The coding complexity of diffusion processes under $L^p[0,1]$-norm distortion

by

Steffen 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. We investigate the high resolution quantization and entropy coding problem for solutions of stochastic differential equations under $L^p[0,1]$-norm distortion. We find explicit high resolution formulas in terms of the average diffusion coefficient seen by the process. The proof is based on a decoupling method introduced in a former article by the author. Given that link it remains to analyze the coding problem for a concatenation of Wiener processes and to solve the corresponding rate allocation problem.

Keywords. High-resolution quantization; complexity; stochastic processes; stochastic differential equations.

2000 Mathematics Subject Classification. 60G35, 41A25, 94A15.

1 Introduction

In this article we use a decoupling method derived in a former article by the author to solve the high resolution quantization and entropy coding problem for diffusions X (*original*) under $L^p[0,1]$-norm $\|\cdot\| = \|\cdot\|_{L^p[0,1]}$. We will keep $p > 0$ fixed in the whole article. Sometimes we also need to evaluate different L^s-norms. We shall write $\|\cdot\|_{L^s[a,b)}$ and $\|\cdot\|_{L^s(\mathbb{P})}$ for the L^s-norm on the interval $[a,b)$ and the L^s-norm induced by the measure $\mathbb{P}$, respectively. We also will use this notation for $s \in (0,1)$. Moreover, let $D[a,b]$ denote the space of real-valued cadlag functions defined on $[a,b]$. We shall use analog notation for open and half-open intervals.

The article is devoted to the analysis of the *quantization error*

$$D^{(q)}(r|s) = \inf\{\mathbb{E}[\|X-\hat{X}\|^s]^{1/s} : \hat{X}\ D[0,1]\text{-valued with } |\operatorname{range}\hat{X}| \le e^r\},$$

the *entropy coding error*

$$D^{(e)}(r|s) = \inf\{\mathbb{E}[\|X-\hat{X}\|^s]^{1/s} : \hat{X}\ D[0,1]\text{-valued with } \mathbb{H}(\hat{X}) \le r\},$$

and the *distortion rate function*

$$D(r|s) = \inf\{\mathbb{E}[\|X-\hat{X}\|^s]^{1/s} : \hat{X}\ D[0,1]\text{-valued with } I(X;\hat{X}) \le r\}.$$

Here and elsewhere $\mathbb{H}(\hat{X})$ denotes the entropy of $\hat{X}$ in the natural basis that is

$$\mathbb{H}(\hat{X}) = \begin{cases} \sum_{x\in \text{range}\,(\hat{X})} p_x \log(1/p_x) & \text{if } \hat{X} \text{ is discrete} \\ \infty & \text{otherwise,} \end{cases}$$

where (p_x) denote the probability weights of $\hat{X}$, and I denotes the *Shannon mutual information* defined as

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

Strictly speaking the values of the distortion rate function depend on the underlying probability space. We shall assume the existence of a $[0,1]$-uniformly distributed random variable, that is independent of X. (In this case the distortion rate function attains its minimal value.)

If the original $X = (X_t)_{t\in[0,1]}$ is a Wiener process the approximation quantities satisfy

$$\lim_{r\to\infty} \sqrt{r}\, D^{(q)}(r|s) = \lim_{r\to\infty} \sqrt{r}\, D^{(e)}(r|s) = \lim_{r\to\infty} \sqrt{r}\, D(r|p) = K_p \tag{1}$$

for some constant $K_p \in (0,\infty)$ that does not depend on the moment $s > 0$ (Theorems 1.3 and 6.1 of [3]). So far the only explicitly known value for K_p is $K_2 = \frac{\sqrt{2}}{\pi}$ (see [5], [1])

Our proofs are based on a decoupling argument introduced in [2] to analyze the supremum norm distortion setting. As pointed out in [2] the decoupling method may also be proven for d-dimensional diffusions as long as the diffusion coefficient is *scalar*. However, we only treat the 1-dimensional setting here.

Let us now fix the notation. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P})$ be a complete filtered probability space that satisfies the usual hypotheses, that is $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets of $\mathcal{F}$ and $(\mathcal{F}_t)$ is right continuous. Let $(W_t)_{t\ge 0}$ be a 1-dimensional $(\mathcal{F}_t)$-Wiener process. We denote by $\sigma : \mathbb{R}\times[0,\infty) \to \mathbb{R}$ and $b : \mathbb{R}\times[0,\infty) \to \mathbb{R}$ two deterministic functions, and assume that $(X_t)_{t\ge 0}$ is an $(\mathcal{F}_t)$-adapted semimartingale solving the integral equation

$$X_t = \int_0^t b(X_s, s)\, ds + \int_0^t \sigma(X_s, s)\, dW_s \qquad (t \ge 0). \tag{2}$$

For ease of notation, we abridge $b_t := b(X_t, t)$ and $\sigma_t := \sigma(X_t, t)$ for $t \ge 0$. $(X_t)_{t\in[0,1]}$ represents the original process which is to be approximated by some r.v. $\hat{X}$, the *reconstruction.*

We assume the following technical assumption:

Assumption (C): There exist constants $\beta \in (0,1]$ and $L < \infty$ such that for $x, x' \in \mathbb{R}$ and $t, t' \in [0,1]$:

$$\begin{aligned} |b(x,t)| \le L(|x|+1)\ ,\ |\sigma(0,0)| \le L \text{ and} \\ |\sigma(x,t) - \sigma(x',t')| \le L[|x-x'|^\beta + |x-x'| + |t-t'|^\beta]. \end{aligned} \tag{3}$$

As a consequence of assumption (C) all moments $\mathbb{E}[\|X\|^s]$ $(s \geq 1)$ are finite. We shall use this fact without further mentioning. Additionally, we assume that the process $(\sigma_t)_{t\in[0,1]}$ is not indistinguishable from the constant 0-function, since otherwise the problem is trivial.

