MASTERARBEIT | MASTER'S THESIS

Titel | Title

On the Bass Functional and the Bass Measure in Martingale Optimal Transport

verfasst von | submitted by

Valentin Zlabinger BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 821
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Mathematik
Betreut von | Supervisor:	Assoz. Prof. Dr. Julio Daniel Backhoff-Veraguas

Abstract

This thesis studies continuous-time martingale optimal transport and the distinguished role of Bass martingales, also known as standard stretched Brownian motion. After introducing the necessary background on optimal transport and martingale transport plans, the thesis reviews the martingale Benamou-Brenier problem and its relation to Bass martingales, duality theory and model-independent pricing in mathematical finance.

The main contributions concern the Bass functional and its minimizer, which characterizes the Bass measure of the optimal Bass martingale. Several new structural properties of the minimizer are established, including rotational invariance properties and results concerning its copula structure. Moreover, the minimizer is constructed explicitly in the case of Gaussian marginals, and the theory is illustrated through numerical simulations.

The final part of the thesis studies convexity properties of the Bass functional. In particular, a counterexample is constructed showing that the Bass functional fails to be displacement convex in dimension $d \geq 2$. This contrasts with the one-dimensional case, where displacement convexity was previously established, and resolves an open question posed in earlier work.

Zusammenfassung

Diese Arbeit beschäftigt sich mit dem Martingal-Optimal-Transportproblem und der besonderen Rolle von Bass-Martingalen, auch bekannt als Standard Stretched Brownian Motion. Nach einer Einführung in die Grundlagen des optimalen Transports und der Martingal-Transportpläne behandelt die Arbeit das Martingal-Benamou-Brenier-Problem sowie dessen Zusammenhang mit Bass-Martingalen, der Dualitätstheorie und der modellunabhängigen Preisbewertung in der Finanzmathematik.

Die Hauptbeiträge der Arbeit betreffen das Bass-Funktional und dessen Minimierer, der das Bass-Maß des optimalen Bass-Martingals charakterisiert. Es werden mehrere neue strukturelle Eigenschaften des Minimierers nachgewiesen, darunter Rotationsinvarianzeigenschaften sowie Resultate bezüglich seiner Copula Struktur. Darüber hinaus wird der Minimierer im Fall Gaußscher Randverteilungen explizit konstruiert, und die Theorie wird durch numerische Simulationen veranschaulicht.

Der abschließende Teil der Arbeit untersucht Konvexitätseigenschaften des Bass-Funktionals. Insbesondere wird ein Gegenbeispiel konstruiert, das zeigt, dass das Bass-Funktional in Dimension $d \geq 2$ nicht displacement konvex ist. Dies steht im Gegensatz zum eindimensionalen Fall, in dem die displacement Konvexität zuvor nachgewiesen wurde, und beantwortet damit eine in früheren Arbeiten aufgeworfene Frage.

Contents

1 Introduction

Optimal transport studies the most efficient way to transfer mass between probability distributions by minimizing a prescribed cost. The origins of the optimal transport problem go back to the work of G. Monge [24] in 1781. Almost 150 years later, the problem was reformulated in modern terms by L. Kantorovich [19]. Over the past decades, further major developments were achieved through the seminal works of J. Benamou, Y. Brenier and R. McCann [13, 14, 15, 23], which established many of the cornerstones of the modern theory.

Martingale optimal transport extends the classical transport problem by imposing an additional martingale constraint on admissible transport plans. Such constraints arise naturally in mathematical finance, where no-arbitrage assumptions lead to martingale dynamics and martingale optimal transport leads to model-independent bounds for derivative prices [11].

A continuous-time martingale optimal transport problem was introduced in [8, 7] through a probabilistic analogue of the problem studied by Benamou and Brenier. This optimization problem admits a solution which is unique in law. Under an irreducibility assumption, the optimizer is given by a Bass martingale, also referred to as standard stretched Brownian motion. It may be interpreted as the martingale which correlates as closely as possible with Brownian motion while satisfying prescribed initial and terminal marginals.

Therefore, the class of Bass martingales, which are constructed by transporting Brownian motion through the gradient of a convex function, plays a distinguished role in this theory. Moreover, for the optimal Bass martingale, the initial distribution of the underlying Brownian motion is characterized as the minimizer of the Bass functional. Roughly speaking, this functional is given by the difference between the maximal covariance of the Brownian motion with the prescribed terminal marginal and the maximal covariance with the prescribed initial marginal. Furthermore, the convex function that appears in the construction of the optimal Bass martingale is related to the optimizer of a corresponding dual problem.

This thesis is organized as follows: The first chapter introduces the basic notions of martingale optimal transport and motivates the theory through an application in mathematical finance. The second chapter discusses the martingale Benamou-Brenier problem as well as its connection to Bass martingales. Moreover, it introduces the Bass functional and characterizes its connection to the transport problem. In particular, several useful properties of the minimizer of the Bass functional are established, and the theory is illustrated through numerical simulations. Finally, the third chapter studies convexity properties of the Bass functional, including higher-dimensional obstructions, an L^2-Bass gradient flow and convergence to the minimizer under suitable assumptions.

2 Introduction to martingale optimal transport

This chapter provides a brief introduction to martingale optimal transport and introduces the concepts required in later chapters. The first subsection presents some of the notation and basic ideas of classical optimal transport which will be used throughout the thesis, without attempting to give a comprehensive overview of the theory; for this, see [1, 2, 30, 31]. The second and third subsections focus on the space of martingale transport plans and the relation between their existence and convex order. Finally, the last section motivates the theory through an application to model-independent pricing in mathematical finance.

2.1 Probability measures and Wasserstein spaces

Definition 2.1 (Transport plans)**.** Given $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, define

$$\Pi(\mu,\nu) := \begin{Bmatrix} \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \pi(A \times \mathbb{R}^d) = \mu(A), \pi(\mathbb{R}^d \times B) = \nu(B) \\ \forall \text{ Borel sets } A, B \subset \mathbb{R}^d \end{Bmatrix}.$$

Definition 2.2. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. For $\mu \in \mathcal{P}(X)$ and a measurable map $f : X \to Y$ the pushforward measure of μ is denoted by $f(\mu)$ and defined via

$$f(\mu)(B) = \mu(f^{-1}(B)) \quad \forall B \in \mathcal{B}.$$

The pushforward can also be characterized via the change of variables formula

$$\int_Y \phi \, df(\mu) = \int_X (\phi \circ f) \, d\mu \quad \forall \phi \in C_b(Y).$$

Denoting by p^1, p^2 the coordinate projections, transport plans can also equivalently be characterized by

$$p^1(\pi) = \mu, \quad p^2(\pi) = \nu.$$

Definition 2.3. Let X be a Polish space. The weak topology on $\mathcal{P}(X)$ is the smallest topology such that the maps

$$\mu \mapsto \int_X f(x) \, d\mu(x)$$

are continuous for every bounded continuous function $f \in C_b(X)$.

Therefore, a sequence $(\mu_n) \subset \mathcal{P}(X)$ converges weakly to $\mu \in \mathcal{P}(X)$ if and only if

$$\int_X f(x) \, d\mu_n(x) \to \int_X f(x) \, d\mu(x)$$

for all $f \in C_b(X)$.

In Kantorovich's formulation of optimal transport, one seeks to find

$$\inf_{\pi\in\Pi(\mu,\nu)} \left\{ \int c(x,y)\, d\pi(x,y) \right\}, \tag{2.1}$$

where the cost function c represent the costs of moving mass from x to y. We denote by $\Pi_o(\mu,\nu)$ the possibly empty set of transport plans for which the infimum is attained. A corollary of Prokhorov's theorem shows that with respect to this weak topology, the set of transport plans $\Pi(\mu,\nu)$ is compact. Further, for (X,d) a Polish metric space and $p\in[1,\infty)$ we define the Wasserstein space:

$$\mathcal{P}_p(X) := \left\{ \mu\in\mathcal{P}(X) : \int d(x,x_0)^p\, d\mu(x) < \infty \text{ for some } x_0\in X \right\}$$

In this thesis, we will often consider transportation problems associated with the quadratic cost function $c(x,y)=\frac{1}{2}|x-y|^2$. In this case, we have the following theorem which is due to Y. Brenier [14, 15] and M. Knott and C.S. Smith [20] and plays a central role in the theory of optimal transport.

Theorem 2.4 ([31, Theorem 2.12])**.** *Let $\mu,\nu\in\mathcal{P}_2(\mathbb{R}^d)$ and μ absolutely continuous with respect to the Lebesgue measure. For the quadratic cost $c(x,y)=\frac{1}{2}|x-y|^2$ we have:*

1. *The problem* (2.1) *has a unique solution π. Moreover, π is concentrated on the graph of a measurable map $T:\mathbb{R}^d\mapsto\mathbb{R}^d$ so that $\pi=(\mathrm{id},T)(\mu)$ and $T=\nabla\psi$ μ-a.e., where $\psi:\mathbb{R}^d\mapsto(-\infty,\infty]$ is lower semicontinuous, convex and differentiable μ-almost everywhere. This map T is the unique solution of Monge's transportation problem:*

$$\inf_{\substack{T:\mathbb{R}^d\to\mathbb{R}^d \text{ Borel}\\ T(\mu)=\nu}} \left\{ \int \frac{1}{2}|x-T(x)|^2\, d\mu(x) \right\} \tag{2.2}$$

2. *Conversely, if $T=\nabla\psi$ for some convex, lower semicontinuous and μ-almost everywhere differentiable function with $|\nabla\psi|\in L^2(\mu)$, then T is optimal for* (2.2) *if $\nu=T(\mu)$.*

3. *Finally, if ν is also absolutely continuous with respect to the Lebesgue measure, then for $T^{\mu\to\nu}$ (respectively $T^{\nu\to\mu}$) denoting the unique optimal transport map between μ and ν (respectively ν and μ) in* (2.2) *we have*

$$T^{\nu\to\mu}\circ T^{\mu\to\nu}(x)=x, \quad T^{\mu\to\nu}\circ T^{\nu\to\mu}(y)=y,$$

for μ-a.e. x and ν-a.e. y.

Definition 2.5 (Wasserstein distance)**.** Let (X,d) be a Polish metric space and let $p\in[1,\infty)$. For $\mu,\nu\in\mathcal{P}_p(X)$, the p-th Wasserstein distance between μ and ν is defined by

$$\mathcal{W}_p(\mu,\nu) := \left(\inf_{\pi\in\Pi(\mu,\nu)} \int d(x,y)^p\, d\pi(x,y) \right)^{1/p}.$$

In particular, the Wasserstein distance defines a metric on $\mathcal{P}_p(X)$ and the space $(\mathcal{P}_p(X), \mathcal{W}_p)$ is a Polish metric space. It can also be shown that there is an equivalence between convergence with respect to the p-th Wasserstein distance and weak convergence in $\mathcal{P}(X)$ together with convergence of p-th moments in the sense that for some $x_0 \in X$ we have

$$\int d(x_0, x)^p \, d\mu_n(x) \to \int d(x_0, x)^p \, d\mu(x).$$

Finally, we shall define the term Wasserstein geodesic.

Definition 2.6. A curve $(\mu_t)_{0 \leq t \leq 1} \subset \mathcal{P}_p(X)$ is called a Wasserstein geodesic if

$$\mathcal{W}_p(\mu_s, \mu_t) = (t - s)\mathcal{W}_p(\mu_0, \mu_1) \quad \forall 0 \leq s \leq t \leq 1.$$

2.2 Martingale transport plans

This subsection will introduce the space $\Pi_M(\mu, \nu)$ of martingale couplings. It is important to note that throughout the rest of this thesis we only consider measures in $\mathcal{P}(\mathbb{R}^d)$ with finite first moment, i.e., in $\mathcal{P}_1(\mathbb{R}^d)$.

Definition 2.7. Given $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, define

$$\Pi_M(\mu, \nu) := \{\pi \in \Pi(\mu, \nu) : \mathbb{E}_\pi[(y - x)h(x)] = 0, \forall h : \mathbb{R}^d \to \mathbb{R} \text{ Borel bounded}\}.$$

Equivalently, the martingale property could have been written as

$$\int y \, \pi_x(dy) = x \text{ for } \mu\text{-a.e. } x,$$

where $(\pi_x)_{x \in \mathbb{R}^d}$ is the disintegration of π with respect to its first marginal μ, so that

$$\pi(dx, dy) = \mu(dx)\pi_x(dy).$$

It is clear that this space is also convex. Indeed, if $\pi, \rho \in \Pi_M(\mu, \nu)$ and any $\lambda \in [0, 1]$, then $\gamma := \lambda\pi + (1 - \lambda)\rho$ is still in $\Pi(\mu, \nu)$ and the martingale property also carries over. In addition, the following proposition will show that it is a closed subset of $\Pi(\mu, \nu)$ with respect to the weak topology. Since the latter is compact in the weak topology, this shows that $\Pi_M(\mu, \nu)$ is also weakly compact.

Proposition 2.8 ([11, Proposition 2.4])**.** *$\Pi_M(\mu, \nu)$ is compact in the weak topology.*

Proof. As already mentioned above, we only have to show that it is closed. Therefore, let $(\pi_m) \subset \Pi_M(\mu, \nu)$ and suppose that $\pi_m \to \pi$ weakly for some $\pi \in \Pi(\mu, \nu)$. We need to show that π also has the martingale property. Let h be bounded and Borel. We have

$$\int h(x)(y - x) \, d\pi_m(x, y) = 0.$$

Moreover, by Lusin's theorem we can find $\forall \epsilon > 0$ a compact set $K_\epsilon \subset \mathbb{R}^d$ such that $\mu(\mathbb{R}^d \setminus K_\epsilon) < \epsilon$ and the restriction of h to K_ϵ is continuous. By Tietze's extension theorem, there exists a continuous and bounded extension $\tilde{h}$ of $h \restriction_{K_\epsilon}$ to all of $\mathbb{R}^d$. The error term

$$\int h(x)(y-x)\, d\pi(x,y) - \int \tilde{h}(x)(y-x)\, d\pi(x,y)$$

can be made arbitrarily small by choosing ϵ small enough since h and $\tilde{h}$ are bounded and μ and ν have finite first moments. Therefore, in the following we may only consider h bounded and continuous. Let $\eta_a : \mathbb{R}^d \times \mathbb{R}^d \mapsto [0,1]$ be a continuous cutoff function equal to 1 on $[-a,a]^{2d}$ and 0 outside $[-a-1, a+1]^{2d}$. Since $\Pi(\mu,\nu)$ is weakly compact and $\eta_a(x,y)h(x)(y-x)$ is bounded and continuous,

$$\int \eta_a(x,y)h(x)(y-x)\, d\pi_m(x,y) \to \int \eta_a(x,y)h(x)(y-x)\, d\pi(x,y).$$

Moreover, as μ, ν have finite first moments,

$$\int (1-\eta_a)h(x)(y-x)\, d\pi_m \to 0 \text{ uniformly as a} \to \infty$$

because h was bounded and therefore $h(x)(y-x)$ has at most linear growth. This implies

$$\int h(x)(y-x)\, d\pi(x,y) = 0.$$

Since h was arbitrary, π also has the martingale property. □

2.3 Convex order

This subsection, further characterizes the space $\Pi_M(\mu,\nu)$. A priori, it is not even clear whether this set is nonempty. This is different from the classical transport problem, where the product measure of μ and ν always gives an element of $\Pi(\mu,\nu)$. However, assuming that there is a $\pi \in \Pi_M(\mu,\nu)$ and φ convex then by Jensen's inequality we have

$$\begin{aligned}\int \varphi(y)\, d\nu(y) = \int \varphi\, d\pi(x,y) &= \int\int \varphi(y)\, d\pi_x(y)\, d\mu(x) \\ &\geq \int \varphi\left(\int y\, d\pi_x(y)\right) d\mu(x) = \int \varphi(x)\, d\mu(x).\end{aligned}$$

This motivates the following definition.

Definition 2.9. Two measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ are in *convex order* denoted by $\mu \preceq_c \nu$ if for all convex functions $\varphi : \mathbb{R}^d \to \mathbb{R}$

$$\int \varphi(x)\, d\mu(x) \leq \int \varphi(y)\, d\nu(y).$$

The calculations above showed, that convex order is indeed a necessary condition for the set $\Pi_M(\mu,\nu)$ to be nonempty. That the converse is also true is the content of the following theorem.

Theorem 2.10 ([26, Theorem 8])**.** *For two probability measures in $\mathbb{R}^d$, the set $\Pi_M(\mu,\nu)$ is nonempty if and only if $\mu \preceq_c \nu$.*

In order to further characterize this property, we reproduce part of the proof for probability measures on $\mathbb{R}$ from the second chapter of [12]. The following are some examples of probability measures in convex order.

Example 2.11. For μ a Dirac measure δ_x, the condition $\mu \preceq_c \nu$ is equivalent to ν having barycenter x. This is also exactly the martingale condition and therefore in this case the above theorem already follows. In fact, when considering continuous martingales, the existence of such a martingale is discussed in subsection 3.4.

Example 2.12. If $\mu_i \preceq_c \nu_i$, $i = 1,\ldots,n$ then their sum is also in convex order i.e. $\sum_{i=1}^n \mu_i/n \preceq_c \sum_{i=1}^n \nu_i/n$.

