Asymptotic formulae for coding problems and intermediate optimization problems: a review

by

S. Dereich

Institut für Mathematik, MA 7-5, Fakultät II
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
dereich@math.tu-berlin.de

Summary. In this article we review asymptotic formulae for the three coding problems induced by a range constraint (*quantization*), an entropy constraint (*entropy coding*) and a mutual information constraint (*distortion rate function*) on the approximation $\hat{X}$ to an original signal X. We will consider finite dimensional random variables as well as stochastic processes as original signal. A main objective of the article is to explain relationships between the original problems and certain intermediate convex optimization problems (the *point* or *rate allocation problem*). These intermediate optimization problems often build the basis for the proof of asymptotic formulae.

Keywords. High resolution coding, quantization, entropy coding, distortion rate function, asymptotic formulas, stochastic processes.

1 Introduction

In practice, discretization problems appeared for the first time in the context of Pulse-Code-Modulation (PCM). Provided that a source signal has limited bandwidth, it is possible to reconstruct it perfectly, when knowing its values on an appropriate grid of time points. However, in order to digitize the data, it still remains to *quantize* the signal at these time points.

Today the motivation for studying discretization problems is twofold. The *information theoretic* approach asks for a digital representation for an incoming random signal X. The digital representation is typically a finite number of binary digits, which is thought of as the basis for further reconstructions of the original signal. The digital representation is to be stored on a computer or to be send over a channel, and one needs to find a good trade-off between the quality of the reconstruction and the complexity of the describing representation. When adopting the information theoretic viewpoint, the signal X describes a real world object that is to be translated into an approximation $\hat{X}$ on the basis of measurements or a first very fine discretization.

A second viewpoint is adopted in the context of *simulation* of random vectors. Often a random object X is described implicitly via a probability distribution and one asks for discrete simulations $\hat{X}$ of X. In that case, one rather wants to approximate the distribution $\mathbb{P}_X$ of X (by some discrete distribution $\mathbb{P}_{\hat{X}}$) than the random element X itself. Recently, the finding of finitely supported approximations has attracted much attention since it can be used to define cubature formulas: for a real function f defined on the state space of X, one approximates the expectation $\mathbb{E}f(X)$ by the expectation $\mathbb{E}f(\hat{X})$. Provided one knows all possible outcomes of $\hat{X}$ including the corresponding probability weights, one can explicitly express the latter expectation as a finite sum. Now the aim is to find $\mathbb{P}_{\hat{X}}$ that is supported on a small set and that is close to $\mathbb{P}_X$ in a Wasserstein metric.

Although both problems share a lot of similarities, there are also important differences: when adopting the coding perspective, one is interested to have easily implementable mechanisms that map X into its digital representation, and that reconstruct $\hat{X}$ on the basis of the representation. In this case the algorithm does not need to incorporate the exact probability distribution of $\hat{X}$. On the other hand, the simulation approach needs to be able to simulate a random vector with law $\mathbb{P}_{\hat{X}}$ at least. If the approximation is to be used as a cubature formula, one furthermore needs to get hold of the probability weights of the approximation $\hat{X}$. However, it is not necessary to couple the simulated random element $\hat{X}$ to an original signal, since one only ask for an approximating distribution in that case.

The information theoretic viewpoint has been treated in an abundance of monographs and articles. The most influential contributions were Shannon's works on the Source Coding Theorem (Shannon 1948). General accounts on this topic are for instance the monographs by Cover & Thomas (1991) and by Ihara (1993). Surveys on the development of source coding theory and quantization were published by Kieffer (1993) and by Gray & Neuhoff (1998) in the engineering literature. Moreover, a mathematical account on finite dimensional quantization was provided by Graf & Luschgy (2000).

Applications of discretization schemes in finance are described, for instance, in Pagès & Printems (2004).

In this article, the discretization problem is formalized as a minimization problem of certain objective functions (representing the approximation error) under given complexity constraints on the approximation. Our main focus lies on the development of asymptotic formulae for the best-achievable approximation error when the information content of the approximation tends to infinity. Our problems will be described via non-convex optimization problem, and an explicit evaluation of minimizers is typically not feasible. However, it is often possible to relate the original problem to simpler convex optimization problems. Solving these intermediate optimization problems, often leads to asymptotic formulae in the associated approximation problems. Moreover, the solutions give a crude description of "good" approximation schemes, and they can be used for initializing numerical methods or to define close to optimal approximation schemes. In this article we try to convey intuition on how intermediate optimization problems can be derived and used to solve the asymptotic coding problems. In doing so we will survey several classical and recent results in that direction. Our complexity constraints are given through constraints on the disk space needed to save the binary representation or the number of points maximally attained by the approximation. Note that these constraints do not incorporate the computational complexity of finding a closest representation or to compute the probability weights of the approximating random variable, so that the results presented here can only serve as benchmark (which may show the optimality of certain schemes) or as a basis for the development of feasible approximation schemes.

To be more precise, we think of a problem constituted of a *source signal* (the *original*), modelled as a random variable X taking values in a measurable space E, and an error criterion given as a product measurable map $\rho : E \times E \to [0, \infty]$ (the *distortion measure*). Then the objective is to minimize the expectation

$$\mathbb{E}\rho(X, \hat{X}) \tag{1}$$

over a set of E-valued random variables $\hat{X}$ (the *reconstruction*) satisfying a *complexity constraint* (also called *information constraint*). Often we shall also consider certain moments $s > 0$ of the approximation error as objective function. In that case (1) is replaced by

$$\mathbb{E}[\rho(X, \hat{X})^s]^{1/s} = \big\|\rho(X, \hat{X})\big\|_{L^s(\mathbb{P})}. \tag{2}$$

Mainly, we shall be working with *norm-based distortion measures*, that are distortions ρ admitting the representation $\rho(x, \hat{x}) = \|x - \hat{x}\|$.

The information constraints

We will work with three different complexity constraints that all depend on a parameter $r \geq 0$ (called *rate*) and have been originally suggested by Kolmogorov (1968):

- $\log |\operatorname{range}(\hat{X})| \leq r$ (*quantization constraint*)
- $H(\hat{X}) \leq r$, where H denotes the entropy of $\hat{X}$ (*entropy constraint*)
- $I(X; \hat{X}) \leq r$, where I denotes the Shannon mutual information of X and $\hat{X}$ (*mutual information constraint*).

Here and elsewhere, we use the standard notations for entropy and mutual information:

$$H(\hat{X}) = \begin{cases} -\sum_x p_x \log p_x & \text{if } \hat{X} \text{ is discrete with probability weights } (p_x) \\ \infty & \text{otherwise} \end{cases}$$

and

$$I(X, \hat{X}) = \begin{cases} \int \log \frac{d\mathbb{P}_{X,\hat{X}}}{d\mathbb{P}_X \otimes \mathbb{P}_{\hat{X}}} \, d\mathbb{P}_{X,\hat{X}} & \text{if } \mathbb{P}_{X,\hat{X}} \ll \mathbb{P}_X \otimes \mathbb{P}_{\hat{X}} \\ \infty & \text{otherwise.} \end{cases}$$

In general, we will denote by $\mathbb{P}_Z$ the distribution function of a random variable Z indicated in the subscript. Notice that the constraints do not incorporate the computational complexity that is needed to find $\hat{X}$. So the minimal values obtained in the minimization problem can be conceived as a benchmark. As we shall explain later, the rigorous treatment of the minimization problem often leads to simpler intermediate convex optimization problems. Often one can use the intermediate optimization problem to construct feasible good coding schemes.

In order to code a countable set I, one typically uses *prefix-free codes*, these are maps $\psi : I \to \{0,1\}^*$ ($\{0,1\}^*$ denoting the strings of binary digits of finite length) such that for any $i \neq j$ in I the code $\psi(i)$ is not a prefix of $\psi(j)$. That means a prefix-free code is naturally related to a binary tree, where the leaves of the tree correspond to the elements of I and the code describes the path from the root to the leaf (0 meaning left child, 1 meaning right child). Clearly, such a code allows to decode the original message uniquely. Moreover, a concatenation of prefix-free codes leads to a new prefix-free code.

Suppose for now that the logarithms are taken to the basis 2. Then the quantization constraint implies that the set $\mathcal{C} = \operatorname{range}(\hat{X}) \subset E$ (called *codebook*) is of size $\log|\mathcal{C}| \le r$. Hence, there exists a prefix-free representation for $\hat{X}$ (of for $\mathcal{C}$) of length at most $\lceil r \rceil$. The quantization constraint is a worst-case constraint on the code length. Note that any finite set $\mathcal{C} \subset E$ with $\log|\mathcal{C}| \le r$ induces a *quantizer* $\hat{X}$ for X of rate r via

$$\hat{X} = \operatorname*{argmin}_{\hat{x} \in \mathcal{C}} \rho(X, \hat{X}).$$

Next, suppose that $\hat{X}$ satisfies the entropy constraint. Using Lempel-Ziv coding there exists a code ψ for the range of $\hat{X}$ such that

$$\mathbb{E}\,\mathrm{length}(\psi(\hat{X})) < r + 1.$$

Thus the entropy constraint is an average-case constraint on the bit length needed. The mutual information constraint, is motivated by Shannon's celebrated source coding theorem which will be stated later (see Theorem 4.1). For a general account on information theory we refer the reader to the monographs by Cover & Thomas (1991) and by Ihara (1993).

For a reconstruction $\hat{X}$ the constraints are ordered as follows

$$I(X;\hat{X}) \le H(\hat{X}) \le \log|\operatorname{range}(\hat{X})|,$$

so that the mutual information constraint is the least restrictive. Moreover, our notions of information satisfy the following *additivity property*: suppose that $\hat{X}_1$ and $\hat{X}_2$ are random vectors attaining values in some Borel space and suppose that $\hat{X}$ is given as $\varphi(\hat{X}_1, \hat{X}_2)$ where φ is some measurable function; then

- $\log|\operatorname{range}(\hat{X})| \le \log|\operatorname{range}(\hat{X}_1)| + \log|\operatorname{range}(\hat{X}_2)|$,
- $H(\hat{X}) \le H(\hat{X}_1) + H(\hat{X}_2)$ and
- $I(X;\hat{X}) \le I(X;\hat{X}_1) + H(\hat{X}_2)$.

Essentially, these estimates state that when combining the information contained in two random vectors $\hat{X}_1$ and $\hat{X}_2$ the resulting random vector has information content less than the sum of the single information contents. Note that the mutual information is special in the sense that the property $I(X;\hat{X}) \le I(X;\hat{X}_1) + I(X;\hat{X}_2)$ does not hold in general!

Now let us introduce the notation used for the minimal values in the minimization problems. When minimizing (2) under the quantization constraint for a given rate r and moment s, the minimal value will be denoted by $D^{(q)}(r,s)$:

$$D^{(q)}(r,s) = \inf\big\{\big\|\|X - \hat{X}\|\big\|_{L^s(\mathbb{P})} : \log|\operatorname{range}(\hat{X})| \le r\big\}.$$

Here the (q) in the upper index refers to the quantization constraint. Similarly, we will write $D^{(e)}(r,s)$ and $D(r,s)$ for the minimal values induced by the entropy and mutual information constraint, respectively.

Strictly speaking, the quantities $D^{(e)}(\cdot)$ and $D(\cdot)$ depend on the underlying probability space. Since we rather prefer a notion of complexity that only depends on the distribution μ of the underlying signal X, we allow extensions of the probability space when minimizing over $\hat{X}$ so that the pair $(X, \hat{X})$ can assume any probability distribution on the product space having first marginal μ.

Sometimes the source signal or the distortion measure might not be clear from the context. In that case we include this information in the notation. For instance, we write $D(r, s|\mu, \|\cdot\|)$ when considering a μ-distributed original under the objective function

$$\mathbb{E}[\|X - \hat{X}\|^s]^{1/s},$$

and we write $D(r|\mu, \rho)$ in order to refer to the objective function

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

Thereafter we write $f \sim g$ iff $\lim \frac{f}{g} = 1$, while $f \lesssim g$ stands for $\limsup \frac{f}{g} \leq 1$. Finally, $f \approx g$ means

$$0 < \liminf \frac{f}{g} \leq \limsup \frac{f}{g} < \infty \ ,$$

and $f \precsim g$ means

$$\limsup \frac{f}{g} < \infty.$$

Moreover, we use the Landau symbols o and $\mathcal{O}$.

Synopsis

The article is outlined as follows. In the next section, we provide asymptotic formulae for finite dimensional quantization problems. Historically there the concept of an intermediate convex optimization problem appeared for the first time. We proceed in Section 3 with a treatment of Banach space valued Gaussian signals under norm based distortion measures. In Section 4 we encouter the next intermediate optimization problem when considering Hilbert space valued Gaussian signals. Then it follows a treatment of 1-dimensional diffusions in Section 5. Here again convex minimization problems will play a crucial role in the analysis. Finally, we conclude the article in Section 6 with two further approaches for deriving (weak) asymptotic formulae. One is applicable for 1-dimensional stochastic processes and the other one for Lévy processes.

2 Finite dimensional signals

This section is devoted to some classical and some recent results on finite dimensional quantization. First we will introduce the concept of an intermediate optimization problem in the classical setting. The following section then proceeds with a treatment of Orlicz norm distortions.

2.1 Classical setting

Suppose now that X is a $\mathbb{R}^d$-valued original, where d denotes an arbitrary integer. Moreover, fix a norm $|\cdot|$ on $\mathbb{R}^d$, let ρ be the corresponding norm-based distortion and fix a moment $s > 0$. We shall consider the asymptotic quantization problem under the additional assumptions that the absolutely continuous part μ_c of $\mu = \mathcal{L}(X)$ (w.r.t. Lebesgue measure) does not vanish and that $\mathbb{E}[|X|^{\tilde{s}}] < \infty$ for some $\tilde{s} > s$.