Note that assumption (C) does neither ensure existence nor uniqueness of the solution of the stochastic differential equation (2). More information on existence and uniqueness of stochastic differential equations can be found for instance in [6].

Sometimes we need to consider the above approximation numbers for other originals than X and for different time horizons: For $T > 0$, a $D[0,T]$-valued random vector Z, $s > 0$ and $r \geq 0$, let

$$D^{(q)}(r|Z,T,s) = \inf_{\hat{Z}} \mathbb{E}[\|Z - \hat{Z}\|^s_{L^p[0,T]}]^{1/s},$$

where the infimum is taken over all discrete, $D[0,T]$-valued r.v.'s $\hat{Z}$ with

$$|\,\mathrm{range}\,(\hat{Z})| \leq e^r.$$

We call $D^{(q)}(r|Z,Ts)$ the *s-th moment quantization error* for the *rate* r, *source* Z and time horizon T. We use analog notation for the entropy coding error and the distortion rate function, and we will often omit parameters in the notation that are obvious from the context.

The high resolution formula for diffusion reads as follows:

Theorem 1.1. *For $s > 0$ one has*

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

and

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

The proof of the theorem is based on a decoupling method which we introduce in the following. Let $(\varphi(t))_{t\in[0,1]} = (\int_0^t \sigma_u^2\, du)_{t\in[0,1]}$. Based on a fixed parameter $\alpha \in (0, \beta/2)$ we consider approximations $\hat{\varphi}^{(n)} = (\hat{\varphi}^{(n)}_t)_{t\in[0,1]}$ for $\varphi = (\varphi_t)_{t\in[0,1]}$ that are monotonically increasing, linear on each intervall $[i/n, (i+1)/n]$ $(i = 1, \dots, n)$ and satisfy

$$\hat{\varphi}^{(n)}(i/n) = \operatorname*{argmin}_{y\in\mathbb{I}(n)} |\varphi(i/n) - y| \qquad (i = 0, \dots, n),$$

where $\mathbb{I}(n)$ is defined as

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

Theorem 1.2. *Fix $p > 0$, $\alpha \in (0, \beta/2)$ and $\gamma_1 \in ((1+\alpha)^{-1}, 1)$. Moreover, let $\hat{\varphi}^{(n)}$ be as above, relate n and $r > 0$ via $n = n(r) = \lceil r^{\gamma_1} \rceil$.*
Then there exist $D[0,1]$-valued random elements $\bar{R}^{(n)}$, $\hat{R}^{(r)}$ and $\bar{W}^{(n)}$ such that

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

- $\log|\operatorname{range}(\hat{R}^{(r)}, \hat{\varphi}^{(n)})| = \mathcal{O}(r^\gamma)$, *for some* $\gamma \in (0,1)$,

The article is outlined as follows. In Section 2 we treat the asymptotic coding problem for the Wiener process. The new result is a slight extension of the main result of [3] which we will need later to prove the lower bound of the main theorem. In the proof of the main theorem we relate the coding problem for the diffusion to that of a concatenation of Wiener processes that is a process that is a concatenation of n independent Wiener processes on time intervals of length $1/n$ with possibly different diffusion coefficient on each time interval. The proof is based on a asymptotic analysis for the concatenations that is contained on Section 3. In Section 4 the article concludes with the proof of the main result.

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 $\mathcal{O}$.

2 Coding the Wiener processes

In order to prove the lower bounds we need to strengthen the high resolution estimates for the Wiener process from [3]:

Theorem 2.1. *For any* $s \in (0,\infty)$

$$D(r|W,s) \sim D^{(q)}(r|W,s) \sim K_p \frac{1}{\sqrt{r}}.$$

Remark 2.2. The proof works equally well when replacing W by a fractional Brownian motion.

Proof. It remains to show that for $s < p$

$$D(r|W,s) \gtrsim K_p \frac{1}{\sqrt{r}}.$$

Let $\hat{W}^{(1)} = \hat{W}^{(1,r)}$ $(r \geq 0)$ denote an arbitrary family of reconstructions in $D[0,1]$ such that $\mathbb{P}_{W,\hat{W}^{(1)}} \ll \mathbb{P}_W \otimes \mathbb{P}_{\hat{W}^{(1)}}$. Moreover, denote by $\mathcal{C}_r \subset D[0,1]$ codebooks with at most e^r elements that satisfy

$$\mathbb{E}[\min_{\hat{w}\in\mathcal{C}_r} \|W - \hat{w}\|^{2p}]^{1/2p} \lesssim K_p \frac{1}{\sqrt{r}},$$

and let

$$\hat{W}^{(2)} = \hat{W}^{(2,r)} = \operatorname*{argmin}_{\hat{w}\in\mathcal{C}_r} \|W - \hat{w}\|.$$

