1.

Lemma 4.4.3. *There exist $A, B \in \mathbb{R}_+$ such that, for all $t \ge 0$,*

$$\mathbb{P}(\|X\| \ge t) \le \exp\{-[(t-A)^+]^2/B\},$$

where $(t-A)^+ = (t-A) \vee 0$.

Proof. Denote $X_t = (X_t^{(1)},\dots,X_t^{(d)})$ and $B_t = (B_t^{(1)},\dots,B_t^{(d)})$ and let $\{B_t^{(d+1)}, B_t^{(d+2)}\}_{t\in[0,1]}$ be a 2-dimensional Brownian motion that is independent of X and B. For technical reasons we study the $d+2$-dimensional process

$$\tilde{X}_t = (X_t^{(1)},\dots,X_t^{(d)}, B_t^{(d+1)}, B_t^{(d+2)} + 1)$$

instead of X. Clearly, $\|X\| \leq \|\tilde{X}\|$. By definition, $\tilde{X}$ solves the differential equation

$$d\tilde{X}_t = \tilde{f}(\tilde{X}_t, t)\,dt + d\tilde{B}_t$$

with $\tilde{X}_0 = (0,\dots,0,1)$, $\tilde{B}_t = (B_t^{(1)},\dots,B_t^{(d+2)})$ and $\tilde{f}(x_1,\dots,x_{d+2},t) = f(x_1,\dots,x_d,t)$ for $x \in \mathbb{R}^{d+2}$ and $t \in [0,1]$. Note that $\tilde{f}$ satisfies the $d+2$-dimensional analog to property (4.5).

Since by construction $\tilde{X}_t$ stays a.s. at all times $t \in [0,1]$ in the domain $\mathbb{R}^{d+2}\backslash\{0\}$, one obtains by Itô's formula

$$|\tilde{X}_t| = 1 + \int_0^t \Big[\frac{1}{|\tilde{X}_s|}\,\langle \tilde{X}_s, \tilde{f}(\tilde{X}_s, s)\rangle + \frac{d+1}{|\tilde{X}_s|}\Big]\,ds + \sum_{i=1}^{d+2}\int_0^t \frac{\tilde{X}_s^{(i)}}{|\tilde{X}_s|}\,dB_s^{(i)}.$$

Noting that $\sum_{i=1}^{d+2}\int_0^t \frac{\tilde{X}_s^{(i)}}{|\tilde{X}_s|}\,dB_s^{(i)}$ is a martingale with quadratic variation process $\{t\}_{t\in[0,1]}$, we conclude that it is a standard Brownian motion which we will denote by $\hat{B} = \{\hat{B}_t\}_{t\in[0,1]}$. Hence,

$$|\tilde{X}_t| = 1 + \int_0^t \Big[\frac{1}{|\tilde{X}_s|}\,\langle \tilde{X}_s, \tilde{f}(\tilde{X}_s, s)\rangle + \frac{d+1}{|\tilde{X}_s|}\Big]\,ds + \hat{B}_t.$$

Let now T_2 denote a random time when $|\tilde{X}_\cdot|$ attains its maximum in the time interval $[0,1]$ and denote by T_1 the last time before time T_2 when $|\tilde{X}_\cdot|$ hits 1. Then

$$\begin{aligned}
|\tilde{X}_{T_2}| &\leq 1 + \int_{T_1}^{T_2} [\frac{1}{|\tilde{X}_t|}\,\langle \tilde{X}_t, \tilde{f}(\tilde{X}_t, t)\rangle + (d+1)]\,dt + \hat{B}_{T_2} - \hat{B}_{T_1} \\
&\leq 1 + \int_{T_1}^{T_2} [L\,|\tilde{X}_t| + c + (d+1)]\,dt + \hat{B}_{T_2} - \hat{B}_{T_1} \\
&\leq 1 + c + (d+1) + 2\|\hat{B}\| + L\int_{T_1}^{T_2} |\tilde{X}_t|\,dt,
\end{aligned}$$

where we have used condition (4.5). Set $\tilde{c} = 1 + c + (d+1)$. With the Gronwall Lemma, we obtain

$$|\tilde{X}_{T_2}| \leq (2\|\hat{B}\| + \tilde{c})\,e^L.$$

Hence

$$\|\tilde{X}\| \leq 2\,e^L\|\hat{B}\| + \tilde{c}\,e^L.$$

An application of the concentration inequality given in (2.2) to the Brownian motion $\hat{B}$ implies the existence of constants $A, B \in \mathbb{R}_+$, such that

$$\mathbb{P}(\|X\| \geq t) \leq \exp\big\{-[(t-A)^+]^2/B\big\}$$

for $t \geq 0$. □

Lemma 4.4.4. *Let $\alpha \in (0,1]$ and $\tilde{s} > s > 0$. There exists a constant $\kappa = \kappa(s,\tilde{s})$ such that for all C_0^α-valued processes Z, it holds*

$$D^{(q)}(r|Z, \|\cdot\|_{[0,1]}^s)^{1/s} \le \kappa\, \mathbb{E}[\|Z\|_{C^\alpha}^{\tilde{s}}]^{1/\tilde{s}}\, \frac{1}{r^\alpha}. \tag{4.7}$$

Proof. Let

$$U = B_{C_0^\alpha}(0,1).$$

By Lemma 4.4.2, there exists a constant $c_1 \in \mathbb{R}_+$ such that

$$\log N_e(\varepsilon, U) \le \frac{c_1}{\varepsilon^{1/\alpha}}$$

for all $\varepsilon > 0$. It follows that

$$\log N_e(\varepsilon, rU) \le \frac{c_1\, r^{1/\alpha}}{\varepsilon^{1/\alpha}} \tag{4.8}$$

for every $r, \varepsilon > 0$. We fix $\eta > 0$ such that $(1+\eta)s < \tilde{s}$ and let $\varepsilon > 0$ be arbitrary. Set

$$\varepsilon_i = \varepsilon\, e^{(1+\eta)i}, \quad i \in \mathbb{N}_0.$$

Moreover, let

$$r_i = e^i, \quad i \in \mathbb{N}_0,$$

and $r_{-1} = 0$. We use ε_i-nets of the sets $r_i U$ to build appropriate codebooks. Note that $\varepsilon_i \ge r_i$, if

$$i \ge \left\lceil \frac{1}{\eta} \log(1/\varepsilon) \right\rceil =: N.$$

Since $\|g\|_{C^\alpha} \ge \|g\|$ for $g \in C_0^\alpha$, the set $\{0\}$ is an optimal ε_i-net of $r_i U$ for $i \ge N$. For $\varepsilon > 0$, we consider the codebook

$$\mathcal{C}(\varepsilon) = \{0\} \cup \bigcup_{i=0}^{N-1} \mathcal{C}_i(\varepsilon),$$

where $\mathcal{C}_i(\varepsilon)$ are arbitrary optimal ε_i-nets of $r_i U$ for $i \in \mathbb{N}_0$. By basic analysis, one concludes that

$$\begin{aligned}
\mathbb{E}[d(Z,\mathcal{C}(\varepsilon))^s] &\le \sum_{i=0}^{\infty} \mathbb{E}[1_{\{\|Z\|_{C^\alpha}\in[r_{i-1},r_i]\}}\, d(Z,\mathcal{C}_i(\varepsilon))^s] \\
&\le \sum_{i=0}^{\infty} \mathbb{P}(\|Z\|_{C^\alpha} \ge r_{i-1})\, \varepsilon_i^s \\
&= \varepsilon^s + \sum_{i=1}^{\infty} \mathbb{P}\Big(\frac{\|Z\|_{C^\alpha}^{\tilde{s}}}{r_{i-1}^{\tilde{s}}} \ge 1\Big)\, \varepsilon_i^s \\
&\le \varepsilon^s + \mathbb{E}[\|Z\|_{C^\alpha}^{\tilde{s}}] \sum_{i=1}^{\infty} \frac{\varepsilon_i^s}{r_{i-1}^{\tilde{s}}} \\
&= \varepsilon^s + \mathbb{E}[\|Z\|_{C^\alpha}^{\tilde{s}}]\, \varepsilon^s \sum_{i=1}^{\infty} \frac{e^{(1+\eta)is}}{e^{(i-1)\tilde{s}}} \\
&= \varepsilon^s + \mathbb{E}[\|Z\|_{C^\alpha}^{\tilde{s}}]\, \varepsilon^s \sum_{i=1}^{\infty} e^{\tilde{s}-(\tilde{s}-(1+\eta)s)i}.
\end{aligned}$$

Since $\tilde{s} > (1+\eta)s$, the previous sum converges. Consequently, there exists a constant $c_2 \in \mathbb{R}_+$ such that

$$\mathbb{E}[d(Z,\mathcal{C}(\varepsilon))^s] \le (1 + c_2\, \mathbb{E}[\|Z\|_{C^\alpha}^{\tilde{s}}])\, \varepsilon^s.$$

Now we estimate the number of elements in $\mathcal{C}(\varepsilon)$. If $\varepsilon \ge 1$ then $N \le 0$ and $|\mathcal{C}(\varepsilon)| = 1$. Otherwise, one obtains with equation (4.8)

$$\begin{aligned}
|\mathcal{C}(\varepsilon)| &\le 1 + \sum_{i=0}^{N-1} |\mathcal{C}_i(\varepsilon)| \le 1 + \sum_{i=0}^{N-1} \exp\big\{c_1\, (r_i/\varepsilon_i)^{1/\alpha}\big\} \\
&= 1 + \sum_{i=0}^{N-1} \exp\big\{c_1\, \frac{1}{\varepsilon^{1/\alpha}}\, e^{-i\eta/\alpha}\big\} \\
&\le 1 + N\, \exp\big\{c_1\, \frac{1}{\varepsilon^{1/\alpha}}\big\} \\
&\le 1 + \Big(1 + \frac{1}{\eta}\, \log(1/\varepsilon)\Big)\, \exp\big\{c_1\, \frac{1}{\varepsilon^{1/\alpha}}\big\}.
\end{aligned}$$

The previous expression can be bounded uniformly, for $\varepsilon \in (0,1)$, by

$$|\mathcal{C}(\varepsilon)| \le \exp\big\{c_3\, \frac{1}{\varepsilon^{1/\alpha}}\big\},$$

where $c_3 \in \mathbb{R}_+$ is an appropriate constant. Note that the estimate is also valid for $\varepsilon \ge 1$. Consequently,

$$D^{(q)}\Big(c_3\, \frac{1}{\varepsilon^{1/\alpha}}\, |Z, \|\cdot\|^s\Big) \le [1 + c_2\, \mathbb{E}[\|Z\|_{C^\alpha}^{\tilde{s}}]\, \varepsilon^s$$

for $\varepsilon > 0$ and, hence, for all $r \geq 0$

$$D^{(q)}(r|Z, \|\cdot\|^s)^{1/s} \leq [1 + c_2\,\mathbb{E}[\|Z\|_{C^\alpha}^{\tilde{s}}]^{1/s}\, c_3^\alpha\, \frac{1}{r^\alpha}. \tag{4.9}$$

All constants c_1, c_2 and c_3 depend on s, $\tilde{s}$ and η but not on the process Z. Note that the left hand side of equation (4.7) satisfies

$$D^{(q)}(r|\lambda Z, \|\cdot\|_{[0,1]}^s)^{1/s} = \lambda\, D^{(q)}(r|Z, \|\cdot\|_{[0,1]}^s)^{1/s}$$

for $\lambda \in [0,\infty)$. The same scaling property holds for the right hand side of equation (4.7). Hence, it suffices to prove the inequality (4.7) for processes Z with $\mathbb{E}[\|Z\|_{C^\alpha}^{\tilde{s}}] = 1$. Consequently, the assertion follows from (4.9). □

Proof of Theorem 4.4.1. Let $\|\cdot\| = \|\cdot\|_{[0,1],G}$ or $\|\cdot\| = \|\cdot\|_{L_p([0,1],G)}$, $p \geq 1$. As a consequence of Section 3.6, one has, for any $s \geq 1$,

$$D(r|B, \|\cdot\|^s)^{1/s} \approx D^{(e)}(r|B, \|\cdot\|^s)^{1/s} \approx \frac{1}{\sqrt{r}}.$$

On the other hand, for any $g \in C([0,1], \mathbb{R}^d)$, it holds

$$\|g\| \leq \|g\|_{[0,1],G}.$$

Since all norms on $\mathbb{R}^d$ are equivalent, there exists a constant $c > 0$ such that

$$\|g\| \leq \|g\|_{[0,1],G} \leq c\,\|g\|_{[0,1]}$$

for all $g \in C([0,1], \mathbb{R}^d)$. By the previous lemma, we obtain

$$D^{(q)}(r|Y, \|\cdot\|^s)^{1/s} \leq c\,\kappa\,\mathbb{E}[\|Y\|_{C^1}^{2s}]^{1/2s}\,\frac{1}{r},$$

where κ is as in the lemma. Due to Lemma 4.4.3, $\mathbb{E}[\|Y\|_{C^1}^{2s}]$ is finite. An application of Corollaries 4.1.4 and 4.2.7 completes the proof. □

Chapter 5

General results

5.1 Spread of "good" codebooks

Let μ denote a centered Gaussian measure on a separable Banach space $(E, \|\cdot\|)$. In this section, we are concerned with the quantization problem of the information source $(\mu, \|\cdot\|^s)$, $s > 0$. We study the spread of a "good" codebook. Denote by X a μ-distributed random element and set

$$\sigma = \sigma(X) = \sup_{f \in E' \setminus \{0\}} \frac{(\mathbb{E}[f(X)^2])^{1/2}}{\|f\|_{E'}}.$$

Here, E' denotes the topological dual of E equipped with the norm

$$\|f\|_{E'} = \sup_{x \in B_E(0,1)} |f(x)|, \quad f \in E'.$$

Lemma 5.1.1. *For each $s > 0$, there exist a sequence $\{c_N\}_{N\in\mathbb{N}}$ in $\mathbb{R}_+$ and a sequence of codebooks $\{\mathcal{C}_N\}_{N\in\mathbb{N}}$ in E with $|\mathcal{C}_N| \le N$ such that*

$$\begin{aligned} &i.) \quad c_N \sim \sqrt{2\sigma^2 \log_- \delta^{(q)}(N|\mu, \|\cdot\|^s)} \quad (N \to \infty), \\ &ii.) \quad \mathcal{C}_N \subset B(0, 2c_N) \quad (N \in \mathbb{N}) \quad \textit{and} \\ &iii.) \quad \mathbb{E}[d(X, \mathcal{C}_N)^s] \sim \delta^{(q)}(N|\mu, \|\cdot\|^s) \quad (N \to \infty), \end{aligned}$$

where $\log_- x = (-\log x) \vee 0$ *for* $x > 0$.

For the proof we need

Proposition 5.1.2. *For $\varepsilon > 0$, one has*

$$\mathbb{E}[\|X\|^s 1_{\{\|X\|>t\}}] \lesssim \exp\Big\{-\frac{(1-\varepsilon)t^2}{2\sigma^2}\Big\} \quad (t \to \infty).$$

Proof. By Hölder's inequality, it holds, for $p, q \in (1,\infty)$ with $\frac{1}{p}+\frac{1}{q}=1$,

$$\mathbb{E}[\|X\|^s 1_{\{\|X\|>t\}}] \le \mathbb{E}[\|X\|^{sp}]^{1/p}\, \mathbb{P}(\|X\| \ge t)^{1/q}.$$

By the finiteness of all moments the first expectation is finite. Let $M \in \mathbb{R}$ denote a median of the random variable $\|X\|$. Recall, that due to the isoperimetric inequality (see (2.2)),

$$\mathbb{P}(\|X\| > t) \le \Psi((t-M)/\sigma) \le \exp\Big\{-\frac{(t-M)^2}{2\sigma^2}\Big\}.$$

Hence,

$$\mathbb{E}[\|X\|^s 1_{\{\|X\|>t\}}] \le \mathbb{E}[\|X\|^{sp}]^{1/p} \exp\Big\{-\frac{(t-M)^2}{2q\sigma^2}\Big\}.$$

Since $q > 1$ may be chosen arbitrarily close to 1 the result follows. □

Proof of Lemma 5.1.1. Let $\{\mathcal{C}_N\}_{N\in\mathbb{N}}$ denote a sequence of codebooks in E with $|\mathcal{C}_N| \le N$ and $\mathbb{E}[d(X,\mathcal{C}_N)^s] \sim \delta^{(q)}(N|\mu, \|\cdot\|^s)$ as $N \to \infty$. Fix $\varepsilon > 0$ and consider

$$c_N := (1+\varepsilon)\sqrt{2\sigma^2 \log_- \delta^{(q)}(N|\mu, \|\cdot\|^s)}, \quad N \in \mathbb{N}.$$

We split $\mathcal{C}_N$ into $\mathcal{C}_{N,1} := \mathcal{C}_N \cap B(0, 2c_N)$ and $\mathcal{C}_{N,2} := \mathcal{C}_N \cap B(0,2c_N)^c$. Set

$$\tilde{\mathcal{C}}_N = \begin{cases} \mathcal{C}_N & \text{if } \mathcal{C}_{N,2} = \emptyset, \\ \mathcal{C}_{N,1} \cup \{0\} & \text{else.} \end{cases}$$

Then $|\tilde{\mathcal{C}}_N| \le N$ and $\tilde{\mathcal{C}}_N \subset B(0,2c_N)$. If $\mathcal{C}_{N,2} = \emptyset$, coding with $\tilde{\mathcal{C}}_N$ and $\mathcal{C}_N$ coincide. From now on consider only $N \in \mathbb{N}$ with $\mathcal{C}_{N,2} \neq \emptyset$. By the construction of the set $\tilde{\mathcal{C}}_N$ it follows that

$$\mathbb{E}[d(X,\tilde{\mathcal{C}}_N)^s] \le \mathbb{E}[d(X,\mathcal{C}_N)^s\, 1_{\{\|X\|\le c_N\}}] + \mathbb{E}[\|X\|^s\, 1_{\{\|X\|>c_N\}}].$$

Due to the above proposition,

$$\mathbb{E}[d(X,\tilde{\mathcal{C}}_N)^s] \lesssim \delta^{(q)}(N|\mu, \|\cdot\|^s) + \exp\Big\{-\frac{(1-\varepsilon')c_N^2}{2\sigma^2}\Big\} \quad (N \to \infty)$$

for any $\varepsilon' > 0$. Choose $\varepsilon' > 0$ such that $1+\varepsilon'' := (1-\varepsilon')(1+\varepsilon)^2 > 1$. Then

$$\begin{aligned} \exp\Big\{-\frac{(1-\varepsilon')c_N^2}{2\sigma^2}\Big\} &= \exp\{-(1+\varepsilon'')\log_- \delta^{(q)}(N|\mu, \|\cdot\|^s)\} \\ &\sim \delta^{(q)}(N|\mu, \|\cdot\|^s)^{1+\varepsilon''} = o(\delta^{(q)}(N|\mu, \|\cdot\|^s)), \end{aligned}$$

as $N \to \infty$. Hence, $\mathbb{E}[d(X,\tilde{\mathcal{C}}_N)^s] \sim \delta^{(q)}(N|\mu, \|\cdot\|^s)$. Since $\varepsilon > 0$ was arbitrary, the result follows from a diagonalization argument. □

Lemma 5.1.3. *Let $s > 0$ and $\{\mathcal{C}_N\}_{N\in\mathbb{N}}$ a sequence of codebooks in E. If $\lim_{N\to\infty} \mathbb{E}[d(X,\mathcal{C}_N)^s] = 0$, then $c_N := \max_{x\in\mathcal{C}_N} \|x\|$ satisfies*

$$c_N \gtrsim \sqrt{2\sigma^2 \log_- \mathbb{E}[d(X,\mathcal{C}_N)^s]} \text{ as } N \to \infty.$$

Proof. Let $\varepsilon > 0$ arbitrary and let $f \in B_{E'}(0,1)$ with

$$\tilde{\sigma} := (\mathbb{E}[f(X)^2])^{1/2} \geq (1-\varepsilon)\sigma.$$

Since $f : E \to \mathbb{R}$ is contractive, one has $d(x,\mathcal{C}_N) \geq d_{\mathbb{R}}(f(x),[-c_N,c_N])$ for any $N \in \mathbb{N}$ and $x \in E$. Therefore,

$$\begin{aligned}\mathbb{E}[d(X,\mathcal{C}_N)^s] &\geq \mathbb{E}[d_{\mathbb{R}}(f(X),[-c_N,c_N])^s]\\ &= 2\int_{c_N}^{\infty} (2\pi\tilde{\sigma}^2)^{-1/2}(x-c_N)^s e^{-x^2/(2\tilde{\sigma}^2)}\,dx \geq \Psi\Big(\frac{c_N+1}{\tilde{\sigma}}\Big).\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, it follows $\mathbb{E}[d(X,\mathcal{C}_N)^s] \geq \Psi(\frac{c_N+1}{\sigma})$. By assumption, $\lim_{N\to\infty} \mathbb{E}[d(X,\mathcal{C}_N)^s] = 0$ and, hence. c_N tends to infinity. Recall that $\Psi(t) \sim (2\pi)^{-1/2}\frac{1}{t}e^{-t^2/2}$ as $t \to \infty$. Consequently,

$$\log_- \mathbb{E}[d(X,\mathcal{C}_N)^s] \lesssim \frac{(c_N+1)^2}{2\sigma^2} \quad (N \to \infty)$$

and one obtains

$$c_N \gtrsim \sqrt{2\sigma^2 \log_- \mathbb{E}[d(X,\mathcal{C}_N)^s]}$$

as $N \to \infty$. □

5.2 Random versus deterministic codebooks

Fehringer's upper bound was derived via certain randomly generated codebooks. As seen before, his bound is of the correct weak asymptotic order in many cases (Theorem 3.1.2). A second example where random coding provides an even (asymptotically) optimal coding procedure, is the source coding theorem (see Section 1.5). Motivated by these results we pose the question in how far random codebooks can compete with deterministic codebooks in the high resolution quantization problem.

In the sequel, we are concerned with quantization based on randomly generated codebooks. For $N \in \mathbb{N}$ and $r \geq 0$, we denote the *quantization error with random codebooks* by

$$\delta^{(r)}(N|\mu,\rho) := \inf\Big\{\int_E \int_{E^N} \min_{i=1,\dots,N} \rho(x,y_i)\,d\nu^{\otimes N}(y_1,\dots,y_N)\,d\mu(x) : \\ \nu \in \mathcal{M}_1(E)\Big\}$$

and

$$D^{(r)}(r|\mu,\rho) := \inf\Big\{\int_E \int_{E^N} \min_{i=1,\dots,N} \rho(x,y_i)\, d\nu^{\otimes N}(y_1,\dots,y_N)\, d\mu(x) : \nu \in \mathcal{M}_1(E), N = \lfloor e^r \rfloor\Big\}.$$

Theorem 5.2.1. *Let $s > 0$. Suppose that $\delta^{(q)}(\,\cdot\,|\mu, \|\cdot\|^s)$ is slowly varying and that*

$$\delta^{(r)}(N|\mu, \|\cdot\|^s) \approx \delta^{(q)}(N|\mu, \|\cdot\|^s) \tag{5.1}$$

as $N \to \infty$. Then

$$\delta^{(r)}(N|\mu, \|\cdot\|^s) \sim \delta^{(q)}(N|\mu, \|\cdot\|^s) \tag{5.2}$$

as $N \to \infty$.

Remark 5.2.2. We fix $s \geq 1$ and suppose that the small ball function satisfies

$$\varphi^{-1}(2r) \approx \varphi^{-1}(r) \quad (r \to \infty).$$

Due to Theorem 3.5.2, one has

$$D^{(q)}(r|\mu, \|\cdot\|^s)^{1/s} \approx D^{(R)}(r|\mu, \|\cdot\|^s)^{1/s} \approx \varphi^{-1}(r) \quad (r \to \infty).$$

Therefore, $\delta^{(q)}(N|\mu, \|\cdot\|^s) \approx \delta^{(r)}(N|\mu, \|\cdot\|^s)$ $(r \to \infty)$. To apply the above theorem, it remains to verify that $\delta^{(q)}(\,\cdot\,|\mu, \|\cdot\|^s)$ is slowly varying. Unfortunately, it is not in the scope of the perturbation result for the quantization error (Lemma 4.3.3) to imply this property.

More can be done, if we assume additionally that the asymptotics of the entropy coding error and the quantization error coincide. In that case the perturbation result for entropy coding (Lemma 4.2.4) yields that, for $\kappa > 0$, one has

$$\begin{aligned}\delta^{(q)}(e^r|\mu, \|\cdot\|^s) &\sim D^{(e)}(r|\mu, \|\cdot\|^s)\\ &\sim D^{(e)}(r+\kappa|\mu, \|\cdot\|^s)\\ &\sim \delta^{(q)}(e^\kappa e^r|\mu, \|\cdot\|^s) \quad (r \to \infty).\end{aligned}$$

Consequently, $\delta^{(q)}(\,\cdot\,|\mu, \|\cdot\|^s)$ is slowly varying and one has strong equivalence of $\delta^{(q)}(\,\cdot\,|\mu, \|\cdot\|^s)$ and $\delta^{(r)}(\,\cdot\,|\mu, \|\cdot\|^s)$ due to Theorem 5.2.1.

Later in this thesis we are concerned with asymptotic relations between different coding quantities. From the discussion above, it follows that if $D^{(e)}(r) \sim D^{(q)}(r)$, then this equivalence may also be proven by using random codebooks under certain regularity conditions.

Proof of Theorem 5.2.1. We abbreviate $\delta^{(q)}(N) = \delta^{(q)}(N|\mu, \|\cdot\|^s)$ and $\delta^{(r)}(N) = \delta^{(r)}(N|\mu, \|\cdot\|^s)$. In order to prove the result, we construct asymptotically optimal random codebooks. By assumption, there exists a sequence $\{\nu_N\}_{N\in\mathbb{N}}$ of probability measures on E and a constant $\eta \geq 1$ such that

$$\int_E \int_{E^N} \min_{i=1,\dots,N} \|x - y_i\|^s \, d\nu_N^{\otimes N}(y_1, \dots, y_N) \, d\mu(x) \lesssim \eta \, \delta^{(q)}(N)$$

as $N \in \mathbb{N}$. Moreover, by Lemma 5.1.1, there exists a sequence of codebooks $\{\mathcal{C}_N\}_{N\in\mathbb{N}}$ in E satisfying $|\mathcal{C}_N| = N$, $\max_{x\in\mathcal{C}_N} \|x\| \lesssim \sqrt{8\sigma^2 \log_- \delta^{(q)}(N, s)}$ and

$$\mathbb{E}[d(X, \mathcal{C}_N)^s] \lesssim \delta^{(q)}(N) \text{ as } N \to \infty.$$

For fixed $\kappa \in \mathbb{N}$, we consider random codebooks $\tilde{\mathcal{C}}_N$, $N \in \mathbb{N}$, of size $(\kappa+1)N$ generated by i.i.d. samples with law

$$\tilde{\nu}_N := \frac{\kappa}{\kappa+1} \frac{1}{N} \sum_{x\in\mathcal{C}_N} \delta_x + \frac{1}{\kappa+1} \nu_{\lfloor \frac{N}{2} \rfloor}.$$

Let $\{Z_i^{(N)}\}_{i\in\mathbb{N}}$ and $\{\tilde{Z}_i^{(N)}\}_{i\in\mathbb{N}}$ denote independent sequences of i.i.d. r.e.'s with laws $\frac{1}{N}\sum_{x\in\mathcal{C}_N} \delta_x$ and $\nu_{\lfloor \frac{N}{2} \rfloor}$, respectively. Moreover, assume that $M = M(N)$ is an independent binomial r.v. to the parameters $\frac{1}{\kappa+1}$ and $(\kappa+1)N$ and set for $N \in \mathbb{N}$

$$\tilde{\mathcal{C}}_{N,1} = \{Z_1^{(N)}, \dots, Z_{(\kappa+1)N-M}^{(N)}\} \text{ and } \tilde{\mathcal{C}}_{N,2} = \{\tilde{Z}_1^{(N)}, \dots, \tilde{Z}_M^{(N)}\}.$$

Note that $\tilde{\mathcal{C}}_N := \tilde{\mathcal{C}}_{N,1} \cup \tilde{\mathcal{C}}_{N,2}$ is a codebook constituted by $(\kappa+1)N$ $\tilde{\nu}_N$-distributed r.e.'s. Let $\pi_N(\cdot) : E \to \mathcal{C}_N$, $N \in \mathbb{N}$, be a measurable map with

$$\mathbb{E}[d(X, \mathcal{C}_N)^s] = \mathbb{E}[d(X, \pi_N(X))^s].$$

Then

$$\begin{aligned}
\mathbb{E}[d(X, \tilde{\mathcal{C}}_N)^s] &\leq \mathbb{E}[1_{\tilde{\mathcal{C}}_{N,1}}(\pi_N(X)) \, d(X, \pi_N(X))^s] \\
&\quad + \mathbb{E}[1_{(\tilde{\mathcal{C}}_{N,1})^c}(\pi_N(X)) \, 1_{\{M \geq \lfloor N/2 \rfloor\}} \, d(X, \tilde{\mathcal{C}}_{N,2})^s] \\
&\quad + \mathbb{E}[1_{\{M < \lfloor N/2 \rfloor\}} \, d(X, \tilde{\mathcal{C}}_{N,1})^s] \\
&=: I_1(N) + I_2(N) + I_3(N).
\end{aligned}$$

It remains to find appropriate bounds for $I_1(N)$, $I_2(N)$ and $I_3(N)$. Clearly, $I_1(N) \leq \mathbb{E}[d(X, \mathcal{C}_N)^s] \lesssim \delta^{(q)}(N) \sim \delta^{(q)}((\kappa+1)N)$ $(N \to \infty)$. We control the second summand by

$$I_2(N) \leq \mathbb{P}(\pi_N(X) \notin \tilde{\mathcal{C}}_{N,1}) \, \mathbb{E}[\min_{i=1,\dots,\lfloor \frac{N}{2} \rfloor} d(X, \tilde{Z}_i^{(N)})^s].$$

It follows from the construction of $\tilde{\mathcal{C}}_{N,1}$ that

$$\mathbb{P}(\pi_N(X) \notin \tilde{\mathcal{C}}_{N,1}) = \Big(1 - \frac{\kappa}{(\kappa+1)N}\Big)^{(\kappa+1)N} \leq e^{-\kappa}.$$

Observe that

$$\mathbb{E}[\min_{i=1,\dots,\lfloor \frac{N}{2} \rfloor} d(X, \tilde{Z}_i^{(N)})^s] \lesssim \eta\, \delta^{(q)}(\lfloor N/2 \rfloor) \sim \eta\, \delta^{(q)}((\kappa+1)N).$$

Consequently,

$$I_2(N) \lesssim \eta\, e^{-\kappa}\, \delta^{(q)}((\kappa+1)N).$$

Expression $I_3(N)$ is bounded by

$$\begin{aligned} I_3(N) &\leq \mathbb{P}(M < \lfloor N/2 \rfloor)\, \mathbb{E}[(\|X\| + \max_{x \in \mathcal{C}_N} \|x\|)^s] \\ &\leq 2^s\, \mathbb{P}(M < \lfloor N/2 \rfloor)\, \big(\mathbb{E}[\|X\|^s] + \max_{x \in \mathcal{C}_N} \|x\|^s\big) \\ &\lesssim 2^s\, \mathbb{P}(M < \lfloor N/2 \rfloor) \sqrt{8\sigma^2 \log_-\, \delta^{(q)}(N)}. \end{aligned}$$

By Cramér's Theorem, $\mathbb{P}(M < \lfloor N/2 \rfloor)$ converges exponentially fast to 0 as $N \to \infty$. Since $\delta^{(q)}(N)$ and, hence, $\sqrt{8\sigma^2 \log_-\, \delta^{(q)}(N)}$ are slowly varying, one has $I_3(N) = o(\delta^{(q)}(N))$ as $N \to \infty$. Combining the above estimates yields

$$\mathbb{E}[d(X, \tilde{\mathcal{C}}_N)^s] \lesssim (1 + \eta\, e^{-\kappa})\, \delta^{(q)}((\kappa+1)N) \quad (N \to \infty).$$

Since $\delta^{(q)}$ is slowly varying, one has

$$\delta^{(r)}(N) \lesssim (1 + \eta\, e^{-\kappa})\, \delta^{(q)}(N)$$

as $N \to \infty$. Letting $\kappa \to \infty$ completes the proof. □

Chapter 6

Coding Gaussian measures on Hilbert spaces

This chapter is devoted to the study of the high resolution coding problem for Gaussian measures on Hilbert spaces under norm-based distortions.

