The Gradient Flow of the Bass Functional in Martingale Optimal Transport

Julio Backhoff-Veraguas, Gudmund Pammer, Walter Schachermayer

August 8, 2024

Abstract

Given μ and ν, probability measures on $\mathbb{R}^d$ in convex order, a Bass martingale is arguably the most natural martingale starting with law μ and finishing with law ν. Indeed, this martingale is obtained by *stretching* a reference Brownian motion so as to meet the data μ, ν. Unless μ is a Dirac, the existence of a Bass martingale is a delicate subject, since for instance the reference Brownian motion must be allowed to have a non-trivial initial distribution α, not known in advance. Thus the key to obtaining the Bass martingale, theoretically as well as practically, lies in finding α.

In [7] it has been shown that α is determined as the minimizer of the so-called *Bass functional*. In the present paper we propose to minimize this functional by following its gradient flow, or more precisely, the gradient flow of its L^2-lift. In our main result we show that this gradient flow converges in norm to a minimizer of the Bass functional, and when $d = 1$ we further establish that convergence is exponentially fast.

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 $\mu \leq_{cx} \nu$, 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 a d-dimensional Brownian motion 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 [16], see [4, 5]. The problem (MBB) can be shown to be equivalent to

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

Indeed their optimizer are identical and we have $MT(\mu, \nu) = -2P(\mu, \nu) + \int |x|^2 d(\nu - \mu) + d$.

As shown in [4], the problem (MBB) admits a strong Markov martingale $\hat{M}$ as the 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 [8, 4, 5].

Definition 1.2. *Let $B = (B_t)_{0 \leqslant t \leqslant 1}$ be a 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*

$$\hat{M}_t := \mathbb{E}[\nabla \hat{v}(B_1) \,|\, \sigma(B_s \colon s \leqslant t)] = \mathbb{E}[\nabla \hat{v}(B_1) \,|\, B_t], \qquad 0 \leqslant t \leqslant 1 \tag{2}$$

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

The reason behind this terminology is that Bass [8] 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 $\mu \leqslant_{cx} \nu$ 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 a 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 $\mu \leq_{cx} \nu$, a 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[1] and by γ_1 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 (see [5])

$$(\nabla \hat{v} * \gamma_1)(\hat{\alpha}) = \mu \qquad \text{and} \qquad \nabla \hat{v}(\hat{\alpha} * \gamma_1) = \nu. \tag{3}$$

In terms of (2) this amounts to

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

where γ_s denotes the centred Gaussian with covariance matrix $s I_d$.

1.3 The Bass functional

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

$$\mathrm{MCov}(p_1, p_2) := \sup_{q \in \mathrm{Cpl}(p_1, p_2)} \int \langle x_1, x_2 \rangle \, q(dx_1, dx_2), \tag{4}$$

where $\mathrm{Cpl}(\mu, \nu)$ 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 Wasserstein distance. We follow [7] in defining:

Definition 1.4. *The Bass functional is given by*

$$\mathcal{P}_2(\mathbb{R}^d) \ni \alpha \longmapsto \mathcal{V}(\alpha) := \mathrm{MCov}(\alpha * \gamma_1, \nu) - \mathrm{MCov}(\alpha, \mu). \tag{5}$$

If $\mu \leq_c \nu$ we have $\mathcal{V}(\alpha) \geq 0$ for every $\alpha \in \mathcal{P}_2(\mathbb{R}^d)$.

The main result of [7] was the reformulation of Problem (1), which characterizes the Bass measure $\hat{\alpha}$ in (3) as the optimizer of the Bass functional (5). This is the content of [7, Theorem 1.5]:

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

$$P(\mu, \nu) = \inf_{\alpha \in \mathcal{P}_2(\mathbb{R}^d)} \mathcal{V}(\alpha). \tag{6}$$

The right-hand side of (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 $\hat{\alpha} \in \mathcal{P}_2(\mathbb{R}^d)$.

---

[1]By a minor abuse of notation we write $f * \gamma_1$ for the convolution of a measurable map $f \colon \mathbb{R}^d \to \mathbb{R}^d$ with the density of the Gaussian γ_1 and similarly $\alpha * \gamma_1$ for the convolution of a measure $\alpha \in \mathcal{P}(\mathbb{R}^d)$ with the Gaussian measure γ_1.

1.4 Main Results

Per Theorem 1.5, it is desirable to find the minimizer of the Bass functional. Indeed, doing so will provide us with the Bass measure and hence with the Bass martingale between μ and ν. Moreover, one would like to have a *fast* minimization method for the Bass functional. One idea would be to perform the gradient descent of the Bass functional in the sense of Wasserstein spaces. This seems challenging to the authors, since not much is known about the convexity of the Bass functional in the Wasserstein geometry for d arbitrary (see however [7, Theorem 1.7]). For this reason we take a detour.

We fix a probability space supporting independent random variables $X \sim \mu$ and $\Gamma \sim \gamma_1$, which we fix throughout. We then define the sigma-algebras $\mathcal{G} := \sigma(X)$ and $\mathcal{H} := \sigma(X, \Gamma)$, with their associated L^2-spaces of $\mathbb{R}^d$-valued random variables, respectively $L^2(\mathcal{H}; \mathbb{R}^d)$ and $L^2(\mathcal{G}; \mathbb{R}^d)$.

Definition 1.6. *The lifted Bass functional is given by*

$$L^2(\mathcal{G}; \mathbb{R}^d) \ni Z \mapsto V(Z) := \sup_{\substack{Y \in L^2(\mathcal{H}; \mathbb{R}^d), \\ Y \sim \nu}} \mathbb{E}[(Z + \Gamma) \cdot Y - Z \cdot X]. \tag{7}$$

As we show in Lemma 2.1 below, minimizing $\mathcal{V}$ or V lead to the same problem, and in fact if Z^* minimizes V then $\mathscr{L}(Z^*)$ is a Bass measure. Furthermore, we shall show that V is convex in the conventional sense, Fréchet differentiable, and admits the explicit L^2-gradient,

$$D_Z V = \nabla v * \gamma_1(Z) - X,$$

where ∇v is the Brenier map[2] from $\mathscr{L}(Z) * \gamma_1$ to ν. We can hence define the L^2-Bass gradient flow as follows:

Definition 1.7. *An absolutely continuous curve $\{Z_t\}_{t \in \mathbb{R}_+} \subset L^2(\mathcal{G}; \mathbb{R}^d)$ is called an L^2-Bass gradient flow (started at Z_0) if the curve satisfies for a.e. $t \in \mathbb{R}_+$ the equality*

$$\frac{dZ_t}{dt} = -[\nabla v_t * \gamma_1(Z) - X],$$

where ∇v_t is the Brenier map from $\mathscr{L}(Z_t) * \gamma_1$ to ν.

Invoking the classical theory of gradient flows of convex functionals in Hilbert spaces, one could hope to obtain the existence and uniqueness of the Bass gradient flow, as well as the weak-L^2 convergence of the flow to a minimizer of V. However the story is not that simple, since there is in general no reason for V to attain its minimum at a square integrable random variable. Moreover weak-L^2 convergence of the Bass flow is not strong enough to derive convergence of $\mathscr{L}(Z_t)$ in a meaningful sense, and the abstract theory would not give us convergence rates. For these reasons we need to make further assumptions and argue in a self-contained fashion.

---

[2]The Brenier map from an absolutely continuous measure ρ, to η, is the unique almost-sure gradient of a convex function v satisfying the push-forward constraint $(\nabla v)(\rho) = \eta$, while v is called the Brenier potential.

Under technical assumptions, defined in Section 3.1 below,[3] we obtain our first main result:

Theorem 1.8. *Under Assumption (A), for any r.v. $Z_0 \in L^2(\mathcal{G}; \mathbb{R}^d)$, the L^2-Bass gradient flow $(Z_t)_{t \geq 0}$ exists, is unique, and it converges in L^2-norm to $Z^\star \in L^\infty(\mathcal{G}; \mathbb{R}^d)$, where the latter is the unique minimizer of V with $bary(Z^\star) = bary(Z_0)$. Further, if Z_0 is bounded, then the entire curve $(Z_t)_{t \geq 0}$ is uniformly bounded.*

In our second main result, we sharpen the above by obtaining exponential convergence in the one-dimensional case:

Theorem 1.9. *Let $d = 1$. In the setting of Theorem 1.8, assume additionally that $Z_0 \in L^\infty(\mathcal{G}; \mathbb{R}^d)$, and that ν has convex support and admits a density which is bounded away from zero and infinity on its support. Then we have for some $\kappa > 0$:*

$$\|Z_t - Z^\star\|_{L^2} \leq \|Z_0 - Z^\star\|_{L^2} \cdot \exp\{-\kappa t\}, \tag{8}$$
$$V(Z_t) - \inf V \leq [V(Z_0) - \inf V] \cdot \exp\{-\kappa t\}. \tag{9}$$

In order to obtain the latter result a careful study of the Hessian of V must be carried out, which we deem of independent interest. A corollary of this analysis is the proof that V is strongly convex when we restrict it to random variables which are uniformly bounded. We conjecture the validity of Theorem 1.9 also in higher dimensions. We leave the latter as well as the practical implementation of the Bass gradient flow as an interesting research topic, the exponential convergence result suggesting a fast computational method.

