Submitted to the Annals of Probability

EXISTENCE OF BASS MARTINGALES AND THE MARTINGALE BENAMOU–BRENIER PROBLEM IN $\mathbb{R}^d$

By Julio Backhoff-Veraguas[1,a], Mathias Beiglböck[1,b], Walter Schachermayer[1,c], and Bertram Tschiderer[1,d]

[1] *Faculty of Mathematics, University of Vienna,* [a] julio.backhoff@univie.ac.at; [b] mathias.beiglboeck@univie.ac.at; [c] walter.schachermayer@univie.ac.at; [d] bertram.tschiderer@univie.ac.at

In classical optimal transport, the contributions of Benamou–Brenier and McCann regarding the time-dependent version of the problem are cornerstones of the field and form the basis for a variety of applications in other mathematical areas.

In this article, we characterize solutions to the martingale Benamou–Brenier problem as *Bass martingales*, i.e. transformations of Brownian motion through the gradient of a convex function. Our result is based on a new (static) Brenier-type theorem for a particular weak martingale optimal transport problem. As in the classical case, the structure of the primal optimizer is derived from its dual counterpart, whose derivation forms the technical core of this article. A key challenge is that dual attainment is a subtle issue in martingale optimal transport, where dual optimizers may fail to exist, even in highly regular settings.

1. Introduction. Optimal transport as a field in mathematics goes back to Monge [51] and Kantorovich [44], who established its modern formulation. The seminal results of Benamou, Brenier, and McCann [19, 20, 17, 47] form the basis of the modern theory, with striking applications in a variety of different areas; see the monographs [58, 59, 3, 55].

We are interested in transport problems where the transport plan satisfies an additional martingale constraint. This additional requirement arises naturally in finance, but is of independent mathematical interest. For example, there are notable consequences for the study of martingale inequalities (e.g. [18, 38, 53]) and the Skorokhod embedding problem (e.g. [9, 43, 14]). Early articles on this topic of *martingale optimal transport* include [40, 10, 57, 30, 28, 21].

In view of the central role taken by the results of Benamou–Brenier [17] and McCann [49] on optimal transport for squared (Euclidean) distance, the related continuous-time transport problem and McCann's displacement interpolation, it is paramount to search for similar concepts in the martingale context. This is the main motivation of the present article. Before describing our results, we briefly recap the classical role models.

1.1. *Benamou–Brenier transport problem and McCann interpolation in probabilistic terms.* Given μ, ν in the space $\mathcal{P}_2(\mathbb{R}^d)$ of d-dimensional distributions with finite second moment, Brenier's theorem asserts that if μ is absolutely continuous, the following are equivalent for a coupling $\pi \in \mathsf{Cpl}(\mu, \nu)$, i.e., a probability on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ, ν:

(1) π minimizes the transport costs with respect to the squared distance between μ and ν.
(2) π is concentrated on the graph of a function $f\colon \mathbb{R}^d \to \mathbb{R}^d$ which is *monotone* in the sense that $f = \nabla v$ for a convex function $v\colon \mathbb{R}^d \to (-\infty, +\infty]$.

Brenier's theorem gives a structural description of the optimal transport plan for the most widely used cost function. Another essential consequence of the result is that it provides a

MSC2020 subject classifications: Primary 60G42, 60G44; secondary 91G20.

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

particularly natural way to move probabilities: it implies that μ can be transported to ν via the gradient of a convex function. In fact, in many applications it is the mere existence of this monotone transport map that is required, irrespective of its optimality properties.

Ideally, a martingale counterpart of Brenier's theorem should mimic these aspects. That is, starting from a natural optimization problem, our goal is to define a natural martingale that connects the probabilities μ and ν.

While [12, 39, 31] have proposed martingale versions of Brenier's monotone transport map based on static transport problems, our starting point is a continuous formulation in the spirit of the Benamou–Brenier continuous-time transport problem given in [5, 41].

We first recapitulate the continuous-time formulation of Brenier's theorem in probabilistic language. For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, consider

$$\text{(BB)} \qquad T_2(\mu,\nu) \coloneqq \inf_{\substack{X_0\sim\mu,\, X_1\sim\nu,\\ X_t = X_0 + \int_0^t b_s\, ds}} \mathbb{E}\Big[\int_0^1 |b_t|^2\, dt\Big].$$

Under the above assumptions, (BB) has a unique optimizer (in law), and the following are equivalent for a process $X = (X_t)_{0\leqslant t\leqslant 1}$ with $X_0 \sim \mu$, $X_1 \sim \nu$:

(1) X solves (BB).
(2) $X_1 = f(X_0)$, where f is the gradient of a convex function $v\colon \mathbb{R}^d \to (-\infty, +\infty]$ and all particles move with constant speed, i.e. $X_t = X_0 + bt$, $t \in [0,1]$, for the random variable $b = X_1 - X_0$.

In this case, setting $\mu_t \coloneqq \mathrm{Law}(X_t)$, $t \in [0,1]$, defines McCann's displacement interpolation between μ and ν. This interpolation is time-consistent in the sense that interpolation between μ_s and μ_t, for $0 \leqslant s < t \leqslant 1$, recovers the family $(\mu_r)_{s\leqslant r\leqslant t}$ up to the obvious affine-time transformation.

1.2. *Martingale optimization problem.* Assume that $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ are in convex order, in signs $\mu \preceq_{\mathrm{c}} \nu$, that is, $\int \phi\, d\mu \leqslant \int \phi\, d\nu$ for all convex functions $\phi\colon \mathbb{R}^d \to \mathbb{R}$ with linear growth. We consider the optimization problem

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

where $|\cdot|_{\mathrm{HS}}$ denotes the Hilbert–Schmidt norm. Here, the infimum is taken over all filtered probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$, with σ an $\mathbb{R}^{d\times d}$-valued $\mathcal{F}$-progressive process and B a d-dimensional $\mathcal{F}$-Brownian motion such that M is a martingale. The stochastic integral $\int_0^t \sigma_s\, dB_s$ is in the Itô sense.

The main result of [5] is that (MBB) admits a unique optimizer M^*. Moreover, M^* is a continuous strong Markov martingale. While (BB) implies that particles move along straight lines, the functional in (MBB) stipulates that M^* maximizes the correlation with Brownian motion subject to the given marginal constraints. It is also shown in [5] that M^* is the process whose evolution follows the movement of a Brownian particle as closely as possible with respect to an *adapted Wasserstein distance* (see, e.g., [4, 29]) subject to the marginal conditions $M_0^* \sim \mu$ and $M_1^* \sim \nu$. As in the classical case, setting $\mu_t \coloneqq (\mathrm{Law}(M_t^*))_{0\leqslant t\leqslant 1}$ defines a time-consistent interpolation between μ and ν. These properties motivate to interpret (MBB) as a Martingale Benamou–Brenier problem and to call the martingale M^* *stretched Brownian motion* between μ and ν as in [5].

1.3. *The Bass martingale.* Our main result relates the optimality property in the definition of stretched Brownian motion to a structural description. As motivation we recall a classical construction of Bass [7] which provides a particularly appealing martingale $M = (M_t)_{0\leqslant t\leqslant 1}$ that terminates in a fixed measure ν on the real line and starts at the barycenter of ν. This amounts to a solution of the Skorokhod embedding problem (modulo time change). Let $(B_t)_{0\leqslant t\leqslant 1}$ be Brownian motion (started in $B_0 = 0$), let $\gamma := \mathrm{Law}(B_1)$ denote the standard Gaussian and define $f\colon \mathbb{R} \to \mathbb{R}$ as the γ-a.e. unique increasing mapping that pushes γ to ν. Bass then defines the martingale

$$M_t := \mathbb{E}[f(B_1) \,|\, \sigma(B_s\colon s \leqslant t)] = \mathbb{E}[f(B_1) \,|\, B_t], \qquad 0 \leqslant t \leqslant 1, \tag{1.1}$$

so that M_0 starts at the barycenter of ν and $\mathrm{Law}(M_1) = \nu$.

As we are interested in martingales in multiple dimensions with possibly non-degenerate starting law μ, we consider the construction of Bass in the following generality:

Definition 1.1. Let $B = (B_t)_{0\leqslant t\leqslant 1}$ be Brownian motion on $\mathbb{R}^d$, where $B_0 \sim \alpha \in \mathcal{P}(\mathbb{R}^d)$, and let $v\colon \mathbb{R}^d \to \mathbb{R}$ be convex such that $\nabla v(B_1)$ is square-integrable. Then we call

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

a *Bass martingale* from $\mu := \mathrm{Law}(M_0)$ to $\nu := \mathrm{Law}(M_1)$.

The question arises under which conditions on $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ there is a Bass martingale M from μ to ν, i.e. satisfying $M_0 \sim \mu$ and $M_1 \sim \nu$, for μ and ν prescribed in advance?

1.4. *Structure of stretched Brownian motion.* We need a connectivity assumption on the marginals, known under the name of *irreducibility* in the martingale transport literature.

Definition 1.2. For $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ 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 $\mathbb{P}(X_0 \in A, X_1 \in B) > 0$.

With this definition in hand, we can announce our first main result.

Theorem 1.3. *Let $\mu \preceq_{\mathrm{c}} \nu$ be probabilities on $\mathbb{R}^d$ with finite second moments and suppose that (μ, ν) is irreducible. Then the following are equivalent for a martingale $M = (M_t)_{0\leqslant t\leqslant 1}$ with $M_0 \sim \mu$ and $M_1 \sim \nu$:*

(1) *M is stretched Brownian motion, i.e. the optimizer of* (MBB).
(2) *M is a Bass martingale.*

We briefly provide some context for the irreducibility assumption in Theorem 1.3. In classical optimal transport, the product coupling of the marginals guarantees that mass can be transported from an arbitrary starting position to an arbitrary target position. However, this is not necessarily true in the case of martingale transport, where it can happen that $\mathbb{R}^d$ decomposes into disjoint (convex) regions that do not communicate with each other. The irreducibility assumption excludes this, as it guarantees that for any sets A and B that are charged by μ and ν, respectively, there exists a martingale X connecting them (in the sense that $\mathbb{P}(X_0 \in A, X_1 \in B) > 0$). For $d > 1$, the appearance of more than one irreducible component leads to intricate phenomena, analyzed in the remarkable contributions [31, 27, 52]. This problem is revisited in the follow-up article [56] in terms of Bass martingales. For equivalent characterizations of the irreducibility assumption we refer to Theorem D.1 in Appendix D. In

particular, it is equivalent to consider continuous- or discrete-time martingales in Definition 1.2.

We emphasize that irreducibility is not only a sufficient assumption for Theorem 1.3, but it is in fact necessary. Indeed, the Bass martingale connects any two sets which are charged by μ and ν, see Remark D.3 in Appendix D. In particular, if there is a Bass martingale from μ to ν, then (μ,ν) is irreducible.

An important consequence of Theorem 1.3 is that for any irreducible pair $\mu \preceq_c \nu$ there exists a unique Bass martingale

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

with $M_0 \sim \mu$, $M_1 \sim \nu$ and it is worthwhile to comment on the properties of M. We write γ^t for the d-dimensional centered Gaussian distribution with covariance matrix tI_d and $v_t := v * \gamma^{1-t}\colon \mathbb{R}^d \to \mathbb{R}$ for the convolution of the function v and the measure γ^{1-t}. In these terms, (1.2) amounts to

$$M_t = \nabla v_t(B_t), \qquad 0 \leqslant t \leqslant 1, \tag{1.3}$$

which emphasizes that the Bass martingale is obtained as a monotone transformation of Brownian motion at each time point. Finally, as a Brownian martingale, M is a diffusion which connects μ and ν. Indeed, applying Itô's formula to (1.3) we obtain

$$dM_t = \nabla^2 v_t \circ \nabla v_t^*(M_t)\, dB_t, \qquad 0 \leqslant t \leqslant 1,$$

where v_t^* denotes the convex conjugate of v_t.

McCann [48] managed to extend the validity of Brenier's theorem: using Aleksandrov's lemma he showed that the Brenier map is well defined without any moment assumptions. We leave the question of whether a similar extension is possible in the case of Bass martingales for future research.

