universität wien

MASTERARBEIT

Titel der Masterarbeit

A Connection of Skorokhod Embeddings to the Brascamp-Lieb Moment Inequality

Verfasser

Markus Ausserhofer, BSc.

angestrebter akademischer Grad

Master of Science (MSc)

Wien, im Dezember 2015

Studienkennzahl lt. Studienblatt:	A 066821
Studienrichtung lt. Studienblatt:	Masterstudium Mathematik
Betreuer:	Assoz. Prof. Dr. Mathias Beiglbck

A CONNECTION OF SKOROKHOD EMBEDDINGS TO THE BRASCAMP-LIEB MOMENT INEQUALITY

Abstract. We prove the Brascamp-Lieb moment inequality using the Skorokhod embedding by Bass and a basic result of Stochastic Ordering of random variables. For a special case, we provide a more direct proof of the inequality using stochastic ordering. We also present a property of random variables, that are log-concave with respect to a Gaussian.

Acknowledgements

I would like to thank my advisor Mathias Beiglböck, as well as Marcel Nutz for their support in writing this thesis, and I apology to Mathias for interfering in his fraternity leave. Further thanks go to Harald Rindler, H.G. Feichtinger, and Jiří Černý for giving entertaining lectures and introducing me to mathematics, as well as Burak Gürel from Boğaziçi University for going out of his way to provide excellent guidance in conquering my first scientific paper.

Furthermore, I would like to thank my family for their unconditional support and love, particularly my grandmom, whose love I could enjoy for almost twenty-four years, and to whom this thesis shall be dedicated.

Contents

1. Introduction

1.1. **Introduction.** The central topic of this thesis is the Brascamp-Lieb moment inequality. It relates the expectation of two random variables under a convex function. It requires one random variable to be log-concave to the other.

Definition 1.1. *A random variable X with distribution μ is log-concave with respect to a random variable Y with distribution ν if μ is absolutely continuous with respect to ν and the Radon-Nikódym derivative is given in the following form*

$$\mu(dx) = \frac{1}{Z} e^{-V(x)}\ \nu(dx), \tag{1}$$

where $V : \mathbb{R} \to \mathbb{R}$ is convex and Z is the normalizing constant $Z = \int e^{-V(x)}\ \nu(dx)$.

Our main theorem is the following.

Theorem 1.2. *If X and Y are random variables with values in $\mathbb{R}^n$, Y a Gaussian random variable and X log-concave with respect to Y, then the following inequality holds for all convex functions $\psi : \mathbb{R} \to \mathbb{R}$:*

$$\mathbb{E}[\psi(\phi \cdot Y - \mathbb{E}[\phi \cdot Y])] \geq \mathbb{E}[\psi(\phi \cdot X - \mathbb{E}[\phi \cdot X])]. \tag{2}$$

First proved in [1] 1976, it states that the moments (ψ was assumed to be a power function) of a Gaussian random variable Y are bigger than the moments of a random variable X that is log-concave relative to Y. In 2002 Caffarelli [4] used Optimal Transport Theory, especially his results involving the Monge-Ampere equation, to generalize the above inequation to general convex ψ. In 2014 Yuu Hariya [3] used the Skorokhod Embedding by Bass to give a more elegant proof of the inequality (2).

We were able to give a slightly different version of the proof, alongside the method of Yuu Hariya, which is the substantial content of this thesis.

1.2. **Terminology issues.** There is some confusion regarding the name of the inequality. As stated above, Brascamp and Lieb published the first version of Equation 2 in [1] in 1976 in the following form

$$\mathbb{E}[(\phi \cdot Y - \mathbb{E}[\phi \cdot Y])^n] \geq \mathbb{E}[(\phi \cdot X - \mathbb{E}[\phi \cdot X])^n]. \tag{3}$$

Additionaly Brascamp and Lieb published another article [2], in which they proved another inequality. Until Caffarelli, equation (2) was referred to as Brascamp-Lieb *moment* inequality, and the inequality in [2] simply as Brascamp-Lieb inequality.

With the result of Caffarelli, the generalization from equation (3) to equation (2), the extension *moment* was considered no longer appropriate and hence has been neglected by Hariya. This has created a not neglicable amount of confusion to the author of this thesis, and we therefore prefer to stick to the name Brascamp-Lieb *moment* inequality.

1.3. **How to read this thesis.** This thesis offers a fast and elegant proof of the Brascamp-Lieb moment inequality (2) with the method of Yuu Hariya [3]. We were able to further shorten the proof of the inequality by using the concept of convex order of random variables and a direct relation between random variables in convex order and random variables embedded in Brownian motion with ordered stopping times $S \leq T$.

In Sections 2.1 and 2.2 we review basic concepts of stochastic analysis covered in any graduate course on the subject. In Section 2.3 we state and prove results from potential theory, that will be used in Section 3.2 to present the Skorokhod embedding by Chacon and Walsh. We present the embedding by Bass in Section 3.1. In Section 4.2 we present the concept of convex order and use the Chacon and Walsh embedding to establish the above mentioned direct relation to Skrokhod embeddings. Subsection **??** might be interesting as it gives a direct proof if V is unbounded at plus and minus infinity.

The real essence of this thesis is to be found in Sections 5 and 6. There, we combine the above results to our main result, the proof of the Brascamp-Lieb moment inequality. Theorem 5.10 describes a property of log-concave functions and is new to the author.

2. Preliminaries

For the purpose of our thesis, we need a few results from stochastic analysis. Here we want to repeat the notion of uniform integrability and results involving uniform integrable random variables, namely continuous martingale convergence and optional stopping. In Subsection 2.2 we recall martingale representation theorems, and prove Clark's formula. In Subsection 2.3 we present a theorem that lists properties of potentials in one dimension.

2.1. **Uniform Integrability.** One, if not one of the biggest, of the recurring problems in mathematics is the interchangability of a limit and the integral sign. In every course on measure theory the following two very useful theorems are proved.

Theorem 2.1 (Monotone Convergence Theorem, see [17])**.** *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and $f_n : \Omega \to \mathbb{R}$ be a sequence of measurable functions such that $f_n(\omega) \leq f_{n+1}(\omega)$. Let $f : \Omega \to \mathbb{R}$ be such that $f(\omega) = \lim_n f_n(\omega)$ where the limit exists and infinity otherwise. Then the limit and the integral sign commute. This is,*

$$\lim \int f_n \, d\mu = \int \lim f_n \, d\mu = \int f \, d\mu. \tag{4}$$

For $(\Omega, \mathcal{A}, \mu)$ a probability space, and X_n an increasing sequence of random variables,

$$\lim \mathbb{E}[X_n] = \mathbb{E}[\lim_n X_n]. \tag{5}$$

Theorem 2.2 (Dominated Convergence Theorem, see [17])**.** *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $g : \Omega \to \mathbb{R}$ an integrable function, $\mathbb{E}[|g|] < \infty$, and $f_n : \Omega \to \mathbb{R}$ be a sequence of measurable functions such that $f_n(\omega) \leq g(\omega)$. Let f_n converge almost surely to $f : \Omega \to \mathbb{R}$, $f(\omega) = \lim_n f_n(\omega)$. Then the limit and the integral sign commute. This is,*

$$\lim \int f_n \ d\mu = \int \lim f_n \ d\mu = \int f \ d\mu. \tag{6}$$

For $(\Omega, \mathcal{A}, \mu)$ a probability space and $f_n \leq g$,

$$\lim \mathbb{E}[X_n] = \mathbb{E}[\lim_n X_n]. \tag{7}$$

In practice, sequences of random variables at hand converge pointwise to a limit, but sometimes neither of the two above theorems, Monotone- or Dominated Convergence Theorem, can be applied to justify swapping the limit. We will see in this chapter that uniform integrability is exactly the property needed.

Definition 2.3 (Uniform integrability)**.** *A family $(X_i)_{i \in I}$ of random variables is uniformly integrable if*

$$\lim_{M \to \infty} \sup_{i \in I} \mathbb{E}[|X_i| \mathbb{1}_{|X_i| \geq M}] = 0. \tag{8}$$

As starters, let us examine uniform integrability more closely. If we are looking at one random variable only, or $I = \{0\}$ in Definition 2.3, Equation 8 provides a notion of integrability. We have

$$\lim_{M \to \infty} \mathbb{E}[|X| \mathbb{1}_{|X| \geq M}] = 0, \text{ if and only if } X \in \mathcal{L}^1. \tag{9}$$

Indeed, if $X \in \mathcal{L}^1$, also $|X| \in \mathcal{L}^1$ and $|X|$ can be used as a dominating function in Theorem 2.2 because $|X| \mathbb{1}_{|X| \geq M} \leq |X|$. As M goes to infinity, $|X| \mathbb{1}_{|X| \geq M}$ goes to zero and $\lim_{M \to \infty} \mathbb{E}[|X| \mathbb{1}_{|X| \geq M}] = 0$. On the other hand, if $\lim_{M \to \infty} \mathbb{E}[|X| \mathbb{1}_{|X| \geq M}] = 0$, there is M such that $\mathbb{E}[|X| \mathbb{1}_{|X| \geq M}] \leq 1$ and

$$\begin{aligned} \mathbb{E}[|X|] &= \mathbb{E}[|X| \mathbb{1}_{|X| \geq M}] + \mathbb{E}[|X| \mathbb{1}_{|X| < M}] \\ &\leq 1 + M. \end{aligned}$$

In particular, this shows that every random variable that belongs to a family of uniformly integrable random variables is also in $\mathcal{L}^1$. However, the point of uniform integrability is of course the fact that the expected values of the truncated random variables go to zero *uniformly*. There is an alternative definition of uniform integrability that we will examine now.

Definition 2.4. *A family of random variables* $(X_i)_{i\in I}$ *is uniformly integrable if there is a* $M \in \mathbb{R}$ *such that* $\mathbb{E}[|X_i|] \leq M$ *and for every* $\varepsilon > 0$ *there is a* $\delta > 0$ *such that for every measurable* A*,* $\mathbb{E}[|X_i|\mathbb{1}_A] \leq \varepsilon$ *if* $P(A) \leq \delta$ *.*

Proposition 2.5. *Definitions* 2.3 *and* 2.4 *agree.*

Proof. As before, $\mathbb{E}[|X|] = \mathbb{E}[|X|\mathbb{1}_{|X|\geq M}] + \mathbb{E}[|X|\mathbb{1}_{|X|<M}]$. If $(X_i)_i$ is uniformly integrable according to Definition 2.3, setting $\varepsilon = 1$ gives an M such that

$$\mathbb{E}[|X|\mathbb{1}_{|X|\geq M}] \leq \varepsilon,$$

and $\mathbb{E}[|X|] < M+1$. Also, set $\delta = \frac{\varepsilon}{M}$ to get $\mathbb{E}[|X_i|\mathbb{1}_A] \leq MP(A) + \mathbb{E}[|X|\mathbb{1}_{X\geq M}] \leq M \cdot \frac{\varepsilon}{M} + \varepsilon = 2\varepsilon$, and $(X_i)_i$ is uniformly integrable according to Definition 2.4.

If $(X_i)_i$ is uniformly integrable according to Definition 2.4, define $A_\gamma := \{|X| \geq \gamma\}$. The probability of A_γ decreases to zero as $\gamma \to \infty$ and there is γ_0 such that $P(A_{\gamma_0}) \leq \delta(\varepsilon)$ and therefore $\mathbb{E}[|X|\mathbb{1}_{|X|\geq\gamma_0}] \leq \varepsilon$, and $(X_i)_i$ is uniformly integrable according to Definition 2.3. □

The following lemma is another characterization of uniform integrability. A function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ is called a *test function for uniform integrability* if

$$\lim_{t\to\infty} \frac{\phi(t)}{t} = \infty. \tag{10}$$

Lemma 2.6. *Let* $(X_i)_{i\in I}$ *be a subset of* $\mathcal{L}^1$*. If*

$$\sup_{i\in I} \mathbb{E}[\phi \circ |X_i|] \leq \infty, \tag{11}$$

then $(X_i)_{i\in I}$ *is uniformly integrable.*

There is actually a converse to this lemma. If $(X_i)_{i\in I}$ is uniformly integrable, then there is an increasing test function ϕ such that assertion (11) holds. The corresponding theorem is called de la Vall'ee Poussin-Theorem and can be found in [20]. Here we only prove the lemma.

Proof of Lemma 2.6. Set $A := \sup_{i\in I} \mathbb{E}[\phi \circ |X_i|] \leq \infty$ and $a := \varepsilon^{-1}A$. Then there is t_0 such that $\frac{\phi(t)}{t} \geq a$ for $t \geq t_0$. Then,

$$\begin{aligned}
\mathbb{E}[|X_i|\mathbb{1}_{|X_i|\geq t_0}] &= \mathbb{E}[\frac{\phi(|X_i|)}{a}\mathbb{1}_{|X_i|\geq t_0}] \\
&= \frac{\varepsilon}{A}\mathbb{E}[\phi(|X_i|)\mathbb{1}_{|X_i|\geq t_0}] \\
&\leq \frac{\varepsilon}{A}\sup_{i\in I} \mathbb{E}[\phi \circ |X_i|] \leq \varepsilon.
\end{aligned}$$

This is, $(X_i)_{i\in I}$ is uniformly integrable according to Definition 2.4. □

The above results hold for any class of random variables on any probability space. In stochastic analysis one usually works on a filtred probability space that satisfies the usual conditions. For the sake of completeness these two notions shall be recalled. We also recall the definition of submartingales, supermartingales, and martingales.

Definition 2.7. *A collection* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{T}}, P)$ *is a filtered probability space if* $(\Omega, \mathcal{F}, P)$ *is a probability space and* $(\mathcal{F}_t)_{t\in\mathbb{T}}$ *an increasing familiy of* σ*-algebras, this is* $\mathcal{F}_s \subset \mathcal{F}_t$ *if* $s < t$.

Definition 2.8. *A filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{T}}, P)$ *satisfies the usual conditions if every* $\mathcal{F}_t$ *contains all* P*-nullsets and the filtration is complete,* $F_t = F_{t+} = \cap_{t'>t} F_{t'}$.

Definition 2.9. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* $(\mathcal{F}_t)_{t\geq 0}$ *a filtration. A stochastic process* $(X_t)_{t\geq 0}$ *with* $\mathbb{E}[X_t] < \infty$ *is a* $\mathcal{F}_t$*-martingale if*

$$\mathbb{E}[X_t|\mathcal{F}_s] = X_s \text{ for all } t \geq s, \tag{12}$$

a $\mathcal{F}_t$*-submartingale if*

$$\mathbb{E}[X_t|\mathcal{F}_s] \geq X_s \text{ for all } t \geq s, \tag{13}$$

and a $\mathcal{F}_t$*-supermartingale if*

$$\mathbb{E}[X_t|\mathcal{F}_s] \leq X_s \text{ for all } t \geq s. \tag{14}$$

Martingales, alongside with sub- and supermartingales, are a very important object in probability theory and any attempt to stress their importance would do them unjust. Let us just mention that the central process of this thesis, Brownian motion, is not only itself a martingale, but in the sense of the Dambis-Dubin-Schwarz theorem the *only* martingale in continuous time, see Theorem 2.26.

Working on a filtred probability space that satisfies the usual conditions has various advantages. For example, every submartingale X_t for which the function $t \mapsto \mathbb{E}[X_t]$ is right-continuous has a modification Y that is also a submartingale and the trajectories of Y are right-continuous and possess limits from the left at every point, such martingales are called càdlàg. Càdlàg is a french acronym, continue à droite limite à gauche that translates to right-continuous and existing limits from the left.

A martingale has the property that, when interpreting the parameter t as time and the filtration $\mathcal{F}_t$ as the information at time t, when starting at time s the expected value of the martingale at a later time t equals the value of the martingale at time s. A submartingale on the other hand has a tendency to rise, the value at time s falls short to the expectation of the process at a later time. So, in some ways, submartingales are the stochastic pendant to non-decreasing sequences in $\mathbb{R}$. As bounded non-decreasing sequences in $\mathbb{R}$ converge, we have a similar statement for submartingales. The following theorem can be found in [11], Theorem 2.10.

Theorem 2.10 (Almost sure martingale convergence). *Let $(X_t)_t$ be a submartingale and M a uniform bound, $\mathbb{E}[X_t] \leq M$. Then X_t converges almost surely to some random variable $X < \infty$ a.s.*

Doob's $\mathcal{L}^p$ inequalities are also known from any course on stochastic analysis. We state it without proof, the first part of the theorem with $p = 1$ is sometimes just called Doob's inequality.

Theorem 2.11 (Doob's $\mathcal{L}^p$-inequality, see [11]). *Let X be a right-continuous martingale or a non-negative sub-martingale. Then for $p \geq 1$*

$$\lambda^p \cdot P[\sup_{u \leq t} |X_u| \geq \lambda] \leq \sup_{u \leq t} \mathbb{E}[|(X_u)|^p], \text{ for every } \lambda > 0, \tag{15}$$

and for $p > 1$

$$\mathbb{E}[(\sup_{u \leq t} |X_u|)^p]^{\frac{1}{p}} \leq \frac{p}{p-1} \sup_{u \leq t} \mathbb{E}[|X_t|^p]^{\frac{1}{p}}, \text{ for } p \in (1, \infty). \tag{16}$$

Let us focus on convergence of random variables and the role of uniform integrability. First, we will see that uniform integrability behaves well with linear combinations.

