THE BASS FUNCTIONAL OF MARTINGALE TRANSPORT

JULIO BACKHOFF-VERAGUAS, WALTER SCHACHERMAYER, AND BERTRAM TSCHIDERER

Abstract. An interesting question in the field of martingale optimal transport, is to determine the martingale with prescribed initial and terminal marginals which is most correlated to Brownian motion. Under a necessary and sufficient irreducibility condition, the answer to this question is given by a *Bass martingale*. At an intuitive level, the latter can be imagined as an order-preserving and martingale-preserving space transformation of an underlying Brownian motion starting with an initial law α which is tuned to ensure the marginal constraints.

In this article we study how to determine the aforementioned initial condition α. This is done by a careful study of what we dub the *Bass functional*. In our main result we show the equivalence between the existence of minimizers of the Bass functional and the existence of a Bass martingale with prescribed marginals. This complements the convex duality approach in a companion paper by the present authors together with M. Beiglböck, with a purely variational perspective. We also establish an infinitesimal version of this result, and furthermore prove the displacement convexity of the Bass functional along certain generalized geodesics in the 2-Wasserstein space.

Keywords: Optimal transport, Brenier's theorem, Benamou–Brenier, Stretched Brownian motion, Bass martingale.

Mathematics Subject Classification (2010): Primary 60G42, 60G44; Secondary 91G20.

1. Introduction

1.1. **Martingale optimization problem.** Let μ, ν be elements of $\mathcal{P}_2(\mathbb{R}^d)$, the space of probability measures on $\mathbb{R}^d$ with finite second moments. Assume that μ, ν are in convex order, denoted by $\mu \preceq_{\mathrm{c}} \nu$, and meaning that $\int \phi \, d\mu \leqslant \int \phi \, d\nu$ holds for all convex functions $\phi \colon \mathbb{R}^d \to \mathbb{R}$. As in [4, 5] we consider the martingale optimization problem

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

where B is Brownian motion on $\mathbb{R}^d$ and $|\cdot|_{\mathrm{HS}}$ denotes the Hilbert–Schmidt norm. The abbreviation "MBB" stands for "Martingale Benamou–Brenier" and this designation is motivated from the fact that (MBB) can be seen as a martingale counterpart of the classical formulation in optimal transport by Benamou–Brenier [13], see [4, 5]. The problem (MBB) is equivalent to maximizing the covariance between M and Brownian motion subject to the marginal conditions $M_0 \sim \mu$ and $M_1 \sim \nu$, to wit

$$P(\mu,\nu) := \sup_{\substack{M_0 \sim \mu,\, M_1 \sim \nu, \\ M_t = M_0 + \int_0^t \sigma_s \, dB_s}} \mathbb{E}\Big[\int_0^1 \mathrm{tr}(\sigma_t) \, dt \Big]. \tag{1.1}$$

Both problems have the same optimizer and the values are related via

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

We thank Ben Robinson for his valuable comments during the preparation of this paper. WS and BT acknowledge support by the Austrian Science Fund (FWF) through projects P 35197 and P 35519, and JB acknowledges support by the FWF through projects Y 00782 and P 36835.

As shown in [4], the problem (MBB) admits a strong Markov martingale $\hat{M}$ as a unique optimizer, which is called the *stretched Brownian motion* from μ to ν in [4].

1.2. **Bass martingales and structure of stretched Brownian motion.** Owing to the work [5] it is known that the optimality property of stretched Brownian motion is related to a structural / geometric description. For its formulation we start with the following definition.

Definition 1.1. For probability measures μ, ν we say that the pair (μ, ν) is *irreducible* if for all measurable sets $A, B \subseteq \mathbb{R}^d$ with $\mu(A), \nu(B) > 0$ there is a martingale $X = (X_t)_{0 \leqslant t \leqslant 1}$ with $X_0 \sim \mu$, $X_1 \sim \nu$ such that $\mathbb{P}(X_0 \in A, X_1 \in B) > 0$.

We remark that in the classical theory of optimal transport one can always find couplings (X_0, X_1) of (μ, ν) such that $\mathbb{P}(X_0 \in A, X_1 \in B) > 0$, for all measurable sets $A, B \subseteq \mathbb{R}^d$ with $\mu(A), \nu(B) > 0$; e.g., by letting (X_0, X_1) be independent. In martingale optimal transport this property may fail.

Next we recall the following concept from [6, 4, 5].

Definition 1.2. Let $B = (B_t)_{0 \leqslant t \leqslant 1}$ be Brownian motion on $\mathbb{R}^d$ with $B_0 \sim \hat{\alpha}$, where $\hat{\alpha}$ is an arbitrary element of $\mathcal{P}(\mathbb{R}^d)$, the space of probability measures on $\mathbb{R}^d$. Let $\hat{v} \colon \mathbb{R}^d \to \mathbb{R}$ be convex such that $\nabla \hat{v}(B_1)$ is square-integrable. We call

$$\hat{M}_t := \mathbb{E}[\nabla \hat{v}(B_1) \,|\, \sigma(B_s \colon s \leqslant t)] = \mathbb{E}[\nabla \hat{v}(B_1) \,|\, B_t], \qquad 0 \leqslant t \leqslant 1 \tag{1.2}$$

a *Bass martingale* with *Bass measure* $\hat{\alpha}$ joining $\mu = \operatorname{Law}(\hat{M}_0)$ with $\nu = \operatorname{Law}(\hat{M}_1)$.

The reason behind this terminology is that Bass [6] used this construction (with $d = 1$ and $\hat{\alpha}$ a Dirac measure) in order to derive a solution of the Skorokhod embedding problem.

In [5, Theorem 1.3] it is shown that under the irreducibility assumption on the pair (μ, ν) there is a unique Bass martingale $\hat{M}$ from μ to ν, i.e., satisfying $\hat{M}_0 \sim \mu$ and $\hat{M}_1 \sim \nu$:

Theorem 1.3. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$ and assume that (μ, ν) is irreducible. Then the following are equivalent for a martingale $\hat{M} = (\hat{M}_t)_{0 \leqslant t \leqslant 1}$ with $\hat{M}_0 \sim \mu$ and $\hat{M}_1 \sim \nu$:*

(1) $\hat{M}$ is stretched Brownian motion, i.e., the optimizer of (MBB).

(2) $\hat{M}$ is a Bass martingale.

Since, for probability measures $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$, stretched Brownian motion always exists by [4, Theorem 1.5], the above theorem states that the existence of a Bass martingale follows from — and is in fact equivalent to — the irreducibility assumption on the pair (μ, ν).

Denoting by $*$ the convolution operator and by γ the standard Gaussian measure on $\mathbb{R}^d$, we remark that the convex function $\hat{v}$ and the Bass measure $\hat{\alpha}$ from Definition 1.2 satisfy the identities

$$(\nabla \hat{v} * \gamma)(\hat{\alpha}) = \mu \qquad \text{and} \qquad \nabla \hat{v}(\hat{\alpha} * \gamma) = \nu. \tag{1.3}$$

We formalize the fundamental relations (1.3) and their correspondence with Bass martingales in Lemma 2.3 below.

Throughout we write γ^t for the d-dimensional centered Gaussian distribution with covariance matrix tI_d and set $\hat{v}_t := \hat{v} * \gamma^{1-t} \colon \mathbb{R}^d \to \mathbb{R}$, for $0 \leqslant t \leqslant 1$. In these terms, (1.2) amounts to

$$\hat{M}_t = \nabla \hat{v}_t(B_t), \qquad 0 \leqslant t \leqslant 1.$$

1.3. **Main results.** In the following we denote by MCov the *maximal covariance* between two probability measures $p_1, p_2 \in \mathcal{P}_2(\mathbb{R}^d)$, defined as

$$\operatorname{MCov}(p_1, p_2) := \sup_{q \in \mathsf{Cpl}(p_1, p_2)} \int \langle x_1, x_2 \rangle \, q(dx_1, dx_2), \tag{1.4}$$

where $\mathsf{Cpl}(\mu, \nu)$ is the set of all couplings $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ between μ and ν, i.e., probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with first marginal μ and second marginal ν. As is well known, maximizing the covariance between p_1 and p_2 is equivalent to minimizing their expected squared distance; see also (2.3) below.

Definition 1.4. We introduce the *Bass functional*

$$\mathcal{P}_2(\mathbb{R}^d) \ni \alpha \longmapsto \mathcal{V}(\alpha) := \mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu). \tag{1.5}$$

In our first main result we derive a novel formulation of problem (1.1), which characterizes the Bass measure $\hat{\alpha}$ in (1.3) as the optimizer of the Bass functional (1.5).