1.5   Related literature

Optimal transport as a field in mathematics goes back to Monge [46] and Kantorovich [42], who established its modern formulation. The seminal results of Benamou, Brenier, and McCann [19, 20, 16, 44, 45] form the basis of the modern theory, with striking applications in a variety of different areas, see the monographs [52, 53, 2, 49].

We are interested in transport problems where the transport plan satisfies an additional martingale constraint. This additional requirement arises naturally in finance (e.g. [11]), but is also of independent mathematical interest. For example there are notable consequences for the study of martingale inequalities (e.g. [17, 36, 48]) and the Skorokhod embedding problem (e.g. [10, 41, 14]). Early articles on this topic of *martingale optimal transport* include [37, 11, 50, 30, 28, 23]. The study of irreducibility of a pair of marginals (μ, ν) was initiated by Beiglböck and Juillet [13] in dimension one and extended in the works [31, 27, 47] to multiple dimensions.

Continuous-time martingale optimal transport problems have received much attention in the recent years; see e.g. [12, 26, 33, 35, 32, 24, 34]. In this paper we concern ourselves with

---

[3]In short: the pair (μ, ν) is irreducible, the convex hull of the support of ν, denoted by C, is compact, and the convex hull of the support of μ is compactly contained in the interior of C.

the specific structure given by the martingale Benamou–Brenier problem, introduced in [4] in probabilistic language and in [38] 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 [43]. 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, 29]) subject to the given marginal constraints.

The connection between the martingale Benamou–Brenier problem and Bass martingales was revealed in [4, 5]. If $d = 1$ then an algorithm for obtaining a Bass martingale with given marginals was proposed by Conze and Henry-Labordere in [25]. This algorithm was subsequently studied in detail in [1], and its multidimensional form has been proposed in [39]. Recently, objects akin to Bass martingales have been proposed in [51], [18], and [6, 15].

2 Regularity of the lifted Bass functional

Throughout this paper, let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$. We recall from (5) the associated Bass functional $\mathcal{V} \colon \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ given by

$$\mathcal{V}(\alpha) := \mathrm{MCov}(\alpha * \gamma_1, \nu) - \mathrm{MCov}(\alpha, \mu). \tag{10}$$

Assuming that μ and ν are in convex order ($\mu \leq_{cx} \nu$), this functional takes its values in $[0, \infty)$.

In a first step, we will identify $\mathcal{V}$ with a convex, Fréchet differentiable functional V on a Hilbert space. To this end, we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a random variable $X \sim \mu$ and an independent random variable $\Gamma \sim \gamma_1$ where γ_1 is here the d-dimensional standard Gaussian. Define the sigma-algebras

$$\mathcal{G} := \sigma(X) \text{ and } \mathcal{H} := \sigma(X, \Gamma).$$

We define $V \colon L^2(\mathcal{G}; \mathbb{R}^d) \to \mathbb{R}$ by

$$V(Z) := \sup_{\substack{Y \in L^2(\mathcal{H}; \mathbb{R}^d), \\ Y \sim \nu}} \mathbb{E}[(Z + \Gamma) \cdot Y - Z \cdot X] \tag{11}$$

Lemma 2.1. *Let $Z \in L^2(\mathcal{G}; \mathbb{R}^d)$. The functional V given in (11) enjoys the following properties:*

(i) V is Fréchet differentiable with derivative $D_Z V = \nabla v * \gamma_1(Z) - X \in L^2(\mathcal{G}; \mathbb{R}^d)$ where v is the Brenier potential from $\mathscr{L}(Z + B)$ to ν.

(ii) If the law of (X, Z) is monotone,[4] then $V(Z) = \mathcal{V}(\mathscr{L}(Z))$, else $V(Z) \geq \mathcal{V}(\mathscr{L}(Z))$.

---

[4]Two $\mathbb{R}^d$-valued random variables X and Z with finite second moment are coupled monotonically if $\mathbb{E}[X \cdot Z] = \mathrm{MCov}(\mathscr{L}(X), \mathscr{L}(Z))$.

(iii) V is convex and weakly lower semicontinuous.

(iv) If $\mathcal{V}$ admits an optimizer in $\mathcal{P}_2(\mathbb{R}^d)$, then for every $m \in \mathbb{R}^d$ there is a unique $Z^\star \in L^2(\mathcal{G}; \mathbb{R}^d)$ with $\mathbb{E}[Z^\star] = m$, $V(Z^\star) = \min_{\alpha \in \mathcal{P}_2(\mathbb{R}^d)} \mathcal{V}(\alpha)$, and $(X, Z^\star)$ monotone.

Proof. (i): Let $(Z_t)_{t\in[0,1]}$ be a continuous curve in $L^2(\mathcal{G}; \mathbb{R}^d)$. Since $\mathcal{L}(Z_t + \Gamma) \in \mathcal{P}_2(\mathbb{R}^d)$ is absolutely continuous w.r.t. the Lebesgue measure, there is by Brenier's theorem a convex function $v_t \colon \mathbb{R}^d \to \mathbb{R}$ with $\nabla v_t(Z_t + \Gamma) \sim \nu$. In particular, the coupling

$$\pi^t := \mathcal{L}(Z_t + \Gamma, \nabla v_t(Z_t + \Gamma))$$

is optimal for $\mathrm{MCov}(\beta^t, \nu)$ where $\beta^t := \mathcal{L}(Z_t + \Gamma) = \mathcal{L}(Z_t) * \gamma_1$. Therefore, we deduce from stability of optimal transport maps (see, for example, [53, Theorem 5.20]) and uniqueness of the optimizer to the optimization problem $\mathrm{MCov}(\beta^t, \nu)$, that $\pi^t \to \pi^0$ in distribution for $t \to 0$. At the same time, it is straightforward to see that $\beta^t \to \beta^0$ in total variation. Therefore all requirements of [40, Lemma A.2] are met and we obtain that $\nabla v_t \to \nabla v_0$ in β^0-probability. We want to use the above to show that $\nabla v_t(Z_t + \Gamma) \to \nabla v_0(Z_0 + \Gamma)$ in probability. To this end, fix $\varepsilon > 0$ and note that

$$\mathbb{P}\left(|\nabla v_t(Z_t + \Gamma) - \nabla v_0(Z_0 + \Gamma)| \geq 3\varepsilon\right)$$
$$\leq \underbrace{\mathbb{P}\left(|\nabla v_t(Z_t + \Gamma) - \nabla v_0(Z_t + \Gamma)| \geq \varepsilon\right)}_{=:A(t)} + \underbrace{\mathbb{P}\left(|\nabla v_0(Z_t + \Gamma) - \nabla v_0(Z_0 + \Gamma)| \geq 2\varepsilon\right)}_{=:B(t)}.$$

To deal with the first term of the right-hand side, observe that

$$A(t) \leq \mathbb{P}\left(|\nabla v_t(Z_0 + \Gamma) - \nabla v_0(Z_0 + \Gamma)| \geq \varepsilon\right) + \mathrm{TV}(\beta^t, \beta^0),$$

where both terms of the right-hand side vanish be our previous arguments. In order to handle the second term, we use that $\hat{\pi}^t := \mathcal{L}(Z_t + \Gamma, \nabla v_0(Z_t + \Gamma))$ is an optimal coupling with marginals β^t and $\nu^t := (\nabla v_0)(\beta^t)$. Since $\beta^t \to \beta^0$ in total variation, the same holds true for the convergence of $\nu^t \to \nu^0 = \nu$. As before, using stability of optimal transport maps, we find that $\hat{\pi}^t \to \hat{\pi}^0 = \pi^0$ in distribution. Hence, we can assume w.l.o.g. that there is $\hat{Z}_t \sim \beta^0$ with

$$|Z_t + \Gamma - \hat{Z}_t| + |\nabla v_0(Z_t + \Gamma) - \nabla v_0(\hat{Z}_t))| \to 0 \tag{12}$$

in probability. Consequently,

$$B(t) \leq \mathbb{P}\left(|\nabla v_0(Z_t + \Gamma) - \nabla v_0(\hat{Z}_t)| \geq \varepsilon\right) + \mathbb{P}\left(|\nabla v_0(\hat{Z}_t) - \nabla v_0(Z_0 + \Gamma)| \geq \varepsilon\right).$$

Here, letting $t \to 0$, the first term vanishes on the right-hand side vanishes by (12) while the second term vanishes due to [40, Lemma A.1]. We have shown that $A(t) + B(t) \to 0$ as $t \to 0$, which yields the claim that $\nabla v_t(Z_t + \Gamma) \to \nabla v_0(Z_0 + \Gamma)$ in probability. Since the family $(\|\nabla v_t(Z_t + \Gamma)\|_{L^2}^2)_{t\in[0,1]}$ is uniformly integrable as $\nabla v_t(Z_t + \Gamma) \sim \nu$, we have shown that

$$\nabla v_t(Z_t + \Gamma) \to \nabla v_0(Z_0 + \Gamma) \quad \text{in } L^2(\mathcal{H}; \mathbb{R}^d). \tag{13}$$

