Coding of Brownian Motion by Quantization of Exit Times

Felix Poloczek and Florin Ciucu
T-Labs / TU Berlin
{felix,florin}@net.t-labs.tu-berlin.de

Abstract—We present a concrete coding scheme for standard linear Brownian motion. The coding error under sup-norm decays like $O(r^{-\frac{3}{2}})$ in the entropy $r$, and is thus rate-optimal. Moreover, the multiplicative constant lies below the theoretical maximum, and thus improves the known results. The scheme is based on a novel technique for quantizing the exit times of Brownian motion within a random error.

I. INTRODUCTION

A fundamental problem in information theory concerns the optimal (lossless) data compression, i.e., assigning codewords to the outcomes of a random source $X$ such that the average codeword length is minimal. This problem is completely solved by the Huffman code algorithm which yields the minimum as the entropy $H(X)$ ([4]).

In practice, however, the original source cannot be exactly encoded if the available memory or bandwidth is limited. This is typically the case for sources with infinite entropy, like continuously distributed random variables or stochastic processes. One important example of a source with infinite entropy is Brownian motion which was originally intended to model the trace of a particle moving randomly in space ([15]). More recently, Brownian motion has been applied to various other scientific fields, including the analysis of stock markets and networking ([16], [17]), and has arguably become the most important object in modern probability theory ([13]). This paper is concerned with coding Brownian motion.

In recent years, the coding complexity (i.e., approximability by simpler processes) of Brownian motion and related objects has been investigated ([2], [5], [6]). The focus has been primarily on determining the asymptotic behavior of the entropy coding error ([7])

$$D^{(c)}(r | X, s) := \inf \left\{ \sup_{t \in [0,1]} \left| X_t - \hat{X}_t \right|_{L^r} \right\},$$

where $s > 0$ is the moment under consideration and the infimum is taken over all stochastic processes $\hat{X}$ such that $\mathbb{H}_e(\hat{X}) \leq r$.

In the particular case of a Brownian motion $B$, an important result ([9]) is that

$$\lim_{r \to \infty} D^{(c)}(r | B, s) \sqrt{r} = \kappa,$$

where $\kappa$ is a constant independent of $s$. A further result is that $\kappa \in [\frac{\pi}{\sqrt{8}}, \pi]$ ([8]). These two results, however, are not accompanied by a concrete coding strategy.

In this paper we present such a concrete coding strategy $\hat{B}$ for a Brownian motion $B$. Our strategy is not only rate-optimal, but it also reduces the interval for $\kappa$ to

$$\kappa \in \left[ \frac{\pi}{\sqrt{8}}, 2.11 \right],$$

by explicitly bounding $\kappa \leq 2.11$.2

The main idea of our coding strategy is to represent $\hat{B}$ in terms of piecewise constant paths whose jumping points are the exit times of $B$ of certain intervals. The key observation is that $B$ is bounded between two consecutive exit times, and hence the error $|B_t - \hat{B}_t|$ is bounded as well. We point out, however, that the exit times are continuously distributed and cannot be encoded exactly.

In order to deal with this technical problem we must resort to quantized versions, i.e., discrete approximations $D$ of the exit times $X$. In order to quantize random variables, a typical method is to minimize

$$\mathbb{E}[d(X, D)]$$

(for an appropriate metric $d$), subject to constraints of the form ([11], [12])

$$\text{rng} D \leq n, \quad \text{or} \quad \mathbb{H}(D) \leq H.$$

Unfortunately, such a method fails in our context. The reason is that the maximum of Brownian motion has an unbounded distribution in every (deterministic) interval of positive length, and hence $|B_t - \hat{B}_t|$ cannot be bounded (a.s.) by any deterministic quantization error. Our crucial idea is to suitably construct a random variable $Y \geq 0$ (which is also an exit time and independent of $X$), and which plays the role of a random error. In this way, by bounding Brownian

---

1$\mathbb{H}_e$ and $\mathbb{H}$ denote the entropy measured in nats and bits, respectively; similarly, $\ln := \log_e$ and $\log := \log_2$.

2Another coding strategy was recently proposed in [1] with $\kappa \leq 5.19$, and hence outside the interval established in [8].
motion with a (random) interval, the supremum error of the approximation $\hat{B}$ can also be bounded.

In more concrete terms, given two independent random variables $X$ and $Y \geq 0$, our goal is to define a discrete random variable $D$ with minimal finite entropy such that

$$X \leq D \leq X + Y \quad \text{a.s.} \quad (2)$$

The rest of this paper is organized as follows. In Section II we develop solutions for the quantization problem introduced in Eq. (2). We then use these solutions in Section III for our proposed coding strategy of Brownian motion. Finally, we present some brief conclusions in Section IV.

II. QUANTIZATION OF RANDOM VARIABLES

It is possible that no $D$ satisfying Eq. (2) with finite entropy exists. For a quick example take $X$ integer-valued with $\mathbb{H}(X) = \infty$, and let $Y$ such that $Y < 1$ almost surely; note that $D$ exactly encodes $X$ and thus has infinite entropy as well. Therefore, we must enforce some conditions on $Y$.

We will show in particular that the conditions

$$\forall t \geq 0 : \mathbb{P}(Y \leq t) \leq \beta t^n \quad \text{and} \quad \mathbb{E}[|X|] < \infty$$

(for some $n \in \mathbb{N}$, $\beta > 0$) suffice for the existence of a $D$ with finite entropy and satisfying Eq. (2).

The following two scaling properties will be frequently used in the sequel:

1) For $c > 0$, $a \in \mathbb{R}$ the problem for $(cx, cy)$ and $(x + a, y)$ is just as hard as for $(X, Y)$ (simply take the random variables $cD$ and $a + D$ instead of $D$, which have the same entropy).

2) If $\mathbb{P}(Y_1 \leq t) \leq \mathbb{P}(Y_2 \leq t)$ for every $t \geq 0$, the problem for $(X, Y_1)$ is not harder than for $(X, Y_2)$ (by the independence assumption one may assume that $Y_2 \leq Y_1$ a.s., so that $X \leq D \leq X + Y_2 \leq X + Y_1$).

The second property is crucial: If the probability of $Y$ is too large, for small values of $t$, then the resulting entropy of $D$ is also large. We therefore confine ourselves to the case when this probability is bounded by $\mathbb{P}(Y \leq t) \leq \beta_n t^n$.