Example 2.13. If μ is concentrated on $[a,b]$ and ν is concentrated on $\mathbb{R}\setminus[a,b]$ with μ and ν having the same barycenter. Then for each convex function φ we have

$$\int \varphi\, d\mu \leq \int \psi\, d\mu = \int \psi\, d\nu \leq \int \varphi\, d\nu,$$

where ψ is the linear function defined via $\varphi(a) = \psi(a)$ and $\varphi(b) = \psi(b)$. This shows that $\mu \preceq_c \nu$.

For the following proposition note that the notion of convex order can be generalized to arbitrary finite measures, provided that they have finite first moments. Moreover, Section 4.1 in [12], shows that if μ and ν have the same mass and barycenter, then for convex order it suffices to check against convex functions φ which are positive and with at most linear growth.

Proposition 2.14 ([12, Proposition 2.7])**.** *For μ a sum of Dirac measures $\mu = \sum_{i=1}^n \alpha_i\delta_i$ on $\mathbb{R}$, then $\Pi_M(\mu,\nu)$ is nonempty if and only if $\mu \preceq_c \nu$.*

Proof. The proof is via induction over the number of atoms of μ. For $n=1$ the claim follows from Example 2.11. Now, assume the result is known for measures with $n-1$ atoms, then we can split $\mu = \alpha_1\delta_1 + \tilde{\mu}$. The goal is to show that we can also split $\nu = \nu' + \tilde{\nu}$ so that $\alpha_1\delta_1 \preceq_c \nu'$, $\tilde{\mu} \preceq_c \tilde{\nu}$ and then conclude as in Example 2.12. Therefore, let

$$\nu_s := G(\lambda_{[s,s+\alpha_1]}),$$

be the pushforward of the Lebesgue measure restricted to $[s, s+\alpha_1]$ via a quantile function G of ν. Via testing the convex order with the functions $(\cdot - G(\alpha_1))_-$ and $(\cdot - G(\nu(\mathbb{R}) - \alpha_1))_+$ it can be seen that there exists a $s' \in [0, \nu(\mathbb{R}) - \alpha_1]$ such that

$\nu' := \nu_{s'}$ has the same mass and barycenter as $\alpha_1\delta_1$. For $\tilde{\nu} := \nu - \nu'$, we have $\tilde{\nu}(I) = 0$ by construction, where $I := \mathrm{conv}(\mathrm{spt}(\nu'))^\circ$ is the interior of the smallest interval containing the support of ν'. In order to check $\tilde{\mu} \preceq_c \tilde{\nu}$, let φ be a positive, convex function with at most linear growth. Then we can construct a function ψ such that

$$\psi = \varphi \text{ on } \mathbb{R} \setminus I, \quad \psi \text{ is affine on } I, \quad \varphi \leq \psi.$$

We then have

$$\int \varphi \, d\tilde{\mu} \leq \int \psi \, d\tilde{\mu} = \int \psi \, d\mu - \int \psi \, d\alpha_1\delta_1 \leq \int \psi \, d\nu - \int \psi \, d\nu' = \int \psi \, d\tilde{\nu} = \int \varphi \, d\tilde{\nu}.$$

For the second inequality we used that $\mu \preceq_c \nu$ and the fact that ψ is affine on I and therefore $\int \psi \, d\alpha_1\delta_1 = \int \psi \, d\nu'$. The last equality follows because $\tilde{\nu}$ is concentrated on $\mathbb{R} \setminus I$. $\square$

With the help of an approximation argument this proposition can be generalized and used to prove Theorem 2.10 in $\mathbb{R}$.

2.4 Model-independent pricing

The goal of this subsection is to motivate martingale optimal transport by deriving model-independent bounds for option prices. This is done by introducing the main result from [11]. Since the introduction of the Black-Scholes framework, many more sophisticated models have been developed to determine the risk of exotic options, such as local volatility models, stochastic volatility models, jump-diffusion models or mixed local stochastic volatility models. However, even if all of these models are calibrated to the same market prices for vanilla options, they can still produce significantly different prices for exotic derivatives. Indeed, the no-arbitrage assumption only requires the dynamics of forward prices to be (local) martingales, which are not uniquely determined by such a calibration.

For this reason, it is important to study model-independent approaches that derive upper and lower bounds for exotic option prices using only the no-arbitrage assumption and observable market data. Such bounds would provide robust pricing intervals that are valid across all models consistent with the given marginals and provided that these bounds have an interpretation as investment strategies they would be useful to detect the possibility of arbitrage in market prices.

Throughout, we fix a single asset S, and consider an exotic option whose payoff depends only on the values of S at times $t_1 < \cdots < t_n$. The payoff is specified by a measurable function Φ and is given by $\Phi(S_1, \ldots, S_n)$. Under the no-arbitrage assumption, the standard approach is to consider a probability measure $\mathbb{Q}$ on $\mathbb{R}^n$ under which the coordinate process $(S_i)_{i=1}^n$ is required to be a discrete-time martingale with respect to its natural filtration. A fair price of this option is then given by its expected payoff

$$\mathbb{E}_{\mathbb{Q}}[\Phi].$$

Further, we require the model to be calibrated to a continuum of call options with payoffs $(S_i - K)_+$, $K \in \mathbb{R}$, $i = 1, \dots, n$ and observed market prices $C(t_i, K)$. This is done by imposing

$$C(t_i, K) = \mathbb{E}_{\mathbb{Q}}[(S_i - K)_+] \quad K \in \mathbb{R}, \quad i = 1, \dots, n.$$

Equivalently, if $\mu_i := \mathrm{Law}(S_i)$ denotes the law of S_i, then the one-dimensional marginals of $\mathbb{Q}$ satisfy

$$\mathbb{Q}_i = \mu_i \quad i = 1, \dots, n,$$

where μ_i are uniquely determined by the identities $C(t_i, K) = \int (x - K)_+ \, \mu_i(dx)$ $(\forall K)$. Extending the notation from before we denote the set of these martingale measures $\mathbb{Q}$ by $\Pi_M(\mu_1, \dots, \mu_n)$. The *primal problem* is then defined as

$$P = \inf\{\mathbb{E}_{\mathbb{Q}}[\Phi] : \mathbb{Q} \in \Pi_M(\mu_1, \dots, \mu_n)\} \tag{2.3}$$

We shall see that this corresponds to the following dual formulation. We are interested in payoffs of the form

$$\Psi_{(u_i),(\Delta_j)}(s_1, \dots, s_n) := \sum_{i=1}^{n} u_i(s_i) + \sum_{j=1}^{n-1} \Delta_j(s_1, \dots, s_j)(s_{j+1} - s_j),$$

where $u_i : \mathbb{R} \mapsto \mathbb{R}$ satisfy $u_i \in L^1(\mu_i)$ for $i = 1, \dots, n$, and the functions $\Delta_j : \mathbb{R}^j \mapsto \mathbb{R}$ are bounded and measurable for $j = 1, \dots, n-1$. Additionally, we impose

$$\Phi \geq \Psi_{(u_i),(\Delta_j)}.$$

This leads to the following inequality for every $\mathbb{Q} \in \Pi_M(\mu_1, \dots, \mu_n)$

$$\mathbb{E}_{\mathbb{Q}}[\Phi] \geq \mathbb{E}_{\mathbb{Q}}[\Psi_{(u_i),(\Delta_j)}] = \mathbb{E}_{\mathbb{Q}}\left[\sum_{i=1}^{n} u_i(S_i)\right] = \sum_{i=1}^{n} \mathbb{E}_{\mu_i}[u_i]. \tag{2.4}$$

Therefore, it makes sense to consider the *dual problem*

$$D = \sup\{\sum_{i=1}^{n} \mathbb{E}_{\mu_i}[u_i] : \exists \Delta_1, \dots, \Delta_{n-1} \text{ s.t. } \Psi_{(u_i),(\Delta_j)} \leq \Phi\}. \tag{2.5}$$

Indeed, this dual formulation admits a natural financial interpretation in terms of semi-static subhedging strategies. Such a strategy consists of a static portfolio of vanilla options $u_i(S_i)$ together with dynamic trading in the underlying asset according to a self-financing strategy $(\Delta_j)_{j=1}^{n-1}$. Moreover, suppose that the exotic option Φ is offered in the market at a price $p < D$. Then there exist $(u_i), (\Delta_j)$ with $\Psi_{(u_i),(\Delta_j)} \leq \Phi$ while the cost $\sum_{i=1}^{n} \mathbb{E}_{\mu_i}[u_i]$ is strictly larger than p. Hence, one may buy the option Φ and simultaneously short the subhedging portfolio $\Psi_{(u_i),(\Delta_j)}$, thereby locking in an arbitrage profit. This is reflected by the following inequality which follows from (2.4).

$$P \geq D \tag{2.6}$$

The following theorem shows that we have in fact equality in (2.6) under relatively mild assumptions. Therefore any option priced below P gives rise to an arbitrage opportunity through semi-static subhedging.

Theorem 2.15 ([11, Theorem 1])**.** *Assume that $\mu_1, \dots, \mu_n \in \mathcal{P}(\mathbb{R})$ satisfy $\mu_1 \preceq_c \dots \preceq_c \mu_n$, and hence $\Pi_M(\mu_1, \dots, \mu_n) \neq \emptyset$. Let $\Phi : \mathbb{R}^n \to (-\infty, \infty]$ be a lower semicontinuous function such that*

$$\Phi(s_1, \dots, s_n) \geq -K \cdot (1 + |s_1| + \dots + |s_n|) \tag{2.7}$$

for some constant K. Then there is no duality gap, i.e. P=D. Moreover, the infimum for the primal problem P is attained.

Similarly, it is possible to give an upper bound for the price of the option Φ via semi-static superhedging. Finally, a related duality framework for continuous-time semimartingale transport was developed in [27].

3 Bass martingales

First, I will give a quick introduction into the martingale optimization problem that I will be considering from now on. In the beginning, I will follow [8]. Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the space of probability measures on $\mathbb{R}^d$ with finite second moment. As discussed in the previous chapter, we will require that the measures are in convex order, denoted by $\mu \preceq_c \nu$. The optimization problem is

$$MT(\mu, \nu) := \sup_{\substack{M_t = M_0 + \int_0^t \sigma_s \, dB_s \\ M_0 \sim \mu, M_1 \sim \nu}} \mathbb{E}\left[\int_0^1 |\sigma_t - I_d|_{\text{HS}}^2 \, dt\right], \tag{MBB}$$

where $|\cdot|_{\text{HS}}$ denotes the Hilbert-Schmidt norm, B is a Brownian motion on $\mathbb{R}^d$, $(\sigma_t)_{0 \leq t \leq 1}$ are progressively measurable $\mathbb{R}^{d \times d}$-valued processes and the supremum is taken over the class of all filtered probability spaces $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in [0,1]})$ supporting a martingale M satisfying $M_0 \sim \mu$ and $M_1 \sim \nu$. The optimizer of this problem is often referred to as *stretched Brownian motion.* By using the identity

$$|\sigma_t - I_d|_{\text{HS}}^2 = \text{tr}(\sigma_t \sigma_t^\top) - 2 \, \text{tr}(\sigma_t) + d$$

and Itô's formula it can be seen that we have

$$MT(\mu, \nu) = \int |y|^2 \, d\nu - \int |x|^2 \, d\mu - 2P(\mu, \nu) + d,$$

where P describes the optimization problem

$$P(\mu,\nu) := \sup_{\substack{M_t = M_0 + \int_0^t \sigma_s \, dB_s \\ M_0 \sim \mu, M_1 \sim \nu}} \mathbb{E}\left[\int_0^1 \mathrm{tr}(\sigma_t)\, dt\right]. \tag{3.1}$$

Hence (MBB) is equivalent to maximizing the covariance between a Brownian motion and M.

3.1 Existence and uniqueness of solutions

In order to deduce the existence and uniqueness of solutions, we will actually look at a further, static analogue of this optimization problem. First, the maximal covariance MCov between two probability measures $p_1, p_2 \in \mathcal{P}_2(\mathbb{R}^d)$ is defined as

$$\mathrm{MCov}(p_1, p_2) := \sup_{q \in \Pi(p_1, p_2)} \int \langle x_1, x_2 \rangle \, q(dx_1, dx_2).$$

Let γ denote the standard Gaussian measure, we define the weak optimal transport problem as

$$WOT(\mu,\nu) := \sup_{\pi \in \Pi_M(\mu,\nu)} \int \mathrm{MCov}(\pi_x, \gamma)\, \mu(dx). \tag{WOT}$$

Here, the family of probability measures $\{\pi_x\}_{x \in \mathbb{R}^d} \subseteq \mathcal{P}_2(\mathbb{R}^d)$ is obtained via disintegration of π with respect to the first marginal μ. So the supremum could also be written as over the set $\{\{\pi_x\}_x : \int \pi_x \, \mu(dx) = \nu \text{ and } \int y \, \pi_x(dy) = x\}$.

The optimization problem is referred to as a *weak* optimal transport problem, because the mapping $\pi \to \int \mathrm{MCov}(\pi_x, \gamma)\, \mu(dx)$ is non-linear, whereas in the classical optimal transport setting, linear problems of the form $\pi \to \int c(x,y)\, \pi(dx, dy)$ are studied.

Expanding the square, we immediately have the following relation for the usual quadratic Wasserstein distance associated with the cost $c(x,y) = |x-y|^2$

$$\mathcal{W}_2^2(p_1, p_2) = \int |x|^2 \, p_1(dx) + \int |y|^2 \, p_2(dy) - 2\,\mathrm{MCov}(p_1, p_2). \tag{3.2}$$

Applying the identity with $p_1 = \pi_x$ and $p_2 = \gamma$, then integrating with respect to $\mu(dx)$ and taking the infimum over $\pi \in \Pi_M(\mu,\nu)$, yields

$$\inf_{\pi \in \Pi_M(\mu,\nu)} \int \mathcal{W}_2^2(\pi_x, \gamma)\, \mu(dx) = d + \int |y|^2 \, \nu(dy) - 2\,\mathrm{WOT}(\mu,\nu). \tag{3.3}$$

Again, this leads to an analogous characterization: In this static formulation, maximizing the maximal covariance is equivalent to minimizing the quadratic Wasserstein distance.

Theorem 3.1 ([8, Theorem 2.2])**.** *The problem* (MBB) *and* (WOT) *are equivalent. In particular we have:*

1. $WOT = P(\mu, \nu) < \infty$.

2. (WOT) *has a unique optimizer* π^*.

3. (MBB) *has a unique-in-law optimizer* M^*.

4. $\pi^* = \mathrm{Law}(M_0^*, M_1^*)$ *and* $M^* = G(\pi^*)$ *for some function* G.

Proof. We will only prove 1. The rest of the proof can be found in [8], or alternatively follows from the more general theory developed in [29]. First, let M be a candidate Martingale for (MBB), then we have

$$\mathbb{E}\left[\int_0^1 \mathrm{tr}(\sigma_\mathrm{t})\,\mathrm{dt}\right] = \mathbb{E}[M_1 \cdot B_1 - M_0 \cdot B_0] = \mathbb{E}[\mathbb{E}[M_1 \cdot (B_1 - B_0)]|M_0],$$

where we used Itô's formula for the first and the martingale property for the second equality. Recalling that with this notation we have $\pi_x = \mathrm{Law}(M_1|M_0)$, we can define a $q_x \in \Pi(\pi_x, \gamma)$ via $q_x = \mathrm{Law}(M_1, B_1 - B_0|M_0 = x)$. Hence, we have

$$\mathbb{E}\left[\int_0^1 \mathrm{tr}(\sigma_\mathrm{t})\,\mathrm{dt}\right] = \int\int \langle x_1, x_2\rangle\, q_x(dx_1, dx_2)\,\mu(dx).$$

This implies (3.1) $\leq$ (WOT). For the other direction, again let π be a candidate coupling for (WOT). Note that by Brenier's theorem for each π_x there is a convex $F^x(\cdot)$, such that $\nabla F^x(\gamma) = \pi_x$. For a standard Brownian motion with filtration $\mathcal{F}^B$ we can then define $M_t^x := \mathbb{E}[\nabla F^x(B_1)|\mathcal{F}_t^B]$. In particular, after possibly enlarging our probability space, there exists a random variable $X \sim \mu$ independent of the Brownian motion B. By our assumptions on π, we have that $M_0^x = \int y\,\pi_x(dy) = x$ and $\int \pi_x(dy)\,\mu(dx) = \nu(dy)$. Therefore, $\{M_t^X\}_{t\in[0,1]}$ is a continuous martingale from μ to ν, which satisfies

$$\begin{aligned}\int \mathrm{MCov}(\pi_x, \gamma)\,\mu(dx) = \int\int \langle \nabla F^x(b), b\rangle\,\gamma(db)\,\mu(dx) &= \mathbb{E}[\mathbb{E}[B_1 \cdot M_1^X|X]] \\ &= \mathbb{E}\left[\int_0^1 \mathrm{tr}(\sigma_\mathrm{t})\,\mathrm{dt}\right].\end{aligned}$$

Where the last equality follows as above by Itô's formula and the martingale property and σ can be computed from ∇F^x. This implies (3.1) $\geq$ (WOT). The finiteness of (WOT) follows from μ and ν having finite second moment and the inequality $m \cdot b \leq |m|^2 + |b|^2$ $\square$

Remark 3.2. The above theorem now also shows that the precise underlying probability space in (MBB) is irrelevant, as long as it is rich enough, e.g. it supports a $\mathcal{F}_0$ measurable random variable with continuous distribution. Additionally, the proof also shows how one could construct an optimizer and how the particular function G in the theorem looks like. Let γ^t be the law of a centered standard Brownian motion at time t, i.e. the d-dimensional centered Gaussian distribution with covariance matrix tI_d. In order to find the optimizer one has to:

1. Find the unique optimizer π^* of (WOT).
2. Find the convex Brenier potentials, unique up to a constant, F^x such that $\nabla F^x(\gamma) = \pi^*_x$.
3. Define $M^x_t := \mathbb{E}[\nabla F^x(B_1)|B_t] = \int \nabla F^x(y + B_t)\, \gamma^{1-t}(dy)$.
4. Find a random variable $X \sim \mu$, independent of B and define $M_t := M^X_t$.