Next, let $Z_1 \in L^2(\mathcal{G}; \mathbb{R}^d)$ and consider in the particular curve $(Z_t)_{t \in [0,1]}$ with

$$Z_t := (1-t)Z + tZ_1 \quad \text{for } t \in [0,1].$$

As π^0 maximizes $\mathrm{MCov}(\beta^0, \nu)$ we get the upper bound

$$
\begin{aligned}
V(Z_t) - V(Z_0) &= \mathbb{E}\left[(Z_t + \Gamma) \cdot \nabla v_t(Z_t + \Gamma) - (Z_0 + \Gamma) \cdot \nabla v_0(Z_0 + \Gamma) - (Z_t - Z_0) \cdot X\right] \\
&\leq \mathbb{E}\left[(Z_t + \Gamma) \cdot \nabla v_t(Z_t + \Gamma) - (Z_0 + \Gamma) \cdot \nabla v_t(Z_t + \Gamma) - (Z_t - Z_0) \cdot X\right] \\
&= t\mathbb{E}\left[(Z_1 - Z_0) \cdot \nabla v_t(Z_t + \Gamma) - (Z_1 - Z_0) \cdot X\right] \\
&= t\mathbb{E}\left[(Z_1 - Z_0) \cdot ((\nabla v_t * \gamma_1)(Z_t) - X)\right],
\end{aligned}
$$

where the right-hand side used independence of Z_t and Γ for the last equality. Similarly, we obtain the lower bound

$$
\begin{aligned}
V(Z_t) - V(Z_0) &\geq t\mathbb{E}\left[(Z_1 - Z_0) \cdot (\nabla v_0(Z_0 + \Gamma) - X)\right] \\
&= t\mathbb{E}\left[(Z_1 - Z_0) \cdot ((\nabla v_0 * \gamma_1)(Z_0) - X)\right]. \tag{14}
\end{aligned}
$$

Combining these two inequalities, dividing by t and then sending t to zero yields

$$\lim_{t \to 0} \frac{1}{t} \left(V(Z_t) - V(Z_0)\right) = \mathbb{E}\left[(Z_1 - Z_0) \cdot ((\nabla v_0 * \gamma_1)(Z_0) - X)\right],$$

where we invoked (13) for the upper bound. We have shown that V is Gateaux differentiable with derivative $D_Z V = \nabla v_0 * \gamma_1(Z) - X \in L^2(\mathcal{G}; \mathbb{R}^d)$. Furthermore, since as consequence of (13) the map $Z \mapsto D_Z V : L^2(\mathcal{G}; \mathbb{R}^d) \to L^2(\mathcal{G}; \mathbb{R}^d)$ is continuous, we conclude that V is Fréchet differentiable.

(ii): If $\mathscr{L}(X, Z)$ is monotone, then $\mathbb{E}[X \cdot Z] = \mathrm{MCov}(\mu, \alpha)$ where $\alpha := \mathscr{L}(Z)$. Thus,

$$
\begin{aligned}
V(Z) &= \sup_{\substack{Y \in L^2(\mathcal{H}; \mathbb{R}^d), \\ Y \sim \nu}} \mathbb{E}\left[(Z + \Gamma) \cdot Y\right] - \mathrm{MCov}(\mu, \alpha) \\
&= \mathrm{MCov}(\nu, \alpha * \gamma_1) - \mathrm{MCov}(\mu, \alpha) = \mathcal{V}(\alpha),
\end{aligned}
$$

proving the claim.

(iii): We have seen in (14) that $V(Z_1) \geq V(Z_0) + D_{Z_0}V(Z_1 - Z_0)$ for every $Z_0, Z_1 \in L^2(\mathcal{G}; \mathbb{R}^d)$. It readily follows that V is convex. Thus, $V : L^2(\mathcal{G}; \mathbb{R}^d) \to \mathbb{R}$ is continuous and convex, which implies by [9, Theorem 9.1] that V is also lower semicontinuous.

(iv): If $\mathcal{V}$ admits an optimizer in $\mathcal{P}_2(\mathbb{R}^d)$, then by [7, Theorem 1.5] there is a Bass martingale from μ to ν with Bass measure $\alpha^\star \in \mathcal{P}_2(\mathbb{R}^d)$, where $\alpha^\star$ is a minimizer of $\mathcal{V}$. By [5, Theorem 1.4], $\alpha^\star$ is the push-forward of μ via a Brenier map ∇u. Defining $Z^\star = \nabla u(X)$ we see that $V(Z^\star) = \mathcal{V}(\alpha^\star)$, and by construction $(X, Z^\star)$ is monotone. Since V is translation invariant, we can set the mean of $Z^\star$ to be any real number, without affecting its optimality or the monotonicity of $(X, Z^\star)$. $\qquad \square$

Remark 2.2. It is somewhat remarkable that despite $\mathcal{V}$ being not $\mathcal{W}_2$-differentiable ($\mathrm{MCov}(\cdot, \mu)$ may fail to be differentiable), by passing on to the Hilbert space setting, we obtain a Fréchet differentiable functional V. The reason for this is that while there may be multiple optimal couplings from μ to $\alpha := \mathscr{L}(Z)$ for $Z \in L^2(\mathcal{G}; \mathbb{R}^d)$, by selecting the pair (X, Z) we have already fixed (in the Hilbert space setting) the coupling and thereby avoid this issue.

Lemma 2.3. *Let $(Z_t)_{t \geq 0}$ be an absolutely continuous curve in $L^2(\mathcal{G}; \mathbb{R}^d)$ with derivative $\frac{d}{dt} Z_t = -D_{Z_t} V$. Then $t \mapsto \mathbb{E}[Z_t]$ is a constant.*

Proof. We have $D_{Z_t} V = \nabla v_t * \gamma_1(Z_t) - X$ where ∇v_t is the Brenier map from $\mathscr{L}(Z_t + \Gamma)$ to ν. Hence

$$\mathbb{E}[\nabla v_t * \gamma_1(Z_t)] = \mathbb{E}[\mathbb{E}[\nabla v_t(Z_t + \Gamma)|\mathcal{G}]] = \mathbb{E}[\nabla v_t(Z_t + \Gamma)] = \text{bary}(\nu).$$

Since also $\text{bary}(\nu) = \text{bary}(\mu) = \mathbb{E}[X]$, we conclude that $\frac{d}{dt} \mathbb{E}[Z_t] = 0$ for almost every t. $\qquad \square$

3 The Gradient flow

3.1 Existence and strong convergence

To show L^2-convergence of the gradient flow, we will impose further assumptions on the marginals (μ, ν). Throughout this section we denote by

$$C_\rho := \text{co}(\text{supp } \rho),$$

the convex hull of the support of a measure $\rho \in \mathcal{P}(\mathbb{R}^d)$. Since the set C_ν features prominently in the subsequent parts of our work, we will simply write

$$C := C_\nu.$$

Without loss of generality we assume that $dim(C) = d$, otherwise restricting to the affine space spanned by C to achieve this. We also write

$$I := \text{int}(C),$$

for the interior of C, which is thus a nonempty open subset of $\mathbb{R}^d$. If $\mu \leq_{cx} \nu$ then $C_\mu \subset C$ and $\text{bary}(\nu) = \text{bary}(\mu)$. Hence we may assume w.l.o.g. that $\text{bary}(\nu) = \text{bary}(\mu) = 0 \in I$ and write $\mathcal{P}_2^0(\mathbb{R}^2)$ for the subset of $\mathcal{P}_2(\mathbb{R}^2)$ consisting of centred measures.

Subsequently we will work with the following assumptions:

Assumption (A): We have $\mu \leq_{cx} \nu$ and

(A1) the pair (μ, ν) is irreducible;

(A2) ν is compactly supported, i.e. C is compact;

(A3) $\text{supp}(\mu) \subseteq I$, i.e. C_μ is compactly contained in I.

To stress the importance of this assumption, we recall the following fact. If we only assume (A1) and (A2), i.e., the pair (μ, ν) is irreducible and ν is in addition compactly supported, the Bass functional may fail to admit a minimizer in $\mathcal{P}_2(\mathbb{R}^d)$ and minimizing sequences $(\alpha_n)_{n \in \mathbb{N}}$ may even fail to be tight (see [5, Example 6.7]). On the other hand, assuming (A3) in addition to (A1) and (A2), this drawback cannot happen anymore, as shown in the subsequent lemma. For reasons of presentation we present these results already here while the proofs rely on some technicalities, which are shown in the subsequent section, and are therefore deferred to Subsection 3.3.

9

Lemma 3.1. *Under Assumption (A) the functional V admits a minimizer in L^2 and $\mathcal{V}$ a minimizer in $\mathcal{P}_2(\mathbb{R}^d)$. Furthermore, any minimizer of V is bounded.*

As we have existence of a minimizer to the Bass functional under Assumption (A), the classical theory of gradient flows in Hilbert spaces provides us with the next result.

