Asymptotic behavior of the distortion-rate function for Gaussian processes in Banach spaces

Steffen Dereich

Fakultät II, Institut für Mathematik, MA 7-5, Technische Universität Berlin, Straße des 17. Juni 136, D-10623 Berlin, Germany

Received 15 April 2005; accepted 17 April 2005
Available online 1 August 2005

Abstract

Let \( \mu \) be a Gaussian measure on a separable Banach space. We prove a tight link between the logarithmic small ball probabilities of \( \mu \) and certain moment generating functions. Based upon this link we provide a new lower bound for the distortion-rate function (DRF) against the small ball function. This allows us to use results of the theory of small ball probabilities to deduce lower bounds for the DRF. In particular, we obtain the correct weak asymptotics of the distortion rate function in many important cases (e.g. Brownian motion).

© 2005 Elsevier SAS. All rights reserved.

Keywords: High-resolution coding; Small ball probabilities; Distortion-rate function; Gaussian process

1. Introduction and summary of results

We study the high resolution coding problem for (infinite-dimensional) Gaussian measures on separable Banach spaces. In the sequel, \((E, \| \cdot \|)\) denotes a separable Banach space and \( \mu \) denotes a centered Gaussian measure on the Borel sets of \( E \). Moreover, we let \( X \) be an \( E \)-valued \( \mu \)-distributed random vector (r.v.) defined on a probability space \((\Omega, \mathcal{F}, P)\). For \( q \in [1, \infty) \), we denote by \( \| \cdot \|_{L_q(P)} \) the norm given by the functional \( \|Y\|_{L_q(P)} = \mathbb{E}[\|Y\|^q]^{1/q} \) for \( E \)-valued r.v.’s \( Y \).
We study the best achievable quality of an approximation \( \hat{X} \), the reconstruction, for the original \( X \), under certain information constraints parameterized by the rate \( r \geq 0 \):

1. \( \hat{X} \) takes at most \( e^r \) distinct values (quantization).
2. \( \hat{X} \) has entropy less than \( r \) (entropy coding).
3. \( \hat{X} \) is such that the Shannon mutual information between \( X \) and \( \hat{X} \) is less than \( r \) (Shannon coding).

The constraints above are ordered increasingly in the sense that reconstructions satisfying condition (1) also satisfy condition (2), and those satisfying condition (2) satisfy condition (3).

In [1] (see also [2, Theorem 3.1.2]) the quantization problem was related to small ball probabilities. It was found that, if the small ball function

\[ \varphi(\varepsilon) := -\log \mu(B(0, \varepsilon)), \quad \varepsilon > 0, \]

satisfies \( \varphi^{-1}(\varepsilon) \approx \varphi^{-1}(2\varepsilon) \) as \( \varepsilon \downarrow 0 \), then for all moments \( q \geq 1 \) one has

\[ \varphi^{-1}(r) \lesssim \inf_{\hat{X}} \|X - \hat{X}\|_{L_q(P)} \lesssim 2\varphi^{-1}(r/2), \quad r \to \infty, \]

where the infimum is taken over all \( E \)-valued random vectors \( \hat{X} \) satisfying the quantization constraint (1).

Here and elsewhere we write \( f \sim g \) iff \( \lim f/g = 1 \), while \( f \lesssim g \) stands for \( \limsup f/g \leq 1 \). Finally, \( f \approx g \) means

\[ 0 < \liminf f/g \leq \limsup f/g < \infty. \]

If \( \hat{X} \) satisfies the quantization constraint, then there exists a prefix-free code \( \Upsilon : \text{range}(\hat{X}) \to \{0, 1\}^* \), with \( \text{length}(\Upsilon(\hat{X})) \leq 1 + r/\log 2 \) a.s., which gives a worst case bound on the complexity. Here, \( \text{range}(\hat{X}) \) denotes the range of the discrete r.v. \( \hat{X} \). A broader class of reconstructions is admitted when considering entropy coding. In that case, Huffman coding yields a prefix-free code \( \Upsilon : \text{range}(\hat{X}) \to \{0, 1\}^* \) with \( \mathbb{E}[\text{length}(\Upsilon(\hat{X}))] < 1 + r/\log 2 \), which is an average case complexity bound. The last information constraint (mutual information constraint) leads to the distortion rate function (DRF). Due to Shannon’s source coding theorem, this is the asymptotically best achievable average distortion induced by certain block codes under single letter distortion measures. For more details concerning information theory we refer the reader to the monograph by Cover and Thomas [3].