Next, let

$$J = \begin{cases} 1 & \text{if } \log \frac{d\mathbb{P}_{W,\hat{W}^{(1)}}}{d\mathbb{P}_W \otimes \mathbb{P}_{\hat{W}^{(1)}}} \leq r \text{ and } \|W - \hat{W}^{(1)}\| \leq (1-\varepsilon)\, K_p \frac{1}{\sqrt{r}} \\ 2 & \text{otherwise} \end{cases}$$

and consider the reconstruction $\hat{W} = \hat{W}^{(J)}$. Then

$$\begin{aligned} I(W;\hat{W}) &\le I(W;\hat{W},J) \le \log 2 + \int_{\{J=1\}} \log \frac{d\mathbb{P}_{W,\hat{W}^{(1)}}}{d\mathbb{P}_W \otimes \mathbb{P}_{\hat{W}^{(1)}}}\, d\mathbb{P} + \int_{\{J=2\}} \log \frac{d\mathbb{P}_{W,\hat{W}^{(2)}}}{d\mathbb{P}_W \otimes \mathbb{P}_{\hat{W}^{(2)}}}\, d\mathbb{P} \\ &\le \log 2 + \mathbb{P}(J=1)r + \mathbb{P}(J=2)r = r + \log 2. \end{aligned}$$

Due to the equivalence of norms in the quantization problem for the Wiener process (see (1)), it follows that the rescaled random error $\sqrt{r}\|W - \hat{W}^{(2)}\|$ converges to K_p in $L^{2p}(\mathbb{P})$ (see also the proof of Lemma A.1 in [3]). Consequently,

$$\mathbb{E}\|W - \hat{W}\|^p \lesssim (\mathbb{P}(J=1)(1-\varepsilon) + \mathbb{P}(J=2))\big(K_p \frac{1}{\sqrt{r}}\big)^p.$$

On the other hand, due to (1) the estimate $I(W;\hat{W}) \lesssim r$ implies that

$$\mathbb{E}\|W - \hat{W}\|^p \gtrsim \big(K_p \frac{1}{\sqrt{r}}\big)^p.$$

Consequently, $\lim_{r\to\infty} \mathbb{P}(J=1) = 0$. Let now $\mathcal{F}$ denote the family of $[0,\infty]^2$-valued random variables of the form

$$(A,B) = \Big(\|W - \hat{W}^{(1)}\|^s, \log_+ \frac{d\mathbb{P}_{W,\hat{W}^{(1)}}}{d\mathbb{P}_W \otimes \mathbb{P}_{\hat{W}^{(1)}}}\Big),$$

where $\hat{W}^{(1)}$ is an arbitary $D[0,1]$-valued random variable such that $\mathbb{P}_{W,\hat{W}^{(1)}} \ll \mathbb{P}_W \otimes \mathbb{P}_{\hat{W}^{(1)}}$. The above argument implies that

$$\lim_{r\to\infty} \sup_{(A,B)\in\mathcal{F}} \mathbb{P}\big(A \le \big((1-\varepsilon)K_p \frac{1}{\sqrt{r}}\big)^s, B \le r\big) = 0.$$

Note that for general $\hat{W}^{(1)}$ it is true that

$$\begin{aligned} I(W;\hat{W}^{(1)}) &= \int \frac{d\mathbb{P}_{W,\hat{W}^{(1)}}}{d\mathbb{P}_W \otimes \mathbb{P}_{\hat{W}^{(1)}}} \log \frac{d\mathbb{P}_{W,\hat{W}^{(1)}}}{d\mathbb{P}_W \otimes \mathbb{P}_{\hat{W}^{(1)}}}\, d\mathbb{P}_W \otimes \mathbb{P}_{\hat{W}^{(1)}} \\ &\ge \int \frac{d\mathbb{P}_{W,\hat{W}^{(1)}}}{d\mathbb{P}_W \otimes \mathbb{P}_{\hat{W}^{(1)}}} \log_+ \frac{d\mathbb{P}_{W,\hat{W}^{(1)}}}{d\mathbb{P}_W \otimes \mathbb{P}_{\hat{W}^{(1)}}}\, d\mathbb{P}_W \otimes \mathbb{P}_{\hat{W}^{(1)}} - \frac{1}{e} \end{aligned}$$

so that an application of Lemma A.3 of [3] implies that

$$D(r|W,s)^s \ge \inf_{\substack{(A,B)\in\mathcal{F}: \\ \mathbb{E}B \le r+1/e}} \mathbb{E}A \gtrsim \big((1-\varepsilon)K_p \frac{1}{\sqrt{r}}\big)^s,$$

and the assertion follows since $\varepsilon > 0$ is arbitrary. □

3 Concatenation of Wiener processes under $L^p[0,1]$-distortion

Now we treat the coding problem for concatenations of Wiener processes with non-constant diffusion coefficients. For $n \in \mathbb{N}$, let $(Z_t^{(i)})_{t\in[0,1/n)}$ $(i = 0,\dots,n-1)$ denote independent Wiener processes with diffusion coefficients σ_i and set

$$Y_{t+i/n} = Z_t^{(i)}$$

for $i = 0,\dots,n-1$ and $t \in [0,1/n)$. We shall call the process $Y = (Y_t)_{t\in[0,1)}$ a (σ_i)-concatenation of Wiener processes or short (σ_i)-concatenation.