Lemma 3.2. *Let $Z_0 \in L^2(\mathcal{G};\mathbb{R}^d)$. Under Assumption (A) there exists a curve $(Z_t)_{t\geq 0}$ in $L^2(\mathcal{G};\mathbb{R}^d)$ starting in Z_0 such that*

$$\frac{d}{dt}Z_t = -D_{Z_t}V \text{ for a.e. } t > 0, \quad and \quad \frac{d^+}{dt}Z_t = -D_{Z_t}V \text{ for } t > 0. \tag{15}$$

Further, there exists a minimizer $Z^\star \in L^2(\mathcal{G};\mathbb{R}^d)$ of V such that $\lim_{t\to\infty} Z_t = Z^\star$ weakly in L^2 and $\lim_{t\to\infty} V(Z_t) = V(Z^\star)$.

We have already seen in Lemma 3.2 that, under Assumption (A), for any initial datum $Z_0 \in L^2(\mathcal{G};\mathbb{R}^d)$ the L^2-gradient flow $(Z_t)_{t\geq 0}$ of V exists and converges weakly to a minimizer of V, which exists by Lemma 3.1. But even beyond that, Assumption (A) is sufficiently strong to ensure strong convergence of the gradient flow. We now state the main result of this section.

Theorem 3.3. *Under Assumption (A), let $(Z_t)_{t\geq 0}$ be an L^2-gradient flow of V. Then $(Z_t)_{t\geq 0}$ converges strongly to the unique minimizer $Z^\star$ of V with $\mathbb{E}[Z^\star] = \mathbb{E}[Z_0]$.*

For the purpose of Proposition 3.4, which is key in showing Theorem 3.3, we need to pass to a pathwise absolutely continuous version of the gradient flow:

Fix a representative $\tilde{Z}_0$ of Z_0 and, for fixed ω, consider the curve $\tilde{Z} = (\tilde{Z}_t)_{t\geq 0}$ with

$$\tilde{Z}_t := \tilde{Z}_0 - \int_0^t D_{Z_s}V \, ds, \tag{16}$$

which is $\mathbb{P}$-almost surely well-defined since $|D_{Z_s}V|$ is uniformly bounded. Consequently, $\tilde{Z}_t$ is a representative of Z_t with absolutely continuous paths having pathwise the derivative

$$\frac{d\tilde{Z}_t}{dt} = -D_{Z_t}V = -D_{\tilde{Z}_t}V. \tag{17}$$

Proposition 3.4. *Under Assumption (A) let $(Z_t)_{t\geq 0}$ be an L^2-gradient flow of V and $(\tilde{Z}_t)_{t\geq 0}$ the representative given in (16). Then there is $M > 0$ such that the random time $\tau := \inf\{t > 0 : |\tilde{Z}_t| \leq M\}$ is a.s. finite and satisfies almost surely*

(i) $t \mapsto |\tilde{Z}_t|$ is strictly decreasing for $t \in [0, \tau)$;

(ii) $|\tilde{Z}_t| \leq M$ for all $t \geq \tau$.

In particular, $(|Z_t|^2)_{t\geq 0}$ is uniformly integrable.

The proofs of Theorem 3.3 and Proposition 3.4 need some preparations and will be given at the end of this section.

3.2 Geometric auxiliary results

As it turns out, under Assumption (A) and assuming that the starting variable Z_0 is in $L^\infty(\mathcal{G}; \mathbb{R}^d)$, one can show that the L^2-Bass gradient flow is confined to a bounded subset of $\mathbb{R}^d$, which eventually allows us to establish its convergence in the norm of L^2. Therefore, we need to control the gradient of V, which we do by analysing properties of the Brenier map $\nabla v : \mathbb{R}^d \to \mathbb{R}^d$ that pushes a measure $\beta \in \mathcal{P}_2(\mathbb{R}^d)$ with $\beta \ll \lambda$ to ν. To this end, we consider the support function $\ell : \mathbb{R}^d \to \mathbb{R}$ of C given by

$$\ell(a) := \max\{y \cdot a : y \in C\}. \tag{18}$$

We define the slice $S_{a,\delta}$ of C consisting of points in C that are at most $\delta > 0$ away from the supporting hyperplane with normal direction a, that is,

$$S_{a,\delta} := \{y \in C : y \cdot a > \ell(a) - \delta\}. \tag{19}$$

Clearly, ℓ is Lipschitz continuous on the $(d-1)$-sphere S^{d-1}. For the next lemma, we recall that we have assumed (w.l.o.g.) that C spans $\mathbb{R}^d$.

Lemma 3.5. *Let $\delta > 0$ and $r \geq 0$. Under Assumption (A), we have*

(i) there is $\varepsilon = \varepsilon(\delta, C) > 0$ such that for all $a, b \in S^{d-1}$

$$|a - b| \leq \varepsilon \implies S_{a,\delta/2} \subseteq S_{b,\delta}; \tag{20}$$

(ii) for $a \in S^{d-1}$ the set $\{y \in C : \exists y' \in S_{a,\delta} \text{ with } |y - y'| \leq r\}$ is contained in $S_{a,\delta+r}$;

(iii) the ν-measure of each δ-slice is uniformly bounded away from zero, i.e.,

$$\eta(\delta) := \inf\{\nu(S_{a,\delta}) : a \in S^{d-1}\} > 0;$$

(iv) we have

$$\Delta := \inf\{\ell(a) - x \cdot a : a \in S^{d-1}, x \in C_\mu\} > 0, \tag{21}$$

so that, for $0 < \delta < \Delta$ the sets C_μ and $S_{a,\delta}$ are disjoint, for all $a \in S^{d-1}$.

Proof. Ad (i), let $a, b \in S^{d-1}$ and $\varepsilon := \frac{\delta}{4L}$ where $L := \max\{|y| : y \in C\}$. We claim that then (20) holds. Indeed, if $|a - b| \leq \varepsilon$ and $y \in S_{a,\delta/2}$ we have

$$y \cdot b \geq y \cdot a - L|b - a| > \ell(a) - \frac{\delta}{2} - L|b - a| \geq \ell(b) - \frac{\delta}{2} - 2L|b - a| \geq \ell(b) - \delta.$$

Ad (ii), let $a \in S^{d-1}$, $y \in C$ and $y' \in S_{a,\delta}$ with $|y - y'| \leq r$. We merely observe that

$$y \cdot a \geq y' \cdot a - r > \ell(y') - \delta - r.$$

Ad (iii), first observe that $\nu(S_{a,\delta}) > 0$ for all $a \in S^{d-1}$, since C is the smallest closed convex set containing the support of ν and $C \setminus S_{a,\delta}$ is a closed convex subset of C. We proceed by

contradiction, that means, we assume by compactness that there is a sequence $(a_n)_{n\in\mathbb{N}}$ in S^{d-1} converging to a with $\nu(S_{a_n,\delta}) \to 0$. By (i) we have that whenever $|a_n - a| \leq \varepsilon$ we have $S_{a,\delta/2} \subseteq S_{a_n,\delta}$ which means that

$$\liminf_{n\to\infty} \nu(S_{a_n,\delta}) \geq \nu(S_{a,\delta/2}) > 0,$$

yielding a contradiction, hence, $\eta(\delta) > 0$.

Ad (iv), assume that $\Delta = 0$ which means by compactness that there is $a \in S^{d-1}$ and $x \in C_\mu$ with $\ell(a) - x \cdot a = 0$. Consequently, x is a support point of C with normal vector a and thus, by [9, Corollary 7.6], x is in the boundary of C which contradicts that $x \in C_\mu \subseteq I$. $\square$

Before continuing we would like to recall that as ν is compactly supported under Assumption (A), hence the Brenier map from any absolutely continuous $\beta \in \mathcal{P}_1(\mathbb{R}^d)$ to ν exists. Thus we can work in this section with probability measures having only finite first moments.

Lemma 3.6. *Let $\Lambda \subseteq \mathcal{P}_1(\mathbb{R}^d)$ be a tight family consisting of measures equivalent to the Lebesgue measure and write ∇v_β for the Brenier map from $\beta \in \Lambda$ to ν. Under Assumption (A), for $\delta > 0$, there is a constant $M = M(\nu, \Lambda, \delta) > 0$ such that*

$$z \in \mathbb{R}^d, |z| \geq M \implies \forall \beta \in \Lambda, \nabla v_\beta(z) \in S_{\frac{z}{|z|},\delta}.$$

Proof. Using Lemma 3.5 (iii) and tightness of Λ, there exists a radius $r > 0$ such that

$$\inf_{\beta \in \Lambda} \beta(B(r)) > 1 - \eta(\delta/2), \tag{22}$$

where $B(r) := \{y \in \mathbb{R}^d : |y| \leq r\}$ is the centred ball with radius r.