Finally, we note that the implication "(2) $\Rightarrow$ (1)" is comparably easy and follows from [5, Theorem 1.10]. We will establish the reverse implication that every optimizer of (MBB) is a Bass martingale, based on a duality result, which we describe in the next section.

1.5. *Weak transport formulation and dual viewpoint.* Our arguments rely on a novel dual viewpoint on stretched Brownian motion, which is in turn based on a reformulation of problem (MBB) as a weak transport problem.

In classical transport, minimization of the squared distance is equivalent to maximization of correlation and the latter formulation yields a dual problem which is simpler to interpret. A similar fact holds true in the present martingale setting: problem (MBB) is equivalent to maximizing the covariance with Brownian motion

$$P(\mu,\nu) := \sup_{\substack{M_0\sim\mu,\, M_1\sim\nu,\\ M_t = M_0 + \int_0^t \sigma_s\, dB_s}} \mathbb{E}\Big[\int_0^1 \operatorname{tr}(\sigma_t)\, dt\Big], \tag{1.4}$$

in the sense that 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 go on to reformulate (1.4) as a weak (martingale) transport problem in the sense of [35]. Indeed

$$P(\mu,\nu) = \sup_{\pi\in \mathsf{MT}(\mu,\nu)} \int \operatorname{MCov}(\pi_x,\gamma)\, \mu(dx), \tag{1.5}$$

where the *maximal covariance* (MCov) between $p_1, p_2 \in \mathcal{P}_2(\mathbb{R}^d)$ is defined as

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

The optimization problem (1.5) is a *weak* martingale optimal transport problem since the function $\pi \mapsto \int \mathrm{MCov}(\pi_x, \gamma)\, \mu(dx)$ is non-linear, as opposed to classical optimal transport, where linear problems of the form $\pi \mapsto \int c(x,y)\, \pi(dx, dy)$ are studied.

The equality (1.5) was established in [5, Theorem 2.2]. Furthermore, there exists a unique optimizer $\pi^{\mathrm{SBM}} \in \mathsf{MT}(\mu, \nu)$ of (1.5) and if $(M_t)_{0 \leqslant t \leqslant 1}$ is the stretched Brownian motion from μ to ν, then the law of (M_0, M_1) equals π^{SBM}. Indeed, the discrete-time formulation (1.5) will play a central role in this paper. In particular, it will be the basis for the following duality result, which implies the difficult implication "(1) ⇒ (2)" in Theorem 1.3.

THEOREM 1.4. *Assume that $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ are in convex order. The value $P(\mu, \nu)$ of the continuous-time optimization problem (1.4) is equal to*

$$D(\mu, \nu) := \inf_{\substack{\psi \in L^1(\nu), \\ \psi \text{ convex}}} \Big(\int \psi \, d\nu - \int (\psi^* * \gamma)^* \, d\mu \Big). \tag{1.6}$$

The infimum is attained by a lower semicontinuous convex function $\psi_{\mathrm{opt}} \colon \mathbb{R}^d \to (-\infty, +\infty]$ satisfying $\mu(\mathrm{ri}(\mathrm{dom}\, \psi_{\mathrm{opt}})) = 1$ if and only if (μ, ν) is irreducible. In this case the (unique) optimizer to (MBB) is given by the Bass martingale

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

where $v = \psi^*_{\mathrm{opt}}$ and $B_0 \sim \nabla(\psi^*_{\mathrm{opt}} * \gamma)^*(\mu)$.

Here, $\mathrm{ri}(\mathrm{dom}\, \psi_{\mathrm{opt}})$ denotes the relative interior of the domain of ψ_{opt}, i.e. the set on which ψ_{opt} is finite. The symbol $*$ used as a superscript denotes the convex conjugate of a function, otherwise it is the standard convolution operator.

It turns out that the optimizer ψ_{opt} is not necessarily ν-integrable. Therefore, in order for the difference of the integrals in (1.6) to be well defined for $\psi = \psi_{\mathrm{opt}}$, attainment of $D(\mu, \nu)$ has to be understood in a "relaxed" sense frequently encountered in martingale transport problems; see [12, 14, 15] and Propositions 4.1, 4.2 below.

Note that, since the primal optimizer π^{SBM} equals the law of (M_0, M_1), Theorem 1.4 describes the optimizer of the static weak martingale transport problem (1.5) in terms of the gradient of a convex function, which originates from its dual problem (1.6). More precisely, we have

$$\pi^{\mathrm{SBM}} = \mathrm{Law}\big((\nabla v * \gamma)(B_0), \nabla v(B_1)\big).$$

This is analogous to the classical Brenier theorem and hence one might view Theorem 1.4 as a Brenier-type theorem for weak martingale transport.

We emphasize that while Brenier's theorem is a direct consequence of attainment for the dual transport problem (which is now well understood), the situation is more delicate in case of Theorem 1.4. The main reason is that dual attainment for martingale transport can fail even in very regular settings, e.g. for Lipschitz costs and compactly supported measures on the real line, see [10, Section 4.3]. Positive results are only available for $d = 1$ and under strong assumptions [12, 13, 15]. In addition, the duality theory for weak optimal transport is known to be significantly more complicated than its classical counterpart. Specifically first results for dual attainment have appeared only recently [16] and are not applicable to weak martingale transport problems. Accordingly, the technical core of our work is to establish dual attainment in the framework of Theorem 1.4.

1.6. *Bass martingales for specific marginals.* Theorem 1.3 is only an existence result, which does not provide an explicit Bass martingale connecting the given marginals. This is parallel to the situation in optimal transport, where the question to numerically approximate optimal transport plans has generated enormous interest. Already in dimension one, the problem to explicitly determine Bass martingales is non-trivial (as well as highly important for financial applications). Here [25] gives an efficient iteration algorithm. See [1] for a rigorous proof of the algorithm's linear convergence and [42] for a multi-dimensional variant of the algorithm.

On the other hand (as in the case of Brenier's theorem), one can easily generate examples of Bass martingales by specifying a convex function v and the starting distribution of B_0. We give here one such example that also serves to illustrate some of the intricacies related to irreducible decompositions. Fix parameters $\rho, k > 0$ and consider the radially symmetric function $v(x,y) = h(|(x,y)|)$, where h is continuous and piecewise affine with slope $1/2$ on $[0,k]$, slope $8/5$ on $[k,\infty)$, and $h(0) = 0$. We consider the Bass martingale $M_t = \mathbb{E}[\nabla v(B_1) \,|\, \mathcal{F}_t]$, where B_0 is uniformly distributed on the centered circle with radius ρ. Then $\nabla v(x) = \frac{x}{2|x|}$, $|x| \in (0,k)$ and $\nabla v(x) = \frac{8x}{5|x|}$, $|x| \in (k,\infty)$, from which it follows easily that M terminates on either the circle with radius $1/2$ or the circle with radius $8/5$ and that M starts uniformly distributed on some centered circle. Specifying $\rho = 3$, $k = 3.17$, we find (numerically) that M_0 is uniformly distributed on the unit circle, while M_1 charges the circles with radius $1/2$ and $8/5$ equally, see Figures 1 and 2.

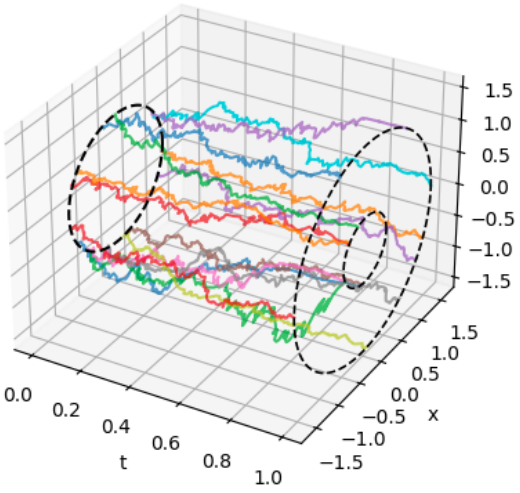

Fig 1. Sample paths of M, different starting values.

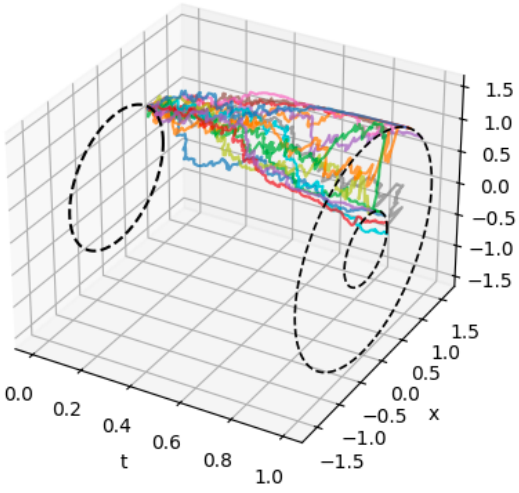

Fig 2. Sample paths of M, identical starting value.

Setting $\mu = \operatorname{Law} M_0$, $\nu = \operatorname{Law} M_1$, we have, of course, $\mathbb{P}((M_0, M_1) \in A \times B) > 0$ whenever $\mu(A), \nu(B) > 0$.

Notably this is not the case for every martingale M starting in μ and terminating in ν. Specifically we may consider martingales that move only along the (violet) segments indicated in Figure 3, we refer to [56, Example 6.2] for a detailed justification of this claim.

This example illustrates that even if (μ, ν) is irreducible, there can be martingales which are entirely confined to smaller subpartitions. This phenomenon cannot appear in the trivial case $d = 1$, see [12, Appendix A], a fact that is heavily exploited in dual attainment results for martingale transport in $d = 1$. On the other hand, we believe that the existence of martingales confined to smaller subpartitions present a significant obstacle to general dual attainment results for martingale transport (and weak martingale transport) for $d \geqslant 2$.

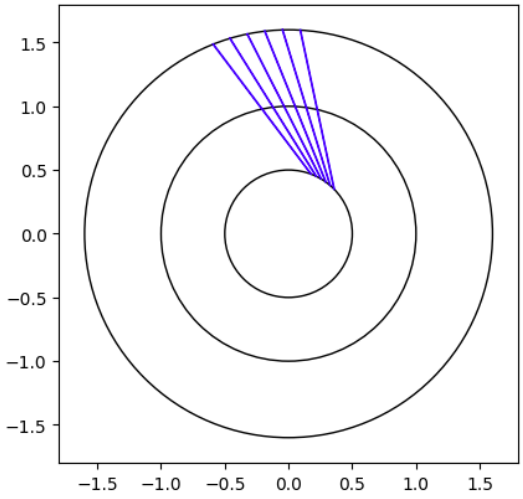

Fig 3. There is a martingale which starts uniformly distributed on the middle circle ($r = 1$), terminates with equal probability on the inner circle ($r = 1/2$) or outer circle ($r = 8/5$), and moves only on the violet line segments. The angle between the segments and the middle circle is chosen so that the lengths to the inner and outer circle are equal.

1.7. *Literature.* The first article that provides structural results for the martingale transport problem in general dimensions is the work [31] of Ghoussoub–Kim–Lim. They obtain descriptions for the minimizers and maximizers for the cost function $c(x, y) = |x - y|$, when marginals are supported on $\mathbb{R}^2$, as well as for marginals on higher-dimensional state spaces that are in subharmonic order. Given a specific martingale X, Ghoussoub–Kim–Lim also define a finest paving of the source space into cells that are invariant under the martingale X.

The contributions of De March–Touzi [27] and Obłój–Siorpaes [52] put the theme of irreducible decompositions center stage. In contrast to the work of Ghoussoub–Kim–Lim, their interest lies in pavings that are invariant under *all* martingales which start in μ and terminate in ν. Specifically it is shown in [27] that there exists a unique finest paving with this property. This De March–Touzi paving is revisited in the follow-up paper [56], where it is characterized in terms of Bass martingales. While the Ghoussoub–Kim–Lim paving and the De March–Touzi paving agree for $d = 1$, they can be different for $d \geqslant 2$, as discussed in the previous section. The interested reader is referred to Ciosmak's works [23, 24], where a more general notion of irreducibility (going beyond martingale transports) is studied.