Lemma 3.1. *For fixed $s \in (0,\infty)$ there exists a function $g = g_{s,p} : \mathbb{R}_+ \to \mathbb{R}_+$ with $\lim_{t\to\infty} g(t) = 1$ such that the following statement is valid.*

For a $(\sigma_i)_{i=0,\dots,n-1}$-concatenation Y and an $\mathbb{R}_+$-valued vector $(r_i)_{i=0,\dots,n-1}$, there exists a codebook $\mathcal{C} \subset D[0,1)$ with $\log|\mathcal{C}| \le \sum_{i=0}^{n-1} r_i$ and

$$\mathbb{E}[\min_{\hat y\in\mathcal{C}} \|Y-\hat y\|^s]^{1/s} \le g(r_*)\, K_p \Big(\frac{1}{n}\sum_{i=0}^{n-1} \frac{|\sigma_i|^p}{(nr_i)^{p/2}}\Big)^{1/p},$$

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

Proof. Due to the monotonicity of the coding error in the moment s, it suffices to consider the case $s \ge p$. For $\tilde r \ge 0$, $\tilde s > 0$, $\sigma \in \mathbb{R}$ and $T \in [0,1]$ let

$$f(\tilde r,\tilde s;\sigma,T) = D^{(q)}(\tilde r|\sigma W,[0,T],\tilde s) \text{ and } f(\tilde r,\tilde s) = D^{(q)}(\tilde r|W,\tilde s),$$

and set

$$h(r_*) = \sup_{\tilde r\ge r_*} \max\Big(\frac{f(\tilde r,s)}{K_p/\sqrt{\tilde r}}, \frac{f(\tilde r,s)}{f(\tilde r,p)}\Big)\Big(1+\frac{1}{r_*}\Big).$$

Due to Theorem 2.1 one has $\lim_{r_*\to\infty} h(r_*) = 1$. Now fix $n \in \mathbb{N}$, (σ_i) and (r_i) and let $r_* = \min_{i=0,\dots,n-1} r_i$. Furthermore, fix codebooks $\mathcal{C}_i \subset D[0,1/n)$ of size e^{r_i} with

$$\mathbb{E}[\min_{\hat z\in\mathcal{C}_i} \|Z^{(i)}-\hat z\|^s_{L^p[0,1/n)}]^{1/s} \le \Big(1+\frac{1}{r_i}\Big) f(r_i,s;\sigma,1/n).$$

Based on the codebooks $\mathcal{C}_i$ we define

$$\mathcal{C} = \mathcal{C}_0 * \cdots * \mathcal{C}_{n-1},$$

where $*$ denotes the concatenation of the functions of the codebooks. Certainly, the codebook $\mathcal{C}$ satisfies

$$\log|\mathcal{C}| \le \sum_{i=0}^{n-1} r_i.$$

Next, denote by $\hat Y$ an $L^p[0,1)$ optimal approximation from $\mathcal{C}$ for Y and denote

$$\Delta_i = \mathbb{E}\Big[\int_{i/n}^{(i+1)/n} |Y_t-\hat Y_t|^p\,ds\Big]^{1/p} \quad\text{and}\quad \Delta = \Big(\sum_{i=1}^{n-1} \Delta_i^p\Big)^{1/p} = \mathbb{E}[\|Y-\hat Y\|^p_{L^p[0,1)}]^{1/p}$$

Due to Jensen's inequality

$$\begin{aligned}\mathbb{E}[\|Y-\hat{Y}\|^s_{L^p[0,1)}] &= \Delta^s\,\mathbb{E}\Big[\Big(\sum_{i=0}^{n-1}\frac{\Delta_i^p}{\Delta^p}\frac{1}{\Delta_i^p}\int_{i/n}^{(i+1)/n}|Y_t-\hat{Y}_t|^p\,dt\Big)^{s/p}\Big]\\ &\le \Delta^s\,\mathbb{E}\Big[\sum_{i=0}^{n-1}\frac{\Delta_i^p}{\Delta^p}\Big(\frac{1}{\Delta_i^p}\int_{i/n}^{(i+1)/n}|Y_t-\hat{Y}_t|^p\,dt\Big)^{s/p}\Big]\\ &\le \Delta^s\sum_{i=0}^{n-1}\frac{\Delta_i^p}{\Delta^p}\Big(\Big(1+\frac{1}{r_*}\Big)\frac{f(r_i,s;\sigma_i,1/n)}{\Delta_i}\Big)^s.\end{aligned} \tag{4}$$