6.1 Introduction and known results

Let μ be a centered Gaussian measure on a separable real Hilbert space $(H, \langle \cdot , \cdot \rangle)$ and denote by X a μ-distributed r.e. We denote by $\| \cdot \|$ the norm associated to the scalar product $\langle \cdot , \cdot \rangle$. We concentrate on measures μ that have infinite dimensional support.

Let $\tilde{C}_\mu : H' \to H$ denote the covariance operator of μ. By the Riesz representation theorem the map

$$i : H \to H',\ f \mapsto \langle f, \cdot \rangle$$

is an isometric isomorphism between H and its dual space H'. Consider now

$$C_\mu = \tilde{C}_\mu \circ i : H \to H.$$

In the case that the underlying space is a Hilbert space it is convenient to call C_μ the covariance operator of μ. In particular when $H = \mathbb{R}^d$, we obtain the covariance matrix of the underlying Gaussian measure. For $f, g \in H$ define

$$\begin{aligned} f \otimes g : H &\to H, \\ h &\mapsto f \otimes g(h) = i(g)(h) \cdot f = \langle g, h \rangle f. \end{aligned}$$

Since

$$\langle f, C_\mu g \rangle = \mathbb{E}[i(f)X \cdot i(g)X] = \langle C_\mu f, g \rangle,$$

it follows that C_μ is self-adjoint. Taking $f = g$ in the previous equation shows that C_μ is positive semidefinite. Furthermore, C_μ is compact, since $\tilde{C}_\mu$ is compact. By the spectral theorem for compact operators, there exists an at most countable index set I and an orthonormal system of eigenfunctions $\{e_j\}_{j\in I}$ corresponding to the strictly positive eigenvalues $\{\lambda_j\}_{j\in I}$ such that

$$C_\mu = \sum_{j\in I} \lambda_j \cdot e_j \otimes e_j,$$

where convergence occurs in the operator norm. For each $j \in I$, we denote

$$X_j = \frac{1}{\sqrt{\lambda_j}} \langle e_j, X\rangle.$$

Since $\langle e_i, C_\mu e_j\rangle = \delta_{ij}\lambda_j$ for $i, j \in I$ (here δ denotes the Kronecker-delta), the sequence $\{X_j\}_{j\in\mathbb{N}}$ is a sequence of independent standard normals. We can extend the system $\{e_j\}_{j\in I}$ by countably many orthonormal elements $\{f_j\}_{j\in J}$ (J countable index set) such that the union forms a complete orthonormal system of H. Then $\langle f_j, X\rangle = 0$ for all $j \in J$ almost surely. Hence, one has a.s.

$$X = \sum_{j\in I} \sqrt{\lambda_j} X_j e_j. \tag{6.1}$$

This representation is called *Karhunen-Loève expansion*. Without changing the distribution of X we will assume strict equality in equation (6.1).

In the sequel, we assume that μ has infinite dimensional support. Hence, the index set I is infinite and countable and we assume without loss of generality that $I = \mathbb{N}$. Since C_μ is a compact operator, there exist only finitely many eigenvalues greater than ε for arbitrary $\varepsilon > 0$. Therefore, we can assume without loss of generality that the eigenvalues are ordered decreasingly in j.

The distortion rate function $D(r|\mu, \|\cdot\|^2)$ for squared norm distortion was derived by Kolmogorov in 1956 [35]. Consider the following system of equations for $d \in (0, \|\lambda\|_{l_1}]$, $d_c \in (0, \lambda_1]$ and $r \in [0, \infty)$

$$\begin{cases} d = \sum_j d_c \wedge \lambda_j \\ r = \sum_j \frac{1}{2} \log_+(\frac{\lambda_j}{d_c}). \end{cases} \tag{6.2}$$

Here, we denote $\log_+ x = 0 \vee \log x$ for $x > 0$. When we fix one parameter d, d_c or r in the allowed domains, the system (6.2) possesses a unique solution (d, d_c, r) with the determined value in the fixed parameter. Hence, the system induces unique maps

$$d : [0, \infty) \to (0, \|\lambda\|_{l_1}],\ r \mapsto d(r) \text{ and}$$
$$d_c : [0, \infty) \to (0, \lambda_1],\ r \mapsto d_c(r).$$

such that $(d(r), d_c(r), r)$ solves the system (6.2) for any $r \geq 0$. Then it holds

$$D(r|\mu, \|\cdot\|^2) = d(r), \quad r \geq 0.$$

This result can be found also in Cover and Thomas [13] (Theorem 13.3.3) and Ihara [32] (Theorem 6.9.1).

Recall that the SCT is typically proven via a random coding argument. There the codebook is constituted by a sequence of i.i.d. random elements. In the case where the information source is $(\mu, \|\cdot\|^2)$, the optimal choice for the underlying distribution is known explicitly. It is centered Gaussian in H with covariance operator

$$\sum_{j\in\mathbb{N}} \tilde{\lambda}_j(r) \cdot e_j \otimes e_j,$$

where

$$\tilde{\lambda}_j(r) = (\lambda_j - d_c(r)) \vee 0, \quad j \in \mathbb{N}.$$

Note that $\tilde{\lambda}_j(r) = 0$ for all $j > N(r) := \sup\{i \in \mathbb{N} : \lambda_i > d_c\}$. Here the supremum of the empty set is assumed to be 0. In the following our aim is to use similar random codebooks in order to obtain result on the high resolution quantization problem.

The relation between the functions d, d_c, N and $\tilde{\lambda}$ is illustrated in Figure 6.1. This link is called *inverse water filling principle* in the literature. Similar principles play an important role, for instance, in the theory of the capacity of Gaussian channels. First results of that kind were obtained by Shannon (1949, [61]) and Pinsker (1954, [56]; 1956, [57]). See also Chapters 5 and 6 of Ihara [32] and the references therein.

A close relation between the high resolution quantization problem and the distortion rate function on infinite dimensional Hilbert spaces was pointed out in the introduction of a technical report by Donoho [20]. There it is stated that

$$D^{(q)}(r|\mu, \|\cdot\|^2) \sim D(r|\mu, \|\cdot\|^2) \quad (r \to \infty), \tag{6.3}$$

if the eigenvalues of μ decrease polynomially, i.e. there exist $\kappa > 1$ and $c > 0$ such that

$$\lambda_j \sim c\, j^{-\kappa} \quad (j \to \infty).$$

In Luschgy and Pagès [51], the latter equivalence is proven for regularly varying eigenvalues, i.e. there exist $\kappa \geq 1$ and a slowly varying function $l : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\lambda_j \sim j^{-\kappa} l(j) \quad (j \to \infty).$$

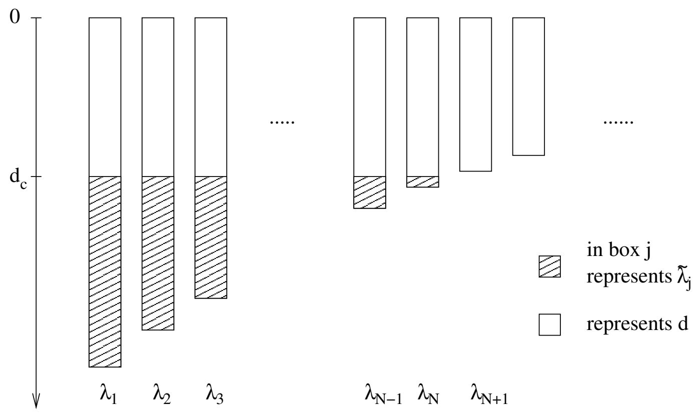

Figure 6.1: Inverse water filling principle

Donoho's report does not contain a proof of his assertion. Nevertheless, one of the key techniques used in the article is a *subband decomposition* which was used later by Luschgy and Pagès to prove equivalence (6.3).

Subband coding

In the sequel, we shortly describe the ideas of subband coding and give a sketch of the proof of equivalence (6.3). Denote by l_2 the Hilbert space of square summable real-valued sequences. The map

$$\begin{aligned} \pi : H &\to l_2 \\ x &\mapsto \{\langle e_j, x\rangle\}_{j\in\mathbb{N}} \end{aligned}$$

is a contraction. Moreover, π restricted to the linear subspace $H_0 = \mathrm{supp}(\mu)$ is an isometric isomorphism into l_2. It follows that

$$D^{(q)}(r|X, \|\cdot\|^s) = D^{(q)}(r|\pi(X), \|\cdot\|_{l_2}^s).$$

The analog statement holds also for the coding quantities $D^{(e)}$ and D. Note that $\pi(X) = \{\sqrt{\lambda_j}X_j\}_{j\in\mathbb{N}}$. Consequently, we may assume without loss of generality that $X = \{\sqrt{\lambda_j}X_j\}_{j\in\mathbb{N}}$ and that the underlying Hilbert space is l_2. In particular, we can and will assume that $\mathrm{supp}(\mu) = H$.

We divide the ordered sequence of eigenvalues $\{\lambda_j\}_{j\in\mathbb{N}}$ into subbands

$$\{\lambda_j\}_{j\in I^{(m)}}, \quad m \in \mathbb{N},$$

where $I^{(m)} = \{k_{m-1}, \dots, k_m - 1\}$ and $\{k_m\}_{m\in\mathbb{N}_0}$ is a monotonically increasing sequence in $\mathbb{N}$ with $k_0 = 1$. We assume that the subband decomposition satisfies

- $\lim_{m\to\infty} k_m - k_{m-1} = \infty$ and
- $\lim_{m\to\infty} \frac{\lambda_{k_m - 1}}{\lambda_{k_{m-1}}} = 1$.

An appropriate decomposition can be found, for instance, if the eigenvalues are regularly varying. Note that the eigenvalues in each block become more and more similar and that the block size tends to ∞. Set

$$X^{(m)} := \{\sqrt{\lambda_j}\, X_j\}_{j\in I^{(m)}}, \quad m \in \mathbb{N}.$$

Due to the SCT, one expects that, for large $m \in \mathbb{N}$, $D^{(q)}(r|X^{(m)}, |\cdot|^2)$ is close to $D(r|X^{(m)}, |\cdot|^2)$. Here, $|\cdot|$ denotes Euclidean norm on the appropriate finite dimensional spaces. As a consequence of Luschgy and Pagès ([51], Proposition 4.4) it follows that

$$\kappa(m) := \sup_{\substack{r\geq 0,\\ \sigma^2>0}} \frac{D^{(q)}(r|\mathcal{N}(0,\sigma^2)^{\otimes m}, |\cdot|^2)}{D(r|\mathcal{N}(0,\sigma^2)^{\otimes m}, |\cdot|^2)} \tag{6.4}$$

is finite for all $m \in \mathbb{N}$ and satisfies $\lim_{m\to\infty} \kappa(m) = 1$. It remains to combine this result with standard estimates.

For fixed $r \geq 0$, we let

$$r_m = r_m(r) = \frac{1}{2} \sum_{j\in I^{(m)}} \log_+ \frac{\lambda_j}{d_c(r)}, \quad m \in \mathbb{N}.$$

Then $\sum_{m=1}^{\infty} r_m = r$ and

$$D(r|X, \|\cdot\|_{l_2}^2) = \sum_{m=1}^{\infty} D(r_m|X^{(m)}, |\cdot|^2).$$

We estimate

$$\begin{aligned}
D^{(q)}(r|X, \|\cdot\|_{l_2}^2) &\overset{(a)}{\leq} \sum_{m=1}^{\infty} D^{(q)}(r_m|X^{(m)}, |\cdot|^2) \\
&\overset{(b)}{\leq} \sum_{m=1}^{\infty} D^{(q)}(r_m|\mathcal{N}(0,\lambda_{k_{m-1}})^{\otimes|I^{(m)}|}, |\cdot|^2) \\
&\overset{(c)}{\leq} \sum_{m=1}^{\infty} \kappa(|I^{(m)}|)\, D(r_m|\mathcal{N}(0,\lambda_{k_{m-1}})^{\otimes|I^{(m)}|}, |\cdot|^2) \\
&\overset{(d)}{=} \sum_{m=1}^{\infty} \kappa(|I^{(m)}|)\, \frac{\lambda_{k_{m-1}}}{\lambda_{k_m-1}}\, D(r_m|\mathcal{N}(0,\lambda_{k_m-1})^{\otimes|I^{(m)}|}, |\cdot|^2) \\
&\overset{(e)}{\leq} \sum_{m=1}^{\infty} \kappa(|I^{(m)}|)\, \frac{\lambda_{k_{m-1}}}{\lambda_{k_m-1}}\, D(r_m|X^{(m)}, |\cdot|^2).
\end{aligned} \tag{6.5}$$

Inequality (a) follows by considering a product codebook which assigns rate r_m to each block $X^{(m)}$. Inequalities (b) and (e) hold due to the fact that blowing up the variances in the normal distribution increases the quantization error and the distortion rate function. Inequality (c) holds by definition of $\kappa(\cdot)$ in (6.4), and equality (d) holds due to scaling properties of the normal distribution.

In (6.5), the coefficients of $D(\cdot)$ converge to 1 as $m \to \infty$. It can be shown that finitely many coordinates (or blocks) have no influence on the asymptotic behavior of the quantization error and may be neglected. Then one concludes

$$D^{(q)}(r|X, \|\cdot\|_{l_2}^2) \lesssim \sum_{m=1}^{\infty} D(r_m|X^{(m)}, |\cdot|^2) = D(r|X, \|\cdot\|_{l_2}^2)$$

as $r \to \infty$.

6.2 The high resolution quantization problem

The known proofs of the equivalence between the DRF and the quantization error are based on a subband decomposition. It is not known whether the equivalence holds if it is not possible to find a subband decomposition that fulfills the required assumptions. For instance, it is typically not possible to find such a decomposition if the eigenvalues are decaying exponentially or faster than exponentially. Or the eigenvalues may decrease very irregularly so that there does not exist an appropriate decomposition. A second open problem is to find the asymptotics of the quantization error if the distortion measure is of the form $\|\cdot\|^s$ where $s > 0$ is not equal to 2.

The aim of this section is to prove

Theorem 6.2.1. *Assume that*

$$\lim_{n\to\infty} \frac{\log\log(1/\lambda_n)}{n} = 0.$$

Then, for any $s \in (0, \infty)$,

$$D^{(q)}(r|\mu, \|\cdot\|^s)^{1/s} \sim D(r|\mu, \|\cdot\|^2)^{1/2}$$

as $r \to \infty$.

As a consequence of the Hölder inequality, one has

$$D(r|\mu, \|\cdot\|^s)^{1/s} \geq D(r|\mu, \|\cdot\|^2)^{1/2}$$

for any $s \geq 2$ and $r \geq 0$. Hence, Theorem 6.2.1 implies

Corollary 6.2.2. *Suppose that*

$$\lim_{n\to\infty} \frac{\log\log(1/\lambda_n)}{n} = 0.$$

Then, for any $s \in [2,\infty)$,

$$D^{(e)}(r|\mu, \|\cdot\|^s)^{1/s} \sim D(r|\mu, \|\cdot\|^s)^{1/s} \sim D(r|\mu, \|\cdot\|^2)^{1/2}$$

as $r \to \infty$.

Theorem 6.2.1 yields strong equivalence of the moments in the asymptotic quantization problem. This indicates that, for "good" codebooks $\mathcal{C}$ of high rate, the (random) quantization error $d(X, \mathcal{C})$ is concentrated at some typical value.

We maintain the definitions for X, $d(\cdot), d_c(\cdot), N(\cdot)$ and $\tilde\lambda_j(\cdot)$ of the introduction throughout the whole chapter unless stated otherwise. Thereafter, $r \ge 0$ denotes the (approximate) rate of the coding strategy considered. We will always evaluate the latter functions at r. To simplify notations we will omit the parameter r. We write d, d_c, N and $\tilde\lambda_j$ when we mean $d(r), d_c(r), N(r)$ and $\tilde\lambda_j(r)$, respectively.

In the following, we consider the efficiency of certain randomly generated codebooks. The underlying codebook distribution is chosen in analogy to the source coding theorem. The codebook depends on two parameters $r \ge 0$ and $\delta > 0$. Since we keep δ fixed most of the time, we omit the parameter in the notations. Set

$$\mathcal{C}_r = \mathcal{C}_r^{(\delta)} = \{\tilde X_i^{(r)} : i = 1, \ldots, \lceil e^{r + 2\delta \frac{d}{d_c}} \rceil\}, \tag{6.6}$$

where the elements are i.i.d. centered Gaussian r.e.'s in H with covariance operator

$$\sum_{j\in\mathbb{N}} \tilde\lambda_j \cdot e_j \otimes e_j.$$

We need a number of propositions for the proof of Theorem 6.2.1. For ease of notation, we abbreviate $D(r) = D(r|\mu, \|\cdot\|^2)$, $r \ge 0$.

Proposition 6.2.3. *The function* $D(\cdot)$ *is differentiable on* $[0,\infty)$ *with*

$$\frac{\partial}{\partial r} D(r) = -2 d_c(r).$$

Proof. Let

$$\tilde d : (0, \lambda_1] \to (0, \|\lambda\|_{l_1}],\ x \mapsto \sum_{j\in\mathbb{N}} \lambda_j \wedge x \text{ and}$$

$$\tilde r : (0, \lambda_1] \to \mathbb{R}_+,\ x \mapsto \frac{1}{2} \sum_{j\in\mathbb{N}} \log_+ \frac{\lambda_j}{x}.$$

Then for any $x \in (0, \lambda_1]$, $(\tilde{d}(x), x, \tilde{r}(x))$ solves (6.2) and, hence, one has $D(r) = \tilde{d}(\tilde{r}^{-1}(r))$ for any $r \geq 0$. Furthermore, $\tilde{d}$ and $\tilde{r}$ are differentiable on $\mathcal{D} = (0, \lambda_1] \backslash \{\lambda_j : j \in \mathbb{N}\}$ with

$$\frac{\partial}{\partial x}\,\tilde{d}(x) = N(x) \quad \text{ and } \quad \frac{\partial}{\partial x}\,\tilde{r}(x) = -\frac{N(x)}{2x},$$

where $N(x) := \sup\{i \in \mathbb{N} : \lambda_i > x\}$, $x \in \mathbb{R}_+$, and the supremum of the empty set is 0. Note that $\tilde{r}|_{\mathcal{D}}$ is a diffeomorphism on its image. Using the chain rule and the inversion formula we obtain for any $r \in \tilde{r}(\mathcal{D})$

$$\frac{\partial}{\partial r}\,D(r) = \frac{\partial \tilde{d}}{\partial x}\,(d_c(r))\Big(\frac{\partial \tilde{r}}{\partial x}\,(d_c(r))\Big)^{-1} = -N(d_c(r))\,\frac{2d_c(r)}{N(d_c(r))} = -2d_c(r).$$

Since $D(\cdot)$ is continuous and the derivative $\frac{\partial}{\partial r}\,D(\cdot)$ can be continuously extended onto $\overline{\tilde{r}(\mathcal{D})} = [0, \infty)$, the function $D(\cdot)$ is differentiable on $[0, \infty)$ with

$$\frac{\partial}{\partial r}\,D(r) = -2d_c(r).$$

□

Corollary 6.2.4. *One has for $\delta \in (0, \frac{1}{2})$ and $r \geq 0$,*

$$D\Big(r + \delta\frac{d}{d_c}\Big) \geq (1 - 2\delta)\,d.$$

Proof. By the convexity of D and Proposition 6.2.3, one has

$$D\Big(r + \delta\frac{d}{d_c}\Big) \geq D(r) + \frac{\partial}{\partial r}D(r)\delta\frac{d}{d_c} = (1 - 2\delta)d.$$

□

For the parameter $\delta > 0$, we let $\mathcal{T}(r) = \mathcal{T}^{(\delta)}(r)$ be the event

$$\mathcal{T}(r) = \big\{\min_{\hat{x} \in \mathcal{C}_r^{(\delta)}} \|X - \hat{x}\|^2 \leq (1 + \delta)d\big\}, \quad r \geq 0. \tag{6.7}$$

We will show that the event occurs almost with probability 1 when r is large. The proof is based on the link between SBPs and certain moment generating functions given in Theorem 3.4.1 and Corollary 3.4.3. For $\theta \leq 0$, $x \in H$ and $r \geq 0$, set

$$\Lambda_x(\theta|r) = \log \mathbb{E}[e^{\theta\|\tilde{X}_1^{(r)} - x\|^2}]$$

and

$$\Lambda'_x(\theta|r) = \frac{\partial}{\partial \theta}\,\Lambda_x(\theta|r).$$

Proposition 6.2.5. *Let* $x = \sum_{j\in\mathbb{N}} \sqrt{\lambda_j} x_j e_j \in H$. *Then, for* $\theta \le 0$ *and* $r \ge 0$,

$$\Lambda_x(\theta|r) = \sum_{j\in\mathbb{N}} \Big[-\frac{1}{2}\log(1-2\theta\tilde{\lambda}_j) + \frac{\theta\lambda_j}{1-2\theta\tilde{\lambda}_j}\, x_j^2 \Big] \tag{6.8}$$

$$\Lambda_x'(\theta|r) = \sum_{j\in\mathbb{N}} \Big[\frac{\tilde{\lambda}_j}{1-2\theta\tilde{\lambda}_j} + \frac{\lambda_j}{(1-2\theta\tilde{\lambda}_j)^2}\, x_j^2 \Big]. \tag{6.9}$$

Proof. Let $r \ge 0$ and let

$$Z = \sum_{j\in\mathbb{N}} \sqrt{\tilde{\lambda}_j} Z_j e_j,$$

where $\{Z_j\}_{j\in\mathbb{N}}$ is a sequence of i.i.d. standard normals. Then Z is centered Gaussian with covariance operator $\sum_{j\in\mathbb{N}} \tilde{\lambda}_j\, e_j \otimes e_j$. Hence,

$$\mathcal{L}(\tilde{X}_1^{(r)}) = \mathcal{L}(Z).$$

Let $x = \sum_{j\in\mathbb{N}} \sqrt{\lambda_j} x_j e_j \in H$ and $\theta \le 0$. Then

$$\begin{aligned}\Lambda_x(\theta|r) &= \log \mathbb{E}\big[\exp\{\sum_{j\in\mathbb{N}} \theta\big(\sqrt{\tilde{\lambda}_j} Z_j - \sqrt{\lambda_j} x_j\big)^2\}\big] \\ &= \sum_{j=1}^{N(r)} \log \mathbb{E}\big[\exp\{\theta\tilde{\lambda}_j\big(Z_j - \sqrt{\tfrac{\lambda_j}{\tilde{\lambda}_j}} x_j\big)^2\}\big] + \sum_{N(r)+1}^{\infty} \theta\lambda_j x_j^2,\end{aligned}$$

where we have used the independence of the random variables Z_j, $j \in \mathbb{N}$. Note that, for any $z_1 \in \mathbb{R}$ and $\theta \le 0$,

$$\log \mathbb{E}[e^{\theta(Z_1-z_1)^2}] = -\frac{1}{2}\log(1-2\theta) + \frac{\theta}{1-2\theta}\, z_1^2.$$

Hence,

$$\Lambda_x(\theta|r) = \sum_{j=1}^{N(r)} \Big[-\frac{1}{2}\log(1-2\theta\tilde{\lambda}_j) + \frac{\theta\lambda_j}{1-2\theta\tilde{\lambda}_j}\, x_j^2 \Big] + \sum_{j=N(r)+1}^{\infty} \theta\lambda_j x_j^2$$

and the first statement follows. Equation (6.9) is obtained immediately from (6.8). □

For convenience, we omit the parameter r in the expressions $\Lambda_x(\theta|r)$ and $\Lambda_x'(\theta|r)$. For instance, we write $\Lambda_x(\theta)$ when we mean $\Lambda_x(\theta|r)$.

Proposition 6.2.6. *Let $r \geq 0$ and*

$$\theta_0(r) = -\frac{1}{2d_c(r)}.$$

One has

$$\Lambda'_X(\theta_0(r)) = d + d_c\Big[\sum_{j=1}^{N} \frac{d_c}{\lambda_j}(X_j^2 - 1) + \sum_{N+1}^{\infty} \frac{\lambda_j}{d_c}(X_j^2 - 1)\Big] \quad \textit{and}$$

$$\Lambda^*_X(\Lambda'_X(\theta_0(r))) = r + \sum_{j=1}^{N} \frac{\tilde{\lambda}_j}{2\lambda_j}(X_j^2 - 1).$$

Proof. By the representation given in Proposition 6.2.5 and straightforward calculations, one has

$$\begin{aligned}
\Lambda'_X(\theta_0(r)) &= \sum_{j\in\mathbb{N}}\Big[\frac{\tilde{\lambda}_j}{1+\tilde{\lambda}_j/d_c} + \frac{\lambda_j}{(1+\tilde{\lambda}_j/d_c)^2}\, X_j^2\Big] \\
&= \sum_{j=1}^{N(r)}\Big[\frac{d_c\tilde{\lambda}_j}{\lambda_j} + \frac{d_c^2\lambda_j}{\lambda_j^2}\, X_j^2\Big] + \sum_{j=N(r)+1}^{\infty} \lambda_j X_j^2 \\
&= \sum_{j=1}^{N(r)}\Big[d_c + \frac{d_c^2}{\lambda_j}\,(X_j^2 - 1)\Big] + \sum_{j=N(r)+1}^{\infty}\Big[\lambda_j + \lambda_j(X_j^2 - 1)\Big] \\
&= d + d_c\Big[\sum_{j=1}^{N(r)} \frac{d_c}{\lambda_j}\,(X_j^2 - 1) + \sum_{j=N(r)+1}^{\infty} \frac{\lambda_j}{d_c}(X_j^2 - 1)\Big].
\end{aligned}$$

Hence, $\mathbb{E}[\Lambda'_X(\theta_0)] = d$. Note that

$$\begin{aligned}
\Lambda_X(\theta_0(r)) &= -\sum_{j=1}^{\infty} \frac{1}{2}\log(1+\tilde{\lambda}_j/d_c) - \sum_{j=1}^{N(r)}\Big[\frac{1}{2} + \frac{1}{2}(X_j^2 - 1)\Big] \\
&\quad - \sum_{N(r)+1}^{\infty}\Big[\frac{\lambda_j}{2d_c} + \frac{\lambda_j}{2d_c}\,(X_j^2 - 1)\Big] \\
&= -r - \frac{d}{2d_c} - \sum_{j=1}^{N(r)} \frac{1}{2}(X_j^2 - 1) - \sum_{N(r)+1}^{\infty} \frac{\lambda_j}{2d_c}\,(X_j^2 - 1).
\end{aligned}$$

Consequently,

$$\begin{aligned}\Lambda_X^*(\Lambda_X'(\theta_0(r))) &= \theta_0(r)\Lambda_X'(\theta_0(r)) - \Lambda_X(\theta_0(r)) \\ &= r + \sum_{j=1}^{N(r)} \frac{1}{2}(X_j^2 - 1) + \sum_{N(r)+1}^{\infty} \frac{\lambda_j}{2d_c}(X_j^2 - 1) \\ &\quad - \frac{1}{2}\Big[\sum_{j=1}^{N(r)} \frac{d_c}{\lambda_j}(X_j^2 - 1) + \sum_{j=N(r)+1}^{\infty} \frac{\lambda_j}{d_c}(X_j^2 - 1)\Big] \\ &= r + \sum_{j=1}^{N(r)} \frac{\tilde{\lambda}_j}{2\lambda_j}(X_j^2 - 1).\end{aligned}$$

□

We keep the function θ_0 as defined in the above lemma and abbreviate $\theta_0 = \theta_0(r)$.

To show mass concentration on $\mathcal{T}(r)$ we use the link between the moment generating function and small ball probabilities. More explicitly, we show that the random variables

$$\Lambda_X^*(\Lambda_X'(\theta_0)) \quad \text{and} \quad \Lambda_X'(\theta_0)$$

are typically close to their means r and d, respectively. Consequently, it follows that

$$-\log \mathbb{P}\big(\|X - \tilde{X}_1\|^2 \le (1+\delta)d \big| X\big) \le r + \frac{\delta d}{d_c}$$

with high probability. Consequently, a random codebook with $\lceil e^{r+2\delta d/d_c}\rceil$ elements contains a $(1+\delta)d$-close representation w.r.t. $\|\cdot\|^2$ for "most" realizations of X. Due to Proposition 6.2.6, the expressions of interest are weighted sums of i.i.d. random variables.

Lemma 6.2.7. *Let $\{Z_j\}_{j\in\mathbb{N}}$ be a sequence of zero-mean real-valued i.i.d. random variables in $L_2(\mathbb{P})$ and denote by Λ_{Z_1} the logarithmic moment generating function of Z_1,*

$$\Lambda_{Z_1}(\theta) = \log \mathbb{E}e^{\theta Z_1}, \quad \theta \in \mathbb{R}.$$

Let $\{a_j\}_{j\in\mathbb{N}} \in l_2$ be a sequence of positive numbers and set $a_{\max} = \max_{j\in\mathbb{N}} a_j$. Suppose $\eta > 0$ is such that Λ_{Z_1} is finite on $[0,\eta]$. Then, for each $t \ge 0$ and $\zeta \in [0, \eta/a_{\max}]$, it holds

$$\log \mathbb{P}\Big(\sum_{j=1}^{\infty} a_j Z_j \ge t\Big) \le -\zeta\Big(t - \frac{1}{2}\zeta\xi\|a\|_{l_2}^2\Big), \tag{6.10}$$

where $\xi = \sup\{\Lambda_{Z_1}''(x) : x \in (0,\eta)\}$.

Proof. By the Markov inequality (exponential Chebyshev inequality), one has for every $\zeta \in [0, \eta/a_{\max}]$

$$\mathbb{P}\Big(\sum_{j=1}^{\infty} a_j Z_j \geq t\Big) \leq \frac{\mathbb{E}[\exp\{\sum_{j\in\mathbb{N}} a_j Z_j \zeta\}]}{\exp\{t\zeta\}}.$$

Since $\{Z_j\}_{j\in\mathbb{N}}$ is a sequence of i.i.d. random variables, we obtain

$$\log \mathbb{P}\Big(\sum_{j=1}^{\infty} a_j Z_j \geq t\Big) \leq \sum_{j\in\mathbb{N}} \Lambda_{Z_1}(a_j\zeta) - t\zeta. \tag{6.11}$$

We use Taylor's formula for estimating $\tilde{\Lambda} = \Lambda|_{[0,\eta]}$. Recall, that by assumption $\tilde{\Lambda}$ is finite. Hence, $\tilde{\Lambda}$ is continuous on $[0,\eta]$ and C^∞ on $(0,\eta)$. Note that $\tilde{\Lambda}(0) = 0$ and

$$\tilde{\Lambda}'(\theta) = \frac{\mathbb{E}[Z_1 e^{\theta Z_1}]}{\mathbb{E}[e^{\theta Z_1}]}, \quad \theta \in (0,\eta),$$

where the equality follows by interchanging differentiation and integration. By dominated convergence,

$$\lim_{\theta\downarrow 0} \tilde{\Lambda}'(\theta) = \mathbb{E}[Z_1],$$

and hence $\tilde{\Lambda}'(0) = 0$. Combining all results, we obtain by Taylor's formula

$$\tilde{\Lambda}(\theta) \leq \frac{1}{2}\xi\theta^2, \quad \theta \in [0,\eta].$$

With equation (6.11),

$$\log \mathbb{P}\Big(\sum_{j=1}^{\infty} a_j Z_j \geq t\Big) \leq \sum_{j=1}^{\infty} \frac{1}{2}\xi(a_j\zeta)^2 - t\zeta.$$

Rewriting this estimate yields (6.10). □

Proposition 6.2.8. *For all $\delta > 0$ there exists $\kappa > 0$ such that*

$$\mathbb{P}(\Lambda'_X(\theta_0) \geq d + \delta d) \leq e^{-\kappa\frac{d}{d_c}} \text{ and} \tag{6.12}$$

$$\mathbb{P}\Big(\Lambda^*_X(\Lambda'_X(\theta_0)) \geq r + \delta\frac{d}{d_c}\Big) \leq e^{-\kappa\frac{d}{d_c}} \tag{6.13}$$

for all $r \geq 0$.