If the underlying space \( E \) is a Hilbert space, then the comparison of coding quantities induced by the different constraints is possible. In that case the asymptotics of the problems are determined by the eigenvalues, \( (\lambda_n)_{n \in \mathbb{N}} \), of the covariance operator of \( \mu \). Essentially, one sees that the strong asymptotics coincide for any moment \( q \geq 2 \) and any of the above constraints, if the eigenvalues satisfy

\[ \lim_{n \to \infty} \frac{\log \log(1/\lambda_n)}{n} = 0, \]

which is the case for many important examples. These results were derived in the author’s dissertation [2] (see also [4]) and will be published in a forthcoming article. The main
results of this article are a new lower bound for the distortion rate function and a tight relation between small ball probabilities (SBPs) and certain Legendre transforms. We shall see that in our framework, the known lower bound for the quantization problem remains valid for the distortion rate function. In particular, the technique of using block codes (as done in the source coding theorem) does not change the asymptotic rate of the coding problem.

Let us now give the main notation. We denote by $H(\hat{X})$ the entropy of a random vector $\hat{X}$, i.e., $H(\hat{X}) = -\sum_x P(\hat{X} = x) \log P(\hat{X} = x)$ if $\hat{X}$ is supported by a discrete set and $H(\hat{X}) = \infty$ otherwise. Moreover, the Shannon mutual information between $X$ and $\hat{X}$ is denoted by $I(X; \hat{X})$, i.e.,

$$I(X; \hat{X}) = \left\{ \int \log \frac{dP_{X,\hat{X}}}{dP_X \otimes P_{\hat{X}}} \, dP_X, \hat{X} \text{ r.v. in } E \text{ with } |\text{range}(\hat{X})| \leq e^r \right\} \quad \text{else.}$$

The information constraints (1) and (3) from above induce the following approximation quantities for the rate $r \geq 0$ and the moment $q > 0$,

$$D(q)(r, q) = \inf \left\{ \|X - \hat{X}\|_{L^q(P)} : \hat{X} \text{ r.v. in } E \text{ with } I(X, \hat{X}) \leq r \right\},$$

$$D(r, q) = \inf \left\{ \|X - \hat{X}\|_{L^q(P)} : \hat{X} \text{ r.v. in } E \text{ with } I(X, \hat{X}) \leq r \right\}.$$  

We will not consider entropy coding any further, since the corresponding approximation quantity lies between $D$ and $D(q)$. Strictly speaking, the quantity $D$ depends on the underlying probability space $(\Omega, \mathcal{F}, P)$. In order to suppress this dependence, we assume that $(\Omega, \mathcal{F}, P)$ is sufficiently rich. More explicitly, we assume that for any probability kernel $K$ from $E$ to $E$ there exists an $E$-valued random vector $\hat{X}$ such that for two Borel sets $A$ and $B$ of $E$,

$$P(X, \hat{X}) \in A \times B) = \int_A K(x, B) P_X(dx).$$

Here $P_X$ denotes the probability distribution of $X$.

Now we are in a position to state the main theorems.

**Theorem 1.1.** For every $\varepsilon > 0$ and $q \geq 1$, there exists $r_0 \geq 1$ such that

$$D(r, q) \geq D(r, 1) \geq \varphi^{-1} \left( r + \frac{1 + \varepsilon}{2} \log r \right)$$

for all $r \geq r_0$.