Originally, this problem has been addressed by Zador (1966) and by Bucklew & Wise (1982). More recently, Graf & Luschgy (2000) presented a slightly extended version of the original results in their mathematical account on finite dimensional quantization. Let us state the asymptotic formula:

Theorem 2.1.
$$\lim_{n\to\infty} n^{1/d}\, D^{(q)}(\log n, s) = c(|\cdot|, s) \left\| \frac{d\mu_c}{d\lambda^d} \right\|_{L^{d/(d+s)}(\lambda^d)}^{1/s}.$$
Here, $c = c(|\cdot|, s) \in (0,\infty)$ is a constant depending on the norm $|\cdot|$ and the moment s only.

Note that the singular part of μ has no influence on the asymptotic quantization problem. Unfortunately, the constant c is known explicitly only in a few cases (e.g. when $|\cdot|$ denotes 2-dimensional Euclidean norm or a d-dimensional supremum norm). Notice that it is convenient to state the result in the number of approximating points rather than in the rate.

A heuristic derivation of the point allocation problem

Let us shortly give the main ideas along which one can prove the theorem. First one considers a $[0,1)^d$-uniformly distributed random variable X as original. Based on the self similarity of the uniform distribution one can show that

$$\lim_{n\to\infty} n^{1/d}\, D^{(q)}(\log n, s) = \inf_{n\in\mathbb{N}} n^{1/d}\, D^{(q)}(\log n, s) = c, \tag{3}$$

where the constant c lies in $(0,\infty)$ and agrees with the c in the asymptotic formula above.

In order to explain the findings for the general case, first assume that μ is absolutely continuous w.r.t. Lebesgue measure and that its density $h = \frac{d\mu}{d\lambda^d}$ can be represented as

$$h(x) = \sum_{i=1}^{m} \mathbb{1}_{B_i}(x)\, h_i, \tag{4}$$

where $m \in \mathbb{N}$ and $B_i \subset \mathbb{R}^d$ $(i = 1,\dots,m)$ denote disjoint cuboids having side length l_i and $(h_i)_{i=1,\dots,m}$ is a $\mathbb{R}_+^m$-valued vector. We want to analyze a quantization scheme based on a combination of optimal codebooks for the measures $\mathcal{U}(B_i)$, where in general $\mathcal{U}(B)$ denotes the uniform distribution on the set B. We fix a function $\xi : \mathbb{R}^d \to [0,\infty)$ with $\int \xi = 1$ of the form

$$\xi(x) = \sum_{i=1}^{m} \mathbb{1}_{B_i}(x)\, \xi_i \tag{5}$$

and denote by $\mathcal{C}_i$ an $\mathcal{U}(B_i)$-optimal codebook of size $n\,\xi_i\,\lambda^d(B_i) = n\int_{B_i}\xi$. Here, n parametrizes the size constraint of the global codebook $\mathcal{C}$ defined as $\mathcal{C} = \bigcup_{i=1}^m \mathcal{C}_i$. Notice that the inferred error $\mathbb{E}\min_{\hat{x}\in\mathcal{C}} |X - \hat{x}|^s$ is for large n approximately equal to

$$\sum_{i=1}^{m} \underbrace{l_i^s c^s (n\xi_i\lambda^d(B_i))^{-s/d}}_{\text{av. error in box } i}\ \underbrace{h_i\lambda^d(B_i)}_{\mathbb{P}(X\in B_i)} = \left(\frac{c}{n^{1/d}}\right)^s \sum_{i=1}^{m} \xi_i^{-s/d}\, h_i\lambda^d(B_i) = \left(\frac{c}{n^{1/d}}\right)^s \int \xi(x)^{-s/d} h(x)\, dx.$$

Here the average error in box i can be explained as follows: the terms $c^s(n\xi_i\lambda^d(B_i))^{-s/d}$ represent the error obtained for the $\mathcal{U}([0,1)^d)$-distribution when applying an optimal codebook of size $n\xi_i\lambda^d(B_i)$ (due to (3)). The term l_i^s arises from the scaling needed to get from $\mathcal{U}([0,1)^d)$ to $\mathcal{U}(B_i)$ and the fact that we consider the s-th moment.

We arrive at the minimization problem

$$\int \xi(x)^{-s/d} h(x)\, dx = \min! \tag{6}$$

where the infimum is taken over all non-negative ξ with $\int \xi = 1$. This convex optimization problem has a unique solution that will be given explicitly below. In particular, its solution is of the form (5) in our particular setting.

On the other hand, accepting that for "good" codebooks a typical realization of X that has fallen into B_i is also approximated by an element of B_i allows one to conclude that the above construction is close to optimal. In particular, one might correctly guess that the minimal value of the minimization problem (6) leads to the optimal asymptotic rate in the quantization problem.

The function ξ will be called *point density* and the minimization problem shall be called *point allocation problem*.

Rigorous results related to the point allocation problem

The importance of the point allocation problem in the classical setting was firstly conjectured by Lloyd and Gersho (see Gersho (1979)). First rigorous proofs are due to Bucklew (1984). In the general setting (as introduced in the first lines of this section) one can prove rigorously the following statement.

Result 2.2. *For each $n \in \mathbb{N}$ fix a codebook $\mathcal{C}(n) \subset \mathbb{R}^d$ with at most n elements and consider the associated empirical measure*

$$\nu_n = \frac{1}{n} \sum_{\hat{x} \in \mathcal{C}(n)} \delta_{\hat{x}}.$$

Supposing that for some infinite index set $I \subset \mathbb{N}$ the measures $(\nu_n)_{n \in I}$ converge vaguely to a measure ν, one can prove that

$$\liminf_{\substack{n\to\infty \\ n \in I}} n^{s/d}\, \mathbb{E}[\min_{\hat{x}\in\mathcal{C}(n)} |X - \hat{x}|^s] \;\geq\; c^s \int_{\mathbb{R}^d} \xi(x)^{-s/d}\, h(x)\, dx, \tag{7}$$

where $\xi = \frac{d\nu_c}{d\lambda^d}$ and $h = \frac{d\mu_c}{d\lambda^d}$. On the other hand, one can prove that for an arbitrary measurable function $\xi : \mathbb{R}^d \to [0, \infty)$ with $\int \xi \leq 1$,

$$\limsup_{n\to\infty} n^{s/d}\, D^{(q)}(\log n, s)^s \;\leq\; c^s \int_{\mathbb{R}^d} \xi(x)^{-s/d}\, h(x)\, dx. \tag{8}$$

Consequently, the solution of the point allocation problem leads to the asymptotics of the quantization error. Equations (7) and (8) are even more powerful: they show that for an asymptotically optimal family $(\mathcal{C}(n))_{n\in\mathbb{N}}$ of codebooks, in the sense that

$$|\mathcal{C}(n)| \leq n \;\text{ and }\; \mathbb{E}[\min_{\hat{x}\in\mathcal{C}(n)} |X - \hat{x}|^s]^{1/s} \lesssim D^{(q)}(\log n, s),$$

any accumulation point of $(\nu_n)_{n\in\mathbb{N}}$ is a minimizer of the point allocation problem (more explicitly $\xi = \frac{d\nu_c}{d\lambda^d}$ with ν_c denoting the continuous part of the accumulation point is a minimizer). This follows from the fact that the set of non-negative finite measures on $\mathbb{R}^d$ with total mass less or equal to 1 is compact in the vague topology. Since the objective function is even *strictly convex* in ξ, the minimizer is unique up to Lebesgue null-sets. Together with the property that the optimal ξ satisfies $\int \xi = 1$, one concludes that $(\nu_n)_{n\in\mathbb{N}}$ converges in the weak topology to a measure having as density the solution ξ to the point allocation problem.

It remains to solve the point allocation problem. Applying the reverse Hölder inequality with adjoint indices $p = -d/s$ and $q = d/(s+d)$ one gets for ξ with $\int \xi \le 1$

$$\int_{\mathbb{R}^d} \xi(x)^{-s/d}\, h(x)\, dx \ge \|\xi^{-s/d}\|_{L^p(\mathbb{R}^d)}\, \|h\|_{L^q(\mathbb{R}^d)} = \|\xi\|_{L^1(\mathbb{R}^d)}^{-s/d} \|h\|_{L^q(\mathbb{R}^d)} \ge \Big\| \frac{d\mu_c}{d\lambda^d} \Big\|_{L^{d/(d+s)}(\mathbb{R}^d)}. \quad (9)$$

The achievability of this lower bound can be easily verified for the optimal point density:

$$\xi(x) = \frac{1}{\int h^{d/(d+s)}}\, h(x)^{d/(d+s)}. \quad (10)$$

Consequently, it follows that for an asymptotically optimal family of codebooks $(\mathcal{C}(n))_{n\in\mathbb{N}}$ (in the sense mentioned above) the empirical measures ν_n ($n \in \mathbb{N}$) converge to a probability measure having density ξ given by (10).

Moreover, estimate (7) gives immediately a lower bound for the efficiency of *mismatched codebooks*: when using asymptotically optimal codebooks $(\mathcal{C}(n))_{n\in\mathbb{N}}$ for the moment s in the case where the underlying distortion is taken to a different moment $s' > 0$, one has:

$$\liminf_{n\to\infty} n^{1/d}\, \mathbb{E}[\min_{\hat{x}\in\mathcal{C}(n)} |X-\hat{x}|^{s'}]^{1/s'} \;\ge\; c(|\cdot|, s') \Big(\int h^{d/(d+s)}\Big)^{1/d} \Big(\int h^{1-\frac{s'}{d+s}}\Big)^{1/s'}.$$

Interestingly, the latter integral is infinite when $\lambda^d(\{h > 0\}) = \infty$ and $s' > d + s$, so that in that case the rate of convergence to zero is of a different order. For further reading concerning mismatched codebooks, we refer to the article by Graf, Luschgy & Pagès (2006).

We have presented this classical quantization result in detail, since it represents a stereotype of a coding result. Typically, the optimization problem (related to a coding problem) is non-convex and it is not possible to give explicit solutions. In practice, one needs to apply numerical or probabilistic methods to obtain good solutions. In order to analyze the problem, a powerful tool is to relate the original problem to an intermediate convex minimization problem. Such a relation then typically allows to derive rigorously asymptotic formulae. The intermediate problem is also of practical interest: for instance in the above example, the optimal point density can be used to initialize procedures used for generating close to optimal codebooks.

2.2 Orlicz norm distortion

In the classical setting, the objective function for the approximation loss is given as

$$\mathbb{E}[|X-\hat{X}|^s]^{1/s}.$$

Let now $f : [0,\infty) \to [0,\infty)$ be an increasing left continuous function with $\lim_{t\downarrow 0} f(t) = 0$. It is natural to pose the question what happens when replacing the objective function by

$$\mathbb{E}[f(|X-\hat{X}|)].$$

This problem has been treated by Delattre, Graf, Luschgy & Pagès (2004) in the case where f behaves like a polynomial at 0, that means

$$\lim_{\varepsilon\downarrow 0} \frac{f(\varepsilon)}{\kappa\varepsilon^\alpha} = 1,$$

for two parameters $\alpha, \kappa > 0$. They found that under certain concentration assumptions on X, the asymptotic quantization problem behaves as in the classical case with $s = \alpha$. In particular, the optimal point density does not differ from the one derived before.

Somehow this result is non-satisfying, in the sense that if $f(\varepsilon)$ and $\kappa\varepsilon^\alpha$ differ strongly in a reasonable range (the approximation error one typically expects for a given point constraint), then the corresponding optimal point density does not give a reasonable description of good codebooks. A possibility to remedy this effect and to get a whole family of point density functions is by considering Orlicz-norm distortions instead. Let us first introduce the necessary notation.

Let $\varphi : [0,\infty) \to [0,\infty)$ be a monotonically increasing, left continuous function with $\lim_{t\downarrow 0}\varphi(t) = 0$. Note that this implies that φ is lower semicontinuous. We assume that $\varphi \neq 0$ and let $|\cdot|$ denote an arbitrary norm on $\mathbb{R}^d$. For any $\mathbb{R}^d$-valued r.v. Z, the Orlicz norm $\|\cdot\|_\varphi$ is defined as

$$\|Z\|_\varphi = \inf\Big\{t \geq 0 : \mathbb{E}\,\varphi\Big(\frac{|Z|}{t}\Big) \leq 1\Big\},$$

with the convention that the infimum of the empty set is equal to infinity. Actually, the left continuity of φ together with monotone convergence implies that the infimum is attained, whenever the set is nonempty. We set

$$L^\varphi(\mathbb{P}) = \{Z : Z\ \mathbb{R}^d\text{-valued r.v. with } \|Z\|_\varphi < \infty\}.$$

Then $\|\cdot\|_\varphi$ defines a norm on $L^\varphi(\mathbb{P})$ when φ is convex (otherwise the triangle inequality fails to hold). We will not assume convexity for φ. Nevertheless, with a slight abuse of notation, we will allow ourselves to call $\|\cdot\|_\varphi$ an Orlicz norm. Choosing $\varphi(t) = t^p$, $p \geq 1$, yields the usual $L^p(\mathbb{P})$-norm.

Now we choose as objective function

$$\|X - \hat{X}\|_\varphi$$

and we denote by $D^{(q)}(r)$ the corresponding minimal quantization error of rate $r \geq 0$. The concept of introducing an intermediate optimization problem is also applicable in the Orlicz-norm setting. Let us quote the main results taken from Dereich & Vormoor (2005).

The constant c of the classical setting is now replaced by a convex decreasing function $g : (0,\infty) \to [0,\infty)$ which may be defined via

$$g(\zeta) = \lim_{n\to\infty} \inf_{\mathcal{C}(n)} \mathbb{E}\varphi\big((n/\zeta)^{1/d}\, d(U, \mathcal{C}(n))\big), \tag{11}$$

where the infima are taken over all finite sets $\mathcal{C}(n) \subset \mathbb{R}^d$ with $|\mathcal{C}(n)| \leq n$ and U denotes a uniformly distributed r.v. on the unit cube $[0,1)^d$.

Theorem 2.3. *Suppose that $\mathbb{E}\psi(|X|) < \infty$ for some function ψ satisfying a growth condition (G) depending on φ. Then*

$$\lim_{n\to\infty} n^{1/d}\, D^{(q)}(\log n) = J^{1/d}$$

where J is the minimal value in the optimization problem

$$\int_{\mathbb{R}^d} \xi(x)\, dx = \min!$$

among all non-negative ξ with

$$\int_{\mathbb{R}^d} g\big(\xi(x)\big)\, h(x)\, dx \leq 1,$$

where h denotes again the density of μ_c.