Lemma 2.12. *If $(X_i)_{i \in I}$ and $(Y_j)_{j \in J}$ are uniform integrable families of random variables, α and β real numbers, then the familiy $(\alpha X_i + \beta Y_j)_{i \in I, j \in J}$ is also uniformly integrable.*

Proof. It is easily seen that $\alpha(X_i)_i$ is uniformly integrable, because

$$\mathbb{E}[\alpha |X| \mathbb{1}_{\alpha |X| \geq \alpha M}] = \alpha \mathbb{E}[|X| \mathbb{1}_{|X| \geq M}],$$

the latter expression goes to zero as $M \to \infty$ and trivially αM goes to infinity when M does. The following pointwise inequality is true

$$|X_i + Y_j| \mathbb{1}_{|X_i + Y_j| \geq 2M} \leq 2|X_i| \mathbb{1}_{|X_i| \geq M} + 2|Y_j| \mathbb{1}_{|Y_j| \geq M}.$$

Taking the expectation, sup over $i \in I$ and $j \in J$, and the limit we find

$$\lim_{M \to \infty} \sup_{i \in I, j \in J} \mathbb{E}[|X_i + Y_j| \mathbb{1}_{|X+Y| \geq M}] = 0.$$

□

This particularly implies the following useful corollary.

Corollary 2.13. *If $(X_i)_i$ is uniformly integrable and $X \in \mathcal{L}^1$, then $(X_i - X)_i$ is also uniformly integrable.*

Using this corollary we can relate uniform integrability to $\mathcal{L}^1$-convergence.

Proposition 2.14. *$X_t \to X$ almost surely and $(X_t)_t$ is uniformly integrable, if and only if $X_t \to X$ in $\mathcal{L}^1$.*

Proof. ($\Rightarrow$) Using Fatou's Lemma we show for X

$$\begin{aligned}\mathbb{E}[|X|] &= \mathbb{E}[|\lim X_t|]\\ &= \liminf \mathbb{E}[|X_t|]\\ &\leq \sup \mathbb{E}[|X_t|] < \infty.\end{aligned}$$

Therefore, X is integrable, and from Corollary 2.13 $(X_t - X)_t$ is uniformly integrable. Consider,

$$\mathbb{E}[|X_t - X|] = \mathbb{E}[|X_t - X|\mathbb{1}_{|X_t - X| \geq M}] + \mathbb{E}[|X_t - X|\mathbb{1}_{|X_t - X| < M}].$$

As $X_t - X$ is uniformly integrable, with choosing M we can make the first term arbitrarily small. We claim the second term to be zero. Indeed, as t goes to infinity, we can apply the Dominated Convergence Theorem to the second term with dominating function $Y \equiv M$, and the term goes to zero.

($\Leftarrow$) $X_t - X$ is uniformly integrable, because

$$\mathbb{E}[|X_t - X|\mathbb{1}_{|X_t - X| \geq M]} \leq \mathbb{E}[|X_t - X|]$$

and the second term goes to zero by assumption. As X_t goes to X in $\mathcal{L}^1$, all X_t and X are in $\mathcal{L}^1$. By Corollary 2.13 $(X_t - X + X)_t = (X_t)_t$ is uniformly integrable. $X_t \to X$ almost surely, because $\mathcal{L}^1$-convergence implies almost sure convergence. □

We are ready to state our main theorem on martingale convergence.

Theorem 2.15 (Martingale Convergence Theorem)**.** *For a cádlàg martingale $(X_t)_{t \in \mathbb{R}^+}$, the following three conditions are equivalent,*

(1) X_t converges in $\mathcal{L}^1$,
(2) there exists a random variable $X_\infty \in \mathcal{L}^1$ such that $X_t = \mathbb{E}[X_\infty|\mathcal{F}_t]$,
(3) the family $(X_t)_{t \in \mathbb{R}^+}$ is uniformly integrable.

If these conditions hold, then X_∞ is also the pointwise limit of X_t, $\lim X_t = X$ a.s.

Moreover, if for some $p > 1$ the martingale is bounded in $\mathcal{L}^p$, then the equivalent conditions above are satisfied and the convergence holds in $\mathcal{L}^p$-sense.

Proof. (3) $\Rightarrow$ (1) : If $(X_t)_t$ is uniformly integrable, also $\sup_t \mathbb{E}[X_t] < \infty$ and by Theorem 2.10 X_t converges to a random variable X almost surely. Finally, Proposition 2.14 ensures $\mathcal{L}^1$-convergence.

(1) $\Rightarrow$ (2) : X_t is a martingale by assumption, so we have

$$X_t = \mathbb{E}[X_{t+h}|\mathcal{F}_t].$$

$\mathcal{L}^1$-convergence of $X_t \to X$ implies $\lim \mathbb{E}[X_t] = \mathbb{E}[X]$. The same is true for the conditional expectation, because conditional expectation as an operator from $\mathcal{L}^0$

to $\mathcal{L}^0$ is a contraction (see [18]). Let X_∞ be the $\mathcal{L}^1$-limit of X_t. Then

$$X_t = \lim \mathbb{E}[X_{t+h}|\mathcal{F}_t] = \mathbb{E}[X_\infty|\mathcal{F}_t].$$

(2) ⇒ (3) : We have to show that $\mathbb{E}[|\mathbb{E}[X_\infty|\mathcal{F}_t]|\mathbb{1}_{\mathbb{E}[X_\infty|\mathcal{F}_t]\geq M}]$ goes to zero as M goes to infinity. Due to the general fact $|\mathbb{E}[X|\mathcal{F}]| \leq \mathbb{E}[|X|\ |\mathcal{F}]$ and the averaging property of conditional expectation

$$\begin{aligned}\mathbb{E}[|\mathbb{E}[X_\infty|\mathcal{F}_t]|\mathbb{1}_{\mathbb{E}[X_\infty|\mathcal{F}_t]\geq M}] &\leq \mathbb{E}[\mathbb{E}[|X_\infty||\mathcal{F}_t]\mathbb{1}_{\mathbb{E}[X_\infty|\mathcal{F}_t]\geq M}] \\ &= \mathbb{E}[|X_\infty|\mathbb{1}_{\mathbb{E}[X_\infty|\mathcal{F}_t]\geq M}] \\ &\leq \mathbb{E}[|X_\infty|] \cdot P(\mathbb{E}[|X_\infty||F_\infty] \geq M).\end{aligned}$$

The first factor is bounded, because X_∞ is in $\mathcal{L}^1$. For the second term we use the Markov inequality that states if Z is a non-negative integrable random variable that the probability of Z being large can be bound with its expectation in the following way,

$$P(Z \geq M) \leq \frac{\mathbb{E}[Z]}{M}.$$

Therefore

$$\mathbb{E}[|\mathbb{E}[X_\infty|\mathcal{F}_t]|\mathbb{1}_{\mathbb{E}[X_\infty|\mathcal{F}_t]\geq M}] \leq \frac{\mathbb{E}[|X_\infty|]^2}{M}.$$

This is, the truncated expectations decay uniformly to zero as M goes to infinity.

Finally, if $\sup_t \mathbb{E}[|X_t|^p] < \infty$, by Doob's L^p-inequality Theorem 2.11, $\sup_t |X_t|$ is in $\mathcal{L}^p$ and the family $(X_t)_t^p$ is uniformly integrable. □

The above theorem shows again the importance of uniform integrability in $\mathcal{L}^1$-convergence. Another important concept that is used throughout this thesis is optional stopping. One can find more than one version of the optional stopping theorem in the literature, but all express the same idea. We state and prove the version as in [11], Theorem 3.2. For a uniformly integrable martingale X_t and a stopping time S that is inifinte, we define on $S = \infty$ $X_S = X_\infty$.

Theorem 2.16 (Optional Stopping Theorem, cf. [11], Theorem II 3.2)**.** *On a filtred probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ let X be a uniformly integrable cádlàg martingale, the family $\{X_S : S$ is a stopping time$\}$ is uniformly integrable and if $S \leq T$*

$$X_S = \mathbb{E}[X_T|F_S] = \mathbb{E}[X_\infty|\mathcal{F}_S]. \tag{17}$$

For the proof we use a more elementary version of the optional stopping theorem, that is where the stopping times S and T are bounded by some real number M.

Theorem 2.17 (Optional Stopping for bounded stopping times, cf. [11] Proposition II 1.4)**.** *Let X be a submartingale, and S and T be stopping times such that*

$$S \leq T \leq M < \infty. \tag{18}$$

Then,

$$\mathbb{E}[X_T|\mathcal{F}_S] \geq X_S.$$

Furthermore, X is a martingale, if and only if $\mathbb{E}[X_T] = \mathbb{E}[X_S]$ for every two stopping times $S \leq T \leq M < \infty$.

Proof of Theorem 2.17. Define the stochastic process $H := \mathbb{1}_{S \leq t \leq T}$. Then,

$$\int_0^t H(s)\, dX_s = X_T - X_S, \text{ for all } t \geq M.$$

The integral of a bounded process with respect to a martingale is again a martingale and hence,

$$\mathbb{E}[X_T] = \mathbb{E}[X_S]. \tag{19}$$

We use that T is bounded by a nonrandom time M with the following. Define,

$$\begin{aligned} S^B &= S\mathbb{1}_B + M\mathbb{1}_{B^c} \\ T^B &= T\mathbb{1}_B + M\mathbb{1}_{B^c}. \end{aligned}$$

S^B and T^B are stopping times, and if $B \in \mathcal{F}_S \subset \mathcal{F}_T$,

$$\mathbb{E}[X_T\mathbb{1}_B + X_M\mathbb{1}_{B^c}] = \mathbb{E}[X_S\mathbb{1}_B + X_M\mathbb{1}_{B^c}].$$

And therefore

$$\mathbb{E}[X_T\mathbb{1}_B] = \mathbb{E}[X_S\mathbb{1}_B] \text{ for all } B \in \mathcal{F}_S,$$

and this is $\mathbb{E}[X_T|\mathcal{F}_S] = X_S$.

If X is a stochastic process with $\mathbb{E}[X_0] \leq \infty$ and Equation (19) holds for bounded stopping times S and T, we can deduce that the stochastic process X is a martingale by setting $T = t\mathbb{1}_A + M\mathbb{1}_{A^c}$ and $S = s\mathbb{1}_A + M\mathbb{1}_{A^c}$.

The formulae for submartingales follow by taking $Y_t = \max(a, X_t)$, which is a integrable process and by the same reasonings as above. Sending $a \to -\infty$ completes the proof. □

Proof of Theorem 2.16. First, we show that the family $U := \{\mathbb{E}[X_\infty|\mathcal{G}] : \mathcal{G} \subset \mathcal{F}, \mathcal{G} \text{ is a } \sigma-\text{algebra}\}$ is uniformly integrable. Due to Jensen inequality and averaging property of conditional expectation we have

$$\begin{aligned} \mathbb{E}[|\mathbb{E}[X_\infty|\mathcal{G}]|\mathbb{1}_{|\mathbb{E}[X_\infty|\mathcal{G}]|\geq M}] &\leq \mathbb{E}[\mathbb{E}[|X_\infty||\mathcal{G}]\mathbb{1}_{|\mathbb{E}[X_\infty|\mathcal{G}]|\geq M}] \\ &\leq \mathbb{E}[|X_\infty|\mathbb{1}_{|\mathbb{E}[X_\infty|\mathcal{G}]|\geq M}] \\ &\leq \mathbb{E}[|X_\infty|] \cdot P(|\mathbb{E}[X_\infty|\mathcal{G}]| \geq M). \end{aligned}$$

As above we can use Markov inequality to bound the probability

$$\begin{aligned} P(|\mathbb{E}[X_\infty|\mathcal{G}]| \geq M) &\leq \frac{\mathbb{E}[\mathbb{E}[X_\infty|\mathcal{G}]|]}{M} \\ &= \frac{\mathbb{E}[X_\infty]}{M}. \end{aligned}$$

Hence, the expectation of the truncated random variables decays to zero uniformly.

We claim that the closure in $\mathcal{L}^1$ of U is also uniformly integrable. Indeed, if $X \in \bar{U}$ there must be $X_n \in U$ such that $X_n \to X$ in $\mathcal{L}^1$. Then

$$\mathbb{E}[|X|\mathbb{1}_{|X|\geq M}] \leq \mathbb{E}[|X_n|\mathbb{1}_{|X_n|\geq M}] + \mathbb{E}\left[|X|\mathbb{1}_{|X|\geq M} - |X_n|\mathbb{1}_{|X_n|\geq M}\right].$$

The first term on the right-hand side is bounded by ε with choosing M big enough. The second term can be controlled by looking at the three cases $M < |X_n| \wedge |X|$, $|X_n| > M$, and $|X| > M$. This yields the bound

$$\mathbb{E}\left[||X_n| - |X||\,\mathbb{1}_{|X_n|\wedge|X|>M}\right] + \mathbb{E}[|X_n|\mathbb{1}_{|X_n|\geq M}] + \mathbb{E}[|X|\mathbb{1}_{|X|\geq M}].$$

The first term goes to zero as n goes to infinity. The second term is less than ε for M big enough, as is the third term, because $X \in \mathcal{L}^1$. Note, that we can use the same M for the third term as we do for the second. This uniform choice of M ensures that the $\mathcal{L}^1$-closure of U is also uniform integrable.

Now, for any stopping time S there are stopping times S_k eventually decay to S, such that

$$S_k = \sum_{i=1}^{n_k} y_i^k \mathbb{1}_{A_i} + \infty \cdot \mathbb{1}_{A^c},$$

for sets A_i and $A = \cup A_i$. Then,

$$\begin{aligned} \mathbb{E}[X_\infty|\mathcal{F}_s] &= \mathbb{E}[X_\infty \mathbb{1}_{A^c} + X_\infty \mathbb{1}_A|\mathcal{F}_s] \\ &= X_{S_k}\mathbb{1}_{A^c} + X_{S_k}\mathbb{1}_A \\ &= X_{S_k}. \end{aligned}$$

Since S_k eventually go to S from above and X is càdlàg, $X_{S_k} \to X_S$ almost surely. But X_{S_k} is also uniformly integrable, so $X_{S_k} \to X_S$ also in $\mathcal{L}^1$.

Finally, let $C \in \mathcal{F}_S \subset \mathcal{F}_{S_k}$, then

$$\mathbb{E}[\mathbb{1}_C X_\infty] = \mathbb{E}[\mathbb{1}_A X_{S_k}] \to \mathbb{E}[\mathbb{1}_A X_S].$$

This is, $\mathbb{E}[X_\infty|\mathcal{F}_S] = X_S$. □

We have seen that uniform integrability is exactly the property needed to ensure $\mathcal{L}^1$-conergence of a stochastic processes that is known to converge almost surely. This result is especially useful for martingales since uniform integrability implies uniform boundedness of the martingale, and due to Theorem 2.10 the almost sure limit exists.

One way to interpret Brownian motion is to view it as a collection of normal random variables. Particularly, Brownian motion $(B_t)_{t\geq 0}$ stopped at a non-random time T is just a normal random variable with the following two properties

$$\mathbb{E}[B_T] = 0, \tag{20}$$

$$\mathbb{E}[B_T^2] = \mathbb{E}[T]. \tag{21}$$

The above two identities are called Wald's identities.

Let us examine what happens when we stop Brownian motion at a *random* time T. We would expect the same equations to hold, due to the martingale property of Brownian motion. However, this is not quite true. Consider

$$T^a = \inf\{t \geq 0 : X_t = a\}.$$

Brownian motion hits every point almost surely so that $\mathbb{E}[X_{T^a}] = \mathbb{E}[a] = a$ and Equation (20) is not satisfied. But it also takes Brownian motion in expectation infinitely long to hit a point, hence $\mathbb{E}[T_a] = \infty$. However, if we impose restriction on T, Equations (20) and (21) will hold.

For a stopping time T and a stochastic process $(X_t)_{t\geq 0}$ define the stopped process

$$(X_t^T)_t = (X_{t\wedge T})_t. \tag{22}$$

The stopped process X^T is a martingale if X is a martingale, because $X_t^T = \int H\ dX$, for $H = \mathbb{1}_{T\geq t}$ and a stochastic integral with respect to a martingale is still a martingale.

Let T be a stopping time, such that T is almost surely bounded by $N \in \mathbb{R}$, $T \leq N$ a.s. Then the process X^T is a uniformly integrable martingale. Indeed, X^T is a martingale, therefore using Jensen's inequality on taking absolute value, $\mathbb{E}[|M_t^T||\mathcal{F}_s] \geq |\mathbb{E}[M_t^T|\mathcal{F}_s]| = |M_s^T|$, yields that $|M_t|$ is a submartingale. Hence,

$$\mathbb{E}[|M_t^T|] \leq \mathbb{E}[|M_T|] = \mathbb{E}[|M_N^T|] < \infty,$$

because M^T is a martingale. The same reasoning can be applied to the process $(M^T)^2$, because $x \mapsto x^2$ is convex. Therefore, we can use Optional Stopping Theorem 2.16 on X^T and $(X_t^T)^2 - (t \wedge T)$ to establish both of Wald's identities, equations (20) and (21).

We can even let T be integrable, allowing arbitrarily large values of T.

Proposition 2.18 (cf. [16])**.** *Let T be an integrable stopping time, $\mathbb{E}[T] < \infty$. Then both of Wald's identities hold,*

$$\mathbb{E}[B_T] = 0, \tag{23}$$

$$\mathbb{E}[B_T^2] = \mathbb{E}[T], \tag{24}$$

and $(B_t^T)_t$ and $\big((B_t^T)^2 - (T \wedge t)\big)_t$ are uniformly integrable.