The new lower bound implies

**Theorem 1.2.** If

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

then for any $q \geq 1$

$$\varphi^{-1}(r) \lesssim D(r, q) \leq D(q)(r, q) \lesssim 2\varphi^{-1}(r/2)$$

as $r \to \infty$. 
The proof of Theorem 1.1 relies on a tight relation between a particular moment generating function and small ball probabilities for Gaussian measures, obtained in Section 3.

**Theorem 1.3.** For \( \eta > 0 \), there exists a universal constant \( r_0 = r_0(\eta) \geq 0 \) such that the following holds: Let \( X \) be an arbitrary Gaussian random vector in an arbitrary separable Banach space \((E, \| \cdot \|)\), and let \( x \in E, \varepsilon > 0 \) and \( q \geq 1 \), then one has

\[-\log P(\|X - x\|^q \leq \varepsilon) \leq \left[ \Lambda^*_x(\varepsilon) + \frac{1 + \eta}{2} \log \Lambda^*_x(\varepsilon) \right] \vee r_0,
\]

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

In order to prove Theorem 1.1, a weaker statement than Theorem 1.3 would suffice. However, the link provided above is useful beyond the application in this article. In fact, it is one of the main tools needed to infer the strong asymptotics in the Hilbert space setting [2, Section 6].

The article is organized as follows. In Section 2, we use results of information theory to lower bound the DRF against some particular Legendre transform which represents a measure for the mass concentration of \( \mu \) around 0. In the following section, a concentration property of Gaussian measures (the Ehrhard inequality) is used to relate the small ball function to the former Legendre transform. Then follows a proof of Theorem 1.3. Next, we combine all results and prove Theorems 1.1 and 1.2. In the last section, we state some of the numerous known results on small ball probabilities and give the corresponding estimates for the distortion-rate function.

2. A lower bound for the distortion-rate function

In this section we provide an estimate of \( D(r, 1) \) against an inverse of a particular Legendre transform.

**Lemma 2.1.** For \( r \geq 0 \) it is true that

\[ D(r, 1) \geq \inf \{ d \geq 0 : \hat{\Lambda}^*(d) \leq r \}, \]

where \( \hat{\Lambda}^*(d) = \sup_{\theta \leq 0} [\theta d - \hat{\Lambda}(\theta)] \) is the Legendre transform of

\[ \hat{\Lambda}(\theta) = \log \left( \int_E e^{\theta \|x\|} d\mu(x) \right), \theta \leq 0. \]

In order to prove the lemma, we use results of information theory which can be found, for instance, in Dembo and Kontoyiannis [5]. In particular, we adopt their notation.

Let \( P \) be an arbitrary measure on the Borel sets of \( E \) and let \( \rho : E \times E \to [0, \infty) \) be a Borel measurable function. For \( d \geq 0 \), we define the rate-distortion function of the information source \((P, \rho)\) by

\[ R(d) = \inf_{(X,Y)} I(X; Y), \]

(1)
where the infimum is taken over all $E^2$-valued r.v.’s $(X, Y)$ such that $X$ has distribution $P$ and $\mathbb{E}[\rho(X, Y)] \leq d$. Moreover, for two probability measures $Q$ and $W$ defined on a common measurable space, we denote by $H(Q\|R)$ the relative entropy of $Q$ w.r.t. $R$, i.e.

$$H(Q\|R) = \left\{ \begin{array}{ll} \int \log \frac{dQ}{dP} dQ, & \text{if } Q \ll R, \\ \infty, & \text{else.} \end{array} \right.$$  

**Lemma 2.2.** For $d \geq 0$ one has

$$R(d) \geq \Lambda^*(d),$$

where $\Lambda^*(d) = \sup_{\theta \leq 0} [\theta d - \Lambda(\theta)]$ is the Legendre transform of

$$\Lambda(\theta) := \sup_{y \in E} \log \left( \int_{E} e^{\theta \rho(x, y)} dP(x) \right).$$