Theorem 1.5. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$. Then*

$$P(\mu, \nu) = \inf_{\alpha \in \mathcal{P}_2(\mathbb{R}^d)} \mathcal{V}(\alpha). \tag{1.6}$$

The right-hand side of (1.6) is attained by $\hat{\alpha} \in \mathcal{P}_2(\mathbb{R}^d)$ 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 Theorem 1.5 is given in Section 3. In Section 4 we will show the following infinitesimal version of Theorem 1.5, which constitutes our second main result:

Theorem 1.6. *Let $(M_t)_{0 \leqslant t \leqslant 1}$ be an $\mathbb{R}^d$-valued martingale bounded in L^2, which is given by the stochastic integral*

$$M_t = M_0 + \int_0^t \sigma_s \, dB_s, \qquad 0 \leqslant t \leqslant 1,$$

where $(\sigma_t)_{0 \leqslant t \leqslant 1}$ is a progressively measurable process. Denote by μ_t the law of M_t. For Lebesgue-a.e. $0 \leqslant t \leqslant 1$ we have, for each $\alpha \in \mathcal{P}_2(\mathbb{R}^d)$, the inequality

$$\mathbb{E}\big[\mathrm{tr}(\sigma_t)\big] \leqslant \liminf_{h \to 0} \tfrac{1}{h}\Big(\mathrm{MCov}(\alpha * \gamma^h, \mu_{t+h}) - \mathrm{MCov}(\alpha, \mu_t)\Big). \tag{1.7}$$

We note that, for a Bass martingale $(\hat{M}_t)_{0 \leqslant t \leqslant 1}$ of the form

$$d\hat{M}_t = \hat{\sigma}_t(\hat{M}_t) \, dB_t,$$

with associated Bass measure $\hat{\alpha} \in \mathcal{P}_2(\mathbb{R}^d)$ and diffusion function $\hat{\sigma}_t \colon \mathbb{R}^d \to \mathbb{R}^{d \times d}$, we have, for Lebesgue-a.e. $0 \leqslant t \leqslant 1$, the equality

$$\mathbb{E}\big[\mathrm{tr}\big(\hat{\sigma}_t(\hat{M}_t)\big)\big] = \frac{d}{dt} \mathrm{MCov}(\hat{\alpha} * \gamma^t, \hat{\mu}_t),$$

where $\hat{\mu}_t = \mathrm{Law}(\hat{M}_t)$. This exhibits the sharpness of (1.7) and shows that Theorem 1.6 is an infinitesimal analogue of Theorem 1.5.

In our final main result we discuss convexity properties of the Bass functional $\alpha \mapsto \mathcal{V}(\alpha)$ defined in (1.5).

Theorem 1.7. *We have the following results:*

1. *If $d = 1$, then $\mathcal{V}$ is displacement convex, i.e., convex along the geodesics given by McCann interpolations* [37].
2. *If $d \geqslant 1$, then $\mathcal{V}$ is displacement convex along generalized geodesics with base μ.*

The proof of this result, together with a discussion on the various forms of convexity stated therein (see e.g. [2, 43, 37]), and a treatment of the strict convexity of $\mathcal{V}$, are deferred to Section 5. We merely stress here that the Bass functional fails to be convex, and can even be concave, if we consider convex combinations of measures in the usual linear sense.

1.4. **Related literature.** Optimal transport as a field in mathematics goes back to Monge [38] and Kantorovich [33], who established its modern formulation. The seminal results of Benamou, Brenier, and McCann [15, 16, 13, 35, 36] form the basis of the modern theory, with striking applications in a variety of different areas, see the monographs [43, 44, 1, 41].

We are interested in transport problems where the transport plan satisfies an additional martingale constraint. This additional requirement arises naturally in finance (e.g. [8]), but is of independent mathematical interest. For example there are notable consequences for the study of martingale inequalities (e.g. [14, 29, 40]) and the Skorokhod embedding problem (e.g. [7, 32, 12]). Early articles on this topic of *martingale optimal transport* include [30, 8, 42, 23, 21, 17]. The study of irreducibility of a pair of marginals (μ, ν) was initiated by Beiglböck and Juillet [11] in dimension one and extended in the works [24, 20, 39] to multiple dimensions.

Continuous-time martingale optimal transport problems have received much attention in the recent years; see e.g. [9, 19, 26, 28, 25, 18, 27]. In this paper we concern ourselves with the specific structure given by the martingale Benamou–Brenier problem, introduced in [4] in probabilistic language and in [31] in PDE language, and subsequently studied through the point of view of duality theory in [5]. In the context of market impact in finance, the same kind of problem appeared independently in a work by Loeper [34].

It was also shown in [4] that the optimizer $\hat{M}$ of the problem (MBB) is the process whose evolution follows the movement of Brownian motion as closely as possible with respect to an *adapted Wasserstein distance* (see e.g. [3, 22]) subject to the given marginal constraints.

Contents

2. Preliminaries

In this short section we give a more detailed review of some of the main results in [5], which will be useful for the coming discussions and proofs.

2.1. **Dual viewpoint.** As established in [5, Theorem 1.4], the problem (1.1) admits a dual formulation with a particularly appealing structure:

Theorem 2.1. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$. The value $P(\mu,\nu)$ of the problem (1.1) is equal to*

$$D(\mu,\nu) := \inf_{\substack{\psi \in L^1(\nu), \\ \psi \text{ convex}}} \left(\int \psi \, d\nu - \int (\psi^* * \gamma)^* \, d\mu \right) \tag{2.1}$$

and is attained by a convex function $\hat{\psi}$ if and only if (μ,ν) is irreducible. In this case, the (unique) optimizer to (MBB), (1.1) is given by the Bass martingale with associated convex function $\hat{v} = \hat{\psi}^*$ and Bass measure $\hat{\alpha} = \nabla(\hat{\psi}^* * \gamma)^*(\mu)$.

Note that the symbol $*$ used as a superscript denotes the convex conjugate of a function. We also remark that attainment of $D(\mu,\nu)$ has to be understood in a "relaxed" sense, since the optimizer $\hat{\psi}$ is not necessarily ν-integrable; see [5, Proposition 4.2].

2.2. **Static martingale optimal transport.** We fix $\mu,\nu\in\mathcal{P}_2(\mathbb{R}^d)$ with $\mu\preceq_{\mathrm{c}}\nu$ and consider a static / discrete-time version of the continuous-time martingale optimization problem (1.1), to wit

$$\tilde{P}(\mu,\nu):=\sup_{\pi\in\mathsf{MT}(\mu,\nu)}\int \mathrm{MCov}(\pi_x,\gamma)\,\mu(dx). \tag{2.2}$$

The collection of martingale transports $\mathsf{MT}(\mu,\nu)$ consists of those couplings $\pi\in\mathsf{Cpl}(\mu,\nu)$ that satisfy $\mathrm{bary}(\pi_x):=\int y\,\pi_x(dy)=x$, for μ-a.e. $x\in\mathbb{R}^d$. Here, the family of probability measures $\{\pi_x\}_{x\in\mathbb{R}^d}\subseteq\mathcal{P}_2(\mathbb{R}^d)$ is obtained by disintegrating the coupling π with respect to its first marginal μ, i.e., $\pi(dx,dy)=\pi_x(dy)\,\mu(dx)$.

By [4, Theorem 2.2] the value $\tilde{P}(\mu,\nu)$ of (2.2) is finite and equals $P(\mu,\nu)$, as defined in (1.1). Furthermore, there exists a unique optimizer $\hat{\pi}\in\mathsf{MT}(\mu,\nu)$ of (2.2) and if $(\hat{M}_t)_{0\leqslant t\leqslant 1}$ is the stretched Brownian motion from μ to ν, then the law of $(\hat{M}_0,\hat{M}_1)$ equals $\hat{\pi}$.

As already alluded to, maximizing the maximal covariance is equivalent to minimizing the squared quadratic Wasserstein distance, modulo adding constants. More precisely, in the present setting we have the relation

$$\inf_{\pi\in\mathsf{MT}(\mu,\nu)}\int \mathcal{W}_2^2(\pi_x,\gamma)\,\mu(dx)=d+\int |y|^2\,d\nu(y)-2\tilde{P}(\mu,\nu),$$

where the quadratic Wasserstein distance $\mathcal{W}_2(\,\cdot\,,\,\cdot\,)$ between two probability measures $p_1,p_2\in\mathcal{P}_2(\mathbb{R}^d)$ is defined as

$$\mathcal{W}_2(p_1,p_2):=\sqrt{\inf_{q\in\mathsf{Cpl}(p_1,p_2)}\int |x_1-x_2|^2\,q(dx_1,dx_2)}. \tag{2.3}$$

In these terms, the value of (MBB) can be expressed as

$$MT(\mu,\nu)=\inf_{\pi\in\mathsf{MT}(\mu,\nu)}\int \mathcal{W}_2^2(\pi_x,\gamma)\,\mu(dx)-\int |x|^2\,d\mu(x).$$

2.3. **Structure of optimizers.** From [5, Theorem 6.6] we recall the following characterization of the dual optimizer $\hat{\psi}$ of (2.1) and of the primal optimizer $\hat{\pi}\in\mathsf{MT}(\mu,\nu)$ of (2.2).