Among all distributions satisfying this constraint, the worst case is represented by those satisfying the above line with equality. It is sufficient to work with these worst-case distributions, which we denote for convenience by $\mathcal{M}_{\beta_n}$.

In the following we first consider the simpler case when $X$ is a bounded random variable, and then we consider the general case.

A. Bounded random variable $X$

In this section, in addition to $Y \sim \mathcal{M}_{\beta_n}$ (for some $n \in \mathbb{N}$ and $\beta_n > 0$), we also assume that $X$ is bounded on some interval $[0, \alpha]$.

Given $X$ and $Y$, the main idea to construct a quantized version $D$ of $X$ is as follows. First estimate $Y$ from below, i.e., find $d \leq Y$. Then divide the range of $X$ (i.e., $[0, \alpha]$) into subintervals of length $d$ and determine the one where $X$ lies in. The right endpoint of this interval is finally chosen as the value for $D$.

We introduce two integer-valued random variables $\hat{Y}$ and $\hat{X}_{\hat{Y}}$ corresponding to the first two steps:

$$\hat{Y} = i \Leftrightarrow Y \in [2^{-i}, 2^{-i+1})$$

$$\hat{X}_{\hat{Y}} = j \Leftrightarrow X \in [(j-1)2^{-\hat{Y}}, j2^{-\hat{Y}}), \quad i, j \geq 1.$$ 

With these random variables we can easily define $D$:

**Definition 1:** For $\hat{Y}$, $\hat{X}_{\hat{Y}}$ as above, define

$$D := \frac{\hat{X}_{\hat{Y}}}{2^{\hat{Y}}},$$

$D$ is the right endpoint of the interval in the definition of $\hat{X}_{\hat{Y}}$.

Some further simplifications are useful. By applying the previous scaling property 1) with $c = \sqrt{\beta_n}$, we obtain an equivalent problem with

$$\mathbb{P}(cY \leq t) = t^n \quad \text{and} \quad cX \in [0, \alpha \sqrt{\beta_n}].$$

Further, if $m \in \mathbb{Z}$ denotes the unique integer such that $2^{m-1} < \alpha \sqrt{\beta_n} \leq 2^m$, and applying both scaling properties, the quantization problem becomes harder when the range of $X$ is replaced by $[0, 2^m]$.

The advantage of this setting (i.e., $X \in [0, 2^m]$ and $Y \sim \mathcal{M}_{1}^\alpha$) is that it provides a “worst case” distribution for $X$: the entropy $\mathbb{H}(D)$ is maximal if $X$ is uniformly distributed. To see this, consider the entropy of $D$ conditioned on $\{Y = i\}$ (for $i \geq 1$):

$$\mathbb{H}(D | \hat{Y} = i) = \mathbb{H}(\hat{X}_{\hat{Y}} | \hat{Y} = i)$$

Clearly, if $X$ is uniformly distributed, so is $\hat{X}_{\hat{Y}} | \{\hat{Y} = i\}$. As the (discrete) uniform distribution maximizes the entropy among all distribution with the same (bounded) support, all the conditioned entropies $\mathbb{H}(D | \hat{Y} = i)$ and thus $\mathbb{H}(D)$ are maximal.

Therefore, for the rest of this section, we assume

$$Y \sim \mathcal{M}_{1}^\alpha \quad \text{and} \quad X \sim \mathcal{U}_{[0, 2^m]},$$

for some $n \geq 1$ and $m \in \mathbb{Z}$.

Suppose first the special case $m = -1$. We calculate the (joint) distribution of the corresponding $\hat{Y}$ and $\hat{X}_{\hat{Y}}$. For $\hat{Y}$ it clearly holds:

$$\mathbb{P}(\hat{Y} = i) = \mathbb{P}(Y \leq 2^{-i+1}) - \mathbb{P}(Y \leq 2^{-i}) = (2^n - 1)2^{-mi}.$$ 

Since $X$ and $Y$ are assumed to be independent, the conditioned random variable $\hat{X}_{\hat{Y}} | \{\hat{Y} = i\}$ is uniformly distributed

$$\mathbb{P}(\hat{X}_{\hat{Y}} = j) = (2^n - 1)2^{-(m+(n+1)i)}.$$ 

As $D$ is completely determined by the values of $\hat{Y}$ and $\hat{X}_{\hat{Y}}$, its entropy is bounded by $\mathbb{H}(D) \leq \mathbb{H}(\hat{Y}, \hat{X}_{\hat{Y}}) \leq
Clearly, \( \Xi = \mathbb{Y} + \mathbb{X} \). But as \( \Xi = \mathbb{Y} + \mathbb{X} \) cannot be reconstructed out of \( D \), this bound can be improved by calculating \( D \)'s distribution.

The range of \( D \) consists of all the dyadic rationals in the interval \([0, 2^n] \). We next introduce the following useful representation. Let \( R_k := \{ 2^{-k} \} \) and for \( k \geq 1 \)

\[
R_k := \{ l2^{-k-1} \mid l \text{ is odd, } l \leq 2^k \} .
\]

Clearly, \( \text{rng}(D) = \bigcup_{k \geq 0} R_k \) and the \( R_k \)'s are pairwise disjoint.

**Lemma 1:** Let \( Y \sim \mathcal{M}_1^n \), \( X \sim \mathcal{U}_{(0,2^{-1})} \). Then

\[
\mathbb{P}(D = r) = \begin{cases} \frac{2^n + 1}{2^n + 2} & r = \frac{1}{2} \\ \frac{2^n + 1}{2^n + 2} 2^{-2} & r \in R_k, k \geq 1. \end{cases}
\]

**Proof:** We have for \( r = 2^{-1} \):

\[
\mathbb{P}(D = 2^{-1}) = \sum_{i=1}^{\infty} \mathbb{P}(D = 2^{-1}, \hat{Y} = i) = \sum_{i=1}^{\infty} (2^n - 1) 2^{1-(n+1)i} = \frac{2^n + 1 - 2}{2^n + 1 - 1} .
\]

and for \( r = \frac{1}{2^k} \in R_k \) (\( k \geq 1 \), \( l \) odd, \( l \leq 2^k \)):

\[
\mathbb{P}(D = r) = \sum_{i=k+1}^{\infty} \mathbb{P}(D = r, \hat{Y} = i) = \sum_{i=k+1}^{\infty} (2^n - 1) 2^{1-(n+1)i} = \frac{2^n + 1 - 2}{2^n + 1 - 1} 2^{-(n+1)k} .
\]