Proof. 1.) We start with proving inequality (6.12). Recall that by Proposition 6.2.6,

$$\Lambda'_X(\theta_0) = d + d_c\Big[\sum_{j=1}^{N} \frac{d_c}{\lambda_j}(X_j^2 - 1) + \sum_{j=N+1}^{\infty} \frac{\lambda_j}{d_c}(X_j^2 - 1)\Big].$$

Denote

$$\psi_1(X) = \psi_1(X|r) = \sum_{j=1}^{N} \frac{d_c}{\lambda_j}(X_j^2 - 1) + \sum_{j=N+1}^{\infty} \frac{\lambda_j}{d_c}(X_j^2 - 1)$$

and

$$\Sigma_1(r) = \sum_{j=1}^{N} \Big(\frac{d_c}{\lambda_j}\Big)^2 + \sum_{j=N+1}^{\infty} \Big(\frac{\lambda_j}{d_c}\Big)^2.$$

Then

$$\mathbb{P}(\Lambda'_X(\theta_0) \geq d + \delta d) = \mathbb{P}\Big(\psi_1(X) \geq \delta \frac{d}{d_c}\Big).$$

Note that the coefficients in the term $\psi_1(X)$ are bounded by 1. We will use Lemma 6.2.7 to bound the probability of large deviations. Note that $\tilde{\Lambda}(\theta) = \log \mathbb{E}[e^{\theta(X_1^2-1)}]$ is finite on $(-\infty, 1/2)$. Fix $\eta \in (0, 1/2)$ and let

$$\xi = \sup\{\tilde{\Lambda}''(x) : x \in (0, \eta)\} < \infty.$$

Then

$$\log \mathbb{P}\Big(\psi_1(X) \geq \delta \frac{d}{d_c}\Big) \leq -\zeta\Big(\delta \frac{d}{d_c} - \frac{1}{2}\zeta\xi\Sigma_1(r)\Big)$$

for all $\zeta \in [0, \eta]$. With the estimate

$$\Sigma_1(r) \leq \frac{d}{d_c}$$

one obtains

$$\log \mathbb{P}\Big(\psi_1(X) \geq \delta \frac{d}{d_c}\Big) \leq -\zeta(\delta - \zeta\xi)\frac{d}{d_c}.$$

Choosing ζ sufficiently implies the existence of a constant $\kappa > 0$ such that

$$\mathbb{P}(\Lambda'_X(\theta_0) \geq d + \delta d) \leq e^{-\kappa \frac{d}{d_c}}$$

for all $r \geq 0$.
2.) We proceed with the proof of inequality (6.13). Recall that, by Proposition 6.2.6,

$$\Lambda^*_X(\Lambda'_X(\theta_0)) = r + \sum_{j=1}^{N} \frac{\tilde{\lambda}_j}{2\lambda_j}(X_j^2 - 1).$$

We estimate the probability of large deviations of

$$\psi_2(X) := \psi_2(X|r) := \sum_{j=1}^{N} \frac{\tilde{\lambda}_j}{2\lambda_j}(X_j^2 - 1)$$

with Lemma 6.2.7. Set

$$\Sigma_2(r) = \frac{1}{4}\sum_{j=1}^{N}\Big(\frac{\tilde{\lambda}_j}{\lambda_j}\Big)^2.$$

Then

$$\log \mathbb{P}\Big(\psi_2(X) \geq \delta\frac{d}{d_c}\Big) \leq -\zeta\Big(\delta\frac{d}{d_c} - \frac{1}{2}\zeta\xi\Sigma_2(r)\Big)$$

for $\zeta \in [0, 2\eta]$ where ξ and η are as above. Since $\Sigma_2(r) \leq \frac{d}{4d_c}$, it follows that

$$\log \mathbb{P}\Big(\psi_2(X) \geq \delta\frac{d}{d_c}\Big) \leq -\zeta(\delta - \zeta\xi)\frac{d}{d_c}.$$

Therefore,

$$\mathbb{P}\Big(\Lambda_X^*(\Lambda_X'(\theta_0)) \geq r + \delta\frac{d}{d_c}\Big) \leq e^{-\kappa\frac{d}{d_c}}$$

for all $r \geq 0$, where κ is as above. □

Remark 6.2.9. Analogously to the above proof, one can show that for all $\delta > 0$ there exists $\kappa > 0$ such that

$$\mathbb{P}(\Lambda_X'(\theta_0) \leq (1-\delta)d) \leq e^{-\kappa\frac{d}{d_c}} \text{ and}$$
$$\mathbb{P}\Big(\Lambda_X^*(\Lambda_X'(\theta_0)) \leq r - \delta\frac{d}{d_c}\Big) \leq e^{-\kappa\frac{d}{d_c}}.$$

Note that in general $\frac{d}{d_c} \geq N$. If the eigenvalues decay rapidly, then N is a good approximation for the term $\frac{d}{d_c}$. The estimates show that the random variables of interest are concentrated at their means and the tail probabilities decay at least exponentially fast in N.

The estimates of the previous proposition lead to an estimate of $\mathbb{P}(\mathcal{T}(r)^c)$.

Proposition 6.2.10. *Suppose that*

$$\log r = o\Big(\frac{d}{d_c}\Big) \quad (r \to \infty).$$

For $\delta > 0$, one has

$$\mathbb{P}(\mathcal{T}(r)^c) \lesssim 2e^{-\kappa\frac{d}{d_c}} \quad (r \to \infty),$$

where $\kappa = \kappa(\delta)$ is as in Proposition 6.2.8 and $\mathcal{T}(r) = \mathcal{T}^{(\delta)}(r)$ is as in (6.7).

Proof. Recall that

$$\mathcal{C}_r = \{\tilde{X}_i^{(r)} : i = 1, \dots, \lceil e^{r+2\delta \frac{d}{d_c}} \rceil\},$$

where the elements are i.i.d. centered Gaussian random elements with covariance operator

$$\sum_{j\in\mathbb{N}} \tilde{\lambda}_j \, e_j \otimes e_j.$$

For ease of notation, we denote by $\Delta : [0,\infty) \to [0,\infty),\ r \mapsto \Delta r$ the unique function with $\lceil e^{r+2\delta \frac{d}{d_c}} \rceil = e^{r+\Delta r}$. For $r \geq 0$, consider the set

$$\mathcal{T}_0(r) = \Big\{ x \in H_0 : \Lambda'_x(\theta_0) \leq d + \delta d \text{ and } \Lambda^*_x(\Lambda'_x(\theta_0)) \leq r + \delta \frac{d}{d_c} \Big\}.$$

Due to the previous proposition there exists $\kappa > 0$ such that

$$\mathbb{P}(X \notin \mathcal{T}_0(r)) \leq 2e^{-\kappa \frac{d}{d_c}}$$

for all $r \geq 0$. Note that

$$\mathbb{P}(\mathcal{T}(r)^c | X) = (1 - \mathbb{P}(\|X - \tilde{X}_1^{(r)}\|^2 \leq d + \delta d | X))^{\exp\{r+\Delta r\}}.$$

Using that $(1 - u/v)^v \leq e^{-u}$ for any $v \geq u \geq 0$, we obtain

$$\begin{aligned} \mathbb{P}(\mathcal{T}(r)^c | X) &\leq \exp\big\{ -\mathbb{P}\big(\|X - \tilde{X}_1^{(r)}\|^2 \leq d + \delta d | X\big) \exp\{r + \Delta r\} \big\} \\ &= \exp\big\{ -\exp\big\{ \log \mathbb{P}\big(\|X - \tilde{X}_1^{(r)}\|^2 \leq d + \delta d | X\big) + r + \Delta r \big\}\big\}. \end{aligned} \tag{6.14}$$

Applying the relation between the moment generating function and SBPs (Corollary 3.4.3) yields that

$$-\log \mathbb{P}\big(\|X - \tilde{X}_1^{(r)}\|^2 \leq \Lambda'_X(\theta_0) | X\big) \leq r_0 \vee \big[\Lambda^*_X(\Lambda'_X(\theta_0)) + \log(\Lambda^*_X(\Lambda'_X(\theta_0)))\big]$$

for some universal constant $r_0 \geq 1$. For $X \in \mathcal{T}_0(r)$, we obtain

$$-\log \mathbb{P}\big(\|X - \tilde{X}_1^{(r)}\|^2 \leq d + \delta d | X\big) \leq r_0 \vee \big[r + \delta \frac{d}{d_c} + \log(r + \delta \frac{d}{d_c})\big]. \tag{6.15}$$

Hence, with estimate (6.14),

$$\begin{aligned} \mathbb{P}(\mathcal{T}(r)^c) &\leq \mathbb{P}(X \notin \mathcal{T}_0(r)) \\ &\quad + \exp\Big\{ -\exp\big\{ r + \Delta r - [r_0 \vee (r + \delta \frac{d}{d_c} + \log(r + \delta \frac{d}{d_c}))] \big\}\Big\} \end{aligned}$$

and, for all $r \geq r_0$,

$$\mathbb{P}(\mathcal{T}(r)^c) \leq 2e^{-\kappa \frac{d}{d_c}} + \exp\big\{ -\exp\{ \delta \frac{d}{d_c} - \log(r + \delta \frac{d}{d_c}) \}\big\}. \tag{6.16}$$

Using that $\log r = o(d/d_c)$, one concludes that

$$\mathbb{P}(\mathcal{T}(r)^c) \lesssim 2e^{-\kappa\frac{d}{d_c}} \quad (r \to \infty).$$

□

Lemma 6.2.11. *If the eigenvalues $\{\lambda_j\}_{j\in\mathbb{N}}$ satisfy*

$$\lim_{n\to\infty} \frac{\log\log(1/\lambda_n)}{n} = 0, \tag{6.17}$$

then it holds

$$\log r = o(d/d_c).$$

Proof. Note that $d \geq N d_c$ and, hence, for $r \geq 1$,

$$\begin{aligned}\frac{d_c}{d}\log r &= \frac{d_c}{d}\log\Big(\frac{1}{2}\sum_{j=1}^{N}\log\frac{\lambda_j}{d_c}\Big)\\ &\leq \frac{1}{N}\log\Big(\frac{N}{2}\log\frac{\lambda_1}{\lambda_{N+1}}\Big) \to 0\end{aligned}$$

as $r \to \infty$. □

The properties that we have derived so far show that we can find in $\mathcal{C}_r$ a $\sqrt{d+\delta d}$-close representation for X with high probability. In order to prove the main theorem, we still need an estimate for the weak asymptotics of the quantization error.

Proposition 6.2.12. *For $s > 0$ there exists $\kappa = \kappa(s) > 0$ such that*

$$D^{(q)}(r|\mathcal{N}(0,t), |\cdot|^s)^{1/s} \leq \kappa\, D(r|\mathcal{N}(0,t), |\cdot|^2)^{1/2}$$

for all $r \geq 0$ and $t \geq 0$.

Proof. Observe that

$$\begin{aligned}&\sup_{N\in[1,\infty)} \frac{\delta^{(q)}(N|\mathcal{N}(0,1), |\cdot|^s)^{1/s}}{D(\log N|\mathcal{N}(0,1), |\cdot|^2)^{1/2}}\\ &= \sup_{N\in[1,\infty)} \frac{\delta^{(q)}(N|\mathcal{N}(0,1), |\cdot|^s)^{1/s}}{1/N}\\ &\leq \sup_{N\in\mathbb{N}} \frac{\delta^{(q)}(N|\mathcal{N}(0,1), |\cdot|^s)^{1/s}}{1/(N+1)}.\end{aligned}$$

By Theorem 1.2.1, the limit

$$\lim_{N\to\infty} N\,\delta^{(q)}(N|\mathcal{N}(0,1), |\cdot|^s)^{1/s}$$

exists and is equal to some constant $c \in \mathbb{R}_+$. Hence,

$$\frac{\delta^{(q)}(N|\mathcal{N}(0,1),|\cdot|^s)^{1/s}}{1/(N+1)}$$

is less than $2c$ for all but finitely many $N \in \mathbb{N}$. Consequently, the supremum of this expression taken over all $N \in \mathbb{N}$, say κ, is finite and it follows that

$$D^{(q)}(r|\mathcal{N}(0,1),|\cdot|^s)^{1/s} \leq \kappa\, e^{-r} = \kappa\, D(r|\mathcal{N}(0,1),|\cdot|^2)^{1/2}$$

for $r \geq 0$. Due to the scaling properties of the quantization error and the DRF, the result carries over to general normal distributions $\mathcal{N}(0,t)$, $t > 0$, with the same constant κ. For $t = 0$, the estimate is trivial. □

Proposition 6.2.13. *For $s \geq 2$ there exists $\kappa = \kappa(s) > 0$ such that*

$$D^{(q)}(r|\mu,\|\cdot\|^s)^{1/s} \leq \kappa\, D(r|\mu,\|\cdot\|^2)^{1/2}$$

for all $r \geq 0$.

Proof. As in the section on subband coding we assume without loss of generality that the underlying Hilbert space H is l_2 and X is the random element $X = \{\sqrt{\lambda_j} X_j\}_{j\in\mathbb{N}}$. With $Z_j := \sqrt{\lambda_j} X_j$, $j \in \mathbb{N}$, one has $X = \{Z_j\}_{j\in\mathbb{N}}$. Let $r \geq 0$ and associate with r

$$r_j = r_j(r) = \frac{1}{2}\log_+ \frac{\lambda_j}{d_c}, \quad j \in \mathbb{N}.$$

We construct an appropriate product codebook $\mathcal{C}_r$ on H. By Proposition 6.2.12, there exist $\kappa = \kappa(s) > 0$ and codebooks $\mathcal{C}_r^{(j)}$, $j \in \mathbb{N}$, $r \geq 0$, on $\mathbb{R}$ such that $|\mathcal{C}_r^{(j)}| \leq e^{r_j}$ and

$$\mathbb{E}[\min_{\hat z \in \mathcal{C}_r^{(j)}} |Z_j - \hat z|^s]^{1/s} \leq \kappa\, D(r_j|\mathcal{N}(0,\lambda_j),|\cdot|^2)^{1/2} \tag{6.18}$$

for all $r \geq 0$. Set $\mathcal{C}_r = \prod_{j\in\mathbb{N}} \mathcal{C}_r^{(j)}$. Since $\sum_{j\in\mathbb{N}} r_j = r$, the set $\mathcal{C}_r$ contains at most e^r elements. For $j \in \mathbb{N}$, let $\hat Z_j$ denote the closest (in $|\cdot|$) representation in $\mathcal{C}_r^{(j)}$ for Z_j and let $\hat X = \{\hat Z_j\}_{j\in\mathbb{N}}$. Then

$$\begin{aligned} D^{(q)}(r|\mu,\|\cdot\|^s)^{2/s} &\leq \mathbb{E}[\|X - \hat X\|^s]^{2/s} \\ &= \mathbb{E}[(\sum_{j\in\mathbb{N}} |Z_j - \hat Z_j|^2)^{s/2}]^{2/s} \\ &\overset{(a)}{\leq} \sum_{j\in\mathbb{N}} \mathbb{E}[|Z_j - \hat Z_j|^s]^{2/s} \\ &\overset{(b)}{\leq} \kappa^2 \sum_{j\in\mathbb{N}} D(r_i|Z_j,|\cdot|^2) \\ &= \kappa^2\, D(r|\mu,\|\cdot\|^2). \end{aligned}$$

Here, (a) follows from the triangle inequality in the $L_{s/2}(\mathbb{P})$-space. (b) holds due to equation (6.18). □

We are now able to prove Theorem 6.2.1.

Proof of Theorem 6.2.1. **1.)** We start with proving inequality "$\lesssim$". By the Hölder inequality, it follows that

$$D^{(q)}(r|\mu, \|\cdot\|^{s'})^{1/s'} \leq D^{(q)}(r|\mu, \|\cdot\|^{s})^{1/s}$$

for $s \geq s' > 0$ and $r \geq 0$. Hence, it suffices to consider only $s \geq 2$. By Proposition 6.2.13, there exist a sequence of codebooks $\{\tilde{\mathcal{C}}_r\}_{r\geq 0}$ and a constant $\kappa > 1$ such that $|\tilde{\mathcal{C}}_r| \leq e^r$ and

$$\mathbb{E}[\min_{\hat{x}\in\tilde{\mathcal{C}}_r} \|X - \hat{x}\|^{2s}]^{1/2s} \leq \kappa\, D(r)^{1/2} \tag{6.19}$$

for all $r \geq 0$. Let $\mathcal{C}_r = \mathcal{C}_r^{(\delta)}$, $r \geq 0$, denote the random codebooks as defined in (6.6) for some fixed $\delta \in (0, 1/4)$. We consider coding with codebook

$$\hat{\mathcal{C}}_r := \mathcal{C}_r \cup \tilde{\mathcal{C}}_r.$$

Let $\delta' \in (2\delta, 1/2)$. Then there exists $r_0 \geq 0$ such that

$$|\hat{\mathcal{C}}_r| \leq e^r + \lceil e^{r+2\delta\frac{d}{d_c}} \rceil \leq e^{r+\delta'\frac{d}{d_c}}$$

for all $r \geq r_0$. From now on let $r \geq r_0$. One has

$$\begin{aligned} D^{(q)}(r + \delta' d/d_c|\mu, \|\cdot\|^s) &\leq \mathbb{E}[\min_{\hat{x}\in\hat{\mathcal{C}}_r} \|X - \hat{x}\|^s] \\ &\leq \mathbb{E}[1_{\mathcal{T}(r)} \min_{\hat{x}\in\mathcal{C}_r} \|X - \hat{x}\|^s] + \mathbb{E}[1_{\mathcal{T}(r)^c} \min_{\hat{x}\in\tilde{\mathcal{C}}_r} \|X - \hat{x}\|^s] \\ &=: I_1(r) + I_2(r), \end{aligned}$$

where $\mathcal{T}(r) = \mathcal{T}^{(\delta)}(r)$ is defined as in (6.7). Clearly $I_1(r) \leq [(1+\delta)d]^{s/2}$. Moreover, by the Cauchy-Schwarz inequality,

$$I_2(r) \leq \mathbb{P}(\mathcal{T}(r)^c)^{1/2}\, \mathbb{E}[\min_{\tilde{x}\in\tilde{\mathcal{C}}_r} \|X - \hat{x}\|^{2s}]^{1/2}.$$

Property (6.19) implies that

$$I_2(r) \leq \kappa^s\, \mathbb{P}(\mathcal{T}(r)^c)^{1/2}\, d^{s/2}.$$

Due to Proposition 6.2.10, $\mathbb{P}(\mathcal{T}(r)^c)$ converges to 0. We conclude that $I_2(r) = o(d^{s/2})$ as $r \to \infty$. Consequently,

$$D^{(q)}(r + \delta' d/d_c|\mu, \|\cdot\|^s)^{1/s} \leq (I_1(r) + I_2(r))^{1/s} \lesssim (1+\delta)^{1/2} d^{1/2}.$$

On the other hand, by Corollary 6.2.4,

$$D(r+\delta' d/d_c) \geq (1-2\delta')d.$$

Putting everything together gives

$$\begin{aligned} D^{(q)}(r+\delta' d/d_c|\mu, \|\cdot\|^s)^{1/s} &\lesssim (1+\delta)^{1/2} d^{1/2} \\ &\leq \frac{(1+\delta)^{1/2}}{(1-2\delta')^{1/2}} D(r+\delta' d/d_c)^{1/2} \end{aligned}$$

as $r \to \infty$. We can choose $\delta \in (0, 1/4)$ and, hence, $\delta' \in (2\delta, 1/2)$ arbitrarily small. We conclude that

$$D^{(q)}(r|\mu, \|\cdot\|^s)^{1/s} \lesssim D(r)^{1/2} \quad (r \to \infty).$$

2.) Now we prove inequality "$\gtrsim$". Let $\{\check{\mathcal{C}}_r\}_{r\geq 0}$ be an arbitrary sequence of (deterministic) codebooks in H with $|\check{\mathcal{C}}_r| \leq e^r$. Fix $\delta \in (0, \frac{1}{4})$ and let $\hat{\mathcal{C}}_r$, $r \geq 0$, as in the first part of the proof. Note that $\hat{\mathcal{C}}_r$ depends on the parameters δ and s. We take $s = 2$ and denote

$$Z_1(r) = \min_{\hat{x}\in\check{\mathcal{C}}_r} \|X - \hat{x}\| \text{ and } Z_2(r) = \min_{\hat{x}\in\hat{\mathcal{C}}_r} \|X - \hat{x}\|.$$

Note that $|\check{\mathcal{C}}_r \cup \hat{\mathcal{C}}_r| \leq 2e^r + \lceil e^{r+2\delta d/d_c} \rceil$ For $\delta' \in (2\delta, \frac{1}{2})$ there exists $r_0 \geq 0$ such that

$$|\check{\mathcal{C}}_r \cup \hat{\mathcal{C}}_r| \leq e^{r+\delta' \frac{d}{d_c}} \quad (r \geq r_0).$$

Hence, for $r \geq r_0$,

$$\mathbb{E}[Z_1(r)^2 \wedge Z_2(r)^2] \geq D^{(q)}\Big(r+\delta' \frac{d}{d_c}|\mu, \|\cdot\|^2\Big) \geq D\Big(r+\delta' \frac{d}{d_c}\Big).$$

Due to Corollary 6.2.4, one has

$$\mathbb{E}[Z_1(r)^2 \wedge Z_2(r)^2] \geq (1-2\delta')d \tag{6.20}$$

for all $r \geq r_0$. On the other hand

$$\begin{aligned} &\mathbb{E}[Z_1(r)^2 \wedge Z_2(r)^2] \\ &\qquad \leq \mathbb{E}\big[1_{\{Z_1(r)^2\wedge Z_2(r)^2 \leq (1+\delta)d\}}\big(Z_1(r)^2 \wedge [(1+\delta)d]\big)\big] \\ &\qquad\quad + \mathbb{E}\big[1_{\{Z_1(r)^2\wedge Z_2(r)^2 > (1+\delta)d\}} Z_2(r)^2\big] \\ &\qquad =: I_1(r) + I_2(r). \end{aligned} \tag{6.21}$$

As a consequence of part 1.), it holds

$$I_2(r) \leq \mathbb{E}[1_{\{Z_2(r)^2 \geq d+\delta d\}} Z_2(r)^2] = o(d) \quad (r \to \infty).$$

With equations (6.20) and (6.21), we conclude

$$I_1(r) \geq (1-2\delta')d - I_2(r) \sim (1-2\delta')d \quad (r \to \infty).$$

Therefore,

$$\mathbb{E}\big[Z_1(r)^2 \wedge [(1+\delta)d]\big] \gtrsim (1-2\delta')d \quad (r \to \infty).$$

Recall that we can choose $\delta \in (0, 1/4)$ and $\delta' \in (2\delta, 1/2)$ arbitrarily small. Consequently, for any $\varepsilon > 0$,

$$\lim_{r\to\infty} \mathbb{P}\big[Z_1(r) < (1-\varepsilon)d^{1/2}\big] = 0.$$

Hence, for any $s > 0$, we have

$$\mathbb{E}[Z_1(r)^s]^{1/s} \gtrsim D(r)^{1/2} \text{ as } r \to \infty.$$

□

6.3 The high resolution quantization problem for random codebooks

In this section, we study the quantization problem with random codebooks, i.e. the quantity $D^{(r)}$. Recall that

$$D^{(r)}(r|\mu, \|\cdot\|^s) = \inf\Big\{\int\int \min_{i=1,\dots,N} \|x-y_i\|^s \, d\nu^{\otimes N}(y_1,\dots,y_N)\, d\mu(x) : \\ N = \lfloor e^r \rfloor, \, \nu \in \mathcal{M}_1(E)\Big\}$$

for $r \geq 0$. We keep the notations and definitions of the previous section.

The main aim of this section is to strengthen Theorem 6.2.1 to

Theorem 6.3.1. *Suppose that*

$$\lim_{n\to\infty} \frac{\log\log(1/\lambda_n)}{n} = 0. \tag{6.22}$$

Then for any $s \in (0, \infty)$

$$D^{(r)}(r|\mu, \|\cdot\|^s)^{1/s} \sim D^{(q)}(r|\mu, \|\cdot\|^s)^{1/s} \sim D(r|\mu, \|\cdot\|^2)^{1/2}$$

as $r \to \infty$.

We adopt the notations of d and d_c of the previous section. In order to strengthen the result to the above theorem, we apply Theorem 5.2.1. It remains to verify its assumptions. Due to Corollary 6.2.4 and Theorem 6.2.1, perturbations $\Delta : [0,\infty) \to [0,\infty)$, $r \mapsto \Delta r$ with $\Delta r = o(d/d_c)$ do not have an influence on the strong asymptotics of the quantization problem, i.e.

$$D^{(q)}(r+\Delta r|\mu, \|\cdot\|^s) \sim D^{(q)}(r|\mu, \|\cdot\|^s) \quad (r \to \infty)$$

for arbitrarily fixed $s > 0$. Consequently, $\delta^{(q)}(\,\cdot\,|\mu, \|\cdot\|^s)$ is slowly varying. Hence, it remains to prove the weak equivalence

$$D^{(r)}(r|\mu, \|\cdot\|^s) \approx D^{(q)}(r|\mu, \|\cdot\|^s) \quad (r \to \infty).$$

Since $D^{(q)}$ is in general dominated by $D^{(r)}$, it suffices to find an appropriate upper bound for $D^{(r)}$. In analogy to the proof of Theorem 6.2.1, we only need to consider $s \geq 2$. Then Theorem 6.3.1 is a consequence of the following lemma.

Lemma 6.3.2. *For $s \geq 2$ there exists a constant $\kappa = \kappa(s) > 0$ such that*

$$D^{(r)}(r|\mu, \|\cdot\|^s)^{1/s} \leq \kappa\, D(r|\mu, \|\cdot\|^2)^{1/2}$$

for all $r \geq 0$.

The lemma does not require assumption (6.22). Consequently, the expressions $D^{(r)}(r|\mu, \|\cdot\|^s)^{1/s}$ and $D(r|\mu, \|\cdot\|^2)^{1/2}$ are weakly asymptotically equivalent for any $s \geq 2$ without any assumptions on the eigenvalues. However, if Theorem 6.2.1 is not applicable, then it is not known whether the function

$$N \mapsto D^{(q)}(\log N|\mu, \|\cdot\|^s)$$

is slowly varying and we cannot conclude strong equivalence in the general case. This problem remains unsolved.

Lemma 6.3.3. *Let $\{A_i\}_{i\in\mathbb{N}}$ and $\{B_i\}_{i\in\mathbb{N}}$ be independent sequences of i.i.d. non-negative real-valued r.v.'s. Here, the underlying laws of the sequences may differ. It holds, for any $N, M \in \mathbb{N}$ and $s > 0$,*

$$\mathbb{E}\Big[\min_{i=1,\dots,NM}[A_i+B_i]^s\Big] \leq \mathbb{E}\Big[\min_{\substack{i=1,\dots,N;\\ j=1,\dots,M}}[A_i+B_j]^s\Big].$$

Proof. Fix $N, M \in \mathbb{N}$ and $s > 0$, and denote

$$I_m = \{(m-1)N+1, \dots, mN\}, \quad m = 1, \dots, M.$$

For each $m \in \{1, \ldots, M\}$, we denote by j_m the minimal random index in I_m such that

$$A_{j_m} = \min_{i \in I_m} A_i.$$

Let us denote by J the random set of indices $\{j_1, \ldots, j_M\}$. Then the sequence $\{A_j\}_{j \in J}$ contains M i.i.d. r.v.'s with law $\mathcal{L}(\min_{i=1,\ldots,N} A_i)$. By construction, the sequences $\{A_j\}_{j \in J}$ and $\{B_j\}_{j \in J}$ are mutually independent. Moreover, $\{B_j\}_{j \in J}$ is an i.i.d. sequence with law $\mathcal{L}(B_1)$. Denote by $\hat{m}$ the minimal random index in $\{1, \ldots, M\}$ with

$$B_{j_{\hat{m}}} = \min_{i \in J} B_i.$$

Then $\mathcal{L}(B_{j_{\hat{m}}}) = \mathcal{L}(\min_{i=1,\ldots,M} B_i)$. Since $\{A_i\}_{i \in J}$ and $\hat{m}$ are independent, it follows that

$$\mathcal{L}(A_{j_{\hat{m}}} + B_{j_{\hat{m}}}) = \mathcal{L}\big(\min_{\substack{i=1,\ldots,N;\\ j=1,\ldots,M}} [A_i + B_j]\big).$$

The assertion is obtained by noting that

$$\min_{i=1,\ldots,NM} [A_i + B_i] \le A_{j_{\hat{m}}} + B_{j_{\hat{m}}}.$$

□

Proposition 6.3.4. *Let $\tilde{s} > s \ge 1$. There exists a constant $\kappa = \kappa(s, \tilde{s})$ such that*

$$D^{(r)}(r|Z, |\cdot|^s)^{1/s} \le \kappa \, \|Z\|_{L_{\tilde{s}}(\mathbb{P})} \, e^{-r} \tag{6.23}$$

for all $r \ge 0$ and all real-valued random variables Z in $L_{\tilde{s}}(\mathbb{P})$.

The proposition is a consequence of a result of Pierce [55] (see also Graf and Luschgy [27], Lemma 6.6). He found that, for $s \ge 1$ and $\varepsilon > 0$ there exist constants $C_1, C_2, C_3 > 0$ such that, for any distribution $\xi \in \mathcal{M}_1(\mathbb{R})$,

$$\delta^{(q)}(N|\xi, |\cdot|^s) \le \Big(C_1 \int |x|^{s+\varepsilon} d\xi(x) + C_2\Big) \frac{1}{N^s}, \quad N \ge C_3.$$

The proof of this result uses a random coding argument. Moreover, the proof contained in Graf and Luschgy [27] implies

Lemma 6.3.5. *Let $\tilde{s} > s \ge 1$ and denote by ν_P^+ the Pareto distribution with*

$$\nu_P^+(-\infty, t] = 1 - ((t+1) \vee 1)^{-\frac{\tilde{s}}{s}+1}, \quad t \in \mathbb{R}.$$

Then there exist constants $C_1, C_2, C_3 > 0$ such that

$$\int\int \min_{i=1,\ldots,N} |x - y_i|^s \, d\big(\nu_P^+\big)^{\otimes N}(y_1, \ldots, y_N) \, d\xi(x) \le \frac{C_1 \int x^{\tilde{s}} \, d\xi(x) + C_2}{N^s}$$

for all $N \ge C_3$ and any $\xi \in \mathcal{M}_1[0, \infty)$ with finite $\tilde{s}$-th moment.

Proof of Proposition 6.3.4. Let $Z \in L_{\tilde{s}}(\mathbb{P})$ and observe that, for all $\lambda \in [0,\infty)$ and $r \geq 0$,

$$D^{(r)}(r|\lambda Z, |\cdot|^s)^{1/s} = \lambda\, D^{(r)}(r|Z, |\cdot|^s)^{1/s}.$$

The same scaling property holds for the right hand side of (6.23). Therefore, it suffices to prove estimate (6.23) for r.e.'s Z with $\|Z\|_{L_{\tilde{s}}(\mathbb{P})} = 1$.