Huesmann–Trevisan [41] investigate Benamou–Brenier-type formulations for the martingale transport problem on $\mathbb{R}^d$. In particular, they provide equivalent PDE-formulations and establish existence and duality results. In the context of market impact in finance Loeper [46] arrives independently to problem (MBB).

For further contributions to the martingale transport in continuous time, we mention [28, 11, 26, 33, 37, 32, 22, 36] among many others.

Finally, [5] is a predecessor of the present article in which it is established that the martingale transport problem (MBB) admits a unique solution, solves further related optimization problems and has properties of time consistency in the spirit of classical optimal transport for the squared distance cost function. Furthermore, the counterpart of our main result is established in dimension one (without reference to duality). As discussed in length above, the main difficulty in higher dimensions stems from the subtleties surrounding the concept of irreducible component.

1.8. *Structure of the paper.* In Section 2 we introduce some definitions and frequently used notation. Duality results for stretched Brownian motion and the important role of convexity are discussed in Section 3. The proof of the first part of Theorem 1.4, namely that there is no duality gap between the primal problem (1.4) and the dual problem (1.6), is given in Section 4. In Section 5 we outline and prepare the proof of the second part of Theorem 1.4, which gives a necessary and sufficient condition for dual attainment in terms of the irreducibility assumption. To this end, we analyze the connection between the existence of dual optimizers and Bass martingales, which is the content of Section 6. In Section 7 we show that the irreducibility assumption implies the existence of a dual optimizer. After these preparations we are in a position to prove Theorem 1.3 in Subsection 7.4 and complete the proof of Theorem 1.4 in Subsection 7.5.

In Appendix A we prove Theorem 3.3, which shows that there is no duality gap between the auxiliary optimization problems (1.5) and (3.1). The rather technical proofs of Lemmas 4.4, 5.1, 6.2, 6.4 and 6.5 are collected in Appendix B. In Appendix C we provide the proof of Proposition 7.20, a result which will be of crucial importance in the follow-up paper [56] on the non-irreducible case, but which also seems of independent interest. Finally, in Appendix D we give equivalent characterizations of irreducibility, as introduced in Definition 1.2.

CONTENTS

1 Introduction . . . 1
1.1 Benamou–Brenier transport problem and McCann interpolation in probabilistic terms . . . 1
1.2 Martingale optimization problem . . . 2
1.3 The Bass martingale . . . 3
1.4 Structure of stretched Brownian motion . . . 3
1.5 Weak transport formulation and dual viewpoint . . . 4
1.6 Bass martingales for specific marginals . . . 6
1.7 Literature . . . 7
1.8 Structure of the paper . . . 8
2 Definitions and notation . . . 10
3 Duality for stretched Brownian motion . . . 11
4 Proof of the first part of Theorem 1.4 . . . 14
5 Preparation for the proof of the second part of Theorem 1.4 . . . 18
6 Existence of dual optimizers and Bass martingales . . . 22
7 Irreducibility and existence of dual optimizers . . . 26
7.1 Outline and objectives . . . 26
7.2 Irreducibility and dual attainment . . . 28
7.3 Proof of Theorem 7.6 . . . 29
7.3.1 Step 1 of the proof of Theorem 7.6 . . . 29
7.3.2 Step 2 of the proof of Theorem 7.6 . . . 31
7.4 Proof of Theorem 1.3 . . . 36
7.5 Proof of the second part of Theorem 1.4 . . . 37
A Proof of Theorem 3.3 . . . 37
B Proofs of Lemmas 4.4, 5.1, 6.2, 6.4 and 6.5 . . . 38
C Proof of Proposition 7.20 . . . 42
D Characterization of irreducibility . . . 45
Acknowledgments . . . 46
Funding . . . 46
References . . . 46

In response to the thoughtful suggestions of a diligent referee, we provide in Figure 4 an overview of the logical structure of our paper. This is intended to help readers navigate the dependencies between the results of this work, as well as the related literature, more effectively.

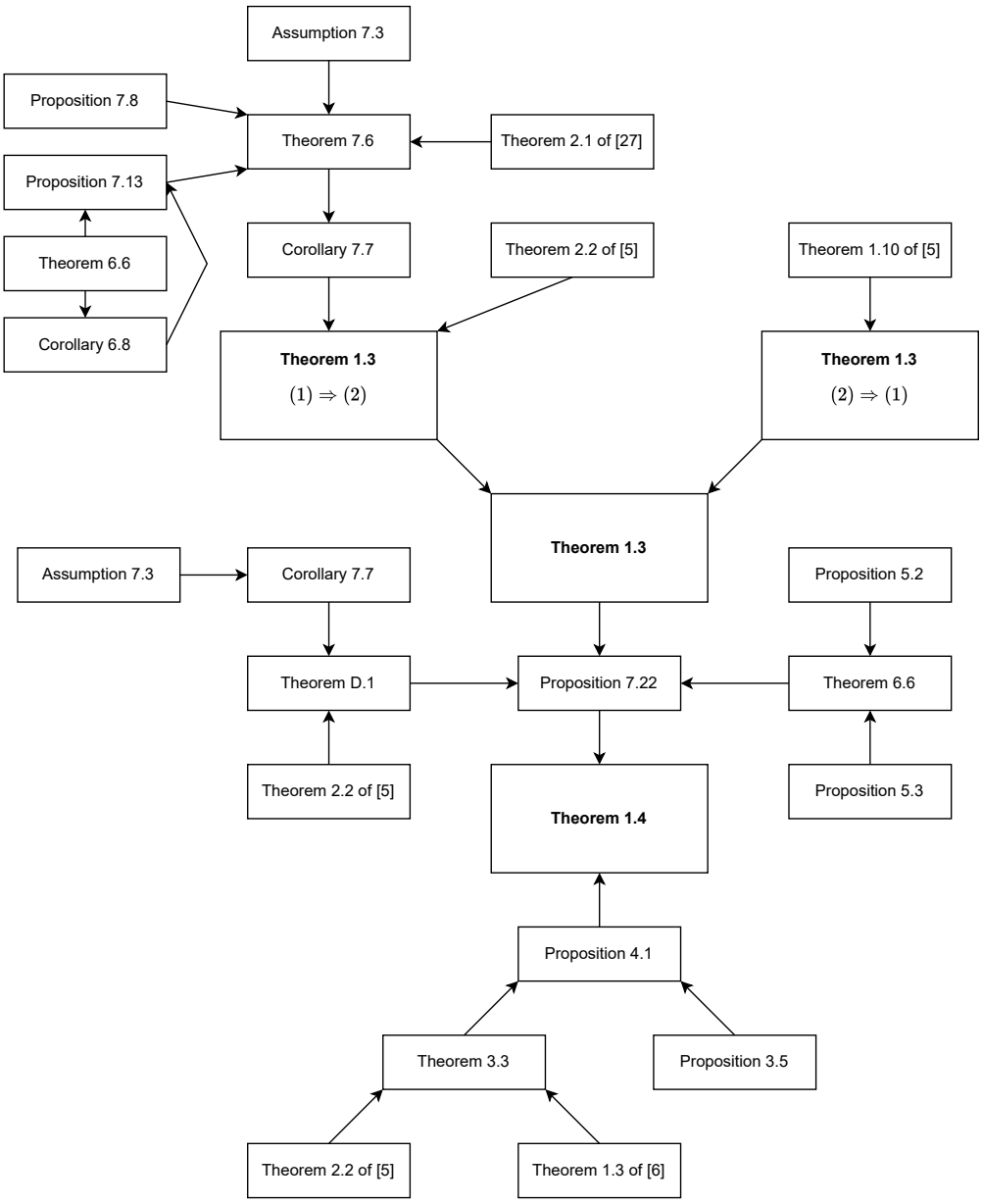

Fig 4. Dependencies between results.

2. Definitions and notation.

- We write $\mathcal{P}(\mathbb{R}^d)$ for the probability measures on $\mathbb{R}^d$, $\mathcal{P}_p(\mathbb{R}^d)$ for the subset of probability measures satisfying $\int |x|^p \, d\mu < +\infty$, and $\mathcal{P}_p^x(\mathbb{R}^d)$ for the elements of $\mathcal{P}_p(\mathbb{R}^d)$ with barycenter $x \in \mathbb{R}^d$, for $p \in [1, +\infty)$.
- For $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, we denote by $\mathsf{Cpl}(\mu, \nu)$ the set of all couplings $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ between μ and ν, i.e., probability measures π on $\mathbb{R}^d \times \mathbb{R}^d$ with first marginal μ and second marginal ν.
- We say that $\mu \in \mathcal{P}_1(\mathbb{R}^d)$ is dominated by $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ in convex order and write $\mu \preceq_{\mathrm{c}} \nu$, if for all convex functions $\phi \colon \mathbb{R}^d \to \mathbb{R}$ with linear growth we have $\int \phi \, d\mu \leqslant \int \phi \, d\nu$.
- For $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$ we define the collection of martingale transports $\mathsf{MT}(\mu, \nu)$ as those couplings $\pi \in \mathsf{Cpl}(\mu, \nu)$ satisfying $\mathrm{bary}(\pi_x) \coloneqq \int y \, \pi_x(dy) = x$, for μ-a.e. $x \in \mathbb{R}^d$. Here, the family of probability measures $\{\pi_x\}_{x \in \mathbb{R}^d} \subseteq \mathcal{P}(\mathbb{R}^d)$ is obtained by disintegrating the coupling π with respect to its first marginal μ, i.e., $\pi(dx, dy) = \pi_x(dy) \, \mu(dx)$.
- The d-dimensional Gaussian distribution with barycenter $x \in \mathbb{R}^d$ and covariance matrix tI_d is denoted by γ_x^t and we set $\gamma^t \coloneqq \gamma_0^t$, $\gamma_x \coloneqq \gamma_x^1$ as well as $\gamma \coloneqq \gamma_0$.
- We denote by $C_{\mathrm{q}}(\mathbb{R}^d)$ the set of continuous functions $\psi \colon \mathbb{R}^d \to \mathbb{R}$ with quadratic growth, meaning that there are constants $a, k, \ell \in \mathbb{R}$ with

$$\ell + \tfrac{|\cdot|^2}{2} \leqslant \psi(\cdot) \leqslant a + k|\cdot|^2.$$

  We also introduce the set

$$C_{\mathrm{q}}^{\mathrm{aff}}(\mathbb{R}^d) \coloneqq \big\{\psi(\cdot) + \mathrm{aff}(\cdot) \colon \ \psi \in C_{\mathrm{q}}(\mathbb{R}^d), \, \mathrm{aff} \colon \mathbb{R}^d \to \mathbb{R} \text{ affine}\big\}. \tag{2.1}$$

- The pushforward of a measure μ under a function f is denoted by $f(\mu)$.
- For two measures ϱ and ρ we write $\varrho * \rho$ for their convolution. If f is a function, the convolution of f and ρ is defined as

$$(f * \rho)(x) \coloneqq \int f(x - y) \, \rho(dy).$$

  In particular, $(f * \gamma)(x) = \int f(x + y) \, \gamma(dy)$.
- For a function $f \colon \mathbb{R}^d \to (-\infty, +\infty]$, its convex conjugate is given by

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

  and we write

$$\mathrm{dom}\, f \coloneqq \{x \in \mathbb{R}^d \colon \, f(x) < +\infty\}$$

  for the domain of f. If f is convex and $\mathrm{dom}\, f \neq \varnothing$, we say that f is a proper convex function. If f is additionally lower semicontinuous, then $f = f^{**}$ by the Fenchel–Moreau theorem, a fact that we will use repeatedly in this work.
- The operators int and ri denote the interior and relative interior of a set, respectively. We write $\overline{A}$ for the closure of a set $A \subseteq \mathbb{R}^d$.
- The operator conv applied to a function or a set denotes the convex hull. The convex hull $\mathrm{conv}\, \psi$ of a function ψ is the greatest convex function smaller or equal to ψ.
- The support and the closed convex hull of the support of a measure ρ are denoted by $\mathrm{supp}(\rho)$ and $\widehat{\mathrm{supp}}(\rho)$, respectively.
- The symbol ∂, applied to a convex function, denotes its subdifferential or — by abuse of notation — also a subgradient.