Now it is easy to compute the entropy of \( D \):

**Lemma 2:** With the situation as in Lemma 1 the entropy of \( D \) is given by

\[
\mathbb{H}(D) = \Lambda_n - 1 := \log \frac{2^n + 1 - 2}{2^n - 1} + \frac{(n + 1)2^n}{(2^n - 1)(2^n + 1 - 1)} - 1.
\]

**Proof:** We have \( |R_0| = 1 \), and \( |R_k| = 2^{k-1} \) for \( k \geq 1 \). Therefore

\[
H_0 := -\mathbb{P}(D = 2^{-1}) \log \mathbb{P}(D = 2^{-1}) = \frac{2^n + 1 - 2}{2^n + 1 - 1} \log \frac{2^n + 1 - 1}{2^n + 1 - 2}
\]

and

\[
H_{\geq 1} := -\sum_{k=1}^{\infty} \sum_{r \in R_k} \mathbb{P}(D = r) \log \mathbb{P}(D = r) = \sum_{k=1}^{\infty} \frac{2^{k-1}2^n + 1 - 2}{2^n + 1 - 1} 2^{-(n+1)k} \log \frac{2^n + 1 - 1}{2^n + 2} 2^{(n+1)k}.
\]

where we used the well known fact that \( \sum_{i=1}^{\infty} ia^{-i} = \frac{a}{(a-1)^2} \). Clearly, \( \mathbb{H}(D) = H_0 + H_{\geq 1} \) and hence:

\[
\mathbb{H}(D) = \log \frac{2^n + 1 - 1}{2^n + 2} + \frac{(n + 1)2^n}{(2^n - 1)(2^n + 1 - 1)} .
\]

The result easily extends to the case \( m \geq 0 \). Concretely, by conditioning on the events \( \{ X \in [(j-1)2^{-1}, j2^{-1}] \} \), one obtains:

**Theorem 1:** Let \( m \geq -1 \), \( Y \sim \mathcal{M}_1^n \), \( X \sim \mathcal{U}_{(0,2^{-m})} \). Then for the entropy \( \mathbb{H}(D) \) it holds

\[
\mathbb{H}(D) = \Lambda_n + m .
\]

For the remaining case (i.e., \( m \leq -2 \)), let us remark that Definition 1 is inadequate. To see that we let \( m = -2 \) and consider two outcomes where \( Y \in \left[ \frac{1}{2}, 1 \right] \) (i.e., \( \hat{Y} = 1 \)) and \( Y \in \left[ \frac{1}{2}, \frac{1}{2} \right] \) (i.e., \( \hat{Y} = 0 \)). Since \( X \in [0,2^m] = [0, \frac{1}{4}] \), we have that \( \hat{X} = 1 \) in both cases. Although the value of \( 2^{-2} \) is possible for both outcomes, Definition 1 separates them by giving \( D = 2^{-1} \) and \( D = 2^{-2} \), respectively, which increases the entropy. To avoid this problem, we modify the slightly modify Definition 1:

**Definition 2:** For \( m \leq -2 \) and \( \hat{Y}, \hat{X} \) as above, let

\[
D := \begin{cases} \frac{2m}{2^m} \hat{Y} \leq |m| \\ \frac{\hat{X}}{2^m} \text{ otherwise} . \end{cases}
\]

Let us also introduce a new random variable \( B \), indicating which branch in the above definition is used:

\[
B := \begin{cases} +1 \hat{Y} \leq |m| \\ -1 \text{ otherwise} . \end{cases}
\]

Since \( \mathbb{P}(B = -1) = 1 \), \( \mathbb{P}(B = +1) = 2^m \), the conditional entropy \( \mathbb{H}(D \mid B) \) is given by

\[
\mathbb{H}(D \mid B) = (1 - 2^m) \mathbb{H}(D \mid B = +1) + 2^m \mathbb{H}(D \mid B = -1) .
\]

By definition, \( D \) is constant on \( \{ B = +1 \} \), and thus has entropy 0 on this set. In turn, for \( \{ B = -1 \} \), note that for \( t \leq 2^m \)

\[
\mathbb{P}(Y \leq t \mid B = -1) = 2^{-m} \mathbb{P}(Y \leq t) = 2^{-mn} .
\]

i.e., on \( \{ B = -1 \} \) it holds that \( X \sim \mathcal{U}_{(0,2^{-m})} \), \( Y \sim \mathcal{M}_{1,2^{-m}}^n \). By applying the scaling properties 1) and 2), we obtain an equivalence to the case when \( X \sim \mathcal{U}_{(0,1)} \) and \( Y \sim \mathcal{M}_{1}^n \). Hence, by Theorem 1:

\[
\mathbb{H}(D \mid B) = 2^m \Lambda_n .
\]

**Theorem 2:** If \( m \leq -2 \) the entropy of \( D \) is bounded by

\[
\mathbb{H}(D) \leq 2^m (2n |m| + \Lambda_n) .
\]
**Proof:** By the monotonicity of the log-function we get
\[
\mathbb{H}(B) = -2^{nm} \log 2^{nm} - (1 - 2^{nm}) \log(1 - 2^{nm}) \\
\leq -2(2^{nm} \log 2^{nm}) = 2^{nm} 2n \log \frac{1}{m}
\]
Applying the chain rule, \(\mathbb{H}(D) \leq \mathbb{H}(D, B) = \mathbb{H}(B) + \mathbb{H}(D \mid B)\) yields the result.

For large \(|m|\) this bound is quite sharp since \(\mathbb{H}(D \mid B) \leq \mathbb{H}(D)\) and the term \(2n \log |m|\) has a small influence compared to \(2^{nm}\).

Summarizing, for random variables \(X\) and \(Y\) such that \(X \in [0, \alpha]\) and \(\mathbb{P}(Y \leq t) \leq \beta_n t^n\) there is a solution \(D\) of the quantization problem for \((X, Y)\). Its entropy grows linearly in \(m := \lceil \log \alpha \sqrt[2n]{m} \rceil\) for \(m \to \infty\) (Theorem 1) and decays exponentially in \(m\) for \(m \to -\infty\) (Theorem 2).

### B. Unbounded random variable \(X\)

Here we still assume that \(Y \sim \mathcal{M}_\alpha^n\), but we no drop the condition that \(X\) is bounded. By the example above, without further conditions on \(X\), there is in general no \(D\) with finite entropy. Thus an analogue of the uniform distribution as a worst case distribution for \(X\) is lacking, and we cannot compute the entropy of the corresponding \(D\) directly.