Fix $\beta \in \Lambda$ and $z \in \mathbb{R}^d$ with $|z| \geq \sup\{|y| : y \in C\} =: L$, and write $a := \frac{z}{|z|} \in S^{d-1}$. We proceed to show that if $x := \nabla v_\beta(z) \notin S_{a,\delta}$ then $|z| \leq M := 2(r+L)^2/\delta$. It follows from $(\nabla v_\beta)(\beta) = \nu$ that

$$\beta(\{\tilde{z} \in \mathbb{R}^d : \nabla v_\beta(\tilde{z}) \notin S_{a,\delta/2}\}) = 1 - \nu(S_{a,\delta/2}) \leq 1 - \eta(\delta/2),$$

whence, by (22) there exists $z' \in B(r)$ with $x' := \nabla v_\beta(z') \in S_{a,\delta/2}$. We have

$$|z' - x| \leq r + L, \tag{23}$$
$$|z - x| \geq a \cdot (z - x) = |z| - x \cdot a \geq |z| - \ell(a) + \delta > 0, \tag{24}$$

where we used in (24) that $|z| \geq L \geq \ell(a)$ and $x \notin S_{a,\delta}$. By cyclical monotonicity we must have

$$|x - z|^2 + |x' - z'|^2 \leq |x' - z|^2 + |x - z'|^2,$$

implying that

$$\begin{aligned}
(|z| - \ell(a) + \delta)^2 &\leq |z|^2 - 2z \cdot (x') + |x'|^2 + (L+r)^2 \\
&= |z|^2 - 2|z|a \cdot x' + |x'|^2 + (L+r)^2 \\
&\leq |z|^2 - 2|z|(\ell(a) - \delta/2) + |x'|^2 + (L+r)^2,
\end{aligned}$$

12

where the first equality comes from (23) and (24) and the last is due to $x' \in S_{a,\delta/2}$. Rearranging the terms leads to

$$|z|\delta + \delta^2 \leq L^2 + (L+r)^2,$$

from where we readily derive the claim. □

Lemma 3.7. *Under the assumptions of Lemma 3.6, there is a (possibly larger) constant $M = M(\nu, \Lambda, \delta) > 0$ such that*

$$z \in \mathbb{R}^d, |z| \geq M \implies \forall \beta \in \Lambda, \nabla v_\beta * \gamma_1(z) \in S_{\frac{z}{|z|},\delta}. \tag{25}$$

Proof. By Lemma 3.6 there is $\tilde{M} > 0$ such that

$$z \in \mathbb{R}^d, |z| \geq \tilde{M} \implies \forall \beta \in \Lambda, \nabla v_\beta(z) \in S_{\frac{z}{|z|},\delta/4}. \tag{26}$$

Further, by Lemma 3.5 (i) there is $\varepsilon > 0$ such that for all $a, b \in S^{d-1}$

$$|a - b| \leq \varepsilon \implies S_{a,\delta/4} \subseteq S_{b,\delta/2}. \tag{27}$$

Next, we choose r sufficiently large such that $\gamma_1(B(r)^c) \leq \delta/(4L)$ and set

$$M := \max\left(\tilde{M}, \frac{2r}{\varepsilon} + r\right).$$

It remains to show that M has the property described in (25). Fix $z \in \mathbb{R}^d$ with $|z| \geq M$ and $\beta \in \Lambda$, and write $x := \nabla v_\beta * \gamma_1(z)$. We have

$$x = \gamma_1(B(r)) \underbrace{\int_{B(r)} \nabla v_\beta(z-y)\frac{\gamma_1(dy)}{\gamma_1(B(r))}}_{=:x_1} + \gamma_1(B(r)^c)\underbrace{\int_{y \in B(r)^c} \nabla v_\beta(z-y)\frac{\gamma_1(dy}{\gamma_1(B(r)^c)}}_{=:x_2}.$$

When $|y| \leq r$ then $b = \frac{z}{|z|}$ and $a = \frac{z-y}{|z-y|}$ satisfy

$$|a - b| \leq \left|\frac{z}{|z|} - \frac{z}{|z-y|}\right| + \frac{r}{|z-y|} \leq \frac{||z-y| - |z||}{|z-y|} + \frac{r}{M-r} \leq \frac{2r}{M-r} \leq \varepsilon.$$

Therefore, (26) and (27) yield that in this case $\nabla v_\beta(z-y) \in S_{b,\delta/2}$, hence, also $x_1 \in S_{b,\delta/2}$ as the latter is a convex set. On the other hand, we have

$$|x_2|\gamma_1(B(r)^c) \leq L\gamma_1(B(r)^c) \leq \frac{\delta}{4}.$$

We claim that $x \in S_{b,\delta}$. Indeed,

$$x \cdot b \geq \gamma_1(B(r))(x_1 \cdot b) - \frac{\delta}{4} \geq (\gamma_1(B(r))x_1 + \gamma_1(B(r)^c)x_2) \cdot b$$

$$> \left(1 - \frac{\delta}{4L}\right)\left(\ell(b) - \frac{\delta}{2}\right) - \frac{\delta}{4} \geq \ell(b) - \frac{3\delta}{4} - \frac{\delta\ell(b)}{4L} \geq \ell(b) - \delta,$$

where we used that $\ell(b) \leq L$ and $x_1 \in S_{b,\delta/2}$. This concludes the proof. □

Lemma 3.8. *Under the assumptions of Lemma 3.6, let $M > 0$ satisfy (25). Then we have*

$$x \in C_\mu, z \in \mathbb{R}^d, |z| \geq M \implies \forall \beta \in \Lambda, z \cdot (\nabla v_\beta * \gamma_1(z) - x) \geq |z|(\Delta - \delta), \qquad (28)$$

where Δ is given as in (21).

Proof. Let $x \in C_\mu$, $\beta \in \Lambda$ and $z \in R^d$ with $|z| \geq M$, and write $a = \frac{z}{|z|}$.

$$\begin{aligned}
z \cdot (\nabla v_\beta * \gamma_1(z) - x) &= |z| a \cdot (\nabla v_\beta * \gamma_1(z) - x) \\
&\geq |z| \left(a \cdot \nabla v_\beta * \gamma_1(z) - \ell(a) + \Delta \right) \\
&\geq |z|(\Delta - \delta),
\end{aligned}$$

where we use Lemma 3.5 (iv) for the first inequality and Lemma 3.7 for the second. $\qquad \square$

3.3 Postponed proofs

Proof of Lemma 3.1. We have established in Item (ii) of Lemma 2.1 that $V(Z) \geq \mathcal{V}(\mathscr{L}(Z))$ for all $Z \in L^2(\mathcal{G}; \mathbb{R}^d)$ with equality if and only if the law of (X, Z) is monotone. Hence, Z minimizes V if and only if $\mathscr{L}(X, Z)$ is monotone and $\mathscr{L}(Z)$ is a Bass measure.

Since ν is compactly supported and (μ, ν) is irreducible, there exists by [5, Theorem 1.4] a convex function $v : \mathbb{R}^d \to \mathbb{R}$, $u := v * \gamma_1$ such that $\alpha := (\nabla u^*)(\mu)$ is the Bass measure of the Bass martingale joining μ and ν. Using Lemma 3.5 (iv), we pick $\delta > 0$ such that $\delta < \Delta$. Then we have $C_\mu \cap S_{a,\delta} = \emptyset$ for all $a \in S^{d-1}$. By Lemma 3.7 there is $M > 0$ such that

$$\{z \in \mathbb{R}^d : |z| \geq M\} \subseteq \bigcup_{a \in S^{d-1}} \{z \in \mathbb{R}^d : \nabla u(z) \in S_{a,\delta}\}.$$

We find that $\{z \in \mathbb{R}^d : \nabla u(z) \in C_\mu\} \subseteq \{z \in \mathbb{R}^d : |z| < M\}$, from where we conclude that α is supported on $\{z \in \mathbb{R}^d : |z| \leq M\}$ and $\alpha \in \mathcal{P}_2(\mathbb{R}^d)$. Hence, $Z = \nabla u^*(X) \in L^2(\mathcal{G}; \mathbb{R}^d)$ minimizes V. $\qquad \square$

Proof of Lemma 3.2. The existence of the gradient flow satisfying (15) follows by [21, Theorems 3.1]. The weak convergence to $Z^\star$, a minimizer of V, is a consequence of [22, Theorem 4]. To check that $V(Z_t) \to V(Z^\star)$, we will use that $\|\frac{d^+}{dt} Z_t\|_{L^2} \to 0$, where the latter property is due to [21, Theorem 3.7], and argue by contradiction. Indeed, if $(t_n)_{n \in \mathbb{N}}$ is an increasing, non-negative sequence with $t_n \to \infty$ and $a > 0$ exist such that $V(Z_{t_n}) \geq a + V(Z^\star)$, then by convexity

$$-a \geq V(Z^\star) - V(Z_{t_n}) \geq D_{Z_{t_n}} V \cdot (Z^\star - Z_{t_n}) = -\frac{d^+}{dt} Z_{t_n} \cdot (Z^\star - Z_{t_n}) \to 0,$$

for $n \to \infty$, as $(Z_t)_{t \geq 0}$ is L^2-bounded. This contradicts $a > 0$. $\qquad \square$

Proof of Proposition 3.4. Due to Lemma 3.5 (iv) we can pick $\delta > 0$ such that $\delta < \Delta$. We know by Lemma 3.2 that the gradient flow is weakly convergent to a minimizer of V.