Proof. We first show that B^T has a majorizing function. To this end define

$$M_k := \sup_{t\in[0,1]} |B(k+t) - B(k)]$$

$$M := \sum_{k=1}^{\lceil T\rceil} M_k$$

Observe that each B_t^T is bounded by M

$$\begin{aligned} |B_t^T| &= |\sum^{\lceil T\rceil}(B(k+1) - B(k)) + B(T) - B(k)| \\ &\leq \sum^{\lceil T\rceil} |B(k+1) - B(k)| + |B(T) - B(k)| \\ &\leq M. \end{aligned}$$

It remains to check that M is integrable,

$$\mathbb{E}\left[|M|\right] = \mathbb{E}\left[\sum_{k=0}^{\infty} \mathbb{1}_{T>k} M_k\right].$$

By Beppo-Levi, an immediate consequence of the monotone convergence theorem,

$$\mathbb{E}\left[\sum_{k=0}^{\infty} \mathbb{1}_{T>k} M_k\right] \leq \sum_{k=0}^{\infty} \mathbb{E}\left[M_k \mathbb{1}_{T>k}\right].$$

We use the Markov Property of Brownian motion on $\mathbb{E}[M_k]\mathbb{1}_{T>k}]$. Using the averaging property of conditional expectation, the facts that $\mathbb{1}_{T>k} \in \mathcal{F}_k$ and $\mathbb{E}[M_k|\mathcal{F}_k]$ is independent of $\mathcal{F}_k$ and equals M_0, we compute

$$\begin{aligned} \mathbb{E}[M_k \mathbb{1}_{T>k}] &= \mathbb{E}[\mathbb{E}[M_k \mathbb{1}_{T>k}|\mathcal{F}_k]] \\ &= \mathbb{E}[\mathbb{1}_{T>k}\mathbb{E}[M_k|\mathcal{F}_k]] \\ &= \mathbb{E}[\mathbb{1}_{T>k}\mathbb{E}[M_0]] \\ &= P(T>k) \cdot \mathbb{E}[M_0]. \end{aligned}$$

Hence,

$$\mathbb{E}[|M|] = \mathbb{E}[\sup_{0\leq s\leq 1} |B_s|] \cdot \mathbb{E}[T].$$

The second term is finite by assumption. For the first term we use Doob's $\mathcal{L}^p$-inequality for $p = 2$, to get $P(\sup_{0\leq s\leq 1} |B_s| \geq \alpha) \leq \frac{\mathbb{E}[|B_1|^2}{\alpha^2}$.And consequently,

$$\begin{aligned} \mathbb{E}[\sup_{0\leq s\leq 1} |B_s|] &\leq \mathbb{E}[|B_1|^2] \cdot \left(1\frac{1}{1^2} + 1\frac{1}{2^2} + 1\frac{1}{3^2} + \dots\right) \\ &\leq 1 \cdot \text{const}\,. \end{aligned}$$

Hence, B^T is uniformly integrable and we can use optimal stopping, note that $B^T_\infty = B^T$, to establish the first Wald's identity.

If we also show that $\big(B_t^T)^2 - (T \wedge t)\big)_t$ is uniformly integrable, we are done by Optimal-Stopping-Theorem.

Fix t, then we can use Doob's $\mathcal{L}^p$-inequalities with $p = 2$ to show

$$\mathbb{E}\left[(\sup_{s\leq t} B_t^T)^2\right] \leq 4\mathbb{E}[(B_t^T)^2] = 4\mathbb{E}[T \wedge t], \tag{25}$$

where the last equality follows because $(B_t^T)^2 - (T \wedge t)$ is a martingale, since T is a stopping time and $B_t^2 - t$ is a martingale. The running maximum $\sup_{s\leq t} B_s$ of any process is monotone in t, and hence the left-hand side converges by Monotone Convergence Theorem to $\mathbb{E}[(\sup_{s\leq T} B_s)^2]$. The right-hand side, also by Monotone Convergence Theorem, converges to $4\mathbb{E}[T]$. This yields

$$\mathbb{E}[|(B_t^T)^2 - (T \wedge t)|] \leq 4\mathbb{E}[T] + \mathbb{E}[T] = 5\mathbb{E}[T] < \infty.$$

□

2.2. **Martingale Representation and Clark's formula.** Richard F. Bass 1983 in [13] employed stochastic integrals to construct a solution to the Skorokhod Embedding Problem. In [13] he uses a special type of Martingale Representation, and we shall take this opportunity to review general Martingale Representation and introduce Clark's formula, the specific type used by Bass.

In general, Martingale Representation Theorems assert that a martingale can be represented as a stochastic integral of a suitable random variable with respect to Brownian motion. The discrete time analouge of continuous time Brownian motion is a symmetric random walk. A discrete stochastic integral with respect to a symmetric random walk is a summation over suitable random variables γ_i times Bernoullie $(1/2, 1/2)$ random variables β_i.

Let $\Omega = \{-1, 1\}^{\mathbb{N}}$ the space of infinicte sequences of ones and zeros. Let $\pi_i(\omega) = \omega_i$ be the natural projections, $\mathcal{F} = \sigma(\pi_i, i \in \mathbb{N})$ and $\mathcal{F}_t = \sigma(\pi_i, i = 1, .., t)$. Let P be the unique probability on $(\Omega, \mathcal{F})$ such that $P(\pi_i = 1) = P(\pi_i = -1) = 1/2$ for all $i \in \mathbb{N}$. Recall that a process γ_t is predictable with respect to $\mathcal{F}_t$ if γ_t is F_{t-1} measurable. Then the discrete version of the martingale representation theorem reads as follows.

Theorem 2.19 (Discrete Time Martingale Representation)**.** *Let $(X_n)_n$ be a $\mathcal{F}_t$-martingale on $(\Omega, \mathcal{F}, P)$, then for every simple random walk $Y_k = \sum_{i=1}^k \beta_i$ there is a predicatable process $(\gamma_n)_n$ such that $X_n = X_0 + \sum_{i=1}^n \beta_i\gamma_i = X_0 + \int_{i=1}^n \gamma \; dY$.*

In the discrete time case we had to impose only one restriction on the martingale. This was to be defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=1}^N)$, that is the natural measurable space for a random walk. This ensures that the given martingale X inherits at most as much randomness as the simple random walk, a crucial fact necessary for the theorem to hold. If the martingale depends on something the simple random walk does not, the random walk is too poor to imitate the martingale. This observation is also crucial in the continuous time case. Due to increased complexities - one might say technicalities - further restrictions on the regularity of the martingale are necessary.

Theorem 2.20 (Continuous Time Martingale Representation)**.** *Let B be a Brownian motion on $(\Omega, \mathcal{F}, P)$ and let $\mathcal{F}_t$ be natural the filtration generated by B. Then, for any square-integrable (continuous) martingale M_t with $M_0 = 0$, there is a progressively measurable process ψ such that $E[\int_0^T \psi_t^2 \, dt] < \infty$, and*

$$M_t = \int_0^t \psi_s \, dB_s, 0 \leq t < \infty.$$

We will work on the canonical space for Brownian motion $(C, \mathcal{F}_\infty, \mathcal{F}_t, W)$ with $C = C([0,\infty), \mathbb{R})$ the continuous functions from the positive reals into the reals, $\mathcal{F}_\infty$ the sigma algebra generated by cylinders, $\mathcal{F}_t$ their restrictions to $[0,t]$, and W the Wiener measure. A Brownian Functional F is a random variable on (C, F_∞, W). F is square-integrable if $\mathbb{E}[F^2] < \infty$. For such functionals we have the following theorem:

Theorem 2.21 (Brownian Functionals as Stochastic Integrals)**.** *Let F be a $\mathcal{F}_\infty$-measurable square-integrable Brownian Functional, then there is a progressively measurable ψ with $\int_0^T \psi_t^2 \, dt < \infty$ such that*

$$F - \mathbb{E}[F] = \int_0^\infty \psi_t \, dB_t. \tag{26}$$

In various applications (such as hedging in mathematical finance) it has proven useful to know the above process ψ explicitly. An explicit expression is also of vital importance in the embedding by Bass, and we will be using Clark's formula for the specific functional $F(\omega) = g(B_1)$.

Now, we are ready to state Clark's formula for simple functionals, originally proven in [14]. A general version of Clark's formula holds for any square-integrable functional.

Theorem 2.22 (Clark's formula for simple functionals)**.** *Let $g \in C^2, B_t = \omega(t)$ be the canonical coordinates, and $F = g(B_1)$ a square-integrable functional. Then ψ in* (26) *is the conditional expectation of $g'(B_1)$. This is,*

$$g(B_1) - \mathbb{E}[g(B_1)] = \int_0^\infty \mathbb{E}[g'(B_1)|\mathcal{F}_t] \, dB_t. \tag{27}$$

The following is taken from [12], reproducing an approach by Bismut.

Definition 2.23. *In addition to square-integrability, we will require two more properties:*

- *We say that a Brownian functional F is quasi-continuous, if there is a non-negative and square-integrable Brownian functional h and a function $k : [0,\infty) \to [0,\infty)$ with $\limsup_{\varepsilon\to 0} \frac{k(\varepsilon)}{\varepsilon} < \infty$, such that $\forall(\omega,\varphi) \in \Omega^2$:*

$$|F(\omega+\varphi) - F(\omega)| \leq h(\omega)\ k(\|\varphi\|_\infty). \tag{28}$$

- *We say that a Brownian functional F is quasi-differentiable, if there is a measurable mapping $\omega \mapsto \partial F(\omega;\cdot) : (\Omega, \mathcal{F}_T) \to (\boldsymbol{M}, \mathcal{M})$, where $\boldsymbol{M}$ is the set of finite Borel measures on $\mathcal{B}([0,T])$ and $\mathcal{M}$ is the σ-algebra generated by the topology of weak convergence on $\boldsymbol{M}$, such that $\forall\varphi \in C^1([0,T])$:*

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[F(\omega+\varepsilon\varphi) - F(\omega)\right] = \int_0^T \varphi(t)\ \partial F(\omega;dt),\ \textit{for } P\textit{-a.e. } \omega.$$

Quasi-continuity and quasi-differentiability describe similar properties as their well-known counterparts in standard analysis. We note that for $g \in C^2$, $F(\omega) = g(B_1)$ is both quasi-continuous and quasi-differentiable. For the first, choose $h \equiv 2$ and $k(x) = x$. For the latter, choose $\partial F(\omega,\cdot) = g'(B_1)\delta_1$.

Theorem 2.24 (Clark's formula for general functionals)**.** *Let F be a square-integrable, quasi-continuous, and -differentiable Brownian functional. ψ in* (26) *is the predictable projection of the (not necessarily adapted) process $\partial F(B;(t,T])$. This is, for Lebesgue-almost-every $t \in [0,T]$, we have*

$$\psi_t = \mathbb{E}\left[\partial F(B;(t,T])|\mathcal{F}_t\right]\ \ P\textit{-a.s.} \tag{29}$$

Proof. The theorem will be shown with showing

$$\mathbb{E}\left[\int_0^T X_t\psi_t\ dt\right] = \mathbb{E}\left[\int_0^T X_t \partial F(B;(t,T])\ dt\right] \tag{30}$$

for every bounded, continuous, and adapted process X. For this end, let $\varphi_t = \int_0^t X_s ds$ and define the following exponential martingales and probability measures:

$$Z^\varepsilon = \exp\left\{\varepsilon \int_0^t X_s\ dB_s - \frac{\varepsilon^2}{2}\int_0^t X_s^2\ ds\right\}$$
$$P^\varepsilon(A) = \mathbb{E}\left[Z^\varepsilon \mathbb{1}_A\right],\ \text{for } A \in F_\infty$$

The Girsanov theorem (as presented e.g. in [10] 3.5) gives that $(B_t - \varepsilon\varphi_t)$ is a P^ε-Brownian motion and thus,

$$\mathbb{E}[F(B)] = \mathbb{E}^\varepsilon[F(B-\varepsilon\varphi)] = \mathbb{E}[Z^\varepsilon F(B-\varepsilon\varphi)] \tag{31}$$

Using (31):

$$\mathbb{E}[F(B) - F(B-\varepsilon\varphi)] + \mathbb{E}[(Z^\varepsilon - 1)(-F(B-\varepsilon\varphi))] = 0$$

And thus, adding one term and dividing by ε,

(32)

$$\frac{1}{\varepsilon}\mathbb{E}\left[F(B)-F(B-\varepsilon\varphi)\right]+\mathbb{E}\left[\frac{Z^\varepsilon-1}{\varepsilon}\left(F(B)-F(B-\varepsilon\varphi)\right)\right]=\mathbb{E}\left[F(B)\frac{Z^\varepsilon-1}{\varepsilon}\right].$$

The exponential martingales Z^ε solve the SDE $dZ^\varepsilon=\varepsilon X\ dB$ with initial condition $Z_0=1$, and so Z_t can be written as

$$Z_t^\varepsilon=1+\int_0^t \varepsilon Z_s^\varepsilon X_s\ dB_s.$$

Using this expression, one can deduce the following two

$$\lim_{\varepsilon\to 0}\mathbb{E}\left[\int_0^T (Z_s^\varepsilon-1)^2\ ds\right]=0 \tag{33}$$

$$\frac{Z_t^\varepsilon-1}{\varepsilon}\overset{L^2}{\to}\int_0^t X_s\ dB_s. \tag{34}$$

We send $\varepsilon\to 0$ in (32), differentiating F along φ, and control the three terms individually:

Using quasi-continuity of F, the first term on the left $\frac{1}{\varepsilon}\left(F(B)-F(B-\varepsilon\varphi)\right)$ is bounded, and using quasi-differentiability its limit is equal to $\int_0^T\varphi(t)\partial F(B;dt)$. Using quasi-continuity, the second term is bounded by $\mathbb{E}\left[|\frac{Z^\varepsilon-1}{\varepsilon}|h(B)\right]k(\varepsilon\|\varphi\|)$. By Cauchy-Schwarz the expected value goes to zero, as does $k(\varepsilon\|\varphi\|)$. The right hand side goes to $\mathbb{E}\left[F(B)\int_0^T X_s\ dB_s\right]$, because of (34).

So, (32) goes in L^2 to

$$\mathbb{E}\left[\int_0^T\varphi(t)\partial F(B;dt)\right]=\mathbb{E}\left[F(B)\int_0^T X_s\ dB_s\right]. \tag{35}$$

Using the definition of φ and Fubini, the left hand side is:

$$\mathbb{E}\left[\int_0^T\int_0^T \mathbb{1}_{0\le t<s\le T}X_t\ dt\partial F(B;dt)\right]=\mathbb{E}\left[\int_0^T X_t\partial F(B;(t,T])\ dt\right]. \tag{36}$$

Using Theorem 2.21 and properties of L^2-martingales, the right hand side equals

$$\mathbb{E}\left[\left\{E[F]+\int_0^T\psi_s\ dB_s\right\}\int_0^T X_s\ dB_s\right]=\mathbb{E}\left[X_t\psi_t\ dt\right]. \tag{37}$$

We have shown (30), and with that the theorem. □

Remark 2.25. *According to* [12] *equation* (31) *can be taken as a starting point for a stochastic calculus of variations, known as Malliavin calculus.*

Another type of martingale representation is the above mentioned Dambis, Dubin-Schwarz Theorem. It states that if we index a continuous time martingale with its quadratic variation, it is a martingale. In this sense, the quadratic variation process can be seen as the natural clock of a martingale and continuous time martingales are all the same, they just differ in terms of the velocity with which they evolve. We state without proof.

Theorem 2.26 (Dambis, Dubin-Schwarz, see [11], Theorem 1.6)**.** *Let $(M_t)_{t\geq 0}$ be a continuous local martingale with $M_0 = 0$ and almost sure unbounded quadratic variation process, $\langle M\rangle_\infty = \infty$. Then the process*

$$B_t := M_{T_t},$$

whith $T_t := \inf\{s \geq 0 : \langle M\rangle_s > t\}$ is a Brownian motion.

2.3. **Potential Theory.** On the real line, the potential U_μ of a measure μ is defined as $U_\mu(x) = \int_{\mathbb{R}} |x-t|\ \mu(dt)$. In the probabilistic context we can define the potential of a random variable X as

$$U_X(x) = \mathbb{E}[|X - x|]. \tag{38}$$

The potential U can be seen as an operator from $L^1 \to C(\mathbb{R}, \mathbb{R}_-)$. This operator possesses some surprising features which we will use in the Chacon-Walsh embedding. Let $L^{\mathbb{1}}_\mu$ be the space of integrable random variables with mean μ.

Theorem 2.27. *Let $\mu \in \mathbb{R}$ and $X \in L^{\mathbb{1}}_\mu$. Then*

(1) U_X is concave and finite
(2) $U_X(x) \leq U_{\delta(\mu)}(x) = -|x - \mu|$
(3) $X_1, X_2 \in L^{\mathbb{1}}_\mu, \lim_{|x|\to\infty} |U_{X_1}(x) - U_{X_2}(x)| = 0$
(4) $\mathbb{E}\left[(X-\mu)^2\right] = \int_{\mathbb{R}} \big||x| + U_X(x+\mu)\big|\ dx$
(5) $X_i \to X$ in distribution, if and only if $U_{X_i}(x) \to U_X(x)$ for all $x \in \mathbb{R}$
(6) Let $(B_t)_{t\in\mathbb{R}^+}$ be a martingale and $T_{a,b}$ the hitting time of $(a,b)^c$. Then $U_{B_T} = U_{B_0}$ on $(a,b)^c$ and U_{B_T} is linear on $[a,b]$.