Our strategy consists of, first, bounding \(X\) to some interval \(I\) and, second, applying the previously developed solution to \(Y\) and \(X \mid \{X \in I\}\). The entropy of \(D\) can then be estimated by the entropy of the bounding procedure for \(X\) plus the entropy as calculated in the previous section.

For a given length \(d > 0\), we let the random variable \(\hat{X}\) indicate in which interval \(X\) lies in, i.e.,
\[
\hat{X} = n \Leftrightarrow X \in [nd, (n+1)d), \quad n \in \mathbb{Z}.
\]

Next we modify the definition of \(\hat{X}_Y\) to
\[
\hat{X}_Y = j \Leftrightarrow X \in [d\hat{X} + (j-1)2^{-Y}, d\hat{X} + j2^{-Y})
\]
such that we can give a formal definition of \(D\):

**Definition 3:** With \(\hat{X}\) and \(\hat{X}_Y\) as above and \(\hat{Y}\) as before, define:
\[
D := d\hat{X} + \hat{X}_Y 2^{-\hat{Y}}.
\]
Again, \(D\) is the right endpoint of the interval in the definition of \(\hat{X}_Y\).

Applying the chain rule to \(D\) and \(\hat{X}\) leads to
\[
\mathbb{H}(D) \leq \mathbb{H}(\hat{X}) + \mathbb{H}(D \mid \hat{X}).
\]

As the random variable \(D \mid \hat{X} = n\) is bounded for \(n \geq 1\), we can apply Theorem 1. So all we have to do is to estimate \(\mathbb{H}(\hat{X})\). This will be done by the use of maximum entropy distributions. As mentioned, a universal \(X\) such that \(\mathbb{H}(X)\) is maximal does not exist. Therefore we consider only those \(X\) satisfying some moment constraint \(\mathbb{E}[h(X)] \leq \mu\) and determine the distribution which maximizes the entropy in that class.

As \(\hat{X} \leq \frac{X}{2}\), it holds for the moments that \(\mathbb{E}[|\hat{X}|^p] \leq \frac{\mathbb{E}[|X|^p]}{d^p}\). Since maximum entropy distributions are hard to compute in the discrete case, we follow an approach from [10] and define a continuous random variable \(\Xi\) by
\[
\Xi := \hat{X} + U,
\]
where \(U \sim U_{[0,1]}\) is independent of \(\hat{X}\). Clearly, the density of \(\Xi\) is given by \(f(t) = \mathbb{P}(\hat{X} = [t])\) and we have:
\[
\mathbb{H}(\Xi) = \mathbb{H}(\hat{X}) + \mathbb{H}(U).
\]

It is well known (see, e.g., [14]) that the distributions maximizing the entropy in \(C_p^n := \{Z \mid \mathbb{E}[|Z|^p] \leq \mu\}\), \(p = 1, 2\) are the centered Laplacian and normal distribution, respectively. Thus:
\[
\max\{\mathbb{H}(Z) \mid Z \in C_p^n\} = \begin{cases}
\log(2e\mu) & p = 1 \\
\frac{1}{2} \log(2\pi e\mu) & p = 2.
\end{cases}
\]

A straightforward calculation gives the relationship between \(\mathbb{E}[|\Xi|^p]\) and \(\mathbb{E}[|\hat{X}|^p]\):

**Lemma 3:** Let \(\hat{X}, \Xi\) as above, then:
\[
\mathbb{E}[|\Xi|^p] = \frac{1}{3} + \mathbb{E}[|\hat{X}|^p] + \mathbb{E}[|\hat{X}|^p] \leq \frac{1}{3} + \frac{\mathbb{E}[|X|^p]}{d} + \frac{\mathbb{E}[|X|^p]}{d^2}.
\]

In order to combine the preceding results with those of Section II-A we have to choose a specific value for \(d\). It seems reasonable to set \(d = 2^n\) for some \(m \in \mathbb{Z}\), which corresponds to the \(m\) from the previous section. We additionally assume \(m \geq -1\), as we have exact results only for this case. Then \(X \mid \{X = n\}\) is a random variable bounded in \([n2^n, (n+1)2^n)\), which is, due to the scaling properties, equivalent to an \(X\) being bounded in \([0, 2^m]\).

In order to find the optimal \(m \geq -1\), note that by applying Eq. (3) with the entropy in Theorem 1 and the maximum entropy Eq. (4) with \(\mu\) from Lemma 3, the entropy of our discrete \(D\) can be bounded by:
\[
\mathbb{H}(D) \leq \begin{cases}
\log(2e(\frac{1}{2} + \frac{\mathbb{E}[X]}{2\mathbb{E}[X]^2})) + \Lambda_n & p = 1 \\
\log(2\pi e(\frac{1}{2} + \frac{\mathbb{E}[X]}{2\mathbb{E}[X]^2}) + \frac{\mathbb{E}[X]^2}{4\mathbb{E}[X]^2}) + \Lambda_n & p = 2.
\end{cases}
\]
Replacing \(m\) by a continuous variable \(t \in [-1, \infty)\), it is easy to see that the derivative of the right hand side is \(> 0\) in both cases. Thus, it is monotonically increasing and therefore \(m = -1\) (or equivalently \(d = \frac{1}{2}\)) is always the best choice.

Combining the preceding results we have:

**Theorem 3:** Let \(p = 1, 2, X \in L^p\), and \(Y\) a random variable with \(\mathbb{P}(Y \leq t) \leq \beta_n t^n\), for some \(n \geq 1, \beta_n > 0\). Then there is a discrete \(D\) with \(X \leq D \leq X + Y\). Its entropy is bounded by:
\[
\mathbb{H}(D) \leq \begin{cases}
\Lambda_n + \frac{1}{2} \log \left( 2\pi e \left( \frac{1}{2} + 2\beta_n E[X] \right) \right) & p = 1 \\
\Lambda_n + \frac{1}{2} \log \left( 2\pi e \left( \frac{1}{2} + 2\beta_n E[X] + 4\beta_n^2 E[X^2] \right) \right) & p = 2.
\end{cases}
\]
Proof: By the scaling properties we may work with $Y \sim M^n_\tau$ and $\beta^n \tau X$ instead of the original $X$ and $Y$. Let $m = -1$ by the remark above. By Eq. (3), now add up the entropy from Lemma 2 and the maximum entropy in Eq. (4) with the moments from Lemma 3, which concludes the proof. ■

For the application from Section III, the encoding of Brownian motion, the special case $X \geq 0$ will be particularly important. Clearly, the corresponding maximum entropy distributions turn into its absolute value, i.e., the exponential and truncated normal. Thus, the entropy reduces by 1.