Note that $D^{(r)}(r|Z, |\cdot|^s)^{1/s} \leq \|Z\|_{L_{\tilde{s}}(\mathbb{P})}$ for all $r \geq 0$. It remains to find an estimate applicable for large rates $r \geq 0$. We consider quantization with random codebooks generated by the distribution

$$\nu = \frac{1}{2}(\nu_P^+ + \nu_P^-),$$

where ν_p^+ is the Pareto distribution as in Lemma 6.3.5 and ν_P^- is the at 0 flipped version of ν_P^+. Let $\mathcal{C}_N = \{Y_1, \dots, Y_N\}$, $N \in \mathbb{N}$, be constituted by N ν-distributed random variables. We divide $\mathcal{C}_N$ into two parts, $\mathcal{C}_{1,N} = \mathcal{C}_N \cap (-\infty, 0]$ and $\mathcal{C}_{2,N} = \mathcal{C}_N \cap [0,\infty)$. The size of each set $\mathcal{C}_{i,N}$, $i = 1,2$, is binomially distributed with parameters N and $1/2$. Hence, by the Markov inequality there exists a constant $c > 0$ such that, for all $N \in \mathbb{N}$ and $i = 1,2$,

$$\mathbb{P}(|\mathcal{C}_{i,N}| < N/3) \leq e^{-cN}.$$

Let $\mathcal{T}(N)$, $N \in \mathbb{N}$, denote the event

$$\mathcal{T}(N) = \{|\mathcal{C}_{i,N}| \geq N/3 \text{ for } i = 1,2\}.$$

Then $\mathbb{P}(\mathcal{T}(N)^c) \leq 2e^{-cN}$, $N \in \mathbb{N}$. We estimate

$$\begin{aligned}\mathbb{E}[d(Z,\mathcal{C}_N)^s]^{1/s} &\leq \mathbb{E}[1_{\mathcal{T}(N)} d(Z,\mathcal{C}_N)^s]^{1/s} + \mathbb{E}[1_{\mathcal{T}(N)^c}\, d(Z,\mathcal{C}_N)^s]^{1/s} \\ &=: I_1(N) + I_2(N).\end{aligned}$$

By Lemma 6.3.5, the first term is bounded by

$$I_1(N) = \mathbb{E}[1_{\mathcal{T}(N)}\, d(Z,\mathcal{C}_N)^s]^{1/s} \leq (C_1 \mathbb{E}[|Z|^{\tilde{s}}] + C_2)^{1/s}\, \frac{3}{N}$$

for $N \geq C_3$ where $C_1, C_2, C_3 \in \mathbb{R}_+$ are appropriate constants. Moreover,

$$\begin{aligned}I_2(N) = \mathbb{E}[1_{\mathcal{T}(N)^c}\, d(Z,\mathcal{C}_N)^s]^{1/s} &\leq \mathbb{E}[1_{\mathcal{T}(N)^c} \big(|Z| + \min_{i=1,\dots,N} |Y_i|\big)^s]^{1/s} \\ &\overset{(a)}{=} \mathbb{P}\{\mathcal{T}(N)^c\}^{1/s}\, \mathbb{E}\big[\big(|Z| + \min_{i=1,\dots,N} |Y_i|\big)^s\big]^{1/s} \\ &\overset{(b)}{\leq} \big(2e^{-cN}\big)^{1/s} \big(\|Z\|_{L_s(\mathbb{P})} + \|\min_{i=1,\dots,N} |Y_i|^s\|_{L_s(\mathbb{P})}\big).\end{aligned}$$

In step (a), we used the independence of the r.v. $1_{\mathcal{T}(N)}$ and the sequence $\{|Y_i|\}_{i=1,\dots,N}$. Estimate (b) is a consequence of the estimate for $\mathbb{P}(\mathcal{T}(N)^c)$ and the triangle inequality in $L_s(\mathbb{P})$. According to Lemma 6.3.5, one has in particular that $\mathbb{E}[\min_{i=1,\dots,N}|Y_i|^s] < \infty$ for $N \geq C_3$. In fact, the finiteness is immediately obtained when taking $\xi = \delta_0$. Combining the estimates for I_1 and I_2, yields the existence of constants $C_4, C_5 > 0$ such that

$$\mathbb{E}[d(Z, \mathcal{C}_N)^s] \leq (C_4 \mathbb{E}[|Z|^{s+\varepsilon}] + C_5)^{1/s} \frac{1}{N}$$

for all $N \geq C_3$. The estimate is valid for all processes $Z \in L_{\tilde{s}}(\mathbb{P})$ and the proof is complete. □

Proof of Lemma 6.3.2. By the reasoning in the section on subband coding, we may assume without loss of generality that the underlying Hilbert space H is l_2 and X is the random element $X = \{\sqrt{\lambda_j} X_j\}_{j\in\mathbb{N}}$. With $Z_j = \sqrt{\lambda_j} X_j$, $j \in \mathbb{N}$, one can write $X = \{Z_j\}_{j\in\mathbb{N}}$. Let $r \geq 0$ and associate with r

$$r_j = r_j(r) = \frac{1}{2} \log_+ \frac{\lambda_j}{d_c} \quad (j \in \mathbb{N})$$

and $N_j = N_j(r) = \lfloor e^{r_j} \rfloor$ and denote $N = N(r) = \lfloor e^r \rfloor$. Fix $s \geq 2$. By Proposition 6.3.4, there exist distributions $\nu_j = \nu_j^{(r)} \in \mathcal{M}_1(\mathbb{R})$, $j \in \mathbb{N}$, and a constant $\kappa = \kappa(s) \in \mathbb{R}_+$ such that, for any $j \in \mathbb{N}$ and $r \geq 0$,

$$\int\int \min_{i=1,\dots,N_j} |x - y_i|^s \, d\nu_j^{\otimes N_j}(y_1, \dots, y_{N_j}) \, d\mathcal{N}(0, \lambda_j)(x) \leq \big(\kappa \sqrt{\lambda_j} e^{-r_j}\big)^s. \tag{6.24}$$

Moreover, we can choose $\nu_j^{(r)} = \delta_0$ if $N_j(r) = 1$. In fact, one obtains immediately by the Anderson inequality that for $N_j(r) = 1$ the distribution $\nu_j^{(r)} = \delta_0$ minimizes the left hand side of (6.24).

We consider coding with codebooks $\mathcal{C}_r$, $r \geq 0$, constituted by $N(r)$ independent

$$\nu^{(r)} = \bigotimes_{j\in\mathbb{N}} \nu_j^{(r)}$$

distributed random elements. Fix $r \geq 0$ and let $\{Y_j^{(r)}(i)\}_{i\in\mathbb{N}}$, $j \in \mathbb{N}$, denote independent sequences of independent $\nu_j^{(r)}$-distributed random variables. Moreover, we assume independence of X. We consider

$$\mathcal{C}_r = \big\{Y^{(r)}(i) : i = 1, \dots, N(r)\big\},$$

where we let $Y^{(r)}(i) = \{Y_j^{(r)}(i)\}_{j\in\mathbb{N}} \in l_2$, $i \in \mathbb{N}$. Then one has

$$\begin{aligned}\mathbb{E}[d(X,\mathcal{C}_r)^s] &= \int \mathbb{E}\Big[\min_{i=1,\dots,N(r)} \|x - Y^{(r)}(i)\|^s\Big]\,d\mu(x)\\ &= \int \mathbb{E}\Big[\min_{i=1,\dots,N(r)} \Big(\sum_{j\in\mathbb{N}} |x_j - Y_j^{(r)}(i)|^2\Big)^{s/2}\Big]\,d\mu(x).\end{aligned}$$

Note that for all but finitely many $j \in \mathbb{N}$, one has $Y_j^{(r)}(i) = 0$ for all $i \in \mathbb{N}$, almost surely. Moreover, $\prod_{j\in\mathbb{N}} N_j(r) \le N(r)$. Therefore, finitely many applications of Lemma 6.3.3 give

$$\mathbb{E}\Big[\min_{i=1,\dots,N(r)} \Big(\sum_{j\in\mathbb{N}} |x_j - Y_j^{(r)}(i)|^2\Big)^{s/2}\Big] \le \mathbb{E}\Big[\Big(\sum_{j\in\mathbb{N}} \min_{i=1,\dots,N_j(r)} |x_j - Y_j^{(r)}(i)|^2\Big)^{s/2}\Big]$$

for any $x = \{x_j\}_{j\in\mathbb{N}} \in l_2$. Consequently,

$$\begin{aligned}\mathbb{E}[d(X,\mathcal{C}_r)^s]^{2/s} &\le \Big(\int \mathbb{E}\Big[\Big(\sum_{j\in\mathbb{N}} \min_{i=1,\dots,N_j(r)} |x_j - Y_j^{(r)}(i)|^2\Big)^{s/2}\Big]\,d\mu(x)\Big)^{2/s}\\ &= \mathbb{E}\Big[\Big(\sum_{j\in\mathbb{N}} \min_{i=1,\dots,N_j(r)} |X_j - Y_j^{(r)}(i)|^2\Big)^{s/2}\Big]^{2/s}\\ &\le \sum_{j\in\mathbb{N}} \mathbb{E}\Big[\min_{i=1,\dots,N_j(r)} |X_j - Y_j^{(r)}(i)|^s\Big]^{2/s},\end{aligned}$$

where the last inequality follows from the triangle inequality in the $L_{s/2}(\mathbb{P})$-space. With (6.24) we obtain

$$\mathbb{E}[d(X,\mathcal{C}_r)^s]^{2/s} \le \kappa^2 \sum_{j\in\mathbb{N}} \lambda_j\, e^{-2r_j} = \kappa^2 D(r|\mu, \|\cdot\|^2).$$

□

6.4 Applications

In the case where the eigenvalues $\{\lambda_j\}_{j\in\mathbb{N}}$ are regularly varying there is an elegant formula for the asymptotics of $D(r|\mu, \|\cdot\|^2)$. Here, regularly varying means that the function $r \mapsto \lambda_{\lceil r\rceil}$ is regularly varying.

Lemma 6.4.1. *Let $\lambda_j = j^{-\alpha} l(j)$ be regularly varying with index $-\alpha < -1$. Then*

$$D(r|\mu, \|\cdot\|^2) \sim \frac{\alpha^\alpha}{2^{\alpha-1}(\alpha-1)}\, r\lambda_{\lceil r\rceil} \quad (r \to \infty).$$

The same result is contained in Luschgy and Pagès [51], Theorem 2.2. Our proof is similar to the one provided in this reference.

Proof. In contrast to before, we define functions $d : (0, \lambda_1] \to (0, \|\lambda\|_{l_1}]$ and $r : (0, \lambda_1] \to [0, \infty)$ by

$$d(x) = \sum_{j \in \mathbb{N}} (x \wedge \lambda_j) \text{ and}$$
$$r(x) = \frac{1}{2} \sum_{j \in \mathbb{N}} \log_+ \frac{\lambda_j}{x}.$$

Furthermore, d_c denotes an arbitrary real value in $(0, \lambda_1]$. Then $(d(d_c), d_c, r(d_c))$ solves the system of equations (6.2). We extend the function λ on $\mathbb{R}_+ \backslash \mathbb{N}$ by $\lambda_t = \lambda_{\lceil t \rceil}$, $t > 0$. Then for $N \in \mathbb{N}$

$$\sum_{j=1}^{N} \log \lambda_j = \int_0^N \log \lambda_t \, dt.$$

When we approximate the latter integral by replacing λ by a function $\tilde{\lambda} : \mathbb{R}_+ \to \mathbb{R}_+$ with $\log \circ \tilde{\lambda}$ locally integrable and $\tilde{\lambda}_t \sim \lambda_t$, we can control the error for any $T \in \mathbb{R}_+$ by

$$\Big| \int_0^T \log \lambda_t \, dt - \int_0^T \log \tilde{\lambda}_t \, dt \Big| \le \int_0^T \Big| \log \frac{\lambda_t}{\tilde{\lambda}_t} \Big| \, dt = o(T) \quad (T \to \infty).$$

By basic facts of regularly varying functions (see Theorem A.3) there exists a function $\tilde{\lambda}_t = t^{-\alpha} \tilde{l}(t)$ in $C^\infty(\mathbb{R}_+, \mathbb{R}_+)$ that is asymptotically equivalent to λ and satisfies $\tilde{l}'(t) = o(\tilde{l}(t)/t)$ as $t \to \infty$. We estimate

$$\begin{aligned} \int_0^T \log \tilde{\lambda}_t \, dt &= \int_0^T \log t^{-\alpha} \, dt + \int_0^T \log \tilde{l}(t) \, dt \\ &= -\alpha [T \log T - T] + [t \log \tilde{l}(t)]_0^T - \int_0^T \frac{t \tilde{l}'(t)}{\tilde{l}(t)} \, dt \\ &= T \log(T^{-\alpha} \tilde{l}(T)) + \alpha T + o(T), \end{aligned}$$

where we have used partial integration and $\tilde{l}'(t) = o(\tilde{l}(t)/t)$. For $d_c \le \lambda_1$ let

$$N(d_c) := \max\{j \in \mathbb{N} : \lambda_j \ge d_c\}.$$

Using the above results, we conclude that

$$\begin{aligned} 2\,r(d_c) &= \sum_{j\in\mathbb{N}} \log_+ \frac{\lambda_j}{d_c} = \int_0^{N(d_c)} \log \lambda_t\, dt - N(d_c)\log d_c \\ &= \int_0^{N(d_c)} \log \tilde{\lambda}_t\, dt - N(d_c)\log d_c + o(N(d_c)) \\ &= N(d_c)\log \frac{\tilde{\lambda}_{N(d_c)}}{d_c} + \alpha N(d_c) + o(N(d_c)) \\ &= N(d_c)\log \frac{\lambda_{N(d_c)}}{d_c} + \alpha N(d_c) + o(N(d_c)), \end{aligned}$$

where the asymptotic estimates are for $d_c \downarrow 0$. By basic properties of regularly varying functions (see Lemma A.6), $N(d_c)$ is an asymptotic inverse of $\{\lambda_t\}_{t>0}$, one has $\lambda_{N(d_c)} \sim d_c$ as $d_c \downarrow 0$. Consequently,

$$2\,r(d_c) = \alpha N(d_c) + o(N(d_c)). \tag{6.25}$$

Consider now $d(d_c)$. One has

$$\begin{aligned} d(d_c) &= \sum_{j\in\mathbb{N}} (d_c \wedge \lambda_j) = N(d_c)d_c + \int_{N(d_c)}^{\infty} \lambda_t\, dt \\ &\sim N(d_c)d_c + \int_{N(d_c)}^{\infty} \tilde{\lambda}_t\, dt \quad (d_c \downarrow 0), \end{aligned}$$

with $\{\tilde{\lambda}_t\}$ as above. On the other hand,

$$\int_{N(d_c)}^{\infty} t^{-\alpha}\tilde{l}(t)\, dt = \Big[-\frac{1}{\alpha-1} t^{-\alpha+1}\tilde{l}(t)\Big]_{N(d_c)}^{\infty} + \int_{N(d_c)}^{\infty} \frac{1}{\alpha-1} t^{-\alpha+1}\tilde{l}'(t)\, dt.$$

Since by assumption $t^{-\alpha+1}\tilde{l}'(t) = o(t^{-\alpha}\tilde{l}(t))$ and $N(d_c) \to \infty$ as $d_c \downarrow 0$, we conclude

$$\begin{aligned} d(d_c) &\sim N(d_c)d_c + \frac{1}{\alpha-1} N(d_c)\tilde{\lambda}_{N(d_c)} \\ &\sim \frac{\alpha}{\alpha-1} N(d_c)d_c \quad (d_c \downarrow 0). \end{aligned} \tag{6.26}$$

Combining (6.25) and (6.26) gives

$$d(d_c) \sim \frac{2}{\alpha-1} r(d_c)d_c \sim \frac{2}{\alpha-1} r(d_c)\lambda_{2r(d_c)/\alpha} \sim \frac{\alpha^\alpha}{2^{\alpha-1}(\alpha-1)}\, r(d_c)\lambda_{r(d_c)}$$

as $d_c \downarrow 0$, which is equivalent to the statement in the lemma. □

As a consequence of Theorem 6.3.1 and Corollary 6.2.2, one obtains

Corollary 6.4.2. *Let $\{\lambda_j\}_{j\in\mathbb{N}}$ be regularly varying with index $-\alpha < -1$. Then for any $s > 0$ and $\tilde{s} \geq 2$, one has*

$$D^{(r)}(r|\mu, \|\cdot\|^s)^{1/s} \sim D^{(q)}(r|\mu, \|\cdot\|^s)^{1/s} \sim D^{(e)}(r|\mu, \|\cdot\|^{\tilde{s}})^{1/\tilde{s}}$$
$$\sim D(r|\mu, \|\cdot\|^{\tilde{s}})^{1/\tilde{s}} \sim \sqrt{\frac{\alpha^\alpha}{2^{\alpha-1}(\alpha-1)}\, r\lambda_{\lceil r\rceil}}.$$

In the sequel, we provide some examples for which the eigenvalues are known and give the corresponding asymptotics of $D(r|\mu, \|\cdot\|^2)$.

Fractional Brownian motion

Let $X = \{X_t\}_{t\in[0,1]}$ be fractional Brownian motion with Hurst exponent $\gamma/2$, $0 < \gamma < 2$, in $C[0,1]$, i.e. the centered continuous Gaussian process X with covariance kernel

$$\mathbb{E}[X_t X_s] = \frac{1}{2}\big[t^\gamma + s^\gamma - |t-s|^\gamma\big], \quad t, s \in [0,1].$$

We consider X as a random element in the Hilbert space $L_2[0,1]$ and denote $\|\cdot\| = \|\cdot\|_{L_2[0,1]}$.

The asymptotic behavior of the ordered eigenvalues has been determined by Bronski [9]. He found

$$\lambda_n \sim \frac{\sin(\pi\gamma/2)\,\Gamma(\gamma+1)}{(n\pi)^{\gamma+1}} \quad (n \to \infty),$$

where Γ is the Euler gamma function. Hence, by Corollary 6.4.2

$$D(r|X, \|\cdot\|^2)^{1/2} \sim \sqrt{\frac{(\gamma+1)^{\gamma+1}\sin(\pi\gamma/2)\,\Gamma(\gamma+1)}{2^\gamma\,\gamma\,\pi^{\gamma+1}}}\, r^{-\gamma/2}$$

as $r \to \infty$. In particular,

$$D(r|X, \|\cdot\|^2)^{1/2} \sim \frac{\sqrt{2}}{\pi}\, r^{-1/2},$$

when X denotes standard Brownian motion.

Fractional Brownian sheet

Fix $d \in \mathbb{N}$ and let $X = \{X_t\}_{t\in[0,1]^d}$ denote fractional Brownian sheet with realizations in $C([0,1]^d)$ and parameter $\gamma = (\gamma_1, \dots, \gamma_d)$, $0 < \gamma_j < 2$, i.e. X is a centered continuous Gaussian process with covariance kernel

$$\mathbb{E}[X_t X_s] = \frac{1}{2^d}\prod_{j=1}^d [t_j^{\gamma_j} + s_j^{\gamma_j} - |t_j - s_j|^{\gamma_j}] =: K(t,s), \quad t, s \in [0,1]^d.$$

We consider X as Gaussian random element in the Hilbert space $L_2[0,1]^d$. Then the covariance operator of X is the integral operator corresponding to the kernel $K(\cdot,\cdot)$, i.e.

$$\begin{aligned} C_\gamma : L_2[0,1]^d &\to L_2[0,1]^d \\ f \quad &\mapsto C_\gamma f = \int_{[0,1]^d} K(\cdot,s) f(s)\, ds. \end{aligned}$$

Denote by C_{γ_j}, $j=1,\ldots,d$, the integral operators on $L_2[0,1]$ corresponding to the kernels

$$K_{\gamma_j}(s,t) = \frac{1}{2}[t^{\gamma_j} + s^{\gamma_j} - |t-s|^{\gamma_j}], \quad t,s \in [0,1].$$

One can write C as the tensor product of the bounded (even compact) operators $C_{\gamma_1},\ldots,C_{\gamma_d}$, i.e.

$$C_\gamma = C_{\gamma_1} \otimes \cdots \otimes C_{\gamma_d}.$$

Here, we denote the tensor product to be the unique operator on $L_2[0,1]^d$ satisfying

$$C_{\gamma_1} \otimes \cdots \otimes C_{\gamma_d}(f_1 \otimes \cdots \otimes f_d) = (C_{\gamma_1} f_1) \otimes \cdots \otimes (C_{\gamma_d} f_d),$$

where we mean by $(f_1 \otimes \cdots \otimes f_d)(x_1,\ldots,x_d) = f_1(x_1)\cdots f_d(x_d)$. It is known that the eigenvalues of the operator C_γ can be deduced from the eigenvalues of the operators C_{γ_j}, $j=1,\ldots,d$. Let $\{\lambda_i^{(j)}\}_{i\in\mathbb{N}}$ denote the eigenvalues of C_{γ_j}. Then, each combination $i_1,\ldots,i_d \in \mathbb{N}$ yields an eigenvalue

$$\lambda_{i_1}^{(1)} \cdots \lambda_{i_d}^{(d)}$$

of C_γ. Furthermore, all eigenvalues of C_γ are obtained with the correct multiplicity by that construction. Consequently, for $t>0$,

$$|\{i \in \mathbb{N} : \lambda_i \geq t\}| = |\{(i_1,\ldots,i_d) \in \mathbb{N}^d : \lambda_{i_1}^{(1)} \cdots \lambda_{i_d}^{(d)} \geq t\}|,$$

where $\{\lambda_i\}_{i\in\mathbb{N}}$ denotes the sequence of ordered eigenvalues with correct multiplicity of C_γ. Hence, one can deduce the asymptotic of $\{\lambda_i\}$ from the sequences $\{\lambda_i^{(j)}\}$. The operation which assigns the sequences $\lambda^{(1)},\ldots,\lambda^{(d)}$ to the unique ordered sequence $\{\lambda_i\}_{i\in\mathbb{N}}$ will be denoted by π_d, i.e.

$$\pi_d(\lambda^{(1)},\ldots,\lambda^{(d)}) := \{\lambda_i\}_{i\in\mathbb{N}}.$$

A similar discussion to that above is contained in Ritter [58] (Section 6.2). There it is proven (Lemma 34)

Lemma 6.4.3. *Let $\alpha > 1$, $m \in \mathbb{N}$ and assume that $\eta^{(j)} = \{\eta_i^{(j)}\}_{i\in\mathbb{N}}$ are sequences in $\mathbb{R}_+$ with $\eta_i^{(j)} \sim i^{-\alpha}$ $(i \to \infty)$ for all $j \in \{1,\dots,m\}$. Then*

$$\pi_m(\eta^{(1)},\dots,\eta^{(m)})(i) \sim ((d-1)!)^{-\alpha}((\log i)^{d-1}/i)^{\alpha}$$

as $i \to \infty$.

Suppose that $\gamma_1 \le \gamma_2 \le \cdots \le \gamma_d$ and denote by $m \in \mathbb{N}$ the maximal number with $\gamma_1 = \gamma_m$. We conclude

Corollary 6.4.4. *If $\gamma_1 = \cdots = \gamma_d$ (i.e. $m = d$), then*

$$\lambda_i \sim \Big(\frac{\sin(\pi\gamma_1/2)\Gamma(\gamma_1+1)}{\pi^{\gamma_1+1}}\Big)^d ((d-1)!)^{-(\gamma_1+1)} \Big(\frac{(\log i)^{d-1}}{i}\Big)^{\gamma_1+1}.$$

Remark 6.4.5. Suppose now that $m \ne d$. It can be shown that the tensorization of covariance operators with eigenvalues decreasing on different scales (i.e. of different polynomial orders) yields eigenvalues that have the same weak asymptotic order as the operator with the slowest decaying eigenvalues. Moreover, straight forward calculations lead to

$$\lambda_i \sim c_\gamma \Big(\frac{(\log i)^{m-1}}{i}\Big)^{\gamma_1+1} \quad (i \to \infty),$$

where c_γ is a constant in $\mathbb{R}_+$ depending on the vector γ. Unfortunately, it is typically not possible to compute c_γ. It does not suffice to know the strong asymptotics of $\{\lambda_i^{(j)}\}_{i\in\mathbb{N}}$, $j = 1,\dots,d$, in order to compute c_γ.

In the case that $m = d$, Corollary 6.4.2 gives

$$D(r|X,\|\cdot\|^2) \sim \Big(\frac{\gamma_1+1}{(d-1)!}\Big)^{\gamma_1+1} \Big(\frac{\sin(\pi\gamma_1/2)\Gamma(\gamma_1+1)}{\pi^{\gamma_1+1}}\Big)^d \frac{1}{\gamma_1\, 2^{\gamma_1}} \frac{(\log r)^{(d-1)(\gamma_1+1)}}{r^{\gamma_1}}$$

as $r \to \infty$. For general $m \le d$, we still have

$$D(r|X,\|\cdot\|^2) \sim c_\gamma \frac{(\gamma_1+1)^{\gamma_1+1}}{\gamma_1\, 2^{\gamma_1}} \frac{(\log r)^{(m-1)(\gamma_1+1)}}{r^{\gamma_1}}$$

with $c_\gamma > 0$ as above.

Integrated Brownian motion

Let $m \in \mathbb{N}$ and let X be m-times integrated Brownian motion on $C[0,1]$, i.e.

$$X_t = \int_0^t \int_0^{t_1} \dots \int_0^{t_{m-1}} B_{t_m} dt_m \dots dt_2\, dt_1, \quad t \in [0,1],$$

where B denotes Brownian motion in $C[0,1]$. X is a Gaussian process in $L_2[0,1]$. Gao et al. [26] computed the asymptotics of the eigenvalues of the corresponding covariance operator. They found

$$\lambda_n \sim \frac{1}{(\pi n)^{2(m+1)}} \text{ as } n \to \infty.$$

Hence,

$$D\big(r|X, \|\cdot\|^2_{L_2[0,1]}\big) \sim \frac{1}{(2m+1)\,2^{2m+1}} \Big(\frac{2(m+1)}{\pi}\Big)^{2(m+1)} \frac{1}{r^{2m+1}}$$

as $r \to \infty$.

6.5 Convex combinations of measures

In the typical Hilbert space setting, we found that the asymptotics in the quantization problem do not depend on the moment $s > 0$. Moreover, we proved equivalence of the coding quantities D, $D^{(e)}$ and $D^{(q)}$ for all moments $s \geq 2$. In this section, we show that these equivalences are untypical and cannot be expected for most non-Gaussian originals.

We assume the general setting. For $n \in \mathbb{N}$, let μ_i, $i = 1, \dots, n$, be probability measures on a Polish space E and let ρ be a distortion measure on E. In the sequel, we consider the coding complexity of the measure

$$\mu := \sum_{i=1}^n p_i\, \mu_i,$$

where $p_i \in (0,1)$ for $i = 1, \dots, n$ and $\sum_{i=1}^n p_i = 1$.

Quantization

We consider the quantization problem of μ under the assumption that

$$\lim_{r\to\infty} r^\alpha\, D^{(q)}(r|\mu_i, \rho) = \kappa_i, \quad i = 1, \dots, n,$$

where $\alpha > 0$ and $\kappa_i \in [0, \infty)$, $i = 1, \dots, n$. Moreover, we suppose that $\kappa_i > 0$ for at least one $i \in \{1, \dots, n\}$. Then

$$D^{(q)}(r|\mu, \rho) \geq \sum_{i=1}^n p_i\, D^{(q)}(r|\mu_i, \rho) \sim \sum_{i=1}^n p_i\, \kappa_i\, \frac{1}{r^\alpha}$$

as $r \to \infty$. On the other hand,

$$D^{(q)}(r|\mu, \rho) \leq \sum_{i=1}^n p_i\, D^{(q)}(r - \log n|\mu_i, \rho) \sim \sum_{i=1}^n p_i\, \kappa_i\, \frac{1}{r^\alpha}.$$

Hence,

$$\lim_{r\to\infty} r^{\alpha}\, D^{(q)}(r|\mu,\rho) = \|\kappa_\cdot\|_{l_1(p)},$$

where we denote

$$\|a\|_{l_q(p)} = \Big(\sum_{i=1}^{n} p_i\,|a_i|^q\Big)^{1/q}$$

for any $q \in \mathbb{R}\backslash\{0\}$ and any real-valued sequence $a = \{a_i\}_{i=1,\dots,n}$.

Example 6.5.1. For $i = 1,\dots,n$, let $\sigma_i \in \mathbb{R}_+$ and let $X^{(i)}$ be independent Brownian motions with diffusion coefficients σ_i, i.e. $\mathcal{L}(X^{(i)}) = \mathcal{L}(\sigma_i B)$, where B denotes the standard Brownian motion. Let $p_i \in (0,1)$, $i = 1,\dots,n$, with $\sum_{i=1}^n p_i = 1$ and denote by ξ an independent r.v. (of $X^{(1)},\dots,X^{(n)}$) with $\mathbb{P}(\xi = i) = p_i$.

For $s > 0$ and $i \in \{1,\dots,n\}$, it holds

$$\lim_{r\to\infty} r^{1/2}\, D^{(q)}(r|X^{(i)}, \|\cdot\|^s_{L_2[0,1]})^{1/s} = \frac{\sqrt{2}\sigma_i}{\pi}.$$

We apply the previous result and obtain

$$\lim_{r\to\infty} r^{1/2}\, D^{(q)}(r|X^{(\xi)}, \|\cdot\|^s_{L_2[0,1]})^{1/s} = \Big\|\frac{\sqrt{2}\sigma_\cdot}{\pi}\Big\|_{l_s(p)}.$$

In particular, one has no longer equivalence of the moments for the asymptotic quantization problem for $X^{(\xi)}$.

Entropy coding

We consider entropy coding of $\mu = \sum_{i=1}^n p_i\,\mu_i$.

Lemma 6.5.2. *Let*

$$f(r) := \inf \sum_{i=1}^{n} p_i\, D^{(e)}(r_i|\mu_i,\rho), \tag{6.27}$$

where the infimum is taken over all $r_i \in [0,\infty)$, $i = 1,\dots,n$, with

$$\sum_{i=1}^{n} p_i\, r_i \le r.$$

For $r \ge \log n$, it holds

$$f(r) \le D^{(e)}(r|\mu,\rho) \le f(r - \log n).$$

Proof. Let ξ be a r.v. in $\{1,\dots,n\}$ with $\mathbb{P}(\xi=i)=p_i$ and let $X^{(i)}$, $i=1,\dots,n$, be independent μ_i-distributed r.e.'s. Moreover, we assume independence of $\{X^{(i)}\}_{i=1,\dots,n}$ and ξ. Then $\mathcal{L}(X^{(\xi)})=\mu$.

We start with proving the lower bound. For an arbitrary reconstruction $\hat X$ of rate $r\ge 0$ (i.e. $\mathbb{H}(\hat X)\le r$), we estimate

$$r\ge \mathbb{H}(\hat X)\ge \mathbb{H}(\hat X|\xi)=\sum_{i=1}^n p_i\,\mathbb{H}(\hat X|\xi=i).$$

Moreover,

$$\mathbb{E}[\rho(X,\hat X)]=\sum_{i=1}^n p_i\,\mathbb{E}[\rho(X^{(i)},\hat X)|\xi=i]\ge\sum_{i=1}^n p_i\,D^{(e)}(\mathbb{H}(\hat X|\xi=i)|\mu_i,\rho).$$

Consequently, $D^{(e)}(r|\mu,\rho)$ is greater than $f(r)$.