Proof. we can assume without loss of generality, $\mu = 0$:

for (1): $U_X(x) < \infty$, because $\mathbb{E}|X - x| \leq \mathbb{E}|X| + \mathbb{E}|x| < \infty + |x|$. U_X is concave, because $\mathbb{E}|X - (\alpha x + \beta y)| \leq \mathbb{E}|\alpha X - \alpha x| + \mathbb{E}|\beta X - \beta y| = \alpha U_X(x) + \beta U_X(y)$.

for (2): for $x \leq 0$: $-\mathbb{E}|X - x| \leq -\mathbb{E}(X - x) = x = -|x|$. Likewise for $x \geq 0$.

for (3): observe that $\mathbb{E}X_1 = \mathbb{E}X_2$ implies $\mathbb{E}[(X_1 - x) - (X_2 - x)] = 0$. For $x \to +\infty$ one computes

$$\begin{aligned} U_{X_1}(x) - U_{X_2}(x) &= \mathbb{E}\left[|X_1 - x| - |X_2 - x|\right] \\ &= \mathbb{E}\left[|X_1 - x| + (X_1 - x) - (X_2 - x) + |X_2 - x|\right] \\ &= \mathbb{E}\left[2(X_1 - x)\ \mathbb{1}_{X_1 \geq x}\right] + \mathbb{E}\left[2(X_2 - x)\ \mathbb{1}_{X_2 \geq x}\right] \end{aligned}$$

As $\mathbb{E}|X_1| < \infty$, both expressions go to zero when x goes to $+\infty$. For $X \to -\infty$ similarly.

for (4): observe that

$$-|y| - U_X(y) = \begin{cases} 2\mathbb{E}[X-y; X \geq y], & y \geq 0 \\ 2\mathbb{E}[y-X; X < y] & y < 0. \end{cases}$$

Using that X is centred, $\mathbb{E}[X; X < y] = -\mathbb{E}[X; X \geq y]$ and trivially $P(X \geq y) = 1 - P(X < y)$. Using these one proves for $y > 0$:

$$\begin{aligned} -y + \mathbb{E}[|X-y|] &= -y + \mathbb{E}[X-y; X \geq y] + \mathbb{E}[y-X; X < y] \\ &= \mathbb{E}[X; X \geq y] - \mathbb{E}[X; X < y] + y\big(-1 - P(X \geq y) + P(X < y)\big) \\ &= 2\ \mathbb{E}[X; X \geq y] - 2\ \mathbb{E}[-y; X \geq y] \end{aligned}$$

Similarly, for $y < 0$. Then, using the above (4) is proven with the following computation:

$$\begin{aligned} &\tfrac{1}{2}\int_{-\infty}^{\infty} -|y| - U_X(y)\ dy \\ &= \int_{-\infty}^{0} \mathbb{E}\left[y-X; X<y\right]\ dy + \int_{0}^{\infty} \mathbb{E}\left[X-y; X \geq y\right]\ dy \\ &= \mathbb{E}\left[\int_X^0 y - X\ dy; X < 0\right] + \mathbb{E}\left[\int_0^X X - y\ dy; X \geq 0\right] \\ &= \mathbb{E}\left[\int_X^0 y\ dy - X\int_X^0 1\ dy; X < 0\right] + \mathbb{E}\left[X\int_0^X 1\ dy - \int_0^X y\ dy; X \geq 0\right] \\ &= \mathbb{E}\left[-\tfrac{X^2}{2} + X^2; X < 0\right] + \mathbb{E}\left[X^2 - \tfrac{X^2}{2}; X \geq 0\right] = \mathbb{E}\left[X^2\right]\tfrac{1}{2}. \end{aligned}$$

for (6): Define $I_1 = \mathbb{1}_{B_0 \leq a}, I_2 = \mathbb{1}_{a<B_0<b}, I_3 = \mathbb{1}_{b \leq B_0}$ and $B^i = B * I_i, i = 1,2,3$, so that $B_t = B_t^{\mathbb{1}} + B_t^2 + B_t^3$. Define $V_X^i(x) = -\mathbb{E}|I_i(X-x)|$, then $U_{B_0} = \sum_i V^i_{B_0^i} I_i$, and $B_T = \sum_i V_{B_T^i} I_i$. We check that $U_{B_0} = U_{B_T}$ for every B^i separately. Observe that $T * I_1 = T * I_3 = 0$, so the potentials of B^1 and B^3 agree for all x. For B_2, if $x \leq a$, $\mathbb{E}|B_0^2 - x| = \mathbb{E}B_0^2 - x = \mathbb{E}B_T^2 - x = \mathbb{E}|B_T^2 - x|$. For $x \geq b$ similarly. This is, $U_{B_0} = U_{B_T}$ for $x \in (a,b)^c$.

To prove the linearity of U_{B_T} for $x \in [a,b]$, check that

$$I_i * V^i_{X_T} = \begin{cases} I_1 * (x - \mathbb{E}B_0^{\mathbb{1}}), & \text{if } i = 1 \\ I_2 * \left(\frac{b-\mathbb{E}B_0^2}{b-a}(x-a) + \frac{\mathbb{E}B_0^2 - a}{b-a}(b-x)\right) & \text{if } i = 2 \\ I_3 * (\mathbb{E}B_0^3 - x) & \text{if } i = 3. \end{cases}$$

Observe that each of the three functions is linear in x and that $U_{B_T}(x)$ is the sum of these three linear functions. □

3. Skorokhod embedding

We first present the original version of the Skorokhod embedding problem as proposed by Skorokhod. Suppose T is an integrable stopping time and B is a Brownian motion. If we stop B at the stopping time T, the resulting random variable B_T has mean zero and second moment equal to $\mathbb{E}[T]$ as assured by Wald's identities, Proposition 2.18.

Skorokhod proposed 1965 the reversed question: given a distribution μ with mean zero and finite second moment, find a stopping time T such that $B_T \sim \mu$!

A first answer to this question may be a bit disappointing as it somehow cheating. Let F_μ be the distribution function of μ, F_μ^{-1} its right-continuous inverse and Φ the distribution function of a standard normal random variable. Observe that $(\mathcal{F}_\mu^{-1} \circ \Phi)(B_1) \sim \mu$ and that therefore $T = \inf\{t \geq 1 | B_t = (\mathcal{F}_\mu^{-1} \circ \Phi)(B_1)\}$ embeds μ in Brownian motion. T is almost surely finite, but unless $F_\mu = \Phi$, $\mathbb{E}T = \infty$, because hitting times of single points of Brownian motion are infinite in expectation.

Of vital importance in the construction of the above stopping time was that $g := (\mathcal{F}_\mu^{-1} \circ \Phi)(B_1) \sim \mu$. Let ν be the distribution of a standard normal random variable, then one says g transports ν to μ. A relatively modern brand of mathematics, called optimal transport theory, is examining transport problems of this form. The interested reader may be directed to the standard opus on optimal transportation by Cédric Villani [7]. Whilst most of present-day research takes place in more complicated spaces than the real line, the latter is much easier to handle and it is well known that g is the optimal transportation plan in a broad spectrum of cases of cost functions.

The above construction works for any probability μ not necessarily centred or of finite second moment. It is therefore clear that we should impose some restrictions on T. Skorokhod required μ to have finite second moment σ^2 and T be integrable. Due to Wald's identities, Proposition 2.18, this readily implied for a stopping T embedding μ, $B_T \sim \mu$, that the expectation of T to be equal the second moment of μ, $\mathbb{E}[T] = \sigma^2$. Problem 3.1 is the original form of the problem as proposed by Skorokhod.

Problem 3.1 (Original Skorokhod Embedding Problem)**.** *For a given centred probability measure μ on $\mathbb{R}$ with finite second moment, find a integrable stopping time T such that $B_T \sim \mu$.*

Obłój [6], finds that B_t^T is uniformly integrable is a more natural restriction. Below we state the modern version Skorokhod embedding problem formally and according to [6]:

Problem 3.2 (Modern Version of Skorokhod Embedding Problem)**.** *For a given centred probability measure μ on $\mathbb{R}$ with finite first moment, find a stopping time T such that $B_T \sim \mu$ and $(B_{t\wedge T} : t \geq 0)$ is uniformly integrable.*

Since the proposition of the problem in 1961 many different solutions have been developed using different tools from different areas in mathematics. The first solution, presented by Skorokhod required external randomization (the stopping time T was allowed to depend on a bigger filtration than the natural filtration of the Brownian motion). Other authors managed to find results which did not require external randomization and featured different properties. For example, Root's embedding allows for a visualisation, or Bass' solution gives an easy-to-handle expression for calculations, which we will exploit in proving the Brascamp-Lieb moment inequality. In this document we will present two solutions. The embedding by Bass, that uses martingale methods, and the embedding by Chacon and Walsh, which relies on potential theory.

3.1. **Embedding by Bass.** Given a centred probability measure μ on $\mathbb{R}$ with finite first moment. Our aim is to find a stopping time T such that $B_T \sim \mu$ and $\{B_{t\wedge T} | t \geq 0\}$ is uniformly integrable.

For any distribution μ, let $F_\mu(t) = \mu\left((-\infty, t]\right)$ be the cumulative distribution function of a random variable with distribution μ. Let F_μ^{-1} be the generalized inverse of F_μ, this is $F_\mu^{-1}(x) = \inf\{t \in \mathbb{R} | F_\mu(t) \geq x\}$. Let Φ denote the cumulative distribution function of a standard normal variable. Then define

$$g(x) = (F_\mu^{-1} \circ \Phi)(x). \tag{39}$$

This function g transports $\mathcal{N}(0,1)$ to μ in the sense of $B_1 \sim \mathcal{N}(0,1)$ and $g(B_1) \sim \mu$.

The idea of the Bass embedding is the following: if g is smooth and squared-integrable, using Clark's formula (Theorem 2.24) we can write:

$$g(B_1) - \mathbb{E}[g(B_1)] = \int_0^1 \mathbb{E}[g'(B_1)|\mathcal{F}_s] dB_s = \int_0^1 a(s, B_s) dB_s, \tag{40}$$

for $a(s,y) = \mathbb{E}[g'(B_1)|B_s = y]$. Also define a martingale

$$M_t = \int_0^t a(s, B_s)\; dB_s,$$

with $M_1 \sim g(B_1) - \mathbb{E}[g(B_1)]$. Then, using Dambis-Dubins-Schwarz, we get a Brownian motion X with $X_t = M_{\tau(t)}$, where $\tau(t)$ denotes the time until which M has acquired t quadratic variation,

$$\int_0^{\tau(t)} a^2(s, B_s)\; ds = t.$$

Then for

$$T = \int_0^1 a^2(s, B_s)\ ds,$$

the stopped Brownian motion X has distribution equal to $g(B_1) - \mathbb{E}[g(B_1)]$ and for g as in (39) equal to μ. This is,

$$X_T \sim \mu.$$

In order to justify the above, the following steps are to be done:

(1) identify functions g to which this procedure can be generalized
(2) show that M_t is a martingale
(3) show that T is a stopping time with respect to the natural filtration of X
(4) for an arbitrary Brownian motion W_t find a stopping time S such that $W_S \sim \mu$

Let us examine the term $\mathbb{E}[g'(B_1)|\mathcal{F}_t]$ in Clark's formula more closly. Observe, due to the Markov property of Brownian motion $\mathbb{E}[g'(B_1)|\mathcal{F}_t] = \mathbb{E}_{X_t}[g'(B_{1-t})]$. Then the function $a : [0,1] \times \mathbb{R} \to \mathbb{R}, (t, y) \mapsto \mathbb{E}[g'(B_{1-t} + y)]$ satisfies the following

$$\begin{aligned} \mathbb{E}\left[g'(B_1)|\mathcal{F}_s\right] &= a(s, B_s) \\ a(s, y) &= \int g'(z) p_{1-s}(z - y)\ dz, \end{aligned}$$

where $p_s(t)$ is the density function of a centred Gaussian random variable with variance s.

The function a can be computed explicitly

$$\begin{aligned} a(s, y) &= \int_{\mathbb{R}} g'(z)\ p_{1-s}(z - y)\ dz \\ &= -\int_{\mathbb{R}} g(z)\ q_{1-s}(z - y)\ dz \\ &= \int_{\mathbb{R}} g(z)\ q_{1-s}(y - z)\ dz, \end{aligned}$$

using an integration by parts, $q_s = p'_s$ and $q(t) = -q(-t)$.

Similarly, due to the Markov property of Brownian motion

$$\mathbb{E}[g(B_1)|\mathcal{F}_t] = \mathbb{E}_{B_t}[g(B_{1-t})],$$

and the function $b : [0,1] \times \mathbb{R} \to \mathbb{R}, (t, y) \mapsto \mathbb{E}[g(X_{1-t} + y)]$ satisfies the following,

$$\begin{aligned} \mathbb{E}[g(B_1)|\mathcal{F}_t] &= b(s, B_t) \\ b(s, y) &= \int_{\mathbb{R}} p_{1-s}(z - y)\ g(z)\ dz. \end{aligned}$$

Trivially conditional expectations of a integrable random variable is a martingale. Consider $Y_t = \mathbb{E}[Z|\mathcal{F}_t]$, since the conditional expectation of a conditional expectation is equals the condtional expectation with respect to the smaller σ-algebra, for $s \leq t$ we have: $\mathbb{E}[Y_t|\mathcal{F}_s] = \mathbb{E}[\mathbb{E}[Z|\mathcal{F}_t]|\mathcal{F}_s] = \mathbb{E}[Z|F_s] = Y_s$. Hence, $M_t = \mathbb{E}[g(B_1)|\mathcal{F}_t]$ is a martingale. Applying Ito's formula (see e.g. [10] or any book on stochastic calculus) to $b(t, B_t)$ the dt-terms cancel, because $b(t, B_t)$ is a martingale, and hence

$$b(t, B_t) - b(0, B_0) = \int_0^t b'(s, B_s)\ dB_s, \tag{41}$$

where $b'(s, B_s) = \frac{\partial}{\partial t} b(t, B_t)$ is the partial derivative of b with respect to its first coordinate. Replacing $b(t, B_t)$ with $\mathbb{E}[g(B_1)|\mathcal{F}_t]$ and $b(0, B_0)$ with $\mathbb{E}[g(B_1)|\mathcal{F}_0] = \mathbb{E}[g(B_1)]$ we can write

$$\mathbb{E}[g(B_1)|\mathcal{F}_t] - \mathbb{E}[g(B_1)] = \int_0^t b'(s, B_s)\ dB_s. \tag{42}$$

In order to link $b'(s, B_s)$ to the equation in Clark's formula we compute

$$\begin{aligned} b(t, y) &= \mathbb{E}[g(B_{1-t} + y)] \\ &= \int_{\mathbb{R}} g(x + y) p_{1-t}(x)\ dx \\ &= \int_{\mathbb{R}} g(x) p_{1-t}(x - y)\ dx \\ &=: \int_{\mathbb{R}} k(x, y, t)\ dx, \end{aligned}$$

where the function k of course equals $k(x, y, t) = g(x)p_{1-t}(x - y)$. We employ the following lemma.

Theorem 3.3 (Differentiation lemma, cf. Lemma 16.2 in [19])**.** *Let I be a non-empty open interval. The function $f : I \times \Omega \to \mathbb{R}$ be such that*

(1) $x \mapsto f(t, x)$ is integrable for every t,
(2) $t \mapsto f(t, x)$ is differentiable for all x with derivative $\frac{\partial f}{\partial t}$, and
(3) there is a Lebesgue integrable function $h \geq 0 : \mathbb{R} \to \mathbb{R}$, such that $|\frac{\partial f}{\partial t}(t, x)| \leq h(x)$.

Then for all x in $\mathbb{R}$ the following holds,

$$\frac{d}{dt} \int_{\mathbb{R}} f(x, t)\ dx = \int_{\mathbb{R}} \frac{\partial}{\partial t} f(x, t)\ dx. \tag{43}$$

Proof. Let $F(t,x) := \int f(t,x)\ dx$. Then

$$\begin{aligned}\lim_h \frac{F(t+h,x)-F(t,x)}{h} &= \lim_h \int \frac{f(t+h,x)-f(t,x)}{h}\ dx \\ &= \lim_h \int f'(\xi(t,h,x),x)\ dx,\end{aligned}$$

and $\xi \in (t, t+h)$. Due to assumption 3, $f' \le h$ and we can use Dominated Convergence Theorem,

$$\begin{aligned}\lim_h \int f'(\xi(t,h,x),x)\ dx &= \int \lim_h f'(\xi,x)\ dx \\ &= \int f'(t,x)\ dx.\end{aligned}$$

The above expressions are well-defined due to assumptions 1 and 2. □

The Gaussian density $p_s(t)$ is smooth, so is its derivative q. Hence, the function k and its partial derivative in t are continuous if g is continuous. For assumption 3 in Theorem 3.3, observe that there is Z_0 and γ such that for $k \ge 0, t \le \gamma < 1$, and $z \ge Z_0$ we have $|z-y|^k \exp\{\frac{(z-y)^2}{2t}\} \le \exp\{-\frac{z^2}{2}\}$. And hence on intervals J bounded away from zero $f'(t,x) \le h(x) = \mathbb{1}_{|x|<Z_0} \sup f' + \mathbb{1}_{|x|\ge Z_0} \exp\{-\frac{z^2}{2}\}$. This justifies differentiating under the integral sign for all $t < 1$.