**Theorem 4.** Consider the hypotheses from Theorem 3, and restrict $X \geq 0$. Then the entropy $\mathbb{H}(D)$ is bounded by:

$$
\mathbb{H}(D) \leq \begin{cases} 
\Lambda_n + \log \left( \frac{e}{2} \left( \frac{1}{2} + 2\beta \frac{1}{n} \mathbb{E}|X| \right) \right) \\
\Lambda_n + \frac{1}{2} \log \left( \frac{e}{\sqrt{2\pi}} \left( \frac{1}{2} + 2\beta \frac{1}{n} \mathbb{E}|X| + 4\beta \frac{2}{n} \mathbb{E}|X^2| \right) \right)
\end{cases} .
$$

**III. CODING OF BROWNIAN MOTION**

Encoding Brownian motion $B$ on $[0,1]$ with an error of $\varepsilon$ is, due to the scaling property (of Brownian motion), the same as encoding $B$ on $[0,\frac{1}{\sqrt{\varepsilon}}]$ with an error of 1. We will approximate $(B_t)_{t \in [0,\frac{1}{\sqrt{\varepsilon}}]}$ by a cadlag process $\hat{B}$, whose paths are piecewise constant. Its jumping points will depend on the exit times of $B$ of intervals of size 1.

**Definition 4.** Let $(\tau_n)_{n \in \mathbb{N}}$ denote the exit times of Brownian motion on a grid of size 1, i.e., let

$$
\tau_0 := 0 \\
\tau_{n+1} := \inf \left\{ t \geq 0 \mid B_{\sum_{i=1}^n \tau_i + t} - B_{\sum_{i=1}^n \tau_i} = 1 \right\} , \quad n \geq 0
$$

Further, let $\xi_0 = 0$ and $\xi_{n+1} := B_{\sum_{i=1}^{n+1} \tau_i} - B_{\sum_{i=1}^n \tau_i} = \pm 1$ for $n \geq 0$, indicate whether the Brownian motion moves up or down at the specific time.

Suppose we have constructed discrete random variables $Z_i$ such that for every $n \in \mathbb{N}$

$$
\sum_{i=0}^n \tau_i \leq \sum_{i=0}^n Z_i \leq \sum_{i=0}^n \tau_i + \tau_{n+1} ,
$$

and its entropy is uniformly bounded by some $H \geq 0$. An equivalent definition will be given below. Given these $Z_i$’s we can directly define our process $\hat{B}$.

**Definition 5.** With the $\xi_n$ as in Definition 4, define the approximating process $\hat{B}$ by

$$
\hat{B}_t := \sum_{n=0}^{\infty} \frac{\xi_n + \xi_{n+1}}{2} I_{\sum_{i=0}^n Z_i, \infty}(t) .
$$

In every (random) interval $I^n := [\sum_{i=0}^n \tau_i, \sum_{i=0}^{n+1} \tau_i]$ at most one jump of $\hat{B}$ occurs. If we let

$$
I^n_l := \left[ \sum_{i=0}^n \tau_i, \sum_{i=0}^{n+1} Z_i \right] , \quad I^n_r := \left[ \sum_{i=0}^n Z_i, \sum_{i=0}^{n+1} \tau_i \right]
$$

the subintervals left and right of the jump, we have for every $n \in \mathbb{N}$

$$
\hat{B} \mid I^n_l = \frac{B_{\sum_{i=1}^{n-1} \tau_i} + B_{\sum_{i=1}^{n+1} \tau_i}}{2} ,
$$

$$
\hat{B} \mid I^n_r = \frac{B_{\sum_{i=1}^{n-1} \tau_i} + B_{\sum_{i=1}^{n+1} \tau_i}}{2} ,
$$

and clearly

$$
B \mid I^n \in (B_{\sum_{i=1}^{n-1} \tau_i} - 1, B_{\sum_{i=1}^{n+1} \tau_i} + 1)
$$

so that for the difference $B - \hat{B}$ it holds:

$$(B_t - \hat{B}_t) \mid I^n \in (-1, +1) + \frac{B_{\sum_{i=1}^{n-1} \tau_i} - B_{\sum_{i=1}^{n+1} \tau_i}}{2} , \quad (B_t - \hat{B}_t) \mid I^n \in (-1, +1) + \frac{B_{\sum_{i=1}^{n-1} \tau_i} - B_{\sum_{i=1}^{n+1} \tau_i}}{2} .
$$

Thus we have shown the following upper bound for the supremum error of our coding strategy $\hat{B}$:

**Lemma 4.** For the coding error it holds that $\mathbb{P}_{t \geq 0} |B_t - \hat{B}_t| \leq \frac{3}{2}$ almost surely, and hence for any $s > 0$:

$$
\left\| \sup_{t \in [0,1]} (B_t - \hat{B}_t) \right\| \leq \frac{3}{2} .
$$

In Figure 1 a sample path of Brownian motion $B$ and the corresponding path of $\hat{B}$ are shown. We see that the process $\hat{B}$ does not necessarily jump in every interval $I^n$. Surely, this happens exactly when the coefficient $\frac{\xi_n + \xi_{n+1}}{2}$ vanishes, i.e., when $\xi_n = -\xi_{n+1}$. If this is case, then the coding of the corresponding random variable $Z_n$ is in that sense dispensable, as for $\hat{B}$ we only need the values of

$$
\ldots \sum_{i=1}^{n-1} Z_i, \sum_{i=1}^{n+1} Z_i, \ldots
$$

and not of $Z_n$ itself.

Therefore, we will define a sequence of random times $(\sigma_n)_{n \in \mathbb{N}}$ corresponding to those $\tau_n$ where $\hat{B}$ does jump within the following interval. These times together with its coefficients fully determine $\hat{B}$.
First, we define the integer-valued random variables $N_n$’s as follows: $N_0 = 0$, $N_1$ denotes the number of $\tau$-intervals before $B_t$ moves for the second time in the same direction, $N_2$ denotes the number of intervals before $B_t$ moves two times in the same direction after $\sum_{i=0}^{N_1} \tau_i$, etc. Finally, let $M_n = \sum_{i=0}^{n} N_i$ the cumulative sum of the $N_i$’s.