Lemma 2.2. *Let $\mu,\nu\in\mathcal{P}_2(\mathbb{R}^d)$ with $\mu\preceq_{\mathrm{c}}\nu$. Suppose that a Bass martingale $(\hat{M}_t)_{0\leqslant t\leqslant 1}$ from μ to ν with Bass measure $\hat{\alpha}\in\mathcal{P}(\mathbb{R}^d)$ and associated convex function $\hat{v}$ exists. Then the Legendre transform $\hat{v}^*$ is equal to the dual optimizer $\hat{\psi}$ of (2.1) and* $\mathrm{Law}(\hat{M}_0,\hat{M}_1)$ *is equal to the primal optimizer $\hat{\pi}$ of (2.2). Furthermore, we have $\hat{\alpha}=\nabla\hat{\varphi}(\mu)$, where*

$$\nabla\hat{\varphi}(x)=(\nabla\hat{v}*\gamma)^{-1}(x)=\nabla(\hat{v}*\gamma)^*(x), \tag{2.4}$$

for μ-a.e. $x\in\mathbb{R}^d$, and

$$\hat{\pi}_x=\mathrm{Law}(\hat{M}_1\,|\,\hat{M}_0=x)=\nabla\hat{v}(\gamma_{\nabla\hat{\varphi}(x)}), \tag{2.5}$$

where $\gamma_{\nabla\hat{\varphi}(x)}$ denotes the d-dimensional Gaussian distribution with barycenter $\nabla\hat{\varphi}(x)$ and covariance matrix I_d.

We set $\hat{u} := \hat{v} * \gamma$, so that $\nabla\hat{u}(\hat{\alpha}) = \mu$. Recalling (1.3), we summarize the relationships between the optimizers in the following diagram:

$$\begin{tikzcd}
\hat{\alpha} * \gamma \arrow[r, shift right, "\nabla\hat{v}"'] & \nu \arrow[l, shift right, "\nabla\hat{\psi}"'] \\
\hat{\alpha} \arrow[u, "*"] \arrow[r, shift right, "\nabla\hat{u}"'] & \mu \arrow[l, shift right, "\nabla\hat{\varphi}"'] \arrow[u, "\hat{\pi}."']
\end{tikzcd}$$

Finally, we prove the equivalence between the identities (1.3) and the existence of a Bass martingale from μ to ν.

Lemma 2.3. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. There is a Bass martingale $\hat{M}$ with Bass measure $\hat{\alpha} \in \mathcal{P}(\mathbb{R}^d)$ from $\mu = \mathrm{Law}(\hat{M}_0)$ to $\nu = \mathrm{Law}(\hat{M}_1)$ if and only if there is a convex function $\hat{v} \colon \mathbb{R}^d \to \mathbb{R}$ satisfying the identities*

$$(\nabla\hat{v} * \gamma)(\hat{\alpha}) = \mu \qquad \text{and} \qquad \nabla\hat{v}(\hat{\alpha} * \gamma) = \nu. \tag{2.6}$$

Moreover, the Bass martingale $\hat{M}$ can be expressed as

$$\hat{M}_t = \nabla\hat{v}_t(B_t), \qquad 0 \leqslant t \leqslant 1. \tag{2.7}$$

Proof. Let $\hat{M}$ be a Bass martingale in the sense of Definition 1.2. We first prove (2.7). Let $A \subseteq \mathbb{R}^d$ be a Borel set. We have to show that

$$\mathbb{E}\big[\nabla\hat{v}(B_1)\, \mathbf{1}_{\{B_t \in A\}}\big] = \mathbb{E}\big[(\nabla\hat{v} * \gamma^{1-t})(B_t)\, \mathbf{1}_{\{B_t \in A\}}\big]. \tag{2.8}$$

Denote by $\varphi_t(x,y)$ the Gaussian kernel, for $t \in (0,1]$ and $x, y \in \mathbb{R}^d$. Then the left-hand side of (2.8) can be expressed as

$$\int \hat{\alpha}(dx_0) \int_A \varphi_t(x_0, dx_t) \int \nabla\hat{v}(x_1)\, \varphi_{1-t}(x_t, dx_1),$$

while the right-hand side is equal to

$$\int \hat{\alpha}(dx_0) \int_A (\nabla\hat{v} * \gamma^{1-t})(x_t)\, \varphi_t(x_0, dx_t).$$

Now we see that (2.8) follows from

$$\int \nabla\hat{v}(x_1)\, \varphi_{1-t}(x_t, dx_1) = \int \nabla\hat{v}(x_1)\, \gamma_{x_t}^{1-t}(dx_1) = (\nabla\hat{v} * \gamma^{1-t})(x_t),$$

where $\gamma_{x_t}^{1-t}$ denotes the d-dimensional Gaussian distribution with barycenter x_t and covariance matrix $(1-t)I_d$. This completes the proof of (2.7).

In particular, at times $t = 0$ and $t = 1$ we obtain from (2.7) that $\hat{M}_0 = (\nabla\hat{v} * \gamma)(B_0)$ and $\hat{M}_1 = \nabla\hat{v}(B_1)$, respectively. If $\hat{M}$ is a Bass martingale from $\mu = \mathrm{Law}(\hat{M}_0)$ to $\nu = \mathrm{Law}(\hat{M}_1)$, this readily gives (2.6)

Conversely, suppose that $\mu, \nu, \hat{\alpha}, \hat{v}$ satisfy the identities (2.6). Let $(B_t)_{0 \leqslant t \leqslant 1}$ be Brownian motion on $\mathbb{R}^d$ with $\mathrm{Law}(B_0) = \hat{\alpha}$. We then define a process $(\hat{M}_t)_{0 \leqslant t \leqslant 1}$ by (1.2). In light of the previous argument, $\hat{M}$ is characterized by (2.7). Since by assumption the identities (2.6) are satisfied, we see that $\mathrm{Law}(\hat{M}_0) = \mu$ and $\mathrm{Law}(\hat{M}_1) = \nu$. Thus $\hat{M}$ is indeed a Bass martingale from μ to ν. ◻

3. A variational characterization of Bass measures

Throughout this section we fix $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$ and provide the proof of Theorem 1.5. This is done in several steps.

Lemma 3.1. *We have the weak duality*

$$P(\mu, \nu) \leqslant \inf_{\alpha \in \mathcal{P}_2(\mathbb{R}^d)} \mathcal{V}(\alpha). \tag{3.1}$$

Proof. Let $\alpha \in \mathcal{P}_2(\mathbb{R}^d)$ be arbitrary. By Brenier's theorem [43, Theorem 2.12] there is a convex function v such that $\nabla v(\alpha * \gamma) = \nu$. Hence from the Kantorovich duality [44, Theorem 5.10] it follows that

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

Since $v * \gamma$ is convex, applying once more the Kantorovich duality yields

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

Finally, from Theorem 2.1 we deduce that

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

which gives the inequality (3.1). We remark that it is immaterial whether in (2.1) we optimize over convex functions ψ which are elements of $L^1(\nu)$ or which are just μ-a.s. finite, see [5, Section 4]. □

Lemma 3.2. *Suppose that there exists a Bass martingale from μ to ν with Bass measure $\hat{\alpha} \in \mathcal{P}_2(\mathbb{R}^d)$. Then the right-hand side of (3.1) is attained by $\hat{\alpha}$ and is equal to*

$$\mathcal{V}(\hat{\alpha}) = \int \mathrm{MCov}(\hat{\pi}_x, \gamma) \, \mu(dx), \tag{3.2}$$

where $\hat{\pi} \in \mathsf{MT}(\mu, \nu)$ is the optimizer of (2.2).

Proof. By assumption there exists a Bass martingale from μ to ν, with Bass measure $\hat{\alpha} \in \mathcal{P}_2(\mathbb{R}^d)$ and associated convex function $\hat{v}$ satisfying (recall Lemma 2.3) the identities (2.6). According to Lemma 2.2, we have that $\hat{\alpha} = \nabla \hat{\varphi}(\mu)$ and

$$\hat{\pi}_x = \nabla \hat{v}(\gamma_{\nabla \hat{\varphi}(x)}),$$

for μ-a.e. $x \in \mathbb{R}^d$. Applying Brenier's theorem, we deduce that

$$\begin{aligned} \int \mathrm{MCov}(\hat{\pi}_x, \gamma) \, \mu(dx) &= \int \int \big\langle \nabla \hat{v}\big(\nabla \hat{\varphi}(x) + z\big), z \big\rangle \, \gamma(dz) \, \mu(dx) \\ &= \int \int \Big(\big\langle \nabla \hat{v}\big(\nabla \hat{\varphi}(x) + z\big), \nabla \hat{\varphi}(x) + z \big\rangle - \big\langle \nabla \hat{v}\big(\nabla \hat{\varphi}(x) + z\big), \nabla \hat{\varphi}(x) \big\rangle \Big) \, \gamma(dz) \, \mu(dx) \\ &= \int \int \big(\langle \nabla \hat{v}(a + z), a + z \rangle - \langle \nabla \hat{v}(a + z), a \rangle \big) \, \gamma(dz) \, \hat{\alpha}(da) \\ &= \int \langle \nabla \hat{v}, \mathrm{Id} \rangle \, d(\hat{\alpha} * \gamma) - \int \langle (\nabla \hat{v} * \gamma), \mathrm{Id} \rangle \, d\hat{\alpha} \\ &= \mathrm{MCov}(\hat{\alpha} * \gamma, \nu) - \mathrm{MCov}(\hat{\alpha}, \mu) = \mathcal{V}(\hat{\alpha}), \end{aligned}$$

which shows (3.2). Together with the weak duality (3.1) of Lemma 3.1 above, and recalling from Subsection 2.2 that the right-hand side of (3.2) is equal to $\tilde{P}(\mu, \nu) = P(\mu, \nu)$, we conclude the assertion of Lemma 3.2. □