The previous remark can be used to prove the following lemma, which shows that (MBB) obeys a dynamic programming principle and that the unique optimizer M^* is a strong Markov martingale. Similarly as before, assume that $\{B_t\}_{t\geq 0}$ is a Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_t)$. Additionally, suppose that the space Ω carries a continuously distributed random variable which is independent of the filtration $(\mathcal{F}_t)_t$. For a $(\mathcal{F}_t)_t$-stopping time $0 \leq \tau \leq T$, define

$$V(\tau, T, \overline{\mu}, \overline{\nu}) := \sup_{\substack{M_s = X + \int_\tau^s \sigma_r\, dB_r,\, \tau \leq s \leq T \\ X \sim \overline{\mu},\, M_T \sim \overline{\nu}}} \mathbb{E}\left[\int_\tau^T \operatorname{tr}(\sigma_r)\, dr\right]. \tag{3.4}$$

In particular, we have $P(\mu, \nu) = V(0, 1, \mu, \nu)$. Note that, similarly as before, when taking the supremum, it does not matter whether one works on the original probability space or on an extension of it.

Lemma 3.3 ([8, Lemma 2.4])**.** *For every stopping time $0 \leq \tau \leq 1$ we have*

$$V(0, 1, \mu, \nu) = \sup_{\substack{M_s = M_0 + \int_0^s \sigma_r\, dB_r \\ 0 \leq s \leq \tau,\, M_0 \sim \mu}} \left\{\mathbb{E}\left[\int_0^\tau \operatorname{tr}(\sigma_r)\, dr\right] + V(\tau, 1, \operatorname{Law}(M_\tau), \nu)\right\}, \tag{3.5}$$

with the usual convention that $\sup \emptyset = -\infty$. Moreover, for M^*, which is also an optimizer for $V(0, 1, \mu, \nu)$, we have

1. *$M^*|_{[\tau,1]}$ is optimal for $V(\tau, 1, \operatorname{Law}(M^*_\tau), \nu)$,*
2. *$M^*|_{[0,\tau]}$ is optimal for $V(0, \tau, \mu, \operatorname{Law}(M^*_\tau))$,*
3. *A.s. we have $\operatorname{Law}(M^*_1 | M^*_s, s \leq \tau) = \operatorname{Law}(M^*_1 | M^*_\tau)$.*

Proof. In order to show (3.5), we only have to show $\geq$, since the l.h.s. is obviously smaller. Therefore, let M^1 be admissible for the right-hand side, so that M^1 is adapted to a filtration $\{\mathcal{F}_{s\wedge\tau}\}_{s\geq 0}$ and $dM^1 = \sigma^{(1)} dB^1$, where B^1 is a Brownian motion on $[0, \tau]$ with respect to this filtration. Next, let M^2 be optimal for $V(\tau, 1, \operatorname{Law}(M_\tau), \nu)$. By the previous remark, we can construct such a M^2 from the starting distribution M^1_τ with a filtration $\mathcal{F}^2$ and a Brownian motion B^2 independent of $\mathcal{F}^1$, such that $dM^2 = \sigma^{(2)} dB^2$.

We can now construct a continuous martingale M on $[0,1]$ by setting it to M^1 on $[0,\tau]$ and to M^2 on $[\tau,1]$. With this construction we clearly have

$$\mathbb{E}\left[\int_0^\tau \operatorname{tr}(\sigma_r^{(1)})\,dr\right] + V(\tau, \operatorname{Law}(M_\tau^1), \nu) = \mathbb{E}\left[\int_0^\tau \operatorname{tr}(\sigma_r^{(1)})\,dr + \int_\tau^1 \operatorname{tr}(\sigma_r^{(2)})\,dr\right]. \tag{3.6}$$

Similarly, we can construct a Brownian motion $W_s = 1_{[0,\tau]}(s)B_s^1 + 1_{(\tau,1]}(s)[B_\tau^1 + B_s^2 - B_\tau^2]$ with respect to the concatenation of the filtrations $\mathcal{F}^1$ and $\mathcal{F}^2$. Where we have $dM = (1_{[0,\tau]}(s)\sigma_s^1 + 1_{(\tau,1]}(s)\sigma_s^{(2)})\,dW$. Therefore, M is a candidate martingale for the left-hand side of (3.5) and as the right-hand side of (3.6) is the cost of M, which is therefore smaller than $V(0,1,\mu,\nu)$, we have proven the desired inequality.

Since M^* was assumed optimal for $V(0,1,\mu,\nu)$. The points (1) and (2) follow immediately from (3.5). By the same construction as in the previous paragraph, we can then compose $M^*|_{[0,\tau]}$ and $M^*|_{[\tau,1]}$, which are now known to be optimal by point (2) and point (1) respectively, to obtain an optimizer M for $V(0,1,\mu,\nu)$. However, by uniqueness, this M has to coincide with M^* almost surely. For this new M, we know that M_1 only depends on M_τ^* and a Brownian motion independent of $\{M_s^* : s \leq \tau\}$. Therefore, $\operatorname{Law}(M_1 | M_s^*, s \leq \tau) = \operatorname{Law}(M_1 | M_\tau^*)$ and point (3) follows as well. □

It is an immediate corollary of point (3) of the previous lemma, that the unique optimizer M^* of (MBB) has the strong Markov property. Finally, I shall state one more proposition without proof, whose identity (3.7) characterizes M^* as a constant-speed geodesic, when the distance is measured with respect to the square of our cost functional from (3.1).

Proposition 3.4 ([8, Proposition 2.6])**.** *For M^*, the unique optimizer of* (MBB)*, we define*

$$[\mu,\nu]_t^M := \operatorname{Law}(M_t^*).$$

Then $\operatorname{Law}(M_0^*, M_t^*)$ *is an optimizer of* (WOT) *for the marginals μ and $[\mu,\nu]_t^M$. Additionally, the optimizer of* (MBB) *for these marginals is the time-changed martingale $s \in [0,1] \mapsto M_{st}^*$. Finally, for $0 \leq r \leq t \leq 1$, we have*

$$WOT([\mu,\nu]_r^M, [\mu,\nu]_t^M) = P([\mu,\nu]_r^M, [\mu,\nu]_t^M) = \sqrt{t-r}\,P(\mu,\nu) = \sqrt{t-r}\,WOT(\mu,\nu). \tag{3.7}$$

3.2 Comparison with deterministic problem

In this chapter I shall give a quick comparison between the (MBB) and its deterministic counterpart. MBB stands for Martingale-Benamou-Brenier. In fact, the equivalent formulation (3.1) can be seen as a martingale counterpart of the classical continuous-time optimal transport problem considered in Benamou-Brenier[13], which in a Lagrangian form is given by,

$$T_2(\mu,\nu) := \inf_{X_t = X_0 + \int_0^t v_s\,ds,\, X_0 \sim \mu,\, X_1 \sim \nu} \mathbb{E}\left[\int_0^1 |v_t|^2\,dt\right], \tag{BB}$$

where they only consider measures μ and ν absolutely continuous with respect to Lebesgue. We then have the following theorem.

Theorem 3.5 ([13, Proposition 1.1])**.** *For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ and absolutely continuous with respect to Lebesgue,* (BB) *has a unique optimizer X^*. In particular, $T_2(\mu,\nu) = W_2^2(\mu,\nu)$. Moreover, $X_1^* = f(X_0^*)$ where f is the gradient of a convex function $\Psi : \mathbb{R}^d \to \mathbb{R}$ and all particles move at a constant speed, therefore $X_t = X_0 + t(X_1 - X_0)$*

In the setting of this theorem a further property, which is again analogous to Proposition 3.4, can be given via McCann's [23] displacement interpolations defined as

$$[\mu,\nu]_t := \mathrm{Law}(X_t^*), \quad 0 \le t \le 1.$$

Theorem 3.6 ([30, Corollary 7.22])**.** *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ and absolutely continuous with respect to Lebesgue. Then we have for $\lambda \in [0,1]$, $0 \le r < t \le 1$*

$$\left[[\mu,\nu]_r, [\mu,\nu]_t\right]_\lambda = [\mu,\nu]_{(1-\lambda)r+\lambda t}$$

and

$$(t-r)\sqrt{T_2(\mu,\nu)} = \sqrt{T_2([\mu,\nu]_r, [\mu,\nu]_t)}.$$

Consequently, solutions of (BB) move along straight-line trajectories. In contrast, in (MBB), the optimal M^* is expected to maximize its correlation with Brownian motion, subject to the prescribed marginal constraints.

Additionally, via [17, Exercise 4.29], it can be seen that for the case of (BB) and $\Omega = C([0,1];\mathbb{R}^d)$ we have that

$$\mathrm{Law}(X^*) = \underset{\substack{\mathbb{P}\in\mathcal{P}(\Omega)\\ \mathbb{P}_0=\mu,\, \mathbb{P}_1=\nu}}{\arg\min}\; W^2(\mathbb{P}, \mathbb{P}^{cs}),$$

where W^2 represents the Wasserstein distance defined using the squared Cameron-Martin norm and $\mathbb{P}^{cs}$ is the law of a constant speed reference process.

Using the terminology of bicausal couplings, considered for example in [22], it is possible to define an adapted Wasserstein distance, see for example [6]. Let $\mathbb{W}$ denote the Wiener measure. In section 6 of [8] it is shown that for such an adapted analogue W_c^2 we have the following relation for the optimizer M^* of (MBB).

$$\mathrm{Law}(M^*) = \underset{\mathbb{Q}\in M^c(\mu,\nu)}{\arg\min}\; W_c^2(\mathbb{Q}, \mathbb{W}),$$

where $M^c(\mu,\nu)$ denotes the set of laws of continuous martingales on $[0,1]$ starting in μ and ending in ν.

3.3 Bass martingales

Definition 3.7. Let $B = \{B_t\}_{0 \le t \le 1}$ denote a Brownian motion on $\mathbb{R}^d$ with initial distribution $B_0 \sim \alpha$, where $\alpha \in \mathcal{P}(\mathbb{R}^d)$. Equivalently, $B = \tilde{B} + A$, where A is $\mathcal{F}_0$ measurable, $A \sim \alpha$ and $\tilde{B}$ is the standard Brownian motion with $\tilde{B}_0 = 0$ independent of $\mathcal{F}_0$. Let then $v : \mathbb{R}^d \to \mathbb{R}$ be a convex function, such that $\nabla v(B_1)$ is square-integrable. We then define

$$M_t := \mathbb{E}[\nabla v(B_1)|\sigma(B_s : s \le t)] = \mathbb{E}[\nabla v(B_1)|B_t], \quad 0 \le t \le 1 \tag{3.8}$$

as the Bass martingale or also called standard stretched Brownian motion from $\mu = \mathrm{Law}(M_0)$ to $\nu = \mathrm{Law}(M_1)$ with Bass measure α.

Remark 3.8. As a consequence of Brenier's theorem, there always exists a convex function F, such that for $\alpha, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ we have $\nabla F(\alpha * \gamma) = \nu$, where F is $\alpha * \gamma$ unique up to an additive constant. Here $*$ denotes the convolution operator and hence, $\alpha * \gamma$ is the law of B_1 from the definition above. Additionally, Brownian motion and geometric Brownian motion are both examples of Bass martingales.

First, it will be important to give a slightly different characterization of Bass martingales. In particular, we have the following fundamental relations:

$$(\nabla v * \gamma)(\alpha) = \mu \quad \text{and} \quad \nabla v(\alpha * \gamma) = \nu \tag{3.9}$$

The second identity is clear, since we have $\mathbb{E}[\nabla v(B_1)|B_1] = \nabla v(B_1)$. For the first identity, if we have as in the definition above, that $B_t = A + \tilde{B}$ and if we fix $A = x$, then we have

$$\mathbb{E}[\nabla v(B_1)|B_0 = x] = \mathbb{E}[\nabla v(\tilde{B}_1 + x)] = \int \nabla v(x+y)\,\gamma(dy) = (\nabla v * \gamma)(x).$$

Since $A = B_0 \sim \alpha$, we have that $M_0 \sim (\nabla v * \gamma)(\alpha)$. Recalling that γ^t is the law of a centered standard Brownian motion at time t, we can define $v_t := v * \gamma^{1-t} : \mathbb{R}^d \to \mathbb{R}$, for $0 \le t \le 1$. With this notation, we can summarize these properties in the following lemma.

Lemma 3.9 ([5, Lemma 2.3])**.** *For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$. The following are equivalent:*

1. *There is a Bass martingale M with Bass measure $\alpha \in \mathcal{P}(\mathbb{R}^d)$ from μ to ν*
2. *There is a convex function $v : \mathbb{R}^d \to \mathbb{R}$ satisfying the identities* (3.9)*.*

Additionally, the Bass martingale M can be expressed as

$$M_t = \nabla v_t(B_t), \quad 0 \le t \le 1. \tag{3.10}$$

Proof. The short verification that the conditional expectation given by (3.8) can actually be expressed by (3.10) can be found in [5] and is somewhat similar to the calculations above. For the equivalence itself, by the considerations from above, it is clear that if M is a Bass martingale from μ to ν, then the function v will satisfy the identities (3.9). For the other direction, given a Bass martingale constructed from a function v and measure α, the identities (3.9) ensure that $\mu = \text{Law}(M_0)$ and $\nu = \text{Law}(M_1)$ i.e. M is a Bass martingale from μ to ν. $\square$

3.4 The Bass solution to the Skorokhod embedding problem

The reason for the name Bass martingale is, that this construction was used by Richard F. Bass in [9] to deliver a particularly appealing solution of the Skorokhod embedding problem (SEP). Following [18], I will give a quick summary of this approach. First, the following definition will properly introduce the SEP.

Definition 3.10. Given a Brownian motion B_t and a probability measure ν on $\mathbb{R}$ with barycenter 0, a Skorokhod imbedding of ν is a stopping time τ adapted to the sigma fields of B_t such that B_τ has distribution ν.

Defining $g(z) = F_\nu^{-1}(\Phi(z))$, where F_ν is the distribution function of ν and Φ the standard normal CDF, then g is increasing and $g(B_1) \sim \nu$. The next step is to define

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

Because B_t is Markov, there exists a deterministic function h such that $X_t = h(B_t, t)$. Moreover, h is increasing in the first variable and admits an inverse H such that $B_t = H(X_t, t)$. At time $t = 1$ we have $X_1 = g(B_1) \sim \nu$ and for $t = 0$ it follows that $X_0 = \mathbb{E}[g(B_1)] = 0$. Then defining $a(x,t) := h_x(H(x,t), t)$ we have

$$dX_t = h_x(B_t, t)\, dB_t = a(X_t, t)\, dB_t.$$

Given any Brownian motion W, define the increasing additive functional Γ by

$$\Gamma_0 = 0, \quad \frac{d\Gamma_u}{du} = \frac{1}{a(W_u, \Gamma_u)^2}$$

and its inverse

$$A_t = \inf\{u : \Gamma_u > t\}.$$

Set $X_t := W_{A_t}$. For $u = A_t$ we have

$$\frac{dA_t}{dt} = \frac{1}{d\Gamma_u/du} = a(W_u, \Gamma_u)^2 = a(X_t, t)^2$$

Hence X_t satisfies

$$dX_t = \sqrt{dA_t/dt}\, d\hat{W}_t = a(X_t, t)\, d\hat{W}_t$$

for some Brownian motion $\hat{W}_t$. Therefore, X_t has the same law as the process constructed earlier from B_t. Therefore, we have $X_1 \sim \nu$ and also $W_{A_1} \sim \nu$. Hence $\tau = A_1$ is a solution of the Skorokhod embedding problem for (W, ν).

The Skorokhod embedding problem and its solution have numerous applications, particularly in mathematical finance, which lie beyond the scope of this thesis. However, it is worth noting one simple application. As already mentioned in the previous chapter, the SEP can be used to show that the continuous version of $\Pi_M(\mu, \nu)$ is not empty in the simple one-dimensional case where $\mu = \delta_{x_0}$ for some constant $x_0 \in \mathbb{R}$. Additionally, let ν be such that $\int |x| \, \nu(dx) < \infty$ and $\int x \, \nu(dx) = 0$. Then given a Brownian motion W, by Skorokhod embedding, there exists a stopping time τ such that $W_\tau \sim \nu$ and $\{W_{t\wedge\tau}\}_{t\geq 0}$ is uniformly integrable. Therefore, the process $\{M_t\}_{0\leq t\leq 1}$ defined by

$$M_t := W_{\tau \wedge \frac{t}{(1-t)}}$$

is a continuous Martingale with $M_1 \sim \nu$. Hence, this construction defines a probability measure in $\Pi_M(\mu, \nu)$. In the present setting, this application of the Skorokhod embedding is somewhat redundant, as the Bass construction already yields a continuous martingale with terminal law ν.