As a consequence of a standard scaling argument one obtains that in general

$$f(\tilde{r},\tilde{s};\sigma,T)=|\sigma|\,T^{(2+p)/2p}f(\tilde{r},\tilde{s}), \tag{5}$$

so that

$$\frac{f(r_i,s;\sigma_i,1/n)}{\Delta_i}\le\frac{f(r_i,s)}{f(r_i,p)}.$$

Consequently, equation (4) gives

$$\mathbb{E}[\|Y-\hat{Y}\|^s_{L^p[0,1)}]^{1/s}\le h(r_*)\,\Delta,$$

and together with

$$\begin{aligned}\Delta_i &\le \Big(1+\frac{1}{r_i}\Big)f(r_i,s;\sigma_i,1/n)=\Big(1+\frac{1}{r_i}\Big)|\sigma_i|\frac{1}{n^{(2+p)/2p}}f(r_i,s)\\ &\le h(r_*)|\sigma_i|\frac{1}{n^{1/p}}\frac{K_p}{\sqrt{nr_i}}\end{aligned}$$

we arrive at

$$\mathbb{E}[\|Y-\hat{Y}\|^s_{L^p[0,1)}]^{1/s}\le h(r_*)^2\,K_p\,\Big(\frac{1}{n}\sum_{i=0}^{n-1}\frac{|\sigma_i|^p}{(nr_i)^{p/2}}\Big)^{1/p}.$$

□

Next, we derive a converse estimate for concatenations of Wiener processes.

Lemma 3.2. *For fixed $s\in(0,\infty)$ there exists a real valued function $g=g_{s.p}:\mathbb{R}_+\to\mathbb{R}_+$ with $\lim_{r_*\to\infty}g(r_*)=1$ such that the following statement is valid.*

Let Y be a (σ_i)-concatenation, and let $\hat{Y}$ denote some reconstruction with $I(Y;\hat{Y})\le r$. Then there exists an $[0,\infty)$-valued sequence $(r_i)_{i=0,\dots,n-1}$ with $\sum_{i=0}^{n-1}r_i\le r$ such that for any $r_*>0$:

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

The proof uses the concept of conditional mutual information (see for instance [4] for basic results): For three random elements A,B and C attaining values in standard measurable spaces one defines

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

which is uniquely defined up to $\mathbb{P}_C$-null sets, and lets

$$I(A;B|C)=\int I(A;B|C=c)\,\mathbb{P}_C(dc).$$

Proof. It suffices to prove the assertion for $s\in(0,p]$. For $\tilde r\ge 0$, $\tilde s>0$, $\sigma\in\mathbb{R}$ and $T\ge 0$ let

$$f(\tilde r,\tilde s;\sigma,T)=D(\tilde r|\sigma W,[0,T],\tilde s) \text{ and } f(\tilde r,\tilde s)=D(\tilde r|W,\tilde s),$$

and set

$$h(r_*)=\inf_{\tilde r\ge r_*}\min\Big(\frac{f(\tilde r,s)}{f(\tilde r,p)},\frac{f(\tilde r,p)}{K_p\sqrt{\tilde r}}\Big).$$

Note that $\lim_{r_*\to\infty}h(r_*)=1$. Next, represent $\hat Y$ as the concatenation of n processes $\hat Z^{(0)},\dots,\hat Z^{(n-1)}$, denote $R_i(z^{(i+1)},\dots,z^{(n-1)})=I(Z^{(i)};\hat Z^{(i)}|Z^{(i+1)}=z^{(i+1)},\dots,Z^{(n-1)}=z^{(n-1)})$ and let $r_i=\mathbb{E}R_i(Z^{(i+1)},\dots,Z^{(n-1)})$. From the independence of the sequence $(Z^{(i)})_{i=0,\dots,n-1}$ and the convexity of the distortion rate function (see for instance [4], Theorem 1.7.1), it follows that

$$\begin{aligned}\mathbb{E}[\|Z^{(i)}-\hat Z^{(i)}\|^s_{L^p[0,1/n)}]&=\mathbb{E}\big[\mathbb{E}[\|Z^{(i)}-\hat Z^{(i)}\|^s_{L^p[0,1/n)}|Z^{(i+1)},\dots,Z^{(n-1)}]\big]\\&\ge\mathbb{E}\big[f(R(Z^{(i+1)},\dots,Z^{(n-1)}),s;\sigma_i,1/n)^s\big]\ge f(r_i,s;\sigma_i,1/n)^s.\end{aligned}$$

Moreover, for $i=0,\dots,n-1$ let $\Delta_i=f(r_i+r_*,p;\sigma_i,1/n)$ and $\Delta=\big(\sum_{i=0}^{n-1}\Delta_i^p\big)^{1/p}$. As in (4) one gets that

$$\mathbb{E}[\|Y-\hat Y\|^s]\ge\Delta^s\sum_{i=0}^{n-1}\frac{\Delta_i^p}{\Delta^p}\Big(\frac{1}{\Delta_i^s}\mathbb{E}[\|Y-\hat Y\|^s_{L^p[i/n,(i+1)/n)}]\Big)^s.$$

The same scaling argument from before gives that in general

$$f(\tilde r,\tilde s;\sigma,T)=|\sigma|\,T^{(2+p)/2p}f(\tilde r,\tilde s),$$

so that $\mathbb{E}[\|Y-\hat Y\|^s_{L^p[i/n,(i+1)/n)}]/\Delta_i^s\ge h(r_*)^s$. Therefore,

$$\mathbb{E}[\|Y-\hat Y\|^s]^{1/s}\ge h(r_*)\Delta\ge h(r_*)^2K_p\Big(\frac1n\sum_{i=0}^{n-1}\frac{|\sigma_i|^p}{(n(r_i+r_*))^{p/2}}\Big)^{1/p}.$$