Lemma 3.3. *We have the duality result*

$$P(\mu,\nu)=\inf_{\alpha\in\mathcal{P}_2(\mathbb{R}^d)}\mathcal{V}(\alpha). \tag{3.3}$$

Proof. For $\varepsilon>0$ we define $\mu^\varepsilon:=\mu*\gamma^\varepsilon$ and $\nu^\varepsilon:=\nu*\gamma^{2\varepsilon}$. Then $\mu^\varepsilon\preceq_{\mathrm{c}}\nu^\varepsilon$ and the pair $(\mu^\varepsilon,\nu^\varepsilon)$ is irreducible. Hence by Theorem 1.3 there is a Bass martingale from μ^ε to ν^ε, so that by Lemma 3.2 we have

$$\sup_{\pi\in\mathsf{MT}(\mu^\varepsilon,\nu^\varepsilon)}\int\mathrm{MCov}(\pi_x,\gamma)\,\mu^\varepsilon(dx)=\inf_{\alpha\in\mathcal{P}_2(\mathbb{R}^d)}\Big(\mathrm{MCov}(\alpha*\gamma,\nu^\varepsilon)-\mathrm{MCov}(\alpha,\mu^\varepsilon)\Big). \tag{3.4}$$

By weak optimal transport arguments (see [10, Theorem 2.3]) we know

$$\limsup_{\varepsilon\to0}\sup_{\pi\in\mathsf{MT}(\mu^\varepsilon,\nu^\varepsilon)}\int\mathrm{MCov}(\pi_x,\gamma)\,\mu^\varepsilon(dx)\leqslant\sup_{\pi\in\mathsf{MT}(\mu,\nu)}\int\mathrm{MCov}(\pi_x,\gamma)\,\mu(dx).$$

Therefore, if we can show that the right-hand side of (3.4) converges to the right-hand side of (3.3), we will obtain the inequality

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

which, together with the weak duality of Lemma 3.1, establishes (3.3). But this follows easily from

$$|\mathrm{MCov}(\alpha,\mu^\varepsilon)-\mathrm{MCov}(\alpha,\mu)|\leqslant c_1\varepsilon+\tfrac12|\mathcal{W}_2^2(\alpha,\mu^\varepsilon)-\mathcal{W}_2^2(\alpha,\mu)|\leqslant c_2(\varepsilon+\varepsilon^2)$$

and a similar estimate for $|\mathrm{MCov}(\alpha*\gamma,\nu^\varepsilon)-\mathrm{MCov}(\alpha*\gamma,\nu)|$. □

Lemma 3.4. *Suppose that the right-hand side of* (3.3) *is attained by* $\hat\alpha\in\mathcal{P}_2(\mathbb{R}^d)$. *Then there exists a Bass martingale from* μ *to* ν *with Bass measure* $\hat\alpha$.

Proof. By Brenier's theorem there is a convex function $\hat v$ such that $\nabla\hat v(\hat\alpha*\gamma)=\nu$. According to Lemma 2.3, for the existence of a Bass martingale from μ to ν, it remains to show the first equality in (2.6), i.e.,

$$(\nabla\hat v*\gamma)(\hat\alpha)=\mu. \tag{3.5}$$

Let $\hat Z$ and X be random variables with laws $\hat\alpha$ and μ, respectively, such that

$$\mathrm{MCov}(\hat\alpha,\mu)=\mathbb{E}\big[\langle\hat Z,X\rangle\big]. \tag{3.6}$$

Denote by $\hat q(dz,dx)$ the law of the coupling $(\hat Z,X)$. Let $\boldsymbol{w}\colon\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}^d$ be a smooth function with compact support and define probability measures $(\alpha_u)_{u\in\mathbb{R}}\subseteq\mathcal{P}_2(\mathbb{R}^d)$ by

$$\int f\,d\alpha_u:=\int\int f\big(z+u\boldsymbol{w}(z,x)\big)\,q(dz,dx),\qquad f\in C_b(\mathbb{R}^d). \tag{3.7}$$

We claim that

$$\liminf_{u\to0}\tfrac1u\Big(\mathrm{MCov}(\alpha_u,\mu)-\mathrm{MCov}(\hat\alpha,\mu)\Big)\geqslant\mathbb{E}\Big[\Big\langle\boldsymbol{w}(\hat Z,X),X\Big\rangle\Big] \tag{3.8}$$

and

$$\lim_{u\to0}\tfrac1u\Big(\mathrm{MCov}(\alpha_u*\gamma,\nu)-\mathrm{MCov}(\hat\alpha*\gamma,\nu)\Big)=\mathbb{E}\Big[\Big\langle\boldsymbol{w}(\hat Z,X),(\nabla\hat v*\gamma)(\hat Z)\Big\rangle\Big]. \tag{3.9}$$

Using the optimality of $\hat\alpha\in\mathcal{P}_2(\mathbb{R}^d)$ for the right-hand side of (3.3) and admitting the two claims (3.8), (3.9), we deduce that

$$\begin{aligned}0&\leqslant\liminf_{u\to0}\tfrac1u\Big(\Big(\mathrm{MCov}(\alpha_u*\gamma,\nu)-\mathrm{MCov}(\hat\alpha*\gamma,\nu)\Big)-\Big(\mathrm{MCov}(\alpha_u,\mu)-\mathrm{MCov}(\hat\alpha,\mu)\Big)\Big)\\&\leqslant\mathbb{E}\Big[\Big\langle\boldsymbol{w}(\hat Z,X),(\nabla\hat v*\gamma)(\hat Z)-X\Big\rangle\Big].\end{aligned}$$

Since $\boldsymbol{w}$ was arbitrary, it follows that the random variable $(\nabla\hat v*\gamma)(\hat Z)$ has the same law as X, which readily gives (3.5).

We now turn to the proof of the claim (3.8). By the definition of α_u in (3.7), the random variable $Z_u \coloneqq \hat{Z} + u\boldsymbol{w}(\hat{Z}, X)$ has law α_u. Consequently,

$$\mathrm{MCov}(\alpha_u, \mu) \geqslant \mathbb{E}\big[\langle Z_u, X\rangle\big]. \tag{3.10}$$

Combining (3.6) and (3.10) yields (3.8).

It remains to show the claim (3.9). By analogy with the proof of (3.8), we obtain the inequality "$\geqslant$" in (3.9). For the reverse inequality, we note that by the Kantorovich duality we have

$$\begin{aligned}\mathrm{MCov}(\alpha_u * \gamma, \nu) &= \inf_{v \text{ convex}} \Big(\int v \, d(\alpha_u * \gamma) - \int v^* \, d\nu \Big) \\ &\leqslant \int \hat{v} \, d(\alpha_u * \gamma) - \int \hat{v}^* \, d\nu \\ &= \int (\hat{v} * \gamma) \, d\alpha_u - \int \hat{v}^* \, d\nu\end{aligned}$$

and

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

Therefore

$$\begin{aligned}\mathrm{MCov}(\alpha_u * \gamma, \nu) - \mathrm{MCov}(\hat{\alpha} * \gamma, \nu) &\leqslant \int (\hat{v} * \gamma) \, d\alpha_u - \int (\hat{v} * \gamma) \, d\hat{\alpha} \\ &= \mathbb{E}\big[(\hat{v} * \gamma)\big(\hat{Z} + u\boldsymbol{w}(\hat{Z}, X)\big) - (\hat{v} * \gamma)(\hat{Z})\big]\end{aligned}$$

Using the convexity of the function $\hat{v} * \gamma$, we deduce that

$$\tfrac{1}{u}\Big(\mathrm{MCov}(\alpha_u * \gamma, \nu) - \mathrm{MCov}(\hat{\alpha} * \gamma, \nu)\Big) \leqslant \mathbb{E}\big[\big\langle \boldsymbol{w}(\hat{Z}, X), (\nabla\hat{v} * \gamma)\big(\hat{Z} + u\boldsymbol{w}(\hat{Z}, X)\big)\big\rangle\big].$$