**Proof.** Let $Q \in \mathcal{M}_1(E)$, $d \geq 0$ and

$$R_1(Q, d) = \inf_{W} H(W\|P \otimes Q),$$

where the infimum is taken over all probability measures $W \in \mathcal{M}_1(E \times E)$ such that the first marginal of $W$ is $P$ and $\int_{E \times E} \rho(x, y) dW(x, y) \leq d$. By Yang and Kieffer (see [5, Remark 1]) one has

$$R_1(Q, d) = \inf_{(X, Y)} \left[ I(X; Y) + H(P_Y \| Q) \right],$$

where the infimum is taken over all $E^2$-valued random vectors $(X, Y)$ with $\mathcal{L}(X) = P$ and $\mathbb{E}[\rho(X, Y)] \leq d$. It follows that

$$R(d) = \inf_{Q \in \mathcal{M}_1(E)} R_1(Q, d). \quad (2)$$

Let now

$$\Lambda(\theta; Q) = \int_{E} \log \left( \int_{E} e^{\theta \rho(x, y)} dQ(y) \right) dP(x), \quad \lambda \leq 0.$$  

Then Theorem 2 of [5] implies that for $d \geq 0$,

$$R_1(d; Q) \geq \Lambda^*(d; Q), \quad (3)$$

where

$$\Lambda^*(d; Q) := \sup_{\theta \leq 0} [\theta d - \Lambda(\theta; Q)].$$

Due to Jensen’s inequality one obtains that, for any $Q \in \mathcal{M}_1(E)$ and $\theta \leq 0$,

$$\Lambda(\theta; Q) \leq \log \left( \int_{E \times E} e^{\theta \rho(x, y)} dP \otimes Q(x, y) \right)$$

$$\leq \sup_{y \in E} \log \left( \int_{E} e^{\theta \rho(x, y)} dP(x) \right) = \Lambda^*(\theta).$$

Consequently, $\Lambda^*(d; Q) \geq \hat{\Lambda}^*(d)$, and with (2) and (3) we arrive at

$$R(d) \geq \hat{\Lambda}^*(d).$$

**Proof of Lemma 2.1.** Choose $\rho(x, y) = \|x - y\|$, $x, y \in E$, and $P = \mu$ and let $R(\cdot)$ as in (1). Note that for $\theta \leq 0$

$$\sup_{y \in E} \int_\mathcal{E} e^\theta \|x - y\| \, dP(x) = \int_\mathcal{E} e^\theta \|x\| \, dP(x)$$

due to the Anderson inequality. Hence, the definitions of $\hat{\Lambda}$ of Lemmas 2.2 and 2.1 coincide. Next, the definition of $R(\cdot)$ implies that

$$D(r, 1) \geq \inf \left\{ d \geq 0 : R(d) \leq r \right\}.$$

Consequently,

$$D(r, 1) \geq \inf \left\{ d \geq 0 : \hat{\Lambda}^*(d) \leq r \right\}. \quad \square$$

3. SBPs and moment generating functions

The objective of this section is to prove Theorem 1.3. We need the following notation. Let $x \in E$ and let $\tau : [0, \infty) \to [0, \infty)$ be a Young function, i.e. $\tau$ is convex, one-to-one and satisfies $\tau(0) = 0$. Denote, for $t > 0$ and $\theta \in \mathbb{R}$,

$$Z = \tau\left(\|X - x\|\right),$$

$$F(t) = \mathbb{P}(Z \leq t),$$

$$\Lambda(\theta) = \log \mathbb{E}[e^{\theta Z}],$$

$$\Lambda^*(t) = \sup_{\theta \leq 0} \left[ \theta t - \Lambda(\theta) \right].$$

The proof of Theorem 1.3 is based on the following two lemmas.