Therefore, using Equation (42), and the above Theorem 3.3 on $\frac{d}{dt}\int k(x,y,t)\ dx$ so that

$$\begin{aligned}b'(t,y) &= \int \frac{\partial}{\partial t} k(x,y,t)\ dx \\ &= \int g(x) q_{1-t}(x-y)\ dx \\ &= a(t,y),\end{aligned}$$

we have shown the following:

Theorem 3.4. *Let the function $g : \mathbb{R} \to \mathbb{R}$ be continuous and such that $g(B_1)$ is integrable. Then*

$$M_t = \mathbb{E}[g(B_1)|\mathcal{F}_t] - \mathbb{E}[g(B_1)]$$

is a martingale and the following holds.

$$\mathbb{E}[g(B_1)|\mathcal{F}_t] - \mathbb{E}[g(B_1)] = \int_0^t \bar{a}(s,B_s)\ dB_s, \tag{44}$$

where $\bar{a} = \int_{\mathbb{R}} g(z)\ q_{1-s}(y-z)\ dz$. Additionally, if g is $\mathcal{C}^1$, $\bar{a} = a$ and we recover Equation (40)

$$g(B_1) - \mathbb{E}[g(B_1)] = \int_0^1 \mathbb{E}[g'(B_1)|\mathcal{F}_s] dB_s = \int_0^1 a(s,B_s) dB_s. \tag{45}$$

We particularly have established the following corollary.

Corollary 3.5. *If* $\mathbb{E}[g(B_1)] = 0$*, such as for centred* μ *in* (39),

$$M_t = \mathbb{E}[g(B_1)|F_t] = \int_0^1 a(s, B_s)\ dB_s, \tag{46}$$

and $M_t := \int_0^t a(s, B_s)\ dB_s$ *is a martingale.*

Next, we want to transform the martingale to a Brownian motion and establish that the transformed stopping time 1 is a stopping time in the natural filtration of the Brownian motion. To this end, let μ be a centred probability measure on $\mathbb{R}$.

Note that there are stopping times T_n such that M^{T_n} is a squared-integrable local martingale. Therefore, the quadratic variation process exists for all $(M_t)_{t\geq 0}$ and let it be denoted by $\langle\ \rangle$. Then, we define the stochastic processes η and τ

$$\eta(t) = \langle M \rangle_t = \int_0^t a^2(s, B_s)\ ds,$$
$$\tau(t) = \inf\{r|\ \eta(r) \geq t\}.$$

Let

$$N_t := M_{\tau(t)}.$$

N_t has quadratic variation of t and Dambis-Dubin-Schwarz Theorem 2.26 ensures that N_t is a Brownian motion. Observe, that

$$N_{\eta(1)} = M_1 \sim \mu.$$

With proving the following theorem, we will have established that $\eta(1)$ is a stopping time in the right-continuous completion of the natural filtration of the Brownian motion $N_t, t \geq 0$.

Theorem 3.6. *Let* $\mathcal{F}_u = \sigma(N_t\ |t \leq u)$*.* $\{\eta(1) \geq u\}$ *is in the right continuous completion* $\mathcal{F}_u^+ := \bigcap_{\varepsilon>0} \mathcal{F}_{u+\varepsilon}$*.*

Corollary 3.7. $\eta(1)$ *is a* $\mathcal{F}_t^+$*-stopping time.*

Proof. We have to show that $\{\eta(1) \leq u\}$ is in the right continuous completion $\mathcal{F}_u^+ = \bigcap_{\varepsilon>0} \mathcal{F}_{u+\varepsilon}$. Given $\bar{\varepsilon}, \{T \leq u\} = \bigcap_{0<\delta<\bar{\varepsilon}/2}\{T < u + \delta\}$. But for every δ, due to the above Theorem 3.6, $\{T < u+\delta\} = \{T \geq u+\delta\}$ is in $\mathcal{F}_{u+\delta}^+$ and since $\delta < \bar{\varepsilon}/2$ in $\mathcal{F}_{u+2\delta} \subset \mathcal{F}_{u+\bar{\varepsilon}}$. ☐

In order to proof the theorem we need a technical lemma from [13] on the regularity of the functions $a(s, y), b(s, y)$ and the inverse of $b(s, \cdot)$. Indeed, b can be seen as a function from the possible states of the Brownian motion at time s to the expectation of the random variable $g(B_1)$. The lemma states that the inverse exists and is regular.

Lemma 3.8 (cf. [13], Lemma 2). *Let g be a non-decreasing function $g : \mathbb{R} \to \mathbb{R}$ and not identically constant. Then*

(1) On compact subsets of $[0, 1) \times \mathbb{R}$, $a(s, y)$ is bounded above, bounded below away from 0, and uniformly Lipschitz in s and y.
(2) For each $s < 1, b(s, y)$ is continuous and strictly increasing as a function of y
(3) For each $s < 1$, let $b^{-1}(s, \cdot)$ be the inverse of $b(s, \cdot)$. Then on compact subsets of its domain, $b^{-1}(s, y)$ is uniformly Lipschitz in s and jointly continuous in s and y.

We continue with the proof of the above theorem.

Proof of Theorem 3.6. Bass uses a differential equation argument from Yershov [15]. Due to the lemma $\eta(1) < \infty$, also observe that $a^2(s, B_s)$ is continuous for almost every ω. Therefore, one can use standard integration theory and Monotone Convergence Theorem to write:

$$\eta(1) = \int_0^1 a^2(s, B_s)\ ds = \lim_{t\to 1} \int_0^t a^2(s, B_s)\ ds = \lim_{t\to 1} \eta(t).$$

Therefore, it suffices to consider only $\{\eta(t) \geq u\}$ for $s < 1$ due to the following

$$\{\eta(1) \geq u\} = \bigcup_{n\in\mathbb{N}} \{\eta(1 - \tfrac{1}{n}) \geq u\}.$$

Note, that η is continuously differentiable for almost every ω, and the following equation holds pathwise,

$$d\eta(t) = a^2(t, B_t)\ dt.$$

Due to Lemma 3.8 $a(s, y)$ is bounded away from zero, therefore η is strictly increasing and C^1. Therefore, τ is its true inverse and one computes

$$\frac{d\tau(t)}{dt} = \frac{1}{d\eta(\tau(t))/dt} = \frac{1}{a^2\left(\tau(t), B_{\tau(t)}\right)}.$$

Using the inverse of the function $b(s, y)$, $B_{\tau(t)}$ can be written as

$$B_{\tau(t)} = b^{-1}(\tau(t), M_{\tau(t)}) = b^{-1}(\tau(t), N_t),$$

and thus, for every ω, $\tau(t)$ satisfies the ordinary differential equation

$$\frac{d\tau(t)}{dt} = \frac{1}{a^2\left(\tau(t), b^{-1}(\tau(t), N_t)\right)}. \tag{47}$$

Consider the set $G := \{(\tau(t), N_t) \mid \tau(t) \leq u\}$. This set is contained in the rectangle $R := [0, u] \times [N_{\min}, N_{\max}]$, with min and max denoting points in $[0, u]$ where N_t attains its maximum and minimum. They exist because N_t is continuous.

G therefore is contained in a compact subset of the domain of b^{-1}. This, the regularity of coefficients in equation (47) by Lemma 3.8, and a theorem on uniqueness of solutions of differential equations show that there is a unique solution to (47) up to some t_0 such that $\tau(t_0) = u$. Furthermore, the solution can be constructed with Picard iteration.

This means, if $u \leq t_0$, τ can be constructed from $N_t, 0 \leq t \leq u$ to find $\tau(u) \leq s$. If $u > t_0$, τ can be constructed from $N_t, 0 \leq t \leq t_0$ to find $\tau(t_0) = s$ and conclude $\tau(u) > \tau(t_0) = s$. This is, $\{\tau(u) \leq s\}$ is in $\mathcal{F}_s$. □

We have proven the following theorem:

Theorem 3.9 (Bass embedding for a given Brownian motion)**.** *If μ is a square-integrable probability distribution on $\mathbb{R}$, there is a Brownian motion $\{N_t, t \geq 0\}$ and a stopping time $\eta(1)$ such that the stopped Brownian motion has distribution μ, that is $N_{\eta(1)} \sim \mu$.*

We do not know yet that the stopping time $\eta(1)$ is integrable. Let us give a formal proof.

Theorem 3.10. *If μ is square integrable, $\int x^2 \mu(dx)$, then the stopping time*

$$T = \int_0^1 a^2(s, B_s)\ ds$$

is integrable and

$$\mathbb{E}[T] = \int x^2 \mu(dx).$$

We will need the Burkholder-Davis-Gundy Inequalities. We state without proof.

Theorem 3.11 (Burkholder-Davis-Gundy, see [11], Theorem 4.1)**.** *Let $\langle\ \rangle$ denote the quadrativ variation process and $p \in (0, \infty)$. There are constants c_p and C_p such that for all continuous local martingales M with $M_0 = 0$,*

$$c_p \mathbb{E}[\langle M \rangle_\infty^{\frac{p}{2}}] \leq \mathbb{E}[(M_\infty^*)^p] \leq C_p \mathbb{E}[\langle M \rangle_\infty^{\frac{p}{2}}].$$

Proof of Theorem 3.10 cf. [13]*, page 5.* We use Doob's $\mathcal{L}^p$ inequality for $p = 2$ to get

$$\mathbb{E}[\sup_{s \leq 1} M_s^2] \leq \sup_{s \leq 1} \mathbb{E}[M_s^2] = \mathbb{E}[M_1^2].$$

For the non-random stopping time 1 the stopped process is M^1 and $M_\infty^{\mathbb{1}} = M_1 \sim \mu$. Theorem 3.11 gives a constant $c_2 < \infty$ such that

$$c_2 \mathbb{E}[\langle M_\infty^{\mathbb{1}} \rangle] = c_p \mathbb{E}[\langle M_1 \rangle] \leq \mathbb{E}[\sup_{t \leq 1} M_t^2] \leq \mathbb{E}[M_1^2] < \infty.$$

Since $\mathbb{E}[\langle M_\infty^{\mathbb{1}} \rangle] = \int_0^1 a^2(s, B_s) = T$ the proof is complete. □

But we want more. We only have found a stopping time for a specific Brownian motion, that was constructed from the martingale M_t. We are interested in stopping times for an arbitrary Brownian motion.

Theorem 3.12 (Bass embedding for an arbitrary Brownian motion)**.** *Let μ be a square-integrable probability distribution on $\mathbb{R}$ and $\{X_t, t \geq 0\}$ a Brownian motion. Then there is a integrable stopping time T such that the stopped Brownian motion has distribution μ, this is $X_T \sim \mu$ and T is a solution to the original Skorokhod embedding problem.*

Proof. We use an argument based on the differential equation considerations above.

For every ω, let $\bar{\tau}(t) = \bar{\tau}(t)(\omega)$ be the unique solution to

$$\frac{d\bar{\tau}(t)}{dt} = \frac{1}{a^2\left(\bar{\tau}(t), b^{-1}\left(\bar{\tau}(t), X_t\right)\right)}. \tag{48}$$

Let $\bar{\eta}$ be the inverse to $\bar{\tau}, \bar{\eta}(t) := \bar{\tau}^{-1}(t)$ for $t < 1$ and let $T = \lim_{t\to 1} \bar{\eta}(t)$. Then note that the law of (X, T) is the same as the law of $(N, \eta(1))$, implying $X_T \sim N_{\eta(1)} \sim \mu$. ☐

3.2. **Embedding Chacon-Walsh.** The Chacon-Walsh embedding is one representative of the family of embeddings that use potential theory.

Using potential theory - and in particular the results of theorem (2.27) - Chacon and Walsh constructed a solution to the Skorokhod embedding problem. Let $T_{(a_i,b_i)}$ be the hitting time of $(a,b)^c$ and θ the standard shift operator on $C([0,\infty),\mathbb{R})$, $\theta_s \circ \omega(t) = \omega(t+s)$.

Theorem 3.13 (Embedding by Chacon and Walsh, cf. [8])**.** *Let μ be a square-integrable measure on $\mathbb{R}$ and X a random variable with distribution μ, let*

$$U_0(t) := U_{\delta_o}(t) = -|t|,$$

and let a_n, b_n be points such that the line segment connecting $(a_n, U_n(a_n))$ and $(b_n, U_n(b_n))$ is above U_X and for $n \geq 1$, let U_n be such that $U_{n+1} = U_n$ outside $[a_n, b_n]$ and linear in between.

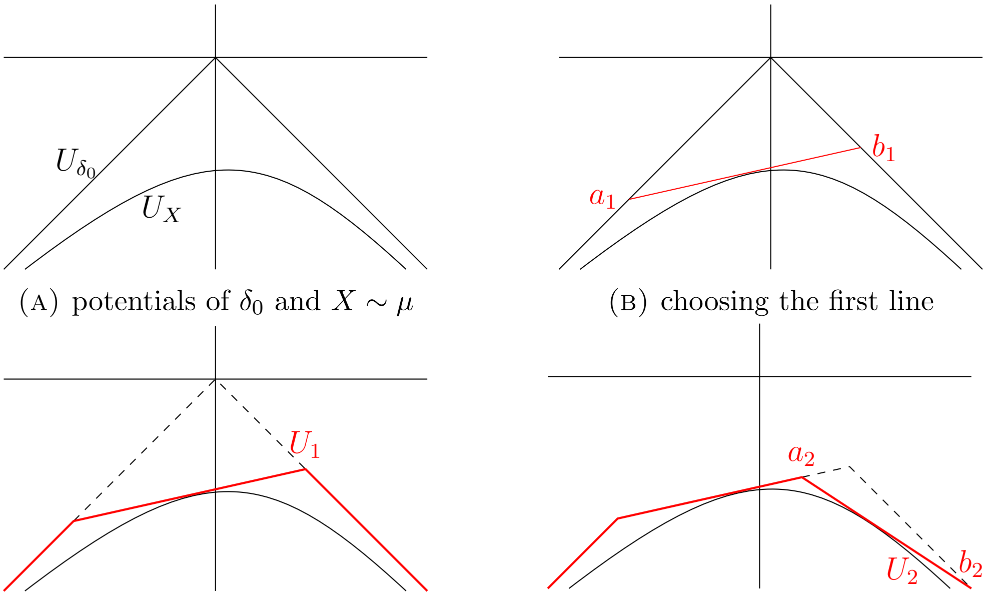
Then for $T_i = T_{(a_i,b_i)} \circ \theta_{T_{i-1}} + T_{i-1}$, the stopping time

$$T := \lim_{i\to\infty} T_i,$$

is well defined, integrable and $X_T \sim \mu$. This is, T is a solution to the original Skorokhod embedding problem (3.2)*.*

Proof. We reproduce a proof similar and in the same spirit as in [8], using the Figures 1a to 1d to explain the main idea:

And the following words explain the Figures 1a to 1d... Due to theorem (2.27)-1 $U_{\delta_0} \geq U_X$. Therefore, we can find a line above U_X intersecting U_{δ_0} at two points a_1



(A) potentials of δ_0 and $X \sim \mu$ (B) choosing the first line (C) potential of B_{T_1} (D) potential of B_{T_2}

FIGURE 1

and b_1. We then define a new function U_1 equal to $U_0 = U_{\delta_0}$ and linear in between. Due to theorem (2.27)-6 U_1 is the potential of B_{T_1}. We continue in this manner and choose lines above U_X, and define $U_n = U_{n-1}$ outside (a_n, b_n) and linear in between.

It is natural to choose tangent lines to U_X as indicated in the pictures, but, in fact, it is sufficient to choose lines above U_X (and therefore a_i and b_i) such that $U_{T_i}(x) \to U_X(x)$ for every $x \in \mathbb{R}$. This is also possible because U_X is concave, and concave functions can be written as a supremum over a countable number of affine linear functions. Using theorem (2.27)-3 we get that B_{T_i} convergence in distribution to a μ-distributed random variable, and because Brownian motion has continuous paths and $T = \lim T_i$ this random variable must be B_T. Particularly, $B_T \sim \mu$.