Now observe that the expectation on the right-hand side of the above inequality is equal to the expectation of the random variable

$$Y_u \coloneqq \Big\langle \boldsymbol{w}(\hat{Z}, X)\,, \nabla\hat{v}(\hat{Z} + \Gamma) \exp\big(u\big\langle\Gamma, \boldsymbol{w}(\hat{Z}, X)\big\rangle - \tfrac{u^2}{2}|\boldsymbol{w}(\hat{Z}, X)|^2\big)\Big\rangle,$$

where Γ is a standard Gaussian random vector on $\mathbb{R}^d$, independent of $\hat{Z}$ as well as of X. Clearly by continuity

$$\lim_{u \to 0} Y_u = \big\langle \boldsymbol{w}(\hat{Z}, X), \nabla\hat{v}(\hat{Z} + \Gamma)\big\rangle, \qquad \mathbb{P}\text{-a.s.}$$

As $\boldsymbol{w}$ is smooth with compact support, for $\delta > 0$ we can find constants c_1, c_2 such that

$$\forall u \in [-\delta, \delta]: \quad |Y_u| \leqslant c_1 \, |\nabla\hat{v}(\hat{Z} + \Gamma)| \, \mathrm{e}^{c_2|\Gamma|}.$$

By the Cauchy–Schwarz inequality and since $\nabla\hat{v}(\hat{\alpha} * \gamma) = \nu \in \mathcal{P}_2(\mathbb{R}^d)$, we have the bound

$$\mathbb{E}\big[|\nabla\hat{v}(\hat{Z} + \Gamma)| \, \mathrm{e}^{|\Gamma|}\big] \leqslant \sqrt{\int |y|^2 \, d\nu(y)} \, \sqrt{\mathbb{E}\big[\mathrm{e}^{2|\Gamma|}\big]} < +\infty.$$

Therefore we can apply the dominated convergence theorem and conclude that

$$\limsup_{u \to 0} \tfrac{1}{u}\Big(\mathrm{MCov}(\alpha_u * \gamma, \nu) - \mathrm{MCov}(\hat{\alpha} * \gamma, \nu)\Big) \leqslant \mathbb{E}\big[\big\langle \boldsymbol{w}(\hat{Z}, X), (\nabla\hat{v} * \gamma)(\hat{Z})\big\rangle\big],$$

which completes the proof of the claim (3.9). □

Proof of Theorem 1.5. The assertion of the theorem follows from Lemmas 3.2 – 3.4. □

The reader has certainly noticed that the proof of Lemma 3.2 was given in an analytic style while the proof of Lemma 3.4 was given in a more probabilistic language. In the remainder of this section we give an alternative probabilistic proof of Lemma 3.2 and sketch how to translate the proof of Lemma 3.4 into a more analytic language.

The following probabilistic proof of Lemma 3.2 does not require the duality results developed in [5], but only relies on the definition of Bass martingales.

Probabilistic proof of Lemma 3.2. By assumption there exists a Bass martingale from μ to ν, with Bass measure $\hat{\alpha} \in \mathcal{P}_2(\mathbb{R}^d)$ and associated convex function $\hat{v}$ satisfying (recall Lemma 2.3) the identities (2.6). Let $\alpha \in \mathcal{P}_2(\mathbb{R}^d)$ be arbitrary. We have to show that

$$\begin{aligned}\mathcal{V}(\hat{\alpha}) &= \mathrm{MCov}(\hat{\alpha} * \gamma, \nu) - \mathrm{MCov}(\hat{\alpha}, \mu) \leqslant \\ &\leqslant \mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu) = \mathcal{V}(\alpha).\end{aligned} \tag{3.11}$$

Take a random variable $\hat{Z}$ with law $\hat{\alpha}$ and define

$$X := (\nabla \hat{v} * \gamma)(\hat{Z}). \tag{3.12}$$

By Brenier's theorem the coupling $(\hat{Z}, X)$ is optimal and according to (2.6) the random variable X has law μ. Now choose a random variable Z with law α such that the coupling (Z, X) is optimal with respect to the maximal covariance (equivalently, with respect to the quadratic Wasserstein distance). Clearly

$$\mathrm{MCov}(\alpha, \mu) - \mathrm{MCov}(\hat{\alpha}, \mu) = \mathbb{E}\big[\langle Z - \hat{Z}, X\rangle\big]. \tag{3.13}$$

Take a standard Gaussian random vector Γ on $\mathbb{R}^d$, independent of Z as well as of $\hat{Z}$. The random variables $\hat{Z} + \Gamma$ and

$$Y := \nabla \hat{v}(\hat{Z} + \Gamma) \tag{3.14}$$

have laws $\hat{\alpha} * \gamma$ and ν, respectively. As by Brenier's theorem the coupling $(\hat{Z} + \Gamma, Y)$ is optimal, we have

$$\mathrm{MCov}(\hat{\alpha} * \gamma, \nu) = \mathbb{E}\big[\langle \hat{Z} + \Gamma, Y\rangle\big]. \tag{3.15}$$

Since the random variable $Z + \Gamma$ has law $\alpha * \gamma$ we conclude that $(Z + \Gamma, Y)$ is some coupling between $\alpha * \gamma$ and ν, i.e.,

$$\mathrm{MCov}(\alpha * \gamma, \nu) \geqslant \mathbb{E}\big[\langle Z + \Gamma, Y\rangle\big]. \tag{3.16}$$

From (3.13) – (3.16) we obtain the inequality

$$\mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\hat{\alpha} * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu) + \mathrm{MCov}(\hat{\alpha}, \mu) \geqslant \mathbb{E}\big[\langle Z - \hat{Z}, Y - X\rangle\big].$$

Therefore, in order to establish the inequality (3.11), it remains to show that

$$\mathbb{E}\big[\langle Z - \hat{Z}, Y - X\rangle\big] = 0. \tag{3.17}$$

For that purpose, we condition $Y - X$ on the random variables Z as well as $\hat{Z}$, so that by (3.12) and (3.14) we obtain

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

which implies (3.17). □

We finally give an alternative heuristic argument for Lemma 3.4, which is based on differentiating the maximal covariance along a continuity equation.

Alternative heuristic proof of Lemma 3.4. Suppose that the right-hand side of (3.3) is attained by $\hat{\alpha} \in \mathcal{P}_2(\mathbb{R}^d)$. We want to show that there exists a Bass martingale from μ to ν with Bass measure $\hat{\alpha}$. The idea is to perturb $\hat{\alpha}$ along a continuity equation

$$\partial_t \alpha_t + \mathrm{div}(\boldsymbol{v}_t \alpha_t) = 0, \qquad t \in (-h, h),$$

with $h > 0$, $\alpha_0 := \hat{\alpha}$, and where $\boldsymbol{v}_t$ is a velocity field. Observe that

$$\begin{aligned}\partial_t|_{t=0} \mathcal{V}(\alpha_t) &= \partial_t|_{t=0} \Big(\mathrm{MCov}(\alpha_t * \gamma, \nu) - \mathrm{MCov}(\alpha_t, \mu)\Big) \\ &= \partial_t|_{t=0} \int \hat{v} \, d(\alpha_t * \gamma) - \partial_t|_{t=0} \int \hat{u} \, d\alpha_t,\end{aligned}$$

where $\nabla \hat{v}(\hat{\alpha} * \gamma) = \nu$ is optimal and likewise $\nabla \hat{u}(\hat{\alpha}) = \mu$ is optimal. By the continuity equation we obtain

$$\partial_t|_{t=0} \int \hat{u} \, d\alpha_t = \int \langle \nabla \hat{u}, \boldsymbol{v}_0\rangle \, d\hat{\alpha}.$$

With similar computations we have

$$\partial_t|_{t=0} \int \hat{v}\, d(\alpha_t * \gamma) = \int \langle \nabla\hat{v} * \gamma, v_0 \rangle \, d\hat{\alpha}.$$

As v_0 was arbitrary and $\hat{\alpha}$ was optimal, we conclude that

$$0 = \int \big\langle \nabla\hat{v} * \gamma - \nabla\hat{u}, v_0 \big\rangle \, d\hat{\alpha},$$

so that $\nabla\hat{u}$, the optimal map from $\hat{\alpha}$ to μ, is $\hat{\alpha}$-a.s. equal to $\nabla\hat{v} * \gamma$, where $\nabla\hat{v}$ is the optimal map from $\hat{\alpha} * \gamma$ to ν. Recalling (2.6) and Lemma 2.3, this is precisely the structure of the Bass martingale. □

4. An infinitesimal version of Theorem 1.5

We provide the proof of Theorem 1.6, an infinitesimal version of Theorem 1.5.