Now let $r_i\in[0,\infty)$ with $\sum_{i=1}^n r_i\le r-\log n$ and $\varepsilon>0$. We denote by $\hat X^{(i)}$, $i=1,\dots,n$, reconstructions of $X^{(i)}$ that are independent of ξ and satisfy

$$\mathbb{H}(\hat X^{(i)})\le r_i \text{ and } \mathbb{E}[\rho(X^{(i)},\hat X^{(i)})]\le D^{(e)}(r_i|\mu_i,\rho)+\varepsilon.$$

Then

$$\begin{aligned}\mathbb{H}(\hat X^{(\xi)})&\le \mathbb{H}(\hat X^{(\xi)},\xi)=\mathbb{H}(\xi)+\sum_{i=1}^n p_i\,\mathbb{H}(\hat X^{(i)})\\&\le \log n+\sum_{i=1}^n r_i\le r.\end{aligned}$$

Hence,

$$D^{(e)}(r|\mu,\rho)\le \mathbb{E}[\rho(X^{(\xi)},\hat X^{(\xi)})]\le\sum_{i=1}^n p_i\,D^{(e)}(r_i|\mu_i,\rho)+\varepsilon.$$

Since $\varepsilon>0$ was arbitrary, we conclude that $D^{(e)}(r|\mu,\rho)\le f(r-\log n)$. □

Lemma 6.5.3. *We assume that there exist $\alpha>0$ and $\kappa_i\in\mathbb{R}_+$, $i=1,\dots,n$, such that*

$$\lim_{r\to\infty} r^\alpha\,D^{(e)}(r|\mu_i,\rho)=\kappa_i.$$

Then

$$\lim_{r\to\infty} r^\alpha\,D^{(e)}(r|\mu,\rho)=\|\kappa_\cdot\|_{l_{1/1+\alpha}(p)}. \tag{6.28}$$

Proof. Let f as in (6.27) and let, for $r \geq 0$, $r_i = r_i(r) \in [0, \infty)$ such that

$$f(r) \sim \sum_{i=1}^{n} p_i \, D^{(e)}(r_i | \mu_i, \rho) \quad (r \to \infty)$$

and $\sum_{i=1}^{n} r_i = r$. Note that necessarily $\lim_{r \to \infty} r_i = \infty$ and, hence,

$$f(r) \sim \sum_{i=1}^{n} p_i \, \kappa_i \, r_i^{-\alpha} \quad (r \to \infty).$$

Applying the Hölder inequality for $q = -\frac{1}{\alpha} \in (-\infty, 0)$ and $q^* = 1/(1+\alpha)$ gives

$$\begin{aligned} \sum_{i=1}^{n} p_i \, \kappa_i \, r_i^{-\alpha} &\geq \|\kappa_\cdot\|_{l_{q^*}(p)} \, \|r_\cdot^{-\alpha}\|_{l_q(p)} \\ &= \|\kappa_\cdot\|_{l_{1/1+\alpha}(p)} \, r^{-\alpha}. \end{aligned} \tag{6.29}$$

On the other hand, for fixed $r \geq 0$, we can find $c \geq 0$ such that

$$r_i := r_i(r) := c \, \kappa_i^{1/1+\alpha}$$

solves $\sum_{i=1}^{n} r_i = r$. For this choice of $\{r_i\}_{i=1,\dots,n}$, one obtains equality in equation (6.29). Consequently,

$$\sum_{i=1}^{n} p_i \, D^{(e)}(r_i | \mu_i, \rho) \sim \|\kappa_\cdot\|_{l_{1/1+\alpha}(p)} \, r^{-\alpha} \quad (r \to \infty).$$

Hence, $f(r) \sim \|\kappa_\cdot\|_{l_{1/1+\alpha}(p)} \, r^{-\alpha}$ and the assertion follows from the previous lemma. □

Example 6.5.4. As in the above example, we let $X^{(i)}$, $i = 1, \dots, n$, be independent Brownian motions with diffusion coefficients $\sigma_i > 0$ and ξ be an independent r.v. with $\mathbb{P}(\xi = i) = p_i$. Then, for $i = 1, \dots, n$ and $s \geq 2$,

$$\lim_{r \to \infty} r^{1/2} \, D^{(e)}(r | X^{(i)}, \| \cdot \|_{L_2[0,1]}^s)^{1/s} = \frac{\sqrt{2}\sigma_i}{\pi}.$$

An application of the previous lemma gives

$$\lim_{r \to \infty} r^{1/2} \, D^{(e)}(r | X^{(\xi)}, \| \cdot \|_{L_2[0,1]}^s)^{1/s} = \Big\| \frac{\sqrt{2}\sigma_\cdot}{\pi} \Big\|_{l_{2s/s+2}(p)}.$$

Again the equivalence of moments does not hold for the process $X^{(\xi)}$. Moreover, there is no equivalence of entropy coding and quantization.

Suppose now that the stochastic process $X = \{X_t\}_{t\in[0,1]} \in C[0,1]$ solves the SDE

$$dX_t = \sigma(X_t)\, dB_t$$

for a "nice" function $\sigma : \mathbb{R} \to \mathbb{R}$. Note that the diffusion coefficient seen by X will for some realizations be rather high and for some others rather low. The above results suggest that there is typically no equivalence of moments and no equivalence of quantization and entropy coding in the high resolution coding problem. We will not further pursue this issue here.

Chapter 7

Random small ball probabilities

Let μ be a centered Gaussian measure on a separable Banach space $(E, \|\cdot\|)$ and denote by X a random element in E with law μ. Denote

$$\varphi(\varepsilon) = -\log \mu(B(0,\varepsilon)), \qquad \varepsilon > 0.$$

The asymptotics of the small ball function has attracted much interest in the last decades. Recall, for instance, the results given in Section 3.6. A detailed survey on this topic is contained in the review article of Li and Shao [47].

In the case that the underlying space is a Hilbert space, there is an explicit formula that gives the asymptotics of the small ball probabilities for an arbitrary center. Beside that particular case, the small ball problem for centers not in the Cameron-Martin space is unsolved.

In this chapter, we consider the r.e. X as center and ask for the asymptotics of the random small ball function

$$-\log \mu(B(X,\varepsilon)), \quad \varepsilon > 0,$$

as ε tends to 0. We will see that the measure concentration around random centers satisfies a number of nice properties. These new insights will lead to a tight link between random small ball probabilities and the coding quantity $D^{(R)}$.

7.1 General results

Theorem 7.1.1. *One has*

$$-\log \mu(B(x,\varepsilon)) \lesssim 2\,\varphi(\varepsilon/2) \text{ as } \varepsilon \downarrow 0$$

for μ-almost all $x \in E$.

We denote by Φ the distribution function of the standard normal distribution and set $\Psi = 1 - \Phi$. For the proof we need

Lemma 7.1.2. *For $y \in (0, 1/2]$,*

$$\Psi^{-1}(y) \le \sqrt{-2\log y}.$$

The lemma is an immediate consequence of the inequality

$$\Psi(x) \le \frac{1}{2} e^{-x^2/2}, \quad x \ge 0.$$

Proof of Theorem 7.1.1. For $n \in \mathbb{N}$, denote $c_n = n$ and $\varepsilon_n = \varphi^{-1}(n^3)$. Let K and $\mathcal{K}$ be the closed unit balls in E and the reproducing Hilbert space of μ, respectively. Consider

$$A_n = \varepsilon_n K + \Big(c_n + \Psi^{-1}(\mu(B(0, \varepsilon_n)))\Big)\mathcal{K}, \quad n \in \mathbb{N}.$$

Then by Borell's inequality (see Lemma 2.3.1)

$$\mu(A_n) \ge \Phi\big[c_n + \Psi^{-1}(\mu(B(0,\varepsilon_n))) + \Phi^{-1}(\mu(B(0,\varepsilon_n)))\big] = \Phi(c_n).$$

The tail probabilities of standard normal random variables are known to satisfy $\Psi(c_n) \sim \frac{1}{\sqrt{2\pi} n} e^{-n^2/2}$ as $n \to \infty$. Hence,

$$\sum_{n \in \mathbb{N}} \mu(A_n^c) \le \sum_{n \in \mathbb{N}} \Psi(c_n) < \infty.$$

Therefore, Borel-Cantelli's Lemma gives that a.s. all but finitely many events A_n, $n \in \mathbb{N}$, occur.

On the other hand, by the estimate of shifted balls (Lemma 2.2.4), we get for every $x \in A_n$

$$\begin{aligned}\mu(B(x, 2\varepsilon_n)) &\ge \exp\{-I(x, \varepsilon_n) - \varphi(\varepsilon_n)\}\\ &\ge \exp\Big\{-\frac{1}{2}\big[c_n + \Psi^{-1}(\mu(B(0,\varepsilon_n)))\big]^2 - \varphi(\varepsilon_n)\Big\},\end{aligned}$$

where $I(x,\varepsilon) = \inf_{z \in B(x,\varepsilon)} \|z\|_{H_\mu}^2/2$. For $n \in \mathbb{N}$ with $\mu(B(0,\varepsilon_n)) \le 1/2$, one has by the previous lemma,

$$\begin{aligned}-\log \mu(B(x, 2\varepsilon_n)) &\le \frac{1}{2}\big[c_n + \Psi^{-1}(\mu(B(0,\varepsilon_n)))\big]^2 + \varphi(\varepsilon_n)\\ &\le \frac{1}{2}\big[c_n + \sqrt{2\varphi(\varepsilon_n)}\big]^2 + \varphi(\varepsilon_n)\\ &= c_n^2/2 + c_n\sqrt{2\varphi(\varepsilon_n)} + 2\varphi(\varepsilon_n),\end{aligned}$$

where the estimate holds uniformly for all $x \in A_n$. Observe that by definition, $c_n = o(\varphi(\varepsilon_n))$ $(n \to \infty)$. Consequently,

$$\sup_{x \in A_n} -\log \mu(B(x, 2\varepsilon_n)) \lesssim 2\varphi(\varepsilon_n) \quad (n \to \infty).$$

Since $\lim_{n\to\infty} \varphi(\varepsilon_{n+1})/\varphi(\varepsilon_n) = 1$ and the small ball probabilities are monotone, the lemma is proven. □

Combining Theorem 7.1.1 with Anderson's inequality shows that the small ball probabilities around a random center are enclosed by two deterministic functions a.s., i.e.

$$\varphi(\varepsilon) \le -\log \mu(B(X, \varepsilon)) \lesssim 2\,\varphi(\varepsilon/2) \quad (\varepsilon \downarrow 0,\ \text{a.s.})$$

In the case where the small ball function φ is regularly varying at 0 with index $-\alpha \le 0$, one has

$$\varphi(\varepsilon) \le -\log \mu(B(X, \varepsilon)) \lesssim 2^{1+\alpha}\,\varphi(\varepsilon) \quad (\varepsilon \downarrow 0,\ \text{a.s.}).$$

In the following, we address the question of existence of a deterministic function $\varphi_R : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying

$$-\log \mu(B(X, \varepsilon)) \sim \varphi_R(\varepsilon) \quad (\varepsilon \downarrow 0,\ \text{a.s.}).$$

Lemma 7.1.3. *For any continuous function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$, there exist constants $c_\psi, C_\psi \in [0, \infty]$ such that*

$$\liminf_{\varepsilon\downarrow 0} \frac{-\log \mu(B(x, \varepsilon))}{\psi(\varepsilon)} = c_\psi$$

and

$$\limsup_{\varepsilon\downarrow 0} \frac{-\log \mu(B(x, \varepsilon))}{\psi(\varepsilon)} = C_\psi$$

for μ-almost all $x \in E$.

The proof is based on the zero-one law for Gaussian measures (Lemma 2.2.3).

Proof. Let $h = C_\mu(z) \in C_\mu(E')$ and $x \in E$. The Cameron-Martin formula (Lemma 2.2.5) gives

$$\mu(B(x - h, \varepsilon)) = \int_{B(x,\varepsilon)} \exp\bigl\{z(y) - \frac{1}{2}\|z\|^2_{L_2(\mu)}\bigr\}\, d\mu(y).$$

The continuity of z implies that

$$\mu(B(x - h, \varepsilon)) \sim \exp\bigl\{z(x) - \frac{1}{2}\|z\|^2_{L_2(\mu)}\bigr\}\, \mu(B(x, \varepsilon))$$

as $\varepsilon \downarrow 0$. In particular,

$$-\log \mu(B(x-h,\varepsilon)) \sim -\log \mu(B(x,\varepsilon)) \quad (\varepsilon \downarrow 0).$$

Therefore, for any $s \geq 0$, the set

$$A_s = \Big\{ x \in E : \liminf_{\varepsilon \downarrow 0} \frac{-\log \mu(B(x,\varepsilon))}{\psi(\varepsilon)} \leq s \Big\}$$

is invariant under an arbitrary shift $h \in C_\mu(E')$. Since ψ is continuous, the sets A_s, $s \geq 0$, are measurable. Consequently, by the zero-one law for Gaussian measures (Lemma 2.2.3), the sets A_s have μ-measure 0 or 1. The first statement follows. Analogously one proves the second statement. □

Remark 7.1.4. As can be easily seen the previous lemma holds also when replacing $-\log$ by a continuous and at 0 slowly varying function l (i.e. $l(1/x)$ is slowly varying).

Corollary 7.1.5. *If $\varphi(\varepsilon) = \varepsilon^{-\alpha} l(1/\varepsilon)$ with $\alpha > 0$ and l slowly varying, then*

$$1 \leq c_\varphi \leq C_\varphi \leq 2^{1+\alpha},$$

where c_φ and C_φ are as in Lemma 7.1.3.

The following lemma will prove to be useful in the later discussion.

Lemma 7.1.6. *Let Z denote a standard normal r.v. For any $s \geq 1$ and $\varepsilon > 0$ with $\mu(B(0,\varepsilon)) \leq 1/2$, one has*

$$\| -\log \mu(B(X,2\varepsilon))\|_{L_s(\mathbb{P})} \leq \varphi(\varepsilon) + \frac{1}{2}\Big(\sqrt{2\varphi(\varepsilon)} + \|Z\|_{L_{2s}(\mathbb{P})}\Big)^2.$$

The upper bound given in the lemma is asymptotically equivalent to $2\varphi(\varepsilon)$ as $\varepsilon \downarrow 0$.

Proof. The proof is similar to that of Theorem 7.1.1. We fix $\varepsilon > 0$ with $\mu(B(0,\varepsilon)) \leq 1/2$ and let

$$A_t = \varepsilon K + (t + \Psi^{-1}(\mu(B(0,\varepsilon))))\mathcal{K}, \quad t \geq 0.$$

Here, K and $\mathcal{K}$ are again the closed unit balls in E and the reproducing kernel Hilbert space H_μ of μ, respectively. By Borell's inequality (Lemma 2.3.1) one has

$$\mu(A_t) \geq \Phi\big[t + \Psi^{-1}(\mu(B(0,\varepsilon))) + \Phi^{-1}(\mu(B(0,\varepsilon)))\big] = \Phi(t) \tag{7.1}$$

for any $t \geq 0$. Applying the small ball estimate for shifted balls (Lemma 2.2.4) gives for any $t \geq 0$ and $x \in A_t$

$$\begin{aligned}\mu(B(x,2\varepsilon)) &\geq \exp\{-I(x,\varepsilon) - \varphi(\varepsilon_n)\}\\ &\geq \exp\Big\{-\frac{1}{2}[t + \Psi^{-1}(\mu(B(0,\varepsilon)))]^2 - \varphi(\varepsilon)\Big\}.\end{aligned}$$

By Lemma 7.1.2, it follows

$$\begin{aligned}-\log \mu(B(x,2\varepsilon)) &\leq \frac{1}{2}[t + \Psi^{-1}(\mu(B(0,\varepsilon)))]^2 + \varphi(\varepsilon)\\ &\leq \frac{1}{2}[t + \sqrt{2\varphi(\varepsilon)}]^2 + \varphi(\varepsilon).\end{aligned}$$

Combining this estimate with (7.1) gives

$$\mathbb{P}\big\{-\log \mu(B(x,2\varepsilon)) > \frac{1}{2}[t + \sqrt{2\varphi(\varepsilon)}]^2 + \varphi(\varepsilon)\big\} \leq \Psi(t)$$

for all $t \geq 0$. Hence, with $Z^+ = Z \vee 0$ it follows that

$$\begin{aligned}\| -\log \mu(B(X,2\varepsilon))\|_{L_s(\mathbb{P})} &= \mathbb{E}\big[(-\log \mu(B(X,2\varepsilon)))^s\big]^{1/s}\\ &\leq \mathbb{E}\big[(\frac{1}{2}[Z^+ + \sqrt{2\varphi(\varepsilon)}]^2 + \varphi(\varepsilon))^s\big]^{1/s}.\end{aligned}$$

Applying the triangle inequality twice yields

$$\begin{aligned}\| -\log \mu(B(X,2\varepsilon))\|_{L_s(\mathbb{P})} &\leq \frac{1}{2}\mathbb{E}\big[\big(Z^+ + \sqrt{2\varphi(\varepsilon)}\big)^{2s}\big]^{1/s} + \varphi(\varepsilon)\\ &\leq \frac{1}{2}\big(\mathbb{E}[(Z^+)^{2s}]^{1/2s} + \sqrt{2\varphi(\varepsilon)}\big)^2 + \varphi(\varepsilon)\end{aligned}$$

and the assertion follows. □

In the rest of this paragraph, we assume the existence of a deterministic function $\varphi_R : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying

$$\varphi_R(\varepsilon) \sim -\log \mu(B(X,\varepsilon)) \quad \text{as } \varepsilon \downarrow 0\text{, in probability.} \tag{7.2}$$

Here, equivalence in probability means that

$$\lim_{\varepsilon \downarrow 0} \frac{-\log \mu(B(X,\varepsilon))}{\varphi_R(\varepsilon)} = 1\text{, in probability.}$$

We will see that φ_R exhibits some nice properties.

Theorem 7.1.7. *Suppose that*

$$\varphi(\varepsilon) \approx \varphi(2\varepsilon) \quad (\varepsilon \downarrow 0)$$

and $\varphi_R : \mathbb{R}_+ \to \mathbb{R}_+$ *satisfies (7.2). Then*

$$\lim_{\varepsilon \downarrow 0} \frac{-\log \mu(B(X,\varepsilon))}{\varphi_R(\varepsilon)} = 1$$

in $L_s(\mathbb{P})$ *for any* $s \geq 1$*. In particular,*

$$\varphi_R(\varepsilon) \sim \| -\log \mu(B(X,\varepsilon))\|_{L_s(\mathbb{P})}$$

as $\varepsilon \downarrow 0$*.*

Proof. Fix $\eta \in (0,1)$ and let

$$\mathcal{T}(\varepsilon) = \Big\{ x \in E : \Big| \frac{-\log \mu(B(x,\varepsilon))}{\varphi_R(\varepsilon)} - 1 \Big| \leq \eta \Big\}.$$

Then

$$\begin{aligned}
\Big\| \frac{-\log \mu(B(X,\varepsilon))}{\varphi_R(\varepsilon)} - 1 \Big\|_{L_s(\mathbb{P})} &\leq \Big\| 1_{\mathcal{T}(\varepsilon)}(X) \Big(\frac{-\log \mu(B(X,\varepsilon))}{\varphi_R(\varepsilon)} - 1 \Big) \Big\|_{L_s(\mathbb{P})} \\
&\quad + \Big\| 1_{\mathcal{T}(\varepsilon)^c}(X) \Big(\frac{-\log \mu(B(X,\varepsilon))}{\varphi_R(\varepsilon)} - 1 \Big) \Big\|_{L_s(\mathbb{P})} \\
&\leq \Big\| 1_{\mathcal{T}(\varepsilon)}(X) \Big(\frac{-\log \mu(B(X,\varepsilon))}{\varphi_R(\varepsilon)} - 1 \Big) \Big\|_{L_s(\mathbb{P})} \\
&\quad + \Big\| 1_{\mathcal{T}(\varepsilon)^c}(X) \, \frac{-\log \mu(B(X,\varepsilon))}{\varphi_R(\varepsilon)} \Big\|_{L_s(\mathbb{P})} \\
&\quad + \| 1_{\mathcal{T}(\varepsilon)^c}(X) \|_{L_s(\mathbb{P})} \\
&=: I_1(\varepsilon) + I_2(\varepsilon) + I_3(\varepsilon).
\end{aligned}$$

Clearly $I_1(\varepsilon) \leq \eta$. Using the Cauchy-Schwarz inequality, we estimate the second term by

$$\begin{aligned}
I_2(\varepsilon) &= \frac{1}{\varphi_R(\varepsilon)} \| 1_{\mathcal{T}(\varepsilon)^c}(X) \, \log \mu(B(X,\varepsilon)) \|_{L_s(\mathbb{P})} \\
&\qquad\qquad \leq \frac{1}{\varphi_R(\varepsilon)} \mathbb{P}(X \in \mathcal{T}(\varepsilon)^c)^{1/2s} \, \| \log \mu(B(X,\varepsilon)) \|_{L_{2s}(\mathbb{P})}.
\end{aligned}$$

Note that by the previous lemma, $\| \log \mu(B(X,\varepsilon)) \|_{L_{2s}(\mathbb{P})} \lesssim 2\varphi(\varepsilon/2)$ as $\varepsilon \downarrow 0$. Due to Anderson's inequality one has $\varphi_R(\varepsilon) \gtrsim \varphi(\varepsilon)$ ($\varepsilon \downarrow 0$). Since

$$\varphi(\varepsilon) \approx \varphi(\varepsilon/2) \text{ and } \lim_{\varepsilon \downarrow 0} \mathbb{P}(X \in \mathcal{T}(\varepsilon)^c) = 0,$$

it follows that

$$\lim_{\varepsilon \downarrow 0} I_2(\varepsilon) = 0.$$

Furthermore, $\lim_{\varepsilon\downarrow 0} I_3(\varepsilon) = 0$ by assumption (7.2). Hence,

$$\Big\|\frac{-\log\mu(B(X,\varepsilon))}{\varphi_R(\varepsilon)} - 1\Big\|_{L_s(\mathbb{P})} \lesssim \eta \quad (\varepsilon \downarrow 0).$$

Since $\eta \in (0,1)$ was arbitrary, we are done. □

Remark 7.1.8. Let us assume that $\varphi(\varepsilon) \approx \varphi(2\varepsilon)$ $(\varepsilon \downarrow 0)$. As a consequence of the previous theorem, we conclude that whenever there exists a gauge function φ_R satisfying (7.2), then one may choose φ_R equal to

$$\tilde\varphi_R(\varepsilon) := \mathbb{E}[-\log\mu(B(X,\varepsilon))], \quad \varepsilon > 0.$$

We keep the definition of $\tilde\varphi_R$.

Lemma 7.1.9. *The function $\tilde\varphi_R : \mathbb{R}_+ \to \mathbb{R}_+$ is convex, one-to-one and onto.*

Proof. Denote by ϕ_x, $x \in E$, the function

$$\phi_x : \mathbb{R}_+ \to (0,\infty],\ \varepsilon \mapsto -\log\mu(B(x,\varepsilon)).$$

It is classical that the function ϕ_x is convex. In fact, one has by Corollary 2.3.4 and basic analysis that the maps $\varepsilon \mapsto \Phi^{-1}(\mu(B(x,\varepsilon)))$ and $\log\circ\Phi$ are concave on $\mathbb{R}_+$ and $[-\infty,\infty)$, respectively. Consequently,

$$\varepsilon \mapsto \phi_x(\varepsilon) = -\log\circ\Phi\circ\Phi^{-1}(\mu(B(x,\varepsilon)))$$

is convex on $\mathbb{R}_+$ for all $x \in E$. We conclude that $\varepsilon \mapsto \tilde\varphi_R(\varepsilon) = \mathbb{E}[\phi_X(\varepsilon)]$ is convex on $\mathbb{R}_+$. The finiteness of $\tilde\varphi_R$ on $\mathbb{R}_+$ follows from Lemma 7.1.6. By Anderson's inequality, one has $\tilde\varphi_R \geq \varphi$. Therefore, $\tilde\varphi_R$ maps into $\mathbb{R}_+$ and $\lim_{\varepsilon\downarrow 0}\tilde\varphi_R(\varepsilon) = \infty$. Note that $\lim_{t\to\infty}\phi_x(t) = 0$. By dominated convergence, it follows that $\lim_{t\to\infty}\tilde\varphi_R(t) = 0$. Putting everything together implies the assertion. □

Lemma 7.1.10. *For $\varepsilon > 0$, one has*

$$\tilde\varphi_R(\varepsilon) \geq \varphi(\varepsilon/\sqrt{2}).$$

Proof. Denote by $\tilde X$ a μ-distributed r.e. that is independent of X. One has for $\varepsilon > 0$,

$$\begin{aligned}\mathbb{E}[\log\mu(B(X,\varepsilon))] &= \mathbb{E}[\log\mathbb{P}(\|X-\tilde X\| \leq \varepsilon|X)]\\ &\leq \log\mathbb{E}[\mathbb{P}(\|X-\tilde X\| \leq \varepsilon|X)]\\ &= \log\mathbb{P}(\|X-\tilde X\| \leq \varepsilon),\end{aligned}$$

where the inequality follows from Jensen's inequality. Note that $X - \tilde{X}$ and $\sqrt{2}X$ are both centered Gaussian processes with the same covariance operator. Therefore, $\mathcal{L}(\sqrt{2}X) = \mathcal{L}(X - \tilde{X})$ and one gets

$$\mathbb{E}[-\log \mu(B(X, \varepsilon))] \geq -\log \mu(B(0, \varepsilon/\sqrt{2})) = \varphi(\varepsilon/\sqrt{2}).$$

□

Corollary 7.1.11. *Suppose φ is regularly varying at 0 with index $-\alpha < 0$, i.e. there exists a slowly varying function l such that $\varphi(\varepsilon) = \varepsilon^{-\alpha} l(1/\varepsilon)$. Moreover, we assume existence of a function φ_R satisfying (7.2). Let c_φ and C_φ as in Lemma 7.1.3 when taking $\psi = \varphi$. Then*

$$2^{\alpha/2} \leq c_\varphi \leq C_\varphi \leq 2^{1+\alpha}.$$

In particular, the small ball probabilities around a random center have different asymptotics than those around 0.

7.2 Small ball probabilities in Hilbert spaces

Let μ denote a centered Gaussian measure on a separable Hilbert space $(H, \langle\cdot,\cdot\rangle)$ and let X be a μ-distributed random element. Without loss of generality, we assume that $\mathrm{supp}(\mu) = H$. Otherwise, we can shrink the space H to $\mathrm{supp}(\mu)$, which is a closed linear subspace of H. We denote by $C_\mu : H \to H$ the corresponding covariance operator, i.e. the self-adjoint, positive semidefinite and compact operator C_μ satisfying

$$\langle f, C_\mu g\rangle = \mathbb{E}\big[\langle f, X\rangle\langle g, X\rangle\big], \quad f, g \in H.$$

By the spectral theorem, there exist a countable index set $I \subset \mathbb{N}$ and a complete orthonormal system $\{e_j\}_{j\in I}$ of eigenvectors in H with corresponding positive eigenvalues $\{\lambda_j\}_{j\in I}$. We assume that the eigenvalues λ_j, $j \in I$, are ordered by their size.

The exact asymptotic behavior of the small ball probabilities for fixed centers was derived by Sytaya in 1974 [63] (see also Lifshits [49]). To state the theorem, we need some more notations. Let $x = \sum_{j\in I} \lambda_j^{1/2} e_j\, x_j \in H$, $x_j \in \mathbb{R}$, be an element of H and let

$$\begin{aligned}\Lambda_x(\theta) &= \log \mathbb{E}[e^{\theta\|X-x\|^2}] \\ &= \sum_{j\in I}\Big[-\frac{1}{2}\log(1-2\theta\lambda_j) + \frac{\theta\lambda_j}{1-2\theta\lambda_j}\, x_j^2\Big], \quad \theta \leq 0.\end{aligned}$$

The second equality holds due to Proposition 6.2.5. Then for $\theta \le 0$

$$\Lambda_x'(\theta) = \frac{d}{d\theta}\Lambda_x(\theta) = \sum_{j\in I}\Big[\frac{\lambda_j}{1-2\theta\lambda_j} + \frac{\lambda_j}{(1-2\theta\lambda_j)^2}\,x_j^2\Big] \text{ and}$$

$$\Lambda_x''(\theta) = \frac{d^2}{d\theta^2}\Lambda_x(\theta) = \sum_{j\in I}\Big[\frac{2\lambda_j^2}{(1-2\theta\lambda_j)^2} + \frac{4\lambda_j^2}{(1-2\theta\lambda_j)^3}\,x_j^2\Big].$$

Theorem 7.2.1. *(Sytaya). Let $x \in H$. For $\varepsilon \in (0, \Lambda_x'(0)]$, let $\theta_0(\varepsilon)$ denote the unique value $\theta_0(\varepsilon) \in (-\infty, 0]$ solving $\Lambda_x'(\theta_0(\varepsilon)) = \varepsilon$. Then one has*

$$\mathbb{P}(\|X-x\|^2 \le \varepsilon) \sim \frac{1}{\sqrt{2\pi\theta_0(\varepsilon)^2\Lambda_x''(\theta_0(\varepsilon))}}\exp\{-\Lambda_x^*(\varepsilon)\} \quad (\varepsilon \downarrow 0),$$

where $\Lambda_x^*(\varepsilon) = \sup_{\theta\le 0}[\theta\varepsilon - \Lambda_x(\theta)]$, $\varepsilon > 0$, denotes the Legendre transform of Λ_x. In particular, $\Lambda_x^*(\varepsilon) = \theta_0(\varepsilon)\varepsilon - \Lambda(\theta_0(\varepsilon))$ for $\varepsilon \in (0, \Lambda_x(0)]$.

In this chapter, we study the mass concentration of μ at the random center X. In order to obtain results, we consider logarithmic small ball probabilities. On a non-logarithmic scale, the asymptotics of small ball probabilities depend strongly on the realization of X and a statement as the following theorem does not hold.

Theorem 7.2.2. *For μ-almost all $x \in H$, one has*

$$-\log\mu(B(x,\varepsilon)) \sim \Lambda^*(\varepsilon^2) \text{ as } \varepsilon \downarrow 0, \tag{7.3}$$

where $\Lambda^*(\varepsilon) := \sup_{\theta\le 0}[\theta\varepsilon - \Lambda(\theta)]$, $\varepsilon > 0$, is the Legendre transform of

$$\Lambda(\theta) := \mathbb{E}[\Lambda_X(\theta)] = \sum_{j\in I}\Big[-\frac{1}{2}\log(1-2\theta\lambda_j) + \frac{\theta\lambda_j}{1-2\theta\lambda_j}\Big], \quad \theta \le 0.$$

Remark 7.2.3. Suppose μ is a Gaussian measure with finite dimensional support, say of dimension n. Then the logarithmic small ball function around a random center Xsatisfies

$$-\log\mu(B(X,\varepsilon)) \sim n\log(1/\varepsilon) \quad (\varepsilon \downarrow 0).$$

It is not hard to verify that in this case (7.3) gives the correct asymptotic behavior. Thereafter, we assume that μ has infinite dimensional support and that $I = \mathbb{N}$.

Lemma 7.2.4. *Let $\{Y_j\}_{j\in\mathbb{N}}$ be a sequence of real-valued i.i.d. random variables with mean 0. Assume that the moment generating function*

$$\tilde\Lambda(\theta) = \log\mathbb{E}[e^{\theta Y_1}], \quad \theta \in \mathbb{R},$$

is finite in an open two-sided neighborhood of zero. Let

$$a_j : [0,\infty) \to [0,1], \quad j \in \mathbb{N},$$

be increasing functions such that

$$[0,\infty) \to [0,\infty),\ \gamma \mapsto \|a.(\gamma)\|_{l_1} = \sum_{j\in\mathbb{N}} a_j(\gamma)$$

is continuous and converges to infinity as $\gamma \to \infty$. Then, for every $\beta > 1/2$,

$$\lim_{\gamma\to\infty} \frac{1}{\|a.(\gamma)\|_{l_1}^{\beta}} \sum_{j\in\mathbb{N}} a_j(\gamma)\, Y_j \;=\; 0, \quad a.s.$$

Proof. Choose $\tilde{\beta}$ with $1/2 < \tilde{\beta} < \beta$ and set $t_0 = \|a.(0)\|_{l_1}$. By the continuity of $\gamma \mapsto \|a.(\gamma)\|_{l_1}$, there exist values $\gamma_t \in [0,\infty)$, $t \geq t_0$, such that

$$\|a(\gamma_t)\|_{l_1} = t. \tag{7.4}$$

Consider $S_t = \sum_{j\in\mathbb{N}} a_j(\gamma_t)\, Y_j$, $t \geq t_0$. We estimate the logarithmic moment generating function of S_t for arguments $\theta \geq 0$ by

$$\log \mathbb{E}e^{\theta S_t} = \sum_{j\in\mathbb{N}} \tilde{\Lambda}(\theta a_j(\gamma_t)) \leq \sum_{j\in\mathbb{N}} a_j(\gamma_t)\, \tilde{\Lambda}(\theta) = t\, \tilde{\Lambda}(\theta).$$

Here, the inequality follows from the convexity of $\tilde{\Lambda}$ combined with the properties $\tilde{\Lambda}(0) = 0$ and $a_j(\gamma) \in [0,1]$, $\gamma \geq 0$. Applying the exponential Chebyshev inequality gives for every $\theta \geq 0$

$$\log \mathbb{P}(S_t \geq t^{\tilde{\beta}}) \leq \log \mathbb{E}[e^{\theta S_t}] - t^{\tilde{\beta}}\theta \leq t\tilde{\Lambda}(\theta) - t^{\tilde{\beta}}\theta.$$

Now choose $\theta = \theta(t) = t^{-1/2}$. Then

$$\log \mathbb{P}(S_t \geq t^{\tilde{\beta}}) \leq t\tilde{\Lambda}(t^{-1/2}) - t^{\tilde{\beta}-1/2}.$$

Note that $\tilde{\Lambda}'(0) = \mathbb{E}Y_1 = 0$ and $\tilde{\Lambda}''(0) = \mathrm{var}(Y_1)$, and $\tilde{\Lambda}$ is at least twice continuously differentiable in some neighborhood around 0. By Taylor's formula, we conclude that

$$-\log \mathbb{P}(S_t \geq t^{\tilde{\beta}}) \gtrsim t^{\tilde{\beta}-1/2} \text{ as } t \to \infty.$$

Set $I = \mathbb{N} \cap [t_0,\infty)$. By the previous estimate, $\sum_{n\in I} \mathbb{P}(S_n \geq n^{\tilde{\beta}})$ is finite and a.s. only finitely many of the events $\{S_n \geq n^{\tilde{\beta}}\}$, $n \in I$, occur. Therefore, $\limsup_{n\to\infty} S_n/n^{\beta} \leq 0$. By symmetry, this statement is also true for $\{-S_n\}_{n\in I}$. Hence, $\{S_n/n^{\beta}\}_{n\in I}$ converges a.s. to zero.