It remains to show that T is integrable. We compute the area between the potential $U_{\delta_0}(x) = -|x|$ and U_X. Since μ is squared-integrable and centred we have

$$\int x\mu(dx) = 0,$$
$$\int x^2\mu(dx) < \infty,$$
$$\int |x|\mu(dx) < \infty.$$

Therefore, for $y < 0$,

$$
\begin{aligned}
-|y| - U_\mu(y) &= -|y| - \int_{\mathbb{R}} |y - x| \ \mu(dx) \\
&= -\int_{-\infty}^{y} -y - (y - x) \ \mu(dx) - \int_{y}^{\infty} -y + (y - x) \ \mu(dx) \\
&= \int_{-\infty}^{y} 2y - x \ \mu(dx) + \int_{y}^{\infty} x\mu(dx) \\
&= \int_{-\infty}^{y} 2y - x \ \mu(dx) - \int_{-\infty}^{y} x\mu(dx) \\
&= 2\int_{-\infty}^{y} y - x \ \mu(dx).
\end{aligned}
$$

Similarly for $y \geq 0$,

$$
-|y| - U_\mu(y) = 2\int_{y}^{\infty} (x - y) \ \mu(dx).
$$

Using these two formulas and a change in order of integration we can compute the area between the two potentials,

$$
\begin{aligned}
&\int_{-\infty}^{\infty} - |y| - U_\mu(y) \ dy \\
&\quad = 2\int_{-\infty}^{0}\int_{-\infty}^{y} (y - x) \ \mu(dx) \ dy + 2\int_{0}^{\infty}\int_{y}^{\infty} (x - y) \ \mu(dx) \ dy \\
&\quad = 2\int_{-\infty}^{0}\int_{x}^{0} (y - x) \ dy \ \mu(dx) + 2\int_{0}^{\infty}\int_{0}^{y} (x - y) \ dy \ \mu(dx) \\
&\quad = 2\int_{-\infty}^{0} \frac{x^2}{2} \ \mu(dx) + 2\int_{0}^{\infty} \frac{x^2}{2} \ \mu(dx) \\
&\quad = \int_{-\infty}^{\infty} x^2 \ \mu(dx).
\end{aligned}
$$

Observe that $B_{t\wedge T_n}$ is bounded and hence $B^2_{t\wedge T_n} - (t \wedge T_n)$ is a uniformly integrable martingale, letting t go to infinity, we have

$$
\mathbb{E}[T_n] = \mathbb{E}[B^2_{T_n}].
$$

From the above computation we know $\mathbb{E}[B^2_{T_n}]$ is the area between U_{δ_0} and U_n, and it is clear that this area is bounded by $\mathbb{E}[B^2_T]$. Hence,

$$
\mathbb{E}[T] = \lim_n \mathbb{E}[T_n] \leq \int x^2 \ \mu(dx).
$$

Due to Proposition 2.18 there also equality in the limit. □

4. Order of Random Variables

4.1. **Stochastic Order.** There are different approaches to the ordering of random variables. The standard stochastic order compares the distribution functions of them. Let X, Y be two random variables and $F_X(a) = P(X \leq a), F_Y = P(Y \leq a)$. The assertion

$$F_X(a) \geq F_Y(a), \text{ for all } a \in \mathbb{R} \tag{49}$$

means that Y is less likely than X to be smaller than a given number a. If this is true for all $a \in \mathbb{R}$, X is said to be smaller in stochastic order than Y.

For the stochastic ordering we have the following lemma:

Lemma 4.1. *Let X and Y be random variables, U the class of all non-decreasing functions on $\mathbb{R}$. X is smaller than Y in the stochastic order, if and only if,*

$$\mathbb{E}[u(X)] \leq \mathbb{E}[u(Y)], \text{ for all } u \in U. \tag{50}$$

Proof. Let X be smaller than Y in stochastic order. As non-decreasing functions are measurable, we can approximate any u with $u^n(x) = \sum_{i=0}^n u(t_i^n) * \mathbb{1}_{(t_i^n, t_{i+1}^n]}(x)$ with choosing $(t_i^n)_{i=0}^{K_n}$ as a refining sequence, such as $t_i^n = u^{-1}((i - n2^n)2^{-n})$ $[i = 0, .., 2n2^n]$. We show that (50) is true for such u^n by induction over i. The result then follows my monotone convergence.

$u(t_0)P(X \leq t_1) + u(t_1)P(X > t_1) \leq u(t_0)P(y \leq t_1) + u(t_1)P(Y > t_1)$ because u non-decreasing, and $P(X \leq t_1) \geq P(Y \leq t_1)$. For the induction step $N \to N+1$ observe the following:

$$\begin{aligned}
&\Theta_Y^N := u(t_N)P(Y > t_N) + u(t_{N-1})P(t_{N-1} < Y \leq t_N) - u(t_{N-1})P(Y > t_{N-1}) \\
&= (u(t_N) - u(t_{N-1}))\ \ P(Y > t_N) \\
&\geq (u(t_N) - u(t_{N-1}))\ \ P(X > t_N) \\
&= u(t_N)P(X > t_N) + u(t_{N-1})P(t_{N-1} < X \leq t_N) - u(t_{N-1})P(X > t_{N-1}) \\
&=: \Theta_X^N
\end{aligned}$$

Let Φ_X^N denote the sum corresponding to X until the n$^{\text{th}}$ summand,

$$\Phi_X^N = \sum_{i=0}^{N-1} u(t_i)P(X_i < X \leq t_{i+1}) + u(t_N)\ P(t_N < X).$$

Let Φ_Y^N be the same for Y. Then

$$\Phi_Y^N = \Phi_Y^{N-1} + \Theta_Y^N \geq \Phi_X^{N-1} + \Theta_X^N = \Phi_X^N.$$

This proves one way. The other way follows with choosing u as $u(x) = \mathbb{1}_{(-\infty,a]}(x)$. □

4.2. **Convex Order.** We have seen that the stochastic order can also be defined through comparing the expected values of images of X and Y under the class of non-decreasing functions. Similarly one can introduce the concept of a different ordering of random variables, which will prove more useful to us.

Definition 4.2. *Let X and Y be random variables and Φ be the class of all convex functions $\Phi = \{f : \mathbb{R} \to \mathbb{R} | f$ is convex$\}$. We say that X is smaller than Y in the convex order if*

$$\mathbb{E}[\psi(X)] \leq \mathbb{E}[\psi(Y)] \text{ for all } \phi \in \Phi. \tag{51}$$

If so, we write $X \prec_{cx} Y$.

Another way to state the Brascamp-Lieb moment inequality (2) (in one dimension) is:

Theorem 4.3. *If X is log-concave to Y, then $X - \mathbb{E}X \prec_{cx} Y - \mathbb{E}Y$.*

Following the book of Shaked and Shantikumar on Stochastic Orders [9] we examine the convex order. Let $X \prec_{cx} Y$, first note that $x \mapsto x$ and $x \mapsto -x$ are convex. Therefore $\mathbb{E}X \leq \mathbb{E}Y \leq \mathbb{E}X$, and $\mathbb{E}X = \mathbb{E}Y$. Furthermore, $x \mapsto x^2$ is convex, and $\operatorname{Var} X \leq \operatorname{Var} Y$.

It is obvious that the function $\phi_a(x) = (x-a)^+ = \max(x-a, 0)$ is convex. Therefore, $X \prec_{cx} Y$ implies that $\mathbb{E}(X-a)^+ \leq \mathbb{E}(Y-a)^+$. It is also known that any convex function can be approximated with an increasing sequence of functions $K_n(x) = \lambda_0 x + \sum_{i=1}^{L_n} \lambda_i \phi_{a_i}(x)$, meaning $\phi(x) = \lim_n K_n(x)\ \forall x$. This and monotone convergence theorem prove the other direction of Lemma 4.4.

Lemma 4.4. *Let X and Y be such that $\mathbb{E}X = \mathbb{E}Y$.*

$$X \prec_{cx} Y \text{ if, and only if } \mathbb{E}(X-a)^+ \leq \mathbb{E}(Y-a)^+ \ \forall a \in \mathbb{R}. \tag{52}$$

Let us recall how handy the decomposition in positive and negative part of an expression can be. Observe,

$$|X-a| = (X-a)^+ + (X-a)^- \qquad \text{and } (X-a) = (X-a)^+ - (X-a)^-,$$

which yields

$$\mathbb{E}|X-a| = \mathbb{E}(X+a)^- + \mathbb{E}(X-a)^+ = 2\mathbb{E}(X-a)^+ + \mathbb{E}X - a,$$

and therefore

$$\mathbb{E}(X-a)^+ \leq \mathbb{E}(Y-a)^+, \text{ if and only if } \mathbb{E}|X-a| \leq \mathbb{E}|Y-a|.$$

Furthermore, a direct computation shows

$$\begin{aligned} \mathbb{E}[|X-a|] &= a + 2\int_a^\infty (1 - F(u)\ du \\ &= -a + 2\int_{-\infty}^a F(u)\ du. \end{aligned}$$

Recalling the definition of the potential of a random variable, we achieved the following theorem:

Theorem 4.5. *Let X and Y be such that $\mathbb{E}X = \mathbb{E}Y$, F_X and F_Y the distribution functions of X and Y, respectively.Also let $U_X(x)$ denote the potential of the random variable X as defined in* (38). *Then,*

$$X \prec_{cx} Y, \text{ if and only if } U_X(x) \geq U_Y(x), \text{for all} x \in \mathbb{R} \tag{53}$$

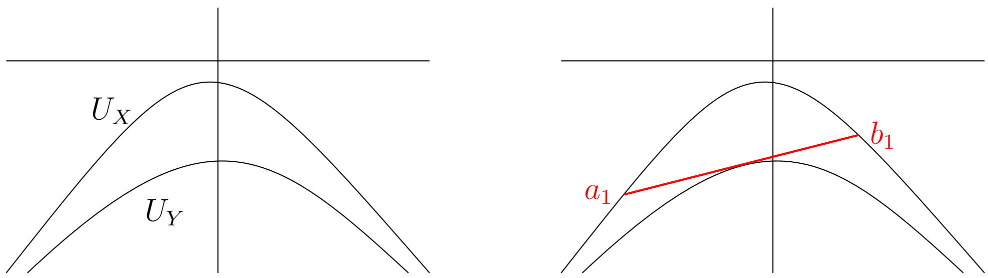
$$\text{if and only if } \int^x F_X(u)\, du \leq \int^x F_Y(u)\, du, \text{ for all} x \in \mathbb{R}. \tag{54}$$

Ultimately, we exploit a connection between Brownian motion and convex ordering that is our main result of this section.

Theorem 4.6. *Let X and Y be centred random variables. $X \prec_{cx} Y$, if and only if there exist two stopping times T_1, T_2 such that $T_1 \leq T_2$ a.s. and $X \sim B_{T_1}, Y \sim B_{T_2}$.*

Proof. Suppose $X \sim B_{T_1}, Y \sim B_{T_2}$ and $T_1 \leq T_2$. Observe that $|B_t^{T_2} - x|$ is a submartingale with final element $|B_{T_2}|$, because $T_2 < \infty$ a.s. Optional Stopping Theorem 2.16 applied to the submartingale $(B_t^{T_2})_t$ yields $\mathbb{E}\left[|B_{T_2}^{T_2} - x|\big|F_{T_1}\right] \geq |B_{T_1} - x|$. Applying expectation on both sides yields that $U_X \geq U_Y$, that is $X \prec_{cx} Y$.

On the other hand, assume that $U_X \geq U_Y$. As X is a centred random variable, due to the Skorokhod embedding by Chacon Walsh there is a stopping time T_1 such that $B_{T_1} \sim X$.



(A) potentials of X and Y with $U_X \geq U_Y$ (B) choosing the first line

FIGURE 2

Because $U_X \geq U_Y$, we can employ the same arguments as in the proof of the embedding of Chacon-Walsh to find a stopping time S such that $B_{T_1+S} = B_{T_2} \sim Y$. □

4.3. **A first proof of the Brascamp-Lieb moment inequality for special V.** It is also the book of Shaked and Shantikumar [9] where the content of this section was found. Here we give a first proof of the Brascamp-Lieb moment inequality, based on stochastic ordering, that is however only valid for convex functions V in (1) with $\lim_{x\to\infty} V(x) = \infty$ and $\lim_{x\to-\infty} V(x) = \infty$.

We first state sufficient conditions for the convex order. Let $S^-(h)$ denote the number of sign changes of the function $h : \mathbb{R} \to \mathbb{R}$,

$$S^-(h) := \sup\{n \in \mathbb{N} : h(x_1), h(x_2), ..., h(x_n), x_i < x_{i+1}, h(x_i)h(x_{i+1}) < 0\} - 1.$$

Lemma 4.7 (cf. [9], Theorem 3.A.44). *Let X and Y be one dimensional random variables on $\mathbb{R}$ with equal means, with densities f and g, distribution functions F and G, respectively. Then X is smaller than Y in convex order if any of the two conditions below holds:*

i. $S^-(g - f) = 2$, *with sign sequence*$+, -, +$,
ii. $S^-(G - F) = 1$, *with sign sequence*$+, -$.

Proof. Assume (i) $S^-(g - f) = 2$ with sign sequence $+, -, +$. Let $I_1 = (-\infty, a]$ and $I_3 = (b, \infty)$ be the intervals where $g - f \geq 0$ and $I_2 = (a, b]$ be the interval where $g - f \leq 0$. Then $(G - F)(x)$ is increasing on I_1 and I_3, decreasing on I_2. Furthermore, note that

$$\lim_{x\to\infty} F(x) = \lim_{x\to\infty} G(x) = 1,$$
$$\lim_{x\to-\infty F(x)} = \lim_{x\to-\infty} G(x) = 0.$$

Hence,

$$\lim_{x\to\pm\infty} (F - G)(x) = 0.$$

Because X and Y have the same mean, $F > G$ everywhere is impossible. Hence, (ii) follows. If (ii) holds, then there are intervals $J_1 = (-\infty, c]$ and $J_2 = (c, \infty)$, such that $G - F \geq 0$ on J_1, and $G - F \leq 0$ on J_2. Then

$$\int^x (G - F)(u)\ du \geq 0. \tag{55}$$

Indeed, $x \leq c$, (55) holds because of monotonicity of the intergral if. If on the other hand, $x \geq c$, we know that $\mathbb{E}[X] = \mathbb{E}[Y]$ implies

$$\int_{-\infty}^{\infty} G - F\ du = 0.$$

Hence,

$$\int_{-\infty}^{x} (G - F)(u)\ du = -\int_{x}^{\infty} (G - F)(u)\ du \geq 0,$$

because of monotonicity of the integral and $G - F \leq 0$ on J_2. □

The following theorem is a direct consequence of V convex in (1).

Theorem 4.8 (cf. [9], Theorem 3.A.54)**.** *Let X, Y, f, g, be as in Lemma 4.7 and $S^-(g-f) = 2$. If X is log-concave with respect to Y, the corresponding sequence to $S(g-f)$ is $+, -, +$ and*

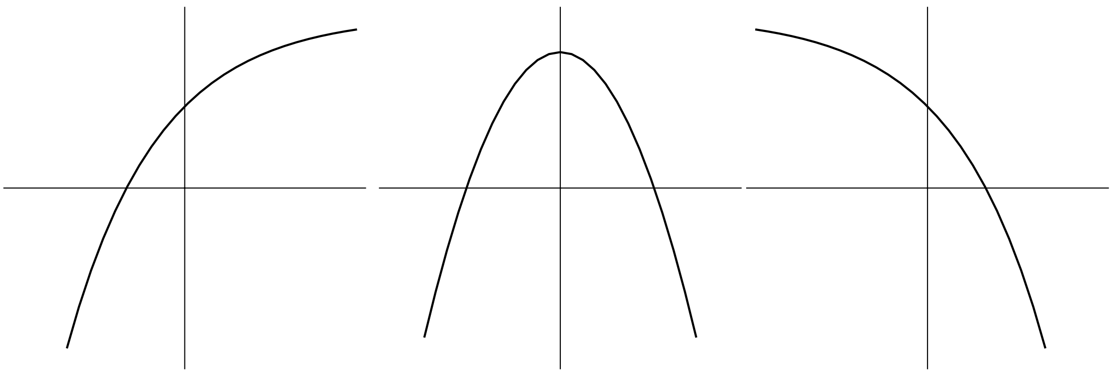
$$X \prec_{cx} Y.$$

And this readily imp lies the following.

Corollary 4.9 (Brascamp-Lieb moment inequality in dimension one for V unbounded at $\pm\infty$)**.** *If X, Y are one-dimensional random variables, X log-concave with respect to Y and V goes to infinity for $x \to -\infty$ and $x \to \infty$, then $S^-(g-f) = 2$ and particularly Inequation* (2) *holds in the one-dimensional sense, that is, for all $\psi : \mathbb{R} \to \mathbb{R}$ convex,*

$$\mathbb{E}[\psi(X - \mathbb{E}[X])] \leq \mathbb{E}[\psi(Y - \mathbb{E}[Y])].$$

We note three things in the above corollary. Firstly, the random variable Y does not need to be Gaussian. It can be anything, as long as X is log-concave with respect to it. Secondly, if Y however is Gaussian, the inequation can be extented to random variables in $\mathbb{R}^n$ with a later Theorem (5.4). Thirdly, this is a special case of general convex V. Indeed, V can be either non-decreasing, non-increasing, or unbounded at $\pm\infty$.



(A) V non-decreasing (B) V unbounded at $\pm\infty$ (C) V non-increasing

FIGURE 3

5. The Brascamp-Lieb moment inequality and proof

The Brascamp-Lieb moment inequality relates the expectation of two random variables under a convex function. The two random variables are related in the following sense.

Definition 5.1. *Let X and Y be two random variables on a probability space $(\Omega, \mathcal{A}, P)$, μ be the distribution of X, ν the distribution of Y. We say X is log-concave with respect to Y if the Radon-Nikodým derivative exists and it is given by*

$$\mu(dx) = \frac{1}{Z} e^{-V(x)} \nu(dx), \tag{56}$$

where $Z = \int_\Omega e^{-V(x)}\ \nu(dx)$ is a normalization constant and $V : \mathbb{R}^d \to \mathbb{R}$ convex.

X is called log-concave to Y, because the logarithm of the density of X with respect to Y is concave.