Proof of Theorem 1.6. For a partition $\Pi = \{t_0, t_1, \dots, t_n\}$ of the interval $[0, 1]$ with

$$0 = t_0 < t_1 < \ldots < t_n = 1$$

we denote by Σ^Π the collection of all progressively measurable and L^2-bounded processes $(\sigma_t^\Pi)_{0 \leqslant t \leqslant 1}$ such that the stochastic integral

$$M_t^\Pi := M_0 + \int_0^t \sigma_s^\Pi \, dB_s, \qquad 0 \leqslant t \leqslant 1$$

defines an L^2-bounded martingale with $\mathrm{Law}(M_{t_k}^\Pi) = \mu_{t_k}$, for $k = 0, \dots, n$. We define

$$m^\Pi([t_{k-1}, t_k]) := \sup_{\sigma^\Pi \in \Sigma^\Pi} \mathbb{E}\Big[\int_{t_{k-1}}^{t_k} \mathrm{tr}(\sigma_s^\Pi) \, ds \Big]. \tag{4.1}$$

By [4], we know that the optimizer of

$$m^\Pi([0, 1]) = \sup_{\sigma^\Pi \in \Sigma^\Pi} \mathbb{E}\Big[\int_0^1 \mathrm{tr}(\sigma_s^\Pi) \, ds \Big]$$

is given, on each interval $[t_{k-1}, t_k]$, by the stretched Brownian motion from $\mu_{t_{k-1}}$ to μ_{t_k}. By Theorem 1.5 we have

$$m^\Pi([t_{k-1}, t_k]) = \inf_{\alpha \in \mathcal{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(\alpha * \gamma^{t_k - t_{k-1}}, \mu_{t_k}) - \mathrm{MCov}(\alpha, \mu_{t_{k-1}}) \Big). \tag{4.2}$$

For $t_k \in \Pi$ and a refinement Π_1 of Π we have

$$m^\Pi([0, t_k]) \geqslant m^{\Pi_1}([0, t_k]),$$

as the process $(\sigma_t^{\Pi_1})_{0 \leqslant t \leqslant 1}$ has to satisfy more requirements than the process $(\sigma_t^\Pi)_{0 \leqslant t \leqslant 1}$. We therefore may pass to a limit $m := \lim m^\Pi$ along the net of finite partitions Π of the interval $[0, 1]$, which extends to a finite measure on $[0, 1]$, still denoted by m. Clearly the measure m is absolutely continuous with respect to Lebesgue measure on $[0, 1]$ and we denote the corresponding density by $g(t)$, for $0 \leqslant t \leqslant 1$. We claim that, for $0 \leqslant r \leqslant u \leqslant 1$, we have

$$\mathbb{E}\Big[\int_r^u \mathrm{tr}(\sigma_s) \, ds \Big] \leqslant m([r, u]). \tag{4.3}$$

Indeed, otherwise we could find a partition Π with $r, u \in \Pi$, such that

$$\mathbb{E}\Big[\int_r^u \mathrm{tr}(\sigma_s^\Pi) \, ds \Big] > m^\Pi([r, u]),$$

which yields a contradiction to the definition of $m^\Pi(\,\cdot\,)$ in (4.1). Since (4.3) holds for all intervals $[r, u] \subseteq [0, 1]$, we deduce that

$$\mathbb{E}\big[\mathrm{tr}(\sigma_t) \big] \leqslant g(t), \tag{4.4}$$

for Lebesgue-a.e. $0 \leqslant t \leqslant 1$. From (4.2) we conclude, for Lebesgue-a.e. $0 \leqslant t \leqslant 1$ and for each $\alpha \in \mathcal{P}_2(\mathbb{R}^d)$, the inequality

$$g(t) \leqslant \liminf_{h \to 0} \tfrac{1}{h}\Big(\mathrm{MCov}(\alpha * \gamma^h, \mu_{t+h}) - \mathrm{MCov}(\alpha, \mu_t)\Big).$$

Together with (4.4), this finishes the proof of (1.7). □

Again we provide a more analytic argument for Theorem 1.6, at least on a formal level.

Alternative heuristic proof of Theorem 1.6. We will use the Kantorovich duality and the Fokker–Planck equations to get a hold of $\frac{d}{dh}\mathrm{MCov}(\alpha * \gamma^h, \mu_{t+h})$. By a change of variables we then get an equivalent expression which, when minimized, gives the left-hand side of (1.7). We suppose here that M is a strong solution of the stochastic differential equation $dM_u = \sigma_u(M_u)\, dB_u$, with σ as benevolent as needed, so that in particular μ_u admits a density for each u.

We set $\rho_h := \alpha * \gamma^h$, $\Sigma := \sigma\sigma'$, and notice that for fixed t we have

$$\partial_h \rho_h(x) = \tfrac{1}{2}\Delta\rho_h(x), \qquad \rho_0 = \alpha;$$
$$\partial_h \mu_{t+h}(x) = \tfrac{1}{2}\sum_{i,k}\partial^2_{ik}\big(\Sigma_{ik}\mu_{t+h}(x)\big).$$

By the Kantorovich duality we have

$$\begin{aligned}\mathrm{MCov}(\rho_h, \mu_{t+h}) &= \inf_{\phi \text{ convex}} \int \phi \, d\rho_h + \int \phi^* \, d\mu_{t+h} \\ &= \int \phi_{\rho_h}^{\mu_{t+h}} \, d\rho_h + \int \phi_{\mu_{t+h}}^{\rho_h} \, d\mu_{t+h},\end{aligned}$$

where we denote by $\phi_p^q(\,\cdot\,)$ the convex function, which is unique up to a constant, such that $\nabla\phi_p^q(p) = q$. Using this, or more directly [44, Theorem 23.9], we have

$$\begin{aligned}\frac{d}{dh}\mathrm{MCov}(\rho_h, \mu_{t+h}) &= \int \phi_{\rho_h}^{\mu_{t+h}} \, \partial_h\rho_h \, d\lambda + \int \phi_{\mu_{t+h}}^{\rho_h} \, \partial_h \mu_{t+h} \, d\lambda \\ &= \int \phi_{\rho_h}^{\mu_{t+h}} \, \tfrac{1}{2}\Delta\rho_h \, d\lambda + \int \phi_{\mu_{t+h}}^{\rho_h} \, \tfrac{1}{2}\sum_{i,k}\partial^2_{ik}(\Sigma_{ik}\mu_{t+h}) \, d\lambda \\ &= \tfrac{1}{2}\int \sum_{i,k}\partial^2_{i,k}(\phi_{\rho_h}^{\mu_{t+h}}) \, I_{ik} \, d\rho_h + \tfrac{1}{2}\int \sum_{i,k}\partial^2_{i,k}(\phi_{\mu_{t+h}}^{\rho_h}) \, \Sigma_{ik} \, d\mu_{t+h} \\ &= \tfrac{1}{2}\int \mathrm{tr}\big(D^2(\phi_{\rho_h}^{\mu_{t+h}})\big) \, \rho_h \, d\lambda + \tfrac{1}{2}\int \mathrm{tr}\big(D^2(\phi_{\mu_{t+h}}^{\rho_h})\Sigma\big) \, \mu_{t+h} \, d\lambda,\end{aligned}$$

where we denote by D and D^2 the Jacobian and Hessian matrix, respectively. During this proof we will use the convention that if $x \mapsto a(x) \in \mathbb{R}^d$ is an invertible vector-valued function, then $a^{-1}(x)$ denotes the inverse function, whereas if $x \mapsto A(x) \in \mathbb{R}^{d\times d}$ is a matrix-valued function, then $[A(x)]^{-1}$ denotes the matrix inverse of $A(x)$. Now observe that

$$D^2(\phi_{\rho_h}^{\mu_{t+h}})(x) = D(\nabla\phi_{\rho_h}^{\mu_{t+h}})(x) = D\big((\nabla\phi_{\mu_{t+h}}^{\rho_h})^{-1}\big)(x) = [D\nabla\phi_{\mu_{t+h}}^{\rho_h} \circ \nabla\phi_{\rho_h}^{\mu_{t+h}}(x)]^{-1},$$

so that

$$\begin{aligned}\int \mathrm{tr}\big(D^2(\phi_{\rho_h}^{\mu_{t+h}})(x)\big) \, \rho_h \, d\lambda &= \int \mathrm{tr}\big([D\nabla\phi_{\mu_{t+h}}^{\rho_h} \circ \nabla\phi_{\rho_h}^{\mu_{t+h}}(x)]^{-1}\big) \, \rho_h \, d\lambda \\ &= \int \mathrm{tr}\big([D^2\phi_{\mu_{t+h}}^{\rho_h}(y)]^{-1}\big) \, \mu_{t+h} \, d\lambda.\end{aligned}$$

Altogether we have

$$\frac{d}{dh}\mathrm{MCov}(\rho_h, \mu_{t+h}) = \tfrac{1}{2}\int \Big(\mathrm{tr}\big([D^2\phi_{\mu_{t+h}}^{\rho_h}]^{-1}\big) + \mathrm{tr}\big(D^2(\phi_{\mu_{t+h}}^{\rho_h})\Sigma\big)\Big) \, \mu_{t+h} \, d\lambda.$$

Define now the functional on invertible, positive-semidefinite symmetric matrices

$$A \mapsto J(A) := \operatorname{tr}(A^{-1}) + \operatorname{tr}(A\Sigma).$$

We remark that $J(A) \geqslant 2\operatorname{tr}(\Sigma^{1/2})$, since this is equivalent to the trivial statement

$$|A^{-1/2} - A^{1/2}\Sigma^{1/2}|_{\mathrm{HS}} \geqslant 0.$$

Hence in fact the minimum of $J(\,\cdot\,)$ is attained at $A = \Sigma^{-1/2}$. We conclude that

$$\frac{d}{dh}\Big|_{h=0} \operatorname{MCov}(\rho_h, \mu_{t+h}) \geqslant \int \operatorname{tr}\big(\sigma_t(y)\big)\,\mu_t(dy) = \mathbb{E}\big[\operatorname{tr}\big(\sigma_t(M_t)\big)\big],$$

which completes the proof of Theorem 1.6. □

5. Displacement convexity of the Bass functional

We observe that the Bass functional $\alpha \mapsto \mathcal{V}(\alpha)$ provides a novel example of a convex functional with respect to the almost-Riemannian structure of the quadratic Wasserstein space $\mathcal{P}_2$. As mentioned in [43, Open Problem 5.17], there are only few known examples of so-called displacement convex functionals (see [43, Definition 5.10], [2, Definition 9.1.1], [37]), and it is desirable to find new ones.