3.5 Duality results and characterization of solutions

From now on, I will follow [7] to introduce one of its main theorems, characterizing the link between Bass martingales and the optimizer of (MBB), justifying the terminology, standard stretched Brownian motion for the Bass martingale and stretched Brownian motion for the optimizer. First, we will need to define a property characterizing the connectedness of the marginals in a martingale sense, which we will call irreducibility, often used hereinafter as an assumption.

Definition 3.11. Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$. The pair (μ, ν) is called *irreducible* if we have that for each pair of measurable sets $A, B \subseteq \mathbb{R}^d$ with $\mu(A) > 0$, $\nu(B) > 0$, there exists a martingale $X = \{X_t\}_{0\leq t\leq 1}$ with $X_0 \sim \mu$ and $X_1 \sim \nu$ and $\mathbb{P}(X_0 \in A, X_1 \in B) > 0$.

The appendix of [7] further characterizes this irreducibility assumption. In particular, it does not matter whether we consider discrete- or continuous-time martingales in the definition. Additionally, when considering Bass martingales, it is also a necessary condition. It is shown that a Bass martingale connects any two sets with positive μ, ν measure. Hence, if there is a Bass martingale from μ to ν, they have to be irreducible. In classical optimal transport, the product measure not only gives existence of a transport plan between to measures, but it also makes such an irreducibility assumption superfluous. In fact, it guarantees that mass can be moved everywhere independent of its starting position. However, this is not the case in martingale transport, where $\mathbb{R}^d$ could be split into several convex components, which do not communicate with each other. A simple example for this would be $\mu = \frac{1}{2}(\delta_{-1} + \delta_1)$ and $\nu = \frac{1}{4}(\delta_{-2} + 2\delta_0 + \delta_2)$. With this definition, we can now state the announced theorem.

Theorem 3.12 ([7, Theorem 1.3])**.** *For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$, let the pair (μ, ν) be irreducible. Then, the following are equivalent for a martingale $M = \{M_t\}_{0 \le t \le 1}$ with $M_0 \sim \mu$ and $M_1 \sim \nu$:*

1. *M is the optimizer of* (MBB)*.*
2. *M is a Bass martingale.*

In particular, for $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$ the existence of an optimizer is given by Theorem 3.1. Hence, in this case and in light of the preceding considerations, the existence of a Bass measure from μ to ν is equivalent to irreducibility.

Theorem 3.13 ([7, Theorem 1.4])**.** *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ be with $\mu \preceq_c \nu$. Then the value $P(\mu, \nu)$ from* (3.1) *is equal to*

$$D(\mu, \nu) := \inf_{\substack{\psi \in L^1(\nu), \\ \psi \text{ convex}}} \left(\int \psi \, d\nu - \int (\psi^* * \gamma)^* \, d\mu \right). \tag{3.11}$$

Where the infimum is attained by a lower semicontinuous convex function $\psi_{opt} : \mathbb{R}^d \to (-\infty, +\infty]$ satisfying $\mu(\mathrm{ri}(\mathrm{dom}\, \psi_{opt}) = 1$ if and only if the pair (μ, ν) is irreducible. In this case the unique optimizer to (MBB) *is given by the Bass martingale constructed with the function $v = \psi^*_{opt}$ and Bass measure $\alpha = \nabla(\psi^*_{opt} * \gamma)^*(\mu)$.*

Here $\mathrm{ri}(\mathrm{dom}\, f))$ denotes the relative interior of the finiteness domain of a function f. The symbol $*$ in this case does not denote the convolution but the convex conjugate. For a function $f : \mathbb{R}^d \to (-\infty, +\infty]$ it is defined as

$$f^*(y) := \sup_{x \in \mathbb{R}^d} (\langle x, y \rangle - f(x)).$$

The proof of Theorem 3.13 and Theorem 3.12 is the subject of [7]. For the proof of Theorem 3.12, the direction (2) $\implies$ (1) is exactly given by [8, Theorem 1.10]. The other direction follows from combining [7, Theorem 6.6] and [7, Theorem 7.6]. Theorem 7.6 shows that a solution for a dual problem exists. Then, by Theorem 6.6, the existence of such a dual optimizer is equivalent to the existence of a Bass martingale, which then also solves the primal problem. Combining this with the uniqueness from 3.1 shows that any optimizer is given by this Bass martingale.

The following diagram will give a nice summary in the context of Theorem 3.13 of the relationships between the optimizer ψ_{opt} of the dual problem and the Bass measure α with associated convex function $v = \psi^*_{\text{opt}}$ of the Bass martingale $\{M_t\}_{0 \le t \le 1}$ optimizing (MBB). We will set $u := v * \gamma$ and $\varphi(x) = \nabla(v * \gamma)^*(x)$. As is also shown in [5] we have that $\nabla u = (\nabla v) * \gamma$ and for μ almost every $x \in \mathbb{R}^d$ we have the identity $(\nabla v * \gamma)^{-1} = \nabla(v * \gamma)^*$. Therefore, we have $\alpha = \nabla\varphi(\mu)$.

$$\begin{tikzcd}
\alpha * \gamma \arrow[r, shift left, leftarrow, "\nabla\psi_{\text{opt}}"] \arrow[r, shift right, "\nabla v"'] & \nu \\
\alpha \arrow[u, "*"] \arrow[r, shift left, leftarrow, "\nabla\varphi"] \arrow[r, shift right, "\nabla u"'] & \mu \arrow[u, "\pi"']
\end{tikzcd}$$

Additionally, we have that $\mathrm{Law}(M_0, M_1) = \pi$ where π is the optimizer of (WOT) and therefore

$$\pi_x = \mathrm{Law}(M_1|M_0 = x) = \nabla v(\gamma_{\nabla\varphi(x)}),$$

where $\gamma_{\nabla\varphi(x)}$ denotes the d-dimensional Gaussian measure with mean vector $\nabla\varphi(x)$ and covariance equal to I_d.

Remark 3.14. It is also worth noting, that this dual problem allows us to find the function v (unique up to a constant and a translation) independently from the primal problem, i.e. without having to find the Bass martingale or its Bass measure α. A priori we only knew that the Bass martingale is unique, but not how many different Bass measures with different associated convex functions would lead to this unique Bass martingale. Indeed, in [7] Theorem 6.6 shows, that for any such optimal Bass martingale construction, v^* will be a dual optimizer. Additionally, Lemma 7.19 shows that a lower semicontinuous convex optimizer is unique modulo adding an affine function. Taking the convex conjugate it is immediate that this translates to uniqueness of $\nabla v(x)$, modulo a translation in the x variable in the sense of $\nabla v(x - a)$. This of course coincides with the fact that one can always translate the measure α, which together with the translated v would construct the same unique martingale. In this sense, if we impose that α should have the same barycenter as μ and ν. Then both, ∇v and α are unique as well. This only works when α cannot live on a higher dimensional space, which cannot happen if the support of ν affinely spans $\mathbb{R}^d$. This is exactly Assumption 3.1 in [7] where it is also shown, that this can be done without loss of generality.

3.6 Bass functional

In this subsection I will introduce the Bass functional and the main results from [5] showing a certain equivalence between the original optimization problem and minimizing the Bass functional. For this section, this is also the paper I will use as a reference.

Definition 3.15. The *Bass functional* is defined as

$$\mathcal{P}_2(\mathbb{R}^d) \in \alpha \longmapsto \mathcal{V}(\alpha) := \mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu). \tag{3.12}$$

The first result shows that minimizing the Bass functional gives the Bass measure for the optimal Bass martingale.

Theorem 3.16 ([5, Theorem 1.5])**.** *For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$. We have*

$$P(\mu, \nu) = \inf_{\alpha \in \mathcal{P}_2(\mathbb{R}^d)} \mathcal{V}(\alpha), \tag{3.13}$$

where the right-hand side of (3.13) *is attained by an $\hat{\alpha} \in \mathcal{P}_2(\mathbb{R}^d)$ if and only if there is a Bass martingale from μ to ν with Bass measure $\hat{\alpha} \in \mathcal{P}_2(\mathbb{R}^d)$.*

The proof of the theorem will follow from the following two Lemmas.

Lemma 3.17 ([5, Lemma 3.1])**.** *In the setting of Theorem 3.16, we have the weak duality*

$$P(\mu,\nu) \leq \inf_{\alpha \in \mathcal{P}_2(\mathbb{R}^d)} \mathcal{V}(\alpha). \tag{3.14}$$

Proof. Fix any $\alpha \in \mathcal{P}_2(\mathbb{R}^d)$. Then by Brenier's theorem there is a convex function v such that $\nabla v(\alpha * \gamma) = \nu$. By the Kantorovich duality given in [[30, Particular Case 5.17]] we have

$$\mathrm{MCov}(\alpha * \gamma, \nu) = \int v \, d(\alpha * \gamma) + \int v^* \, d\nu = \int (v * \gamma) \, d(\alpha) + \int v^* \, d\nu.$$

Since $v * \gamma$ is still convex, we can apply the Kantorovich duality once more and obtain

$$\begin{aligned} \mathrm{MCov}(\alpha * \gamma, \nu) &= \int (v * \gamma) \, d(\alpha) + \int v^* \, d\nu + \int (v * \gamma)^* \, d(\mu) - \int (v * \gamma)^* \, d(\mu) \\ &\geq \mathrm{MCov}(\alpha, \mu) + \int v^* \, d\nu - \int (v * \gamma)^* \, d(\mu). \end{aligned}$$

Applying Theorem 3.13, will now yield the desired inequality.

$$\begin{aligned} \mathrm{MCov}(\alpha * \gamma, \nu) &\geq \mathrm{MCov}(\alpha, \mu) + \inf_{\psi \text{ convex}} \left(\int \psi \, d\nu - \int (\psi^* * \gamma)^* \, d(\mu) \right) \\ &= \mathrm{MCov}(\alpha, \mu) + P(\mu, \nu) \end{aligned}$$

□

The next theorem will now give the first implication from the theorem above, characterizing the Bass measure as a minimizer of the Bass functional.

Lemma 3.18 ([5, Lemma 3.2])**.** *In the setting of Theorem 3.16, if there exists a Bass martingale from μ to ν with Bass measure $\hat{\alpha} \in \mathcal{P}_2(\mathbb{R}^d)$, then $\hat{\alpha}$ is a minimizer of the Bass functional. We also have*

$$\mathcal{V}(\hat{\alpha}) = \int \mathrm{MCov}(\hat{\pi}_x, \gamma) \, \mu(dx), \tag{3.15}$$

where $\hat{\pi} \in \Pi_M(\mu,\nu)$ is the optimizer of (WOT).

Proof. By assumption there exists a Bass measure from μ to ν constructed with a Bass measure $\hat{\alpha}$ and associated convex function $\hat{v}$. Then again fixing an arbitrary $\alpha \in \mathcal{P}_2(\mathbb{R}^d)$ we need to show that $\mathcal{V}(\hat{\alpha}) \leq \mathcal{V}(\alpha)$. First, let $\hat{Z}$ be a random variable with law $\hat{\alpha}$. Then defining

$$X := (\nabla \hat{v} * \gamma)(\hat{Z}) \tag{3.16}$$

we obtain a random variable which has law μ according to (3.9). As above $\nabla \hat{v} * \gamma$ is convex and thus by Brenier's theorem the coupling $(\hat{Z}, X)$ is optimal with respect to the quadratic Wasserstein distance and as shown in (3.2) equivalently with respect to the maximal covariance. Next, let Z be a random variable with law α for which the coupling (Z, X) is optimal with respect to the maximal covariance. Since both couplings are optimal, we have

$$\mathrm{MCov}(\alpha, \mu) - \mathrm{MCov}(\hat{\alpha}, \mu) = \mathbb{E}\left[\langle Z - \hat{Z}, X\rangle\right]. \tag{3.17}$$

Now let Γ be a standard Gaussian random vector on $\mathbb{R}^d$, independent of Z and $\hat{Z}$. Then $\hat{Z} + \Gamma$ has law $\hat{\alpha} * \gamma$ and we define

$$Y := \nabla\hat{v}(\hat{Z} + \Gamma). \tag{3.18}$$

Then with the same reasoning as before, Y has law ν according to (3.9) and by Brenier's theorem the coupling $(\hat{Z} + \Gamma, Y)$ is optimal with respect to the maximal covariance. So we have

$$\mathrm{MCov}(\hat{\alpha} * \gamma, \nu) = \mathbb{E}\left[\langle \hat{Z} + \Gamma, X\rangle\right]. \tag{3.19}$$

There is no reason for the coupling $(Z + \Gamma, Y)$ to be optimal. However, since $Z + \Gamma$ has law $\alpha * \gamma$, we still have

$$\mathrm{MCov}(\alpha * \gamma, \nu) \geq \mathbb{E}\left[\langle Z + \Gamma, X\rangle\right]. \tag{3.20}$$

Combining (3.17)-(3.20) we obtain that

$$\begin{aligned}\mathcal{V}(\alpha) - \mathcal{V}(\hat{\alpha}) &= \mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu) - \mathrm{MCov}(\hat{\alpha} * \gamma, \nu) + \mathrm{MCov}(\hat{\alpha}, \mu) \\ &\geq \mathbb{E}\left[\langle Z - \hat{Z}, Y - X\rangle\right]\end{aligned}$$

For desired inequality it suffices to show that

$$\mathbb{E}\left[\langle Z - \hat{Z}, Y - X\rangle\right] = 0. \tag{3.21}$$

If we condition $Y - X$ on Z and $\hat{Z}$, then by (3.16) and (3.18) we have that

$$\mathbb{E}[Y - X | Z, \hat{Z}] = 0$$

which implies (3.21). Together with the weak duality (3.14) and recalling from Theorem 3.1, that the right-hand side of (3.15) is equal to $WOT(\mu, \nu) = P(\mu, \nu)$, we conclude the proof of Lemma 3.18. □

To finish the proof of Theorem 3.16, we can define the measures $\mu^\epsilon := \mu * \gamma^\epsilon$ and $\nu^\epsilon := \nu * \gamma^\epsilon$. Then $\mu^\epsilon \preceq_c \nu^\epsilon$ and the pair $(\mu^\epsilon, \nu^\epsilon)$ will be irreducible. This is, as we have seen in Theorem 3.12, equivalent to the existence of a Bass martingale from μ^ϵ to ν^ϵ. Therefore we can apply the previous lemma and obtain duality for the pair $(\mu^\epsilon, \nu^\epsilon)$. It then remains to apply corresponding convergence results to derive duality for the original pair (μ, ν). This, is the content of [5, Lemma 3.3]. Finally, it remains to be shown that the converse of Lemma 3.18 is also true. In other words, that if $\hat{\alpha} \in \mathcal{P}_2(\mathbb{R}^d)$ is the optimizer of the Bass functional for μ and ν, then it is also the Bass measure for a Bass martingale from μ to ν. It is clear that the associated convex function $\hat{v}$ should be chosen via Brenier's theorem such that $\nabla\hat{v}(\hat{\alpha} * \gamma) = \nu$. It then remains to show that the first relation in (3.9), namely that $(\nabla\hat{v} * \gamma)(\hat{\alpha}) = \mu$, corresponds to the minimality of $\hat{\alpha} \in \mathcal{P}_2(\mathbb{R}^d)$ on the right-hand side of (3.13), this is the content of the proof of [5, Lemma 3.4].

The following argument provides another proof of the duality (3.13). It can be viewed as an application of Toland's duality [28].

Alternative proof [3, Proof of Lemma 2.3].

$$\begin{aligned}
\inf_\alpha \mathcal{V}(\alpha) &= \inf_\alpha \left\{ \inf_\psi \left[\int \psi^* \, d(\alpha * \gamma) + \int \psi \, d\nu \right] - \sup_{A\sim\alpha,\, X\sim\mu} \mathbb{E}[A \cdot X] \right\} \\
&= \inf_\psi \left\{ \inf_{A,\, X\sim\mu} \mathbb{E}[(\psi^* * \gamma)(A) - A \cdot X] + \int \psi \, d\nu \right\} \\
&= \inf_\psi \left\{ \int \psi \, d\nu - \sup_{A,\, X\sim\mu} \mathbb{E}[A \cdot X - (\psi^* * \gamma)(A)] \right\} \\
&= \inf_\psi \left\{ \int \psi \, d\nu - \int (\psi^* * \gamma)^* \, d\mu \right\} = P(\mu, \nu)
\end{aligned}$$

For the first equality, we used the Kantorovich duality. For the second line, we used $\int \psi^* \, d(\alpha * \gamma) = \int (\psi^* * \gamma) \, d\alpha$ and observed that minimizing over α and then over $A \sim \alpha$, is equivalent to minimizing over all couplings (A, X) with $X \sim \mu$. To get to the penultimate equality we used that by definition $\sup_{a\in R^d}(\langle a, x\rangle - (\psi^* * \gamma)(a)) = (\psi^* * \gamma)^*(x)$. The last equality follows from Theorem 3.13. $\square$

3.7 Gaussian case

In this subsection, I will give a general instruction on how to construct the Bass martingale under the assumption that both μ and ν are Gaussian.