The motivation for this definition is that the $n$-th jumping point of $\hat{B}$ occurs in the interval $I^M$, i.e., between $\sum_{i=1}^{M_n} \tau_i$ and $\sum_{i=1}^{M_n+1} \tau_i$. Moreover, as long as the path goes up and down alternately (i.e., $\xi_1 = -\xi_2 = \xi_3 = \ldots$), the coefficients $\frac{\xi_{k+2} + \xi_{k+1}}{2}$ vanish and thus $\hat{B}$ does not jump.

Additionally, two consecutive jumps of $\hat{B}$ go into the same direction if and only if the corresponding $N_n$ is odd.

**Definition:** Define, for $n \in \mathbb{N}$, the real-valued random variables $\sigma_n$ and $\pi_n$ by $\sigma_0 = 0$, $\pi_0 = 1$ and:

$$\sigma_n := \sum_{i=M_n-1}^{M_n} \tau_i, \quad \pi_n := \tau_{M_n+1}, \quad n \geq 1$$

with $M_n$ as above. Further, let $\eta_0 := \frac{B_0}{2}$ and $\eta_n := \frac{B_{\sum_{i=0}^{n} \sigma_i + \pi_n} - B_{\sum_{i=0}^{n} \sigma_i}}{\tau_n}, \quad n \geq 1$.

Next, the goal is to construct discrete random variables $D_n$’s such that for every $n \in \mathbb{N}$:

$$\sum_{i=0}^{n} \sigma_i \leq \sum_{i=0}^{n} D_i \leq \sum_{i=0}^{n} \sigma_i + \pi_n.$$  \hspace{1cm} (6)

With this property, $\hat{B}$ can be defined equivalently as follows:

$$\hat{B}_t = \sum_{n=0}^{\infty} \eta_n \mathbb{1}_{\{\sum_{i=0}^{n} D_i = \infty\}}(t).$$

We now need to determine the distribution of $\sigma_n$ and $\pi_n$. Due to the Markov property, both $(\tau_n)_n$ and $(N_n)_n$ are i.i.d. families, and hence $(\sigma_n)_n$ and $(\pi_n)_n$ are i.i.d. as well. Since $\pi_n \leq \sigma_{n+1} + 1$, both families are not mutually independent. However, as $\pi_n$ does not depend on $B$ moving up or down in that interval, $\pi_n$ is independent of $M_n$ and hence for any fixed $n \in \mathbb{N}$, $\sigma_n$ and $\pi_n$ are independent.

**Lemma 5:** For $n \geq 1$, the distribution of $\sigma_n$ and $\pi_n$ is the same as the first exit time of Brownian motion of the interval $[-1, 2]$ and $[-1, 1]$ respectively, i.e.,

$$\mathbb{P}(\sigma_n \leq t) = \mathbb{P}(\exists s \leq t : B_s \notin (-1, 2))$$

and

$$\mathbb{P}(\pi_n \leq t) = \mathbb{P}(\exists s \leq t : B_s \notin (-1, 1)).$$

Moreover, the probability $\mathbb{P}(\sigma_n \leq t \wedge \text{sgn}(\eta_{n-1} \cdot \eta_n) = +1)$ is equal to the probability of the exit before $t$ and that the exit takes place at $-1$.

**Proof:** Because of the identical distribution, we may assume without loss of generality that $n = 1$. The assertion for $\pi_1$ is evident from its definition. For $\sigma_1$, note that the exit time can be written as

$$\inf\{s \geq 0 \mid B_s \notin (-1, 2)\} = \sum_{i=1}^{N} \tau_i,$$

where $N = \min\{n \mid B_{\sum_{i=0}^{n} \tau_i} \in \{-1, 2\}\}$. But as $\mathbb{P}(N_1 = k) = 2^{-k} = \mathbb{P}(N = k)$, we have $N = D N_1$.

For the last part, note that Brownian motion leaves $(-1, 2)$ at $-1$ (+2) if and only if $N$ is odd (even). Clearly, $\mathbb{P}(N \text{ is odd}) \leftrightarrow \mathbb{P}(N \text{ is odd})$. The remark above Definition 6 gives the result.

For now on, let $\sigma = D \inf\{t \geq 0 \mid B_t \notin (-1, 2)\}$ and $\pi = D \inf\{t \geq 0 \mid B_t \notin (-1, 1)\}$ be mutually independent copies of the exit times in Lemma 5.

**Lemma 6:** For $\sigma_1, \sigma_2, \ldots, \pi_1, \pi_2, \ldots$ and $\eta_0, \eta_2, \ldots$ as above, there are random variables $D_1, D_2, \ldots$ such that

$$\sum_{i=1}^{n} \sigma_i \leq \sum_{i=1}^{n} D_i \leq \sum_{i=1}^{n} \sigma_i + \pi_n \quad \text{for } n \in \mathbb{N},$$

whose conditional joint entropy can be bounded by

$$\mathbb{H}(D_1, \ldots, D_n \mid \eta_0, \ldots, \eta_n) \leq n \mathbb{H}(D \mid B_\sigma),$$

where $D$ holds $\sigma \leq D \leq \sigma + \pi$ and is constructed as in Section II.

**Proof:** As $\pi_1$ and $\sigma_1$ are independent, we can construct $D_1$ such that $\sigma_1 \leq D_1 \leq \sigma_1 + \pi_1$ as in Section II. Clearly $\mathbb{H}(D_1 \mid \eta_0, \ldots, \eta_n) \leq \mathbb{H}(D_1 \mid \text{sgn} \eta_0 \cdot \eta_1) \leq \mathbb{H}(D \mid B_\sigma)$ (by Lemma 5). For $D_2$ we stipulate that $\sigma_1 + \sigma_2 \leq D_1 + D_2 \leq \sigma_1 + \sigma_2 + \pi_2$, i.e.,

$$\sigma_2 - (D_1 - \sigma_1) \leq D_2 \leq \sigma_2 - (D_1 - \sigma_1) + \pi_2.$$