We shall state two versions of this result. The first one, Proposition 5.1, pertains to the case $d = 1$, while the second one, Proposition 5.2, holds for general $d \in \mathbb{N}$. We also note that, contrary to the rest of this paper, we do not assume that $\mu \preceq_{\mathrm{c}} \nu$.

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

$$\mathcal{P}_2(\mathbb{R}) \ni \alpha \longmapsto \mathcal{V}(\alpha) = \operatorname{MCov}(\alpha * \gamma, \nu) - \operatorname{MCov}(\alpha, \mu) \tag{5.1}$$

is displacement convex. Moreover, if a geodesic $(\alpha_u)_{0\leqslant u\leqslant 1}$ in $\mathcal{P}_2(\mathbb{R})$ is such that α_1 is not a translate of α_0 and if ν is not a Dirac measure, the function $u \mapsto \mathcal{V}(\alpha_u)$ is strictly convex.

Proof. We start by noting that the Bass functional $\mathcal{V}(\,\cdot\,)$ of (5.1) can equivalently be defined in terms of the quadratic Wasserstein distance $\mathcal{W}_2(\,\cdot\,,\,\cdot\,)$ of (2.3) rather than in terms of the maximal covariance $\operatorname{MCov}(\,\cdot\,,\,\cdot\,)$ of (1.4). Indeed, we have the identity

$$\mathcal{V}(\alpha) = \operatorname{MCov}(\alpha * \gamma, \nu) - \operatorname{MCov}(\alpha, \mu) = \tfrac{1}{2}\mathcal{W}_2^2(\alpha, \mu) - \tfrac{1}{2}\mathcal{W}_2^2(\alpha * \gamma, \nu) + \text{const},$$

where the constant

$$\text{const} = \tfrac{d}{2} + \tfrac{1}{2}\int |y|^2\, d\nu(y) - \tfrac{1}{2}\int |x|^2\, d\mu(x)$$

does not depend on α. Therefore showing the (strict) displacement convexity of the Bass functional $\mathcal{V}(\,\cdot\,)$ is equivalent to showing the (strict) displacement convexity of the functional

$$\mathcal{U}(\alpha) := \mathcal{W}_2^2(\alpha, \mu) - \mathcal{W}_2^2(\alpha * \gamma, \nu). \tag{5.2}$$

Fix $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$ and let $(\alpha_u)_{0\leqslant u\leqslant 1}$ be a geodesic in the quadratic Wasserstein space $\mathcal{P}_2(\mathbb{R})$. Using the hypothesis $d = 1$ we can choose mutually comonotone random variables Z_0, Z_1 and X with laws α_0, α_1 and μ, respectively. As $(\alpha_u)_{0\leqslant u\leqslant 1}$ is a geodesic, the random variable $Z_u := (1-u)Z_0 + uZ_1$ has law α_u, for $0 \leqslant u \leqslant 1$. Also note that each Z_u is comonotone with X. Let $u_0, u \in [0,1]$. As regards the first Wasserstein distance in (5.2), a straightforward calculation yields

$$\begin{aligned} \mathcal{W}_2^2(\alpha_u, \mu) - \mathcal{W}_2^2(\alpha_{u_0}, \mu) &= \qquad\qquad &(5.3)\\ &= \mathbb{E}[|Z_u - X|^2] - \mathbb{E}[|Z_{u_0} - X|^2] &(5.4)\\ &= \mathbb{E}[|Z_u - Z_{u_0}|^2] - 2\mathbb{E}[\langle Z_u - Z_{u_0}, X - Z_{u_0}\rangle] &(5.5)\\ &= (u-u_0)^2\mathbb{E}[|Z_1 - Z_0|^2] - 2(u-u_0)\mathbb{E}[\langle Z_1 - Z_0, X - Z_{u_0}\rangle]. &(5.6) \end{aligned}$$

Passing to the second Wasserstein distance in (5.2), we take a standard Gaussian random variable Γ on $\mathbb{R}$, independent of Z_0 as well as of Z_1. Next we choose a random variable

Y_{u_0} such that $(Z_{u_0}+\Gamma, Y_{u_0})$ is an optimal coupling of $(\alpha_{u_0} * \gamma, \nu)$. As $(Z_u * \Gamma, Y_{u_0})$ is a (typically sub-optimal) coupling of $(\alpha_u * \gamma, \nu)$, we obtain the inequality

$$\begin{aligned}
&\mathcal{W}_2^2(\alpha_u * \gamma, \nu) - \mathcal{W}_2^2(\alpha_{u_0} * \gamma, \nu) \leqslant && (5.7)\\
&\qquad \leqslant \mathbb{E}[|Z_u + \Gamma - Y_{u_0}|^2] - \mathbb{E}[|Z_{u_0} + \Gamma - Y_{u_0}|^2] && (5.8)\\
&\qquad = \mathbb{E}[|Z_u - Z_{u_0}|^2] - 2\mathbb{E}[\langle Z_u - Z_{u_0}, Y_{u_0} - (Z_{u_0} + \Gamma)\rangle] && (5.9)\\
&\qquad = (u-u_0)^2\mathbb{E}[|Z_1 - Z_0|^2] - 2(u-u_0)\mathbb{E}[\langle Z_1 - Z_0, Y_{u_0} - (Z_{u_0}+\Gamma)\rangle]. && (5.10)
\end{aligned}$$

Combining (5.3) – (5.10), we deduce that

$$\begin{aligned}
&\mathcal{U}(\alpha_u) - \mathcal{U}(\alpha_{u_0}) = && (5.11)\\
&\quad = \Big(\mathcal{W}_2^2(\alpha_u, \mu) - \mathcal{W}_2^2(\alpha_u * \gamma, \nu)\Big) - \Big(\mathcal{W}_2^2(\alpha_{u_0}, \mu) - \mathcal{W}_2^2(\alpha_{u_0} * \gamma, \nu)\Big) \geqslant && (5.12)\\
&\quad \geqslant 2(u-u_0)\mathbb{E}[\langle Z_1 - Z_0, Y_{u_0} - X - \Gamma\rangle] && (5.13)\\
&\quad = 2(u-u_0)\mathbb{E}[\langle Z_1 - Z_0, Y_{u_0} - X\rangle], && (5.14)
\end{aligned}$$

where the last equation follows from conditioning on Z_0, Z_1. The expression in (5.14) defines a linear function in u, which lies below and touches the function

$$\begin{aligned}
u \longmapsto\ & \mathcal{U}(\alpha_u) - \mathcal{U}(\alpha_{u_0}) = \\
&= \Big(\mathcal{W}_2^2(\alpha_u, \mu) - \mathcal{W}_2^2(\alpha_u * \gamma, \nu)\Big) - \Big(\mathcal{W}_2^2(\alpha_{u_0}, \mu) - \mathcal{W}_2^2(\alpha_{u_0} * \gamma, \nu)\Big)
\end{aligned}$$

at the point $u = u_0$. This readily implies the convexity of the function

$$u \longmapsto \mathcal{U}(\alpha_u) = \mathcal{W}_2^2(\alpha_u, \mu) - \mathcal{W}_2^2(\alpha_u * \gamma, \nu).$$

It remains to show the strict convexity assertion of Proposition 5.1. If α_1 is not a translate of α_0, then α_u is not a translate of α_{u_0} either, provided that $u \neq u_0$. As $Z_{u_0} + \Gamma$ is comonotone with Y_{u_0} and Y_{u_0} is assumed to be non-constant, we may find $y_0 \in \mathbb{R}$ and $z_0 \in \mathbb{R}$ such that $\mathbb{P}[Y_{u_0} < y_0] \in (0,1)$ and

$$\{Z_{u_0} + \Gamma < z_0\} = \{Y_{u_0} < y_0\}.$$

If $Z_u + \Gamma$ were also comonotone with Y_{u_0}, we could find $z \in \mathbb{R}$ such that

$$\{Z_{u_0} + \Gamma < z_0\} = \{Y_{u_0} < y_0\} = \{Z_u + \Gamma < z\},$$

where we have used that the law of $Z_u + \Gamma$ is continuous. Conditioning on $\Gamma = \zeta$ this implies that, for Lebesgue-a.e. $\zeta \in \mathbb{R}$,

$$\{Z_{u_0} < z_0 - \zeta\} = \{Z_u < z - \zeta\},$$

so that Z_{u_0} and Z_u are translates. This gives the desired contradiction, showing that there is a strict inequality in (5.7), (5.8) (thus also in (5.12), (5.13)), which implies the strict convexity assertion of Proposition 5.1. □

We now pass to the case of general $d \in \mathbb{N}$. In Proposition 5.2 below we formulate a convexity property of the Bass functional $\mathcal{V}(\,\cdot\,)$ pertaining to the notion of *generalized geodesics* as analyzed in [2, Definition 9.2.2]. Recall that $(\alpha_u)_{0 \leqslant u \leqslant 1}$ is a generalized geodesic with base μ, joining α_0 to α_1, if there are random variables Z_0, Z_1 and X with laws α_0, α_1 and μ, respectively, such that (Z_0, X) and (Z_1, X) are optimal couplings and such that the random variable $Z_u := uZ_1 + (1-u)Z_0$ has law α_u, for $0 \leqslant u \leqslant 1$.