The point allocation problem in the Orlicz norm setting

In the Orlicz-norm setting, the analogs of (7) and (8) are summarized in the following statement.

Result 2.4. *Let $I \subset (0,\infty)$ denote an index set with $\sup I = \infty$, and denote by $(\mathcal{C}(\eta))_{\eta\in I}$ a family of codebooks such that the associated measures*

$$\nu_\eta = \frac{1}{\eta} \sum_{\hat{x}\in\mathcal{C}(\eta)} \delta_{\hat{x}}$$

converge vaguely to a finite measure ν. Then one can prove that

$$\liminf_{\substack{\eta\to\infty\\ \eta\in I}} \mathbb{E}\,\varphi\big(\eta^{1/d}\, d(X,\mathcal{C}(\eta))\big) \geq \int g\big(\xi(x)\big)\, d\mu_c(x) \tag{12}$$

for $\xi = \frac{d\nu_c}{d\lambda^d}$ where ν_c denotes again the absolutely continuous part of ν. On the other hand, for any non-negative $\xi : \mathbb{R}^d \to [0,\infty)$ with $J := \int \xi < \infty$ there exists a family of codebooks $(\mathcal{C}(\eta))_{\eta\geq 1}$ such that $\limsup_{\eta\to\infty} |\mathcal{C}(\eta)|/\eta \leq J$ and

$$\limsup_{\eta\to\infty} \mathbb{E}\,\varphi\big(\eta^{1/d}\, d(X,\mathcal{C}(\eta))\big) \leq \int g\big(\xi(x)\big)\, d\mu_c(x). \tag{13}$$

Similar as in the classical setting one can heuristically verify these two estimates for μ with density h of the form (4).

Let us show how the estimate (13) can be used to prove the upper bound in the asymptotic formula. Recall that J is the minimal value of

$$\int_{\mathbb{R}^d} \xi(x)\, dx = \min!$$

where the infimum is taken among all non-negative ξ with

$$\int_{\mathbb{R}^d} g\big(\xi(x)\big)\, h(x)\, dx \leq 1,$$

and where h denotes again the density of μ_c.

Fix $\varepsilon > 0$. As one can easily derive from (13) there exists a family of codebooks $(\mathcal{C}(\eta))_{\eta\geq 1}$ such that for sufficiently large η

$$|\mathcal{C}_\eta| \leq (J+\varepsilon)\eta \quad \text{and} \quad \mathbb{E}\,\varphi\big(\eta^{1/d}\, d(X,\mathcal{C}(\eta))\big) \leq 1.$$

Note that the latter estimate is equivalent to $\|d(X, \mathcal{C}(\eta))\|_\varphi \leq \eta^{-1/d}$. Consequently, for sufficiently large η one has

$$D\big(\log\big((J+\varepsilon)\eta\big)\big) \leq \eta^{-1/d}$$

or equivalently switching from η to $\bar{\eta} = (J+\varepsilon)\,\eta$:

$$D\big(\log \bar{\eta}\big) \leq (J+\varepsilon)^{1/d} \bar{\eta}^{-1/d}.$$

This proves the upper bound. The proof of the lower bound is similar and therefore omitted.

Solutions to the point allocation problem

Solutions to the point allocation problem can be represented in terms of the conjugate $\bar{g} : [0,\infty) \to [0,\infty)$ defined as

$$\bar{g}(a) = \inf_{\eta \geq 0} [a\eta + g(\eta)], \qquad a \geq 0.$$

The function $\bar{g}$ is continuous, monotonically increasing and concave, and it satisfies $\bar{g}(0) = 0$.

Result 2.5. *We suppose that*

$$\mu_c(\mathbb{R}^d) \sup_{t \geq 0} \varphi(t) > 1.$$

(Otherwise $J = 0$ and $\xi = 0$ is an optimal point density.) The point allocation problem has an integrable solution iff the integral

$$\int \bar{g}\Big(\frac{\vartheta}{h(x)}\Big)\, d\mu_c(x) \tag{14}$$

is finite for some $\vartheta > 0$. In such a case there exists a parameter $\zeta > 0$ such that the optimal point density ξ satisfies

$$\int g(\xi(x))\, d\mu_c(x) = 1 \quad \text{and} \quad \bar{g}'_+\Big(\frac{\zeta}{h(x)}\Big) \leq \xi(x) \leq \bar{g}'_-\Big(\frac{\zeta}{h(x)}\Big), \qquad x \in \mathbb{R}^d. \tag{15}$$

Here, the functions $\bar{g}'_+$ and $\bar{g}'_-$ denote the right hand side and left hand side derivative of $\bar{g}$, respectively, and we denote $\bar{g}'_+(\infty) = \bar{g}'_-(\infty) = 0$. In particular, the optimal point density is unique, whenever $\bar{g}$ is differentiable or - expressed in terms of g - whenever g is strictly convex.

Back to the original problem

For simplicity we assume that $\bar{g}$ is continuous, and, for any given $\zeta > 0$, we let ξ^ζ be the point density $\xi^\zeta(x) = \bar{g}'\big(\frac{\zeta}{h(x)}\big)$ $(x \in \mathbb{R}^d)$ and denote by $\bar{\xi}^\zeta(x) = \xi^\zeta(x)/\|\xi\|_{L^1(\lambda^d)}$ its normalized version. Suppose now that $\hat{X}^{(n)}$ is a quantizer of rate $\log n$ (assuming at most n different values) that minimizes the objective function

$$\mathbb{E} f(|X - \hat{X}^{(n)}|).$$

We denote by δ its minimal value and set $\bar{\varphi} = \frac{1}{\delta} f$. Then $\hat{X}^{(n)}$ is also an optimal quantizer for the Orlicz-norm objective function

$$\|X - \hat{X}^{(n)}\|_{\bar{\varphi}}.$$

Using the ideas above we can link the problem to a point allocation problem and find an normalized optimal point density $\bar{\xi}$. Recall that the definition of $\bar{\varphi}$ still depends on δ. However, a straight forward analysis shows that $\bar{\xi}$ is also contained in the family $(\bar{\xi}^\zeta)_{\zeta>0}$ when taking $\varphi = f$.

Hence, one can relate, for given $n \in \mathbb{N}$, the original quantization problem to a normalized point density of the family $(\bar{\xi}^\zeta)_{\zeta>0}$. We believe that this approach leads to more reasonable descriptions of optimal codebooks for moderate n.

3 Gaussian signals

In this section we summarize results on the asymptotic coding problems for general Banach space-valued (centered) Gaussian signals. Let us be more precise about our setting: we fix a separable Banach space $(E, \|\cdot\|)$ and call an E-valued (Borel measurable) random variable X *Gaussian signal* iff for any f in the topological dual E' of E the real-valued random variable $f(X)$ is a zero-mean normal distributed random variable. In this context, the Dirac measure in 0 is also conceived as normal distribution. A typical example is, for instance, X being a (fractional) Wiener process considered in the space of continuous functions $E = C[0,1]$ endowed with supremum norm. In general, we will look at the norm-based distortion measure $\rho(x, \hat{x}) = \|x - \hat{x}\|$ so that our objective function is again

$$\mathbb{E}[\|X - \hat{X}\|^s]^{1/s}.$$

3.1 Asymptotic estimates

As was firstly observed in the dissertation by Fehringer (2001) the quantization problem is linked to the behavior of the so called *small ball function*, that is the function

$$\varphi(\varepsilon) = -\log \mathbb{P}(\|X\| \le \varepsilon) \qquad (\varepsilon > 0).$$

We summarize the results in the following theorem.

Theorem 3.1. *Suppose that the small ball function satisfies*

$$\varphi^{-1}(\varepsilon) \approx \varphi^{-1}(2\varepsilon) \quad \text{as} \quad \varepsilon \downarrow 0,$$

where φ^{-1} denotes the inverse of φ. Then for all moments $s \ge 1$ one has

$$\varphi^{-1}(r) \lesssim D(r, s) \le D^{(q)}(r, s) \lesssim 2\varphi^{-1}(r/2), \qquad r \to \infty.$$

The lower and upper bounds were first proved for the quantization error in Dereich, Fehringer, Matoussi & Scheutzow (2003). The remaining lower bound for the distortion rate function was derived in Dereich (2005).

Remark 3.2. i) The upper and lower bound do not depend on s. This suggests that "good" codebooks lead to a random approximation error that is concentrated around a typical value or interval. The problem of proving such a result is still open in the general setting. However, as we will see below one can get stronger results in several particular settings.

ii) The small ball function is a well studied object. Mostly, the small ball function of a functional signal satisfies the assumptions of the theorem and the estimates provided by the theorem agree asymptotically up to some factor. Thus one can immediately infer the rate of convergence in the coding problems for several Gaussian processes. For a general account on small deviations one might consult Li & Shao (2001).

iii) The asymptotic behavior of the small ball function at zero is related to other quantities describing the complexity such as entropy numbers (Kuelbs & Li (1993), Li & Linde (1999)), Kolmogorov width and average Kolmogorov width. A general treatment of such quantities together with the quantization problem can be found in the dissertation by Creutzig (2002) (see also Carl & Stephani (1990)).

The proof of the main result relies heavily on the measure concentration features of Gaussian measures exhibited by the isoperimetric inequality and the Ehrhard inequality. A further important tool is the Cameron-Martin formula combined with the Anderson inequality.

3.2 Some examples

We summarize some results that can be extracted from the link to the small ball function. We only give the results for the approximation problems. For a general account we refer the reader to Li & Shao (2001).

First let $X = (X_t)_{t\in[0,1]}$ denote a Wiener process.

- When E is chosen to be $C[0,1]$ endowed with supremum norm, one gets

$$\frac{\pi}{\sqrt{8r}} \lesssim D(r,s) \leq D^{(q)}(r,s) \lesssim \frac{\pi}{\sqrt{r}}$$

  as $r \to \infty$.

- If $E = L_p[0,1]$ $(p \geq 1)$, then one has

$$\frac{c_p}{\sqrt{r}} \lesssim D(r,s) \leq D^{(q)}(r,s) \lesssim \frac{\sqrt{8}c_p}{\sqrt{r}}$$

  where

$$c_p = 2^{1/p}\sqrt{p}\Big(\frac{\lambda_1(p)}{2+p}\Big)^{(2+p)/2p}$$

  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.

- If $E = C^\alpha$ $(\alpha \in (0,1/2))$ is the space of α-Hölder continuous functions over the time $[0,1]$ endowed with the standard Hölder norm

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

  then

$$\frac{c_\alpha}{r^{(1-2\alpha)/2}} \lesssim D(r,s) \leq D^{(q)}(r,s) \lesssim 2^{(3-2\alpha)/2}\,\frac{c_\alpha}{r^{(1-2\alpha)/2}}$$

  for a constant $c_\alpha > 0$ not known explicitly.

Next, let us consider a fractional Brownian sheet $X = (X_t)_{t\in[0,1]^d}$ with parameter $\gamma = (\gamma_1, \dots, \gamma_d) \in (0,2)^d$ as original signal. The process is characterized as the centered continuous Gaussian process on $[0,1]^d$ with covariance kernel

$$\mathbb{E}[X_t X_u] = \frac{1}{2^d} \prod_{j=1}^d \big[|t_j|^{\gamma_j} + |u_j|^{\gamma_j} - |t_j - u_j|^{\gamma_j}\big], \qquad t, u \in [0,1]^d.$$

As underlying space we consider $E = C([0,1]^d)$ the space of continuous functions endowed with supremum norm.

If there is a unique minimum, say γ_1, in $\gamma = (\gamma_1, \dots, \gamma_d)$, one has

$$D(r,s) \approx D^{(q)}(r,s) \approx r^{-\gamma_1/2}, \qquad r \to \infty$$

(Mason & Shi 2001). Whereas, if there are two minimal coordinates, say γ_1 and γ_2, then

$$D(r,s) \approx D^{(q)}(r,s) \approx r^{-\gamma_1/2} (\log r)^{1+\gamma_1/2}$$

(Belinsky & Linde 2002, Talagrand 1994). For the case that there are more than two minimal elements in the vector γ, it is still an open problem to find the weak asymptotic order of the small ball function.

3.3 A particular random coding strategy

We will now introduce a random coding strategy that has been originally used to prove the upper bound in Theorem 3.1. Let $(Y_i)_{i\in\mathbb{N}}$ be a sequence of independent random vectors with the same law as X. We consider the random set $\mathcal{C}(r) = \{Y_1, \dots, Y_{\lfloor e^r \rfloor}\}$ ($r \geq 0$ indicating again the rate) as codebook for X, and set

$$D^{(r)}(r,s) = \mathbb{E}[\min_{i=1,\dots,\lfloor e^r \rfloor} \|X - Y_i\|^s]^{1/s}$$

A detailed analysis of this approximation error has been carried out in Dereich & Lifshits (2005):

Theorem 3.3. *Assume that there exists $\kappa < \infty$ such that*

$$\Big(1 + \frac{1}{\kappa}\Big)\varphi(2\varepsilon) \leq \varphi(\varepsilon) \leq \kappa\varphi(2\varepsilon)$$

for all sufficiently small $\varepsilon > 0$. Then there exists a continuous, strictly decreasing function $\varphi_* : (0,\infty) \to (0,\infty)$ such that

$$D^{(r)}(r,s) \sim \varphi_*^{-1}(r)$$

for any $s > 0$.

The function φ_* can be represented in terms of a random small ball function. Let

$$\ell_\varepsilon(x) = -\log \mathbb{P}(\|X - x\| \leq \varepsilon) = -\log \mu(B(x,\varepsilon)) \qquad (x \in E, \varepsilon > 0);$$

then φ_* can be chosen as $\varphi_*(\varepsilon) = \mathbb{E}\ell_\varepsilon(X)$.