Again, $\pi_2$ is independent of $\sigma_1$ and $\sigma_2$ (and thus of $D_1$). As $D_1 - \sigma_1 \geq 0$ we apply the scaling property to see that the quantization problem is easier than

$$\sigma_2 \leq D_2 \leq \sigma_2 + \pi_2.$$

Therefore, as for $D_1$:

$$\mathbb{H}(D_1 \mid \eta_0, \ldots, \eta_n) \leq \mathbb{H}(D_2 \mid \text{sgn} \eta_1 \cdot \eta_2) \leq \mathbb{H}(D \mid B_\sigma).$$

Proceeding in this fashion we define the $D_n$, such that for $n \in \mathbb{N}$:

$$\mathbb{H}(D_1, \ldots, D_n \mid \eta_0, \ldots, \eta_n) \leq \sum_{i=1}^{n} \mathbb{H}(D_i \mid \eta_0, \ldots, \eta_n) \leq n \mathbb{H}(D \mid B_\sigma).$$

By the scaling property, we need to encode $\hat{B}$ up to time $\frac{1}{\varepsilon^2}$, therefore we let $\hat{B}^\varepsilon := \hat{B} \mid [0, \varepsilon^2]$. This process depends only on finitely many $D_n$. We introduce the new random variable

$$M^\varepsilon := \max\left\{n \in \mathbb{N} \mid \sum_{i=0}^{n} \sigma_i \leq \varepsilon^{-2}\right\},$$

indicating that number.

Note that, as our first jumping point is at $D_0$, the number of jumps of $\hat{B}^\varepsilon$ is $M^\varepsilon + 1$. As we may choose $D_0 \equiv 0$ deterministically, $M^\varepsilon$ corresponds to the number of $D_n$, which we actually have to encode. The entropy $\mathbb{H}(\hat{B}^\varepsilon)$ is now estimated as follows:
Lemma 7: For the entropy of the coding strategy $\hat{B}^c$ holds
$$H(\hat{B}^c) \leq H(M^c) + H(\hat{\eta}) + H(D | B_\sigma) \mathbb{E}[M^c].$$

**Proof:** We use the conditional entropy $H(M^c | D) = H(M^c) + H(D | M^c)$. Further:
$$H(\hat{B}^c | M^c) = \sum_{n=0}^{\infty} P(M^c = n) [H(D_0, \eta_0, \ldots, D_n, \eta_n) - H(\hat{\eta}) - H(D | B_\sigma)]
\leq \sum_{n=0}^{\infty} P(M^c = n) [H(D_0, \eta_0, \ldots, D_n, | \eta_0) + \sum_{i=1}^{\infty} H(\eta_i | \eta_{i-1})]
= H(\eta_0) + \sum_{n=0}^{\infty} P(M^c = n) \sum_{i=1}^{\infty} H(\eta_i | \eta_{i-1})
= H(\eta_0) + \sum_{n=0}^{\infty} P(M^c = n) \mathbb{E}[M^c].$$

Here we used the fact that we can chose $D_0 \equiv 0$ and that $(\eta_i, \eta_{i-1}) \equiv \sigma(\eta_1, 2\eta_0)$, by the Markov property.

For the rest of this section we determine the constants in this inequality and its asymptotic behavior. We begin with the expectation and entropy of $M^c$:

**Lemma 8:** The expected number of jumping points of $\hat{B}$ per unit interval is asymptotically equal to $\mathbb{E}(\sigma)^{-1}$, i.e.,
$$\lim_{\varepsilon \to 0} \varepsilon^2 \mathbb{E}(M^c) = \frac{1}{\mathbb{E}(\sigma)} = \frac{1}{2}.$$

Moreover, for the entropy of this number holds
$$\lim_{\varepsilon \to 0} \varepsilon^2 H(M^c) = 0.$$

**Proof:** Introducing a time parameter $t \geq 0$, the process $N_t := M^c \mathbb{I}_t$ can be regarded as a renewal process with interarrival times distributed like $\sigma$. By the elementary renewal theorem
$$\lim_{\varepsilon \to 0} \varepsilon^2 \mathbb{E}(M^c) = \lim_{t \to \infty} \frac{\mathbb{E}(N_t)}{t} = \frac{1}{\mathbb{E}(\sigma)}.$$

For the fact that $\mathbb{E}(\sigma) = 2$ see, e.g., [3, p. 212]. For the second result, we can define a continuous random variable $\Xi$ as in the previous section, such that $H(\Xi) = H(N_t)$.

Applying the maximum entropy (with $p = 1$) we thus have:
$$H(N_t) \leq \log \left(2\varepsilon(\mathbb{E}(N_t) + \frac{1}{2})\right).$$

It then follows that
$$\lim_{\varepsilon \to 0} \varepsilon^2 H(M^c) = \lim_{t \to \infty} \frac{H(N_t)}{t} = 0.$$

As the random variables $\eta_n$ only assume different values, its entropy is bounded by 1. In fact, as these $\eta_n$ are not independent, a slightly better estimation holds:

**Lemma 9:** For the (conditional) entropy $H(\eta_0)$ and $H(\eta_1 | \eta_0)$ it holds that
$$H(\eta_0) = 1 \text{ and } H(\eta_1 | \eta_0) < 0.92.$$

**Proof:** By symmetry of Brownian motion, $P(\eta_0 = \pm \frac{1}{2}) = \frac{1}{2}$, and thus $H(\eta_0) = 1$. By the remark above Definition 6:
$$\mathbb{P}(\eta_1 = 1 | \eta_0 = \frac{1}{2}) = \mathbb{P}(\eta_1 = -1 | \eta_0 = \frac{1}{2}) = \mathbb{P}(N_1 \text{ odd}) = \sum_{n=0}^{\infty} 2^{-(2n+1)} = \frac{2}{3}.$$

Thus, $H(\eta_1 | \eta_0) < \frac{2}{3} \log \frac{3}{2} + \frac{1}{4} \log 3 < 0.92$.

In order to estimate $H(D | B_\sigma)$ we need the probabilities $\mathbb{P}(\pi \leq t)$ and some moment of $\sigma | \{B_\sigma = -1\}$ and $\sigma | \{B_\sigma = 2\}$.