3. Duality for stretched Brownian motion. We fix $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$. Recall from (1.5) that our main focus will lie on the (primal) *weak martingale transport problem*

$$P(\mu,\nu) \coloneqq \sup_{\pi \in \mathsf{MT}(\mu,\nu)} \int \mathrm{MCov}(\pi_x, \gamma)\, \mu(dx),$$

where the maximal covariance (MCov) between $p_1, p_2 \in \mathcal{P}_2(\mathbb{R}^d)$ is given by

$$\mathrm{MCov}(p_1,p_2) \coloneqq \sup_{q \in \mathsf{Cpl}(p_1,p_2)} \int \langle x_1, x_2 \rangle\, q(dx_1, dx_2).$$

Throughout this paper we shall make the following assumption.

Assumption 3.1. The support of ν affinely spans $\mathbb{R}^d$.

This assumption will be convenient, mainly for notational reasons. But it does not restrict the generality of the results of this paper. Indeed, let A denote the affine space spanned by $\mathrm{supp}(\nu)$. For $\pi(dx,dy) = \pi_x(dy)\,\mu(dx) \in \mathsf{MT}(\mu,\nu)$ we have, for μ-a.e. $x \in \mathbb{R}^d$, that $\mathrm{supp}(\pi_x) \subseteq \mathrm{supp}(\nu) \subseteq A$. Choose an arbitrary point $x_0 \in A$ and denote by $\gamma^A_{x_0}$ the standard Gaussian distribution on the affine space A centered at x_0. It is easy to see that there is a constant $c = c(x_0, A)$ such that, for $\pi \in \mathsf{MT}(\mu,\nu)$, we have

$$\int \mathrm{MCov}(\pi_x, \gamma)\, \mu(dx) = \int \mathrm{MCov}(\pi_x, \gamma^A_{x_0})\, \mu(dx) + c.$$

In other words, replacing γ by $\gamma^A_{x_0}$ in (1.5) only changes the primal problem in a trivial way by adding a constant. Denoting by m the affine dimension of A, we can choose an isometry from A to $\mathbb{R}^m$ which maps x_0 to 0 and which maps μ, ν to measures $\overline{\mu}, \overline{\nu}$ on $\mathbb{R}^m$. In this way we have transformed the primal problem (1.5) for general $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$ to the problem (1.5) for $\overline{\mu}, \overline{\nu} \in \mathcal{P}_2(\mathbb{R}^m)$ with $\overline{\mu} \preceq_c \overline{\nu}$ such that $\mathrm{supp}(\overline{\nu})$ affinely spans $\mathbb{R}^m$. In conclusion, we shall assume without loss of generality Assumption 3.1 throughout this paper.

We define the dual problem

$$\tilde{D}(\mu,\nu) \coloneqq \inf_{\psi \in C_q(\mathbb{R}^d)} \Big(\int \psi\, d\nu - \int \varphi^\psi\, d\mu \Big), \tag{3.1}$$

where the function $\mathbb{R}^d \ni x \mapsto \varphi^\psi(x)$ is given by

$$\varphi^\psi(x) \coloneqq \inf_{p \in \mathcal{P}_2^x(\mathbb{R}^d)} \Big(\int \psi\, dp - \mathrm{MCov}(p,\gamma) \Big). \tag{3.2}$$

Lemma 3.2. *Let $\psi : \mathbb{R}^d \to (-\infty, +\infty]$ be a proper, lower semicontinuous convex function. The function φ^ψ is convex on* $\mathrm{dom}\,\psi$ *and $\varphi^\psi \leqslant \psi$.*

Proof. The convexity of $x \mapsto \varphi^\psi(x)$ is repeated and proved in Lemma 5.1 below. By taking $p = \delta_x$ in (3.2), we immediately see that $\varphi^\psi(x) \leqslant \psi(x)$. □

Theorem 3.3. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$. There is no duality gap between the primal problem (1.5) and the dual problem (3.1), i.e., $P(\mu,\nu) = \tilde{D}(\mu,\nu)$. Moreover, the primal problem is uniquely attained and has a finite value, i.e., there exists a unique $\pi^{\mathrm{SBM}} \in \mathsf{MT}(\mu,\nu)$ such that*

$$P(\mu,\nu) = \int \mathrm{MCov}(\pi^{\mathrm{SBM}}_x, \gamma)\, \mu(dx) < +\infty. \tag{3.3}$$

Unique attainment of the primal problem is a consequence of the results in [5], while duality can be derived from the general duality results for weak optimal transport provided in [6], see Appendix A for details. To motivate (3.1) as a plausible dual formulation of (1.5) we provide some heuristics:

Indeed, considering

$$\chi(\pi) := \inf_{\psi \in C_q(\mathbb{R}^d)} \left(\int \psi \, d\nu - \int \psi \, d\pi_x \, d\mu(x) \right) = \begin{cases} 0, & \text{if } \int \pi_x \, d\mu(x) = \nu, \\ -\infty, & \text{else,} \end{cases}$$

we *formally* obtain the desired duality relation by interchanging inf and sup:

$$\begin{aligned} P(\mu,\nu) &= \sup_{\substack{\pi(dx,dy)=\pi_x(dy)\,\mu(dx),\\ \pi_x \in \mathcal{P}_2^x(\mathbb{R}^d)}} \left(\int \mathrm{MCov}(\pi_x,\gamma)\, d\mu(x) + \chi(\pi) \right) \\ &= \inf_{\psi \in C_q(\mathbb{R}^d)} \left(\int \sup_{\pi_x \in \mathcal{P}_2^x(\mathbb{R}^d)} \left(\mathrm{MCov}(\pi_x,\gamma) + \int \psi \, d(\nu - \pi_x) \right) d\mu(x) \right) = \tilde{D}(\mu,\nu). \end{aligned}$$

In the remainder of this section we make the important observation that in the dual problem (3.1) it suffices to optimize over the class of functions $\psi \in C_q(\mathbb{R}^d)$ which are *convex*. We refer to [34, Theorem 1.1] for an analogous result. Another similar restriction to convex functions in the setting of weak transport costs has been shown in [35, Theorem 2.11, (3)].

We then show in Lemma 3.7 below that it is also equivalent to optimize over all convex functions $\psi : \mathbb{R}^d \to (-\infty, +\infty]$ which are only μ-a.s. finite, but not necessarily of quadratic growth.

Given some $\psi \in C_q(\mathbb{R}^d)$, it will be convenient to have an explicit representation for the convex hull of ψ, as in (3.4) below. This identity is usually stated in the more specific form

$$(\operatorname{conv}\psi)(y) = \inf \left\{ \sum_{i=1}^{d+1} \lambda_i \psi(y_i) : \sum_{i=1}^{d+1} \lambda_i y_i = y \right\},$$

where the infimum is taken over all expressions of y as a convex combination of $d+1$ points, see [54, Corollary 17.1.5].

Lemma 3.4. *Let $\psi \in C_q(\mathbb{R}^d)$. Then the convex hull* $\operatorname{conv}\psi$ *satisfies*

$$(\operatorname{conv}\psi)(y) = \inf_{p \in \mathcal{P}_2^y(\mathbb{R}^d)} \int \psi \, dp, \qquad y \in \mathbb{R}^d \tag{3.4}$$

and again $\operatorname{conv}\psi \in C_q(\mathbb{R}^d)$.

Recalling the dual problem (3.1), we define the dual function

$$\mathcal{D}(\psi) := \int \psi \, d\nu - \int \varphi^\psi \, d\mu, \tag{3.5}$$

for $\psi \in C_q(\mathbb{R}^d)$. Now we can prove our crucial observation, that it suffices to optimize the dual function over the class of functions $\psi \in C_q(\mathbb{R}^d)$ which are *convex*.

Proposition 3.5. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$. Then $\mathcal{D}(\operatorname{conv}\psi) \leqslant \mathcal{D}(\psi)$ for all $\psi \in C_q(\mathbb{R}^d)$ and consequently*

$$\tilde{D}(\mu,\nu) = \inf_{\psi \in C_q(\mathbb{R}^d)} \mathcal{D}(\psi) = \inf_{\substack{\psi \in C_q(\mathbb{R}^d),\\ \psi \text{ convex}}} \mathcal{D}(\psi). \tag{3.6}$$

Proof. Let $\varepsilon > 0$, $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$ and $\{p_x\}_{x\in\mathbb{R}^d} \subseteq \mathcal{P}_2(\mathbb{R}^d)$ be a measurable collection of probability measures with $\mathrm{bary}(p_x) = x$. To show the claim, it is sufficient to construct a measurable family $\{\bar{p}_x\}_{x\in\mathbb{R}^d} \subseteq \mathcal{P}_2(\mathbb{R}^d)$ with $\mathrm{bary}(\bar{p}_x) = x$ such that

$$\mathrm{MCov}(p_x,\gamma) + \int \mathrm{conv}\,\psi\, d(\nu - p_x) \leqslant \mathrm{MCov}(\bar{p}_x,\gamma) + \int \psi\, d(\nu - \bar{p}_x) + \varepsilon. \tag{3.7}$$

Let us construct appropriate probability measures $\{\bar{p}_x\}_{x\in\mathbb{R}^d}$. By Lemma 3.4 and a measurable selection argument we can choose a measurable collection of probability measures $\{\tilde{p}_y\}_{y\in\mathbb{R}^d} \subseteq \mathcal{P}_2(\mathbb{R}^d)$ with $\mathrm{bary}(\tilde{p}_y) = y$ such that

$$\int \psi\, d\tilde{p}_y \leqslant (\mathrm{conv}\,\psi)(y) + \varepsilon. \tag{3.8}$$

Then we define $\bar{p}_x(dz) \coloneqq \int_y \tilde{p}_y(dz)\, p_x(dy)$, so that $\mathrm{bary}(\bar{p}_x) = x$. Integrating (3.8) with respect to $p_x(dy)$ yields

$$\int \psi\, d\bar{p}_x \leqslant \int \mathrm{conv}\,\psi\, dp_x + \varepsilon. \tag{3.9}$$

Since $\psi, \mathrm{conv}\,\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$ and $p_x \in \mathcal{P}_2^x(\mathbb{R}^d)$ we conclude

$$\ell + \tfrac{1}{2}\int |y|^2\, \bar{p}_x(dy) \leqslant a + k\int |y|^2\, p_x(dy) + \varepsilon < +\infty,$$

so that $\bar{p}_x \in \mathcal{P}_2^x(\mathbb{R}^d)$.

In order to show the inequality (3.7), we first observe that $p_x \preceq_{\mathrm{c}} \bar{p}_x$ by Jensen's inequality. Together with the Kantorovich duality (see, e.g. [59, Theorem 5.10]), we conclude that[1]

$$\begin{aligned} \mathrm{MCov}(p_x,\gamma) &= \inf_{\substack{f\colon \mathbb{R}^d\to\mathbb{R}\\ \text{convex}}} \Big(\int f\, dp_x + \int f^*\, d\gamma \Big) \\ &\leqslant \inf_{\substack{f\colon \mathbb{R}^d\to\mathbb{R}\\ \text{convex}}} \Big(\int f\, d\bar{p}_x + \int f^*\, d\gamma \Big) = \mathrm{MCov}(\bar{p}_x,\gamma). \end{aligned} \tag{3.10}$$

On the other hand, from $\mathrm{conv}\,\psi \leqslant \psi$ and (3.9) we have the inequality

$$\int \mathrm{conv}\,\psi\, d(\nu - p_x) \leqslant \int \psi\, d(\nu - \bar{p}_x) + \varepsilon. \tag{3.11}$$

Finally, summing (3.10) and (3.11), we obtain the inequality (3.7). □

Remark 3.6. On the right-hand side of (3.6) we can further require ψ to be smooth. Indeed, if $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$ is convex, then $\psi^\varepsilon \coloneqq \psi * \gamma^\varepsilon$ is a smooth convex function in $C_{\mathrm{q}}(\mathbb{R}^d)$ and

$$\liminf_{\varepsilon\to 0} \Big(\int \psi^\varepsilon\, d\nu - \int \varphi^{\psi^\varepsilon}\, d\mu \Big) \leqslant \int \psi\, d\nu - \int \varphi^\psi\, d\mu,$$

as follows by dominated convergence and the inequality $\psi^\varepsilon \geqslant \psi$.