The proof of the theorem relies on the following strong limit theorem: assuming that there is a constant $\kappa < \infty$ such that $\varphi(\varepsilon) \le \kappa\varphi(2\varepsilon)$ for sufficiently small $\varepsilon > 0$, one has

$$\lim_{\varepsilon \downarrow 0} \frac{\ell_\varepsilon(X)}{\varphi_*(\varepsilon)} = 1, \text{ a.s.}$$

In information theory, the concept of proving a strong limit theorem in order to control the efficiency of a coding procedure is quite common. For instance, the proof of Shannon's Source Coding Theorem can be based on such a result. Since $\ell_\varepsilon(X)$ is concentrated around a typical value, the conditional probability $\mathbb{P}(d(X, \mathcal{C}(r)) \le \varepsilon | X)$ almost does not depend on the realization of X when $\varepsilon > 0$ is small. Moreover, for large rate r, there is a critical value ε_c for which the probability decays fastly from almost 1 to almost 0. Around this critical value the approximation error is highly concentrated, and the moment s does not have an influence on the asymptotics. Such a strong limit theorem is often referred to as *asymptotic equipartition property.* For further reading concerning the asymptotic equipartition problem in the context of Shannon's Source Coding Theorem, we refer the reader to Dembo & Kontoyiannis (2002).

Open Problem 3.4. Can one prove the equivalence of moments in the quantization problem under weak assumption on the small ball function φ?

3.4 The fractional Wiener process

In this subsection we consider a fractional Wiener process $X = (X_t)_{t \in [0,1]}$ with Hurst index $H \in (0,1)$ as underlying signal. Its distribution is characterized as the unique Gaussian measure on $C[0,1]$ with covariance kernel:

$$\mathbb{E}[X_u \, X_v] = \frac{1}{2}\big[u^{2H} + v^{2H} - |u-v|^{2H}\big].$$

We state the main result.

Theorem 3.5. *If $E = L^p[0,1]$ for some $p \ge 1$, there exists a constant $\kappa > 0$ such that for any $s > 0$*

$$\lim_{r \to \infty} r^H \, D^{(q)}(r,s) = \lim_{r \to \infty} r^H \, D(r,s) = \kappa_p. \tag{16}$$

Additionally, if $E = C[0,1]$, there exists a constant $\kappa > 0$ such that for any $s > 0$

$$\lim_{r \to \infty} r^H \, D^{(q)}(r,s) = \lim_{r \to \infty} r^H \, D^{(e)}(r,s) = \kappa_\infty.$$

The results are proved for the entropy and quantization constraint in Dereich & Scheutzow (2006). The extension to the distortion rate function is established in Dereich (2006*a*).

Remark 3.6.
- In the artice the result is proved for 1-dimensional processes only. However, having a careful look at the proof, the result remains true for the multi dimensional Wiener process, too.
- The moment s and the choice of the complexity constraint do not influence the asymptotics of the approximation error. Thus for good approximation schemes the approximation

error is concentrated around a typical value: for fixed $s > 0$ let $(\mathcal{C}(r))_{r\geq 0}$ be a family of asymptotically optimal codebooks in the sense that

$$\log |\mathcal{C}(r)| \leq r \quad \text{and} \quad \mathbb{E}[\min_{\hat{x}\in\mathcal{C}(r)} \|X - \hat{x}\|^s]^{1/s} \sim D^{(q)}(r, s);$$

then one has

$$\min_{\hat{x}\in\mathcal{C}(r)} \|X - \hat{x}\| \sim D^{(q)}(r, s), \quad \text{in probability,}$$

in the sense that, for any $\varepsilon > 0$,

$$\lim_{r\to\infty} \mathbb{P}\big((1-\varepsilon)D^{(q)}(r, s) \leq \min_{\hat{x}\in\mathcal{C}(r)} \|X - \hat{x}\| \leq (1+\varepsilon)D^{(q)}(r, s)\big) = 1.$$

- The equivalence of moments and the equivalence in the coding quantities are very special, and as we will see below more heterogeneous processes as the diffusion processes will not share these features. In Dereich & Scheutzow (2006) one uses the equivalence of moments to prove the equivalence of the entropy and quantization constraint, and both features seem to be related in a general way.

Open Problem 3.7. It is still an open problem whether one can prove (16) in the case where $E = C[0, 1]$.

4 Hilbert space-valued Gaussian signals

Let now $E = H$ denote a (separable) Hilbert space and X be a H-valued Gaussian signal. In that case one can represent X in the Karhunen-Loève expansion, that is as the a.s. limit

$$X = \sum_i \sqrt{\lambda_i}\xi_i e_i, \tag{17}$$

where

- (λ_i) is a $\mathbb{R}_+$-valued sequence (the eigenvalues of the corresponding covariance operator),
- (e_i) is a orthonormal system in H (the corresponding normalized eigenvalues) and
- (ξ_i) is an i.i.d. sequence of standard normals.

Then there is an isometric isomorphism π mapping the range of X (i.e. the smallest closed set in H containing X a.s.) to l^2 such that

$$\pi(X) = (\sqrt{\lambda_1}\xi_1, \sqrt{\lambda_2}\xi_2, \dots).$$

One can prove that applying a contraction on the original signal does not increase its coding complexity under either information constraint. Therefore, the coding quantities are the same for X in H as for $(\sqrt{\lambda_i}\xi_i)$ in l^2. Thus we can and will assume without loss of generality that X is given in the form $(\sqrt{\lambda_i}\xi_i)$.

Before we treat the general coding problem, we first restrict our attention to the distortion rate function $D(r, 2)$. It is one of the few examples that can be given explicitly in terms of a solution to a rate allocation problem. We start with providing some elementary results on mutual information and distortion rate functions.

4.1 The mutual information and Shannon's source coding theorem

Let us introduce the notion of conditional mutual information. For random variables A, B and C taking values in Borel spaces, we denote

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

where in general

$$H(P\|Q) = \begin{cases} \int \log \frac{dP}{dQ}\, dP & \text{if } P \ll Q \\ \infty & \text{else} \end{cases}$$

denotes the *relative entropy*. Then one denotes by $I(A;B|C) = \int I(A;B|C=c)\, d\mathbb{P}_C(c)$ the *mutual information of A and B conditional on C*.

The mutual information can be verified to satisfy the following properties (see for instance Ihara (1993))

(i) $I(A;B|C) \geq 0$ (*positivity*); $I(A;B|C) = 0$ iff A and B are independent given C

(ii) $I(A;B|C) = I(B;A|C)$ (*symmetry*)

(iii) $I(A_1, A_2; B|C) = I(A_1;B|C) + I(A_2;B|A_1, C)$; in particular, $I(A_1;B|C) \leq I(A_1, A_2; B|C)$ (*monotonicity*)

(iv) $I(A;B|C) \geq I(\varphi(A);B|C)$ for a Borel measurable map φ between Borel spaces.

All above results remain valid for the (unconditional) mutual information.

The mutual information constraint has its origin in Shannon's celebrated source coding theorem (see Shannon (1948) for the original version):

Theorem 4.1 (Shannon's Source Coding Theorem)**.** *Let μ be a probability measure on a Borel space E and let $\rho : E \times E \to [0,\infty]$ be a product measurable map. If there exist $\bar{x} \in E$ and $s > 1$ with $\int \rho(x,\bar{x})^s\, d\mu(x) < \infty$, then one has for any continuity point r of the associated distortion rate function $D(\cdot|\mu,\rho)$:*

$$\lim_{m\to\infty} \frac{1}{m} D^{(q)}(mr|\mu^{\otimes m}, \rho_m) = D(r|\mu,\rho),$$

where $\rho_m : E^m \times E^m \to [0,\infty]$ is defined as

$$\rho_m(x,\hat{x}) = \sum_{i=1}^{m} \rho(x_i, \hat{x}_i).$$

Remark 4.2.

- As a consequence of the convexity of the relative entropy, the distortion rate function is convex. Moreover, it is decreasing and bounded from below by 0. Consequently, it has at most one discontinuity in which it jumps from ∞ to some finite value.
- For a given $m \in \mathbb{N}$, the distortion measure ρ_m is called *single letter distortion measure*, since its value can be expressed as a sum of the errors inferred in each single letter (here the term letter refers to the single E-valued entries).

- The distortion rate function can be evaluated explicitly in a few cases only. One of these is the case where the original is normally distributed and the error criterion is given by the mean squared error:
$$D(r|\mathcal{N}(0,\sigma^2),|\cdot|^2)=\sigma^2 e^{-2r}.$$
In that case, there exists a unique minimizer (in the sense that the distribution of $(X;\hat{X})$ is unique) and the distribution can be given explicitly.

4.2 The derivation of the distortion rate function $D(r,2)$

Let us denote by $D(\cdot)$ the distortion rate function of the Gaussian original X under the objective function
$$\mathbb{E}[\|X-\hat{X}\|^2],$$
that is $D(r)=D(r,2)^2$ $(r\geq 0)$. The distortion rate function will be given as a solution to a rate allocation problem.

The lower bound

Suppose that $\hat{X}$ is a reconstruction for X with $I(X;\hat{X})\leq r$ for a given rate r. For $a,b\in\mathbb{N}$ we let $X_a^b=(X_i)_{i=a,\dots,b}$ with the natural extension when $b=\infty$ or $a>b$. With property (iii) one gets
$$I(X;\hat{X})=I(X_1;\hat{X})+I(X_2^\infty;\hat{X}|X_1)\geq I(X_1;\hat{X}_1)+I(X_2^\infty;\hat{X}|X_1).$$
Repeating the argument infinitely often yields
$$I(X;\hat{X})\geq\sum_{i=1}^\infty I(X_i;\hat{X}_i|X_1^{i-1}). \tag{18}$$
For each $i\in\mathbb{N}$ we conceive $r_i=I(X_i;\hat{X}_i|X_1^{i-1})$ as the rate allocated to the i-th coordinate. Note that one has $\sum_i r_i\leq r$.

Let us now fix $i\in\mathbb{N}$ and analyze $\mathbb{E}|X_i-\hat{X}_i|^2$. It can be rewritten as
$$\mathbb{E}|X_i-\hat{X}_i|^2=\mathbb{E}\big[\mathbb{E}[|X_i-\hat{X}_i|^2|X_1^{i-1}]\big] \tag{19}$$
Note that conditional upon X_1^{i-1} the random variable X_i is $\mathcal{N}(0,\lambda_i)$-distributed so that
$$\mathbb{E}[|X_i-\hat{X}_i|^2|X_1^{i-1}=x_1^{i-1}]\geq D\big(I(X_i;\hat{X}_i|X_1^{i-1}=x_1^{i-1})|\mathcal{N}(0,\lambda_i),|\cdot|^2\big). \tag{20}$$
Together with (19) and the convexity of the distortion rate functions one gets
$$\begin{aligned}\mathbb{E}|X_i-\hat{X}_i|^2&\geq\int D\big(I(X_i;\hat{X}_i|X_1^{i-1}=x_1^{i-1})|\mathcal{N}(0,\lambda_i),|\cdot|^2\big)\,d\mathbb{P}_{X_1^{i-1}}(x_1^{i-1})\\&\geq D(r_i|\mathcal{N}(0,\lambda_i),|\cdot|^2)=\lambda_i e^{-2r_i}.\end{aligned} \tag{21}$$
Altogether, we arrive at
$$\mathbb{E}\|X-\hat{X}\|^2\geq\sum_{i=1}^\infty\lambda_i e^{-2r_i}\quad\text{and}\quad\sum_{i=1}^\infty r_i\leq r.$$

The upper bound

Now we fix a non-negative sequence (r_i) with $\sum_i r_i \le r$. As mentioned in Remark 4.2, for each $i \in \mathbb{N}$, there exists a pair $(X_i, \hat{X}_i)$ (on a possibly enlarged probability space) satisfying

$$\mathbb{E}[|X_i - \hat{X}_i|^2] = \lambda_i e^{-2r_i} \quad \text{and} \quad I(X_i; \hat{X}_i) = r_i.$$

These pairs can be chosen in such a way that $(X_i, \hat{X}_i)_{i\in\mathbb{N}}$ form an independent sequence of random variables. Fix $N \in \mathbb{N}$ and note that due to property (iii) of mutual information one has

$$\begin{aligned} I(X_1^N; \hat{X}_1^N) &= I(X_2^N; \hat{X}_1^N) + I(X_1^N; \hat{X}_1^N | X_2^N) \\ &= I(X_2^N; \hat{X}_2^N) + \underbrace{I(X_2^N; \hat{X}_1 | \hat{X}_2^N)}_{=0} + \underbrace{I(X_1; \hat{X}_2^N | X_2^N)}_{=0} + \underbrace{I(X_1; \hat{X}_1 | X_2^N, \hat{X}_2^N)}_{=I(X_1;\hat{X}_1)}. \end{aligned}$$

Here the second and third term vanish due to the independence of the conditional distributions (property (i)). Moreover, one can remove the conditioning in the last term, since $(X_1, \hat{X}_1)$ is independent of $(X_2^N, \hat{X}_2^N)$. Repeating the argument now gives

$$I(X_1^N; \hat{X}_1^N) = \sum_{i=1}^N I(X_i; \hat{X}_i)$$

and a further argument (based on the lower semicontinuity of the relative entropy) leads to

$$I(X; \hat{X}) = \lim_{N\to\infty} I(X_1^N; \hat{X}_1^N) = \sum_{i=1}^\infty I(X_i; \hat{X}_i) = \sum_{i=1}^\infty r_i \le r.$$

Moreover, $\mathbb{E}\|X - \hat{X}\|^2 = \sum_i \lambda_i e^{-2r_i}$.