**Lemma 10:** Let $\pi$ be the exit time of $B$ of the interval $(-1, 1)$. The following estimation holds:
$$\mathbb{P}(\pi \leq t) \leq \beta_2 t^2,$$
where
$$\beta_2 = \frac{4}{\sqrt{2\pi}} \frac{3^3}{e} < 1.86.$$

**Proof:** Clearly,
$$\mathbb{P}(\pi \leq t) \leq 4 \mathbb{P}(B_t > 1) \leq \frac{4}{\sqrt{2\pi}} t^2 e^{-\frac{t^2}{2}} =: \gamma(t) t^2.$$

The derivative of $\gamma$ is given by
$$\frac{d\gamma}{dt}(t) = \frac{4}{\sqrt{2\pi}} t^{-\frac{3}{2}} e^{-\frac{t^2}{2}} \left[\frac{1}{2} t^{-1} - \frac{3}{2}\right].$$

Therefore there is a unique maximum at $t = \frac{1}{3}$. Now
$$\gamma\left(\frac{1}{3}\right) = \frac{4}{\sqrt{2\pi}} \left(\frac{3}{e}\right)^2 = \beta_2.$$

Hence $\mathbb{P}(\pi \leq t) \leq \gamma(t) t^2 \leq \beta_2 t^2$.

Evaluating its Laplace transforms (see, e.g., [3, p. 212]), the first two moments of $\sigma | \{B_\sigma = -1\}$ and $\sigma | \{B_\sigma = 2\}$ can be seen as:
$$E(\sigma | B_\sigma = -1) = \frac{5}{3}, \quad E(\sigma | B_\sigma = 2) = \frac{8}{3};$$
$$E((\sigma | B_\sigma = -1)^2) = \frac{17}{3}, \quad E((\sigma | B_\sigma = 2)^2) = \frac{32}{3}.$$

For $H(D | B_\sigma)$, the last remaining term of the right hand side of Lemma 7 we apply the results of Section II.

**Lemma 11:** The conditional entropy $H(D | B_\sigma)$ is bounded by:
$$H(D | B_\sigma) < 4.74.$$

Proof: We apply Proposition 4 with $p = 1, 2, n = 2$ and Lemma 10. Therefore $\mathbb{H}(D \mid B_\sigma = \bullet)$ is bounded by the minimum of:

$$
\begin{aligned}
\Lambda_2 + \log \left[ \frac{2}{3} + 2\sqrt{\pi} \mathbb{E}^{\left[ \sigma \mid B_\sigma = \bullet \right]} \right] \\
\Lambda_2 + \frac{3}{2} \log \left[ \frac{2}{3} + 2\sqrt{\pi} \mathbb{E}^{\left[ \sigma \mid B_\sigma = \bullet \right]} + 4\sqrt{2} \mathbb{E}^{\left[ \sigma \mid B_\sigma = \bullet \right]} \right]
\end{aligned}
$$

With the moments above we obtain the values:

$$
\begin{aligned}
\mathbb{H}(D \mid B_\sigma = -1) &\leq \min\{4.57, 4.62\} = 4.57 \\
\mathbb{H}(D \mid B_\sigma = +2) &\leq \min\{5.20, 5.06\} = 5.06.
\end{aligned}
$$

By the probabilities calculated in the proof of Lemma 9:

$$
\begin{aligned}
\mathbb{H}(D \mid B_\sigma = +1) &= \frac{2}{3} \mathbb{H}(D \mid B_\sigma = -1) + \frac{1}{3} \mathbb{H}(D \mid B_\sigma = +2) \\
&\leq 4.74.
\end{aligned}
$$

Collecting results, we can now state the central result of this paper:

**Theorem 5:** For the optimal coding error $D^{(e)}(H)$ of Brownian motion $B$ on the interval $[0, 1]$ it holds that

$$
\kappa = \lim_{H \to \infty} D^{(e)}(H) \cdot \sqrt{H} < 2.11.
$$

Proof: Let $\varepsilon > 0$. The coding error of $\hat{B}$ (and thus of $\hat{B}^e = \hat{B} \mid [0, \frac{1}{2\varepsilon}]$) is bounded by $\frac{3}{2}$ (Lemma 4). Thus, by the scaling property there are $\hat{B}^e$ such that $\sup_{\varepsilon \in [0, 1]} |B - \hat{B}^e| \leq \frac{3}{2} \varepsilon$ and $\mathbb{H}(\hat{B}^e) = \mathbb{H}(B^e)$, hence $D^{(e)}(\mathbb{H}(\hat{B}^e)) \leq \frac{3}{2} \varepsilon$. With the substitution

$$
H := \mathbb{H}(\hat{B}^e) = \mathbb{H}(B^e) \ln 2
$$

and according to the Lemmas 7, 8, 9 and 11 we have:

$$
\kappa = \lim_{H \to \infty} D^{(e)}(H) \cdot \sqrt{H} \leq \lim_{\varepsilon \to 0} \frac{3}{2} \varepsilon \sqrt{\mathbb{H}(B^e)} \ln 2
$$

$$
\leq \lim_{\varepsilon \to 0} \frac{3}{2} \varepsilon \left[ \mathbb{H}(M^e) + \mathbb{H}(\eta_\varepsilon) \right]
$$

$$
\left[ \mathbb{H}(\eta_1 \mid \eta_0) + \mathbb{H}(D \mid B_\sigma) \mathbb{E}[M^e] \right] \ln 2 \right]^\frac{1}{2}
$$

$$
= \frac{3}{2} \left[ 0 + \mathbb{H}(\eta_1 \mid \eta_0) + \mathbb{H}(D \mid B_\sigma) \ln 2 \right] \left[ \frac{1}{2} \ln 2 \right] \frac{1}{2}
$$

$$
< \frac{3}{2} \sqrt{\frac{0.92 + 4.74}{2}} \ln 2 < 2.11.
$$

**IV. Conclusions**

We have derived upper bounds for the quantization problem of the exit times of Brownian motion, as formalized in Eq. (2). We have considered the bounded (Theorems 1 and 2), arbitrary (Theorem 3) and nonnegative (Theorem 4) cases for these exit times. These results were then applied to construct and determine the entropy of a coding strategy for Brownian motion (Theorem 5). Together with the result from [9], we have showed that for the optimal constant $\kappa$ from Eq. (1), it holds that $\kappa \in \left[ \frac{\pi}{\sqrt{2}}, 2.11 \right]$. Therefore, we have not only provided a concrete coding strategy for Brownian motion, but we have also improved the range of the multiplicative constant.

Although we could not achieve optimal results in Section II, we speculate that the multiplicative constant $\kappa$ is less than 2.11. In particular, for the estimation of $\mathbb{H}(X)$, one could apply maximum entropy distributions of classes different from $C^p_\mu$ (e.g., moments of higher order, exponential moments, etc.) in order to tighten the bounds.

**References**