Proposition 3.19. *Let $\mu = N(0, \Sigma_\mu)$ and $\nu = N(0, \Sigma_\nu)$ be Gaussian measures on $\mathbb{R}^d$ and in convex order. Then if $A = (\Sigma_\nu - \Sigma_\mu)^{1/2}$ is invertible, there is a Bass martingale from μ to ν with Gaussian Bass measure $\alpha = N(0, A^{-1}\Sigma_\mu A^{-1})$ and linear $\nabla v(x) = Ax$.*

Proof. First, $A^2 = (\Sigma_\nu - \Sigma_\mu)$ is positive semidefinite. In fact, since we have that for any $U \in \mathbb{R}^d$ the function $\varphi(x) := (U^\top x)^2$ is convex, together with $\mathbb{E}_\mu[X] = \mathbb{E}_\mu[X] = 0$, we have that

$$U^\top \Sigma_\mu U = \mathrm{Var}_\mu(U^\top X) + U^\top \mathbb{E}_\mu[X]^2 = \mathbb{E}_\mu[(U^\top X)^2] \leq \mathbb{E}_\nu[(U^\top X)^2] = U^\top \Sigma_\nu U.$$

Which implies, that $\Sigma_\nu - \Sigma_\mu$ is positive semidefinite. This, together with the symmetry of A^2, guarantees that A is well defined and symmetric. Therefore, we can define $\Sigma_\alpha = A^{-1}\Sigma_\mu A^{-1}$ and $\alpha = N(0, \Sigma_\alpha)$. It remains to verify the identities (3.9). First, we have

$$(\nabla v * \gamma)(x) = \int \nabla v(x + y) \, \gamma(dy) = Ax + A \int y \, \gamma(dy) = Ax$$

and therefore

$$(\nabla v * \gamma)(\alpha) = A(\alpha) = N(0, A\Sigma_\alpha A) = N(0, \Sigma_\mu) = \mu.$$

For the second identity,

$$\begin{aligned}\nabla v(\alpha * \gamma) &= A(N(0, \Sigma_\alpha + I)) = N(0, A(\Sigma_\alpha + I)A) = N(0, A\Sigma_\alpha A + A^2) \\ &= N(0, \Sigma_\mu + (\Sigma_\nu - \Sigma_\mu)) = N(0, \Sigma_\nu) = \nu.\end{aligned}$$

□

In fact, in the proof we only needed that $A\Sigma_\alpha A = \Sigma_\mu$. Therefore, we could have worked under the less restrictive assumption that $\mathrm{Ran}(\Sigma_\mu) \subseteq \mathrm{Ran}(A)$ and then used a pseudoinverse $A^\dagger$.

In the one-dimensional Gaussian setting, restricting to Gaussian competitors allows one to verify directly that the Bass measure minimizes the Bass functional. In this setting the following equivalence comes in handy:

Lemma 3.20. *For two Gaussian measures $\mu = N(0, \Sigma_\mu)$ and $\nu = N(0, \Sigma_\nu)$ we have that $\mu \preceq_c \nu$ if and only if $\Sigma_\mu \preceq \Sigma_\nu$*

Proof. The ($\Longrightarrow$) direction was shown in the proof of the previous proposition. For the reverse direction, note that if $\Sigma_\nu - \Sigma_\mu \succeq 0$, then for $X \sim \mu$, there exists a Gaussian random variable $Z \sim N(0, \Sigma_\nu - \Sigma_\mu)$ independent of X. Therefore, the random variable $Y := X + Z$ satisfies $Y \sim \nu$ and for any convex function φ we have

$$\mathbb{E}[\varphi(Y)] = \mathbb{E}[\mathbb{E}[\varphi(X + Z)|X]] \geq \mathbb{E}[\varphi(X + \mathbb{E}[Y])] = \mathbb{E}[\varphi(X)] \tag{3.22}$$

by the tower property and Jensen's inequality. This shows the reverse implication. □

Now let $\mu = N(0, \sigma_\mu^2)$, $\nu = N(0, \sigma_\nu^2)$ with $\sigma_\mu^2 < \sigma_\nu^2$ and therefore $\mu \preceq_c \nu$. If we denote a candidate Gaussian Bass measure via $\beta = N(0, b)$, then $\beta * \gamma \sim N(0, b+1)$. Moreover, for $p_x = N(0, x)$, $x \in \mathbb{R}$ we have

$$\mathrm{MCov}(p_a, p_b) = \sqrt{ab}.$$

Indeed, by Cauchy-Schwarz we have for any coupling

$$\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]} = \sqrt{ab}$$

and therefore $\mathrm{MCov}(p_a, p_b) \leq \sqrt{ab}$. Where equality is attained for the comonotone coupling $Y = \sqrt{\frac{b}{a}}X$. Therefore, we have

$$\mathcal{V}(\beta) = \mathrm{MCov}(\beta * \gamma, \nu) - \mathrm{MCov}(\beta, \mu) = \sigma_\nu\sqrt{b+1} - \sigma_\mu\sqrt{b}.$$

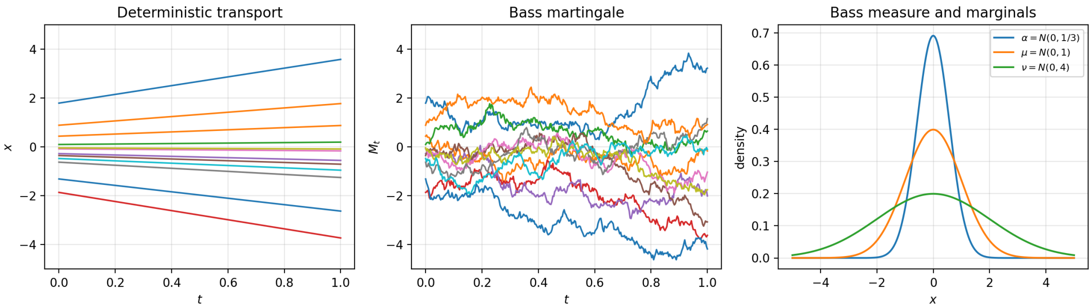

Figure 1: Illustration of the one-dimensional Gaussian case with $\mu = N(0,1)$ and $\nu = N(0,4)$. Left: deterministic transport sends x to $2x$ along straight lines. Middle: the Bass martingale follows paths $M_t = \sqrt{3}(X + B_t)$ with $X \sim N(0,1/3)$ and B_t a standard Brownian motion. Right: the Bass measure $\alpha = N(0,1/3)$ together with the prescribed marginals.

A quick calculation shows that the minimum is attained for

$$b = \frac{\sigma_\mu^2}{\sigma_\nu^2 - \sigma_\mu^2}.$$

Had we taken the approach from the previous proposition then. $A = (\Sigma_\nu - \Sigma_\nu)^{1/2} = (\sigma_\nu^2 - \sigma_\mu^2)^{1/2}$ and therefore

$$\Sigma_\alpha = A^{-1}\Sigma_\mu A^{-1} = \frac{\sigma_\mu^2}{\sigma_\nu^2 - \sigma_\mu^2}.$$

as well. In particular, a simulation shown in Figure 1 for $\sigma_\mu^2 = 1$ and $\sigma_\nu^2 = 4$ and therefore with $A = \sqrt{3}$ and $\Sigma_\alpha = 1/3$ illustrates the statements made in Section 3.2.

Example 3.21. A second example shall show that a Gaussian α can be the Bass measure of a Bass martingale which connects two very different, non-Gaussian measures. Therefore, let

$$\mu = \mathrm{Unif}(0,1), \quad \nu = \frac{1}{2}(\delta_0 + \delta_1).$$

At first, it might be a bit surprising that $\mu \preceq_c \nu$. Apart from the following existence of a Bass martingale, this can be seen by taking a Brownian motion with $B_0 \sim \mu$ and then stopping it at the boundary with $\tau = \inf\{s \geq 0 : B_s \in \{0,1\}\}$. After a time transformation, this yields a martingale on $[0,1]$ from μ to ν, which after applying the usual argument with Jensen's inequality shows that $\mu \preceq_c \nu$. Now let

$$v = (x)_+ = x\chi_{\{x>0\}}, \quad \alpha = N(0,1),$$

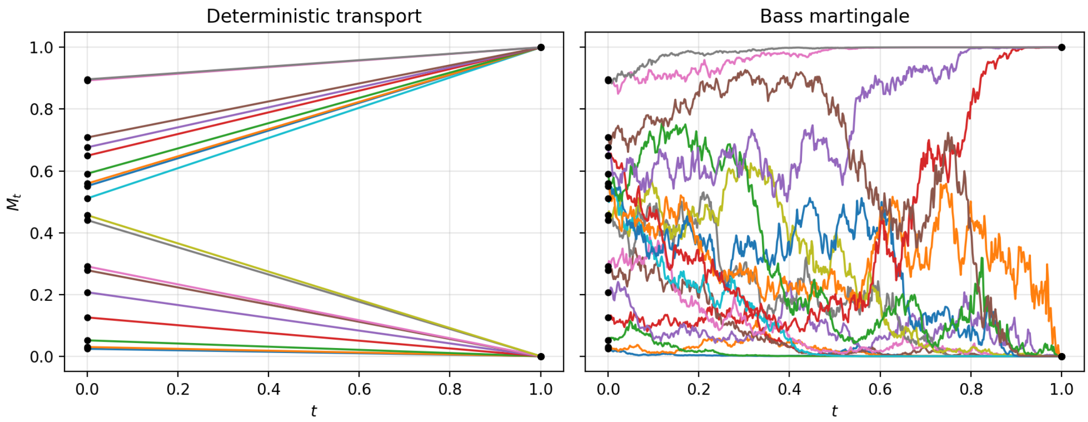

Figure 2: Transport from $\mu = \mathrm{Unif}(0,1)$ to $\nu = \frac{1}{2}(\delta_0 + \delta_1)$. Left: the deterministic transport associated with $v(x) = (x - \frac{1}{2})_+$. Right: the Bass martingale $M_t = \Phi(B_t/\sqrt{1-t})$ with terminal value $M_1 = \chi_{\{B_1>0\}}$. . Both simulations use the same initial positions $M_0 \sim \mathrm{Unif}(0,1)$.

where χ_A denotes the indicator function of a set A. Then the Bass martingale M_t constructed with α and v connects μ and ν. Indeed, the Bass martingale is

$$M_t = \mathbb{E}[\nabla v(B_1)|B_t] = \mathbb{P}(B_1 > 0|B_t) = \Phi\left(\frac{B_t}{\sqrt{1-t}}\right), \quad 0 \leq t < 1$$

where we used that for a Brownian motion $B_1|B_t \sim N(B_t, 1-t)$, $0 \leq t < 1$ and Φ again denotes the CDF of the standard Gaussian. Hence,

$$M_0 = \Phi(B_0) \sim \mathrm{Unif}(0,1) = \mu$$

since $B_0 \sim N(0,1)$ and Φ is therefore exactly its CDF. The identity follows from the fact that for a continuous random variable X and its CDF F we always have $F(X) \sim \mathrm{Unif}(0,1)$. For the second marginal we have $M_1 = \chi_{\{B_1>0\}}$. Since $B_1 \sim N(0,2)$ is centered and symmetric, we have $\mathbb{P}(B_1 > 0) = \frac{1}{2}$, which shows

$$M_1 \sim \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1 = \nu.$$

This example is also illustrated in Figure 2.

More generally, if α is given and $v : \mathbb{R}^d \to \mathbb{R}^d$ is convex, then α is the Bass measure for $\mu := (\nabla v * \gamma)(\alpha)$ and $\nu := \nabla v(\alpha * \gamma)$. Hence, it is associated to infinitely many Bass martingales.

3.8 Properties of the Bass measure

In the previous subsection, we established that the fact that the Bass measure is Gaussian, provided both μ and ν are Gaussian, was very helpful in determining the Bass measure. In this subsection, we shall establish that this is also true for rotational invariance. However, we shall see that this is not necessarily true for other properties of μ and ν.

Proposition 3.22. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ be in convex order, the pair (μ, ν) irreducible and* $\operatorname{supp}(\nu)$ *affinely spans $\mathbb{R}^d$. If both μ and ν are invariant under the pushforward by some $g \in O(d)$, where $O(d)$ denotes the group of orthogonal matrices, then the unique minimizer of the Bass functional with barycenter* 0 *is also invariant under the pushforward by this g.*

Proof. Let $\hat{\alpha}$ be this minimizer of the Bass functional with barycenter 0 and let

$$\alpha_g := g(\hat{\alpha}), \quad v_g(x) = v(g^{-1}(x)),$$

where v_g is also convex because g^{-1} is linear. The goal is to show that the Bass martingale with Bass measure α_g and associated convex function v_g has marginals $g(\mu) = \mu$ and $g(\nu) = \nu$. By uniqueness, it would then follow that this new Bass martingale is equal to the Bass martingale constructed with $\hat{\alpha}$. This new α_g also has barycenter 0, as for $A \sim \hat{\alpha}$ we have $\mathbb{E}[gA] = g\,\mathbb{E}[A] = 0$. Since we have uniqueness of the Bass measure up to a translation of the barycenter (see Remark 3.14), this would show that $g(\hat{\alpha}) = \hat{\alpha}$. Therefore, it remains to show the relations (3.9). We have $\nabla v_g(x) = (g^{-1})^\top \nabla v(g^{-1}(x)) = g\nabla v(g^{-1}(x))$. For the first relation,

$$\begin{aligned}(\nabla v_g * \gamma)(a) &= \mathbb{E}[\nabla v_g(a + \Gamma)] = \mathbb{E}[g\nabla v(g^{-1}(a + \Gamma))] = g\,\mathbb{E}[\nabla v(g^{-1}a + \tilde{\Gamma})] \\ &= g((\nabla v * \gamma)(g^{-1}(a))),\end{aligned}$$

where we used that $\tilde{\Gamma} = g^{-1}(\Gamma) \sim N(g0, g^{-1}Ig) = N(0, I)$ is also a standard Gaussian. Plugging in α_g this yields

$$(\nabla v_g * \gamma)(\alpha_g) = g((\nabla v * \gamma)(g^{-1}(g(\hat{\alpha})))) = g((\nabla v * \gamma)(\hat{\alpha})) = g(\mu).$$

Thus proving the first relation. For the second relation we have

$$\nabla v_g(\alpha_g * \gamma) = g(\nabla v(g^{-1}(g(\alpha) * \gamma))) = g(\nabla v(\alpha * \gamma)) = g(\nu),$$

where we again used that $g^{-1}((\alpha) * \gamma) = (g^{-1}(\alpha) * \gamma)$, i.e. that a standard Gaussian is invariant under the pushforward by an orthogonal map. ☐

This shows that radial symmetries of μ and ν transfer to the optimal $\hat{\alpha}$. In particular if μ and ν is radially symmetric, then so is $\hat{\alpha}$. Additionally, in this case we have $\nabla v(x) = g\nabla v(g^{-1}(x))$, $\forall g \in O(d)$, which means that the map ∇v is equivariant.

The next invariance property we will have a look at are copulas. A precise definition and proofs of the following properties can be found in [25]. By Sklar's theorem a copula is a function C, which satisfies the following relation

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)), \tag{3.23}$$

where F is the CDF of a d-dimensional random variable X and F_i are the CDFs of its marginals X_i. In particular, if the F_i are continuous, then this copula is unique. A natural construction of such a C can be obtained via

$$C(u_1, ..., u_d) = \mathbb{P}(U_1 \leq u_1, ..., U_d \leq u_d),$$

where $U_i := \mathcal{F}_i(X_i, V_i)$ with $\{V_i\}_{1\leq i\leq d}$ independent Unif$(0,1)$ distributed r.v.'s independent of X and $\mathcal{F}_i$ defined via

$$\mathcal{F}_i(x, a) := \mathbb{P}(X_i < x) + a\mathbb{P}(X_i = x).$$

In this case $U_i \sim$ Unif$(0,1)$ and $X_i = F_i^{-1} \circ U_i$, $\mathbb{P}$-almost surely. Furthermore, two probability measures μ, ν have the same dependence structure if there exist random variables $X \sim \mu, Y \sim \nu$ such that, using the same choice of $\{V_i\}$ in the construction above, the associated copulas satisfy $C_X = C_Y$. We then have the following result for the classical optimal transport problem.

Theorem 3.23 ([16, Theorem 2.9])**.** *For $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ and $X \sim \mu$ the following are equivalent:*

1. *μ and ν have the same dependence structure.*
2. *$(X, \tilde{Y})$ is an optimal coupling, where $\tilde{Y}_i = G_i^{-1} \circ \mathcal{F}_i(X_i, V_i)$. With G_i^{-1} the generalized inverse of the CDF of the i-th marginal of ν and $\mathcal{F}_i(X_i, V_i)$ defined as above.*

It is clear that such a property, in general, cannot hold in the martingale optimal transport problem. In particular, if X is continuous, then Y is given as a function of X, which would only satisfy the martingale property if $X = Y$. However, it is true that if X and Y have the same independent copula, then $A \sim \alpha$, the optimal Bass measure, also has the independent copula.