**Lemma 3.1.** For $t \in (0, 1/2)$, set

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

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

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

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

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

$$\Lambda(\theta) = \log \mathbb{E}[e^{\theta Z}] \geq \theta t + \log \mathbb{P}(Z \leq t).$$
Therefore,
\[ A^*(t) = \sup_{\theta \leq 0} \left[ t\theta - A(\theta) \right] \leq -\log \mathbb{P}(Z \leq t). \]

We proceed with the proof of the second inequality. Suppose first that \( t_0 > 0 \) is such that \( \mathbb{P}(Z \leq t_0) = 0 \) and fix \( p \in (0, 1) \) arbitrarily. Then there exists \( \varepsilon > 0 \) such that \( \mathbb{P}(Z < t_0 + \varepsilon) \leq p \). Consequently, for \( \theta \leq 0 \),
\[ A(\theta) = \log \mathbb{E} e^{\theta Z} \leq \log \left[ p e^{\theta t_0} + (1 - p) e^{\theta (t_0 + \varepsilon)} \right] \]
and
\[ A^*(t_0) \geq \limsup_{\theta \to -\infty} \left[ \theta t_0 - A(\theta) \right] \geq -\log p. \]

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

Now let \( t_0 > 0 \) with \( \mathbb{P}(Z \leq t_0) \in (0, 1/2) \). In order to show the second inequality, we let \( G(t) = \mathbb{P}(\|X - x\| \leq t), t > 0 \), and consider the function
\[ f := \Phi^{-1} \circ F = \Phi^{-1} \circ G \circ \tau^{-1} : (0, \infty) \to \mathbb{R} \cup \{-\infty\}. \]

Note that \( \Phi^{-1} \circ G \) is concave. In fact, the Ehrhard inequality (see [6, Theorem 4.2.2]) implies that for \( \gamma \in (0, 1) \) and \( t_1, t_2 \geq 0 \)
\[ \Phi^{-1} \circ G \left( \gamma t_1 + (1 - \gamma) t_2 \right) \]
\[ = \Phi^{-1} \left( \mu \left( \gamma B(x, \gamma t_1) + (1 - \gamma) B(x, t_2) \right) \right) \]
\[ = \Phi^{-1} \left( \mu \left( \gamma B(x, t_1) + (1 - \gamma) B(x, t_2) \right) \right) \]
\[ \geq \gamma \Phi^{-1} \left( \mu \left( B(x, t_1) \right) \right) + (1 - \gamma) \Phi^{-1} \left( \mu \left( B(x, t_2) \right) \right) \]
\[ = \gamma (\Phi^{-1} \circ G)(t_1) + (1 - \gamma) (\Phi^{-1} \circ G)(t_2). \]

Since \( \tau^{-1} \) is concave and \( \Phi^{-1} \circ G \) is monotonically increasing, it follows that \( f \) is concave.

Let now \( q \) denote a tangent of the graph of \( f \) at the point \((t_0, f(t_0))\). Represent \( q \) in the form \( q(t) = (t - m)s \), where \( m, s > 0 \) are appropriate constants. Next, \( N \) denotes a standard normal r.v. and we associate \( q \) with the random variable \( Z_q = q^{-1}(N) = N/s + m \). \( Z_q \) has distribution function \( \Phi \circ q \) and, hence, it is a normal r.v. on \( \mathbb{R} \). Note, that \( F(t) = \Phi \circ f(t) \). Consequently, the distribution function of the r.v. \( f^{-1}(N) \) equals \( F \).

We assume without loss of generality that \( Z = f^{-1}(N) \). Since \( q \) is a tangent of the concave function \( f \), one has \( q \geq f \). Thus we conclude that \( Z_q = q^{-1}(N) \leq f^{-1}(N) = Z \) so that for every \( \theta \leq 0 \)
\[ A_q(\theta) := \log \mathbb{E} e^{\theta Z_q} \geq \log \mathbb{E} e^{\theta Z} = A(\theta). \]

Consequently,
\[ A^*(t) \geq \sup_{\theta \leq 0} \left[ t\theta - A_q(\theta) \right] =: A^*_q(t) \]
for every \( t > 0 \).