For $t \in [n, n+1)$, $n \in I$, we use the monotonicity of $a_j(\cdot)$ to estimate

$$|S_t - S_n| \le \sum_{j\in\mathbb{N}} (a_j(\gamma_t) - a_j(\gamma_n))|Y_j| \le \sum_{j\in\mathbb{N}} (a_j(\gamma_{n+1}) - a_j(\gamma_n))|Y_j| =: Z_n.$$

Then

$$\mathbb{P}(Z_n \ge n^{\tilde\beta}) \le \frac{\mathbb{E}Z_n^2}{n^{2\tilde\beta}}$$

and

$$\begin{aligned}\mathbb{E}Z_n^2 =& \sum_{i\in\mathbb{N}}\sum_{j\in\mathbb{N}} (a_i(\gamma_{n+1}) - a_i(\gamma_n))(a_j(\gamma_{n+1}) - a_j(\gamma_n))\mathbb{E}|Y_iY_j| \\ \le& \sum_{i\in\mathbb{N}}\sum_{j\in\mathbb{N}} (a_i(\gamma_{n+1}) - a_i(\gamma_n))(a_j(\gamma_{n+1}) - a_j(\gamma_n))\mathrm{var}(Y_1) = \mathrm{var}(Y_1),\end{aligned}$$

where the equality follows from (7.4). Consequently,

$$\sum_{n\in\mathbb{N}} \mathbb{P}(Z_n \ge n^{\tilde\beta}) < \infty$$

and $Z_n \le n^{\tilde\beta}$ for all but finitely many $n \in I$, a.s. Therefore, $\{Z_n\}_{n\in I}$ is a.s. of order $o(n^\beta)$ as $n \to \infty$. Since $|S_t| \le |S_n| + Z_n$ for $n \in \mathbb{N}$ and $t \in [n, n+1)$, we conclude that $|S_t| = o(t^\beta)$ as $t \to \infty$, a.s. □

In order to prove Theorem 7.2.2, we use the relation between the moment generating function and small ball probabilities (Theorem 3.4.1). Recall that for arbitrary $x \in H$ one has

$$-\log \mu(B(x, \sqrt{\varepsilon})) \sim \Lambda_x^*(\varepsilon) \quad (\varepsilon \downarrow 0), \tag{7.5}$$

where $\Lambda_x(\theta) = \log \mathbb{E}[\exp\{\theta\|X - x\|^2\}]$, $\theta \le 0$, and $\Lambda_x^*(t) = \sup_{\theta\le 0}[\theta t - \Lambda_x(\theta)]$, $t > 0$. In our case the center X is chosen randomly and, hence, one has to show that the distribution of the random variable $\Lambda_X^*(\varepsilon)$ is concentrated at $\Lambda^*(\varepsilon)$ for small ε.

Proof of Theorem 7.2.2. As noted in Remark 7.2.3, we can restrict our attention to measures μ with infinite dimensional support. By the Karhunen-Loève expansion, one has

$$X = \sum_{j\in\mathbb{N}} \lambda_j^{1/2} X_j e_j,$$

where $\{X_j\}_{j\in\mathbb{N}} = \{\lambda_j^{-1/2}\langle e_j, X\rangle\}_{j\in\mathbb{N}}$ is a sequence of independent $\mathcal{N}(0,1)$-distributed random variables. We consider the random moment generating function

$$\Lambda_X(\theta) = \sum_{j\in\mathbb{N}} \Big[-\frac{1}{2}\log(1 - 2\theta\lambda_j) + \frac{\theta\lambda_j}{1 - 2\theta\lambda_j} X_j^2\Big], \quad \theta \le 0. \tag{7.6}$$

Clearly,

$$\Lambda_X'(\theta) = \sum_{j\in\mathbb{N}} \Big[\frac{\lambda_j}{1-2\theta\lambda_j} + \frac{\lambda_j}{(1-2\theta\lambda_j)^2} X_j^2\Big]. \tag{7.7}$$

Recall that $\Lambda(\theta) = \mathbb{E}\Lambda_X(\theta)$ and denote $\Lambda'(\theta) = \mathbb{E}\Lambda_X'(\theta)$. By interchanging integral and differentiation, one obtains that Λ' is the derivative of Λ.

Note that for $\varepsilon \in (0, \Lambda_X'(0)]$, the supremum in the Legendre transform

$$\Lambda_X^*(\varepsilon) = \sup_{\theta\le 0}[\theta\varepsilon - \Lambda_X(\theta)]$$

is attained for $\theta_0 \le 0$ with $\Lambda_X'(\theta_0) = \varepsilon$. By the continuity of Λ_X' on $(-\infty, 0]$ such a θ_0 always exists. When ε tends to 0, the corresponding parameter θ_0 will tend to $-\infty$. In order to derive results for $\Lambda_X^*(\varepsilon)$ for small ε, we need to send θ to $-\infty$ and study the quantities $\Lambda_X^*(\Lambda_X'(\theta))$ and $\Lambda_X'(\theta)$. For convenience we use $\gamma = -\theta \ge 0$ and send γ to ∞. We first consider the term

$$\begin{aligned}
\xi_X(\gamma) :&= \Lambda_X^*(\Lambda_X'(-\gamma)) = -\gamma\Lambda_X'(-\gamma) - \Lambda_X(-\gamma)\\
&= \sum_{j\in\mathbb{N}}\Big[\frac{1}{2}\log(1+2\gamma\lambda_j) + \frac{\gamma\lambda_j X_j^2}{1+2\gamma\lambda_j} - \frac{\gamma\lambda_j}{1+2\gamma\lambda_j} - \frac{\gamma\lambda_j X_j^2}{(1+2\gamma\lambda_j)^2}\Big]\\
&= \sum_{j\in\mathbb{N}}\Big[\frac{1}{2}\log(1+2\gamma\lambda_j) - \frac{\gamma\lambda_j}{1+2\gamma\lambda_j} + \frac{2(\gamma\lambda_j)^2}{(1+2\gamma\lambda_j)^2}X_j^2\Big].
\end{aligned} \tag{7.8}$$

Since $\log(1+2\gamma\lambda_j) \ge (2\gamma\lambda_j)/(1+2\gamma\lambda_j)$, the deterministic part

$$\sum_{j\in\mathbb{N}}\Big[\frac{1}{2}\log(1+2\gamma\lambda_j) - \frac{\gamma\lambda_j}{1+2\gamma\lambda_j}\Big]$$

is greater than 0. The coefficients of the random terms X_j^2 in (7.8) satisfy the assumptions of Lemma 7.2.4. Therefore, one has

$$\xi_X(\gamma) \sim \mathbb{E}\xi_X(\gamma) =: \xi(\gamma) \text{ as } \gamma \to \infty, \text{ a.s.}$$

It remains to show that $\Lambda_X'(-\gamma)$ is sufficiently close to its expectation. It will be sufficient to show that $|\Lambda_X'(-\gamma) - \Lambda'(-\gamma)| = o(\xi(\gamma)/\gamma)$ as $\gamma \to \infty$, a.s. By Chebyshev's inequality and the estimate

$$\operatorname{var}\big(\gamma\Lambda_X'(-\gamma)\big) = \sum_{j\in\mathbb{N}} \frac{(\gamma\lambda_j)^2}{(1+2\gamma\lambda_j)^4}\operatorname{var}(X_1^2) \le \xi(\gamma),$$

it follows that

$$\mathbb{P}\big(|\Lambda_X'(-\gamma) - \mathbb{E}\Lambda_X'(-\gamma)| \ge \xi(\gamma)^{4/5}/\gamma\big) \le \frac{\operatorname{var}(\gamma\Lambda_X'(-\gamma))}{\xi(\gamma)^{8/5}} \le \xi(\gamma)^{-3/5}.$$

We choose a sequence $\{\gamma_n\}_{n\in\mathbb{N}}$ in $\mathbb{R}_+$ such that $\xi(\gamma_n) = n^2$ for $n \in \mathbb{N}$. Since $\sum_{n\in\mathbb{N}} \xi(\gamma_n)^{-3/5} < \infty$, we conclude by the Borel-Cantelli Lemma that a.s.

$$|\Lambda'_X(-\gamma_n) - \mathbb{E}\Lambda'_X(-\gamma_n)| \leq \xi(\gamma_n)^{4/5}/\gamma_n$$

for all but finitely many $n \in \mathbb{N}$. Hence, using the monotonicity of Λ^*_X on $\mathbb{R}_+$ gives

$$\begin{aligned}\Lambda^*_X\Big(\Lambda'(-\gamma_n) + \frac{\xi(\gamma_n)^{4/5}}{\gamma_n}\Big) &\gtrsim \Lambda^*_X\Big(\Lambda'_X(-\gamma_n) + \frac{2\xi(\gamma_n)^{4/5}}{\gamma_n}\Big)\\ &= \sup_{\theta\leq 0}[(\Lambda'_X(-\gamma_n) + 2\xi(\gamma_n)^{4/5}/\gamma_n)\,\theta - \Lambda_X(\theta)] \\ &\geq \xi_X(\gamma_n) - 2\xi(\gamma_n)^{4/5} \sim \xi(\gamma_n) \quad (n\to\infty,\ \text{a.s.}),\end{aligned} \tag{7.9}$$

where the last inequality is obtained by taking $\theta = -\gamma_n$. On the other hand, one has

$$\Lambda^*_X\Big(\Lambda'(-\gamma_n) + \frac{\xi(\gamma_n)^{4/5}}{\gamma_n}\Big) \lesssim \Lambda^*_X\big(\Lambda'_X(-\gamma_n)\big) = \xi_X(\gamma_n) \sim \xi(\gamma_n) \tag{7.10}$$

as $n \to \infty$, a.s. For $n \in \mathbb{N}$, set $\varepsilon_n = \Lambda'(-\gamma_n) + \xi(\gamma_n)^{4/5}/\gamma_n$. By (7.9) and (7.10) one has

$$\Lambda^*_X(\varepsilon_n) \sim \xi(\gamma_n) \quad (n\to\infty,\ \text{a.s.}).$$

By the monotonicity of Λ^*, one has $\xi(\gamma_n) = \Lambda^*\big(\Lambda'(-\gamma_n)\big) \geq \Lambda^*(\varepsilon_n)$. Moreover,

$$\begin{aligned}\Lambda^*(\varepsilon_n) &= \sup_{\theta\leq 0}\big[\theta\varepsilon_n - \Lambda(\theta)\big] \geq -\gamma_n\varepsilon_n - \Lambda(-\gamma_n)\\ &= -\gamma_n\Lambda'(-\gamma_n) - \Lambda(-\gamma_n) - \xi(\gamma_n)^{4/5}\\ &= \xi(\gamma_n) - \xi(\gamma_n)^{4/5} \sim \xi(\gamma_n) = \Lambda^*(\Lambda'(-\gamma_n)) \quad (n\to\infty).\end{aligned}$$

Hence, $\Lambda^*_X(\varepsilon_n) \sim \Lambda^*(\varepsilon_n)$ ($n \to \infty$, a.s.) and one has with (7.5)

$$-\log(\mu(B(x,\sqrt{\varepsilon_n}))) \sim \Lambda_X(\varepsilon_n) \sim \Lambda^*(\varepsilon_n) \quad (n\to\infty)$$

for μ-almost every $x \in H$. The general result follows by the monotonicity of the small ball probabilities, and the moment generating function and the fact that $\lim_{n\to\infty} \Lambda^*(\varepsilon_{n+1})/\Lambda^*(\varepsilon_n) = 1$. In fact,

$$\frac{\Lambda^*(\varepsilon_{n+1})}{\Lambda^*(\varepsilon_n)} \sim \frac{\xi(\gamma_{n+1})}{\xi(\gamma_n)} = \frac{(n+1)^2}{n^2} \to 1 \quad (n\to\infty).$$

□

The Legendre transform Λ^* is convex. Furthermore, it is non-negative and attains 0 on the set $[\Lambda'(0), \infty)$. Note that $\Lambda'(0) = 2\sum_{j\in\mathbb{N}} \lambda_j$. Consider the map

$$\begin{aligned}\varphi_R : \big(0, (2\sum_{j\in\mathbb{N}} \lambda_j)^{1/2}\big] &\to [0,\infty) \\ \varepsilon \qquad &\mapsto \Lambda^*(\varepsilon^2).\end{aligned}$$

Since Λ^* is finite on $\mathbb{R}_+$ and converges to ∞ as $\varepsilon \downarrow 0$, it follows that φ_R is one-to-one and onto. Motivated by the previous theorem we call φ_R the *small ball function for random centers.*

Theorem 7.2.5. *Assume that the eigenvalues $\{\lambda_j\}_{j\in\mathbb{N}}$ are regularly varying with index $-\alpha < -1$. Then*

$$\varphi_R(\varepsilon) \sim \Big(\frac{\alpha+1}{\alpha}\Big)^{\frac{\alpha}{\alpha-1}} \varphi(\varepsilon) \text{ as } \varepsilon \downarrow 0.$$

In the proof of Theorem 7.2.5, we use de Bruijn's Tauberian Theorem (de Bruijn [10]; Bingham et al. [6], Theorem 4.12.9):

Theorem 7.2.6. *Let ν be a probability measure on $(0,\infty)$ with logarithmic Laplace transform*

$$\Gamma(\gamma) = \log \int_0^\infty e^{-\gamma x} d\nu(x), \quad \gamma > 0.$$

Let ϕ be regularly varying at zero with index $-\beta < 0$. Set $\psi(x) = \phi(x)/x$ and denote by $\phi^{\leftarrow}$ and $\psi^{\leftarrow}$ an asymptotic inverse of $\phi \in \mathcal{R}_{-\beta}(0)$ and $\psi \in \mathcal{R}_{-\beta-1}(0)$, respectively. Then, for $B > 0$,

$$-\log \nu(0,\varepsilon] \sim B/\phi^{\leftarrow}(1/\varepsilon) \text{ as } \varepsilon \downarrow 0$$

if and only if

$$-\Gamma(\gamma) \sim (1+\beta)(B/\beta)^{\beta/1+\beta}/\psi^{\leftarrow}(\gamma) \text{ as } \gamma \to \infty.$$

Corollary 7.2.7. *Let ν_1 and ν_2 be probability measures on the Borel sets of $(0,\infty)$. Let Γ_1 and Γ_2 denote the log-Laplace transforms of ν_1 and ν_2 and suppose that $-\Gamma_1$ is regularly varying with index $\alpha \in (0,1)$. For $\kappa > 0$,*

$$-\Gamma_2(\gamma) \sim -\kappa\,\Gamma_1(\gamma) \quad (\gamma \to \infty)$$

implies

$$-\log \nu_2(0,\varepsilon] \sim -\kappa^{1/(1-\alpha)} \log \nu_1(0,\varepsilon] \quad (\varepsilon \downarrow 0).$$

Proof. Choose

$$\phi(x) = x \Big(\frac{1}{-\Gamma_1}\Big)^{\leftarrow}(x),$$

where $\Big(\frac{1}{-\Gamma_1}\Big)^{\leftarrow} \in \mathcal{R}_{-1/\alpha}(0)$ is an arbitrary asymptotic inverse of $\frac{1}{-\Gamma_1} \in \mathcal{R}_{-\alpha}$. Then $\phi \in \mathcal{R}_{1-1/\alpha}(0)$. Note that $-\beta := 1 - 1/\alpha < 0$ and, hence, the Tauberian Theorem is applicable. With $\phi^{\leftarrow}$, ψ and $\psi^{\leftarrow}$ as in the Tauberian Theorem, we may write

$$-\Gamma_1(\gamma) \sim (1+\beta)(B_1/\beta)^{\beta/1+\beta}/\psi^{\leftarrow}(\gamma)$$

for some appropriate constant $B_1 > 0$. For $B_2 = B_1\kappa^{(1+\beta)/\beta} = B_1\kappa^{1/(1-\alpha)}$, one has

$$\begin{aligned}\Gamma_2(\gamma) &\sim \kappa\,\Gamma_1(\gamma)\\ &\sim \kappa(1+\beta)(B_1/\beta)^{\beta/1+\beta}/\psi^{\leftarrow}(\gamma)\\ &= (1+\beta)(B_2/\beta)^{\beta/1+\beta}/\psi^{\leftarrow}(\gamma).\end{aligned}$$

Consequently, with the Tauberian Theorem,

$$\begin{aligned}-\log\nu_2(0,\varepsilon] &\sim B_2/\phi^{\leftarrow}(1/\varepsilon)\\ &= \kappa^{1/(1-\alpha)}B_1/\phi^{\leftarrow}(1/\varepsilon) \sim -\kappa^{1/(1-\alpha)}\log\nu_1(0,\varepsilon],\end{aligned}$$

as $\varepsilon \downarrow 0$. □

Proof of Theorem 7.2.5. Let X and $\tilde{X}$ be independent μ-distributed random elements in H. As noted before the log-Laplace transform at the random center X is given by

$$\begin{aligned}\Gamma_X(\gamma) :=&\ \log\mathbb{E}[e^{-\gamma\|\tilde{X}-X\|^2}|X]\\ =&\ -\sum_{j\in\mathbb{N}}\Big[\frac{1}{2}\log(1+2\gamma\lambda_j) + \frac{\gamma\lambda_j}{1+2\gamma\lambda_j}\,X_j^2\Big], \quad \gamma \ge 0,\end{aligned}$$

where $\{X_j\}_{j\in\mathbb{N}}$ is a sequence of independent standard normals. By Lemma 7.2.4, one has

$$-\Gamma_X(\gamma) \sim \sum_{j\in\mathbb{N}}\Big[\frac{1}{2}\log(1+2\gamma\lambda_j) + \frac{\gamma\lambda_j}{1+2\gamma\lambda_j}\Big] =: -\Gamma(\gamma) \quad (\gamma\to\infty)$$

for a.e. realization of X. We apply Corollary 7.2.7 to prove the theorem. It remains to show that $-\Gamma_0$ and $-\Gamma$ are regularly varying and to relate the asymptotics of Γ and Γ_0 appropriately.

Consider first

$$\Sigma_1(\gamma) := \sum_{j\in\mathbb{N}}\frac{\gamma\lambda_j}{1+2\gamma\lambda_j}, \quad \gamma \ge 0.$$

Let $\lambda_t := \lambda_{\lceil t \rceil} = t^{-\alpha}\, l(t)$, $t > 0$, and suppose that $\tilde{\lambda} : \mathbb{R}_+ \to \mathbb{R}_+$ is a measurable function with $\tilde{\lambda}_t \sim \lambda_t$ as $t \to \infty$. Then one has for any $t_0 \geq 0$,

$$\Sigma_1(\gamma) = \int_0^\infty \frac{\gamma\lambda_t}{1+2\gamma\lambda_t}\,dt \sim \int_{t_0}^\infty \frac{\gamma\lambda_t}{1+2\gamma\lambda_t}\,dt \quad (\gamma \to \infty), \tag{7.11}$$

since the integrand is bounded by $1/2$ and the integral converges to ∞ when γ tends to ∞. Due to the fact that

$$\frac{\gamma\lambda_t}{1+2\gamma\lambda_t}\,\frac{1+2\gamma\tilde{\lambda}_t}{\gamma\tilde{\lambda}_t}$$

converges uniformly for all $\gamma \geq 0$ to 1 as $t \to \infty$ combined with (7.11), one concludes

$$\Sigma_1(\gamma) \sim \int_0^\infty \frac{\gamma\tilde{\lambda}_t}{1+2\gamma\tilde{\lambda}_t}\,dt \quad (\gamma \to \infty). \tag{7.12}$$

Hence, replacing λ by an asymptotically equivalent $\tilde{\lambda}$ yields the same asymptotics for Σ_1. Consequently, for $\kappa \in \mathbb{R}_+$,

$$\begin{aligned}\Sigma_1(\kappa\gamma) &= \int_0^\infty \frac{\kappa\gamma\lambda_{\tilde{t}}}{1+2\kappa\gamma\lambda_{\tilde{t}}}\,d\tilde{t} \\ &= \kappa^{1/\alpha}\int_0^\infty \frac{\gamma t^{-\alpha} l(\kappa^{1/\alpha}t)}{1+2\gamma t^{-\alpha} l(\kappa^{1/\alpha}t)}\,dt \\ &\sim \kappa^{1/\alpha}\Sigma_1(\gamma) \quad (\gamma \to \infty),\end{aligned}$$

where the second equality follows from substituting $\tilde{t} = \kappa^{1/\alpha}t$. One concludes that Σ_1 is regularly varying with index α^{-1}.

Consider now

$$\Sigma_2(\gamma) := \sum_{j\in\mathbb{N}} \frac{1}{2}\log(1+2\gamma\lambda_j), \quad \gamma \geq 0.$$

By using the inequality $\log(1+x) \geq x/(x-1)$, we see that $\Sigma_2(\gamma) \geq \Sigma_1(\gamma)$. Hence, Σ_2 grows at least as a power. We conclude again that a finite number of summands in Σ_2 has no effect on its asymptotic behavior when $\gamma \to \infty$. In analogy to the argumentation above, one has

$$\Sigma_2(\gamma) \sim \int_0^\infty \frac{1}{2}\log(1+2\gamma\tilde{\lambda}_t)\,dt \quad (\gamma \to \infty) \tag{7.13}$$

for any $\tilde{\lambda} : \mathbb{R}_+ \to \mathbb{R}_+$ that is asymptotically equivalent to λ and locally integrable.

By basic properties of regularly varying functions (see TheoremA.3), there exists a positive slowly varying function $\tilde{l} \in C^1[0,\infty)$ that satisfies $\lambda_t \sim t^{-\alpha}\tilde{l}(t)$ and $\tilde{l}'(t) = o(\tilde{l}(t)/t)$ as $t \to \infty$. Set $\tilde{\lambda}_t := t^{-\alpha}\tilde{l}(t)$, $t \geq 0$. Then

$$\Sigma_2(\gamma) \sim \int_0^\infty \log(1+2\gamma\tilde{\lambda}_t)\,dt = \big[t\frac{1}{2}\log(1+2\gamma\tilde{\lambda}_t)\big]_0^\infty - \int_0^\infty \frac{\gamma t\tilde{\lambda}_t'}{1+2\gamma\tilde{\lambda}_t}\,dt. \tag{7.14}$$

Observe that $\big[t\frac{1}{2}\log(1+2\gamma\tilde{\lambda}_t)\big]_0^\infty = 0$. Since $l'(t) = o(l(t)/t)$, we have

$$\tilde{\lambda}_t' = -\alpha t^{-\alpha-1}\tilde{l}(t) + t^{-\alpha}\tilde{l}'(t) \sim -\alpha\tilde{\lambda}_t/t \text{ as } t \to \infty.$$

Combined with (7.14) and (7.12) this yields

$$\Sigma_2(\gamma) \sim \alpha\Sigma_1(\gamma).$$

Thus,

$$-\Gamma(\gamma) \sim (\alpha+1)\Sigma_1(\gamma) \sim -\frac{\alpha+1}{\alpha}\Gamma_0(\gamma).$$

Moreover, $-\Gamma$ and $-\Gamma_0$ are regularly varying with index $1/\alpha$ and we finish the proof by applying Corollary 7.2.7. □

Recall that the asymptotics of the small ball function can be derived from eigenvalues. As a consequence of Bronski [9], one has

Lemma 7.2.8. *Suppose that the eigenvalues $\{\lambda_j\}_{j\in\mathbb{N}}$ satisfy*

$$\lambda_j \sim \kappa\, j^{-\alpha} \quad (j \to \infty)$$

for some $\kappa > 0$ and $\alpha > 1$. Then

$$\varphi(\varepsilon) \sim \frac{\alpha-1}{2}\Big(\frac{\pi}{\alpha\sin(\frac{\pi}{\alpha})}\Big)^{\frac{\alpha}{\alpha-1}}\kappa^{\frac{1}{\alpha-1}}\varepsilon^{-\frac{2}{\alpha-1}} \quad (\varepsilon \downarrow 0).$$

With Theorem 7.2.5 we obtain

Corollary 7.2.9. *If the eigenvalues are as in the previous lemma, then*

$$\varphi_R(\varepsilon) \sim \frac{\alpha-1}{2}\Big(\frac{(\alpha+1)\pi}{\alpha^2\sin(\frac{\pi}{\alpha})}\Big)^{\frac{\alpha}{\alpha-1}}\kappa^{\frac{1}{\alpha-1}}\varepsilon^{-\frac{2}{\alpha-1}} \quad (\varepsilon \downarrow 0).$$

In Section 6.4, we quoted the asymptotics of the eigenvalues of fractional Brownian motion and integrated Brownian motion. These results yield

$$\varphi_R(\varepsilon) \sim \frac{\gamma}{2}\Big(\frac{(\gamma+2)}{(\gamma+1)^2\sin(\frac{\pi}{\gamma+1})}\Big)^{\frac{\gamma+1}{\gamma}}\big(\sin(\tfrac{\pi\gamma}{2})\Gamma(\gamma+1)\big)^{\frac{1}{\gamma}}\varepsilon^{-\frac{2}{\gamma}}$$

for fractional Brownian motion to the index $\gamma \in (0,2)$ in $L_2[0,1]$ and

$$\varphi_R(\varepsilon) \sim \frac{2m+1}{2}\Big(\frac{2m+3}{4(m+1)^2\sin(\frac{\pi}{2m+2})}\Big)^{\frac{2m+2}{2m+1}}\varepsilon^{-\frac{2}{2m+1}}$$

for m-times integrated Brownian motion in $L_2[0,1]$. In particular, one has

$$\varphi_R(\varepsilon) \sim \frac{9}{32}\,\varepsilon^{-2}$$

for Wiener measure on $L_2[0,1]$.

7.3 Small ball probabilities for Brownian motion

Let $d \in \mathbb{N}$. Suppose $S = \mathbb{R}^d$ is equipped with an arbitrary norm $|\cdot|_S$. For a set I and function $g : I \to \mathbb{R}$, we denote

$$\|g\|_I := \|g\|_{I,S} = \sup_{t\in I}|g(t)|_S.$$

For $x \in \mathbb{R}^d$, let P^x be d-dimensional Wiener measure on $C([0,\infty),S)$ with starting point $x \in \mathbb{R}^d$. Analogously, we let P^{x,t_0}, $x \in \mathbb{R}^d$, $t_0 \geq 0$, be d-dimensional Wiener measure on $C([t_0,\infty),S)$ with starting point $x \in \mathbb{R}^d$ started at time $t_0 \geq 0$. Here, we equip $C([0,\infty),S)$ and $C([t_0,\infty),S)$ with the topology of uniform convergence on compacts. Analogously to the previous section, we study the asymptotics of

$$-\log \mathbb{P}^0(\|W - w\|_{[0,1]} \leq \varepsilon)$$

for a P^0-typical w when ε tends to 0. In order to infer results for this problem we consider the asymptotics of

$$-\log \mathbb{P}^0(\|W - w\|_{[0,t]} \leq 1) \quad \text{as } t \to \infty,$$

for a P^0-typical $w \in C([0,\infty),S)$. Then we use the scaling properties of Brownian motion to conclude on the original problem. The main result of this paragraph is

Theorem 7.3.1. *Let $\lambda_1 > 0$ be the principal eigenvalue of the Dirichlet problem on the domain* $\mathrm{int}(B_S(0,1))$*. There exists $\kappa \in [2\lambda_1, 8\lambda_1]$ such that*

$$-\log \mathbb{P}^0(\|W - w\|_{[0,t]} \leq 1) \sim \kappa t$$

as $t \to \infty$, for P^0-a.e. w in $C([0,\infty),S)$. Furthermore, the maps ($\varepsilon > 0$)

$$\begin{aligned} \phi_\varepsilon : C([0,\infty),S) &\to [0,\infty) \\ w &\mapsto \phi_\varepsilon(w) = -\log \mathbb{P}^0(\|W - w\|_{[0,1]} \leq \varepsilon) \end{aligned}$$

considered as random variables on the canonical Wiener space satisfy

$$\lim_{\varepsilon\downarrow 0} \varepsilon^2 \phi_\varepsilon = \kappa \quad \text{in probability.}$$

Denote by $\mathcal{M}(S)$ the set of positive finite measures on the Borel sets of S. For $\nu \in \mathcal{M}(S)$ and $t_0 \geq 0$, we define P^{ν,t_0} to be the measure on the Borel sets of $C([t_0,\infty),S)$ satisfying

$$P^{\nu,t_0}(A) = \int_{\mathbb{R}^d} P^{z,t_0}(A)\,\nu(dz), \quad A \in \mathcal{B}(C([t_0,\infty),S)).$$

Let $f : [0,\infty) \to S$ be an arbitrary continuous function with $f(0) = 0$ and let $\mu \in \mathcal{M}(S)$ be the measure defined by

$$\mu(A) = \mathbb{E}^0[1_{\{\|W-f\|_{[0,1]}\leq 1\}}\, 1_A(W_1)], \quad A \in \mathcal{B}(\mathbb{R}^d).$$

Before we start with the proof of the theorem, we derive some lemmas.