Therefore, $\sup_{t \geq 0} \mathbb{E}[|Z_t|^2] < \infty$ and $\Lambda := \{\mathscr{L}(Z_t) * \gamma_1 : t \geq 0\}$ is tight. Let $M > 0$ be the constant provided by Lemma 3.8, then we have by (28) that

$$|\tilde{Z}_t| \geq M \implies \tilde{Z}_t \cdot (\nabla v_t * \gamma_1(\tilde{Z}_t) - X) \geq |\tilde{Z}_t|(\Delta - \delta),$$

where $\nabla v_t : \mathbb{R}^d \to \mathbb{R}^d$ denotes the Brenier map from $\mathscr{L}(Z_t) * \gamma_1$ to ν. Recall that by Lemma 2.1 we have $D_{\tilde{Z}_t} V = \nabla v_t * \gamma_1(\tilde{Z}_t) - X$. By (17) we have on a $\mathbb{P}$-full set $\tilde{\Omega}$, for almost every t

$$|\tilde{Z}_t| \geq M \implies \frac{1}{2}\frac{d}{dt}|\tilde{Z}_t|^2 = -\tilde{Z}_t \cdot \left(\nabla v_t * \gamma_1(\tilde{Z}_t) - X\right) \leq -|\tilde{Z}_t|(\Delta - \delta) < 0,$$

where we used for the last inequality that $\Delta - \delta > 0$. Hence, on $\{t \geq 0 : t < \tau\}$ the curve $t \mapsto |\tilde{Z}_t|$ is strictly decreasing on $\tilde{\Omega}$. By the same argument, τ must be finite. On the other hand, let τ_1 be a random time with $\tau_1 \geq \tau$ and $|\tilde{Z}_{\tau_1}| > M$ on $\{\tau_1 > \tau\}$, then we have on $\{\tau_1 > \tau\} \cap \tilde{\Omega}$

$$0 < |\tilde{Z}_{\tau_1}|^2 - |\tilde{Z}_\tau|^2 < -(\tau_1 - \tau)(\Delta - \delta)M,$$

which yields that $\{\tau_1 > \tau\}$ is disjoint with $\tilde{\Omega}$. Consequently, we have shown that $|\tilde{Z}_t| \leq M$ for all $t \geq \tau$ on $\tilde{\Omega}$, which completes the proof. $\qquad\square$

Proof of Theorem 3.3. By Proposition 3.4 we have that $(|Z_t|^2)_{t \geq 0}$ is uniformly integrable. In turn, the family $\{\mathscr{L}(Z_t) : t \geq 0\}$ is $\mathcal{W}_2$-precompact which enables us to find an increasing, non-negative sequence $(t_n)_{n \in \mathbb{N}}$ with $t_n \to \infty$ such that $\mathscr{L}(Z_{t_n}) \to \alpha$ in $\mathcal{W}_2$. Using Lemma 2.1 (ii) and Lemma 3.2 we have

$$\mathcal{V}(\mathscr{L}(Z^\star)) = V(Z^\star) = \lim_{n \to \infty} V(Z_{t_n}) \geq \lim_{n \to \infty} \mathcal{V}(\mathscr{L}(Z_{t_n})) = \mathcal{V}(\alpha).$$

Since $Z^\star$ is the unique minimizer of V with mean $\mathbb{E}[Z_0]$ this shows that $\mathscr{L}(Z^\star) = \alpha$. It follows immediately that

$$\lim_{t \to \infty} \mathbb{E}[|Z_t|^2] = \mathbb{E}[|Z^\star|^2],$$

which, together with weak convergence of $(Z_t)_{t \geq 0}$, yields L^2-norm convergence to $Z^\star$. $\qquad\square$

4 The one-dimensional case

Throughout this section, we assume that $d = 1$.

4.1 Second-order analysis

In this part we assume that $d = 1$, although we will still often use vector and matrix notation to make this part consistent with the other ones. Throughout this section, Φ_1 denotes the standard Gaussian cumulative distribution function and $\phi_1 := \Phi_1'$ its density.

Lemma 4.1. *Let $\Delta Z \in L^1(\mathcal{G}; \mathbb{R}^d)$, Z be $\mathcal{G}$-measurable. Then we have for γ_1-a.e. ζ*

$$\mathbb{E}[\Delta Z | Z + \Gamma = \zeta] = \frac{\mathbb{E}[\phi_1(\zeta - Z)\Delta Z]}{\mathbb{E}[\phi_1(\zeta - Z)]}. \tag{29}$$

Proof. Using independence of $\mathcal{G}$ and Γ, we compute for measurable, bounded $h\colon \mathbb{R} \to \mathbb{R}$ that

$$\mathbb{E}[\Delta Z h(Z + \Gamma)] = \mathbb{E}\left[\Delta Z \mathbb{E}[h(Z + \Gamma)|\mathcal{G}]\right] = \mathbb{E}\left[\Delta Z \int h(\zeta)\phi_1(\zeta - Z)\, d\zeta\right]$$

$$= \int \left(\frac{\mathbb{E}[\Delta Z \phi_1(\zeta - Z)]}{\mathbb{E}[\phi_1(\zeta - Z)]}\right) h(\zeta)\mathbb{E}[\phi_1(\zeta - Z)]\, d\zeta$$

$$= \mathbb{E}\left[\int \left(\frac{\mathbb{E}[\Delta Z \phi_1(\zeta - Z)]}{\mathbb{E}[\phi_1(\zeta - Z)]}\right) h(\zeta)\phi_1(\zeta - Z)\, d\zeta\right]$$

$$= \mathbb{E}\left[\left(\frac{\mathbb{E}[\Delta Z \phi_1(\zeta - Z)]}{\mathbb{E}[\phi_1(\zeta - Z)]}\right)\bigg|_{\zeta = Z + \Gamma} h(Z + \Gamma)\right].$$

From this equality and uniqueness of conditional expectations, we readily derive (29). $\qquad \square$

To proceed with our second-order analysis of V, we require additional regularity of ν.

Assumptions (A'): The pair (μ, ν) satisfies Assumption (A), and in addition

(A4) ν is absolutely continuous with λ-ess $\inf_{x \in I} \frac{d\nu}{d\lambda}(x) > 0$;

(A5) the density of ν satisfies λ-ess $\sup_{x \in \mathbb{R}} \frac{d\nu}{d\lambda}(x) < \infty$.

Using this assumption, we can compute the derivative of $D_Z V$ along continuously differentiable curves. Throughout we denote by Hv the second derivative (Hessian) of v.

Proposition 4.2. *Under Assumption (A'), let $(Z_t)_{t\geq 0}$ be a continuously differentiable curve in $L^2(\mathcal{G}; \mathbb{R}^d)$ with derivative G_t. Then $t \mapsto D_{Z_t} V$ is continuously differentiable with*

$$\frac{d}{dt}(D_{Z_t} V) = \mathbb{E}\left[Hv_t(Z_t + \Gamma)\left(G_t - \mathbb{E}[G_t | Z_t + \Gamma]\right) | \mathcal{G}\right], \tag{30}$$

where $\nabla v_t \colon \mathbb{R} \to \mathbb{R}$ is the Brenier map from $\mathscr{L}(Z_t + \Gamma)$ to ν.

Proof. As we are currently working with $d = 1$, the Brenier map is explicitly given by $\nabla v_t = Q_\nu \circ F_t$ where Q_ν is the quantile function of ν and F_t is the cdf of Z_t. It follows from (A4) that Q_ν is almost surely differentiable on $(0, 1)$. Consequently, $\frac{d}{dx}\nabla v_t(x)$ and $\frac{d}{dt}\nabla v_t(x)$ exist for λ-a.e. $x \in \mathbb{R}$. We compute the space and time derivatives of ∇v_t, and obtain λ-a.s.

$$Hv_t = \frac{d}{dx}(Q_\nu \circ F_t) = \left(\frac{d}{dx}Q_\nu\right) \circ F_t \left(\frac{d}{dx}F_t\right) = \left(\frac{d}{dx}Q_\nu\right) \circ F_t\, \mathbb{E}[\phi_1(\cdot - Z_t)], \tag{31}$$

as well as (γ_1-a.s.)

$$\frac{d}{dt}(\nabla v_t) = \left(\frac{d}{dx}Q_\nu\right) \circ F_t \left(\frac{d}{dt}F_t\right) = -\left(\frac{d}{dx}Q_\nu\right) \circ F_t\, \mathbb{E}[\phi_1(\cdot - Z_t)G_t]$$

$$= -Hv_t\, \mathbb{E}[G_t | Z_t + \Gamma = \cdot],$$

16

where the last equality follows by Lemma 4.1. As $\frac{d\nu}{dx}$ is λ-essentially bounded away from zero on I by some constant $c > 0$, we have that

$$\lambda\text{-}\operatorname*{ess\,sup}_{x\in\mathbb{R}} |Hv_t(x)| \leq \frac{1}{c} < \infty \quad \text{and} \quad \lambda\text{-}\operatorname*{ess\,sup}_{x\in\mathbb{R}} \left|\frac{d}{dt}\nabla v_t(x)\right| \leq \frac{1}{c}\|G_t\|_{L^2} < \infty.$$