For $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$, we recall the definition (3.5) of the dual function $\mathcal{D}(\,\cdot\,)$, which we rewrite as

$$\mathcal{D}(\psi) = \int \Big(\int \psi(y)\, \pi_x^{\mathrm{SBM}}(dy) - \varphi^\psi(x) \Big)\, \mu(dx), \tag{3.12}$$

[1] In fact, $p_x \preceq_{\mathrm{c}} \bar{p}_x$ is equivalent to the inequality $\mathrm{MCov}(p_x, q) \leqslant \mathrm{MCov}(\bar{p}_x, q)$ being valid for all probability measures $q \in \mathcal{P}_2(\mathbb{R}^d)$, see. e.g. [2, Theorem 1] or [60, Corollary 1.2].

where $\pi^{\mathrm{SBM}}(dx,dy) = \pi_x^{\mathrm{SBM}}(dy)\,\mu(dx)$ is the unique optimizer of the primal problem (1.5). The "relaxed" representation (3.12) of (3.5) allows us to extend the definition of the dual function $\mathcal{D}(\,\cdot\,)$ to convex functions $\psi\colon \mathbb{R}^d \to (-\infty,+\infty]$ which are not confined to be in $C_{\mathrm{q}}(\mathbb{R}^d)$, but which are only required to be μ-a.s. finite, i.e., satisfy $\mu(\mathrm{dom}\,\psi) = 1$. This is summarized in the following lemma.

LEMMA 3.7. *Let $\mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$ and let $\pi^{\mathrm{SBM}} \in \mathsf{MT}(\mu,\nu)$ be the unique optimizer of the primal problem (1.5). Let $\psi\colon \mathbb{R}^d \to (-\infty,+\infty]$ be a convex function which is μ-a.s. finite. Formula (3.12) then defines $\mathcal{D}(\psi) \in [0,+\infty]$ and, recalling (3.1), we have the inequality*

$$\mathcal{D}(\psi) \geqslant \tilde{D}(\mu,\nu). \tag{3.13}$$

In particular, recalling (3.6), we have

$$\tilde{D}(\mu,\nu) = \inf_{\substack{\psi\in C_{\mathrm{q}}(\mathbb{R}^d),\\ \psi \text{ convex}}} \mathcal{D}(\psi) = \inf_{\substack{\mu(\mathrm{dom}\,\psi)=1,\\ \psi \text{ convex}}} \mathcal{D}(\psi). \tag{3.14}$$

PROOF. Let $\psi\colon \mathbb{R}^d \to (-\infty,+\infty]$ be a convex function which is μ-a.s. finite. First, note that by Jensen's inequality we have

$$\int \psi(y)\,\pi_x^{\mathrm{SBM}}(dy) \geqslant \psi\Big(\int y\,\pi_x^{\mathrm{SBM}}(dy)\Big) = \psi(x).$$

Together with the inequality $\varphi^{\psi}(x) \leqslant \psi(x)$ (recall Lemma 3.2), we conclude that

$$\int \psi(y)\,\pi_x^{\mathrm{SBM}}(dy) - \varphi^{\psi}(x) \geqslant 0, \tag{3.15}$$

for μ-a.e. $x \in \mathbb{R}^d$, so that $\mathcal{D}(\psi) \in [0,+\infty]$. In order to prove the inequality (3.13), we take $p = \pi_x^{\mathrm{SBM}}$ in (3.2). This gives

$$\varphi^{\psi}(x) \leqslant \int \psi\, d\pi_x^{\mathrm{SBM}} - \mathrm{MCov}(\pi_x^{\mathrm{SBM}},\gamma)$$

and hence

$$\mathcal{D}(\psi) \geqslant \int \mathrm{MCov}(\pi_x^{\mathrm{SBM}},\gamma)\,\mu(dx) = P(\mu,\nu) = \tilde{D}(\mu,\nu),$$

where the last equality follows from Theorem 3.3. Thus (3.13) is satisfied.

Finally, from (3.13) and recalling (3.6), we conclude

$$\tilde{D}(\mu,\nu) \leqslant \inf_{\substack{\mu(\mathrm{dom}\,\psi)=1,\\ \psi \text{ convex}}} \mathcal{D}(\psi) \leqslant \inf_{\substack{\psi\in C_{\mathrm{q}}(\mathbb{R}^d),\\ \psi \text{ convex}}} \mathcal{D}(\psi) = \tilde{D}(\mu,\nu),$$

which shows (3.14). □

4. Proof of the first part of Theorem 1.4. In Subsection 1.5 we already noted (by citing Theorem 2.2 of [5]) that the value of the continuous-time optimization problem (1.4) is equal to the value of the discrete-time formulation (1.5). We also know from Theorem 3.3 that there is no duality gap between the primal problem (1.5) and the dual problem (3.1), i.e., $P(\mu,\nu) = \tilde{D}(\mu,\nu)$. Therefore we can formulate the first part of Theorem 1.4 equivalently as follows.

PROPOSITION 4.1. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$. Then $\tilde{D}(\mu,\nu)$ is equal to*

$$D(\mu,\nu) = \inf_{\substack{\psi \in L^1(\nu),\\ \psi \text{ convex}}} \Big(\int \psi \, d\nu - \int (\psi^* * \gamma)^* \, d\mu \Big). \tag{4.1}$$

In Subsection 7.5 we will prove the second part of Theorem 1.4 and in particular discuss the existence of dual optimizers of (4.1). Since in general we cannot expect a dual optimizer to be integrable with respect to the probability measure ν, we need the following "relaxed" formulation of Proposition 4.1.

PROPOSITION 4.2. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$. Then $\tilde{D}(\mu,\nu)$ is equal to*

$$D_{\mathrm{rel}}(\mu,\nu) := \inf_{\substack{\mu(\operatorname{dom}\psi)=1,\\ \psi \text{ convex}}} \mathcal{E}(\psi), \tag{4.2}$$

where

$$\mathcal{E}(\psi) := \int \Big(\int \psi(y) \, \pi_x^{\mathrm{SBM}}(dy) - (\psi^* * \gamma)^*(x) \Big) \mu(dx), \tag{4.3}$$

with π^{SBM} the unique optimizer of the primal problem (1.5).

The main idea behind the proofs of Propositions 4.1, 4.2 (which we will present at the end of this section) is to apply Proposition 3.5 and then to show that $\varphi^\psi = (\psi^* * \gamma)^*$, for every convex function $\psi \in C_q(\mathbb{R}^d)$. This motivates our next goal, namely to solve the minimization problem (3.2), which we rewrite as a maximization problem

$$-\varphi^\psi(x) = \sup_{p \in \mathcal{P}_2^x(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,\gamma) - \int \psi \, dp \Big). \tag{4.4}$$

As a preliminary step, we consider the simpler problem

$$\varrho^\psi := \sup_{p \in \mathcal{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,\gamma) - \int \psi \, dp \Big), \tag{4.5}$$

where we do not prescribe the barycenter x of $p \in \mathcal{P}_2(\mathbb{R}^d)$. In Lemma 4.4 below we will show for an arbitrary proper convex function $\psi : \mathbb{R}^d \to (-\infty, +\infty]$, that the value ϱ^ψ equals $\int \psi^* \, d\gamma$. In fact, this information is essentially enough to prove Propositions 4.1, 4.2.

Solving the maximization problem (4.5) leads to an interesting connection with Brenier maps in Lemma 4.3 below. By *Brenier's theorem* (see, e.g., [58, Theorem 2.12]), the optimal transport for quadratic cost between γ and $p \in \mathcal{P}_2(\mathbb{R}^d)$ is induced by the γ-a.e. defined gradient ∇v of some convex function $v : \mathbb{R}^d \to \mathbb{R}$ via $(\nabla v)(\gamma) = p$.

LEMMA 4.3 ("reverse Brenier"). *Let $v : \mathbb{R}^d \to \mathbb{R}$ be a finite-valued convex function and $\psi := v^*$ its convex conjugate. Assume that the probability measure $\hat{p} := (\nabla v)(\gamma)$ has finite second moment. Then $\hat{p}$ is the unique maximizer of the optimization problem (4.5) and*

$$\varrho^\psi = \int v \, d\gamma = \int \psi^* \, d\gamma < +\infty.$$

PROOF. We denote by T_γ^p the Brenier map from γ to $p \in \mathcal{P}_2(\mathbb{R}^d)$ and note that $T_\gamma^{\hat{p}} = \nabla v$. Since the convex function v is finite-valued, its gradient exists γ-a.e. and we have that

$$v(z) = \sup_{y \in \mathbb{R}^d} \big(\langle y, z \rangle - v^*(y) \big) = \langle \nabla v(z), z \rangle - v^*(\nabla v(z)),$$

for γ-a.e. $z \in \mathbb{R}^d$. Using these observations, for $p \in \mathcal{P}_2(\mathbb{R}^d)$ we get

$$\begin{aligned}
\mathrm{MCov}(p,\gamma) - \int \psi \, dp &= \int \Big(\big\langle T_\gamma^p(z), z \big\rangle - v^*\big(T_\gamma^p(z)\big) \Big) \, \gamma(dz) \\
&\leqslant \int \sup_{y \in \mathbb{R}^d} \big(\langle y, z \rangle - v^*(y) \big) \, \gamma(dz) = \int v(z) \, \gamma(dz) \\
&= \int \Big(\big\langle \nabla v(z), z \big\rangle - v^*\big(\nabla v(z)\big) \Big) \, \gamma(dz) \\
&= \mathrm{MCov}(\hat{p},\gamma) - \int \psi \, d\hat{p},
\end{aligned}$$

with equality if and only if $T_\gamma^p(z) = T_\gamma^{\hat{p}}(z)$, for γ-a.e. $z \in \mathbb{R}^d$. This in turn is the case if and only if $p = \hat{p}$. Finally, from the convexity of v and the Cauchy–Schwarz inequality we obtain

$$\int |v| \, d\gamma \leqslant |v(0)| + \sqrt{\int |\nabla v|^2 \, d\gamma} \cdot \sqrt{d} < +\infty, \tag{4.6}$$

which proves that $\varrho^\psi < +\infty$. □

Note that in Lemma 4.3 we started with a *finite-valued* convex function $v \colon \mathbb{R}^d \to \mathbb{R}$ and then *defined* $\psi := v^*$. If we additionally *know* that the probability measure $\hat{p} = (\nabla v)(\gamma)$ has finite second moment, then $\hat{p}$ is the unique maximizer of the optimization problem (4.5). These are rather strong assumptions. However, if we are given just a proper convex function $\psi \colon \mathbb{R}^d \to (-\infty, +\infty]$, we can still compute the value of the supremum in (4.5), without explicitly constructing a maximizer of this optimization problem.

Lemma 4.4. *Let $\psi \colon \mathbb{R}^d \to (-\infty, +\infty]$ be a proper convex function. Then*

$$\varrho^\psi = \int \psi^* \, d\gamma. \tag{4.7}$$

We postpone the proof of Lemma 4.4 to Appendix B. As already announced, with the help of (4.7), we are now able to prove Propositions 4.1, 4.2. We first show a simpler variant, where we optimize over the class of convex functions ψ in $C_{\mathrm{q}}(\mathbb{R}^d)$ (i.e., which have quadratic growth), as in (4.8) of Proposition 4.5 below. Proposition 4.2, where we optimize over *all* μ-a.s. finite-valued convex functions $\psi \colon \mathbb{R}^d \to (-\infty, +\infty]$ as in (4.2), is then a straightforward consequence. Proposition 4.1, where we optimize over all ν-integrable convex functions ψ as in (4.1), follows from a "sandwich argument".