On the other hand, one has \( A_q(\theta) = (\theta/s)^2/2 + m\theta \) and, for \( t \in (0, m] \),
\[ \Lambda_q^*(t) = \sup_{\theta \leq 0} \left[ \theta t - \Lambda_q(\theta) \right] \]
\[ = \sup_{\theta \leq 0} \left[ -\frac{1}{2s^2} (\theta + s^2 (m - t))^2 + \frac{s^2(m - t)^2}{2} \right] \]
\[ = \frac{s^2(m - t)^2}{2} = \frac{q(t)^2}{2}. \]
Noticing that \( t_0 < m \), one obtains
\[ \Lambda_q^*(t_0) = \frac{q(t_0)^2}{2} = \frac{f(t_0)^2}{2} = \frac{(\Phi^{-1}(F(t_0)))^2}{2}. \]
Hence,
\[ -\log P(Z \leq t_0) \leq \frac{2 \log(1/F(t_0))}{(\Phi^{-1}(F(t_0)))^2} = h(F(t_0)). \]
The convergence \( \lim_{t \downarrow 0} h(t) = 1 \) is established in the lemma below.

**Lemma 3.2.**
\[ h(\varepsilon) = 1 + \frac{\log \log(1/\varepsilon)}{2 \log(1/\varepsilon)} + o \left( \frac{\log \log(1/\varepsilon)}{\log(1/\varepsilon)} \right), \quad \varepsilon \downarrow 0, \]
where \( o \) denotes the Landau symbol.

**Proof.** In order to prove the lemma, we derive the asymptotics of \( (\Phi^{-1}(\varepsilon))^2 \) as \( \varepsilon \downarrow 0 \).
Consider the functions
\[ g : (0, \infty) \rightarrow (0, 1), \quad t \mapsto e^{-t/2} \text{ and } \]
\[ \tilde{g} : (-\infty, 0) \rightarrow (0, 1), \quad t \mapsto e^{-t^2/2}. \]
Both functions are one-to-one and possess inverse functions \( g^{-1} \) and \( \tilde{g}^{-1} \). We denote by \( \tilde{\Phi} = \Phi_{(-\infty, 0)} \) the function \( \Phi \) restricted to the domain \( (-\infty, 0) \) and observe that
\[ (\tilde{\Phi}^{-1})^2 = g^{-1} \circ g \circ (\tilde{\Phi}^{-1})^2 = g^{-1} \circ \tilde{g} \circ \tilde{\Phi}^{-1} = g^{-1} \circ (\tilde{\Phi} \circ \tilde{g}^{-1})^{-1}. \] (4)
Since \( \Phi(t) \sim (2\pi)^{-1/2} e^{-t^2/2} / (-t) \) as \( t \rightarrow -\infty \), one has,
\[ \tilde{\Phi} \circ \tilde{g}^{-1}(\varepsilon) \sim \frac{\varepsilon}{\sqrt{2\pi \log(1/\varepsilon^2)}}, \quad \varepsilon \downarrow 0. \]
The latter function is regularly varying, hence its inverse is again regularly varying and satisfies (see Bingham et al. [7, p. 28])
\[ (\tilde{\Phi} \circ \tilde{g}^{-1})^{-1}(\varepsilon) \sim \sqrt{2\pi \log(1/\varepsilon^2)} \varepsilon, \quad \varepsilon \downarrow 0. \]
Next, recall that \( (\tilde{\Phi}^{-1})^2 = g^{-1} \circ (\tilde{\Phi} \circ \tilde{g}^{-1})^{-1} \) and \( g^{-1}(u) = -2 \log u, u \in (0, 1) \). Consequently,
\[ \lim_{\varepsilon \downarrow 0} \left[ (\Phi^{-1}(\varepsilon))^2 + 2 \log(\sqrt{2\pi \log(1/\varepsilon^2)} \varepsilon) \right] = 0. \]
We denote by $\eta(\varepsilon)$ the term in the above brackets, i.e. $\eta(\varepsilon) = (\Phi^{-1}(\varepsilon))^2 - [2 \log(1/\varepsilon) - \log(\log(1/\varepsilon)) - \log(4\pi)], \varepsilon \in (0, 1/2)$. Then
\[ h(\varepsilon) = \frac{2 \log(1/\varepsilon)}{(\Phi^{-1}(\varepsilon))^2} = 1 + \frac{\log(\log(1/\varepsilon)) + \log(4\pi) - \eta(\varepsilon)}{2 \log(1/\varepsilon) - \log(\log(1/\varepsilon)) - \log(4\pi) + \eta(\varepsilon)} \]
which implies the assertion. □