Proposition 3.24. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ be in convex order and let the pair (μ, ν) be irreducible. If $X \sim \mu, Y \sim \nu$ both have the independent copula $C(u) = \prod_{i=1}^d u_i$, then the Bass measure α of the optimal Bass martingale from μ to ν is a product measure, i.e. $A \sim \alpha$ also has the independent copula.*

Proof. First, there is an equivalence between X having the independent copula and Law$(X) = \mu$ being a product measure. Indeed, if Law(X) is a product measure, then

$$F(x_1, \dots, x_d) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d) = \prod_{i=1}^d \mathbb{P}(X_i \leq x_i) = \prod_{i=1}^d F_i(x_i).$$

Thus, the independent copula satisfies (3.23). For the converse,

$$F(x_1, \dots, x_d) = C(F_1(x_1), ..., F_d(x_d)) = \prod_{i=1}^d F_i(x_i),$$

showing this simple equivalence. Therefore, $\mu = \bigotimes_{i=1}^d \mu_i$ and $\nu = \bigotimes_{i=1}^d \nu_i$. Furthermore, it is not hard to see that for the marginals, we have $\mu_i \preceq_c \nu_i$ and the pairs (μ_i, ν_i) are irreducible $i = 1, \dots, d$. Therefore, by Theorem 3.12 for each marginal there exists a Bass martingale from μ_i to ν_i with its Bass measure and associated convex function denoted by α_i and v_i. Let

$$\alpha = \bigotimes_{i=1}^{d} \alpha_i, \quad v(x) = \sum_{i=1}^{d} v_i(x_i),$$

where v is convex as a sum of convex functions and $\nabla v(x) = (v_1'(x_1), \dots, v_n'(x_n))$. It suffices to show relations (3.9). For the first relation, let $A \sim \alpha, \Gamma \sim \gamma^d, X \sim (\nabla v * \gamma^d)(\alpha)$

$$\begin{aligned}X = (X_1, \dots, X_d) &= (\mathbb{E}[\nabla v_1(A_1 + \Gamma_1)|A], \dots, (\mathbb{E}[\nabla v_d(A_d + \Gamma_d)|A]) \\ &= (\mathbb{E}[\nabla v_1(A_1 + \Gamma_1)|A_1], \dots, (\mathbb{E}[\nabla v_d(A_d + \Gamma_d)|A_d]) \sim \bigotimes_{i=1}^{d} \mu_i = \mu.\end{aligned}$$

Where we used that the coordinates of A are independent and that coordinate-wise the relation already holds. For the second relation

$$\nabla v(\alpha * \gamma^d) = \bigotimes_{i=1}^{d} v_i'(\alpha_i * \gamma) = \bigotimes_{i=1}^{d} \nu_i = \nu.$$

Where we used that $(\bigotimes_{i=1}^d \alpha_i) * \gamma^d = \bigotimes_{i=1}^d (\alpha_i * \gamma)$, where γ and γ^d denote the standard Gaussian in $\mathbb{R}$ and $\mathbb{R}^d$ respectively. $\square$

However, this property cannot hold for other copulas in general, as the following example will show. First, we need the following proposition, characterizing the copula for a Gaussian in $\mathbb{R}^2$ in terms of its correlation.

Lemma 3.25. *Let X, Y be Gaussian random variables in $\mathbb{R}^2$. Then they have the same copula if and only if they have the same correlation.*

Proof. For

$$(X_1, X_2) \sim N\left(\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right),$$

we set

$$(Z_1, Z_2) = \left(\frac{X_1 - m_1}{\sigma_1}, \frac{X_2 - m_2}{\sigma_2}\right) \sim N\left((0,0), \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Then for F_{X_i} the CDF of the marginals of X and Φ again the CDF of a standard normal Gaussian in $\mathbb{R}$, we have $F_{X_i}(X_i) = \Phi(Z_i)$ $i = 1, 2$. In particular, the copula C_X of $X = (X_1, X_2)$ is given by

$$\begin{aligned}C_X(u, v) &= \mathbb{P}(F_{X_1}(X_1) \le u, F_{X_2}(X_2) \le v) = \mathbb{P}(\Phi(Z_1) \le u, \Phi(Z_2) \le v) \\ &= \mathbb{P}(Z_1 \le \Phi^{-1}(u), Z_2 \le \Phi^{-1}(v)) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)),\end{aligned}$$

where we used the fact that Φ is strictly increasing and Φ_ρ denotes the CDF of (Z_1, Z_2). With this construction the copula depends only on ρ. Therefore, if the laws of X and X have the same correlation, then they will also have the same copula. This already shows the first implication. For the converse, let X and Y have the same copula, then by the construction above, we have

$$\Phi_{\rho_X}(0,0) = C_X\left(\frac{1}{2}, \frac{1}{2}\right) = C_Y\left(\frac{1}{2}, \frac{1}{2}\right) = \Phi_{\rho_Y}(0,0).$$

Intuitively, it is clear that $\Phi_\rho(0,0)$ is strictly increasing in ρ and therefore $\rho_X = \rho_Y$ by the above equality. This intuition is also confirmed by the calculations in [21] (46.47) which state that

$$\Phi_\rho(0,0) = \frac{1}{4} + \frac{1}{2\pi}\arcsin(\rho).$$

This is clearly strictly increasing in ρ and therefore concludes the proof. $\square$

This lemma simplifies the following example, where we will only check if these Gaussians have the same correlation.

Example 3.26. By using the notation from the Gaussian case subsection, we know that we must have $\Sigma_\nu = \Sigma_\mu + A^2$. Choosing

$$A^2 = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix},$$

we can then choose

$$\Sigma_\mu = \begin{pmatrix} \frac{2}{7} & \frac{1}{4} \\ \frac{1}{4} & 1 \end{pmatrix} \quad \Sigma_\nu = \begin{pmatrix} \frac{9}{7} & \frac{3}{4} \\ \frac{3}{4} & 2 \end{pmatrix}.$$

A quick calculation shows that all of these matrices are positive definite and in fact were chosen such that μ and ν have the same correlation $\rho_\mu = \rho_\nu = \frac{\sqrt{14}}{8}$. Further,

$$\Sigma_\alpha = A^{-1}\Sigma_\mu A^{-1} = \begin{pmatrix} \frac{29}{42} - \frac{5\sqrt{3}}{21} & -\frac{2}{21} \\ -\frac{2}{21} & \frac{29}{42} + \frac{5\sqrt{3}}{21} \end{pmatrix}.$$

This is still positive definite, however $\rho_\alpha = -\frac{4}{\sqrt{541}}$. Therefore, we have found a counter example where in fact μ and ν have the same copula, but the optimal Bass measure α has a different copula.

4 Convexity

A further property, which was noticeable in the explicit Gaussian calculations for the previous chapter, was that in fact $\mathcal{V}(\beta)$ was convex with respect to the b variable on some interval around $b = 1/3$. In fact, if we would have parametrized the curve $\beta_t = N(0, b_t)$ via $b_t = ((1-t)\sqrt{a} + t\sqrt{b})^2$, then β_t would have been a Wasserstein geodesic from $p_a = N(0,a)$ to $p_b = N(0,b)$ for $t \in [0,1]$. Indeed b_t is the variance of $X_t = (1-t)X + tT(X)$, where $T(x) = \sqrt{\frac{b}{a}}x$ is the optimal Brenier map from p_a to p_b. Therefore, the fact, that this is a geodesic follows from Theorem 3.5 and Theorem 3.6. Taking for example, $a = 1/4$ and $b = 1$, then $b_t = \frac{(1+t)^2}{4}$. Plugging in, the function $V(\beta_t) = 2\sqrt{\frac{(1+t)^2}{4} + 1} - \sqrt{\frac{(1+t)^2}{4}}$ is clearly convex for $t \in [0,1]$ and attains its minimum at $\hat{t} = \frac{2}{\sqrt{3}} - 1$, where we again have that $b_{\hat{t}} = \frac{1}{3}$. Therefore, the Bass functional is convex along the entire Wasserstein geodesic. In fact, this was not due to a special choice of the geodesic or the restriction to Gaussian measures, but holds in general in one dimension, as the next subsection will show.

4.1 Displacement convexity

Although, the last example made clear, what is meant by displacement convexity, the following definition shall formalize it.

Definition 4.1 (Displacement convexity)**.** Let $\phi : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$. We say that ϕ is displacement convex in $\mathcal{P}_2(\mathbb{R}^d)$ if for every couple $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ there exists an optimal transport plan $\mu \in \Pi_o(\mu_1, \mu_2)$ with respect to quadratic cost such that

$$\phi(\mu_t^{1\to 2}) \leq (1-t)\phi(\mu_1) + t\phi(\mu_2) \quad \forall t \in [0,1],$$

where $\mu_t^{1\to 2} := ((1-t)p^1 + tp^2)(\mu)$ i.e. $\mathrm{Law}(X_t)$ as defined in Theorem 3.5. p^1, p^2 being the projections onto the first and the second variable.

This definition is actually equivalent to the (apparently) stronger statement

$$\text{the map } t \in [0,1] \mapsto \phi(\mu_t^{1\to 2}) \text{ is convex.} \tag{4.1}$$

This can be seen from Theorem 3.6, due to which we have that for $0 < t_1 < t_2 < 1$ the plan $(\pi_{t_1}^{1\to 2} \times \pi_{t_2}^{1\to 2})(\mu)$ is the unique element of $\Pi_o(\mu_{t_1}^{1\to 2}, \mu_{t_2}^{1\to 2})$.

In this chapter it will be important to consider the Bass functional $\mathcal{V}(\cdot)$ with respect to the quadratic Wasserstein distance $\mathcal{W}_2(\cdot,\cdot)$. Indeed, by the relation (3.2) we have

$$\mathcal{V}(\alpha) = \mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu) = \frac{1}{2}\mathcal{W}_2^2(\alpha, \mu) - \frac{1}{2}\mathcal{W}_2^2(\alpha * \gamma, \nu) + C \tag{4.2}$$

where

$$C = \frac{d}{2} + \frac{1}{2}\int |y|^2 \, d\nu(y) - \frac{1}{2}\int |x|^2 \, d\mu(x) \tag{4.3}$$

does not depend on α. Therefore, the functional

$$\mathcal{U}(\alpha) := \mathcal{W}_2^2(\alpha, \mu) - \mathcal{W}_2^2(\alpha * \gamma, \nu) \tag{4.4}$$

is equal (up to a constant) to the Bass functional. In particular, it is equivalent to look at displacement convexity of $\mathcal{U}(\cdot)$ instead of the original $\mathcal{V}(\cdot)$.

Further, it is also interesting to notice, that in the following we will not assume that $\mu \preceq_c \nu$. However, it is important that the theorem restricts itself to the one-dimensional case.

Proposition 4.2 ([5, Proposition 5.1])**.** *Suppose d=1. Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$. The Bass functional*

$$\mathcal{P}_2(\mathbb{R}) \in \alpha \longmapsto \mathcal{V}(\alpha) = \mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu) \tag{4.5}$$

is displacement convex. In particular, if ν is not a Dirac measure and the geodesic $(\alpha_t)_{0\leq t\leq 1}$ is such that α_1 is not a translate of α_0, the function $t \mapsto \mathcal{V}(\alpha_t)$ is strictly convex.

Proof. We fix $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$ and a quadratic Wasserstein geodesic $(\alpha_t)_{0\leq t\leq 1}$ in $\mathcal{P}_2(\mathbb{R}^d)$. Since we are in $d = 1$ we can choose random variables Z_0, Z_1 and X with laws α_0, α_1 and μ such that they are mutually comonotone. For example by taking $X = F_\mu^{-1}(U)$, $Z_0 = F_{\alpha_0}^{-1}(U)$, $Z_1 = F_{\alpha_1}^{-1}(U)$, where $U \sim \mathrm{Unif}(0,1)$ and F denotes the CDF. In particular, the interpolation $Z_t := (1-t)Z_0 + tZ_1$, $t \in [0,1]$ has law α_t and is also comonotone with X. Therefore, the coupling (Z_t, X) of (α_t, μ) is optimal with respect to the quadratic Wasserstein distance, which means

$$\mathcal{W}_2^2(\alpha_t, \mu) = \mathbb{E}|Z_t - X|^2 \quad t \in [0,1]. \tag{4.6}$$

Next we take a $\Gamma \sim \gamma$ independent of Z_0, Z_1. Fixing $t \in [0,1]$ there exists a $Y_t \sim \nu$ such that $(Z_t + \Gamma, Y_t)$ is also an optimal coupling of $(\alpha_t * \gamma, \nu)$ with respect to the quadratic Wasserstein distance. So that again

$$\mathcal{W}_2^2(\alpha_t * \gamma, \nu) = \mathbb{E}|Z_t + \Gamma - Y_t|^2 .$$

Further, $(Z_i+\Gamma, Y_t)$ $i = 0,1$ is also going to be an admissible coupling of $(\alpha_i*\gamma, \nu)$ $i = 0,1$. However, it is in general sub-optimal, meaning we only have

$$\mathcal{W}_2^2((\alpha_i * \gamma, \nu)) \leq \mathbb{E}|Z_i + \Gamma - Y_t|^2 \quad i = 0,1. \tag{4.7}$$

Combining these statements we have

$$\begin{aligned}(1-t)\mathcal{U}(\alpha_0) + t\mathcal{U}(\alpha_1) \geq &(1-t)\,\mathbb{E}|Z_0 - X|^2 + t\,\mathbb{E}|Z_1 - X|^2 \\ &-(1-t)\,\mathbb{E}|Z_0 + \Gamma - Y_t|^2 - t\,\mathbb{E}|Z_1 + \Gamma - Y_t|^2 .\end{aligned}$$

Using the identity

$$(1-t)|a|^2 + t|b|^2 = |(1-t)a + tb|^2 + t(1-t)|a-b|^2$$

twice. Once with $a = Z_0 - X,\ b = Z_1 - X$ and once with $a = Z_0 + \Gamma - Y_t$, $b = Z_1 + \Gamma - Y_t$, the two quadratic terms cancel, since in both cases $a - b = Z_0 - Z_1$. Hence, we obtain

$$\begin{aligned}(1-t)\mathcal{U}(\alpha_0) + t\mathcal{U}(\alpha_1) &\geq \mathbb{E}|Z_t - X|^2 - \mathbb{E}|Z_t + \Gamma - Y_t|^2 \\ &= \mathcal{W}_2^2(\alpha_t, \mu) - \mathcal{W}_2^2(\alpha_t * \gamma, \nu) = \mathcal{U}(\alpha_t).\end{aligned}$$

For the equality case, notice that the only inequality came from (4.7). Therefore, assume that we had equality, meaning that $(Z_i + \Gamma, Y_t)\, i = 0, 1$ were also optimal couplings, which since we are in one dimension is equivalent to them being comonotone. Since we assumed $Y_t \sim \nu$ to be non-constant. There exists $y \in \mathbb{R}$ such that $\mathbb{P}[Y_t < y] \in (0, 1)$. Since the law of $(Z_i + \Gamma, Y_t)\, i = 0, 1$ is continuous we can find $z_0, z_1 \in \mathbb{R}$ such that

$$\{Z_0 + \Gamma < z_0\} = \{Y_t < y\} = \{Z_1 + \Gamma < z_1\}.$$

Conditioning on $\Gamma = \zeta$, we obtain for Lebesgue-a.e. ζ

$$\{Z_0 < z_0 - \zeta\} = \{Z_1 < z_1 - \zeta\}.$$

This implies that Z_0 and Z_1 are translates, which contradicts our assumption and we obtain strict convexity. □

The following example will show, that $\mathcal{W}_2^2(\cdot, \mu^1)$ is in general not convex along geodesics in $\mathbb{R}^d$ for $d \geq 2$.

Example 4.3 ([2, Example 9.1.5])**.** Let

$$\mu^1 := \frac{1}{2}(\delta_{(0,0)} + \delta_{(0,-2)})$$

and

$$\mu^2 := \frac{1}{2}(\delta_{(0,0)} + \delta_{(2,1)}), \quad \mu^3 := \frac{1}{2}(\delta_{(0,0)} + \delta_{(-2,1)}).$$

In the discrete case, where all Dirac masses have the same weight, the optimal plan is induced by a transport map. Therefore, it is not hard to check that among the two options the optimal transport from μ^2 to μ^3 maps $(0, 0)$ to $(-2, 1)$ and $(2, 1)$ to $(0, 0)$. Therefore the unique constant speed geodesic between the two measures is given by

$$\mu_t^{2\to 3} = \frac{1}{2}(\delta_{(-2t,t)} + \delta_{(2-2t,1-t)} \quad t \in [0, 1].$$

In order to calculate $\mathcal{W}_2^2(\mu_t^{2\to 3}, \mu^1)$ we again only have to consider the two plans pushing μ^1 to $\mu_t^{2\to 3}$ which are given by

$$\begin{aligned} f_t(0, 0) &= (-2t, t), & f_t(0, -2) &= (2 - 2t, 1 - t), \\ g_t(0, 0) &= (2 - 2t, 1 - t), & g_t(0, -2) &= (-2t, t). \end{aligned}$$

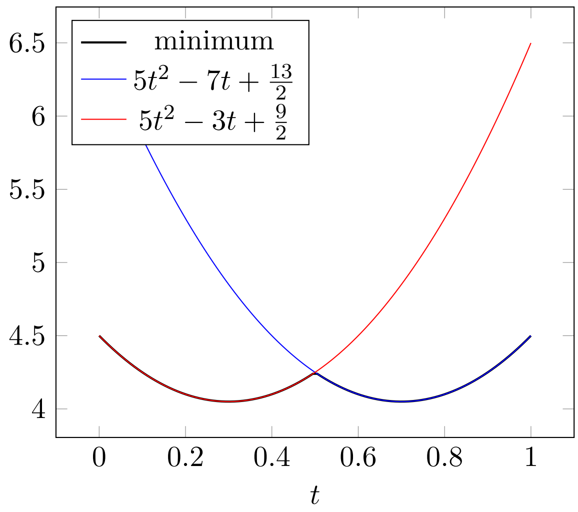

Figure 3: Graph of $\mathcal{W}_2^2(\mu_t^{2\to 3}, \mu^1)$

Hence

$$\begin{aligned}\mathcal{W}_2^2(\mu_t^{2\to 3}, \mu^1) &= \min\left\{\frac{1}{2}(|f_t(0,0)|^2 + |f_t(0,-2)|^2), \frac{1}{2}(|g_t(0,0)|^2 + |g_t(0,-2)|^2)\right\} \\ &= \min\left\{5t^2 - 7t + \frac{13}{2}, 5t^2 - 3t + \frac{9}{2}\right\}.\end{aligned}$$

This a pointwise minimum of convex functions, which is not convex itself, since it has a concave cusp at $t = \frac{1}{2}$ as can be seen from Figure 3.