Proposition 4.5. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. Then $\tilde{D}(\mu,\nu)$ is equal to*

$$D_{\mathrm{q}}(\mu,\nu) := \inf_{\substack{\psi \in C_{\mathrm{q}}(\mathbb{R}^d), \\ \psi \text{ convex}}} \Big(\int \psi \, d\nu - \int (\psi^* * \gamma)^* \, d\mu \Big). \tag{4.8}$$

Proof. By Proposition 3.5 we have

$$\tilde{D}(\mu,\nu) = \inf_{\substack{\psi \in C_{\mathrm{q}}(\mathbb{R}^d), \\ \psi \text{ convex}}} \Big(\int \psi \, d\nu - \int \varphi^\psi \, d\mu \Big).$$

Hence it remains to show that $\varphi^\psi = (\psi^* * \gamma)^*$, for every convex function $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$. In order to do this, we will first prove that $(\varphi^\psi)^* = \psi^* * \gamma$. By definition of the convex conjugate and (4.4), for $\zeta \in \mathbb{R}^d$, we have

$$(\varphi^\psi)^*(\zeta) = \sup_{x\in\mathbb{R}^d} \big(\langle x,\zeta\rangle - \varphi^\psi(x)\big) \tag{4.9}$$

$$= \sup_{x\in\mathbb{R}^d} \sup_{p\in\mathcal{P}_2^x(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,\gamma) - \int \psi_\zeta \, dp\Big) \tag{4.10}$$

$$= \sup_{p\in\mathcal{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,\gamma) - \int \psi_\zeta \, dp\Big) = \varrho^{\psi_\zeta}, \tag{4.11}$$

where the function ψ_ζ is defined by $\psi_\zeta(y) \coloneqq \psi(y) - \langle \zeta, y\rangle$, for $y \in \mathbb{R}^d$. Now applying Lemma 4.4 to the proper convex function ψ_ζ yields

$$(\varphi^\psi)^*(\zeta) = \varrho^{\psi_\zeta} = \int \psi_\zeta^* \, d\gamma = \int \psi^*(\zeta + z)\, d\gamma(z) = (\psi^* * \gamma)(\zeta). \tag{4.12}$$

To complete the proof, we must justify that $(\varphi^\psi)^{**} = \varphi^\psi$, i.e. that the Fenchel–Moreau theorem is applicable. To see this, we first note that the function $x \mapsto \varphi^\psi(x)$ is convex (recall Lemma 3.2) and we have the upper bound $\varphi^\psi \leqslant \psi < +\infty$. Furthermore, for any $p \in \mathcal{P}_2^x(\mathbb{R}^d)$ we have the inequalities

$$\mathrm{MCov}(p,\gamma) \leqslant \tfrac{1}{2}\int |y|^2\, dp(y) + \tfrac{d}{2}$$

as well as

$$\int \psi \, dp \geqslant \ell + \tfrac{1}{2}\int |y|^2\, dp(y),$$

the latter following from the fact that $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$. As a consequence, we get the lower bound $\varphi^\psi \geqslant \ell - \frac{d}{2} > -\infty$. Altogether, φ^ψ is a convex function which is finite everywhere on $\mathbb{R}^d$, thus it is continuous and we indeed have $(\varphi^\psi)^{**} = \varphi^\psi$. □

Proof of Proposition 4.2. We first show that the function $\mathcal{E}(\,\cdot\,)$ in (4.3) is well defined for every convex function $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$, which is μ-a.s. finite. To this end, we will prove the inequality $(\psi^* * \gamma)^* \leqslant \psi$. By Jensen's inequality, this implies that the integrand in (4.3) is μ-a.s. non-negative, and hence $\mathcal{E}(\psi)$ is well defined and $[0, +\infty]$-valued. Recalling the equation (4.12) above, we have $\psi^* * \gamma = (\varphi^\psi)^*$. Taking the convex conjugate and using that $\varphi^\psi \leqslant \psi$, we obtain $(\psi^* * \gamma)^* = (\varphi^\psi)^{**} \leqslant \varphi^\psi \leqslant \psi$, as required.

Now let us turn to the proof of $\tilde{D}(\mu,\nu) = D_{\mathrm{rel}}(\mu,\nu)$. Recalling (4.2), (4.8), and using Proposition 4.5, we have $D_{\mathrm{rel}}(\mu,\nu) \leqslant D_{\mathrm{q}}(\mu,\nu) = \tilde{D}(\mu,\nu)$, so that we need to show the inequality $\tilde{D}(\mu,\nu) \leqslant D_{\mathrm{rel}}(\mu,\nu)$. Let $\pi \in \mathsf{MT}(\mu,\nu)$ and $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ be convex with $\mu(\mathrm{dom}\,\psi) = 1$. Since $\tilde{D}(\mu,\nu) = P(\mu,\nu)$ (recall Theorem 3.3 and (3.3)), we have to verify the inequality

$$\int \mathrm{MCov}(\pi_x^{\mathrm{SBM}},\gamma)\, d\mu(x) \leqslant \int \Big(\int \psi(y)\, \pi_x^{\mathrm{SBM}}(dy) - (\psi^* * \gamma)^*(x)\Big)\, \mu(dx).$$

It is sufficient to prove, for μ-a.e. $x \in \mathbb{R}^d$, that

$$\mathrm{MCov}(\pi_x^{\mathrm{SBM}},\gamma) \leqslant \int \psi \, d\pi_x^{\mathrm{SBM}} - (\psi^* * \gamma)^*(x). \tag{4.13}$$

We express the convex conjugate on the right-hand side of (4.13) as

$$-(\psi^* * \gamma)^*(x) = \inf_{\zeta\in\mathbb{R}^d}\big((\psi^* * \gamma)(\zeta) - \langle\zeta, x\rangle\big) = \inf_{\zeta\in\mathbb{R}^d}\Big(\int \psi_\zeta^*\, d\gamma - \langle\zeta, x\rangle\Big),$$

where the function ψ_ζ is defined by $\psi_\zeta(y) \coloneqq \psi(y) - \langle\zeta, y\rangle$, for $y \in \mathbb{R}^d$. Substituting back into (4.13) yields

$$\mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) \leqslant \inf_{\zeta\in\mathbb{R}^d}\Big(\int \psi_\zeta\, d\pi_x^{\mathrm{SBM}} + \int \psi_\zeta^*\, d\gamma\Big). \tag{4.14}$$

Now observe that by the Fenchel–Young inequality we have

$$\mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) \leqslant \int \phi\, d\pi_x^{\mathrm{SBM}} + \int \phi^*\, d\gamma,$$

for every proper convex function $\phi\colon \mathbb{R}^d \to (-\infty, +\infty]$. In particular, for every $\zeta \in \mathbb{R}^d$, it holds that

$$\mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) \leqslant \int \psi_\zeta\, d\pi_x^{\mathrm{SBM}} + \int \psi_\zeta^*\, d\gamma,$$

which implies (4.14). This completes the proof of Proposition 4.2. □

Proof of Proposition 4.1. The assertion follows immediately from Proposition 4.2 and Proposition 4.5 by a "sandwich argument". Indeed, we have

$$\tilde{D}(\mu,\nu) = D_{\mathrm{rel}}(\mu,\nu) \leqslant D(\mu,\nu) \leqslant D_{\mathrm{q}}(\mu,\nu) = \tilde{D}(\mu,\nu). \tag{4.15}$$

The inequalities in (4.15) are due to the inclusions

$$\{\psi \in C_{\mathrm{q}}(\mathbb{R}^d)\colon \psi \text{ convex}\} \subseteq \{\psi \in L^1(\nu)\colon \psi \text{ convex}\} \subseteq \{\psi \text{ convex}\colon \mu(\mathrm{dom}\,\psi) = 1\};$$

the equalities on the left-hand side and on the right-hand side of (4.15) are justified by Proposition 4.2 and Proposition 4.5, respectively. □

5. Preparation for the proof of the second part of Theorem 1.4. The goal of this rather technical section is to outline and prepare the proof of the second part of Theorem 1.4, for which we will need the results of Sections 6 and 7. In Section 7 we will show that the value $D(\mu,\nu)$ is attained by a convex function ψ_{opt} with $\mu(\mathrm{ri}(\mathrm{dom}\,\psi_{\mathrm{opt}})) = 1$ if and only if (μ,ν) is irreducible. In Section 6 (see Theorem 6.6) we will prove that there is such a dual optimizer ψ_{opt} if and only if there exists a Bass martingale from μ to ν. At a first reading one might skip the present Section 5.

We recall the primal problem (1.5), with optimizer $\pi^{\mathrm{SBM}} \in \mathsf{MT}(\mu,\nu)$, i.e.

$$P(\mu,\nu) = \int \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma)\, \mu(dx); \tag{5.1}$$

the dual problem (3.1), with optimizer ψ_{opt} (supposing that this optimizer exists), i.e.

$$\tilde{D}(\mu,\nu) = \Big(\int \psi_{\mathrm{opt}}\, d\nu - \int \varphi^{\psi_{\mathrm{opt}}}\, d\mu\Big); \tag{5.2}$$

and the fact that there is no duality gap by Theorem 3.3. Therefore, comparing the primal value (5.1) with the dual value (5.2), we see that the minimization problem

$$\varphi^{\psi_{\mathrm{opt}}}(x) = \inf_{p\in\mathcal{P}_2^x(\mathbb{R}^d)}\Big(\int \psi_{\mathrm{opt}}\, dp - \mathrm{MCov}(p,\gamma)\Big)$$

(recall (3.2)) is attained by $\pi_x^{\mathrm{SBM}} \in \mathcal{P}_2^x(\mathbb{R}^d)$, for μ-a.e. $x \in \mathbb{R}^d$. To draw the connection to the Bass martingale, we need to express the optimizer π_x^{SBM} in terms of the (single) convex function ψ_{opt}. More precisely, we will see as a consequence of Proposition 5.3 that $\pi_x^{\mathrm{SBM}} = (\nabla\psi_{\mathrm{opt}}^*)(\gamma_\zeta)$, for some appropriate $\zeta = \zeta(x) \in \mathbb{R}^d$. For this purpose, we have to return to the optimization problem (3.2) and study its optimizers.

First, we need the following technical result, formulated in Lemma 5.1 below. Let $\psi : \mathbb{R}^d \to (-\infty, +\infty]$ be a lower semicontinuous convex function. We recall from convex analysis that ψ is called co-finite if

$$\forall x \in \mathbb{R}^d \setminus \{0\}: \qquad \lim_{t\to+\infty} \frac{\psi(tx)}{t} = +\infty.$$

According to [54, Corollary 13.3.1], the convex function ψ is co-finite if and only if its convex conjugate ψ^* is finite everywhere on $\mathbb{R}^d$. We derive now a sufficient condition for the co-finiteness of ψ, in terms of the function φ^ψ defined in (3.2). We defer the proof of Lemma 5.1 to Appendix B.

Lemma 5.1. *Let $\psi : \mathbb{R}^d \to (-\infty, +\infty]$ be a proper, lower semicontinuous convex function. Then φ^ψ is convex on $\operatorname{dom}\psi$. If we additionally assume that $\varphi^\psi(x) > -\infty$, for some $x \in \operatorname{int}(\operatorname{dom}\psi)$, then $\varphi^\psi > -\infty$ on $\operatorname{int}(\operatorname{dom}\psi)$ and ψ is co-finite.*

Our next result is the analogue of Lemma 4.4. But now, instead of maximizing over all $p \in \mathcal{P}_2(\mathbb{R}^d)$ as in (4.5), we have to maximize over all $p \in \mathcal{P}_2^x(\mathbb{R}^d)$, with a fixed barycenter $x \in \mathbb{R}^d$, as in (4.4).

Proposition 5.2. *Let $\psi : \mathbb{R}^d \to (-\infty, +\infty]$ be a lower semicontinuous convex function and assume that $\varphi^\psi(x) > -\infty$ for some $x \in \operatorname{int}(\operatorname{dom}\psi)$. Then we have the duality formula*

$$\varphi^\psi(x) = \sup_{\zeta\in\mathbb{R}^d} \Big(\langle \zeta, x\rangle - \int \psi^*(\zeta + z)\, d\gamma(z) \Big), \tag{5.3}$$

and the right-hand side admits a unique maximizer $\zeta(x) \in \mathbb{R}^d$.