Kolmogorov's inverse water filling principle

Due to the computations above, the value $D(r)$ can be represented as the minimal value of the strictly convex optimization problem

$$\sum_i \lambda_i e^{-2r_i} = \min!$$

where the infimum is taken over all non-negative sequences (r_i) with $\sum_i r_i = r$. It is common to restate the minimization problem in terms of the errors $d_i = \lambda_i e^{-2r_i}$ inferred in the single coordinates as

$$\sum_i d_i = \min!$$

where (d_i) is an arbitrary non-negative sequence with

$$\sum_i \frac{1}{2} \log_+ \frac{\lambda_i}{d_i} = r. \tag{22}$$

Using Lagrange multipliers one finds that the unique minimizer is of the form

$$d_i = \kappa \wedge \lambda_i,$$

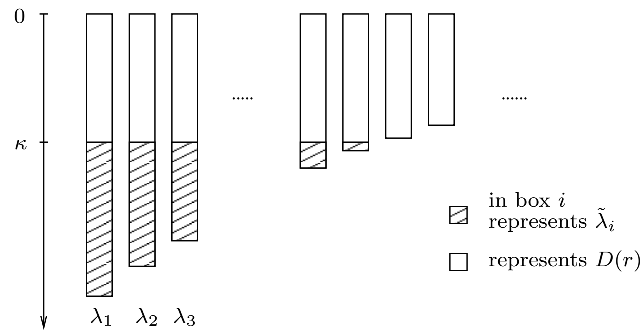

Figure 1: Inverse water filling principle

where $\kappa > 0$ is a parameter that needs to be chosen such that (22) is valid. This link between $r \geq 0$ and $\kappa \in (0, \lambda_1]$ provides a one-to-one correspondence. This result was originally derived by Kolmogorov (1956) and the formula is often referred to as *Kolmogorov's inverse water filling principle*. As illustrated in Figure 1, one can represent each coordinate as a box of size λ_i; then one fills water into the boxes until a level κ is reached for which the corresponding total rate is equal to r. The white area of the boxes represents the coding error. As we will see later also the striped area has an information theoretic meaning.

Résumé

Let us recall the main properties that enabled us to relate the coding problem of the process X to that of a single normal random variable. In order to derive the lower estimate of the error in one coordinate (see (20) and (21)) we used the independence of a single coordinate and the remaining coordinates of X. Then in order to conclude back on the total approximation error, we used that it is given as the sum of the approximation errors in the single coordinates (*single letter distortion measure*), so we could exchange sum and expectation to derive the result. We shall see later that the above ideas can be applied in similar settings provided one considers independent letters under a single letter distortion measure.

Mostly, the eigenvalues (λ_i) are not known explicitly and thus the minimization problem cannot be solved explicitly. However, the asymptotics of $D(\cdot)$ depend only on the asymptotic behavior of (λ_i), and in the case where (λ_i) is regularly varying, both expressions can be linked by an asymptotic formula: one has

$$D(r) \sim \frac{\alpha^\alpha}{2^{\alpha-1}(\alpha-1)} r \lambda_{\lceil r \rceil}, \tag{23}$$

if the sequence (λ_i) is regularly varying with index $-\alpha < -1$, i.e. there exists a continuous function $\ell : [1, \infty) \to \mathbb{R}$ such that

$$\lambda_i \sim i^{-\alpha} \ell(i) \text{ as } i \to \infty$$

and for any $\eta > 1$ one has $\ell(\eta t) \sim \ell(t)$ as $t \to \infty$.

4.3 Generalizations to other coding quantities (1st approach)

Let us now look at the quantization problem for $s = 2$. In this section we want to explain an approach taken by Luschgy & Pagès (2004) (see also Luschgy & Pagès (2002)) to prove the asymptotic equivalence of the quantization error $D^{(q)}(\cdot, 2)$ and the distortion rate function $D(\cdot, 2)$. We adopt a slightly more general setting than that of the original work. However, the generalization can be carried out easily, and we prefer to give the more general assumptions since we believe that these are the natural (most general) ones under which the following arguments work.

Theorem 4.3. *Assume that the eigenvalues are ordered by their size and that there is an increasing $\mathbb{N}_0$-valued sequence $(j_k)_{k\in\mathbb{N}}$ with $j_1 = 0$ such that*

(i) $\lim_{k\to\infty} j_{k+1} - j_k = \infty$ *and*

(ii) $\lim_{k\to\infty} \frac{\lambda_{j_k+1}}{\lambda_{j_{k+1}}} = 1$.

Then one has

$$D(r, 2) \sim D^{(q)}(r, 2)$$

Since the quantization error is always larger than the distortion rate function, it suffices to prove an appropriate upper bound for the quantization error.

Sketch of the proof

In the following, we still consider the single letter distortion measure $\rho(x, \hat{x}) = \|x - \hat{x}\|^2$ and we call the minimal value of the objective function $\mathbb{E}[\|X - \hat{X}\|^2]$ the quantization error of an approximation $\hat{X}$.

Based on the sequence (j_k) we decompose the process X into subbands

$$X_{j_k+1}^{j_{k+1}} = (X_{j_k+1}, \ldots, X_{j_{k+1}}) \qquad (k \in \mathbb{N}).$$

Note that due to property (ii), the eigenvalues corresponding to a subband do only differ by a factor tending to 1 as k tends to ∞. By replacing the eigenvalues of each subband by their largest value, one ends up with a process having a larger quantization error than the original one. Actually, one can show rigorously that the approximation error on each subband increases at most by the factor $\frac{\lambda_{j_k+1}}{\lambda_{j_{k+1}}}$. Since the values of finitely many eigenvalues do not effect the strong asymptotics of the distortion rate function, it remains to show asymptotic equivalence of both coding quantities for the modified process.

Each subband consists of an i.i.d. sequence of growing size and the distortion is just the sum of the distortions in the single letters. Thus we are in a situation where the Source Coding Theorem can be applied: similar to Proposition 4.4 in Luschgy & Pagès (2004) one can prove that

$$\eta(m) := \sup_{\substack{r\geq 0,\\ \sigma^2>0}} \frac{D^{(q)}(r|\mathcal{N}(0,\sigma^2)^{\otimes m}, |\cdot|^2)}{D(r|\mathcal{N}(0,\sigma^2)^{\otimes m}, |\cdot|^2)} \tag{24}$$

is finite for all $m \in \mathbb{N}$ and satisfies $\lim_{m\to\infty} \eta(m) = 1$.

Let us sketch the remaining proof of Theorem 4.3. For a given rate $r \geq 0$ we denote by (r_i) the corresponding solution to the rate allocation problem. Then choose for each subband $X_{j_k+1}^{j_{k+1}}$ an optimal codebook $\mathcal{C}_k \subset \mathbb{R}^{j_{k+1}-j_k}$ of size $\exp(\sum_{i=j_k+1}^{j_{k+1}} r_i)$ and denote by $\mathcal{C}$ the product codebook $\mathcal{C} = \prod_{k=1}^{\infty} \mathcal{C}_k$. It contains at most $\exp(\sum_{i=1}^{\infty} r_i) = \exp(r)$ elements. Moreover, the inferred coding error satisfies

$$\begin{aligned}\mathbb{E}[\min_{\hat{x}\in\mathcal{C}} \|X - \hat{x}\|^2] &= \sum_{k\in\mathbb{N}} \mathbb{E}[\min_{\hat{x}\in\mathcal{C}_k} |X_{j_k+1}^{j_{k+1}} - \hat{x}|^2] \\ &\leq \sum_{k\in\mathbb{N}} \eta(j_{k+1} - j_k)\, D(\sum_{i=1}^{\infty} r_i | X_{j_k+1}^{j_{k+1}}, |\cdot|^2).\end{aligned}$$

For arbitrary $\varepsilon > 0$ one can now fix k_0 such that for all $k \geq k_0$ one has $\eta(j_{k+1} - j_k) \leq 1 + \varepsilon$; then

$$\sum_{k\geq k_0} \eta(j_{k+1} - j_k)\, D\Big(\sum_{i=j_k+1}^{j_{k+1}} r_i \Big| X_{j_k+1}^{j_{k+1}}, |\cdot|^2\Big) \leq (1+\varepsilon)\, D(r|X, |\cdot|^2).$$

The remaining first $k_0 - 1$ summands are of lower order (as $r \to \infty$) and one retrieves the result for the slightly modified process.

4.4 Generalizations to other coding quantities (2nd approach)

A second approach in the analysis of the asymptotic coding errors has been undertaken in Dereich (2003*a*). The results can be stated as follows:

Theorem 4.4. *If the eigenvalues satisfy*

$$\lim_{n\to\infty} \frac{\log\log 1/\lambda_n}{n} = 0,$$

then for any $s > 0$,

$$D^{(q)}(r,s) \sim D(r,s) \sim D(r,2).$$

The analysis of this result is more elaborate and we only want to give the very basic ideas of the proof of the upper bound of the quantization error. A central role is played by an asymptotic equipartition property.

The underlying asymptotic equipartition property

For a fixed rate $r \geq 0$, the solution to the rate allocation problem is linked to a unique parameter $\kappa \in (0, \lambda_1]$ (as explained above). Now set $\tilde{\lambda}_i = (\lambda_i - \kappa) \wedge 0$ and denote by $\tilde{X}^{(r)} = (\tilde{X}_i^{(r)})_{i\in\mathbb{N}}$ an l^2-valued random variable having as entries independent $\mathcal{N}(0, \tilde{\lambda}_i)$-distributed r.v. $\tilde{X}_i^{(r)}$ (the variances $\tilde{\lambda}_i$ $(i \in \mathbb{N})$ are visualized as striped boxes in Figure 1). The asymptotic equipartition property states that for $\varepsilon > 0$ fixed and r going to infinity, the probability that

$$-\log \mathbb{P}\big(\|X - \tilde{X}^{(r)}\|^2 \leq (1+\varepsilon) D(r) \big| X\big) \leq r + \frac{D(r)}{-D'(r)}\varepsilon$$

tends to one. Here $D'(\cdot)$ denotes the right hand side derivative of the convex function $D(\cdot)$.

The proof of the equipartition property relies on a detailed analysis of the random logarithmic moment generating function

$$\begin{aligned}\Lambda_X(\theta) &:= \log \mathbb{E}\big[e^{\theta\|X-\tilde{X}^{(r)}\|^2}\big|X\big]\\ &= \sum_{i\in\mathbb{N}} \Big[-\frac{1}{2}\log(1-2\theta\tilde{\lambda}_i) + \frac{\theta\lambda_i}{1-2\theta\tilde{\lambda}_i}\,\xi_i^2\Big]\end{aligned}$$

(where (ξ_i) is the sequence of independent standard normals from representation (17)), and the close relationship between $-\log \mathbb{P}(\|X-\tilde{X}^{(r)}\|^2 \le \zeta|X)$ and the random Legendre transform

$$\Lambda_X^*(\zeta) := \sup_{\theta\le 0}[\theta\zeta - \Lambda_X(\theta)]$$

given in Dereich (2003*a*, Theorem 3.4.1).

The asymptotic equipartition property implies that the random codebooks $\mathcal{C}(r)$ $(r \ge 0)$ consisting of $\lfloor \exp(r + 2\frac{D(r)}{-D'(r)}\varepsilon)\rfloor$ independent random copies of $\tilde{X}^{(r)}$ satisfy

$$\lim_{r\to\infty} \mathbb{P}\big(\min_{\hat{x}\in\mathcal{C}(r)} \|X-\hat{x}\|^2 \le (1+\varepsilon)D(r)\big) = 1.$$

Moreover, $D\big(r + 2\frac{D(r)}{-D'(r)}\varepsilon\big) \ge (1-2\varepsilon)D(r)$ (due to the convexity of $D(\cdot)$) so that typically the random approximation error satisfies

$$\min_{\hat{x}\in\mathcal{C}(r)} \|X-\hat{x}\|^2 \le \frac{1+\varepsilon}{1-2\varepsilon}\, D\big(\log|\mathcal{C}(r)|\big) \tag{25}$$

for large r and fixed $\varepsilon \in (0, 1/2)$. It remains to control the error inferred in the case when estimate (25) is not valid.

4.5 A particular random quantization procedure

As before we want to compare the coding results quoted so far with the efficiency of the particular random quantization strategy introduced above. Let $(Y_i)_{i\in\mathbb{N}}$ denote a sequence of independent random vectors with the same law as X. We consider the quantization error inferred by the random codebooks $\mathcal{C}(r) = \{Y_1, \dots, Y_{\lfloor e^r\rfloor}\}$. Let

$$D^{(r)}(r,s) = \mathbb{E}[\min_{i=1,\dots,\lfloor e^r\rfloor} \|X - Y_i\|^s]^{1/s}.$$

As mentioned above the asymptotics of $D^{(r)}(\cdot, s)$ are related to a randomly centered small ball function. Let us again denote $\ell_\varepsilon(x) = -\log \mathbb{P}(\|X-x\| \le \varepsilon)$ $(x \in H, \varepsilon > 0)$. We quote the main result of Dereich (2003*b*).

Theorem 4.5. *One has*

$$\lim_{\varepsilon\downarrow 0} \frac{\ell_\varepsilon(X)}{\varphi_*(\varepsilon)} = 1, \ a.s., \tag{26}$$

where $\varphi_*(\varepsilon) = \Lambda^*(\varepsilon^2) = \sup_{\theta\in\mathbb{R}}[\varepsilon^2\theta - \Lambda(\theta)]$ *is the Legendre transform of*

$$\Lambda(\theta) = \sum_i \Big[-\frac{1}{2}\log(1-2\theta\lambda_i) + \frac{\theta\lambda_i}{1-2\theta\lambda_i}\Big]$$

and $\log(z) = -\infty$ *for* $z \le 0$.

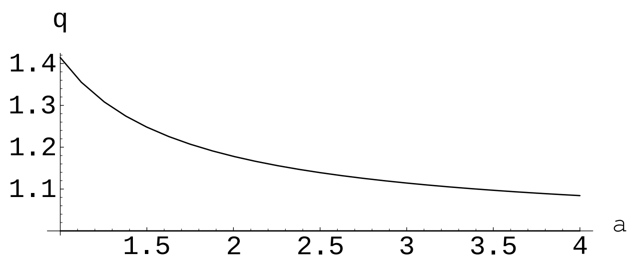

Figure 2: Comparison of $D^{(r)}(\cdot, s)$ and $D^{(q)}(\cdot, s)$ in dependence on α.