**Proof of Theorem 1.3.** For $t \geq \log 2$, let $\tilde{h}(t) = h(e^{-t})$. Due to Lemma 3.2 there exists $t_0 \geq \log 2$ such that
\[ \tilde{h}(t) \leq 1 + \frac{1}{2} \log \left( \frac{\Lambda_x^*(\varepsilon)}{\tilde{F}(\varepsilon)} \right) \Lambda_x^*(\varepsilon) \]
for all $t \geq t_0$. Let now $x \in E$ and $q \geq 1$ arbitrary, and consider $\tau(u) := u^q$, $\tilde{F}(\varepsilon) := -\log P(\|X - x\|^q \leq \varepsilon)$, $\varepsilon > 0$, and $\Lambda_x^*$ as in the statement of the theorem. From now on we assume that $\varepsilon > 0$ is such that $\tilde{F}(\varepsilon) \geq t_0$. Then Lemma 3.1 implies that
\[ \Lambda_x^*(\varepsilon) \leq \tilde{F}(\varepsilon) \leq \tilde{h}(\tilde{F}(\varepsilon)) \Lambda_x^*(\varepsilon) \leq \left( 1 + \frac{1 + \eta \log \tilde{F}(\varepsilon)}{2} \right) \Lambda_x^*(\varepsilon) \]
Next, note that $f : u \mapsto \frac{\log u}{u}$ is monotonically decreasing on the interval $[e, \infty)$. Consequently, if $\Lambda_x^*(\varepsilon) \geq e$, one has
\[ \tilde{F}(\varepsilon) \leq \left( 1 + \frac{1 + \eta \log \Lambda_x^*(\varepsilon)}{2} \right) \Lambda_x^*(\varepsilon) = \Lambda_x^*(\varepsilon) + \frac{1 + \eta}{2} \log \Lambda_x^*(\varepsilon) \]
whereas if $\Lambda_x^*(\theta) < e$, then
\[ \tilde{F}(\varepsilon) \leq \left( 1 + \frac{1 + \eta}{2} \sup_{t \in \mathbb{R}_+} f(t) \right) \Lambda_x^*(\theta) \leq \left( 1 + (1 + \eta)/(2e) \right) e. \]
Altogether, we obtain for $\varepsilon > 0$ arbitrary and $r_0 := t_0 \vee (e + (1 + \eta)/2)$
\[ \tilde{F}(\varepsilon) \leq \left[ \Lambda_x^*(\varepsilon) + \frac{1 + \eta}{2} \log \Lambda_x^*(\varepsilon) \right] \vee r_0. \]
Notice that the value of $r_0$ depends on $\eta > 0$ only so that the proof is complete. □

4. A lower bound for the DRF

In this section, we combine the previous results to prove Theorems 1.1 and 1.2.

**Proof of Theorem 1.1.** Let $\varepsilon > 0$, $\hat{A}(\theta) = \log \mathbb{E}[e^{\theta \|X\|}]$ for $\theta \leq 0$, and denote by $\hat{A}^*$ the Legendre transform of $\hat{A}$. According to Theorem 1.3, there exists $r_0 \geq 1$ such that
\[ \varphi(d) \leq \left[ \hat{A}^*(d) + \frac{1 + \varepsilon}{2} \log \hat{A}^*(d) \right] \vee r_0 \]
for all $d > 0$. Moreover, by Lemma 2.1 it holds for $r \geq r_0$. 
Note that \( \phi : \mathbb{R}_+ \to \mathbb{R}_+ \) is invertible, hence

\[
D(r, 1) \geq \varphi^{-1}\left(r + \frac{1 + \varepsilon}{2} \log r\right)
\]

for all \( r \geq r_0 \). Moreover, for any \( q \geq 1 \) and any \( E \)-valued r.v. \( \hat{X} \) one has \( \|X - \hat{X}\|_{L^q(P)} \geq \|X - \hat{X}\|_{L^1(P)} \) so that

\[
D(r, q) \geq D(r, 1).
\]