Proof. We deploy a similar strategy as in the proof of Proposition 4.5. Recalling equations (4.9) – (4.12), for each $\zeta \in \mathbb{R}^d$, we have

$$(\varphi^\psi)^*(\zeta) = \sup_{p\in\mathcal{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,\gamma) - \int \psi_\zeta\, dp \Big) = \int \psi^*(\zeta + z)\, d\gamma(z), \tag{5.4}$$

where the function ψ_ζ is defined by $\psi_\zeta(y) := \psi(y) - \langle \zeta, y\rangle$, for $y \in \mathbb{R}^d$. Since $\varphi^\psi(x) > -\infty$ for some $x \in \operatorname{int}(\operatorname{dom}\psi)$, the convex function $\xi \mapsto \varphi^\psi(\xi)$ is finite and thus also continuous in a neighbourhood of x. Hence we can apply the variant [8, Proposition 13.44] of the Fenchel–Moreau theorem and obtain

$$\varphi^\psi(x) = (\varphi^\psi)^{**}(x) = \sup_{\zeta\in\mathbb{R}^d} \Big(\langle \zeta, x\rangle - \int \psi^*(\zeta + z)\, d\gamma(z) \Big), \tag{5.5}$$

which proves the duality formula (5.3).

To prove the existence of a maximizer $\zeta(x) \in \mathbb{R}^d$, we define the function

$$f_x(\zeta) := \sup_{p\in\mathcal{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,\gamma) - \int \psi\, dp + \int \langle \zeta, y - x\rangle\, dp(y) \Big), \tag{5.6}$$

so that $-\varphi^{\psi}(x) = \inf_{\zeta\in\mathbb{R}^d} f_x(\zeta)$. Indeed, it follows from (5.4), (5.5) that

$$\begin{aligned} -\varphi^{\psi}(x) &= \inf_{\zeta\in\mathbb{R}^d} \Big(\int \psi^*(\zeta+z)\,d\gamma(z) - \langle \zeta, x\rangle \Big) \\ &= \inf_{\zeta\in\mathbb{R}^d} \sup_{p\in\mathcal{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,\gamma) - \int \psi_\zeta\,dp - \langle \zeta, x\rangle \Big). \end{aligned}$$

By assumption, the function f_x takes values in $(-\infty,+\infty]$ and is not constant equal to $+\infty$. Equivalently, we can express $f_x(\zeta)$ as

$$f_x(\zeta) = \int \big(\psi^*(\zeta+z) + \psi(x) - \langle \zeta+z, x\rangle\big)\,d\gamma(z) - \psi(x). \tag{5.7}$$

We will show that $\zeta \mapsto f_x(\zeta)$ is lower semicontinuous and coercive, implying the existence of an optimizer $\zeta(x)$. It is easy to see from the representation (5.7) that f_x is lower semicontinuous. Indeed, this directly follows from the lower semicontinuity of the non-negative integrand in (5.7) and Fatou's lemma. For the verification of the coercivity, we take a sequence $(\zeta^{(n)})_{n\geqslant 1}$ in $\mathbb{R}^d$ with $|\zeta^{(n)}| \to +\infty$. Then there is a coordinate $k \in \{1,\dots,d\}$ such that $|\zeta_k^{(n)}| \to +\infty$. Since $x \in \operatorname{int}(\operatorname{dom}\psi)$, we can choose $\varepsilon > 0$ small enough such that $x \pm \varepsilon e_k \in \operatorname{dom}\psi$, where e_k denotes the k-th standard basis vector of $\mathbb{R}^d$. Defining

$$y^{(n)} := x + \operatorname{sign}(\zeta_k^{(n)})\,\varepsilon\, e_k \in \operatorname{dom}\psi$$

and taking $p = \delta_{y^{(n)}}$, we conclude from (5.6) that

$$f_x(\zeta^{(n)}) \geqslant -\psi(y^{(n)}) + \varepsilon\,|\zeta_k^{(n)}| \to +\infty,$$

which shows $\lim_{|\zeta|\to+\infty} f_x(\zeta) = +\infty$.

We now show that the maximizer $\zeta(x) \in \mathbb{R}^d$ of the right-hand side of (5.3) is unique. As the assumptions of Lemma 5.1 are satisfied, the convex function ψ is co-finite, and thus ψ^* is finite everywhere on $\mathbb{R}^d$. In particular, ψ^* is continuous everywhere and differentiable Lebesgue-a.e. on $\mathbb{R}^d$. In order to establish the uniqueness of $\zeta(x)$, we show that the function $\zeta \mapsto \int \psi^*(\zeta+z)\,d\gamma(z) \in (-\infty,+\infty]$ is strictly convex on its domain. By contradiction, suppose that there are $\zeta_1,\zeta_2 \in \mathbb{R}^d$ with $\zeta_1 \neq \zeta_2$ such that both $\int \psi^*(\zeta_1+z)\,d\gamma(z)$ and $\int \psi^*(\zeta_2+z)\,d\gamma(z)$ are finite and

$$\int \psi^*\big(t\zeta_1 + (1-t)\zeta_2 + z\big)\,d\gamma(z) = t\int \psi^*(\zeta_1+z)\,d\gamma(z) + (1-t)\int \psi^*(\zeta_2+z)\,d\gamma(z),$$

for some $t \in (0,1)$. Then

$$\psi^*\big(t\zeta_1 + (1-t)\zeta_2 + z\big) = t\psi^*(\zeta_1+z) + (1-t)\psi^*(\zeta_2+z),$$

for γ-a.e. $z \in \mathbb{R}^d$. Consequently, the function ψ^* is affine on the line segment from ζ_1+z to ζ_2+z, for Lebesgue-a.e. $z \in \mathbb{R}^d$. By convexity and finiteness of ψ^* on $\mathbb{R}^d$, we thus have that the function

$$\mathbb{R} \ni \lambda \longmapsto \psi^*\big(\lambda(\zeta_2-\zeta_1)+z\big)$$

is affine, for every $z \in \mathbb{R}^d$. By [54, Theorem 8.8], there exists some $c \in \mathbb{R}^d$ such that

$$\psi^*\big(\lambda(\zeta_2-\zeta_1)+z\big) = \psi^*(z) + \lambda c.$$

Differentiating at the point $\lambda = 1$ yields $\langle \nabla\psi^*(z), \zeta_2-\zeta_1\rangle = c$, for Lebesgue-a.e. $z \in \mathbb{R}^d$. But this is a contradiction to $\varnothing \neq \operatorname{int}(\operatorname{dom}\psi) \subseteq (\partial\psi^*)(\mathbb{R}^d)$, where the symbol ∂, applied to a convex function, denotes its subdifferential. □

The final result of this section is the analogue of Lemma 4.3. The duality formula (5.3) established in Proposition 5.2 enables us to identify the structure of the optimizer in the maximization problem (4.4).

PROPOSITION 5.3. *Let $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ be a lower semicontinuous convex function. For fixed $\zeta \in \mathbb{R}^d$ we define the function ψ_ζ by $\psi_\zeta(y) := \psi(y) - \langle \zeta, y\rangle$, for $y \in \mathbb{R}^d$, so that $\psi^*_\zeta(y) = \psi^*(\zeta + y)$.*

(i) *If there are $x \in \mathbb{R}^d$ and $\zeta = \zeta(x) \in \mathbb{R}^d$ such that $(\nabla\psi^*_\zeta)(\gamma) \in \mathcal{P}_2^x(\mathbb{R}^d)$, then this is actually the unique optimizer of the supremum in (4.4). If additionally $\varphi^\psi(x) > -\infty$ and $x \in \operatorname{int}(\operatorname{dom}\psi)$, then $\zeta = \zeta(x)$ is the optimizer of the supremum in (5.3).*

(ii) *If $\varphi^\psi(x) > -\infty$ for some $x \in \operatorname{int}(\operatorname{dom}\psi)$ and the supremum in (4.4) is attained by some $\hat{p}_x \in \mathcal{P}_2^x(\mathbb{R}^d)$, then $\hat{p}_x = (\nabla\psi^*_\zeta)(\gamma)$ with $\zeta = \zeta(x)$ being the optimizer of the supremum in (5.3).*

PROOF. (i) We set $\hat{p}_x := (\nabla\psi^*_\zeta)(\gamma)$ and let $p_x \in \mathcal{P}_2^x(\mathbb{R}^d)$ be arbitrary. We denote by $T_\gamma^{p_x}$ the Brenier map from γ to p_x and note that $T_\gamma^{\hat{p}_x} = \nabla\psi^*_\zeta$. Similar to the proof of Lemma 4.3, we compute

$$\begin{aligned}\int \psi\, dp_x - \operatorname{MCov}(p_x, \gamma) &= \int \Big(\psi_\zeta\big(T_\gamma^{p_x}(z)\big) + \big\langle \zeta, T_\gamma^{p_x}(z)\big\rangle - \big\langle T_\gamma^{p_x}(z), z\big\rangle\Big)\, d\gamma(z)\\ &\geqslant \langle \zeta, x\rangle + \int \inf_{y\in\mathbb{R}^d} \big(\psi_\zeta(y) - \langle y, z\rangle\big)\, d\gamma(z)\\ &= \langle \zeta, x\rangle - \int \psi^*(\zeta + z)\, d\gamma(z)\\ &= \int \psi\, d\hat{p}_x - \operatorname{MCov}(\hat{p}_x, \gamma),\end{aligned}$$

with equality if and only if $T_\gamma^{p_x}(z) = T_\gamma^{\hat{p}_x}(z)$, for γ-a.e. $z \in \mathbb{R}^d$. This in turn is the case if and only if $p_x = \hat{p}_x$. We conclude that $\hat{p}_x$ is the unique optimizer of the supremum in (4.4), i.e.

$$\varphi^\psi(x) = \int \psi\, d\hat{p}_x - \operatorname{MCov}(\hat{p}_x, \gamma) = \langle \zeta, x\rangle - \int \psi^*(\zeta + z)\, d\gamma(z). \tag{5.8}$$

On the other hand, if additionally $\varphi^\psi(x) > -\infty$ and $x \in \operatorname{int}(\operatorname{dom}\psi)$, by Proposition 5.2 we have that

$$\varphi^\psi(x) = \sup_{\zeta\in\mathbb{R}^d} \Big(\langle \zeta, x\rangle - \int \psi^*(\zeta + z)\, d\gamma(z)\Big),$$

which in light of (5.8) implies that $\zeta = \zeta(x)$ is the optimizer of the supremum in (5.3).

(ii) By assumption and Proposition 5.2, we have that

$$\begin{aligned}\varphi^\psi(x) &= \langle \zeta, x\rangle + \int \Big(\psi_\zeta\big(T_\gamma^{\hat{p}_x}(z)\big) - \big\langle T_\gamma^{\hat{p}_x}(z), z\big\rangle\Big)\, d\gamma(z)\\ &= \langle \zeta, x\rangle - \int \psi^*_\zeta(z)\, d\gamma(z)\end{aligned}$$

and therefore

$$\int \Big(\psi_\zeta\big(T_\gamma^{\hat{p}_x}(z)\big) + \psi^*_\zeta(z) - \big\langle T_\gamma^{\hat{p}_x}(z), z\big\rangle\Big)\, d\gamma(z) = 0.$$

As the integrand is pointwise non-negative, we conclude that

$$\psi\big(T_\gamma^{\hat{p}_x}(z)\big) + \psi^*(\zeta + z) = \big\langle T_\gamma^{\hat{p}_x}(z), \zeta + z \big\rangle$$

and consequently $T_\gamma^{\hat{p}_x}(z) \in \partial\psi^*(\zeta + z)$, for γ-a.e. $z \in \mathbb{R}^d$. Hence $T_\gamma^{\hat{p}_x}(z) = \nabla\psi^*(\zeta + z)$, for γ-a.e. $z \in \mathbb{R}^d$, which implies that $\hat{p}_x = (\nabla\psi^*_\zeta)(\gamma)$. □

6. Existence of dual optimizers and Bass martingales. Throughout this section we fix $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$.