Note that the theorem does not assume any assumptions on the eigenvalues. Given that the eigenvalues are regularly varying it is possible to directly relate the function $\varphi_*(\varepsilon)$ to the standard small ball function $\varphi(\varepsilon) = \ell_\varepsilon(0)$ $(\varepsilon > 0)$:

Theorem 4.6. *If the eigenvalues are regularly varying with index $-\alpha < -1$ (in the sense described above), one has*

$$\lim_{\varepsilon \downarrow 0} \frac{\varphi_*(\varepsilon)}{\varphi(\varepsilon)} = \Big(\frac{\alpha+1}{\alpha}\Big)^{\frac{\alpha}{\alpha-1}}.$$

Let us now compare the error induced by the random coding strategy with the optimal quantization error in the case where the eigenvalues satisfy

$$\lambda_i \sim c\, i^{-\alpha}$$

for two constants $c > 0$ and $\alpha > 1$. Starting from the standard small ball function (see Sytaya (1974) or Lifshits (1997) for a general treatment) one can deduce the asymptotics of the random quantization error $D^{(r)}(\cdot, s)$. A comparison with $D^{(q)}(\cdot, s)$ then gives

$$\lim_{r\to\infty} \frac{D^{(r)}(r,s)}{D^{(q)}(r,s)} = \Big[\frac{(\alpha^2-1)\pi}{\alpha^3 \sin(\frac{\pi}{\alpha})}\Big]^{\alpha/2}.$$

The limiting value on the right hand side is plotted in Figure 2.

4.6 Examples

- The most prominent example is $X = (X_t)_{t\in[0,1]}$ being a Wiener process in $L^2[0,1]$. In that case the Karhunen-Loève expansion is known explicitly: the eigenvalues are given by $\lambda_i = (\pi(i-1/2))^{-2}$ and the corresponding normalized eigenfunctions are

  $$e_i(t) = \sqrt{2}\sin(\pi(n-1/2)t) \qquad (t \in [0,1]).$$

  Thus the eigenvalues are regularly varying with index -2 and one gets

  $$\lim_{r\to\infty} \sqrt{r}\, D(r,2) = \frac{\sqrt{2}}{\pi}.$$

- Next, let $X = (X_t)_{t\in[0,1]}$ be a γ-fractional Wiener process, that is the continuous Gaussian process with

  $$\mathbb{E}[X_t X_u] = \frac{1}{2}[|t|^\gamma + |u|^\gamma - |t-u|^\gamma] \qquad (t, u \in [0,1]).$$

As underlying Hilbert space we consider again $L^2[0,1]$.

In that case the Karhunen-Loéve expansion is not known explicitly and there exist only suboptimal representations (Dzhaparidze & van Zanten 2004, Iglói 2005). However, the asymptotic behavior of the ordered sequence of eigenvalues is known (Bronski 2003): one has

$$\lambda_i \sim \frac{\sin(\pi\gamma/2)\,\Gamma(\gamma+1)}{(i\pi)^{\gamma+1}},$$

where Γ is the Euler gamma function. Hence,

$$\lim_{r\to\infty} r^{\gamma/2}\, D(r,2) = \sqrt{\frac{(\gamma+1)^{\gamma+1}\sin(\pi\gamma/2)\,\Gamma(\gamma+1)}{2^\gamma\,\gamma\,\pi^{\gamma+1}}}.$$

as $r\to\infty$.

- Next, let $X=(X_t)_{t\in[0,1]^d}$ denote a d-dimensional fractional Brownian sheet with parameter $\gamma=(\gamma_1,\dots,\gamma_1)$. Then the sequence of ordered eigenvalues satisfies

$$\lambda_i \sim \Big(\frac{\sin(\pi\gamma_1/2)\Gamma(\gamma_1+1)}{\pi^{\gamma_1+1}}\Big)^d ((d-1)!)^{-(\gamma_1+1)} \Big(\frac{(\log i)^{d-1}}{i}\Big)^{\gamma_1+1}$$

so that

$$\lim_{r\to\infty} \frac{r^{\gamma_1/2}}{(\log r)^{(d-1)(\gamma_1+1)/2}}\, D(r,2) = \sqrt{\frac{(\gamma_1+1)^{\gamma_1+1}}{2^{\gamma_1}\gamma_1((d-1)!)^{\gamma_1+1}}\Big(\frac{\sin(\pi\gamma_1/2)\Gamma(\gamma_1+1)}{\pi^{\gamma_1+1}}\Big)^d}.$$

5 Diffusions

The quantization complexity of diffusion processes was firstly treated in Luschgy & Pagès (2006). There weak asymptotic estimates for the quantization problem were derived for a class of 1-dimensional diffusions.

In the following we want to focus on one approach that has been developed in two articles by the author (Dereich 2006*b*, 2006*a*). It leads to asymptotic formulae for several coding quantities. Moreover, it is based on a rate allocation problem and we believe that it fits best into the context of this article.

We consider as original signal an $\mathbb{R}$-valued process $X=(X_t)_{t\in[0,1]}$ that solves the integral equation

$$X_t = x_0 + \int_0^t \sigma(X_u,u)\,dW_u + \int_0^t b(X_u,u)\,du, \tag{27}$$

where $W=(W_t)_{t\in[0,1]}$ denotes a standard Wiener process, $x_0\in\mathbb{R}$ denotes an arbitrary starting point and $\sigma,b:\mathbb{R}\times[0,1]\to\mathbb{R}$ are continuous functions.

We impose the following regularity assumptions on the functions σ and b: there exist constants $\beta\in(0,1]$ and $L<\infty$ such that for $x,x'\in\mathbb{R}$ and $t,t'\in[0,1]$:

$$\begin{aligned} |b(x,t)| &\le L(|x|+1) \quad\text{and}\\ |\sigma(x,t)-\sigma(x',t')| &\le L[|x-x'|^\beta + |x-x'| + |t-t'|^\beta]. \end{aligned} \tag{28}$$

Note that the assumptions do neither imply existence nor uniqueness of the solution. However, the analysis does not rely on the uniqueness of the solution and we only need to assume the existence of one solution which shall be fixed for the rest of this section.

To simplify notation, we will denote by $(\sigma_t)_{t\in[0,1]}$ and $(b_t)_{t\in[0,1]}$ the stochastic processes $(\sigma(X_t,t))$ and $(b(X_t,t))$, respectively. Let us first state the main results:

Theorem 5.1 (Dereich (2006*b*)). *If $E = C[0,1]$, then for each $s > 0$ one has*

$$\lim_{r\to\infty} \sqrt{r}\, D^{(q)}(r,s) = \kappa_\infty \big\| \|\sigma_\cdot\|_{L^2[0,1]} \big\|_{L^s(\mathbb{P})}$$

and

$$\lim_{r\to\infty} \sqrt{r}\, D^{(e)}(r,s) = \kappa_\infty \big\| \|\sigma_\cdot\|_{L^2[0,1]} \big\|_{L^{2s/(s+2)}(\mathbb{P})},$$

where κ_∞ is the real constant appearing in Theorem 3.5.

Theorem 5.2 (Dereich (2006*a*)). *If $E = L^p[0,1]$ for $p \ge 1$, then for every $s > 0$ one has*

$$\lim_{r\to\infty} \sqrt{r}\, D^{(q)}(r,s) = \kappa_p \big\| \|\sigma_\cdot\|_{L^{2p/(2+p)}[0,1]} \big\|_{L^s(\mathbb{P})}$$

and

$$\lim_{r\to\infty} \sqrt{r}\, D(r,s) = \lim_{r\to\infty} \sqrt{r}\, D^{(e)}(r,s) = \kappa_p \big\| \|\sigma_\cdot\|_{L^{2p/(2+p)}[0,1]} \big\|_{L^{2s/(s+2)}(\mathbb{P})},$$

where κ_p is the constant from Theorem 3.5.

The analysis of the asymptotic coding problem is based on a decoupling argument. The decoupling argument allows us to connect the complexity of the diffusion to that of the Wiener process. After one has applied the decoupling techniques one can prove the asymptotic formulae by considering certain rate allocation problems. In the next section we give a heuristic explanation of the decoupling method. Then we use these results to derive heuristically the lower bound. The solution of the corresponding rate allocation problem can also be used to define asymptotically optimal quantizers or entropy coding schemes.

5.1 The decoupling method

Let us start with an intuitive explanation of the decoupling method. We fix $s \ge 1$ and let the underlying norm $\|\cdot\|$ be supremum norm over the interval $[0,1]$.

The main idea is to write the diffusion as the sum of a decoupled process and a remaining term of negligible complexity. Let us first explain what we understand under negligible terms.

Negligible terms

The approximation error inferred of either complexity constraint leads to an asymptotic error of order $r^{-1/2}$, where r is and will be used to indicate the rate of the approximation. In order to simplify the folowing discussion, we assume that changing the rate r by a term of order $o(r)$ does not have an influence on the strong asymptotics of the coding quantity of interest. This property is valid for the diffusion under either information constraint as one can easily infer from Theorems 5.1 and 5.2. However, the proof of this fact is non-trivial.

Suppose now that $(Y^{(r)})_{r\ge 0}$ is a family of processes such that there exist discrete approximations $\hat{A}^{(r)}$ to $A^{(r)} := X - Y^{(r)}$ satisfying

$$\log|\operatorname{range} \hat{A}^{(r)}| = o(r) \quad \text{and} \quad \mathbb{E}[\|A^{(r)} - \hat{A}^{(r)}\|^s]^{1/s} = o(r^{-1/2}).$$

Then one can replace the process X by $Y^{(r)}$ (actually by the family $(Y^{(r)})_{r\geq 0}$) without changing the strong asymptotics of either of the approximation errors. Indeed, one can relate approximations $\hat{Y}^{(r)}$ of $Y^{(r)}$ to approximations $\hat{X}^{(r)}$ of X via $\hat{X}^{(r)} = \hat{Y}^{(r)} + \hat{A}^{(r)}$ and gets

$$\big|\mathbb{E}[\|X - \hat{X}^{(r)}\|^s]^{1/s} - \mathbb{E}[\|Y^{(r)} - \hat{Y}^{(r)}\|^s]^{1/s}\big| = o(r^{-1/2}). \tag{29}$$

Moreover, reconstructions $\hat{X}^{(r)}$ of rate r give rise to reconstructions $\hat{Y}^{(r)}$ of rate $(1 + o(1))r$ and vice versa (see the additivity properties listed on page 4). Here it does not matter which complexity constraint we use. Moreover, the supremum norm is the strongest norm under consideration and (29) remains valid for the same family $(Y^{(r)})_{r\geq 0}$ when $\|\cdot\|$ denotes an L^p-norm. We call the term $A^{(r)}$ asymptotically negligible.

Relating X to a decoupled time change of a Wiener process

We represent X in its Doob-Meyer decomposition as a sum of a martingale

$$M_t = \int_0^t \underbrace{\sigma(X_u, u)}_{\sigma_u} \, dW_u$$

and a drift term

$$A_t = x_0 + \int_0^t \underbrace{b(X_u, u)}_{b_u} \, du.$$

The drift term is much more regular than the martingale term and one can verify that it is asymptotically negligible: we only need to consider the martingale part. It can be viewed as a time change of a Wiener process:

$$(M_t) = (\tilde{W}_{\varphi(t)}),$$

where $(\tilde{W}_t)_{t\geq 0}$ denotes a Wiener process and $(\varphi(t))_{t\in[0,1]}$ is given by $\varphi(t) = \int_0^t \sigma_u^2 \, du$. Unfortunately, the time change and the Wiener process are not independent so that one cannot immediately apply the asymptotic formulae for the Wiener process. On the other hand, conditional upon $(\varphi(t))$ the process $(\tilde{W}_t)_{t\geq 0}$ is no longer a Wiener process and the process might even be deterministic up to time $\varphi(1)$.

In order to bypass this problem we introduce an approximation $\hat{\varphi} = (\hat{\varphi}(t))_{t\in[0,1]}$ (depending on the rate r) to φ such that the approximation error

$$\mathbb{E}[\|M_{\cdot} - \tilde{W}_{\hat{\varphi}(\cdot)}\|^s]^{1/s} = o(r^{-1/2})$$

is negligible compared to the total coding error (which is of order $r^{-1/2}$). On the other hand, conditional upon $\hat{\varphi}$ the process $(\tilde{W}_t)_{t\geq 0}$ should not significantly deviate from a Wiener process. For controlling the influence of the conditioning, we view the additional information induced by the approximation $\hat{\varphi}$ as an initial enlargement of the canonical filtration $(\mathcal{F}_t^{\tilde{W}})$ induced by the process $\tilde{W}$. Let us be more precise about the estimates used here.

For a weakly differentiable function $f : [0, \infty) \to \mathbb{R}$ with differential $\dot{f}$ we set

$$\|f\|_{\mathcal{H}} = \Big(\int_0^\infty |\dot{f}(u)|^2 \, du\Big)^{1/2}.$$

The Wiener process $\tilde{W}$ is now represented in its $(\mathcal{G}_t) = (\mathcal{F}_t^{\tilde{W}} \vee \sigma(\hat{\varphi}))$-Doob-Meyer decomposition as

$$\tilde{W}_t = \bar{W}_t + \bar{A}_t.$$

Again recall that $\hat{\varphi}$ and thus also $\bar{W}$ and $\bar{A}$ depend on r. Now $(\bar{W}_t)$ is a $(\mathcal{G}_t)$-Wiener process that is independent of $(\hat{\varphi}(t))$. Supposing that the process $(\hat{\varphi}(t))$ has finite range, one can conclude that $(\tilde{W}_t)$ is indeed a $(\mathcal{G}_t)$-semi martingale which justifies the Doob-Meyer decomposition above. Moreover, there exists a constant c depending on $s \geq 1$ only, such that

$$\mathbb{E}[\|\bar{A}\|_{\mathcal{H}}^{2s}]^{1/s} \leq c\,(1 + \log|\operatorname{range}(\hat{\varphi})|)$$

(see Jeulin & Yor (1985), Ankirchner, Dereich & Imkeller (2004)). On the other hand, the Sobolev space $\mathcal{H}$ (the space of weakly differentiable functions with finite $\|\cdot\|_{\mathcal{H}}$-norm) is compactly embedded into $C[0,T]$ for a finite time horizon $T > 0$ and one can use analytic results on this embedding to show the asymptotic negligibility of the term $(\bar{A}_{\hat{\varphi}(t)})_{t\in[0,1]}$. Altogether, one sees that the coding complexity of X and $(\bar{W}_{\hat{\varphi}(t)})_{t\in[0,1]}$ coincide. Here one needs to be aware of the fact that the definition of $(\hat{\varphi}(t))$ and thus of $(\bar{W}_t)$ depend on the parameter r.