The density of X with respect to Y is invariant under adding the same constant to X and Y. Hence, X log-concave to Y is equivalent to $X - \mathbb{E}Y$ log-concave to $Y - \mathbb{E}Y$. Furthermore, for $\alpha \in \mathbb{R}$ let μ_α be the distribution of $X - \alpha$ and let Y be centred. Then $X - \alpha$ with distribution μ_α is log-concave with respect to Y with distribution ν. For the one-dimensional case,

$$\begin{aligned}
\mu_\alpha(dx) &= \mu(d(x+\alpha)) \\
&= e^{-V(x+\alpha)}\ \nu(d(x+\alpha)) \\
&= e^{-V(x+\alpha)} (2\pi\sigma)^{\frac{-1}{2}} e^{\frac{-(x+\alpha)^2}{2}} d(x+\alpha) \\
&= e^{-V(x+\alpha) - \alpha x - \frac{\alpha^2}{2}} \nu(dx) = e^{-V'(x)} \nu(dx).
\end{aligned}$$

This also gives $X + \beta$ is log-concave to $Y + \gamma$, because if $\alpha = \beta - \gamma$ the assertion follows from adding γ to $X + \alpha$ log-concave with resepect to Y. We have shown the following:

Lemma 5.2. *X is log-concave to Y, if and only if $X - \beta$ is log-concave to $Y - \gamma$ for any choice of α and β. This particularly implies that the centred random variables are also log-concave if one the above holds.*

Let us recall the Brascamp-Lieb moment inequality:

Theorem 5.3. *Let X and Y be random variables in $\mathbb{R}^n$ such that X is log-concave to Y and Y is Gaussian. Then the following inequality holds*

$$\mathbb{E}[\psi(\phi \cdot Y - \mathbb{E}[\phi \cdot Y])] \geq \mathbb{E}[\psi(\phi \cdot X - \mathbb{E}[\phi \cdot X])] \tag{57}$$

for all $\psi : \mathbb{R} \to \mathbb{R}$ convex and every $\phi \in \mathbb{R}$.

5.1. **Proof of the Brascamp-Lieb moment inequality.** Using Theorem 5.4 to justify proving the theorem only in one dimension and Caffarelli's Theorem 5.6 to bound the derivative of the transport function, the proof of the Brascamp-Lieb moment inequality is readily obtained.

Theorem 5.4 (Reduction to dimension one)**.** *If X and Y are random variables in $\mathbb{R}^n$ and X is log-concave with respect to Y and Y is Gaussian, then the one-dimensional marginals maintain the log-concave property and the marginal of Y is still Gaussian. This is, for every $\phi \in \mathbb{R}$*

- *$\phi \cdot X$ is log-concave with resepect to $\phi \cdot Y$, and*
- *$\phi \cdot Y$ is Gaussian.*

This particularly implies, that it is enough to show Theorem 5.3 for random variables X and Y in $\mathbb{R}$.

We utilize the following lemma.

Lemma 5.5 (from [1], Corollary 3.5)**.** *Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}, (x, y) \mapsto F(x, y)$ be log-concave. Then the function $G : \mathbb{R}^n \to \mathbb{R}$ defined as*

$$G(x) := \int_{\mathbb{R}^m} F(x, y)\ dy, \tag{58}$$

is log-concave.

Proof of Theorem 5.4. Let $\hat{\mu}$ be the distribution of $\phi \cdot X$. Let $K = K(t)$ be the pre-image of t under the linear transformation $\hat{\phi} : \mathbb{R}^n\ to \mathbb{R}, z \mapsto \phi \cdot z$. Then

$$\begin{aligned} \hat{\mu}(dt) &= \int_K 1\ d\mu \\ &= \int_K \frac{1}{Z} e^{-V} d\nu. \end{aligned}$$

$\mathbb{R}^n$ can be decomposed in the span of ϕ and kernel(ϕ) and $K = \text{kernel}(\phi) + \frac{t}{\|\phi\|_2}$. Choose an orthonormal basis $(e_1, e_2, \ldots, e_n)$ such that $e_1 \in \text{span}(\phi)$ and $(e_2, \ldots, e_n) \in \text{kernel}(\phi)$. Without loss of generality we can assume $\|\phi\|_2 = 1$. Since Y is Gaussian, for $(t, y) \in \text{span}(\phi) \times \text{kernel}(\phi)$ the density of ν is $\nu(d(t, y)) = \text{const} \exp\{-\Psi(t, y)\}$, with $\Psi(t, y) = \langle (t, y), Q(t, y) \rangle, Q = \begin{pmatrix} A & B \\ B' & C \end{pmatrix}$, and $\langle\ ,\ \rangle$ the standard inner product. Then, the above becomes

$$\hat{\mu}(dt) = \int_{\text{kernel}(\phi)} \frac{\text{const}}{Z} e^{-V(t,y)} \exp^{-\Psi(t,y)}\ dy.$$

Let $D = A - BC^{-1}B'$ and $\bar{y} = y + C^{-1}B't$. Then the variables in $\Psi(t, y)$ can be seperated. Indeed,

$$\begin{aligned}
&\langle t, Dt\rangle + \langle \bar{y}, C\bar{y}\rangle \\
&= \langle t, At\rangle + \langle t, BC^{-1}B't\rangle + \\
&\qquad \langle y, Cy\rangle + \langle y, B't\rangle + \langle B't, y\rangle + \langle BC^{-1}B't, t\rangle \\
&= \langle t, At\rangle + \langle y, Cy\rangle + \langle y, B't\rangle + \langle t, By\rangle \\
&= \Psi(t, y).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\hat{\mu}(dt) &= \exp\{-\langle t, Dt\rangle\} \int \frac{\text{const}}{Z} \exp\{-\langle \bar{y}, C\bar{y}\rangle\} \exp\{-V(t, \bar{y})\}\ d\bar{y} \\
&= \exp\{-\frac{t^2}{2d}\} \int \frac{\text{const}}{Z} \exp\{-\langle y, Cy\rangle\} \exp\{-V(t, y)\}\ dy,
\end{aligned}$$

for $d = \frac{1}{2D}$. We know that C is positive definite, because Y is Gaussian. Hence, the integrand is log-concave and Lemma 5.5 applies and

$$\hat{\mu}(dt) = \text{ log-concave function} \cdot \exp\{-\langle t, Dt\rangle\}.$$

Hence, $\phi \cdot X$ is log-concave with respect to a Gaussian random variable with mean zero and variance d^2.

The above computation for $V \equiv 0$ and $Z = 1$ shows that $\phi \cdot Y$ is Gaussian with mean zero and variance d^2, hence, $\phi \cdot X$ is log-concave with respect to $\phi \cdot Y$.

It is clear that $\phi \cdot Y$ is Gaussian, because Y is Gaussian. □

The essential ingredient for the proof of the Brascamp-Lieb moment inequality now is the following theorem. It bounds the derivative of the transport function from $\mathcal{N}(0, 1)$ to μ to the variance of Y. The impact of Y is hidden in the fact that μ is log-concave with respect to Y.

Theorem 5.6 (Caffarelli)**.** *Let X and Y be one-dimensional random variable with distributions μ and ν respectively and Y Gaussian. The transport function g defined as*

$$g(x) = (F_\mu^{-1} \circ \Phi)(x), \tag{59}$$

with F_μ the distribution function of X, F_μ^{-1} its true inverse, and Φ the cumulative distribution function of a standard-normal Gaussian random variable. Let $a = Var(Y)$. Then the following holds true:

$$g'(x) \leq \sqrt{a}. \tag{60}$$

The actual proof of the Brascamp-Lieb moment inequality is now very short.

Proof of Brascamp-Lieb moment inequality. Theorem 5.4 allows us to reduce the proof to one dimension.

Let Y be a one-dimensional normal random variable, and X be a one-dimensional random variable that is log-concave with respect to Y. We know that X is in $\mathcal{L}^2$. We can use Bass' embedding for X, or its law μ.

In Bass' embedding we had for a Brownian motion $(B_t)_{t\geq 0}$ and the function g as defined in Equation (59),

$$g(B_1) - \mathbb{E}[g(B_1)] = \int_0^1 \mathbb{E}[g'(B_1)|\mathcal{F}_s] \, dB_s = \int_0^1 a(s, B_s) \, dB_s.$$

It also gives that $M_t = \mathbb{E}[g(B_1)|\mathcal{F}_t] = \int_0^t a(s, B_s) \, dB_s$ is a martingale. From here we constructed a stopping time

$$T =< M >_1 = \int_0^1 (\mathbb{E}[g'(B_1)|\mathcal{F}_s])^2 \, ds,$$

such that $X \sim B_T$. According to Theorem 5.6, $g' \leq \sqrt{a}$ and hence

$$T \leq \int_0^1 \sqrt{a}^2 \, ds \leq a.$$

Since $a = \mathrm{Var}(Y)$ and Y is a normal random variable, a embeds some $\mathcal{N}(0, a)$ random variable $\bar{Y} \sim Y$. Theorem 4.6 gives

$$X_T \prec_{cx} \bar{Y} \Leftrightarrow X_T \prec_{cx} Y.$$

$X_T \prec_{cx} Y$ is the Brascamp-Lieb-inequality. □

5.2. **Proof of Caffarelli's theorem.** We have defined the transport function from Φ to X to be $g(x) = (F_\mu^{-1} \circ \Phi)(x)$. The assertion of Lemma 5.6 is the special case of the n-dimensional Theorem 11 in [4] for one dimension.

A classical optimal transport problem is the following.

Example 5.7 (A glimpse at Optimal Transport). *Let μ and ν be probabilites on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Let us call the function $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ the cost function. Find a function T such that ν is the push-forward measure of μ, $\nu = T_{\#}\mu$ and minimizes the following expression*

$$\inf_{\nu = S_{\#}\mu} \int c(t, S(t)) \, dt.$$

This is, the transport function T is the cheapest way to transport μ to ν.

It seems natural to recquire the costs $c(x, y)$ to be dependent on the distance between x and y. For example, $c(x, y) = |x - y|^p$. If $p > 1$ the cost function is convex and Caffarelli has contributed greatly (see [4] or [7]) to the development of a regularity theory for convex cost functions in $\mathbb{R}^n$. For very general contextes (cf.

as stated in Theorems 1 and 2 in [4]), it has been shown that the optimal transport function is the gradient of a convex potential φ. Particularly, this asserts that the derivative of the transport function, the second derivative of the convex potential, is positive. For the specific case of measures with densities $f(x) = \exp\{-Q(x)\}$ and $g(x) = \exp\{-(Q(x) + F(x))\}$, with Q a nonnegative quadratic form and F convex, Caffarelli could show more. The density f corresponds (up to a constant) to the density of a Gaussian random variable Y and g is a density of a random variable that is log-concave with respect to Y. He showed the multidimensional version of Theorem 5.6, this is

$$0 \leq \varphi_{x_i x_i} \leq 1.$$

Let us leave this very small excursion to optimal transport, and return to the proof of Caffarelli's Theorem in dimension one.

The careful reader might have questioned the form of g as a transport from a standard-normal random variable to the random variable X, that is log-concave to a normal random variable Y, that is not necessarily standard-normal. The deeper reason for this is that in the above proof of the Brascamp-Lieb moment inequality we needed it in the form of (59) and that expectation and variance of Y do not matter as in the following sense:

Lemma 5.8. *Let X, Y, a, and g be as above. Define $\bar{Y}$ the centred version of Y, that is $\bar{Y} = Y - \mathbb{E}[X]$ and $\hat{Y}$ the standardized version of Y, that is $\hat{Y} = \bar{Y}/\sqrt{a}$. Also, define*

$$\begin{aligned}
\tilde{g}(x) &= (F_\mu^{-1} \circ F_Y)(x) \\
\bar{g}(x) &= (F_\mu^{-1} \circ F_{\bar{Y}})(x) \\
\hat{g}(x) &= (F_\mu^{-1} \circ F_{\hat{Y}})(x),
\end{aligned}$$

the functions that transport Y, $\bar{Y}$, and $\hat{Y}$ to X. Observe $\Phi = F_{\hat{Y}}$ and $g(x) = \hat{g}(x) = (F_\mu^{-1} \circ F_{\hat{Y}})(x)$. Then

$$g' \leq \sqrt{a} \text{ if, and only if } \tilde{g}' \leq 1.$$

Heuristically, it is clear that the two distributions that play a role in the inequality should also be the crucial ingredients for the Theorem 5.6 that now essentially proves the Brascamp-Lieb moment inequality. In inequality (60) this is not apparent in a clear way. Lemma 5.8 gives a clearer view on these matters. It says that the slope of the transport from $\mathcal{N}(0, 1)$ to μ is less than $\sqrt{\mathrm{Var}(Y)}$, if and only if the slope of the transport from $\mathcal{N}(0, a)$ to μ is less than one. We prefer the second interpretation since μ and $\mathcal{N}(0, 1)$ are the distributions the random variables in the Brascamp-Lieb moment inequality.

Remark 5.9 (A more natural proof of the Brascamp-Lieb moment inequality). *In Bass' embedding, μ was centred, and we could have also used the function*

$$\tilde{g} = \bar{g} := F_\mu \circ \Phi_a,$$

where Φ_a is the cumulative distribution function of a centred Gaussian random variable with variance a. Then $g(B_a) \sim \mu$ and similarly as in Theorem 3.4 and Equation (45) *and* (46),

$$\tilde{M}_t = \mathbb{E}[\tilde{g}(B_a)|\mathcal{F}_t] = \int_0^t a(s, B_s)\ dB_s,$$

and

$$\tilde{g}(B_a) = \int_0^a \mathbb{E}[\tilde{g}'(B_a)|\mathcal{F}_t]\ dB_t.$$

Then for the stopping time

$$T := \int_0^a \mathbb{E}[\tilde{g}'(B_a)|\mathcal{F}_t]^2\ dt,$$

the assertion from Caffarelli's Theorem 5.6 and Lemma 5.8

$$\tilde{g} \leq 1,$$

and

$$T \leq a,$$

follows. Hence, we can embed μ with a stopping time T that is strictly smaller than the stopping time a, and Theorem 4.6 yields the Brascamp-Lieb moment inequality.

Consider the following picture. It shows the cumulative distribution function of a Gaussian random variable on the left, and a random variable Y that is log-concave with respect to the Gaussian on the right.

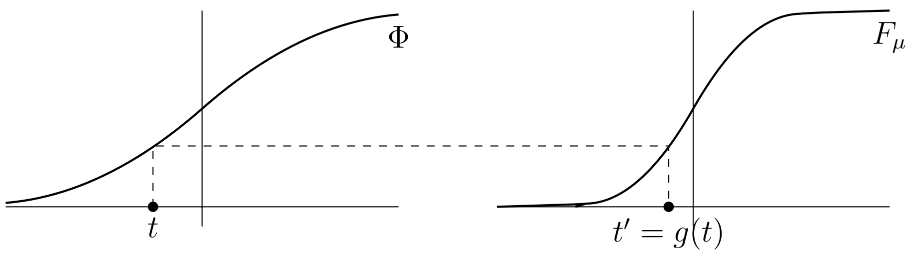


FIGURE 4

The dashed path from t to $g(t)$ indicates the action of g. The line segment from t to $\Phi(t)$ can be interpreted as the question, how much mass of $\mathcal{N}(0,1)$ is located on the left of t? The line segment from $\Phi(t)$ over $F_\mu(g(t))$ to $g(t)$ can be interpreted as the question, where is the point t' so that as much μ-mass is on its left as there is $\mathcal{N}(0,1)$-mass on the left of t? This point is $g(t)$. The assertion

$$g' \leq 1$$

now states that when increasing/decreasing t by δ, $g(t)$ will increase/decrease by at most δ.

Due to Lemma 5.2 we can add a constant to X so that the median is at zero. Then $F_\mu(0) = \Phi(0) = \frac{1}{2}$. Observe that $F_\mu \circ g(t) = \Phi(t)$. Therefore, if we compare $\Phi(g(t))$ to $F_\mu(g(t))$ we see that $F_\mu(g(t)) \geq \Phi(g(t))$ for $g(t) \geq 0$, because Φ is monotone. Similarly, the other inequality. Hence,

$$\text{for } t \geq 0,\ F_\mu(t) \geq \Phi(t),$$
$$\text{and for } t \leq 0,\ F_\mu(t) \leq \Phi(t).$$

Observe, that this is also true for our examples above in Figure 4 as indicated in Figure 5.

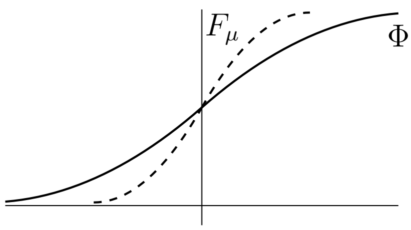


FIGURE 5

The above construction works not only for the $\frac{1}{2}$-quantil. The following corollary is valid by shifting X and Y by their α-quantils (using Lemma **??**) :so that α-quantil is at zero. With this we have proven the following theorem.

Theorem 5.10 (A property of X log-concave with respect to a Gaussian Y)**.** *Let $\alpha \in (0,1)$, X and Y be random variables in $\mathbb{R}$ with distributions μ and ν respectively, and q_μ and q_ν the α-quantiles. If X is log-concave with respect to Y, and Y is Gaussian, then if $\hat{F}_\mu$ and $\hat{F}_\nu$ denote the cumulative distribution functions of $X - q_\mu$ and $Y - q_\nu$ respectively,*

$$\hat{F}_\mu \geq \hat{F}_\nu, \text{ on } \{t \geq 0\}, \text{ and} \tag{61}$$

$$\hat{F}_\mu \leq \hat{F}_\nu, \text{ on } \{t < 0\}. \tag{62}$$

This means that $X - q_\mu$ is concentrated more around zero than $Y - q_\nu$. Therefore, for $\delta_-, \delta_+ > 0$,

$$P(q_x - \delta_- \leq X \leq q_x + \delta_+) \geq P(q_y - \delta_- \leq Y \leq q_y + \delta_+).$$

Remark 5.11. *A corresponding version of the above theorem also holds in n dimensions with using Theorem 2 and Theorem 11 in* [4].