It remains to show that $\sum_{i=0}^{n-1}r_i\le I(Y;\hat Y)$: one has

$$\begin{aligned}I(Y;\hat Y)=I(Z^{(0)},\dots,Z^{(n-1)};\hat Z^{(0)},\dots,\hat Z^{(n-1)})&=I(Z^{(1)},\dots,Z^{(n-1)};\hat Z^{(0)},\dots,\hat Z^{(n-1)})\\&\quad+I(Z^{(0)};\hat Z^{(0)},\dots,\hat Z^{(n-1)}|Z^{(1)},\dots,Z^{(n-1)})\\&\ge I(Z^{(1)},\dots,Z^{(n-1)};\hat Z^{(1)},\dots,\hat Z^{(n-1)})\\&\quad+I(Z^{(0)};\hat Z^{(0)}|Z^{(1)},\dots,Z^{(n-1)})\end{aligned}$$

so that by induction

$$r=I(Y;\hat Y)\ge\sum_{i=0}^{n-1}I(Z^{(i)};\hat Z^{(i)}|Z^{(i+1)},\dots,Z^{(n-1)})=\sum_{i=0}^{n-1}r_i.$$

□

Next, we consider the rate allocation problem for concatenations of Wiener processes. It amounts to studying the convex minimization problem

$$\Big(\frac{1}{n}\sum_{i=0}^{n-1}\frac{|\sigma_i|^p}{(nr_i)^{p/2}}\Big)^{1/p} = \min! \tag{6}$$

under the constraint that $\sum_{i=0}^{n-1} r_i \le r$ for some given rate $r > 0$.

Lemma 3.3. *The minimum in (6) is attained for*

$$r_i = \frac{|\sigma_i|^{2p/(p+2)}}{\sum_{j=0}^{n-1}|\sigma_j|^{2p/(p+2)}}\, r,$$

and it is equal to

$$\Big(\frac{1}{n}\sum_{i=0}^{n-1}|\sigma_i|^{2p/(p+2)}\Big)^{(p+2)/2p}\frac{1}{\sqrt{r}}.$$

Proof. Applying the Hölder inequality (for negative exponents) for $a = -2/p$ and $a^* = 2/(p+2)$ gives

$$\frac{1}{n}\sum_{i=0}^{n-1}\frac{|\sigma_i|^p}{(nr_i)^{p/2}} \ge \Big(\frac{1}{n}\sum_{i=0}^{n-1}|\sigma_i|^{2p/(p+2)}\Big)^{(p+2)/2}\Big(\sum_{i=0}^{n-1} r_i\Big)^{-p/2},$$

and it is now straightforward to verify that the minimum is attained for the above (r_i). □

4 Proof of Theorem 1.1

In this section we prove Theorem 1.1. We will need the notion of *conditional entropy.* For two discrete r.v.'s Z and G, let

$$H(Z|G=g) = \mathbb{E}[\log 1/p_{Z|g}|G=g] \text{ and } H(Z|G) = \mathbb{E}[\log 1/p_{Z|G}],$$

where $p_{z|g}$ denotes the conditional probability $\mathbb{P}(Z=z|G=g)$, which is well defined for $\mathbb{P}_G$-a.a. g. For basic properties of the conditional entropy one might consult [4].

In the rest of this section $s > 0$, $\alpha \in (0, \beta/2)$ and $\gamma_1 \in ((1+\alpha)^{-1}, 1)$ are fixed. Moreover, relate n and $r > 0$ via $n = \lceil r^{\gamma_1}\rceil$ and let $\hat\varphi = \hat\varphi^{(n)}$, $\bar W = \bar W^{(n)}$, $\bar R = \bar R^{(n)}$, $\hat R = \hat R^{(r)}$ be as in Theorem 1.2. For simplicity we omit the parameters n and r in the notations for the stochastic processes. We first turn to be proof of the upper bounds.

Proof of the upper bounds. For $i = 0, \dots, n-1$ and $t \in [i/n, (i+1)/n)$ let

$$Y_t = Y_t^{(n)} = \bar W_{\hat\varphi(t)} - \bar W_{\hat\varphi(i/n)} \text{ and } S_t = S_t^{(n)} = \bar W_{\hat\varphi(t)}$$

and write $\bar W$ as the sum

$$\bar W = Y + S.$$

We start with introducing a coding scheme for S. Let $\mathbb{J} = \mathbb{J}^{(r)} = \frac{1}{r}\mathbb{Z} \cap [-r, r]$ and denote by $\hat{S} = \hat{S}^{(r)}$ a reconstruction for S that is piecewise constant on the time intervals $[i/n, (i+1)/n)$ and satisfies

$$\hat{S}_{i/n} = \operatorname*{argmin}_{x \in \mathbb{J}} |S_{i/n} - x|.$$

Note that

$$\log |\operatorname{range}(\hat{S})| \le n \log(r^2 + 1) = \mathcal{O}(r^{(1+\gamma_1)/2}),$$

and that by straightforward computations:

$$\mathbb{E}[\|S - \hat{S}\|^s]^{1/s} \lesssim \frac{1}{r}.$$

Now consider the coding scheme for Y in the quantization setting. For $i = 0, \dots, n-1$ denote $\hat{\sigma}_i = [n(\hat{\varphi}((i+1)/n) - \hat{\varphi}(i/n))]^{1/2}$, and observe that given $\hat{\varphi}$ the process Y is a $(\hat{\sigma}_i)$-concatenation. Choose $\gamma_2 > 0$ with $\gamma_1 + \gamma_2 < 1$. Due to Lemma 3.1 there exist approximations $\hat{Y} = \hat{Y}^{(r)}$ such that conditional upon $\hat{\varphi}$ the r.v. $\hat{Y}$ attains at most $\exp\{r + nr^{\gamma_2}\}$ different values and satisfies

$$\mathbb{E}[\|Y - \hat{Y}\|^s|\hat{\varphi}]^{1/s} \le g(r^{\gamma_2}) K_p Z_n \frac{1}{\sqrt{r}}, \tag{7}$$

where Z_n is defined as $Z_n = \big(\frac{1}{n}\sum_{i=0}^{n-1} |\hat{\sigma}_i|^{2p/(p+2)}\big)^{(p+2)/2p}$ and g is as in the lemma. Next, define $\bar{\sigma}_t = \hat{\sigma}_i$ for $t \in [i/n, (i+1)/n)$ and $i = 0, \dots, n-1$ and rewrite Z_n in terms of the process $(\bar{\sigma}_t)_{t \in [0,1)}$ as

$$Z_n = \Big(\int_0^1 |\bar{\sigma}_t|^{2p/(p+2)}\, dt\Big)^{(p+2)/2p}.$$

The definition of $\hat{\varphi}$ implies that the process $(\bar{\sigma}_t)$ converges pointwise to (σ_t). Moreover, one has

$$\|\bar{\sigma}_\cdot\|_{L^\infty[0,1]} \le \|\sigma_\cdot\|_{L^\infty[0,1]} + 2$$

so that the dominated convergence theorem applied to (7) implies

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

We consider the process $\hat{X} := \hat{X}^{(r)} = \hat{Y} + \hat{S} + \hat{R}$ as reconstruction for X. Certainly, $\log|\operatorname{range}(\hat{X})| \lesssim r$. Moreover, if $s \ge 1$ the triangle inequality gives

$$\mathbb{E}[\|X - \hat{X}\|^s]^{1/s} \le \mathbb{E}[\|Y - \hat{Y}\|^s]^{1/s} + \mathbb{E}[\|S - \hat{S}\|^s]^{1/s} + \mathbb{E}[\|R - \hat{R}\|^s]^{1/s} \lesssim K_p \Big\| \|\sigma_\cdot\|^s_{L^{2p/(p+2)}[0,1]} \Big\|_{L^s(\mathbb{P})} \frac{1}{\sqrt{r}}.$$

The case $s < 1$ follows by the estimation

$$\mathbb{E}[\|X - \hat{X}\|^s] \le \mathbb{E}[\|Y - \hat{Y}\|^s] + \mathbb{E}[\|S - \hat{S}\|^s] + \mathbb{E}[\|R - \hat{R}\|^s] \lesssim K_p^s \Big\| \|\sigma_\cdot\|^s_{L^{2p/(p+2)}[0,1]} \Big\|^s_{L^s(\mathbb{P})} \frac{1}{\sqrt{r}^s}.$$

In the entropy coding setting the reconstruction $\hat{Y} = \hat{Y}^{(r)}$ is chosen such that

$$\log|\operatorname{range}(\hat{Y}|\hat{\varphi})| \le \frac{Z_n^{2s/(s+2)}}{\mathbb{E} Z_n^{2s/(s+2)}} + nr^{\gamma_2}$$

and

$$\mathbb{E}[\|Y-\hat{Y}\|^s|\hat{\varphi}]^{1/s} \le g(r^{\gamma_2}) K_p Z_n \Big(\frac{Z_n^{2s/(s+2)}}{\mathbb{E} Z_n^{2s/(s+2)}} r\Big)^{-1/2}.$$

Then

$$\mathbb{E}[\|Y-\hat{Y}\|^s]^{1/s} \le g(r^{\gamma_2}) K_p \mathbb{E}[Z_n^{2s/(s+2)}]^{(s+2)/2s} \frac{1}{\sqrt{r}}$$

and

$$\mathbb{H}(\hat{Y}, \hat{S}, \hat{R}, \hat{\varphi}) \le \underbrace{\mathbb{H}(\hat{Y}|\hat{\varphi})}_{\le r + n r^{\gamma_2}} + \underbrace{\mathbb{H}(\hat{S}, \hat{R}, \hat{\varphi})}_{=\mathcal{O}(r^{\gamma_3})}$$

for some $\gamma_3 < 1$. Therefore, $\mathbb{H}(\hat{X}) \lesssim r$ and again by dominated convergence

$$\mathbb{E}[\|Y-\hat{Y}\|^s]^{1/s} \lesssim K_p \big\| \|\sigma.\|_{L^{2p/(p+2)}[0,1]} \big\|_{L^{2s/(s+2)}(\mathbb{P})} \frac{1}{\sqrt{r}}.$$