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

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

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

We do not know whether the above assertion is also true along (non generalized) geodesics $(\alpha_u)_{0 \leqslant u \leqslant 1}$ in $\mathcal{P}_2(\mathbb{R}^d)$, when $d > 1$.

Proof of Proposition 5.2. We follow the lines of the proof of Proposition 5.1 and consider again the functional

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

as in (5.2). Let $(\alpha_u)_{0\leqslant u\leqslant 1}$ be a generalized geodesic with base μ, joining α_0 to α_1. Take Z_0, Z_1, Z_u, X as above such that (Z_0, X) and (Z_1, X) are optimal couplings and by definition $Z_u \sim \alpha_u$. Note that (Z_u, X) is an optimal coupling of (α_u, μ) by [2, Lemma 9.2.1], for $0 \leqslant u \leqslant 1$. The equalities (5.3) – (5.6) and the inequality in (5.7) – (5.10) then carry over verbatim and we again arrive at (5.12) – (5.14), which shows the convexity of the function $[0, 1] \ni u \mapsto \mathcal{U}(\alpha_u)$. ◻

References

[1] L. Ambrosio and N. Gigli. A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, volume 2062 of *Lecture Notes in Math.*, pages 1–155. Springer, Heidelberg, 2013.

[2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[3] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. Adapted Wasserstein distances and stability in mathematical finance. *Finance Stoch.*, 24(3):601–632, 2020.

[4] J. Backhoff-Veraguas, M. Beiglböck, M. Huesmann, and S. Källblad. Martingale Benamou-Brenier: a probabilistic perspective. *Ann. Probab.*, 48(5):2258–2289, 2020.

[5] J. Backhoff-Veraguas, M. Beiglböck, W. Schachermayer, and B. Tschiderer. The structure of martingale Benamou–Brenier in $\mathbb{R}^d$. arXiv:12306.11019, 2023.

[6] R. Bass. Skorokhod embedding via stochastic integrals. In J. Azéma and M. Yor, editors, *Séminaire de Probabilités XVII 1981/82*, number 986 in Lecture Notes in Mathematics, pages 221–224. Springer, 1983.

[7] M. Beiglböck, A. Cox, and M. Huesmann. Optimal transport and Skorokhod embedding. *Invent. Math.*, 208(2):327–400, 2017.

[8] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices: A mass transport approach. *Finance Stoch.*, 17(3):477–501, 2013.

[9] M. Beiglböck, P. Henry-Labordère, and N. Touzi. Monotone martingale transport plans and Skorokhod embedding. *Stochastic Process. Appl.*, 127(9):3005–3013, 2017.

[10] M. Beiglböck, B. Jourdain, W. Margheriti, and G. Pammer. Monotonicity and stability of the weak martingale optimal transport problem. arXiv:2109.06322, 2022.

[11] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.*, 44(1):42–106, 2016.

[12] M. Beiglböck, M. Nutz, and F. Stebegg. Fine properties of the optimal Skorokhod embedding problem. *J. Eur. Math. Soc. (JEMS)*, 24(4):1389–1429, 2022.

[13] J. Benamou and Y. Brenier. A numerical method for the optimal time-continuous mass transport problem and related problems. In *Monge Ampère equation: applications to geometry and optimization (Deerfield Beach, FL, 1997)*, volume 226 of *Contemp. Math.*, pages 1–11. Amer. Math. Soc., Providence, RI, 1999.

[14] B. Bouchard and M. Nutz. Arbitrage and duality in nondominated discrete-time models. *The Annals of Applied Probability*, 25(2):823–859, 2015.

[15] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.*, 305(19):805–808, 1987.

[16] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[17] L. Campi, I. Laachir, and C. Martini. Change of numeraire in the two-marginals martingale transport problem. *Finance Stoch.*, 21(2):471–486, June 2017.

[18] P. Cheridito, M. Kiiski, D. J. Prömel, and H. M. Soner. Martingale optimal transport duality. *Math. Ann.*, 379(3-4):1685–1712, 2021.

[19] A. M. G. Cox, J. Obł ój, and N. Touzi. The Root solution to the multi-marginal embedding problem: an optimal stopping and time-reversal approach. *Probab. Theory Related Fields*, 173(1-2):211–259, 2019.

[20] H. De March and N. Touzi. Irreducible convex paving for decomposition of multidimensional martingale transport plans. *Ann. Probab.*, 47(3):1726–1774, 2019.

[21] Y. Dolinsky and H. M. Soner. Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Relat. Fields*, 160(1-2):391–427, 2014.

[22] H. Föllmer. Optimal couplings on Wiener space and an extension of Talagrand's transport inequality. In *Stochastic analysis, filtering, and stochastic optimization*, pages 147–175. Springer, Cham, 2022.

[23] A. Galichon, P. Henry-Labordère, and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.*, 24(1):312–336, 2014.

[24] N. Ghoussoub, Y.-H. Kim, and T. Lim. Structure of optimal martingale transport plans in general dimensions. *Ann. Probab.*, 47(1):109–164, 2019.

[25] N. Ghoussoub, Y.-H. Kim, and T. Lim. Optimal Brownian stopping when the source and target are radially symmetric distributions. *SIAM J. Control Optim.*, 58(5):2765–2789, 2020.
[26] N. Ghoussoub, Y.-H. Kim, and A. Z. Palmer. PDE methods for optimal Skorokhod embeddings. *Calc. Var. Partial Differential Equations*, 58(3):Paper No. 113, 31, 2019.
[27] I. Guo and G. Loeper. Path dependent optimal transport and model calibration on exotic derivatives. *The Annals of Applied Probability*, 31(3):1232–1263, 2021.
[28] I. Guo, G. Loeper, and S. Wang. Local volatility calibration by optimal transport. In *2017 MATRIX Annals*, pages 51–64. Springer, 2019.
[29] P. Henry-Labordère, J. Obłój, P. Spoida, and N. Touzi. The maximum maximum of a martingale with given n marginals. *Ann. Appl. Probab.*, 26(1):1–44, 2016.
[30] D. Hobson and A. Neuberger. Robust bounds for forward start options. *Math. Finance*, 22(1):31–56, 2012.
[31] M. Huesmann and D. Trevisan. A Benamou-Brenier formulation of martingale optimal transport. *Bernoulli*, 25(4A):2729–2757, 2019.
[32] S. Källblad, X. Tan, and N. Touzi. Optimal Skorokhod embedding given full marginals and Azéma-Yor peacocks. *Ann. Appl. Probab.*, 27(2):686–719, 2017.
[33] L. Kantorovich. On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.
[34] G. Loeper. Option pricing with linear market impact and nonlinear black–scholes equations. *The Annals of Applied Probability*, 28(5):2664–2726, 2018.
[35] R. McCann. A convexity theory for interacting gases and equilibrum crystals. *PhD thesis, Princeton University*, 1994.
[36] R. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80(2):309–323, 1995.
[37] R. J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.
[38] G. Monge. Memoire sur la theorie des deblais et des remblais. *Histoire de l'académie Royale des Sciences de Paris*, 1781.
[39] J. Obłój and P. Siorpaes. Structure of martingale transports in finite dimensions. arXiv:1702.08433, 2017.
[40] J. Obłój, P. Spoida, and N. Touzi. Martingale inequalities for the maximum via pathwise arguments. In *In Memoriam Marc Yor-Séminaire de Probabilités XLVII*, pages 227–247. Springer, 2015.
[41] F. Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.
[42] X. Tan and N. Touzi. Optimal transportation under controlled stochastic dynamics. *Ann. Probab.*, 41(5):3201–3240, 2013.
[43] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
[44] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.

Faculty of Mathematics, University of Vienna
Email address: julio.backhoff@univie.ac.at

Faculty of Mathematics, University of Vienna
Email address: walter.schachermayer@univie.ac.at

Faculty of Mathematics, University of Vienna
Email address: bertram.tschiderer@univie.ac.at