The main result

Let us state the precise decoupling result. Fix $\alpha \in (0, \beta/2)$ and denote by $\hat{\varphi}^{(n)} = (\hat{\varphi}_t^{(n)})_{t\in[0,1]}$ ($n \in \mathbb{N}$) a random increasing and continuous function that is linear on each interval $[i/n, i+1/n]$ ($i = 0, \ldots, n-1$) and satisfies

$$\hat{\varphi}^{(n)}(i/n) = \arg\min_{y\in I(n)} |\varphi(i/n) - y| \qquad (i = 0, \ldots, n),$$

where $I(n)$ is defined as

$$I(n) = \Big\{ j\frac{1}{n^{1+\alpha}} : j \in \mathbb{N}_0, j \leq n^{2(1+\alpha)} \Big\}.$$

Theorem 5.3. *Fix $\zeta \in ((1+\alpha)^{-1}, 1)$, choose $n \in \mathbb{N}$ in dependence on $r > 0$ as $n = n(r) = \lceil r^{\zeta} \rceil$, and denote by $\tilde{W} = \bar{W}^{(n)} + \bar{Y}^{(n)}$ the $(\mathcal{F}_t^{\tilde{W}} \vee \sigma(\hat{\varphi}^{(n)}))$-Doob-Meyer decomposition of $\tilde{W}$. For arbitrarily fixed $s > 0$ there exist $C[0,1]$-valued r.v.'s $\bar{R}^{(n)}$ and $\hat{R}^{(r)}$ such that*

- $X = \bar{W}^{(n)}_{\hat{\varphi}^{(n)}(\cdot)} + \bar{R}^{(n)}$,
- *$\bar{W}^{(n)}$ is a Wiener process that is independent of $\hat{\varphi}^{(n)}$,*
- $\mathbb{E}[\|\bar{R}^{(n)} - \hat{R}^{(r)}\|^s]^{1/s} = \mathcal{O}(r^{-\frac{1}{2}-\delta})$, *for some $\delta > 0$,*
- $\log|\operatorname{range}(\hat{R}^{(r)}, \hat{\varphi}^{(n)})| = \mathcal{O}(r^{\gamma})$, *for some $\gamma \in (0,1)$.*

5.2 The corresponding rate allocation problem

We only consider the case where where $E = L^p[0,1]$ for some $p \geq 1$. Theorem 5.3 states that the process X and the process $(\bar{W}^{(n)}_{\hat{\varphi}^{(n)}(t)})_{t\in[0,1]}$ (actually the family of processes) have the same asymptotic complexity. The number n is still related to r via $\lceil r^{\zeta} \rceil$.

We adopt the notation of Theorem 5.3, and let for $i \in \{0, \dots, n-1\}$ and $t \in [0, 1/n)$

$$\bar{X}^{(n)}_{\frac{i}{n}+t} := \bar{X}^{(n,i)}_t := \bar{W}^{(n)}_{\hat{\varphi}^{(n)}(\frac{i}{n}+t)} - \bar{W}^{(n)}_{\hat{\varphi}^{(n)}(\frac{i}{n})}.$$

Note that r/n tends to infinity and one can verify that the difference between $\bar{X}^{(n)} = (\bar{X}^{(n)}_t)_{t\in[0,1)}$ and $\bar{W}^{(n)}_{\hat{\varphi}^{(n)}(\cdot)}$ is negligible in the sense explained above.

Conditional on $\hat{\varphi}^{(n)}$ the process $\bar{X}^{(n)}$ is a concatenation of n independent scaled Wiener processes $\bar{X}^{(n,0)}, \dots, \bar{X}^{(n,n-1)}$ in the sense that each of the components equals $(\hat{\sigma}_i W_t)_{t\in[0,1/n)}$ in law, where $\hat{\sigma}_i \geq 0$ is given via

$$\hat{\sigma}_i^2 := n(\hat{\varphi}(i/n) - \hat{\varphi}((i-1)/n)).$$

Concatenations of Wiener processes

First we suppose that $\hat{\varphi}^{(n)}$ is deterministic and that $s = p$. The *letters* $\bar{X}^{(n,0)}, \dots, \bar{X}^{(n,n-1)}$ are independent and the objective function

$$\mathbb{E}[\|\bar{X}^{(n)} - \hat{X}^{(r)}\|^p] = \mathbb{E}\Big[\int_0^1 |\bar{X}_t|^p\, dt\Big]$$

can be understood as single letter distortion measure. Thus the discussion of Section 4.2 yields that for an arbitrary rate $\bar{r} \geq 0$ the DRF $D(\bar{r}|\bar{X}^{(n)}, \|\cdot\|^p)$ is naturally related to a rate allocation problem. One has

$$D(\bar{r}|\bar{X}^{(n)}, \|\cdot\|^p) = \inf_{(r_i)} \sum_{i=0}^{n-1} \underbrace{D(r_i|\hat{\sigma}_i W, \|\cdot\|^p_{L^p[0,1/n)})}_{=\hat{\sigma}_i^p D(r_i|W, \|\cdot\|_{L^p[0,1/n)}, p)^p}, \tag{30}$$

where the infimum is taken over all non-negative vectors $(r_i)_{i=0,\dots,n-1}$ with $\sum_{i=0}^{n-1} r_i = \bar{r}$. Moreover, the map

$$\pi : L^p[0, 1/n) \to L^p[0, 1),\ (x_t)_{t\in[0,1/n)} \mapsto (n^{-1/p} x_{t/n})_{t\in[0,1)}$$

is an isometric isomorphism so that

$$D(r_i|W, \|\cdot\|_{L^p[0,1/n)}, p) = D(r_i|n^{-\frac{1}{p}-\frac{1}{2}}\sqrt{n} W_{t/n}, \|\cdot\|_{L^p[0,1)}, p) = n^{-\frac{1}{p}-\frac{1}{2}} \underbrace{D(r_i|W, \|\cdot\|_{L^p[0,1)}, p)}_{\sim \kappa_p r_i^{-1/2}}.$$

Supposing that the rates r_i ($i = 0, \dots, n-1$) are large so that $\kappa_p r_i^{-1/2}$ is a reasonable approximation to the latter DRF, one concludes together with (30) that $D(\bar{r}|\bar{X}^{(n)}, \|\cdot\|, p)$ is approximately equal to

$$\kappa_p \inf_{(r_i)} \Big(\frac{1}{n} \sum_{i=0}^{n-1} \frac{\hat{\sigma}_i^p}{(n r_i)^{p/2}}\Big)^{1/p}. \tag{31}$$

The infimum can be evaluated explicitly by applying the reverse Hölder inequality analogously to the computations in (9). It is equal to

$$\Big(\frac{1}{n} \sum_{i=0}^{n-1} |\hat{\sigma}_i|^{2p/(p+2)}\Big)^{(p+2)/2p} \frac{1}{\sqrt{\bar{r}}}. \tag{32}$$

Next, let $\hat\sigma_t := \hat\sigma_i$ for $i = 0,\dots,n-1$ and $t \in [i/n,(i+1)/n)$ and observe that (32) can be rewritten as

$$\Big(\int_0^1 |\hat\sigma_t|^{2p/(p+2)}\,dt\Big)^{(p+2)/2p}\frac{1}{\sqrt{\bar r}}. \tag{33}$$

Rigorously one can prove that one infers (asymptotically) the same error for the other information constraints and for all other moments s. We quote the exact result:

Lemma 5.4. *For fixed $s \in (0,\infty)$ there exists a real valued function $h = h_{s.p} : \mathbb{R}_+ \to \mathbb{R}_+$ with $\lim_{\bar r\to\infty} h(\bar r) = 1$ such that the following statements are valid.*

Suppose that $\hat\varphi^{(n)}$ is deterministic and let $\hat X$ be a reconstruction for $\bar X^{(n)}$ of rate $\bar r > 0$. Then there exists an $[0,\infty)$-valued sequence $(r_i)_{i=0,\dots,n-1}$ with $\sum_{i=0}^{n-1} r_i \le \bar r$ such that for any $r_* > 0$:

$$\mathbb{E}[\|\bar X^{(n)} - \hat X\|^s]^{1/s} \ge h(r_*)\,\kappa_p\Big(\frac{1}{n}\sum_{i=0}^{n-1}\frac{|\hat\sigma_i|^p}{(n(r_i+r_*))^{p/2}}\Big)^{1/p}.$$

On the other hand, for any $\mathbb{R}_+$-valued vector $(r_i)_{i=0,\dots,n-1}$, there exists a codebook $\mathcal{C} \subset L^p[0,1]$ with $\log|\mathcal{C}| \le \sum_{i=0}^{n-1} r_i$ and

$$\mathbb{E}[\min_{\hat x\in\mathcal{C}}\|\bar X^{(n)} - \hat x\|^s]^{1/s} \le h(r^*)\,\kappa_p\Big(\frac{1}{n}\sum_{i=0}^{n-1}\frac{|\hat\sigma_i|^p}{(nr_i)^{p/2}}\Big)^{1/p},$$

where $r^* = \min_{i=0,\dots,n-1} r_i$.

The lower bound

In the discussion above we have assumed that the time change $\hat\varphi^{(n)}$ is deterministic. Now we switch back to the original problem and suppose that $\hat\varphi^{(n)}$ and $\bar X^{(n)}$ ($n \in \mathbb{N}$) are as introduced in the beginning of this subsection. In particular, $\hat\varphi^{(n)}$ is again assumed to be random. Again the parameter n is linked to $r > 0$ via $n = \lceil r^\zeta\rceil$.

We fix a family $(\hat X^{(r)})_{r\ge 0}$ of reconstructions such that each reconstruction has finite mutual information $I(\bar X^{(n)};\hat X^{(r)})$ to be finite, and we set $r(\phi) = I(\bar X^{(n)};\hat X^{(r)}|\varphi^{(n)} = \phi)$ and $R = r(\hat\varphi^{(n)})$. The random variable R is to be conceived as the random rate reserved for coding $\bar X^{(n)}$ given the time change $\hat\varphi^{(n)}$.

Using the (non-rigorous) lower bound (33) for the approximation error, one gets

$$\mathbb{E}[\|\bar X^{(n)} - \hat X^{(r)}\|^s]^{1/s} = \mathbb{E}\big[\mathbb{E}[\|\bar X^{(n)} - \hat X^{(r)}\|^s|\hat\varphi^{(n)}]\big]^{1/s} \gtrsim \mathbb{E}\Big[\Big(\int_0^1 |\hat\sigma_t|^{2p/(p+2)}\,dt\Big)^{s(p+2)/2p}\frac{1}{R^{s/2}}\Big]^{1/s}.$$

It remains to translate a given information constraint into a condition onto the random rate R.

The quantization constraint

Let us now assume that the family $(\hat X^{(r)})$ satisfies the quantization constraint $\log|\operatorname{range}\hat X^{(r)}| \le r$. Then $R \le r$ almost surely so that one gets the lower bound

$$\mathbb{E}[\|\bar X^{(n)} - \hat X^{(r)}\|^s]^{1/s} \gtrsim \mathbb{E}\Big[\Big(\int_0^1 |\hat\sigma_t|^{2p/(p+2)}\,dt\Big)^{s(p+2)/2p}\Big]^{1/s}\frac{1}{\sqrt{r}}.$$

Since $\int_0^1 |\hat\sigma_t|^{2p/(p+2)}\,dt$ converges to $\int_0^1 |\sigma_t|^{2p/(p+2)}\,dt$ one finally obtains

$$\mathbb{E}[\|\bar X^{(n)} - \hat X^{(r)}\|^s]^{1/s} \gtrsim \big\|\|\sigma_\cdot\|_{L^{2p/(p+2)}[0,1]}\big\|_{L^s(\mathbb{P})}\frac{1}{\sqrt{r}}.$$

The mutual information constraint

Now we assume that $I(\bar X;\hat X^{(r)}) \le r$. Then

$$I(X;\hat X|\hat\varphi) \le I(X;\hat X,\hat\varphi^{(n)}) \le I(\bar X;\hat X^{(r)}) + \log|\operatorname{range}\hat\varphi^{(n)}| \lesssim r.$$

Therefore,

$$\mathbb{E}R \lesssim r. \tag{34}$$

and one gets an asymptotic lower bound for the distortion rate function when minimizing

$$\mathbb{E}\Big[\Big(\int_0^1 |\hat\sigma_t|^{2p/(p+2)}\,dt\Big)^{s(p+2)/2p}\frac{1}{R^{s/2}}\Big]^{1/s}$$

under the constraint (34). Also in this rate allocation problem the Hölder inequality for negative exponents solves the problem and one gets:

$$\mathbb{E}[\|\bar X^{(n)} - \hat X^{(r)}\|^s]^{1/s} \gtrsim \big\|\|\hat\sigma_\cdot\|_{L^{2p/(p+2)}[0,1]}\big\|_{L^{2s/(s+2)}(\mathbb{P})}\frac{1}{\sqrt{\mathbb{E}[R]}} \gtrsim \big\|\|\sigma_\cdot\|_{L^{2p/(p+2)}[0,1]}\big\|_{L^{2s/(s+2)}(\mathbb{P})}\frac{1}{\sqrt{r}}.$$

6 Further results on asymptotic approximation of stochastic processes

As mentioned before quantization of stochastic processes has been a vivid research area within the last years. So far we have restricted ourselves to surveying results that lead to strong asymptotic formulae and that are related to intermediate optimization problems. In this section we want to complete the article with giving two further results which yield the correct weak asymptotics in many cases.

6.1 Estimates based on moment conditions on the increments

Let us next describe a very general approach undertaken in Luschgy & Pagès (2007) to provide weak upper bounds for the quantization error. It is based on moment conditions on the increments of a stochastic process. Let $(X_t)_{t\in[0,1]}$ denote a real-valued stochastic process, that is $X:\Omega\times[0,1]\to\mathbb{R}$ is product measurable. We fix $\zeta>0$ and assume that the marginals of X are in $L^\zeta(\mathbb{P})$.