Lemma 7.3.2. *One has*

$$\operatorname{supp}(\mu) = B_S(f(1),1).$$

Proof. Let $x \in \operatorname{int} B_S(f(1),1)$ and $\varepsilon > 0$. We need to show that $\mu(B_S(x,\varepsilon)) > 0$. Without loss of generality, we assume that $\varepsilon < 1 - |x|_S$. Since the set of functions $C^\infty([0,1],S)$ is dense in $(C([0,1],S), \|\cdot\|_{[0,1]})$, there exists a function $f_1 \in C^\infty([0,1],S)$ with $\|f - f_1\|_{[0,1]} \leq \varepsilon/2$ and $f_1(0) = 0$. Note that

$$\begin{aligned}\mu(B(x,\varepsilon)) &\geq \mathbb{P}^0(\|W_t - (f(t)+tx)\|_{[0,1]} \leq \varepsilon) \\ &\geq \mathbb{P}^0(\|W_t - (f_1(t)+tx)\|_{[0,1]} \leq \varepsilon/2).\end{aligned}$$

By the estimate of shifted balls (Lemma 2.2.4) it follows that

$$\mu(B(x,\varepsilon)) \geq \mathbb{P}^0(\|W\|_{[0,1]} \leq \varepsilon/2)\, \exp\Big\{-\frac{\int_0^1 |\dot{f}_1(s)+x|^2\,ds}{2}\Big\} > 0.$$

□

Lemma 7.3.3. *There exists a log-concave version of the Radon-Nikodym derivative $\frac{d\mu}{d\lambda^d}$, where λ^d denotes d-dimensional Lebesgue measure.*

Proof. Let

$$J(n) = \Big\{\frac{i}{2^n} : i = 0,\ldots,2^n\Big\}, \quad n \in \mathbb{N},$$

and let $\mu^{(n)}$, $n \in \mathbb{N}$, denote the measures on $\mathbb{R}^d$ with

$$\mu^{(n)}(A) = \mathbb{E}^0[1_{\{\|W-f\|_{(J(n))}<1\}} 1_A(W_1)], \quad A \in \mathcal{B}(\mathbb{R}^d).$$

First we show that there exists a log-concave version of the Radon-Nikodym derivative $\frac{d\mu^{(n)}}{d\lambda^d}$. Let

$$p_t(z) = (2\pi t)^{-d/2} e^{-|z|^2/(2t)}, \quad t > 0,\ z \in \mathbb{R}^d,$$

and let $K = \operatorname{int} B_S(0,1)$ be the open unit ball in S. Fix $n \in \mathbb{N}$. The Radon-Nikodym derivative $\frac{d\mu^{(n)}}{d\lambda^d}$ can be constructed as follows. The function

$$g_1^{(n)}(z) := 1_{f(2^{-n})+K}(z) \cdot p_{2^{-n}}(z), \quad z \in \mathbb{R}^d,$$

is the density of the measure $\mu_1^{(n)}$ defined by

$$\mu_1^{(n)}(A) := P^x\big\{w \in C([0,\infty), S) : w_{2^{-n}} \in A \cap (f(2^{-n}) + K)\big\}$$

for $A \in \mathcal{B}(\mathbb{R}^d)$. We continue inductively. Suppose $g_i^{(n)}$, $i \in \{1, \dots, 2^n - 1\}$, is defined already. Then let

$$g_{i+1}^{(n)}(z) := 1_{f((i+1)2^{-n})+K}(z) \cdot \big(p_{2^{-n}} * g_i^{(n)}\big)(z), \quad z \in \mathbb{R}^d,$$

where $*$ denotes the convolution of functions, and

$$\mu_{i+1}^{(n)}(A) := P^{\mu_i, i2^{-n}}\big\{w \in C([i2^{-n},\infty), S) : \\ w_{(i+1)2^{-n}} \in A \cap (f((i+1)2^{-n}) + K)\big\}$$

for $A \in \mathcal{B}(\mathbb{R}^d)$. It follows by construction that $\mu_{2^n}^{(n)}$ is equal to $\mu^{(n)}$. Moreover, $g^{(n)} := g_{2^n}^{(n)}$ is the density of $\mu^{(n)}$. Note that $p_t(\cdot)$ is log-concave for any $t > 0$. By Brascamp and Lieb [8] (Theorem 1.3), the convolution of two log-concave functions is again log-concave. Since log-concavity is also maintained by multiplication with 1_A for convex sets $A \subset S$, it follows that $g^{(n)}$ is log-concave.

Moreover, the construction of $g^{(n)}$ implies that $n \mapsto g^{(n)}(z)$ is monotonically decreasing for all fixed $z \in \mathbb{R}^d$. Set

$$g(z) := \lim_{n\to\infty} g^{(n)}(z) = \inf_{n\to\infty} g^{(n)}(z), \quad z \in \mathbb{R}^d.$$

Since g is an infimum of log-concave functions, it is log-concave as well. It remains to prove that g is a density of μ. Note that for any $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\lim_{n\to\infty} \int g_n 1_A \, d\lambda^d = \int g 1_A \, d\lambda^d,$$

by dominated convergence. On the other hand,

$$\int 1_A \, d\mu^{(n)} = \mathbb{E}^0[1_{\{\|W-f\|_{J(n)}<1\}} \, 1_A(W_1)]$$

converges to $\mathbb{E}^0[1_{\{\|W-f\|_J<1\}} \, 1_A(W_1)]$ with $J = \bigcup_{n\in\mathbb{N}} J_n$. Due to Corollary 2.3.4, one has

$$\mathbb{P}^0(\|W-f\|_{[0,1]} = 1) = 0$$

and, hence,

$$\int 1_A \, g \, d\lambda^d = \mathbb{E}^0[1_{\|W-f\|_{[0,1]}\le 1\}} \, 1_A(W_1)]$$

since J is dense in $[0,1]$ and W and f are continuous. □

Now the previous two lemmas are used to prove

Proposition 7.3.4. *For every $\eta \in (0,1)$ there exists $\varepsilon_0 > 0$ such that*

$$\mu(A) \ge \varepsilon_0 \, \lambda^d(A \cap B(f(1),\eta)), \quad A \in \mathcal{B}(S).$$

Proof. By Lemma 7.3.3, there exists a log-concave function $g : S \to [0,\infty)$ with $g = 0$ on $\operatorname{int}(B_S(f(1),1))^c$ such that

$$\frac{d\mu}{d\lambda^d} = g.$$

Consider the sets

$$A_\varepsilon := g^{-1}(\varepsilon,\infty) \subset B_S(f(1),1), \quad \varepsilon > 0,$$

and denote $A := \bigcup_{\varepsilon>0} A_\varepsilon$. The measure μ is supported on the set $\bar{A}$. Since $A \subset \operatorname{int}(B_S(f(1),1))$, one has $\bar{A} = B_S(f(1),1)$ due to Lemma 7.3.2. The set A is convex due to its definition. Consequently, one has

$$A = \operatorname{int}(B_S(f(1),1))$$

by basic results of convex analysis (see Rockafellar [59], Theorem 6.3). Since $g > 0$ on A, it follows that the function g restricted to the open set A is a continuous function. Hence, for $\varepsilon > 0$ the sets $A_\varepsilon = g^{-1}(\varepsilon,\infty)$ are open subsets of S. Consequently,

$$\bigcup_{\varepsilon>0} g^{-1}(\varepsilon,\infty)$$

is an open cover of the compact set $B_S(f(1),\eta)$ for any $\eta \in (0,1)$. Hence, there exists $\varepsilon_0 > 0$ such that

$$A_{\varepsilon_0} \supset B_S(f(1),\eta).$$

Thus, $g \ge \varepsilon_0$ on $B_S(f(1),\eta)$ and the assertion follows. □

Lemma 7.3.5. *For* $\varepsilon > 0$,

$$-\log \sup_{x\in\mathbb{R}^d} \mathbb{P}^x(\|W-f\|_{[0,t]} \le 1-\varepsilon) \gtrsim -\log \mathbb{P}^0(\|W-f\|_{[0,t]} \le 1),$$

as $t \to \infty$.

Proof. For $x \in \mathbb{R}^d$ and $\varepsilon \in (0,1)$, let $\mu^{x,\varepsilon} \in \mathcal{M}(\mathbb{R}^d)$ with

$$\mu^{x,\varepsilon}(A) = \mathbb{E}^x\big[1_{\{\|W-f\|_{[0,1]}\le 1-\varepsilon\}} 1_A(W_1)\big], \quad A \in \mathcal{B}(\mathbb{R}^d).$$

For $t \ge 1$ one has by the Markov property of Brownian motion that

$$\begin{aligned}&\mathbb{P}^x(\|W-f\|_{[0,t]} \le 1-\varepsilon)\\ &\qquad = P^{\mu^{x,\varepsilon},1}\big\{w \in C([1,\infty),S) : \|w-f\|_{[1,t]} \le 1-\varepsilon\big\}.\end{aligned}$$

Analogously, one has

$$\mathbb{P}^0(\|W-f\|_{[0,t]} \le 1) = P^{\mu,1}\big\{w \in C([1,\infty),S) : \|w-f\|_{[1,t]} \le 1\big\},$$

with μ as above. Consider the set $I_\varepsilon := B_S(f(1), 1-\varepsilon) \subset \mathbb{R}^d$. As W_1 is normally distributed we conclude that there exists a constant $C < \infty$ such that

$$\mu^{x,\varepsilon}(A) \le \mathbb{E}^x\big[1_{A\cap I_\varepsilon}(W_1)\big] \le C\,\lambda^d(A\cap I_\varepsilon)$$

for all $A \in \mathcal{B}(\mathbb{R}^d)$. Note that C can be chosen uniformly for all starting points $x \in S$. On the other hand, by Proposition 7.3.4, there exists a constant $c > 0$ such that

$$\mu(A) \ge c\,\lambda^d(A\cap I_\varepsilon), \quad A \in \mathcal{B}(\mathbb{R}^d).$$

Combining both estimates gives

$$\begin{aligned}\mathbb{P}^0(\|W-f\|_{[0,t]} \le 1) &= P^{\mu,1}\big\{w \in C([1,\infty),S) : \|w-f\|_{[1,t]} \le 1\big\}\\ &\ge \frac{c}{C} \sup_{x\in\mathbb{R}^d} P^{\mu^{x,\varepsilon},1}\big\{w \in C([1,\infty),S) : \|w-f\|_{[1,t]} \le 1-\varepsilon\big\}\\ &= \frac{c}{C} \sup_{x\in\mathbb{R}^d} \mathbb{P}^x(\|W-f\|_{[0,t]} \le 1-\varepsilon).\end{aligned}$$

Due to Anderson's inequality, it follows that $\mathbb{P}^0(\|W-f\|_{[0,t]} \le 1)$ converges to 0 as $t \to \infty$. Consequently,

$$-\log \mathbb{P}^0(\|W-f\|_{[0,t]} \le 1) \lesssim -\log \sup_{x\in\mathbb{R}^d} \mathbb{P}^x(\|W-f\|_{[0,t]} \le 1-\varepsilon)$$

as $t \to \infty$. □

Proof of Theorem 7.3.1. First we prove that the existence of a constant $\kappa \in [0,\infty]$ such that

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{P}(\|W-w\|_{[0,t]} \leq 1) = -\kappa$$

for P^0-a.e. $w \in C[0,\infty)$. Here, convergence is understood in the extended real numbers. The proof is based on Kingman's subadditive ergodic Theorem.

Let $(C([0,\infty),S), \mathcal{B}(C([0,\infty),S)), P^0, \{\theta_t\}_{t\geq 0})$ be the ergodic canonical dynamical system corresponding to the Wiener process. Here,

$$\begin{aligned} \theta_t : C([0,\infty),S) &\to C([0,\infty),S) \\ w \quad &\mapsto (\theta_t w)_s = w_{s+t} - w_t. \end{aligned}$$

For $t \geq 0$, we define the function $\psi_t : C([0,\infty),S) \to (-\infty, 0]$ by

$$\psi_t(w) = \sup_{x\in\mathbb{R}^d} \log \mathbb{P}^x(\|W_u - w\|_{[0,t]} \leq 1).$$

The process $\{\psi_t(w)\}_{t\geq 0}$ is subadditive for any $w \in C([0,\infty),S)$. In fact, one has for $s,t > 0$,

$$\begin{aligned} \psi_{s+t}(w) &= \sup_{x\in S} \log \mathbb{P}^x(\|W-w\|_{[0,s+t]} \leq 1) \\ &= \sup_{x\in\mathbb{R}^d} \log \mathbb{P}^x(\|W-w\|_{[0,s]} \leq 1 \text{ and} \\ &\qquad\qquad \|W_{s+\cdot} - w_s - (\theta_s w)_\cdot\|_{[0,t]} \leq 1) \\ &\leq \sup_{x\in\mathbb{R}^d} \log \mathbb{P}^x(\|W-w\|_{[0,s]} \leq 1) \\ &\quad + \sup_{x\in\mathbb{R}^d} \log \mathbb{P}^x(\|W-\theta_s w\|_{[0,t]} \leq 1) = \psi_s(w) + \psi_t(\theta_s w). \end{aligned}$$

Hence, by Kingman's subadditive ergodic Theorem (see Krengel [39], Theorem 5.3) there exists $\kappa \in [0,\infty]$ such that for P^0-a.e. w, one has

$$\lim_{n\to\infty} \frac{1}{n} \psi_n(w) = -\kappa.$$

Since for $n \in \mathbb{N}$ and $t \in [n, n+1)$

$$\frac{n+1}{n} \frac{\psi_{n+1}(w)}{n+1} \leq \frac{1}{t} \psi_t(w) \leq \frac{n}{n+1} \frac{\psi_n(w)}{n},$$

it follows that

$$\lim_{t\to\infty} \frac{1}{t} \sup_{x\in S} \log \mathbb{P}^x(\|W-w\|_{[0,t]} \leq 1) = -\kappa.$$

We will use the scaling properties of Brownian motion. For $\eta > 0$, the map

$$\begin{aligned}\pi_\eta : C[0,\infty) &\to C[0,\infty)\\ h &\mapsto \pi_\eta(h)(t) = \eta\, h\Big(\frac{t}{\eta^2}\Big)\end{aligned}$$

is linear and satisfies $P^{\eta x} = P^x \circ \pi_\eta^{-1}$. In particular, P^0 is π_η-invariant, i.e. $P^0 = P^0 \circ \pi_\eta^{-1}$. Hence,

$$\begin{aligned}\mathbb{P}^x(\|W-w\|_{[0,t]} \le 1) &= \mathbb{P}^x(\|\pi_\eta(W-w)\|_{[0,\eta^2 t]} \le \eta)\\ &= \mathbb{P}^{\eta x}(\|W-\pi_\eta(w)\|_{[0,\eta^2 t]} \le \eta)\end{aligned} \tag{7.15}$$

and one has for P^0-a.e. $w \in C[0,\infty)$,

$$\lim_{t\to\infty} \frac{1}{t} \sup_{x\in\mathbb{R}^d} \log \mathbb{P}^x(\|W-w\|_{[0,\eta^2 t]} \le \eta) = -\kappa.$$

Now let $\varepsilon \in (0,1)$ and $\eta := 1-\varepsilon$. Then

$$\begin{aligned}-\kappa &= \lim_{t\to\infty} \frac{1}{t} \sup_{x\in\mathbb{R}^d} \log \mathbb{P}^x(\|W-w\|_{[0,(1-\varepsilon)^2 t]} \le 1-\varepsilon)\\ &= (1-\varepsilon)^2 \lim_{t\to\infty} \frac{1}{t} \sup_{x\in\mathbb{R}^d} \log \mathbb{P}^x(\|W-w\|_{[0,t]} \le 1-\varepsilon)\\ &\le (1-\varepsilon)^2 \liminf_{t\to\infty} \frac{1}{t} \log \mathbb{P}^0(\|W-w\|_{[0,t]} \le 1),\end{aligned}$$

where the inequality holds due to Lemma 7.3.5. Letting $\varepsilon \downarrow 0$ gives

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{P}^0(\|W-w\|_{[0,t]} \le 1) = -\kappa$$

for P^0-a.e. w.

Now we show convergence in probability of $\varepsilon^2\phi_\varepsilon$ as $\varepsilon \downarrow 0$. For $t > 0$ consider

$$\begin{aligned}\psi_t^{(1)} : C([0,\infty),S) &\to [0,\infty)\\ w &\mapsto \psi_t^{(1)}(w) = -\frac{1}{t} \log \mathbb{P}^0(\|W-w\|_{[0,t]} \le 1)\end{aligned}$$

and

$$\begin{aligned}\psi_t^{(2)} : C([0,\infty),S) &\to [0,\infty)\\ w &\mapsto \psi_t^{(2)}(w) = -\frac{1}{t} \log \mathbb{P}^0(\|W-w\|_{[0,1]} \le 1/\sqrt{t})\end{aligned}$$

as random variables on the canonical Wiener space. Applying the scaling property (7.15) with $\eta = 1/\sqrt{t}$, gives $\mathcal{L}(\psi_t^{(1)}) = \mathcal{L}(\psi_t^{(2)})$ for any $t > 0$. In particular, $\psi_t^{(2)} = \frac{1}{t}\,\phi_{1/\sqrt{t}}$ converges in probability to κ as $t \to \infty$.

It remains to prove the explicit bounds for κ. By Theorem 7.1.1,

$$\varepsilon^2\,\varphi(\varepsilon) \le \varepsilon^2\,\phi_\varepsilon(w) \lesssim 2\,\varepsilon^2\,\varphi(\varepsilon/2) \quad (\varepsilon \downarrow 0)$$

for P^0-a.e. w, where $\varphi(\varepsilon) = -\log \mathbb{P}^0(\|W\|_{[0,1]} \le \varepsilon)$. Recall that $\varphi(\varepsilon) \sim \frac{\lambda_1}{\varepsilon^2}$, where $\lambda_1 > 0$ is the principle eigenvalue of the Dirichlet problem on the domain $\mathrm{int}(B_S(0,1))$. Hence, κ is in $\mathbb{R}_+$ and Corollary 7.1.11 yields that

$$\kappa \in [2\lambda_1, 8\lambda_1].$$

□

7.4 Monte Carlo approximation of κ

We maintain the notations of the previous section, but confine ourselves to 1-dimensional Brownian motion, i.e. $d = 1$. The norm $\|\cdot\|_I$ $(I \subset \mathbb{R})$ shall denote the standard supremum norm. As we have observed in the previous section there exists $\kappa > 0$ such that for P^0-a.e. $w \in C[0,\infty)$

$$\lim_{t\to\infty} -\frac{1}{t}\log \mathbb{P}^0(\|W - w\|_{[0,t]} \le 1) = \kappa.$$

The principal eigenvalue of the Dirichlet problem on the domain $(-1,1)$ is $\lambda_1 = \frac{\pi^2}{8}$. Hence, $\kappa \in [\frac{\pi^2}{4}, \pi^2]$.

In this section, we estimate the constant κ by using Monte Carlo methods. Note that it is not feasible to construct a perfect P^0-distributed sample w and to afterwards compute

$$-\frac{1}{t}\log \mathbb{P}^0(\|W - w\|_{[0,t]} \le 1).$$

Therefore, we approximate the Wiener process by discrete processes.

Let $U = \{U_i\}_{i\in\mathbb{N}}$ denote a sequence of independent Bernoulli random variables on $\{-1,1\}$ and set

$$S_n = S_n(U) = \sum_{i=1}^n U_i, \quad n \in \mathbb{N}_0.$$

For a fixed parameter $\delta > 0$, we consider the process

$$W_t^{(\delta)} = W_t^{(\delta)}(U) = \delta\, S_{\lfloor t/\delta^2 \rfloor}, \quad t \ge 0.$$

Table 7.1: Results of a Monte Carlo study ($\delta = 1/400$, $T = 10$, $K = 30$)

k	0	1	2	3	4	5
$\hat{\kappa}(k)$	2.7224	2.4381	2.7445	2.7971	2.3199	2.4937
k	6	7	8	9	10	11
$\hat{\kappa}(k)$	2.2441	2.8693	2.7322	2.6318	2.7747	2.3066
k	12	13	14	15	16	17
$\hat{\kappa}(k)$	2.9224	3.0409	2.8994	2.5822	2.6797	2.4320
k	18	19	20	21	22	23
$\hat{\kappa}(k)$	3.2327	2.6069	2.6644	3.0838	2.6450	2.4910
k	24	25	26	27	28	29
$\hat{\kappa}(k)$	2.8788	2.3264	2.9022	2.4058	2.7635	2.3513

We denote by $P^{(\delta)}$ the law of the process $W^{(\delta)} = \{W_t^{(\delta)}\}_{t\geq 0}$ on the Skorohod space, i.e. the space of right continuous functions with left hand limits equipped with the Skorohod topology. Due to Donsker's invariance principle, the family $W^{(\delta)}$, $\delta > 0$, of processes converges weakly to the Wiener process. Motivated by this fact, we investigate the probability

$$-\frac{1}{T} \log \mathbb{P}(\|W^{(\delta)} - w^{(\delta)}\|_{[0,T]} \leq 1)$$

for a $P^{(\delta)}$-typical $w^{(\delta)}$, small $\delta > 0$ and large $T > 0$. Although this approximation is mathematically not rigorous, it gives first insights on the parameter κ.

We let $u = \{u_i\}_{i\in\mathbb{N}}$ be an arbitrary $\{-1,1\}$-valued sequence which will later be thought of as a realization of independent Bernoulli random variables. Analogously to S and $W^{(\delta)}$, we define $s = s(u) = \{s_n(u)\}_{n\in\mathbb{N}}$ and $w^{(\delta)} = w^{(\delta)}(u) = \{w_t^{(\delta)}(u)\}_{t\geq 0}$. Since

$$\begin{aligned}\|W^{(\delta)} - w^{(\delta)}\|_{[0,T]} &= \delta\, \|S_{\lfloor t/\delta^2\rfloor} - s_{\lfloor t/\delta^2\rfloor}\|_{[0,T]} \\ &= \delta \sup_{n=1,\dots,\lfloor T/\delta^2\rfloor} |S_n - s_n|,\end{aligned}$$

we have

$$-\frac{1}{T} \log \mathbb{P}(\|W^{(\delta)} - w^{(\delta)}\|_{[0,T]} \leq 1) = -\frac{1}{T} \log \mathbb{P}\Big(\sup_{n=1,\dots,\lfloor T/\delta^2\rfloor} |S_n - s_n| \leq \frac{1}{\delta}\Big). \tag{7.16}$$

The latter expression may be computed explicitly for any fixed sequence $\{u_n\}_{n\in\mathbb{N}}$ in $\{-1,1\}$. Now one can compute (7.16) for a typical computer generated sequence $\{u_n\}_{n\in\mathbb{N}}$. In order to obtain results on the dispersion of (7.16) we study alternatively

$$\hat{\kappa}(k) := -\frac{1}{T}\log\mathbb{P}\big(\|W^{(\delta)} - w^{(\delta)}\|_{(kT,(k+1)T]} \le 1\big|\|W^{(\delta)} - w^{(\delta)}\|_{(0,kT]} \le 1\big)$$

for $k = 0, \dots, K-1$ where $T > 0$ and $K \in \mathbb{N}$ are fixed parameters.

In Table 7.1, one finds results of a simulation run on Matlab with parameters $\delta = 1/400$, $T = 10$ and $K = 30$. In this sample the average of the values $\hat{\kappa}$ is 2.6661 and one obtains a confidence interval $[2.572, 2.76]$ (95% confidence level) based on a t-test.

Comparing the estimated values with the known bounds $\kappa \ge \frac{\pi^2}{4} = 2.4674...$ and $\kappa \le \pi^2 = 9.8696...$ gives rise to the assumption that the rigorous lower bound is better than the corresponding upper bound.

7.5 Linking random SBPs and quantization

Let μ be a centered Gaussian measure on the separable Banach space $(E, \|\cdot\|)$ and denote by X a μ-distributed r.e. In this section, we study the asymptotics of $D^{(R)}(r|\mu, \|\cdot\|^s)^{1/s}$ and its relations to random small ball probabilities around random centers. Recall that

$$D^{(R)}(r|\mu, \|\cdot\|^s)^{1/s} = \mathbb{E}\big[\min_{i=1,\dots,\lfloor e^r\rfloor} \|X - \tilde{X}_i\|^s\big]^{1/s},$$

where $\{\tilde{X}_i\}_{i\in\mathbb{N}}$ is a sequence of independent (as well of X) μ-distributed r.e.'s in E.

The following calculations are based on

Assumption 7.5.1. There exists a function $\varphi_R : \mathbb{R}_+ \to \mathbb{R}_+$ that is convex, one-to-one and onto, and satisfies

$$\varphi_R(\varepsilon) \sim -\log\mu(B(X,\varepsilon)) \quad \text{as } \varepsilon \downarrow 0, \text{ in probability.} \tag{7.17}$$

Recalling Remark 7.1.8 and Lemma 7.1.9, it is typically not hard to find φ_R such that the regularity conditions beyond (7.17) are fulfilled if there exists a function φ_R satisfying (7.17). In fact, $\tilde{\varphi}_R(\varepsilon) = \mathbb{E}[-\log\mu(B(X,\varepsilon))]$, $\varepsilon > 0$, is an appropriate choice for φ_R in many cases. Moreover, recall that the small ball probabilities are a.s. equivalent to a convex function in the Hilbert space setting.

Theorem 7.5.2. *Suppose that φ_R fulfills Assumption 7.5.1 and assume that*

$$\varphi_R^{-1}(r) \approx \varphi_R^{-1}(2r) \quad (r \to \infty). \tag{7.18}$$

Then

$$D^{(R)}(r|\mu, \|\cdot\|^s)^{1/s} \sim \varphi_R^{-1}(r) \quad (r \to \infty).$$

We need some lemmas.

Lemma 7.5.3. *Under the assumptions of Theorem 7.5.2 one has*

$$\varphi^{-1}(r) \approx \varphi_R^{-1}(r) \quad (r \to \infty).$$

Proof. By assumption (7.17) and Theorem 7.1.1, there exists $\varepsilon_0 > 0$ such that for $\varepsilon \in (0, \varepsilon_0]$,

$$\frac{1}{2}\varphi(\varepsilon) \leq \varphi_R(\varepsilon) \leq 4\varphi(\varepsilon/2).$$

Hence, for $r \geq 4\varphi(\varepsilon_0/2)$,

$$\varphi^{-1}(2r) \leq \varphi_R^{-1}(r) \leq \frac{1}{2}\,\varphi^{-1}(r/4).$$

and assumption (7.18) yields the proof of the lemma. □

Remark 7.5.4. Analogously to the previous proof, one can show that

$$\varphi^{-1}(r) \approx \varphi^{-1}(2r) \quad (r \to \infty) \tag{7.19}$$

implies

$$\varphi_R^{-1}(r) \approx \varphi_R^{-1}(2r) \quad (r \to \infty).$$

Consequently, we can replace assumption (7.18) by assumption (7.19) in Theorem 7.5.2.

Proposition 7.5.5. *Denote*

$$\eta = \limsup_{r\to\infty} \frac{\varphi_R^{-1}(r)}{\varphi_R^{-1}(2r)} < \infty.$$

Fix $\kappa \in (0,1)$ *and consider for* $\delta := \frac{1}{4\eta}\,\kappa$ *and* $r \geq 0$ *the sets*

$$\mathcal{T}_1(r) = \left\{x \in E : -\log \mu(B(x, (1+\kappa)\varphi_R^{-1}(r))) \leq (1-\delta)r\right\}$$

and

$$\mathcal{T}_2(r) = \left\{x \in E : -\log \mu(B(x, (1-\kappa)\varphi_R^{-1}(r))) \geq (1+\delta)r\right\}.$$

Under the assumptions of Theorem 7.5.2, one has

$$\lim_{r\to\infty} \mu(\mathcal{T}_1(r)) = \lim_{r\to\infty} \mu(\mathcal{T}_2(r)) = 1.$$

Proof. Fix $\kappa \in (0,1)$ arbitrarily and let $\delta = \frac{1}{4\eta}\kappa < \frac{1}{4}$ and $r > 0$. Note that by the convexity of φ_R^{-1},

$$\varphi_R^{-1}(r - 2\delta r) - \varphi_R^{-1}(r) \le \frac{2\delta r}{r/2}\big(\varphi_R^{-1}(r/2) - \varphi_R^{-1}(r)\big) \lesssim 4\delta(\eta - 1)\varphi_R^{-1}(r)$$

as $r \to \infty$. Therefore, there exists $r_0 \ge 0$ such that

$$\varphi_R^{-1}(r - 2\delta r) \le (1 + \kappa)\varphi_R^{-1}(r)$$

for all $r \ge r_0$. Consequently, the set $\mathcal{T}_1(r)$ satisfies for $r \ge r_0$

$$\mathcal{T}_1(r) \supset \Big\{x \in E : -\log \mu(B(x, \varphi_R^{-1}((1 - 2\delta)r))) \le \frac{1-\delta}{1-2\delta}(1 - 2\delta)r\Big\}.$$

Since $\varphi_R^{-1}((1 - 2\delta)r)$ converges to 0 and $(1 - \delta)/(1 - 2\delta) > 1$, it holds by assumption (7.17) that

$$\lim_{r\to\infty} \mu(\mathcal{T}_1(r)) = 1.$$

The converse inequality is proved similarly: One has for $r \ge r_0$

$$\varphi_R^{-1}(r) - \varphi_R^{-1}(r + 2\delta r) \le \varphi_R^{-1}(r - 2\delta r) - \varphi_R^{-1}(r) \le \kappa\varphi_R^{-1}(r),$$

where the first inequality is a consequence of the convexity of φ_R^{-1}. Hence, $\varphi_R^{-1}(r + 2\delta r) \ge (1 - \kappa)\varphi_R^{-1}(r)$ for $r \ge r_0$ and it follows

$$\mathcal{T}_2(r) \supset \Big\{x \in E : -\log \mu(B(x, \varphi_R^{-1}(r + 2\delta r))) \ge \frac{1+\delta}{1+2\delta}(1 + 2\delta)r\Big\}.$$

Analogously to above, assumption (7.17) is used to infer that

$$\lim_{r\to\infty} \mu(\mathcal{T}_2(r)) = 1.$$

□

Proposition 7.5.6. *Let $\kappa \in (0,1)$. For $r \ge 0$ consider*

$$Z(r) := \min_{i=1,\dots,\lfloor e^r \rfloor} \|X - \tilde{X}_i\|$$

and the event

$$\mathcal{T}(r) := \big\{Z(r) \in [(1 - \kappa)\varphi_R^{-1}(r), (1 + \kappa)\varphi_R^{-1}(r)]\big\}.$$

Under the assumptions of Theorem 7.5.2 one has

$$\lim_{r\to\infty} \mathbb{P}(\mathcal{T}(r)) = 1.$$

Proof. Recall that φ_R^{-1} is convex. According to Lemma 3.1.4 it suffices to prove the asymptotic equivalence for values $r \in I := \{\log j : j \in \mathbb{N}\}$, i.e. values r for which e^r is an integer. By Proposition 7.5.5, one has $\lim_{r\to\infty} \mu\big(\mathcal{T}_1(r) \cap \mathcal{T}_2(r)\big) = 1$ with $\mathcal{T}_1(r)$ and $\mathcal{T}_2(r)$ as in the proposition. Moreover, for $r \in I$ and $X \in \mathcal{T}_1(r)$, one has

$$\begin{aligned}\mathbb{P}(Z(r) > (1+\kappa)\varphi_R^{-1}(r)|X) &= \big(1-\mu(B(X,(1+\kappa)r))\big)^{e^r}\\ &\le \big(1-e^{-r+\delta r}\big)^{e^r} = \Big(1-\frac{e^{\delta r}}{e^r}\Big)^{e^r}\\ &\le \exp\{-e^{\delta r}\} \to 0 \quad (r\to\infty).\end{aligned}$$

On the other hand, for $X \in \mathcal{T}_2(r)$, $r \in I$, it holds

$$\begin{aligned}\mathbb{P}(Z(r) > (1-\kappa)\varphi_R^{-1}(r)|X) &= \big(1-\mu(B(X,(1-\kappa)r))\big)^{e^r}\\ &\ge \big(1-e^{-r-\delta r}\big)^{e^r} = \Big(1-\frac{e^{-\delta r}}{e^r}\Big)^{e^r} \to 1\end{aligned}$$

as $r \to \infty$. Hence, the events $\mathcal{T}(r)$, $r \ge 0$, satisfy $\lim_{r\to\infty} \mathbb{P}(\mathcal{T}(r)) = 1$. □

Proof of Theorem 7.5.2. Fix $s > 0$. First we prove

$$D^{(R)}(r|\mu, \|\cdot\|^s)^{1/s} \lesssim \varphi_R^{-1}(r) \quad (r\to\infty).$$

Fix $\kappa \in (0,1)$ and let $\mathcal{T}(r)$ and $Z(r)$ as in the previous proposition. Now

$$\mathbb{E}[Z(r)^s] \le \mathbb{E}[1_{\mathcal{T}(r)}\,(1+\kappa)^s\varphi_R^{-1}(r)^s] + \mathbb{E}[1_{\mathcal{T}(r)^c}\,Z(r)^s] =: I_1(r) + I_2(r).$$

One has $I_1(r) \le (1+\kappa)^s\varphi_R^{-1}(r)^s$. Moreover, the Cauchy-Schwarz inequality gives

$$I_2(r) \le \mathbb{P}(\mathcal{T}(r)^c)^{1/2}\,\mathbb{E}[Z(r)^{2s}]^{1/2}.$$

As a consequence of Lemma 7.5.3 and assumption (7.18), φ satisfies

$$\varphi^{-1}(2r) \approx \varphi^{-1}(r) \quad (r\to\infty).$$

Thus, Theorem 3.1.2 is applicable and one obtains that $\mathbb{E}[Z(r)^{2s}]^{1/2}$ is of order $\mathcal{O}(\varphi^{-1}(r)^s)$. By the previous proposition, $\lim_{r\to\infty}\mathbb{P}(\mathcal{T}(r)^c) = 0$. Consequently,

$$I_2(r) = o(\varphi_R^{-1}(r)^s) \quad (r\to\infty).$$

and

$$\mathbb{E}[Z(r)^s]^{1/s} \lesssim (1+\kappa)\,\varphi_R^{-1}(r) \quad (r\to\infty).$$

Since $\kappa \in (0,1)$ was chosen arbitrarily, it follows that

$$\mathbb{E}[Z(r)^s]^{1/s} \lesssim \varphi_R^{-1}(r) \quad (r\to\infty).$$

The converse inequality follows since for fixed $\kappa \in (0,1)$ and $\mathcal{T}(r)$ as above one has

$$\begin{aligned}\mathbb{E}[Z(r)^s]^{1/s} &\geq \mathbb{E}[1_{\mathcal{T}(r)}\, Z(r)^s]^{1/s} \geq \mathbb{P}(\mathcal{T}(r))^{1/s}(1-\kappa)\varphi_R^{-1}(r)\\ &\gtrsim (1-\kappa)\varphi_R^{-1}(r) \quad (r \to \infty).\end{aligned}$$

□

Now we consider the standard Hilbert space setting, i.e. μ is a Gaussian measure on a separable Hilbert space H with infinite dimensional support. Again we denote by $\{\lambda_n\}_{n\in\mathbb{N}}$ the ordered sequence of eigenvalues.