Recall the form of $D_{Z_t}V$ proved in Lemma 2.1. Thanks to the bounds, we can exchange integration with differentiation and get

$$\frac{d}{dt}(D_{Z_t}V) = \mathbb{E}\left[\frac{d}{dt}\left(\nabla v_t(Z_t + \Gamma)\right)\Big|\mathcal{G}\right] = \mathbb{E}\left[\left(\frac{d}{dt}\nabla v_t\right)(Z_t + \Gamma) + Hv_t(Z_t + \Gamma)G_t\Big|\mathcal{G}\right]$$
$$= \mathbb{E}\left[Hv_t(Z_t + \Gamma)\left(G_t - \mathbb{E}[G_t|Z_t + \Gamma]\right)\Big|\mathcal{G}\right].$$

Note that the maps $t \mapsto Hv_t(Z_t + \Gamma)$ and $t \mapsto (\frac{d}{dt}\nabla v_t)(Z_t + \Gamma)$ are L^2-continuous, from where we conclude that $t \mapsto D_{Z_t}V$ is continuously differentiable. $\qquad\square$

Lemma 4.3. *Under Assumption (A'), let $(Z_t)_{t\in[0,1]}$ be a curve in $L^2(\mathcal{G};\mathbb{R}^d)$ with $Z_t = Z_0 + t\Delta Z$ where $\Delta Z := Z_1 - Z_0$. Then we have*

$$\frac{d^2}{dt^2}V(Z_t) = \mathbb{E}\left[\left(\Delta Z - \mathbb{E}[\Delta Z|Z_t + \Gamma]\right)Hv_t(Z_t + \Gamma)\left(\Delta Z - \mathbb{E}[\Delta Z|Z_t + \Gamma]\right)\right], \qquad (32)$$

where ∇v_t is the Brenier map from $\mathscr{L}(Z_t + \Gamma)$ to ν.

Proof. We have shown in Proposition 4.2 that $t \mapsto D_{Z_t}V$ is continuously differentiable. Therefore, the second derivative of $t \mapsto V(Z_t)$ can be represented as

$$\frac{d^2}{dt^2}V(Z_t) = \frac{d}{dt}\mathbb{E}\left[\Delta Z D_{Z_t}V\right] = \mathbb{E}\left[\Delta Z \frac{d}{dt}D_{Z_t}V\right]$$
$$= \mathbb{E}\left[\Delta Z Hv_t(Z_t + \Gamma)\left(\Delta Z - \mathbb{E}[\Delta Z|Z_t + \Gamma]\right)\right],$$

where we used (30) and the tower property for the last equality. Further, since

$$\mathbb{E}\left[\mathbb{E}\left[\Delta Z|Z_t + \Gamma\right]Hv_t(Z_t + \Gamma)\left(\Delta Z - \mathbb{E}\left[\Delta Z|Z_t + \Gamma\right]\right)\right] = 0,$$

we conclude with (32). $\qquad\square$

4.2 Contraction property of conditional expectation

The final ingredient in the proof of exponential convergence of the gradient flow of the lifted Bass functional is the following contraction property of the conditional expectation.

Lemma 4.4. *Let $\Delta Z \in L^2(\mathcal{G};\mathbb{R})$ with $\mathbb{E}[\Delta Z] = 0$ and $R > 0$. Then there is $\varepsilon = \varepsilon(R) \in (0, 1)$ such that, for every $Z \in L^2(\mathcal{G};\mathbb{R})$ with $\|Z\| \leq R$ we have*

$$\|\mathbb{E}[\Delta Z|Z + \Gamma]\|_{L^2}^2 \leq (1 - \varepsilon)\|\Delta Z\|_{L^2}^2.$$

Proof. As preparation we introduce the quantities, for $\zeta \in \mathbb{R}$,

$$g(\zeta) := \min_{(y,z) \in \mathbb{R} \times \mathbb{R},\, |y|,|z| \leq R} \frac{\phi_1(\zeta - y)}{\phi_1(\zeta - z)}$$

Note that g is continuous and $g \in (0, 1)$. Therefore, we can set

$$\varepsilon := \min_{z \in \mathbb{R},\, |z| \leq R} \mathbb{E}[g(z + \Gamma)],$$

and by the previous observations we see that $\varepsilon > 0$. Furthermore, since $\mathbb{E}[\Delta Z] = 0$ we get by Lemma 4.1

$$\mathbb{E}[\Delta Z | Z + \Gamma = \zeta] = \frac{\mathbb{E}[\phi_1(\zeta - Z)\Delta Z]}{\mathbb{E}[\phi_1(\zeta - Z)]} = \mathbb{E}\left[\Delta Z \left(\frac{\phi_1(\zeta - Z)}{\mathbb{E}[\phi_1(\zeta - Z)]} - g(\zeta)\right)\right].$$

Since $\frac{\phi_1(\zeta - Z)}{\mathbb{E}[\phi_1(\zeta - Z)]} \geq g(\zeta)$ and $g(\zeta) < 1$, we have that

$$h(Z) := \left(\frac{\phi_1(\zeta - Z)}{\mathbb{E}[\phi_1(\zeta - Z)]} - g(\zeta)\right) \frac{1}{1 - g(\zeta)},$$

is a probability density. Therefore by Jensen's inequality we get

$$\begin{aligned}
\mathbb{E}[\Delta Z | Z + \Gamma = \zeta]^2 &= (1 - g(\zeta))^2\, \mathbb{E}[\Delta Z h(Z)]^2 \\
&\leq (1 - g(\zeta))^2\, \mathbb{E}[\Delta Z^2 h(Z)] \\
&= (1 - g(\zeta))\mathbb{E}\left[\Delta Z^2 \left(\frac{\phi_1(\zeta - Z)}{\mathbb{E}[\phi_1(\zeta - Z)]} - g(\zeta)\right)\right] \\
&\leq (1 - g(\zeta))\mathbb{E}\left[\Delta Z^2 | Z + \Gamma = \zeta\right].
\end{aligned}$$

Hence, taking expectations leads to

$$\begin{aligned}
\|\mathbb{E}[\Delta Z | Z + \Gamma]\|_{L^2}^2 &\leq \mathbb{E}\left[(1 - g(Z + \Gamma))\mathbb{E}[\Delta Z^2 | Z + \Gamma]\right] \\
&= \mathbb{E}\left[(1 - g(Z + \Gamma))\Delta Z^2\right] \\
&= \mathbb{E}\left[\Delta Z^2 \mathbb{E}[1 - g(Z + \Gamma)|\mathcal{G}]\right] \\
&\leq \mathbb{E}\left[\Delta Z^2 \left(1 - \min_{z \in \mathbb{R},\, |z| \leq R} \mathbb{E}[g(z + \Gamma)]\right)\right] = (1 - \varepsilon)\mathbb{E}[\Delta Z^2],
\end{aligned}$$

which concludes the proof. $\qquad\square$

4.3 Exponential convergence

We consider the setting and notation of Section 4.1, wherein $d = 1$. As in that part, we use here vector and matrix notation nevertheless.

Lemma 4.5. *Let $\nu \ll \lambda$ satisfy $\frac{d\nu}{d\lambda} > 0$ on I, $R > 0$ and $Z \in L^2(\mathcal{G}; \mathbb{R})$ with $|Z| \leq R$. Denote by ∇v_Z the Brenier map from $\mathscr{L}(Z + \Gamma)$ to ν. Then we have for λ-a.e. ζ*

$$Hv_Z(x) \geq \frac{\inf_{\zeta \in \mathbb{R}, |\zeta| \leq R} \phi_1(x - \zeta)}{\|\frac{d\nu}{d\lambda}\|_{L^\infty(\lambda)}}. \tag{33}$$

Proof. Write F for the cdf of $\mathscr{L}(Z + \Gamma)$. Under our assumption on ν the quantile function Q_ν of ν is almost everywhere differentiable on $(0, 1)$ and we have by (31) that for λ-a.e. x

$$Hv_Z(x) = \left(\frac{d}{dx}Q_\nu\right) \circ F(x)\, \mathbb{E}[\phi_1(x - Z)].$$

As consequence of $\frac{d\nu}{d\lambda} > 0$ on I we have λ-a.s. $(\frac{d}{dx}Q_\nu) \circ F \geq \frac{1}{\frac{d\nu}{d\lambda}} \geq \frac{1}{\|\frac{d\nu}{d\lambda}\|_{L^\infty(\lambda)}}$. Further, since $|Z| \leq R$ we get $\mathbb{E}[\phi_1(x - Z)] \geq \inf_{\zeta \in \mathbb{R}, |\zeta| \leq R} \phi_1(x - \zeta)$, we obtain (33). $\square$

Lemma 4.6. *Under Assumption (A') there is for every $R > 0$ an $\varepsilon = \varepsilon(R) > 0$ such that for all $Z_i \in L^2(\mathcal{G}; \mathbb{R})$, $i = 0, 1$ with $\operatorname{ess\,sup} |Z_i| \leq K$, the curve $(Z_t)_{t \in [0,1]}$ with $Z_t := (1-t)Z_1 + tZ_0$ satisfies*

$$\frac{d^2}{dt^2}V(Z_t) \geq \varepsilon \|Z_1 - Z_0 - \mathbb{E}[Z_1 - Z_0 | Z_t + \Gamma]\|_{L^2}^2. \tag{34}$$