**Proof of Theorem 1.2.** It remains to prove the asymptotic lower bound. It is a consequence of Theorem 1.1 and the following property of convex functions: If \( f : \mathbb{R}_+ \to \mathbb{R}_+ \) is a decreasing convex function with \( f(r) \approx f(2r) \) as \( r \to \infty \), then for any function \( \Delta : \mathbb{R}_+ \to \mathbb{R} \) with \( \Delta r = o(r) \) it holds

\[
f(r + \Delta r) \sim f(r), \quad r \to \infty.
\]

The proof is elementary and is contained, for instance, in [2, Lemma 3.1.4]. In fact, this implies the assertion, since \( \varphi^{-1} \) is decreasing and convex. \( \square \)

### 5. Known results about small ball probabilities

In the past years, considerable effort has been put into the determination of the asymptotic behavior of the small ball function for centered Gaussian measures on Banach spaces. Beside quantization, these results can be used to derive certain kinds of laws of the iterated logarithm, and to get hold of certain metric entropies. An overview on the topic can be found in Li and Shao [8]. Below we summarize some results and give the corresponding estimates for the coding quantities.

#### 5.1. Wiener measure

We consider the Wiener measure \( \mu \) on various separable Banach spaces \( E \):

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

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

where \( |\cdot|_G \) is an arbitrary norm on \( \mathbb{R}^d \). Owing to Ledoux [9],

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

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

and, for $q \geq 1$,

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

as $r \to \infty$.

- $E = L_p[0, 1], \; p \geq 1$, equipped with the $L_p$-norm $\| \cdot \|_{L_p[0,1]}$. It is well known (see for instance Li and Shao [8]) that the small ball probabilities satisfy

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

where

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

and

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

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

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

as $r \to \infty$. The small ball probabilities under the $L_p$-norm for general Gaussian Markov processes is treated in Li [10].

- $E = C^\alpha_0, \; \alpha \in (0, 1/2)$, the space of $\alpha$-Hölder continuous functions over the time $[0, 1]$ starting in 0 equipped with the norm

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

Referring to Kuelbs and Li [11], there exists $c_\alpha > 0$ with

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

The constant $c_\alpha$ is not known explicitly although lower and upper bounds are derived in [11]. We obtain, for $q \geq 1$

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

as $r \to \infty$. 
5.2. Gaussian sheets

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

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

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

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

have been studied by many authors. If \( d = 1 \), the process is 1-dimensional fractional Brownian motion and the asymptotics of the SBFs are as stated above. In the case where there is a unique minimum, say \( \gamma_1 \), in \( \gamma = (\gamma_1, \ldots, \gamma_d) \), it was derived by Mason and Shi [12] that

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

Belinsky and Linde [13] studied the case with exactly two minimal elements, say \( \gamma_1 \) and \( \gamma_2 \), in \( \gamma = (\gamma_1, \ldots, \gamma_d) \). They found

\[
\varphi(\varepsilon) \approx \varepsilon^{-2/\gamma_1} \left( \log \left( \frac{1}{\varepsilon} \right) \right)^{1+2/\gamma_1}, \quad \varepsilon \downarrow 0
\]

and extended a result of Talagrand [14], who solved the small ball problem for \( \gamma = (1,1) \), i.e. in the case where \( X \) is a 2-dimensional Brownian sheet.

References

Probab. 6 (3) (1993) 547–577.
[13] E. Belinsky, W. Linde, Small ball probabilities of fractional Brownian sheets via fractional integration oper-