Recall that the asymptotics of φ_R are given in Corollary 7.2.9 under a polynomial decay assumption on the eigenvalues. Combining this result with Theorem 7.5.2 yields immediately

Corollary 7.5.7. *Suppose*

$$\lambda_j \sim \kappa\, j^{-\alpha} \quad (j \to \infty)$$

for some $\kappa > 0$ and $\alpha > 1$. Then for $s > 0$

$$D^{(R)}(r|\mu, \|\cdot\|^s)^{1/s} \sim \Big(\frac{\alpha-1}{2}\Big)^{\frac{\alpha-1}{2}} \Big(\frac{(\alpha+1)\pi}{\alpha^2 \sin\frac{\pi}{\alpha}}\Big)^{\alpha/2} \sqrt{\kappa}\, r^{-(\alpha-1)/2}, \qquad (7.20)$$

as $r \to \infty$.

We compare the coding quantities $D^{(R)}$ and $D^{(q)}$. Recall that under the assumptions of the previous corollary, one has

$$D^{(q)}(r|\mu, \|\cdot\|^s)^{1/s} \sim \frac{\alpha^{\alpha/2}}{2^{(\alpha-1)/2}\sqrt{\alpha-1}} \sqrt{\kappa}\, r^{-(\alpha-1)/2} \quad (r \to \infty) \qquad (7.21)$$

for any $s > 0$ due to Corollary 6.4.2. Hence, the limit

$$\lim_{r\to\infty} \frac{D^{(R)}(r|\mu, \|\cdot\|^s)^{1/s}}{D^{(q)}(r|\mu, \|\cdot\|^s)^{1/s}} = \Big[\frac{(\alpha^2-1)\pi}{\alpha^3 \sin(\frac{\pi}{\alpha})}\Big]^{\alpha/2} =: q(\alpha)$$

exists and does not depend on $s > 0$. Figure 7.1 contains a plot of the ratio $q(\alpha)$ against the parameter α.

By basic analysis, one finds the following

Lemma 7.5.8. *One has*

$$\lim_{\alpha\downarrow 1} q(\alpha) = \sqrt{2} \quad \text{and} \quad \lim_{\alpha\to\infty} q(\alpha) = 1.$$

The lemma indicates that $D^{(R)}$ is closer to $D^{(q)}$ the faster the eigenvalues decay. This gives rise to the conjecture that for certain fast decaying eigenvalues (faster than all polynomials) both quantities have the same asymptotics.

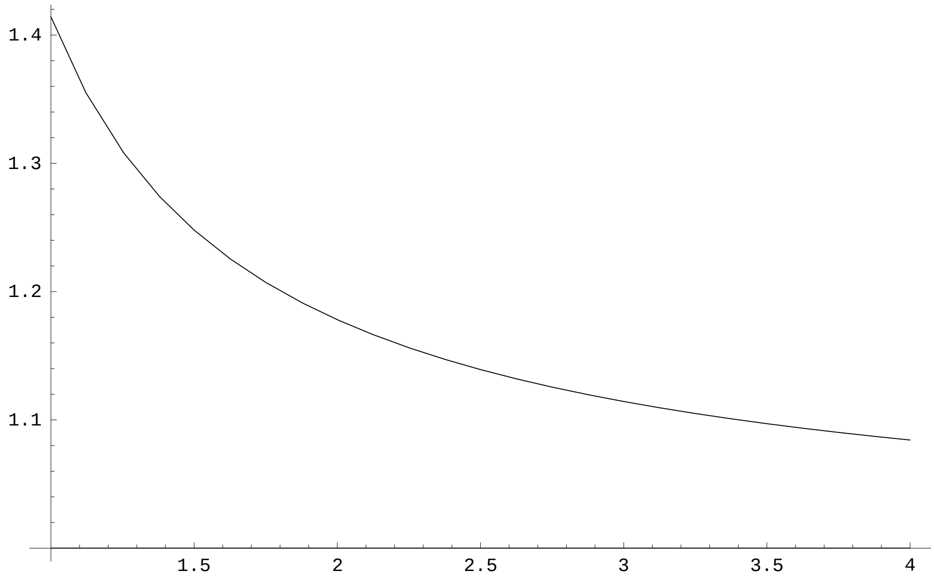

Figure 7.1: The plot shows the graph of the ratio q

Chapter 8

Final remarks and open problems

In this paragraph we recall some results of this dissertation and present some interesting open problems.

8.1 Coding Gaussian measures on Hilbert spaces

The main results of Chapter 6 are Theorem 6.2.1 and Theorem 6.3.1. There the strong asymptotics of $D^{(r)}$ and $D^{(q)}$ for arbitrary moments $s > 0$ have been found under the assumption that the sequence of eigenvalues $\{\lambda_n\}_{n\in\mathbb{N}}$ satisfies

$$\lim_{n\to\infty} \frac{\log\log(1/\lambda_n)}{n} = 0. \tag{8.1}$$

Moreover, the asymptotics of D and $D^{(e)}$ are given for moments $s \geq 2$.

Question 8.1.1. 1.) Can one complement the upper bound for $D(r|\mu, \|\cdot\|^s)^{1/s}$ and $D^{(e)}(r|\mu, \|\cdot\|^s)^{1/s}$ by an appropriate lower bound when $s \in (0, 2)$?
2.) Does a result like Theorem 6.2.1 or Theorem 6.3.1 hold under weaker assumptions on the eigenvalues?

8.2 Coding Gaussian measures on Banach spaces

In Chapter 3, the upper bound derived by Fehringer is complemented by an appropriate lower bound for the DRF (Theorem 3.5.1). Combining these bounds gives the weak asymptotics of the coding quantities $D^{(q)}$, $D^{(e)}$ and

D under the assumption that the small ball function satisfies

$$\varphi^{-1}(r) \approx \varphi^{-1}(2r) \quad (r \to \infty). \tag{8.2}$$

Moreover, all quantities and all moments are weakly equivalent. As one easily verifies in the Hilbert space setting, neither the lower nor the upper bound are optimal in general in the strong sense. If condition (8.2) is not fulfilled, the weak asymptotics of either of the expressions D, $D^{(e)}$ an $D^{(q)}$ is not known explicitly, unless we are in the Hilbert space setting.

Question 8.2.1. 1.) Can one improve the known bounds in the general setting or in particular cases?
2.) What are the weak asymptotics in the high resolution coding problem when (8.2) is not satisfied?

The second main result of this chapter (Theorem 3.2.3) shows that ε-nets of certain compact sets constitute weakly optimal codebooks under assumption (8.2).

Question 8.2.2. Is it possible to find appropriate compact subsets A_r, $r \geq 0$, of E such that ε-nets of A_r constitute asymptotically optimal codebooks of rate r?

8.3 Perturbation of the high resolution coding problem

In Chapter 4 we have treated the effect of perturbations in the rate and distribution parameter for the coding quantities D, $D^{(e)}$ and $D^{(q)}$. It is found that the quantities D and $D^{(e)}$ are quite robust against small perturbations. The perturbation result for the quantization error is weaker than the corresponding results for D and $D^{(e)}$. In fact, it is typically too weak to imply further results. For instance, it is not possible to conclude that the quantity $\delta^{(q)}(\,\cdot\,)$ is slowly varying which is an essential assumption of Theorem 5.2.1.

Question 8.3.1. Do there exist stronger perturbation results for the high resolution quantization problem?

The perturbation results for D and $D^{(e)}$ allowed us to relate the coding problem of diffusion processes with non-constant diffusion coefficient to the coding complexity of Brownian motion. We saw that "nice" drifts have no influence on the asymptotic coding problem. For general diffusions with random diffusion coefficient, the strong asymptotics in the quantization problem are still unknown. The results of Section 6.5 suggest that for this original there is no equivalence of moments and no equivalence of entropy coding and quantization in the high resolution coding problem.

Question 8.3.2. What are the asymptotics of the high resolution coding problem for general diffusion processes?

8.4 Random small ball probabilities

In Theorem 7.2.2, we showed that in the Hilbert space setup there exists a deterministic continuous function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$-\log \mu(B(x,\varepsilon)) \sim \psi(\varepsilon) \quad (\varepsilon \downarrow 0)$$

for μ-a.e. x. In the general setting one still knows that (by Lemma 7.1.3) for μ-a.e. x

$$\liminf_{\varepsilon\downarrow 0} \frac{-\log \mu(B(x,\varepsilon))}{\psi(\varepsilon)} = c_\psi \quad \text{and}$$
$$\limsup_{\varepsilon\downarrow 0} \frac{-\log \mu(B(x,\varepsilon))}{\psi(\varepsilon)} = C_\psi$$

for any continuous gauge function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ with appropriate constants $c_\psi, C_\psi \in [0,\infty]$.

Question 8.4.1. For which Gaussian measures does there exist a gauge function ψ such that the corresponding constants c_ψ and C_ψ satisfy $c_\psi = C_\psi = 1$?

Under the assumption that

$$\varphi(\varepsilon) \approx \varphi(2\varepsilon) \quad (\varepsilon \downarrow 0)$$

one knows that

$$\tilde{\varphi}_R(\varepsilon) = \mathbb{E}[-\log \mu(B(X,\varepsilon))], \quad \varepsilon > 0,$$

is an appropriate choice for ψ whenever there exists an appropriate gauge function (see Remark 7.1.8).

In Theorem 7.5.2 we established an equivalence between the random small ball function and the coding quantity $D^{(R)}$. An application of this result to the standard Hilbert space setup showed that for fast polynomially decaying eigenvalues the ratio between $D^{(R)}$ and $D^{(q)}$ is close to one (see Lemma 7.5.8).

Question 8.4.2. What are the asymptotics of $D^{(R)}$ for fast decaying eigenvalues? Does one have equivalence of $D^{(R)}$ and $D^{(q)}$ under appropriate assumptions on the eigenvalues?

Appendix A

Regular variation

Definition A.1. A function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is said to be *slowly varying* if it is Borel measurable and satisfies

$$\lim_{x\to\infty} \frac{f(\lambda x)}{f(x)} = 1$$

for all $\lambda \in \mathbb{R}_+$.

Definition A.2. A function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is called *regularly varying* with index $\alpha \in \mathbb{R}$ if

$$f(x) = x^\alpha \, l(x), \quad x \in \mathbb{R}_+,$$

for some slowly varying function l. For short we write $f \in \mathcal{R}_\alpha$. Analogously, $g : \mathbb{R}_+ \to \mathbb{R}_+$ is said to be regularly varying at 0 with index $\alpha \in \mathbb{R}$ if

$$g(x) = x^\alpha \, \tilde{l}(1/x), \quad x \in \mathbb{R}_+,$$

for some slowly varying function $\tilde{l}$. Then we write $g \in \mathcal{R}_\alpha(0)$.

Regularly varying functions have some nice properties.

Theorem A.3. *Let $l : \mathbb{R}_+ \to \mathbb{R}_+$ be slowly varying. Then, there exists $l_1 \in C^\infty(0,\infty)$ such that $l(x) \sim l_1(x)$ as $x \to \infty$ and*

$$\lim_{x\to\infty} \frac{x l_1'(x)}{l_1(x)} = 0,$$

where l_1' denotes the derivative of l_1.

Proof. The theorem is a direct consequence of a theorem of de Bruijn [10] (see also Theorem 1.3.3 in [6]). It states that there exists $l_1 \in C^\infty(0,\infty)$ satisfying $l_1(x) \sim l(x)$ with

$$\lim_{x\to\infty} \frac{d}{dx} \log(l_1(e^x)) = 0.$$

Evaluating the latter derivative yields

$$\lim_{x\to\infty} \frac{l_1'(e^x)\, e^x}{l_1(e^x)} = 0$$

which is equivalent to the statement of the Theorem. □

Theorem A.4. *Let $f \in \mathcal{R}_\alpha$ with $\alpha > 0$. Then there exists $g \in \mathcal{R}_{1/\alpha}$ with*

$$f(g(x)) \sim g(f(x)) \sim x$$

as $x \to \infty$. g is called asymptotic inverse of f. It is unique up to asymptotic equivalence. Moreover, two asymptotically equivalent regularly varying function f and $\tilde{f}$ in $\mathcal{R}_\alpha$ have the same functions as asymptotic inverse.

The theorem is taken from [6] (Theorem 1.5.12).

Remark A.5. The previous theorem considers inversion of regularly varying functions with index bigger than zero. Let now $-\alpha < 0$ and $f(x) = x^{-\alpha} l(x)$ for some slowly varying function l. Denote

$$i : \mathbb{R}_+ \to \mathbb{R}_+,\ x \mapsto 1/x.$$

For arbitrary $\beta \in \mathbb{R}$, left hand application of i gives a bijection between $\mathcal{R}_\beta$ and $\mathcal{R}_{-\beta}$ and right hand application of i an bijection between $\mathcal{R}_\beta$ and $\mathcal{R}_{-\beta}(0)$. Moreover both operations maintain asymptotic equivalence. The previous theorem is applicable to the function

$$\tilde{f}(x) = i \circ f(x) = 1/f(x) = x^\alpha / l(x)$$

which is in $\mathcal{R}_\alpha$. Hence, there exists a unique (up to asymptotic equivalence) asymptotic inverse $\tilde{g} \in \mathcal{R}_{1/\alpha}$ of $\tilde{f}$. Denote $g(x) = \tilde{g} \circ i(x)$. Then, $g \in \mathcal{R}_{-1/\alpha}(0)$ and one has

$$g \circ f(x) = \tilde{g} \circ i \circ i \circ \tilde{f}(x) \sim x \tag{A.1}$$

and

$$f \circ g(1/x) = i \circ \tilde{f} \circ \tilde{g}(x) \sim 1/x \tag{A.2}$$

as $x \to \infty$. Consequently, in this case a function $f \in \mathcal{R}_{-\alpha}$ corresponds to an asymptotic inverse $g \in \mathcal{R}_{-1/\alpha}(0)$ and vice versa.

Lemma A.6. *Let $f \in \mathcal{R}_{-\alpha}$ monotonically decreasing for some $\alpha > 0$. Then*

$$f^{\leftarrow}(y) = \sup\{x \in \mathbb{R}_+ : f(x) \ge y\}, \quad y \in \mathbb{R}_+,$$

is an asymptotic inverse of f, i.e. $g = f^{\leftarrow}$ satisfies (A.1) and (A.2). Here, the supremum of the empty set is assumed to be 1 in order to enforce that $f^{\leftarrow}$ maps into $\mathbb{R}_+$.

Lemma A.7. *Let $\alpha, \kappa > 0$ and $\beta \in \mathbb{R}$. Assume that $f \in \mathcal{R}_{-\alpha}(0)$, $\alpha > 0$, satisfies*

$$f(\varepsilon) \sim \kappa \varepsilon^{-\alpha} \left(\log \frac{1}{\varepsilon}\right)^{\beta} \quad (\varepsilon \downarrow 0).$$

Then its asymptotic inverse $f^{\leftarrow} \in \mathcal{R}_{-1/\alpha}$ satisfies,

$$f^{\leftarrow}(t) \sim \left(\frac{\kappa}{\alpha^{\beta}} \frac{(\log t)^{\beta}}{t}\right)^{1/\alpha} \quad (t \to \infty).$$

The previous result follows by basic analysis (see, for instance, Dereich et al. [19], Lemma 4.2).

Notation index

$\mathbb{N}$	natural numbers without 0
$\mathbb{N}_0$	$\mathbb{N} \cup \{0\}$
$\mathbb{R}$	real numbers
$\mathbb{R}_+$	$(0, \infty)$

Real numbers

$\lfloor x \rfloor$	the largest integer $n \leq x$
$\lceil x \rceil$	the smallest integer $n \geq x$
$x \wedge y$	the minimum of x and y
$x \vee y$	the maximum of x and y
x^+	the positive part of x, $x \vee 0$
x^-	the negative part of x, $(-x) \vee 0$
$\sim$	strong asymptotic equivalence, $f(x) \sim g(x)$ $(x \to \infty)$ means $\lim_{x\to\infty} f(x)/g(x) = 1$
$\lesssim$	strong asymptotic domination, $f(x) \lesssim g(x)$ $(x \to \infty)$ means $\limsup_{x\to\infty} f(x)/g(x) \leq 1$
$\gtrsim$	$f(x) \gtrsim g(x)$ $(x \to \infty)$ means $g(x) \lesssim f(x)$ $(x \to \infty)$
$\approx$	weak asymptotic equivalence, $f(x) \approx g(x)$ $(x \to \infty)$ means $\limsup_{x\to\infty}[f(x)/g(x) + g(x)/f(x)] < \infty$

Vectors and sequences

$\lvert x\rvert$	Euclidean norm of x
$\lvert x\rvert_p$	the l_p-norm for vectors, $\big(\lvert x_1\rvert^p + \cdots + \lvert x_d\rvert^p\big)^{1/p}$
$\lVert x\rVert_{l_p}$	the l_p-norm for sequences, $\big(\sum_{j\in\mathbb{N}} \lvert x_j\rvert^p\big)^{1/p}$
l_p	the set of real-valued sequences with finite l_p-norm

Sets in Polish spaces (E, d)

$\bar{A}$	closure of A in E
$B(x,r)$, $B_E(x,r)$	the closed ball of radius r with center x (in E)
$d(x,A)$	the distance between the point x and the set A, $d(x,A) = \inf_{y\in A} d(x,y)$

Sets

$\emptyset$	empty set
$A+B$	the Minkowski sum, $A+B = \{x+y : x\in A, y\in B\}$
$x+A$	$\{x+y : y\in B\}$
A^c	the complement of A
$\lvert A\rvert$	the cardinality of A

Function spaces

$C(E)$, $C(E,F)$	the set of continuous functions mapping from E to $\mathbb{R}$ or F
$\lVert f\rVert_I$, $\lVert f\rVert_{I,G}$	the supremum of $\lvert f\rvert$ ($\lvert f\rvert_G$) over the index set I
$\lVert f\rVert_{L_p(I)}$, $\lVert f\rVert_{L_p(I,G)}$	the L_p-norm of function $\lvert f\rvert$ ($\lvert f\rvert_G$) w.r.t. Lebesgue measure, $\big(\int_I \lvert f\rvert^p \, d\lambda^d\big)^{1/p}$
$L_p(I)$	the set of functions with finite L_p-norm
$\mathcal{R}_\alpha$, $\mathcal{R}_\alpha(0+)$	the set of α-regularly varying functions at ∞ and 0

Special functions

Φ	the distribution function of the standard normal distribution
Ψ	$\Psi(t) = 1 - \Phi(t)$
$\varphi(\varepsilon)$	the logarithmic small ball function of μ, $-\log \mu(B(0, \varepsilon))$

Probability and measures

$\mathcal{B}(E)$	the Borel σ-algebra on E
$\mathcal{M}_1(E)$	the set of probability measures on $(E, \mathcal{B}(E))$
$\mathcal{M}(E)$	the set of finite positive measures on $(E, \mathcal{B}(E))$
$\mathcal{L}(X)$	the law of X
$\mathbb{E}[X]$	the expected value of X
$\\|X\\|_{L_p(\mathbb{P})}$, $\\|X\\|_{L_p(\mathbb{P},E)}$	the p-th moment of X (in E), $\mathbb{E}[\\|X\\|^p]^{1/p}$
$L_p(\mathbb{P})$, $L_p(\mathbb{P}, E)$	the set of real-valued (E-valued) random elements with finite p-th moment
$\mathcal{N}(x, t)$	the normal distribution with mean x and variance t
$\mathcal{U}(A)$	the uniform distribution on the set A
λ^d	d-dimensional Lebesgue measure

Information theory

$\mathbb{H}(X)$	the entropy of X in nats
$\mathbb{H}(\mu\\|\nu)$	the relative entropy of μ w.r.t. ν
$I(X, \hat{X})$	the mutual information between X and $\hat{X}$

Coding quantities for measures μ and distortions ρ

$D(r\|\mu,\rho)$	distortion rate function	p. 15
$D^{(e)}(r\|\mu,\rho)$	entropy coding error	p. 13
$D^{(E)}(r\|\mu,\rho)$	entropy constrained quantization	p. 13
$D^{(q)}(r\|\mu,\rho)$, $\delta^{(q)}(N\|\mu,\rho)$	quantization error	p. 10
$D^{(r)}(r\|\mu,\rho)$, $\delta^{(r)}(N\|\mu,\rho)$	quantization error with random codebooks	p. 65
$D^{(R)}(r\|\mu,\rho)$	quantization error with random codebooks generated by μ	p. 27
$R(d\|\mu,\rho)$	rate distortion function	p. 20

Abbreviations

a.s.	almost surely
i.e.	id est
i.i.d.	independent identically distributed
r.e.	random element
r.v.	random variable
w.r.t.	with respect to
AEP	asymptotic equipartition property
DRF	distortion rate function
SBPs	small ball probabilities
SCT	source coding theorem

Bibliography

[1] de Acosta, A. (1983). Small deviations in the functional central limit theorem with applications to functional laws of the iterated logarithm. *Ann. Probab.* **11**, 78-101.

[2] Anderson, T.W. (1955). The integral of a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6**, 170-176.

[3] Belinsky, E. and Linde, W. (2002). Small ball probabilities of fractional Brownian sheets via fractional integration operators. *J. Theoret. Probab.* **15**, 589-612.

[4] Berger, T. (1971). *Rate distortion theory.* Prentice-Hall Inc., Englewood Cliffs.

[5] Berger, T. and Gibson, J.D. (1998). Lossy source coding. *IEEE Trans. Inform. Theory* **44**, 2693-2723.

[6] Bingham, N.H., Goldie, C.M. and Teugels J.L. (1987). *Regular variation.* Cambridge University Press.

[7] Bogachev, V.I. (1998). *Gaussian measures.* American Mathematical Society.

[8] Brascamp, H.J. and Lieb, E.H. (1975). Some inequalities for Gaussian measures and the long-range order of the one-dimensional plasma. *Funct. Integr. Appl., Proc. int. Conf.*, 1-14.

[9] Bronski, J. (2003). Small Ball Constants and Tight Eigenvalue Asymptotics for Fractional Brownian Motions. *J. Theor. Probab.* **16**, 87-100.

[10] de Bruijn, N.G. (1959). Pairs of slowly oscillating functions occurring in asymptotic problems concerning the Laplace transform. *Nieuw Arch. Wiskd., III.* **7**, 20-26.

[11] Bucklew, J.A. and Wise, G.L. (1982). Multidimensional asymptotic quantization theory with r-th power distortion measures. *IEEE Trans. Inform. Theory* **28**, 239-247.

[12] Carl, B. and Stephani, I. (1990). *Entropy, compactness and the approximation of operators.* Cambridge Tracts in Mathematics.

[13] Cover, T.M. and Thomas, J.A. (1991). *Elements of information theory.* John Wiley & Sons, Inc.

[14] Creutzig, J. (2002). *Approximation of Gaussian random vectors in Banach spaces.* Ph.D. Dissertation, Friedrich-Schiller-Universität Jena.

[15] Dembo, A. and Kontoyiannis, I. (2002). Source coding, large deviations, and approximate pattern matching. *IEEE Trans. IT* **48**, 1590-1615.

[16] Dembo, A. and Zeitouni, O. (1995). Large deviations via parameter dependent change of measure, and an application to the lower tail of Gaussian processes. *Prog. Probab.* **36**, 111-121.

[17] Dembo, A. and Zeitouni, O. (1998). Large deviations techniques and applications. 2nd Ed. Springer, New York.

[18] Dereich, S. (2003). Small ball probabilities around random centers of Gaussian measures and applications to quantization. *J. Theoret. Probab.* **16**, 427-449.

[19] Dereich, S., Fehringer, F., Matoussi, A. and Scheutzow, M. (2003). On the link between small ball probabilities and the quantization problem for Gaussian measures on Banach spaces. *J. Theoret. Probab.* **16**, 249-265.

[20] Donoho, D.L. (2000). Counting bits with Kolmogorov and Shannon. Technical report **38**. Stanford University.

[21] Dudley, R.M., Hoffmann-Jorgenson, J. and Shepp, L.A. (1979). On the lower tail of Gaussian seminorms. *Ann. Prob.* **9**, 319-342.

[22] Elias, P. (1975). Universal codeword sets and representations of the integers. *IEEE Trans. Inform. Theory* **21**, 194-203.

[23] Fehringer, F. (2001). *Kodierung von Gaußmaßen.* Ph.D. Dissertation, Technische Universität Berlin.

[24] Fejes Tóth, L. (1959). Sur la représentation d'une population infinie par un nombre fini d'éléments. *Acta Math. Sci. Hung.* **10**, 299-304.

[25] Fejes Tóth, L. (1972). *Lagerungen in der Ebene, auf der Kugel und im Raum.* Springer, Berlin.

[26] Gao, F., Hannig, J. and Torcaso, F. Preprint.

[27] Graf S. and Luschgy, H. (2000). *Foundations of quantization for probability distributions.* Lecture Notes in Mathematics 1730. Springer-Verlag.

[28] Gray, R.M. (1990). *Entropy and information theory.* Springer, New York.

[29] Gray, R.M. (1998). Quantization. *IEEE Trans. Inform. Theory* **44**, 2325-2383.

[30] Gray, R.M., Linder, T. and Li, J. (2002) A Lagrangian formulation of Zador's entropy-constrained quantization theorem. *IEEE Trans. Inf. Theory* **48**, 695-707.

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

[32] Ihara S. (1993). *Information theory for continuous systems.* World Scientific.

[33] Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes.* Springer-Verlag.

[34] Kallenberg, O. (1997). *Foundations of modern probability.* Springer.

[35] Kolmogorov, A.N. (1956). On the Shannon theory of information transmission in the case of continuous signals. *IRE Trans. Inform. Theory* **2**, 102-108.

[36] Kolmogorov, A.N. (1965). Three approaches to the definition of the notion of amount of information. *Problemy Peredachi Informatsii* **1**, 3-11; translated in [37].

[37] Kolmogorov, A.N. (1993). *Selected works of A. N. Kolmogorov. Volume III: Information theory and the theory of algorithms.* Ed. by A. N. Shiryayev. Kluwer Academic Publishers.

[38] Kolmogorov, A.N. and Tikhomirov, V.M. (1959). ε-entropy and ε-capacity of sets in functional spaces. *Uspekhi Mat. Nauk* **14**, 3-86; translated in [37].

[39] Krengel, U. (1985). *Ergodic theorems.* Walter de Gruyter.

[40] Kuelbs, J. and Li, W. (1993). Metric entropy and the small ball problem for Gaussian measures. *J. Funct. Anal.* **116**, 133-157.

[41] Kuelbs, J. and Li, W. (1993). Small ball estimates for the Brownian motion and the Brownian sheet. *J. Theoret. Prob.* **3**, 547-577.

[42] Kuelbs, J., Li, W.V. and Linde, W. (1994). The Gaussian measure of shifted balls. *Prob. Theor. and Rel. Fields* **98**, 143-162.

[43] Ledoux, M. (1996). *Isoperimetry and Gaussian analysis.* Ecole d'été de probabilités de Saint-Flour XXIV (1994). Lect. Notes Math. 1648, 165-294.

[44] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach spaces. Isoperimetry and processes.* Springer-Verlag.

[45] Li, W.V. (2003). Small Ball Probabilities for Gaussian Markov Processes under the L_p-norm, to appear in Stochastic Processes and Their Applications.

[46] Li, W.V. and Linde W. (1999). Approximation, metric entropy and small ball probabilities for Gaussian measures. *Ann. Prob.* **27**, 1556-1578.

[47] Li, W.V. and Shao, Q.M. (2001). Gaussian Processes: Inequalities, Small Ball Probabilities and Applications. *Stochastic Processes: Theory and Methods.*

[48] Lifshits, M.A. (1995). *Gaussian random functions.* Kluwer, Dordrecht.

[49] Lifshits, M.A. (1997). On the lower tail probabilities of some random series. *Ann. Probab.* **25**, 424-442.

[50] Luschgy, H. and Pagès, G. (2002). Functional quantization of Gaussian processes. *J. Func. Anal.* **196**, 486-531.

[51] Luschgy, H. and Pagès, G. Sharp asymptotics of the functional quantization problem for Gaussian processes. Preprint.

[52] Mason, D.M. and Shi, Z. (2001). Small deviations for some multi-parameter Gaussian processes. *J. Theor. Probab.* **14**, 213-239.

[53] Nyquist, H. (1924). Certain factors affecting telegraph speed. *Bell Sys. Tech. Journal* **3**, 324.

[54] Oliver, B.M., Pierce, J. and Shannon, C.E. (1948). The philosophy of PCM. *Proc. IRE* **36**, 1324-1331.

[55] Pierce, J.N. (1970). Asymptotic quantizing error for unbounded random variables. *IEEE Trans. Inf. Theory* **16**, 81-83.

[56] Pinsker, M.S. (1954). Mutual information between a pair of stationary Gaussian random processes. *Dokl. Akad. Nauk. USSR* **99**, 213-216.

[57] Pinsker, M.S. (1956). Computation of the message rate of a stationary random process and the capacity of a stationary channel. *Dokl. Akad. Nauk. USSR* **111**, 753-756.

[58] Ritter, K. (2000). *Average-case analysis of numerical problems.* Lecture Notes in Mathematics 1733. Springer-Verlag.

[59] Rockafellar, R.T. (1970). *Convex analysis.* Princeton University Press.

[60] Shannon, C.E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379-423, 623-656.

[61] Shannon, C.E. (1949). Communication in the presence of noise. *Proc. IRE* **37**, 10-21.

[62] Shannon, C.E. (1959). Coding theorems for a discrete source with a fidelity criterion. *IRE National Convention Record* **4**, 142-163.

[63] Sytaya, G.N. (1974). On some asymptotic representation of the Gaussian measure in a Hilber space. *Theory of Stochastic Processes* **2**, 94-104.

[64] Talagrand, M. (1994). The small ball problem for the Brownian sheet. *Ann. Probab.* **22**, 1331-1354.

[65] Yang, E.-h. and Kieffer, J.C. (1998). On the performance of data compression algorithms based upon string matching. *IEEE Trans. Inf. Theory* **44**, 47-65.

[66] Zador, P. (1963). *Development and evaluation of procedures for quantizing multivariate distributions.* Ph. D. Dissertation, Stanford University.

[67] Zador, P. (1966). Topics in the asymptotic quantization of continuous random variables. *Bell Laboratories Technical Memorandum.*

[68] Zador, P. (1982). Asymptotic quantization error of continuous signals and the quantization dimension. *IEEE Trans. Inform. Theory* **28**, 139-149.