□

Proof of the lower bounds. Let $\hat{X} = \hat{X}^{(r)}$ denote arbitrary reconstructions for X satisfying the quantization constraint $\log|\,\mathrm{range}\,(\hat{X})| \le r$ and set

$$\hat{Y}_t = \hat{X}_t - \hat{S}_t - \hat{R}_t$$

where $\hat{S} = \hat{S}^{(r)}$ and $\hat{R} = \hat{R}^{(r)}$ are as in the proof of the converse inequality. Then

$$\tilde{r} = \tilde{r}(r) = \log|\,\mathrm{range}\,(\hat{Y})| \le r + \mathcal{O}(r^{\gamma_4})$$

for some appropriate constant $\gamma_4 < 1$. Again we let $\gamma_2 \in (0, \gamma_1)$ and $Z_n = \big(\frac{1}{n}\sum_{i=0}^{n-1} |\hat{\sigma}_i|^{2p/(p+2)}\big)^{(p+2)/2p}$. Observe that conditional on $\hat{\varphi}$ the process Y is a $(\hat{\sigma}_i)$-concatenation where $(\hat{\sigma}_i)_{i=0,\dots,n-1}$ is defined as before. Hence, Lemmas 3.2 and 3.3 imply that

$$\mathbb{E}[\|Y-\hat{Y}\|^s|\hat{\varphi}]^{1/s} \ge g(r^{\gamma_2}) K_p Z_n \frac{1}{\sqrt{\tilde{r} + n r^{\gamma_2}}}$$

so that

$$\mathbb{E}[\|Y-\hat{Y}\|^s]^{1/s} \ge g(r^{\gamma_2}) K_p \mathbb{E}[Z_n^s]^{1/s} \frac{1}{\sqrt{\tilde{r} + n r^{\gamma_2}}} \gtrsim K_p \big\| \|\sigma.\|_{L^{2p/(p+2)}[0,1]} \big\|_{L^s(\mathbb{P})} \frac{1}{\sqrt{r}}.$$

Thus the lower bound follows with

$$\mathbb{E}[\|X-\hat{X}\|^s]^{1/s} \ge \mathbb{E}[\|Y-\hat{Y}\|^s]^{1/s} - \mathbb{E}[\|S-\hat{S}\|^s]^{1/s} - \mathbb{E}[\|\bar{R}-\hat{R}\|^s]^{1/s}$$

It remains to prove the lower bound for the distortion rate function. Let now $\hat{X} = \hat{X}^{(r)}$ denote arbitrary reconstructions with $I(X;\hat{X}) \le r$ and let $\hat{Y}_t = \hat{X}_t - \hat{S}_t - \hat{R}_t$ Then

$$I(Y;\hat{Y}) \le I(Y;\hat{X},\hat{S},\hat{R},\hat{\varphi}) \le I(Y;\hat{X}) + \mathbb{H}(\hat{S},\hat{R},\hat{\varphi})$$

and

$$I(Y;\hat{X}) \le I(X,Y;\hat{X}) = I(X;\hat{X}) + I(Y;\hat{X}|X) = I(X;\hat{X}) \le r,$$

since Y is $\sigma(X)$-measurable. Hence

$$\tilde{r}(\xi) = I(Y; \hat{Y} | \hat{\varphi} = \xi) + nr^{\gamma_2}$$

satisfies $\mathbb{E}\tilde{r}(\hat{\varphi}) \leq I(Y; \hat{Y}) + nr^{\gamma_2} \lesssim r$. Moreover,

$$\mathbb{E}[\|Y - \hat{Y}\|^s | \hat{\varphi}]^{1/s} \geq g(r^{\gamma_2}) K_p Z_n \frac{1}{\sqrt{\tilde{r}(\hat{\varphi})}}$$

so that

$$\mathbb{E}[\|Y - \hat{Y}\|^s]^{1/s} \geq g(r^{\gamma_2}) K_p \mathbb{E}\Big[\Big(Z_n \frac{1}{\sqrt{\tilde{r}(\hat{\varphi})}}\Big)^s\Big]^{1/s}$$

Applying the inverse Hölder inequality as in Lemma 3.3 leads to

$$\mathbb{E}[\|Y - \hat{Y}\|^s]^{1/s} \geq g(r^{\gamma_2}) K_p \mathbb{E}[Z_n^{2s/(s+2)}]^{(s+2)/2s} \frac{1}{\sqrt{\mathbb{E}\tilde{r}(\hat{\varphi})}} \gtrsim K_p \left\| \|\sigma_\cdot\|_{L^{2p/(p+2)}[0,1]} \right\|_{L^{2s/(s+2)}(\mathbb{P})}.$$

□

References

[1] Steffen Dereich. 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, 2003.

[2] Steffen Dereich. The coding complexity of diffusion processes under supremum norm distortion. Preprint, 2006.

[3] Steffen Dereich and Michael Scheutzow. High-resolution quantization and entropy coding for fractional Brownian motion. To appear in Electronic Journal of Probability, 2006.

[4] Shunsuke Ihara. *Information theory for continuous systems.* Singapore: World Scientific, 1993.

[5] Harald Luschgy and Gilles Pagès. Sharp asymptotics of the functional quantization problem for Gaussian processes. *Ann. Probab.*, 32(2):1574–1599, 2004.

[6] Bernt Øksendal. *Stochastic differential equations.* Springer-Verlag, Berlin, sixth edition, 2003.