Proof. We have shown in Lemma 4.3 that, after denoting $\Delta Z := Z_1 - Z_0$,

$$\frac{d^2}{dt^2}V(Z_t) = \mathbb{E}[Hv_t(Z_t + \Gamma)(\Delta Z - \mathbb{E}[\Delta Z | Z_t + \Gamma])^2],$$

where ∇v_t is the Brenier map from $\mathscr{L}(Z_t + \Gamma)$ to ν. Using the lower bound derived in Lemma 4.5 we have

$$\begin{aligned}
\frac{d^2}{dt^2}V(Z_t) &\geq \mathbb{E}\left[g_R(Z_t + \Gamma)(\Delta Z - \mathbb{E}[\Delta Z | Z_t + \Gamma])^2\right] \\
&= \mathbb{E}\left[g_R * \gamma_1(Z_t)(\Delta Z - \mathbb{E}[\Delta Z | Z_t + \Gamma])^2\right],
\end{aligned} \tag{35}$$

where for $x \in \mathbb{R}$

$$g_R(x) := \frac{\inf_{\zeta \in \mathbb{R}, |\zeta| \leq R} \phi_1(x - \zeta)}{\|\frac{d\nu}{d\lambda}\|_{L^\infty(\lambda)}}.$$

Observe that g_R is continuous and strictly positive everywhere. Hence, $g_R * \gamma_1$ is also continuous and strictly positive everywhere. It follows that for $\varepsilon := \inf_{|\zeta| \leq R} g_R * \gamma_1(\zeta) > 0$, so (34) is readily derived from (35). $\square$

Lemma 4.7. *Let $R > 0$. Under Assumption (A') there is $\varepsilon > 0$ such that for all $Z_i \in L^2(\mathcal{G}; \mathbb{R})$, $i = 0, 1$ with $|Z_i| \leq R$ we have*

$$V(Z_1) - V(Z_0) - \langle D_{Z_0}V, Z_1 - Z_0 \rangle_{L^2} \geq \varepsilon \|Z_1 - Z_0\|_{L^2}^2. \tag{36}$$

Proof. Set $\Delta Z := Z_1 - Z_0$ and consider the curve $(Z_t)_{t \in [0,1]}$ with $Z_t := Z_0 + t\Delta Z$. Since V is continuously differentiable by Lemma 2.1 we have

$$V(Z_1) - V(Z_0) = \int_0^1 \langle D_{Z_t}V, \Delta Z \rangle_{L^2}\, dt.$$

Hence, it suffices to prove the inequality

$$\langle D_{Z_t}V - D_{Z_0}V, \Delta Z \rangle_{L^2} \geq 2\varepsilon t \|\Delta Z\|_{L^2}^2.$$

By Lemma 4.6 there is $\varepsilon(R) > 0$ such that

$$\frac{d}{dt}\langle D_{Z_t}V, \Delta Z \rangle_{L^2} = \frac{d^2}{dt^2}V(Z_t) \geq \varepsilon(R)\|\Delta Z - \mathbb{E}[\Delta Z | Z_t + \Gamma]\|_{L^2}^2.$$

Finally, by Lemma 4.4 there is $\varepsilon' > 0$ such that

$$\|\Delta Z - \mathbb{E}[\Delta Z | Z_t + \Gamma]\|_{L^2}^2 = \|\Delta Z\|_{L^2}^2 - \|\mathbb{E}[\Delta Z | Z_t + \Gamma]\|_{L^2}^2 \geq \varepsilon' \|\Delta Z\|_{L^2}^2,$$

where the first equality is due to the tower property. Therefore, (36) holds for $2\varepsilon := \varepsilon(R)\varepsilon'$, which completes the proof. $\qquad\square$

Proof of Theorem 1.9 . Per Proposition 3.4 the whole flow $(Z_t)_{t \geq 0}$ takes values in $B_\infty(M) := \{Z \in L^2(\mathcal{G}; \mathbb{R}) : \operatorname{ess\,sup}|Z| \leq M\}$ for some $M > 0$. Denote by Z^* the L^2-limit of $(Z_t)_{t \geq 0}$ which is a minimizer of V. Hence, throughout this proof we may assume that all random variables are uniformly bounded by M. By Lemma 4.7, we can find ε such that V is (2ε)-strongly convex restricted to $B_\infty(M)$. Using (36) we obtain for $Z \in B_\infty(M)$ that

$$V(Z^*) - V(Z) \geq \inf_{Z'} \langle D_Z V, Z' - Z \rangle_{L^2} + \varepsilon\|Z' - Z\|_{L^2}^2,$$

from where we derive

$$V(Z) - V(Z^\star) \leq \frac{1}{4\varepsilon}\|D_Z V\|_{L^2}^2. \tag{37}$$

On the other hand, adding (36) (where we let $(Z_0, Z_1) = (Z, Z^\star)$) to (36) (where we let $(Z_1, Z_0) = (Z, Z^\star)$) yields

$$\langle D_Z V, Z - Z^\star \rangle_{L^2} \geq 2\varepsilon\|Z - Z^\star\|_{L^2}^2. \tag{38}$$

From (37) and continuous differentiability of $(Z_t)_{t \geq 0}$ with $\frac{d}{dt}Z_t = -D_{Z_t}V$ we get

$$\frac{d}{dt}\big(V(Z_t) - V(Z^\star)\big) = -\|D_{Z_t}V\|_{L^2}^2 \leq -4\varepsilon\big(V(Z_t) - V(Z^\star)\big),$$

while (38) yields

$$\frac{d}{dt}\|Z_t - Z^*\|_{L^2}^2 = -2\langle Z_t - Z^*, D_{Z_t}V \rangle_{L^2} \leq -4\varepsilon\|Z_t - Z^*\|_{L^2}^2.$$

We conclude by Gronwall's inequality. $\qquad\square$

References

[1] B. Acciaio, A. Marini, and G. Pammer. Calibration of the bass local volatility model. *arXiv preprint arXiv:2311.14567*, 2023.

[2] 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.

[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:2306.11019, 2023.

[6] J. Backhoff-Veraguas, G. Loeper, and J. Obloj. Geometric martingale benamou-brenier transport and geometric bass martingales. *arXiv preprint arXiv:2406.04016*, 2024.

[7] J. Backhoff-Veraguas, W. Schachermayer, and B. Tschiderer. The Bass functional of martingale transport. *arXiv preprint arXiv:2309.11181*, 2023.

[8] 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.

[9] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.

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

[11] 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.

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

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

[14] 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.

[15] M. Beiglböck, G. Pammer, and L. Riess. Change of numeraire for weak martingale transport. *arXiv e-prints*, pages arXiv–2406, 2024.

[16] 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.

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

[18] F. Bourgey and J. Guyon. Fast exact joint s&p 500/vix smile calibration in discrete and continuous time. *Risk, February*, 2024.

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

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

[21] H. Brézis. Opérateurs maximaux monotones dans les espaces de hilbert et équations d'évolution. *Lecture Notes*, 5:3–4, 1972.

[22] R. E. Bruck Jr. Asymptotic convergence of nonlinear contraction semigroups in hilbert space. *Journal of Functional Analysis*, 18(1):15–26, 1975.

[23] 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.

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

[25] A. Conze and P. Henry-Labordere. Bass construction with multi-marginals: Lightspeed computation in a new local volatility model. *Available at SSRN 3853085*, 2021.

[26] 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.

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

[28] 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.

[29] 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.

[30] 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.

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

[32] 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.

[33] 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.

[34] 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.

[35] I. Guo, G. Loeper, and S. Wang. Local volatility calibration by optimal transport. In *2017 MATRIX Annals*, pages 51–64. Springer, 2019.

[36] 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.

[37] D. Hobson and A. Neuberger. Robust bounds for forward start options. *Math. Finance*, 22(1):31–56, 2012.

[38] M. Huesmann and D. Trevisan. A Benamou-Brenier formulation of martingale optimal transport. *Bernoulli*, 25(4A):2729–2757, 2019.

[39] B. Joseph, G. Loeper, and J. Obloj. The measure preserving martingale sinkhorn algorithm. *arXiv e-prints*, pages arXiv–2310, 2023.

[40] B. Jourdain and G. Pammer. An extension of martingale transport and stability in robust finance. *Electronic Journal of Probability*, 29:1–30, 2024.

[41] 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.

[42] L. Kantorovich. On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.

[43] G. Loeper. Option pricing with linear market impact and nonlinear Black–Scholes equations. *The Annals of Applied Probability*, 28(5):2664–2726, 2018.

[44] R. McCann. A convexity theory for interacting gases and equilibrum crystals. *PhD thesis, Princeton University*, 1994.

[45] R. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80(2):309–323, 1995.

[46] G. Monge. Memoire sur la theorie des deblais et des remblais. *Histoire de l'académie Royale des Sciences de Paris*, 1781.

[47] J. Obłój and P. Siorpaes. Structure of martingale transports in finite dimensions. arXiv:1702.08433, 2017.

[48] 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.

[49] 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.

[50] X. Tan and N. Touzi. Optimal transportation under controlled stochastic dynamics. *Ann. Probab.*, 41(5):3201–3240, 2013.

[51] B. Tschiderer. q-bass martingales. *arXiv preprint arXiv:2402.05669*, 2024.

[52] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[53] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.