Proof of Lemma 5.8, cf. [3]. First, without loss of generality we can assume $Y = \bar{Y}$, because $\bar{g}(x - \mathbb{E}[Y]) = (F_\mu^{-1} \circ F_{\bar{Y}})(x - \mathbb{E}[Y]) = (F_\mu^{-1} \circ F_Y)(x) = \tilde{g}(x)$.

Recall $g(x) = (F_\mu^{-1} \circ F_{\hat{Y}})(x)$ and therefore $g'(x) = (F_\mu^{-1})'(F_{\hat{Y}}(x)) \cdot (F_{\hat{Y}})'(x)$. Similarly, $\bar{g}'(x) = (F_\mu^{-1})'(F_{\bar{Y}}(x)) \cdot (F_{\bar{Y}})'(x)$. We want to compare g' at $x/\sqrt{a}$ to $\bar{g}'$ at x. To this end, observe that $(F'_{\hat{Y}})(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}}$ and $(F'_{\bar{Y}})(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{a}} e^{\frac{x^2}{2a}}$, and therefore

$$
\begin{aligned}
(F'_{\hat{Y}})(x/\sqrt{a}) &= \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2a}} \\
&= \sqrt{a}(F'_{\bar{Y}})(x).
\end{aligned}
$$

Furthermore, via a change of coordinates $y = \sqrt{a}\ t$,

$$
\begin{aligned}
F_{\bar{Y}}(x) &= \int^x \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{a}} e^{\frac{y^2}{2a}}\ dy \\
&= \int^{\frac{x}{\sqrt{a}}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{a}} e^{\frac{t^2}{2}}\ \sqrt{a}\ dt \\
&= \int^{\frac{x}{\sqrt{a}}} \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}}\ dt \\
&= F_{\hat{Y}}(\frac{x}{\sqrt{a}}).
\end{aligned}
$$

Using the above

$$
\begin{aligned}
g'(\frac{x}{\sqrt{a}}) &= (F_\mu^{-1})'(F_{\hat{Y}}(\frac{x}{\sqrt{a}})) \cdot (F_{\hat{Y}})'(\frac{x}{\sqrt{a}}) \\
&= (F_\mu^{-1})'(F_{\bar{Y}}(x)) \cdot (F'_{\bar{Y}})(x) \cdot \sqrt{a} \\
&= \sqrt{a} \cdot \bar{g}'(x).
\end{aligned}
$$

Particularly, $g'(x) \leq \sqrt{a}$ if, and only if $\bar{g}'(x) = \tilde{g}'(x) \leq 1$. □

In order to prove Theorem 5.6 we employ yet another lemma.

Lemma 5.12 (cf. [3])**.** *For every $x \in \mathbb{R}$ the following assertion holds,*

$$F'_\mu \geq \Phi'(x + V'_-(x)).$$

Proof. μ is log-concave to ν, therefore from (1)

$$
\begin{aligned}
F'_\mu(x) &= \frac{1}{Z} e^{-V(x)}\ \nu(dx) \\
&= \frac{1}{Z} \frac{1}{\sqrt{2\pi}} e^{-V(x)} e^{\frac{x^2}{2}}\ dx \\
&= \frac{1}{\int e^{-\frac{y^2}{2} - V(y)}\ dy} \cdot e^{-V(x) - \frac{x^2}{2}}.
\end{aligned}
$$

Taking the reciprocal, and using $V(y) - V(x) \geq V'_-(x)(y-x)$

$$\begin{aligned}
\frac{1}{F'_\mu(x)} &= \int e^{-\frac{y^2}{2} - V(y)}\, dy \cdot e^{-V(x) - \frac{x^2}{2}} \\
&= e^{\frac{x^2}{2}} \int e^{-\frac{y^2}{2} + V(x) - V(y)}\, dy \\
&\leq e^{\frac{x^2}{2}} \int e^{-\frac{y^2}{2} - V'_-(x)(y-x)}\, dy \\
&= e^{\frac{x^2}{2} + V'_-(x)(x)} \cdot \int e^{-\frac{y^2}{2} - V'_-(x)y - V'_-(x)^2 + V'_-(x)^2}\, dy \\
&= e^{\frac{x^2}{2} + V'_-(x)(x) + V'_-(x)^2} \cdot \int e^{-\frac{y^2}{2} - V'_-(x)y - V'_-(x)^2}\, dy \\
&= e^{\frac{1}{2}(x + V'_-(x))^2} \cdot \int e^{-\frac{1}{2}(y + V'_-(x))^2}\, dy \\
&= e^{\frac{1}{2}(x + V'_-(x))^2} \cdot \sqrt{2\pi} = \frac{1}{\Phi'(x + V'_-(x))}.
\end{aligned}$$

□

We now are ready to prove Theorem 5.6.

Proof of Theorem 5.6, cf. [3]. Using the chain rule and derivative of inverse functions $g'(x)$ can be written as

$$g'(x) = \frac{\Phi'(x)}{F'_\mu \circ F_\mu^{-1}(\Phi(x))}.$$

Due to Lemma 5.8 it is enough to show $g'(x) \leq 1$, and this is equivalent to

$$G(\xi) := F'_\mu \circ F_\mu^{-1}(\xi) - \Phi' \circ \Phi^{-1}(\xi) \geq 0, \text{ for all } \xi \in (0,1). \tag{63}$$

We will show that G is never negative by showing that the limits at $\xi = 0$ and $\xi = 1$ are zero and that every local minimum is greater or equal zero. First, we see that Φ^{-1} goes to minus infinity when ξ goes to zero and that $\Phi'(x)$ goes to zero when x goes to minus infinity. But, since $-V(x)$ is concave we have for some a and b in $\mathbb{R}$ $e^{-V(x)} \leq e^{ax+b}$, and then also $F'_\mu(x) = \text{const} \cdot e^{-V(x)} e^{-(\frac{F^{-1}(x)}{2})^2} \leq \text{const} \cdot e^{ax+b-x^2/2}$ goes to zero as x goes to minus infinty. As ξ goes to zero, $F_\mu^{-1}(\xi)$ goes to minus infinity, and $F'_\mu \circ F_\mu^{-1}$ goes to zero as ξ goes to zero. Therefore, $\lim_{\xi \to o} G(\xi) = 0$. Similarly, $\lim_{\xi \to 1} G(\xi) = 0$.

G is both left- and rightdifferentiable because F_μ^{-1} is and because F_μ^{-1} is monotone. Observe, $\Phi'' = (\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}})' = \frac{1}{2\pi} e^{-\frac{x^2}{2}} \cdot \frac{-2x}{2} = -x \cdot \Phi'$. For $F'_\mu(x) = \frac{1}{Z\sqrt{2\pi}} e^{-V(x) - x^2/2}$, $(F'_\mu(x))'_\pm = \frac{1}{Z\sqrt{2\pi}} e^{-V(x)-x^2/2} \cdot -(x + V'_\pm(x)) = F'_\mu(x) \cdot -(x + V'_\pm(x))$.

Thefore we compute,

$$\begin{aligned} G'_\pm(\xi) &= (F'_\mu)'_\pm \circ F_\mu^{-1}(\xi) \cdot (F_\mu^{-1})'_\pm(\xi) - (\Phi')'_\pm \circ \Phi^{-1}(\xi) \cdot (\Phi^{-1})'(\xi) \\ &= \frac{(F'_\mu)'_\pm}{F'_\mu} \circ F_\mu^{-1}(\xi) + \frac{(\Phi')'_\pm}{\Phi'} \circ \Phi^{-1}(\xi) \\ &= -(x + V'_\pm(x))|_{x=F_\mu^{-1}(\xi)} + \Phi^{-1}(\xi). \end{aligned}$$

This and the fact that for every local minimum at ξ_0 we must have $G'_-(\xi_0) \leq 0$ and $G'_+(\xi_0) \geq 0$ shows

$$(x + V'_+(x))|_{x=F_\mu^{-1}(\xi)} \leq \Phi^{-1}(\xi_0) \leq (x + V'_-(x))|_{x=F_\mu^{-1}(\xi)}.$$

Since V is convex we have $V'_-(x) \leq V'_+(x)$, and thus

$$\Phi^{-1}(\xi_0) = (x + V'_-(x))|_{x=F_\mu^{-1}(\xi)}.$$

Finally for every local minimum at ξ_0,

$$G(\xi_0) = \left(F'_\mu(x) - \Phi'(x + V'_-(x))\right)|_{x=F_\mu^{-1}(\xi_0)} \geq 0,$$

by Lemma 5.12. □

6. Similar results with μ not necessarily log-concave

According to [3] the Brascamp-Lieb moment inequality (2) has been used in the setting of gradient interface models. It is also used to derive the tightness of finite-volume Gibbs measures, that describe the static interface, or strict convexity of the associated surface tension. The potential of a gradient interface model corresponds to the convex function V in Equation (1). Recently, gradient interface models with nonconvex potentials have been studied with great interest. The approach to the proof of the Brascamp-Lieb moment inequality described in this thesis allows for a generalization to nonconvex potentials of gradient interface models. In this section we consider cases with V not necessarily convex, and derive similar results as Equation (2).

For simplicity, we only consider one-dimensional random variables. Let X be a one-dimensional centred random variable with distribution μ such that

$$\mu(dx) = \frac{1}{Z} e^{-U(x)}\, dx. \tag{64}$$

We call the function U in (64) the potential of μ. If we choose the potential U to be of the following form

$$U(x) = \frac{1}{2}|k(x)|^2 - \log k'(x), \text{ for all } x \in \mathbb{R}, \tag{65}$$

the distribution function F_μ of μ takes on a convenient form. If $k \in \mathcal{C}^1$,

$$\begin{aligned} F_\mu(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} k'(y) \exp\{-\frac{1}{2}|k(y)|^2\} \, dy \\ &= \Phi \circ k(x). \end{aligned}$$

Therefore, the function g defined as

$$g(x) := F_\mu^{-1} \circ \Phi(x),$$

equals the inverse of k, k^{-1}, and hence

$$g'(x) = \frac{1}{k'(g(y))}.$$

Exactly as in the proof of the Brascamp-Lieb moment inequality we use the Bass embedding on μ to obtain an explicit expression for a stopping time T with

$$T = \int_0^1 \mathbb{E}^2[g(B_1)|\mathcal{F}_t] \, dt. \tag{66}$$

We, again, use Theorem 4.6 to put the stopping time T in convex order with trivial, that is, non-random stopping times. If we find $\sqrt{\alpha}$ such that

$$k' \geq \sqrt{\alpha}, \tag{67}$$

then

$$g' \leq \frac{1}{\sqrt{\alpha}}. \tag{68}$$

For a centred Gaussian random variable Y_α with variance $\frac{1}{\alpha}$ Theorem 4.6 then gives

$$X \prec_{cx} Y$$

or equivalently for all $\psi : \mathbb{R} \to \mathbb{R}$ convex,

$$\mathbb{E}[\psi(X)] \leq \mathbb{E}[\psi(Y_\alpha)].$$

The above formalism can further be used to find a lower bound for the random variable X. Similarly, for k such that

$$k' \leq \sqrt{\beta}, \tag{69}$$

we have as before with $\geq$ instead of $\leq$,

$$g' \geq \frac{1}{\sqrt{\beta}}, \tag{70}$$

and consequently for a centred Gaussian random variable Y_β with variance $\frac{1}{\beta}$

$$Y_\beta \prec_{cx} X$$

or equivalently for all $\psi : \mathbb{R} \to \mathbb{R}$ convex,

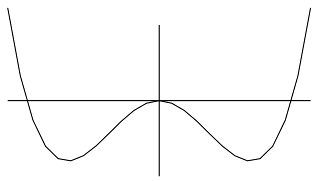
$$\mathbb{E}[\psi(Y_\beta)] \leq \mathbb{E}[\psi(X)].$$

If inequalities (67) and (69) both hold, equations (68) and (70) both hold, and consequently for all $\psi : \mathbb{R} \to \mathbb{R}$ convex,

$$\mathbb{E}[\psi(Y_\beta)] \leq \mathbb{E}[\psi(X)] \leq \mathbb{E}[\psi(Y_\alpha)]. \tag{71}$$

Example 6.1 (cf. [3], Example A.3)**.** *Set $k(x) = x + x^3$, then $k(x) \geq 1 =: \alpha$ and*

$$U(x) = \frac{1}{2}x^2 + x^4 + \frac{1}{2}x^6 - \log(1 + 3x^2).$$



(A) this potential has a double-well near the origin

Example 6.2. *Set $k(x) = \frac{x}{\sigma}$. Then $\alpha = \beta = 1/\sqrt{\sigma}$ and $U(x) = \frac{1}{2\sigma^2 x^2} - \log(1/\sigma)$. For μ we have*

$$\mu(dx) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{ -\frac{x^2}{2\sigma^2} \right\},$$

and μ is the distribution of a standard normal random variable and not suprisingly,

$$\mathcal{N}(0, \sigma^2) \prec_{cx} \mathcal{N}(0, \sigma^2) \prec_{cx} \mathcal{N}(0, \sigma^2).$$

Appendix A. Zusammenfassung auf Deutsch

In der vorliegenden Masterarbeit verwenden wir Resultate aus Stochastischer Analysis, speziell aus der Theorie der Konvexen Ordnung von Zufallsvariablen, Skorokhod-Einbettungen, und Optimaler Transport, um die Brascamp-Lieb Momente Ungleichung zu beweisen. Unser Beweis basiert auf einem Ansatz von Yuu Hariya [3], den wir durch Anwendung der konvexen Ordnung weiters verk´urzen konnten.

Die Arbeit bietet einen klareren Blick auf die Mechanismen hinter dem Beweis und Theorem 5.10 zeigt eine Verbindung von Zufallsvariablen, die log-konkaven zu einer Gauß'schen Zufallsvariable stehen. Zus´atzlich zeigen wir, dass die Brascamp-Lieb Momente Ungleichung f´ur konvexe Funktionen, die auf beiden Seiten unbegrenzt sind, direkt aus der Theorie der Ordnung von Zufallsvariablen folgt.

References

[1] H.J. Brascamp and E.H. Lieb, On extensions of the Brunn-Minkowski and Prkopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation, J. Funct. Anal. 22 (1976), 366389. MR-0450480

[2] H.J. Brascamp and E.H. Lieb, Best constants in Young's inequality, its converse, and its generalization to more than three functions, Advances in Mathematics Volume 20, Issue 2, 151-173

[3] Yuu Hariya, A connection of the Brascamp-Lieb inequality with Skorokhod embedding, Electronic Communications in Probability

[4] L.A. Caffarelli, Monotonicity properties of optimal transportation and the FKG and related inequalities, Commun. Math. Phys. 214 (2000), 547563; Erratum, Commun. Math. Phys. 225 (2002), 449450. MR-1800860 MR-1889232

[5] E.H.Lieb, Gaussian Kernels have only Gaussian Maximizers, Inventiones Mathematicae 102, pp. 179208 (1990)

[6] J. Obłój, The Skorokhod embedding problem and its offspring, Probab. Surv. 1, 321392 (2004)

[7] C. Villani, Topics in Optimal Transportation, Graduate Studies in Mathematics Volume 58, American Mathematical Society, 2003

[8] R. V. Chacon and J. B. Walsh. One-dimensional potential embedding. In Seminaire de Probabilités X, pages 1923. Lecture Notes in Math., Vol. 511. Springer, Berlin, 1976. MR445598

[9] M. Shaked and J.G. Shantikumar, Stochastic Orders, Springer Series in Statistics, 2007

[10] I. Karatzas and S.E. Shreves, Brownian Motion and Stochastic Calculus, Graduate Texts in Mathematics, Springer, 1991

[11] D. Revuz and M. Yor, Continuous Martingales and Brownian Motion, Grundlehren der Mathematischen Wissenschaften 2932, Springer-Verlag Berlin Heidelberg, 1991

[12] I. Karatzas and S.E. Shreves, Methods of Mathematical Finance, Graduate Texts in Mathematics, Springer, 1998

[13] R.F. Bass, Skorokhod Imbedding via Stochastic Integrals, in Séminaire de Probabilités, XVII, 221-224, Lecture Notes in Mathematics 986, Springer, Berlin, 1983 MR-770414

[14] J.M.C. Clark, The representation of functionals of Brownian motion as stochastic integrals. Ann. Math. Statistics 41, 1282,1295. Correction 1778.

[15] M.P. Yershov, On Stochastic Equations, Proceedings 2nd Japan-USSR Symposium on Probability Theory, Lecture Notes in Mathematics 330, 527-530, Springer-Verlag, Berlin, 1973.

[16] Ron Peled and Naomi Feldheim, Lecture Notes Random Walks and Brownian Motion, Tel Aviv University Spring 2011, `http://www.math.tau.ac.il/~peledron/Teaching/RW_and_BM_2011/scribe12.pdf`.

[17] D.L. Cohn, Measure Theory, Birkhäuser, 1997.

[18] R. Durret, Probability: theory and examples, Pacific Grove, Calif.: Wadsworth & Brooks Cole Advanced Books & Software ; 1991

[19] H. Bauer, Maß- und Integrationstheorie, 2. Auflage, Walter de Gruyter, 1992.

[20] P.A. Meyer, Probability and Potentials, Blaisdell Publishing Co, New York, 1966.