The next example will show that at least under somewhat degenerate assumptions, the Bass functional is not convex.

Example 4.4. In the setting of the previous Example (4.3), let $\mu = \mu^1$, and $\alpha_t = \mu_t^{2\to 3}$ be the geodesic between μ^2 and μ^3, then the previous example showed that $\mathcal{W}_2^2(\mu, \alpha_t)$ is not convex. Choosing $\nu = \delta_{(0,0)}$ we have

$$\begin{aligned}\mathcal{W}_2^2(\alpha_t * \gamma, \nu) = \mathcal{W}_2^2(\alpha_t * \gamma, \delta_{(0,0)}) &= \int |x|^2 \, d(\alpha_t * \gamma)(x) = \int |x|^2 \, d\alpha_t(x) + \int |x|^2 \, d\gamma(x) \\ &= \frac{1}{2}\int |x|^2 \, d\delta_{(-2t,t)} + \frac{1}{2}\int |x|^2 \, d\delta_{(2-2t,1-t)} + 2 = 5t^2 - 5t + \frac{9}{2}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\mathcal{U}(\alpha_t) &= \mathcal{W}_2^2(\alpha_t, \mu) - \mathcal{W}_2^2(\alpha_t * \gamma, \nu) = \min\left\{5t^2 - 7t + \frac{13}{2}, 5t^2 - 3t + \frac{9}{2}\right\} - 5t^2 + 5t - \frac{9}{2} \\ &= \min\{2 - 2t, 2t\}.\end{aligned}$$

Figure 4, clearly shows that in this case $\mathcal{U}(\alpha_t)$ is not convex because it has a concave cusp at $t = \frac{1}{2}$.

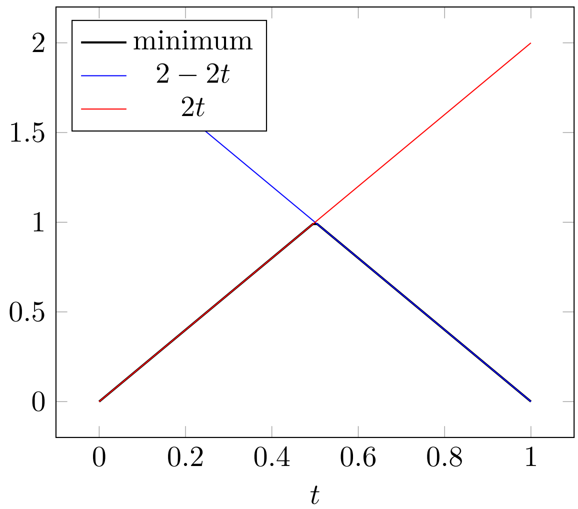

Figure 4: Graph of $\mathcal{U}(\alpha_t)$ from Example 4.4

However, the previous example was not very relevant in the setting of Bass martingales since μ and ν have clearly not been in convex order. The next example will remedy this caveat, with the drawback that it will not be as easy to explicitly calculate the values of $\mathcal{U}(\alpha_t)$. First, the following theorem is needed.

Theorem 4.5 (Derivative of the Wasserstein distance [30, Theorem 23.9])**.**
Let $(\alpha_t)_{0\le t\le 1} \subset \mathcal{P}_2^{ac}(\mathbb{R}^d)$ be a weakly continuous curve (absolutely continuous with respect to Lebesgue) solving the continuity equation

$$\partial_t \alpha_t + \nabla \cdot (\xi_t \alpha_t) = 0,$$

where $\xi_t : \mathbb{R}^d \to \mathbb{R}^d$ is a locally Lipschitz velocity field satisfying

$$\int_0^1 \int_{\mathbb{R}^d} |\xi_t(x)|^2 \, d\alpha_t(x) \, dt < \infty.$$

Then $t \mapsto \alpha_t$ is Hölder-$\frac{1}{2}$ and absolutely continuous in $\mathcal{W}_2$. Moreover, for $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ and a.e. $t \in [0,1]$ we have

$$\frac{d}{dt}\left(\frac{1}{2}\mathcal{W}_2^2(\alpha_t,\nu)\right) = \int \langle \nabla\psi_t(x), \xi_t(x)\rangle \, d\alpha_t(x), \tag{4.8}$$

where $\psi_t : \mathbb{R}^d \to \mathbb{R}$ is the Kantorovich potential. In other words $T_t(x) := x - \nabla\psi_t(x)$ pushes α_t forward to ν, i.e. $(T_t)(\alpha_t) = \nu$.

Proof. The proof in even more generality for Riemannian manifolds can be found in [30]. The following formal calculations will show how one at least could have guessed formula (4.8). Indeed, let $F(t) := \frac{1}{2}\mathcal{W}_2^2(\alpha_t, \nu) = \int \psi_t(x)\, d\alpha_t(x) + \int \psi_t^*(y)\, d\nu(y) = K(t,t)$ by Kantorovich duality, where $K(t_1, t_2) = \int \psi_{t_1}(x)\, d\alpha_{t_2}(x) + \int \psi_{t_1}^*(y)\, d\nu(y)$. Then

$$F'(t) = \frac{\partial K}{\partial t_1}(t,t) + \frac{\partial K}{\partial t_2}(t,t).$$

Here the first term vanishes because ψ_t is assumed to be optimal in the Kantorovich duality. Therefore, we have

$$\begin{aligned}\frac{d}{dt}\left(\frac{1}{2}\mathcal{W}_2^2(\alpha_t, \nu)\right) = \int \psi_t(x)\, d\partial_t \alpha_t(x) &= -\int \psi_t(x)\, d(\nabla \cdot (\xi_t \alpha_t))(x) \\ &= \int \langle \nabla \psi_t(x), \xi_t(x) \rangle\, d\alpha_t(x).\end{aligned}$$

Here first we used the fact that $(\alpha_t)_{0 \le t \le 1}$ solves the continuity equation and then we applied integration by parts. Finally it is worth noting that (4.8) holds for all Lebesgue points of $t \to \int |\xi_t|^2\, d\alpha_t$, since in the following we will only work with smooth ξ, the formula will hold for all $t \in [0,1]$. $\square$

With this result we can now introduce a more general example, where indeed $\mu \preceq_c \nu$.

Example 4.6. Let μ and $(\alpha_t)_{0 \le t \le 1}$ be as in Example 4.4. Taking $\nu = \mu * \gamma$, then with the same calculations as in (3.22) it is easy to see that indeed we have $\mu \preceq_c \nu$. Additionally it is also easy to see that the pair (μ, ν) is irreducible. Since we know from the previous example that $\mathcal{W}_2^2(\mu, \alpha_t)$ has a concave cusp at $t = \frac{1}{2}$, if we are able to show that the second term in $\mathcal{U}(\alpha_t)$, namely $\mathcal{W}_2^2(\mu * \gamma, \alpha_t * \gamma)$ is smooth enough, e.g. C^1 with respect to t, then this second term cannot remedy the concavity of the first term.

First, let us find ξ with respect to which $(\beta_t)_{0 \le t \le 1} := (\alpha_t * \gamma)_{0 \le t \le 1}$ solves the continuity equation. Therefore, let $\rho(x) := \frac{1}{2\pi} e^{-\frac{|x|^2}{2}}$ denote the standard Gaussian density on $\mathbb{R}^2$. We also define $m_1(t) := (-2t, t)$ and $m_2(t) := (2 - 2t, 1 - t)$. Then for $\rho_t^i(x) := \rho(x - m_i(t))$ a straight forward calculation shows

$$0 = \partial_t \rho_t^i + \dot{m}_i \cdot \nabla \rho_t^i = \partial_t \rho_t^i + \nabla \cdot (\rho_t^i \dot{m}_i).$$

Since $\beta_t = \frac{1}{2}(\rho_t^1 + \rho_t^2)$ and summing the equation above we have

$$0 = \partial_t \beta_t + \nabla \cdot (\frac{1}{2}\sum_{i=1}^{2} \rho_t^i \dot{m}_i).$$

Therefore, it is necessary that $2\beta_t \xi_t = \sum \rho_t^i \dot{m}_i$ and the required ξ_t is given by

$$\xi_t = \frac{\rho(x - m_1(t))\dot{m}_1 + \rho(x - m_2(t))\dot{m}_2}{\rho(x - m_1(t)) + \rho(x - m_2(t))}.$$

Clearly, we have $\xi_t \in C^\infty$ and $|\xi_t| \le \sqrt{5}$. Moreover, $(\beta_t)_{0 \le t \le 1}$ is Lipschitz with respect to the Wasserstein distance.

$$\mathcal{W}_2^2(\beta_s, \beta_t) \le \frac{1}{2}\sum_{i=1}^{2} |m_i(t) - m_i(s)|^2 \le 5|t - s|^2$$

where we used the coupling $\frac{1}{2}\sum_{i=1}^{2}(id, id + m_i(t) - m_i(s))(\rho(x - m_i(s)))$ for the first inequality. Additionally, the second moments of $(\beta_t)_{0\le t\le 1}$ are uniformly bounded

$$\int |x|^2 \, d\beta_t(x) = \frac{1}{2}\sum_{i=1}^{2}\int |z + m_i(t)|^2 \, d\gamma(z) = \int |z|^2 \, d\gamma(z) + \frac{1}{2}|m_1(t)|^2 + \frac{1}{2}|m_2(t)|^2$$

since γ has mean 0. Now taking $\pi_t := (id, id - \nabla\psi_t(x))(\beta_t)$ to be the optimal transport plan. Then we have

$$\frac{d}{dt}\left(\frac{1}{2}\mathcal{W}_2^2(\beta_t, \nu)\right) = \int \langle x - y, \xi_t(x)\rangle \, d\pi_t(x).$$

In order to show continuity, we first will restrict the integral from above to B_R. Indeed, $|\langle x - y, \xi_t(x) - \xi_s(x)\rangle|$ is continuous and bounded by $2\sqrt{5}(|x| + |y|)$. Therefore, we have uniformly for any $t \in [0, 1]$

$$\begin{aligned}\left|\int \langle x - y, \xi_t(x)\rangle \, d\pi_t(x) - \int_{B_R} \langle x - y, \xi_t(x)\rangle \, d\pi_t(x)\right| &\le \int_{B_R^c} c(|x| + |y|) \, d\pi_t \\ &\le \frac{c}{R}\int (|x| + |y|)^2 \, d\pi_t \le \frac{\tilde{c}}{R} \to 0 \, (R \to \infty),\end{aligned}$$

where the last inequality follows from the uniform second moment bound of $(\beta_t)_{0\le t\le 1}$. Therefore, after having chosen R sufficiently large, we split

$$\begin{aligned}&\left|\int_{B_R} \langle x - y, \xi_t(x)\rangle \, d\pi_t(x, y) - \int_{B_R} \langle x - y, \xi_s(x)\rangle \, d\pi_s(x, y)\right| \\ &\le \left|\int_{B_R} \langle x - y, \xi_t(x) - \xi_s(x)\rangle \, d\pi_s(x, y)\right| \\ &+ \left|\int_{B_R} \langle x - y, \xi_t(x)\rangle \, d\pi_t(x, y) - \int_{B_R} \langle x - y, \xi_t(x)\rangle \, d\pi_s(x, y)\right|.\end{aligned}$$

The second term, goes to zero because if $\beta_s \to \beta_t$ in $\mathcal{W}_2$, then at least up to a subsequence $\pi_s \to \pi_t$ weakly. Since this holds for any sequence $s_n \to t$, this was not a restriction. For the first term note that

$$\begin{aligned}\left|\int_{B_R} \langle x - y, \xi_t(x) - \xi_s(x)\rangle \, d\pi_s(x, y)\right| &\le \mathcal{W}_2(\beta_s, \nu)\sqrt{\int_{B_R} |\xi_s(x) - \xi_t(x)|^2 \, d\beta_s(x)} \\ &\le \mathcal{W}_2^2(\beta_s, \nu) \sup_{x\in B_R} |\xi_s(x) - \xi_t(x)|.\end{aligned}$$

Here $\mathcal{W}_2(\beta_s, \nu)$ is bounded for $t \in [0, 1]$ because $(\beta_t)_{0\le t\le 1}$ was Lipschitz with respect to the Wasserstein distance and the supremum goes to 0 for $s \to t$ because $[0, 1] \times B_R$ is compact and $\xi_t(x)$ is continuous in (t, x).

4.2 Generalized geodesics

Following [2], the following will introduce a *generalized geodesic*. This will allow us to generalize the previous proposition to the case of $d \in \mathbb{N}$.

Definition 4.7 ([2, Definition 9.2.2])**.** A *generalized geodesic* joining μ_2 to μ_3 with base μ_1 is a curve of the type

$$\mu_t^{2\to 3} = (p_t^{2\to 3})(\mu) \quad t \in [0,1]$$

where $\mu \in \Pi(\mu_1, \mu_2, \mu_3)$, $p^{1,2}(\mu) \in \Pi_o(\mu_1, \mu_2)$, $p^{1,3}(\mu) \in \Pi_o(\mu_1, \mu_3)$ and $p_t^{2\to 3} := (1-t)p^2 + tp^3$.

Lemma 4.8 ([2, Lemma 9.2.1])**.** *For a generalized geodesic, as defined above, we have*

$$\mathcal{W}_2^2(\mu_1, \mu_t^{2\to 3}) = \int |y_1 - y_2|^2 \ d\mu_t^{1,2\to 3}(y_1, y_2), \tag{4.9}$$

where $\mu_t^{1,2\to 3} := ((1-t)p^{1,2} + tp^{1,3})(\mu)$.

Proof. In order to show, that $\mu_t^{1,2\to 3} \in \Pi_o(\mu_1, \mu_t^{2\to 3})$, we will show that its support is cyclically monotone. Since for a continuous function f we have

$$\operatorname{supp}(f(\mu)) = \overline{f(\operatorname{supp}(\mu))}, \tag{4.10}$$

it suffices to show that $p_t^{1,2\to 3}(\operatorname{supp}(\mu))$ is cyclically monotone. Choosing points $(a_i, b_i) \in p_t^{1,2\to 3}(\operatorname{supp}(\mu))$, $i = 1, ..., N$, we can then also find b_i', b_i'', such that

$$(a_i, b_i', b_i'') \in \mu, \quad b_i = (1-t)b_i' + tb_i''.$$

Where again we will also have $(a_i, b_i') \in \operatorname{supp}(\mu^{1,2})$ and $(a_i, b_i'') \in \operatorname{supp}(\mu^{1,3})$ by property (4.10). Since $\operatorname{supp}(\mu^{1,i})$ is cyclically monotone, for any permutation σ

$$\begin{aligned}\sum_{i=1}^N \langle a_i - a_{\sigma(i)}, b_i\rangle &= \sum_{i=1}^N \langle a_i - a_{\sigma(i)}, (1-t)b_i' + tb_i''\rangle \\ &= \sum_{i=1}^N \langle a_i - a_{\sigma(i)}, (1-t)b_i'\rangle + \sum_{i=1}^N \langle a_i - a_{\sigma(i)}, tb_i''\rangle \geq 0.\end{aligned}$$

□

Remark 4.9. In particular in the notational setting of the Bass functional, if for random variables $X \sim \mu$, $Z_0 \sim \alpha_0$ and $Z_1 \sim \alpha_1$, the couplings (Z_0, X) and (Z_1, X) are optimal, then we have that for $Z_t := tZ_1 + (1-t)Z_0$ the couplings (Z_t, X) are optimal as well. In addition, for $\alpha_t := \mathrm{Law}(Z_t)$ we have that $(\alpha_t)_{0\leq t\leq 1}$ is a generalized geodesic joining α_0 to α_1 with base μ.

Remark 4.10. It is also worth noting that this is truly a generalized geodesic, because any true Wasserstein geodesic, i.e. a geodesic in $\mathcal{P}_2(\mathbb{R}^d)$ joining μ_2 and μ_3, is also a generalized geodesic joining μ_2 and μ_3 with base μ_3. Indeed, in this case a curve is a generalized geodesic if and only if it is a true geodesic. The condition $p^{2,3}(\mu) \in \Pi_o(\mu_1, \mu_3)$ for $\mu \in \Pi(\mu_1, \mu_2, \mu_3)$ forces

$$\mu = \int \delta_{x_1}(x_3)\, d\mu^{1,2}(x_1, x_2) \quad \text{for } \mu^{1,2} \in \Pi(\mu_1, \mu_2).$$

Therefore, the second condition for a generalized geodesic $\mu^{1,2} = p^{1,2}\mu \in \Pi_o(\mu_1, \mu_2)$, implies that $\mu_t^{2\to 3}$ is the geodesic interpolation $(tp^1 + (1-t)p^2)(\mu^{1,2})$.

Similar as before, we can now define convexity along generalized geodesics.

Definition 4.11. Let $\phi : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$. We say that ϕ is convex along generalized geodesics in $\mathcal{P}_2(\mathbb{R}^d)$ if for every $\mu_1, \mu_2, \mu_3 \in \mathcal{P}_2(\mathbb{R}^d)$, there exists a generalized geodesic $\mu_t^{2\to 3}$ induced by a plan $\mu \in \Pi(\mu_1, \mu_2, \mu_3)$ such that

$$\phi(\mu_t^{2\to 3}) \le (1-t)\phi(\mu_2) + t\phi(\mu_3) \quad \forall t \in [0, 1].$$

With these definitions we can now prove a generalized analogue of Proposition 4.2, which holds in higher dimensions as well.