We state the main result.

Theorem 6.1. *Let $\varphi:[0,1]\to[0,\infty)$ be a regularly varying function at 0 with index $b>0$, that is φ can be represented as $\varphi(t)=t^b\ell(t)$, where ℓ is continuous and satisfies $\ell(\alpha t)\sim\ell(t)$ as $t\downarrow 0$ for any $\alpha>0$ (ℓ is called slowly varying). Moreover, we assume that for all $0\le u\le t\le 1$*

$$\mathbb{E}[|X_t - X_u|^\zeta]^{1/\zeta} \le \varphi(t-u),$$

in the case where $\zeta \geq 1$*, and*

$$\mathbb{E}[\sup_{v\in[u,t]} |X_v - X_s|^\zeta]^{1/\zeta} \leq \varphi(t-u),$$

otherwise. Then for $p, s \in (0,\zeta)$*, one has*

$$D(r, s|X, \|\cdot\|_{L^p[0,1]}) \precsim \varphi(1/r).$$

This theorem allows to translate moment estimates for the increments into upper bounds for the quantization error. Let us demonstrate the power of this result by applying it to a class of diffusion processes.

Let $X = (X_t)_{t\in[0,1]}$ satisfy

$$X_t = x_0 + \int_0^t G_u\, du + \int_0^t H_u\, dW_u,$$

where $W = (W_t)_{t\in[0,1]}$ denotes a Wiener process, and $G = (G_t)_{t\in[0,1]}$ and $H = (H_t)_{t\in[0,1]}$ are assumed to be progressively measurable processes w.r.t. the canonical filtration induced by W. Supposing that for some $\zeta \geq 2$

$$\sup_{t\in[0,1]} \mathbb{E}[|G_t|^\zeta + |H_t|^\zeta] < \infty,$$

one can infer that for a constant $c \in \mathbb{R}_+$ one has

$$\|X_t - X_u\|_{L^\zeta(\mathbb{P})} \leq c\,(t-u)^{1/2}$$

for all $0 \leq u \leq t \leq 1$. Consequently,

$$D(r, s|X, \|\cdot\|_{L^p[0,1]}) \precsim \frac{1}{\sqrt{r}}.$$

As we have seen above this rate is optimal for the solutions to the autonomous stochastic differential equations studied above.

Moreover, the main result can be used to infer upper bounds for stationary processes and Lévy processes or to recover the weak asymptotics for fractional Wiener processes. For further details we refer the reader to the article by Luschgy & Pagès (2007).

6.2 Approximation of Lévy processes

Let now $X = (X_t)_{t\in[0,1]}$ denote a càdlàg Lévy process. Due to the Lévy-Khintchine formula the marginals X_t ($t \in [0,1]$) admit a representation

$$\mathbb{E}e^{iuX_t} = e^{-t\psi(u)} \qquad (u \in \mathbb{R}),$$

where

$$\psi(u) = \frac{\sigma^2}{2}u^2 - i\alpha u + \int_{\mathbb{R}\backslash\{0\}} (1 - e^{iux} + \mathbb{1}_{\{|x|\leq 1\}} iux)\, \nu(dx),$$

for parameters $\sigma^2 \in [0,\infty)$, $\alpha \in \mathbb{R}$ and a non-negative measure ν on $\mathbb{R}\backslash\{0\}$ with

$$\int_{\mathbb{R}\backslash\{0\}} 1 \wedge x^2 \, \nu(dx) < \infty.$$

The complexity of Lévy processes has been analyzed recently in Aurzada & Dereich (2007) when the underlying distortion measure is induced by the $L^p[0,1]$-norm for a fixed $p \geq 1$. Its complexity is related to the function

$$F(\varepsilon) = \frac{\sigma^2}{\varepsilon^2} + \int_{\mathbb{R}\backslash\{0\}} \Big[\Big(\frac{x^2}{\varepsilon^2} \wedge 1\Big) + \log_+ \frac{|x|}{\varepsilon}\Big] \, \nu(dx) \qquad (\varepsilon > 0).$$

Let us state the main results.

Theorem 6.2 (Upper bound)**.** *There exist positive constants $c_1 = c_1(p)$ and c_2 such that for any $\varepsilon > 0$ and $s > 0$*

$$D^{(e)}(c_1 F(\varepsilon), s) \leq c_2 \varepsilon.$$

The constants c_1 and c_2 can be chosen independently of the choice of the Lévy process. Additionally, if for $s > 0$ one has

- $\mathbb{E}\|X\|^{s'}_{L^p[0,1]} < \infty$ *for some $s' > s$, and*
- *for some $\zeta > 0$,*

$$\limsup_{\varepsilon \downarrow 0} \frac{\int_{|x|>\varepsilon} (|x|/\varepsilon)^\zeta \, \nu(dx)}{\nu([-\varepsilon,\varepsilon]^c)} < \infty,$$

then there exist constants c_1' and c_2' such that for all $\varepsilon > 0$,

$$D^{(q)}(c_1' F(\varepsilon), s) \leq c_2' \varepsilon.$$

For stating the lower bound we will need the function

$$F_1(\varepsilon) = \frac{\sigma^2}{\varepsilon^2} + \int_{\mathbb{R}\backslash\{0\}} \Big(\frac{x^2}{\varepsilon^2} \wedge 1\Big) \, \nu(dx) \qquad (\varepsilon > 0).$$

Theorem 6.3 (Lower bound)**.** *There exist constants $c_1, c_2 > 0$ depending on $p \geq 1$ only such that the following holds. For every $\varepsilon > 0$ with $F_1(\varepsilon) \geq 18$ one has*

$$D(c_1 F(\varepsilon), p) \geq c_2 \, \varepsilon.$$

Moreover, if $\nu(\mathbb{R}\backslash\{0\}) = \infty$ or $\sigma \neq 0$, one has for any $s > 0$,

$$D(c_1 F_1(\varepsilon), s) \gtrsim c_2 \, \varepsilon$$

as $\varepsilon \downarrow 0$.

The upper and lower bound often are of the same weak asymptotic order. Let, for instance, X be a α-stable Lévy process (α being a parameter in $(0,2]$). In that case, the bounds provided above are sharp, and one gets for $s_1 > 0$ and $s_2 \in (0,\alpha)$, that

$$D(r, s_1) \approx D^{(e)}(r, s_1) \approx D^{(q)}(r, s_2) \approx \frac{1}{r^{1/\alpha}}.$$

For more details we refer the reader to the original article.

References

Ankirchner, S., Dereich, S. & Imkeller, P. (2004), Enlargement of filtrations and continuous Girsanov-type embeddings. Preprint.

Aurzada, F. & Dereich, S. (2007), The coding complexity of Lévy processes. Preprint.

Belinsky, E. & Linde, W. (2002), 'Small ball probabilities of fractional Brownian sheets via fractional integration operators', *J. Theoret. Probab.* **15**(3), 589–612.

Bronski, J. C. (2003), 'Small ball constants and tight eigenvalue asymptotics for fractional Brownian motions.', *J. Theor. Probab.* **16**(1), 87–100.

Bucklew, J. A. (1984), 'Two results on the asymptotic performance of quantizers', *IEEE Trans. Inform. Theory* **30**(2, part 2), 341–348.

Bucklew, J. A. & Wise, G. L. (1982), 'Multidimensional asymptotic quantization theory with rth power distortion measures.', *IEEE Trans. Inf. Theory* **28**, 239–247.

Carl, B. & Stephani, I. (1990), *Entropy, compactness and the approximation of operators*, Vol. 98 of *Cambridge Tracts in Mathematics*, Cambridge University Press, Cambridge.

Cover, T. M. & Thomas, J. A. (1991), *Elements of information theory*, Wiley Series in Telecommunications. New York: John Wiley & Sons, Inc.

Creutzig, J. (2002), Approximation of Gaussian random vectors in Banach spaces. Ph.D. Dissertation, Friedrich-Schiller-Universität Jena.

Delattre, S., Graf, S., Luschgy, H. & Pagès, G. (2004), 'Quantization of probability distributions under norm-based distortion measures', *Statist. Decision* **22**, 261–282.

Dembo, A. & Kontoyiannis, I. (2002), 'Source coding, large deviations, and approximate pattern matching', *IEEE Trans. Inform. Theory* **48**(6), 1590–1615. Special issue on Shannon theory: perspective, trends, and applications.

Dereich, S. (2003*a*), High resolution coding of stochastic processes and small ball probabilities. Ph.D. Dissertation, TU Berlin,
URL: http://edocs.tu-berlin.de/diss/2003/dereich_steffen.htm.

Dereich, S. (2003*b*), 'Small ball probabilities around random centers of Gaussian measures and applications to quantization', *J. Theoret. Probab.* **16**(2), 427–449.

Dereich, S. (2005), 'Asymptotic behavior of the distortion-rate function for Gaussian processes in Banach spaces', *Bull. Sci. Math.* **129**(10), 791–803.

Dereich, S. (2006*a*), The coding complexity of diffusion processes under $L^p[0,1]$-norm distortion. Preprint.

Dereich, S. (2006*b*), The coding complexity of diffusion processes under supremum norm distortion. Preprint.

Dereich, S., Fehringer, F., Matoussi, A. & Scheutzow, M. (2003), 'On the link between small ball probabilities and the quantization problem for Gaussian measures on Banach spaces', *J. Theoret. Probab.* **16**(1), 249–265.

Dereich, S. & Lifshits, A. (2005), 'Probabilities of randomly centered small balls and quantization in Banach spaces.', *Ann. Probab.* **33**(4), 1397–1421.

Dereich, S. & Scheutzow, M. (2006), 'High resolution quantization and entropy coding for fractional brownian motion', *Electron. J. Probab.* **11**, 700–722.

Dereich, S. & Vormoor, C. (2005), The high resolution vector quantization problem with Orlicz norm distortion. Preprint.

Dzhaparidze, K. & van Zanten, H. (2004), 'A series expansion of fractional Brownian motion', *Probab. Theory Related Fields* **130**(1), 39–55.

Fehringer, F. (2001), Kodierung von Gaußmaßen. Ph.D. Dissertation, TU Berlin.

Gersho, A. (1979), 'Asymptotically optimal block quantization', *IEEE Trans. Inform. Theory* **25**(4), 373–380.

Graf, S. & Luschgy, H. (2000), *Foundations of quantization for probability distributions*, Lecture Notes in Mathematics 1730, Berlin: Springer.

Graf, S., Luschgy, H. & Pagès, G. (2006), Distortion mismatch in the quantization of probability measures. Preprint, arXiv:math/0602381v1 [math.PR].

Gray, R. M. & Neuhoff, D. L. (1998), 'Quantization', *IEEE Trans. Inf. Theory* **44**(6), 2325–2383.

Iglói, E. (2005), 'A rate-optimal trigonometric series expansion of the fractional Brownian motion', *Electron. J. Probab.* **10**.

Ihara, S. (1993), *Information theory for continuous systems*, Singapore: World Scientific.

Jeulin, T. & Yor, M., eds (1985), *Grossissements de filtrations: exemples et applications*, Vol. 1118 of *Lecture Notes in Mathematics*, Springer-Verlag, Berlin. Papers from the seminar on stochastic calculus held at the Université de Paris VI, Paris, 1982/1983.

Kieffer, J. C. (1993), 'A survey of the theory of source coding with a fidelity criterion', *IEEE Trans. Inform. Theory* **39**(5), 1473–1490.

Kolmogorov, A. (1956), 'On the Shannon theory of information transmission in the case of continuous signals', *IRE Trans. Inform. Theory* **2**, 102–108.

Kolmogorov, A. N. (1968), 'Three approaches to the quantitative definition of information', *Internat. J. Comput. Math.* **2**, 157–168.

Kuelbs, J. & Li, W. V. (1993), 'Metric entropy and the small ball problem for Gaussian measures', *J. Funct. Anal.* **116**(1), 133–157.

Li, W. V. & Linde, W. (1999), 'Approximation, metric entropy and small ball estimates for Gaussian measures', *Ann. Probab.* **27**(3), 1556–1578.

Li, W. V. & Shao, Q.-M. (2001), Gaussian processes: inequalities, small ball probabilities and applications, *in* 'Stochastic processes: theory and methods', Vol. 19 of *Handbook of Statist.*, North-Holland, Amsterdam, pp. 533–597.

Lifshits, M. A. (1997), 'On the lower tail probabilities of some random series', *Ann. Probab.* **25**(1), 424–442.

Luschgy, H. & Pagès, G. (2002), 'Functional quantization of Gaussian processes', *J. Funct. Anal.* **196**(2), 486–531.

Luschgy, H. & Pagès, G. (2004), 'Sharp asymptotics of the functional quantization problem for Gaussian processes', *Ann. Probab.* **32**(2), 1574–1599.

Luschgy, H. & Pagès, G. (2006), 'Functional quantization of a class of Brownian diffusions: a constructive approach', *Stochastic Process. Appl.* **116**(2), 310–336.

Luschgy, H. & Pagès, G. (2007), Functional quantization rate and mean pathwise regularity of processes with an application to Lévy processes. Preprint, arXiv:math/0601774v3 [math.PR].

Mason, D. M. & Shi, Z. (2001), 'Small deviations for some multi-parameter Gaussian processes.', *J. Theor. Probab.* **14**(1), 213–239.

Pagès, G. & Printems, J. (2004), Functional quantization for pricing derivatives. Preprint, Université de Paris VI, LPMA no. 930.

Shannon, C. E. (1948), 'A mathematical theory of communication', *Bell System Tech. J.* **27**, 379–423, 623–656.

Sytaya, G. (1974), 'On some asymptotic representation of the Gaussian measure in a Hilbert space', *Theory of Stochastic Processes* **2**, 94–104.

Talagrand, M. (1994), 'The small ball problem for the Brownian sheet', *Ann. Probab.* **22**(3), 1331–1354.

Zador, P. L. (1966), Topics in the asymptotic quantization of continuous random variables. Bell Laboratories Technical Memorandum.