Proposition 4.12 ([2, Proposition 5.2])**.** *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$. The Bass functional*

$$\mathcal{P}_2(\mathbb{R}^d) \in \alpha \longmapsto \mathcal{V}(\alpha) = \mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu) \tag{4.11}$$

is convex along generalized geodesics $(\alpha_t)_{0\le t\le 1}$ in $\mathcal{P}_2(\mathbb{R}^d)$ with base μ.

Proof. The proof is exactly the same as for Proposition 4.2. The only difference is that optimality of (Z_t, X) follows from the assumption that $(\alpha_t)_{0\le t\le 1}$ is a generalized geodesic with base μ (see Remark 4.9). Therefore, (4.6) holds and the rest then again carries over verbatim. □

Finally, it is worth noting, that the Bass functional also fails to be convex when considering convex combinations of measures in the usual linear sense.

Example 4.13. Consider

$$\mu = \delta_0,\ \nu = \frac{1}{2}(\delta_{-1} + \delta_1) \text{ and } \alpha_t = t\delta_1 + (1-t)\delta_{-1},$$

where $(\alpha_t)_{0\le t\le 1}$ is the linear interpolation between the measures δ_{-1} and δ_1. Clearly $\mathrm{MCov}(\alpha_t, \mu) = 0$. Additionally, $\mathrm{MCov}(\alpha_{1/2} * \gamma, \nu) = \mathbb{E}[|X_{1/2}|]$ for $X_{1/2} \sim \alpha_{1/2} * \gamma$. Indeed, since $\alpha_{1/2} * \gamma$ is centered and symmetric, the optimal coupling is $Y = \mathrm{sign}(X_{1/2})$. This fulfills $Y \sim \nu$ while maximizing $\mathbb{E}[X_{1/2}Y]$. Further, by symmetry of $\Gamma \sim \gamma$ we have

$$\mathbb{E}[|X_{1/2}|] = \frac{1}{2}(\mathbb{E}[|-1+\Gamma|] + \mathbb{E}[|1+\Gamma|]) = \mathbb{E}[|1+\Gamma|].$$

Now for $t = 0$, the optimal comonotone coupling is $Y = \text{sign}(X_0 + 1)$ for $X_0 \sim \alpha_0 * \gamma$. A similar calculation shows that $\mathbb{E}[X_0 Y] = \mathbb{E}\|\Gamma|]$. The same holds for $t = 1$ This shows that

$$\frac{1}{2}(\mathcal{V}(\alpha_0) + \mathcal{V}(\alpha_t)) = \mathbb{E}[|\Gamma|] < \mathbb{E}[|1 + \Gamma|] = \mathcal{V}(\alpha_{1/2}),$$

which contradicts convexity.

4.3 Convexity of the L^2-lift

In this subsection I will follow [4] to introduce a L^2-lift of the Bass functional. This functional is going to be convex and Fréchet differentiable allowing us to define a L^2-Bass gradient flow. Throughout this subsection we will again only consider $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ which are in convex order, i.e. $\mu \preceq_c \nu$. We will fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting the independent random variables $X \sim \mu$ and $\Gamma \sim \gamma$. We define the σ-algebras $\mathcal{G} := \sigma(X)$ and $\mathcal{H} := \sigma(X, \Gamma)$ with their corresponding L^2-spaces of $\mathbb{R}^d$-valued random variables $L^2(\mathcal{H}; \mathbb{R}^d)$ and $L^2(\mathcal{G}; \mathbb{R}^d)$.

Definition 4.14. In this setting the *lifted Bass functional* is given by

$$L^2(\mathcal{G}; \mathbb{R}^d) \in Z \mapsto V(Z) := \sup_{\substack{Y \in L^2(\mathcal{H}; \mathbb{R}^d), \\ Y \sim \nu}} \mathbb{E}[(Z + \Gamma) \cdot Y - Z \cdot X]. \tag{4.12}$$

The next lemma is going to establish a couple of fundamental properties of this functional.

Lemma 4.15 ([4, Definition 2.1])**.** *The lifted Bass functional has the following properties:*

1. *V is Fréchet differentiable. Denoting by v the Brenier potential from* $\text{Law}(Z + \Gamma)$ *to ν, the derivative is given by $D_Z V = (\nabla v * \gamma)(Z) - X \in L^2(\mathcal{G}; \mathbb{R}^d)$.*

2. *$V(Z) \geq \mathcal{V}(\text{Law}(Z))$. Additionally, if the coupling (X, Z) is optimal in the sense that $\mathbb{E}[X \cdot Z] = \text{MCov}(\mu, \text{Law}(Z))$, then $V(Z) = \mathcal{V}(\text{Law}(Z))$.*

3. *V is convex and weakly lower semicontinuous.*

4. *If the Bass functional $\mathcal{V}$ has a minimizer $\hat{\alpha}$ in $\mathcal{P}_2(\mathbb{R}^d)$, then there exists a (up to translation unique) $\hat{Z} \in L^2(\mathcal{G}; \mathbb{R}^d)$ such that the coupling $(X, \hat{Z})$ is optimal and $V(\hat{Z}) = \mathcal{V}(\hat{\alpha})$.*

Proof. The more technical proof of item 1. can be found in [4]. The proof of 2. is straightforward. If (X, Z) is an optimal coupling, then denoting $\alpha := \text{Law}(Z)$, we have

$$\begin{aligned} V(Z) &= \sup_{\substack{Y \in L^2(\mathcal{H}; \mathbb{R}^d), \\ Y \sim \nu}} \mathbb{E}[(Z + \Gamma) \cdot Y] - \text{MCov}(\mu, \alpha) \\ &= \text{MCov}(\nu, \alpha * \gamma) - \text{MCov}(\mu, \alpha) = \mathcal{V}(\alpha). \end{aligned}$$

With the first equality being and inequality if the coupling is not optimal. For the proof of item 3. let $Z_0, Z_1 \in L^2(\mathcal{G};\mathbb{R}^d)$. By Brenier's theorem there exists a convex $v_i : \mathbb{R}^d \to \mathbb{R}$ such that $(Z_i + \Gamma, \nabla(Z_i + \Gamma))$ is an optimal coupling between the laws $\mathrm{Law}(Z_i) * \gamma$ and ν for $i = 0, 1$. Therefore,

$$\begin{aligned} V(Z_1) - V(Z_0) &= \mathbb{E}[(Z_1 + \Gamma) \cdot \nabla v_1(Z_1 + \Gamma) - (Z_0 + \Gamma) \cdot \nabla v_0(Z_0 + \Gamma) - (Z_1 - Z_0) \cdot X] \\ &\geq \mathbb{E}[(Z_1 + \Gamma) \cdot \nabla v_0(Z_0 + \Gamma) - (Z_0 + \Gamma) \cdot \nabla v_0(Z_0 + \Gamma) - (Z_1 - Z_0) \cdot X] \\ &= \mathbb{E}[(Z_1 - Z_0) \cdot \nabla v_0(Z_0 + \Gamma) - (Z_1 - Z_0) \cdot X] \\ &= \mathbb{E}[(Z_1 - Z_0) \cdot (\mathbb{E}[\nabla v_0(Z_0 + \Gamma)|\mathcal{G}] - X)] \\ &= \mathbb{E}[(Z_1 - Z_0) \cdot D_{Z_0} V], \end{aligned}$$

which implies that V is convex. Additionally, a continuous convex functional is also weakly lower semicontinuous [10, Theorem 9.1], which concludes the proof of this item. For the final item 4. if $\mathcal{V}$ has a minimizer $\hat{\alpha}$, then this minimizer is also the Bass measure of a Bass martingale from μ to ν. By Theorem 3.13 this $\hat{\alpha}$ is given as the push-forward of μ under the gradient of a convex map φ, i.e. $\hat{\alpha} = \nabla\varphi(\mu)$. Therefore, defining $\hat{Z} := \nabla\varphi(X)$ yields an optimal coupling $(X, \hat{Z})$. Applying item 2. yields $V(\hat{Z}) = \mathcal{V}(\hat{\alpha})$. As already noted before, this construction is unique up to a translation of the barycenter of α. Additionally, the lifted Bass functional is translation invariant, therefore after specifying the mean of $\hat{Z}$ we indeed have uniqueness. □

If we consider $Z_0, Z_1 \in L^2(\mathcal{G};\mathbb{R}^d)$ such that the pairs (Z_i, X), $i = 0, 1$ are optimal and set $Z_t := tZ_0 + (1-t)Z_1$, then $\mathrm{Law}(Z_t)$ is the generalized geodesic from $\mathrm{Law}(Z_0)$, to $\mathrm{Law}(Z_1)$ with base μ. Thus in this case convexity of the Bass functional along generalized geodesics coincides with convexity of the lifted Bass functional. Further, it is also remarkable, that the lifted Bass functional is differentiable, whereas $\mathcal{V}$ was not $\mathcal{W}_2$-differentiable. Similarly as with convexity, this was exactly due to the term $\mathrm{MCov}(\cdot, \mu)$. This issue is resolved with the lifted Bass functional by fixing the pair (X, Z).

With differentiability and convexity of the Bass functional it makes sense to introduce a L^2-Bass gradient flow. Indeed, $D_Z V$ denotes the direction in which V increases most rapidly. The gradient flow associated to V is flow induced by the following differential equation

$$\frac{dZ_t}{dt} = -D_{Z_t} V = -[(\nabla v_t * \gamma)(Z_t) - X],$$

where $\{Z_t\}_{t \in \mathbb{R}_+} \subset L^2(\mathcal{G};\mathbb{R}^d)$ is an absolutely continuous curve and v_t is the Brenier potential from $\mathrm{Law}(Z_t) * \gamma$ to ν.

The above Lemma shows that if $\hat{Z}$ minimizes V, then $\mathrm{Law}(\hat{Z})$ is a minimizer of $\mathcal{V}$. Therefore, in order to find the optimal Bass measure, one could hope that the L^2-Bass gradient flow converges towards a minimizer. The first main result in [4] shows that under some assumptions this is indeed the case.

Theorem 4.16 ([4, Theorem 1.8])**.** *For any $Z_0 \in L^2(\mathcal{G};\mathbb{R}^d)$ under the assumptions that the measures μ and ν are not only in convex order but also*

1. *the pair* (μ, ν) *is irreducible,*
2. ν *is compactly supported,*
3. $\operatorname{supp}(\mu) \subseteq \operatorname{int}(\operatorname{co}(\operatorname{supp}\nu))$.

Then the L^2*-Bass gradient flow* $\{Z_t\}_{t\geq 0}$ *exists, is unique and it converges in* L^2*-norm towards a* $\hat{Z} \in L^\infty(\mathcal{G}; \mathbb{R}^d)$*. This* $\hat{Z}$ *is the unique minimizer of* V *with* $\mathbb{E}[\hat{Z}] = \mathbb{E}[Z_0]$*. Additionally, if* Z_0 *is bounded, then the entire curve* $\{Z_t\}_{t\geq 0}$ *is uniformly bounded.*

References

[1] Luigi Ambrosio, Elia Brué, and Daniele Semola. *Lectures on optimal transport.* eng. UNITEXT - La matematica per il 3 + 2 ; volume 130. Cham: Springer, 2021. ISBN: 9783030721619.

[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures.* Springer, 2005.

[3] Julio Backhoff et al. *Bridging classical and martingale Schrödinger bridges.* 2026. arXiv: 2604.01299 [math.PR]. URL: https://arxiv.org/abs/2604.01299.

[4] Julio Backhoff-Veraguas, Gudmund Pammer, and Walter Schachermayer. *The Gradient Flow of the Bass Functional in Martingale Optimal Transport.* 2024. arXiv: 2407.18781 [math.PR]. URL: https://arxiv.org/abs/2407.18781.

[5] Julio Backhoff-Veraguas, Walter Schachermayer, and Bertram Tschiderer. *The Bass functional of martingale transport.* 2023. arXiv: 2309.11181 [math.PR]. URL: https://arxiv.org/abs/2309.11181.

[6] Julio Backhoff-Veraguas et al. "Adapted Wasserstein distances and stability in mathematical finance". In: *Finance and Stochastics* 24.3 (2020), pp. 601–632.

[7] Julio Backhoff-Veraguas et al. *Existence of Bass martingales and the martingale Benamou—Brenier problem in* $\mathbb{R}^d$. 2025. arXiv: 2306.11019 [math.PR]. URL: https://arxiv.org/abs/2306.11019.

[8] Julio Backhoff-Veraguas et al. *Martingale Benamou–Brenier: a probabilistic perspective.* 2019. arXiv: 1708.04869 [math.PR]. URL: https://arxiv.org/abs/1708.04869.

[9] Richard F Bass. "Skorokhod imbedding via stochastic integrals". In: *Séminaire de Probabilités* 17 (1983), pp. 221–224.

[10] Heinz H Bauschke and Patrick L Combettes. *Convex analysis and monotone operator theory in Hilbert spaces.* eng. Second edition. CMS books in mathematics. Cham: Spinger, 2017. ISBN: 9783319483108.

[11] Mathias Beiglböck, Pierre Henry-Labordère, and Friedrich Penkner. *Model-independent Bounds for Option Prices: A Mass Transport Approach.* 2013. arXiv: 1106.5929 [q-fin.PR]. URL: https://arxiv.org/abs/1106.5929.

[12] Mathias Beiglböck and Nicolas Juillet. "On a problem of optimal transport under marginal martingale constraints". In: *The Annals of Probability* 44.1 (Jan. 2016). ISSN: 0091-1798. DOI: 10.1214/14-aop966. URL: http://dx.doi.org/10.1214/14-AOP966.

[13] Jean-David Benamou and Yann Brenier. "A numerical method for the optimal time-continuous mass transport problem and related problems". In: *Contemporary mathematics* 226 (1999), pp. 1–12.

[14] Yann Brenier. "Décomposition polaire et réarrangement monotone des champs de vecteurs". In: *CR Acad. Sci. Paris Sér. I Math.* 305 (1987), pp. 805–808.

[15] Yann Brenier. “Polar factorization and monotone rearrangement of vector-valued functions”. In: *Communications on pure and applied mathematics* 44.4 (1991), pp. 375–417.

[16] Juan A Cuestaalbertos, Ludger Ruschendorf, and Araceli Tuerodiaz. “Optimal coupling of multivariate distributions and stochastic processes”. In: *Journal of Multivariate Analysis* 46.2 (1993), pp. 335–361.

[17] Martin Hairer. *An Introduction to Stochastic PDEs*. 2023. arXiv: `0907.4178 [math.PR]`. URL: `https://arxiv.org/abs/0907.4178`.

[18] David Hobson. “The Skorokhod embedding problem and model-independent bounds for option prices”. In: *Paris-Princeton lectures on mathematical finance 2010*. Springer, 2010, pp. 267–318.

[19] Leonid V Kantorovich. “On the translocation of masses”. In: *Dokl. Akad. Nauk. USSR (NS)*. Vol. 37. 1942, pp. 199–201.

[20] Martin Knott and Cyril S Smith. “On the optimal mapping of distributions”. In: *Journal of Optimization Theory and Applications* 43.1 (1984), pp. 39–49.

[21] Samuel Kotz, Narayanaswamy Balakrishnan, and Norman Lloyd Johnson. *Continuous multivariate distributions. 1, Models and applications*. eng. 2. ed. New York, NY [u.a]: Wiley, 2000. ISBN: 9780471183877.

[22] R Lassalle. “Causal Transference Plans and Their Monge–Kantorovich Problems, submitted”. In: *arXiv preprint arXiv:1303.6925* (2015).

[23] Robert John McCann. *A convexity theory for interacting gases and equilibrium crystals*. Princeton University, 1994.

[24] Gaspard Monge. “Mémoire sur la théorie des déblais et des remblais”. In: *Mem. Math. Phys. Acad. Royale Sci.* (1781), pp. 666–704.

[25] Roger B Nelsen. *An introduction to copulas*. eng. 2. ed. Springer series in statistics. New York, NY: Springer, 2006. ISBN: 9780387286594.

[26] Volker Strassen. “The existence of probability measures with given marginals”. In: *The Annals of Mathematical Statistics* 36.2 (1965), pp. 423–439.

[27] Xiaolu Tan and Nizar Touzi. “Optimal transportation under controlled stochastic dynamics”. In: *The Annals of Probability* 41.5 (Sept. 2013). ISSN: 0091-1798. DOI: `10.1214/12-aop797`. URL: `http://dx.doi.org/10.1214/12-AOP797`.

[28] John F Toland. “A duality principle for non-convex optimisation and the calculus of variations”. In: *Archive for Rational Mechanics and Analysis* 71.1 (1979), pp. 41–61.

[29] Julio Backhoff Veraguas, Mathias Beiglboeck, and Gudmund Pammer. *Existence, Duality, and Cyclical monotonicity for weak transport costs*. 2019. arXiv: `1809.05893 [math.OC]`. URL: `https://arxiv.org/abs/1809.05893`.

[30] Cédric Villani. *Optimal transport: old and new*. Vol. 338. Springer, 2009.

[31] Cédric Villani. *Topics in optimal transportation*. eng. Graduate studies in mathematics ; 58. Providence, RI: American Math. Soc., 2003. ISBN: 9780821833124.