Definition 6.1. A lower semicontinuous convex function $\psi_{\text{opt}} \colon \mathbb{R}^d \to (-\infty, +\infty]$ satisfying $\mu(\text{ri}(\text{dom}\,\psi_{\text{opt}})) = 1$ is called an optimizer of the dual problem (3.14) (in short, a dual optimizer) if $\tilde{D}(\mu, \nu) = \mathcal{D}(\psi_{\text{opt}})$, for the dual function $\mathcal{D}(\,\cdot\,)$ as defined in (3.12).

Let us recall that by Assumption 3.1 the support of ν affinely spans $\mathbb{R}^d$. As the next result shows, we could have required in the definition of a dual optimizer ψ_{opt} that also the domain of ψ_{opt} affinely spans $\mathbb{R}^d$; we defer the proof to Appendix B.

Lemma 6.2. *If ψ_{opt} is an optimizer according to Definition 6.1, then the domain of ψ_{opt} has non-empty interior. In particular, by Assumption 3.1, we have that*

$$\dim(\text{dom}\,\psi_{\text{opt}}) = \dim(\text{supp}\,\nu) = d.$$

We recall Definition 1.1 and list some important properties of Bass martingales. Such martingales were introduced in [5] under the name of a "standard stretched Brownian motion". In this paper, we use — with Richard Bass' permission — the term "Bass martingale" instead.

Remark 6.3. Let $M = (M_t)_{0 \leqslant t \leqslant 1}$ be a Bass martingale from μ to ν, with corresponding convex function $v \colon \mathbb{R}^d \to \mathbb{R}$ and initial distribution $\mathcal{P}(\mathbb{R}^d) \ni \alpha \sim B_0$ of the underlying Brownian motion $B = (B_t)_{0 \leqslant t \leqslant 1}$.

(i) As shown in [5] (and as a result of Theorem 6.6 below), the martingale transport

$$\text{Law}(M_0, M_1) \in \mathsf{MT}(\mu, \nu)$$

is equal to the unique optimizer π^{SBM} of (1.5). Furthermore, the knowledge of π^{SBM} already determines the martingale $(M_t)_{0 \leqslant t \leqslant 1}$ as well as the function v, which is $(\alpha * \gamma)$-a.e. (equivalently, Lebesgue-a.e.) unique up to an additive constant.

(ii) The convex function v and the probability measure α satisfy the identities

$$(\nabla v * \gamma)(\alpha) = \mu \qquad \text{and} \qquad \nabla v(\alpha * \gamma) = \nu, \tag{6.1}$$

which we summarize in the following graphic:

$$\begin{array}{ccc} \alpha * \gamma & \xrightarrow{\nabla v} & \nu \\ {\scriptstyle *}\uparrow & & \\ \alpha & \xrightarrow{\nabla v * \gamma} & \mu \end{array}$$

(iii) We also remark (see [5]) that we have

$$M_t = (\nabla v * \gamma^{1-t})(B_t) \quad \text{and} \quad \pi^{\text{SBM}} = \text{Law}\big((\nabla v * \gamma)(B_0), \nabla v(B_1)\big). \tag{6.2}$$

The next result, Theorem 6.6 below, explains how the existence of a dual optimizer ψ_{opt} is related to the existence of a Bass martingale. Along with it we need the following technical lemmas, whose proofs we postpone to Appendix B.

LEMMA 6.4. *Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a finite convex function such that $\nabla f \in L^2(\gamma_\zeta;\mathbb{R}^d)$, for some $\zeta \in \mathbb{R}^d$. Then*

(i) $\nabla(f * \gamma) = (\nabla f) * \gamma$.

If additionally $\operatorname{int}(\operatorname{dom} f^*) \neq \emptyset$, *then*

(ii) $(f * \gamma)^*$ *is differentiable and strictly convex on* $\operatorname{int}(\operatorname{dom}(f * \gamma)^*) \supseteq \operatorname{int}(\operatorname{dom} f^*)$,
(iii) $\nabla(f*\gamma)^*\colon \operatorname{int}(\operatorname{dom}(f*\gamma)^*) \to \mathbb{R}^d$ *and* $\nabla(f*\gamma)\colon \mathbb{R}^d \to \operatorname{int}(\operatorname{dom}(f*\gamma)^*)$ *are bijections, and we have*

$$(\nabla f * \gamma)^{-1} = \nabla(f * \gamma)^*.$$

LEMMA 6.5. *Let $\psi\colon \mathbb{R}^d \to (-\infty,+\infty]$ be a lower semicontinuous convex function and assume that $\varphi^\psi(x) > -\infty$, for some $x \in \operatorname{int}(\operatorname{dom}\psi)$. Furthermore, we suppose that $\nabla\psi^* \in L^2(\gamma_\zeta;\mathbb{R}^d)$, for some $\zeta \in \mathbb{R}^d$. Then, for all $\eta \in \mathbb{R}^d$ and all $t > 0$, we have*

$$\operatorname{int}(\operatorname{dom}\psi) = \operatorname{int}\big((\partial\psi^*)(\mathbb{R}^d)\big) = (\nabla\psi^* * \gamma^t_\eta)(\mathbb{R}^d). \tag{6.3}$$

THEOREM 6.6. *There exists a dual optimizer ψ_{opt} in the sense of Definition 6.1 if and only if there exists a Bass martingale $(M_t)_{0\leqslant t\leqslant 1}$ from μ to ν. In this case,* $\operatorname{Law}(M_0, M_1)$ *is equal to the optimizer $\pi^{\mathrm{SBM}} \in \mathsf{MT}(\mu,\nu)$ of the primal problem (1.5), and $\psi_{\mathrm{opt}}, v, \alpha, \pi^{\mathrm{SBM}}$ are related via $v^* = \psi_{\mathrm{opt}}$, $\alpha = \zeta(\mu)$ and*

$$\pi^{\mathrm{SBM}}_x = \operatorname{Law}(M_1 \mid M_0 = x) = \nabla v\big(\zeta(x) + \cdot\,\big)(\gamma), \tag{6.4}$$

where the function $\zeta\colon X \to \mathbb{R}^d$ is given by

$$\zeta(x) := \operatorname{argmax}_{\zeta\in\mathbb{R}^d}\left(\langle\zeta, x\rangle - \int v(\zeta + z)\,d\gamma(z)\right) \tag{6.5}$$

as in (5.3), and we have

$$\zeta(x) = (\nabla v * \gamma)^{-1}(x) = \nabla(v * \gamma)^*(x), \tag{6.6}$$

for all $x \in X$. The domain X of the function ζ satisfies $X \subseteq \operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})$ and $\mu(X) = 1$.

PROOF. "⇒": Suppose that a dual optimizer ψ_{opt} exists. Then $\tilde{D}(\mu,\nu) - \mathcal{D}(\psi_{\mathrm{opt}}) = 0$ and from (3.3), (3.12) we obtain

$$\int\left(\varphi^{\psi_{\mathrm{opt}}}(x) - \left(\int \psi_{\mathrm{opt}}(y)\,\pi^{\mathrm{SBM}}_x(dy) - \operatorname{MCov}(\pi^{\mathrm{SBM}}_x,\gamma)\right)\right)\mu(dx) = 0.$$

Hence, for μ-a.e. $x \in \mathbb{R}^d$, the infimum in

$$\varphi^{\psi_{\mathrm{opt}}}(x) = \inf_{p\in\mathcal{P}_2^x(\mathbb{R}^d)}\left(\int \psi_{\mathrm{opt}}\,dp - \operatorname{MCov}(p,\gamma)\right)$$

must be attained by $\pi^{\mathrm{SBM}}_x \in \mathcal{P}_2^x(\mathbb{R}^d)$. Furthermore, as $\mathcal{D}(\psi_{\mathrm{opt}})$ is finite, we deduce as in the proof of Lemma 3.7 that $\int \psi_{\mathrm{opt}}\,d\pi^{\mathrm{SBM}}_x < +\infty$, for μ-a.e. $x \in \mathbb{R}^d$. Thus also $\varphi^{\psi_{\mathrm{opt}}}(x)$ is finite, for μ-a.e. $x \in \mathbb{R}^d$. Recalling Lemma 6.2, we have $\mu(\operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})) = 1$ by Definition 6.1. In particular, ψ_{opt} is co-finite by Lemma 5.1. All in all, we can apply part (ii) of Proposition 5.3 to ψ_{opt}, for μ-a.e. $x \in \mathbb{R}^d$. This yields that $\pi^{\mathrm{SBM}}_x = \nabla v(\zeta(x) + \cdot\,)(\gamma)$, with $v := \psi^*_{\mathrm{opt}}$ and $\zeta(x) \in \mathbb{R}^d$ as in (6.5). We now define $\alpha := \zeta(\mu)$. Since π^{SBM} has second marginal equal to ν, we obtain $\nabla v(\alpha * \gamma) = \nu$. On the other hand, as π^{SBM}_x has barycenter x, we have

$$\int \nabla v\big(\zeta(x) + z\big)\,d\gamma(z) = (\nabla v * \gamma)\big(\zeta(x)\big) = x. \tag{6.7}$$

We denote the set of points $x \in \mathbb{R}^d$ for which the identity (6.7) holds by X and note that $X \subseteq \operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})$ as well as $\mu(X) = 1$. Since $\alpha = \zeta(\mu)$, we conclude from (6.7) that $(\nabla v * \gamma)(\alpha) = \mu$. By (6.1), this establishes the existence of a Bass martingale $(M_t)_{0\leqslant t\leqslant 1}$ from μ to ν, which by construction satisfies $\operatorname{Law}(M_0, M_1) = \pi^{\mathrm{SBM}}$, and connects $\psi_{\mathrm{opt}}, v, \alpha, \pi^{\mathrm{SBM}}$ as claimed.

It remains to check the identity (6.6). Since $\pi_x^{\mathrm{SBM}} \in \mathcal{P}_2^x(\mathbb{R}^d)$, and as we already established (6.4), we conclude that

$$\int \big|\nabla v\big(\zeta(x) + z\big)\big|^2 \, d\gamma(z) < +\infty, \tag{6.8}$$

for every $x \in X$. In particular, we can apply part (iii) of Lemma 6.4 to the function $f = v$. Then from (6.7) we derive (6.6).

"$\Leftarrow$": Conversely, suppose that a Bass martingale $M = (M_t)_{0\leqslant t\leqslant 1}$ from μ to ν exists. We denote by $v\colon \mathbb{R}^d \to \mathbb{R}$ the associated convex function and define the proper, lower semicontinuous convex function $\psi_{\mathrm{opt}} \coloneqq v^*$. Recall that $\dim(\operatorname{supp}\nu) = d$ by Assumption 3.1. Since $(\partial v)(\mathbb{R}^d) \subseteq \operatorname{dom} v^*$ and

$$\operatorname{int}\operatorname{conv}\big((\partial v)(\mathbb{R}^d)\big) = \operatorname{int}\overline{\operatorname{conv}}\big((\partial v)(\mathbb{R}^d)\big) = \operatorname{int}\widehat{\operatorname{supp}}(\nu) \neq \varnothing,$$

we also have that $\operatorname{int}(\operatorname{dom} v^*) \neq \varnothing$. In the following we show that ψ_{opt} is indeed a dual optimizer in the sense of Definition 6.1.

First, we verify that $\varphi^{\psi_{\mathrm{opt}}}(x) > -\infty$, for all $x \in \mathbb{R}^d$. As a consequence, we will see that the identity (6.6) holds for all $x \in \operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})$. Note that from (6.1) we have the relation $\nabla v(\alpha * \gamma) = \nu$. Since ν has finite second moment, we conclude that

$$\int |\nabla v(\zeta + z)|^2 \, d\gamma(z) < +\infty, \tag{6.9}$$

for α-a.e. $\zeta \in \mathbb{R}^d$. In particular, by analogy with (4.6), we obtain that

$$(v * \gamma)(\zeta) = \int v(\zeta + z) \, d\gamma(z) \in (-\infty, +\infty), \tag{6.10}$$

for α-a.e. $\zeta \in \mathbb{R}^d$. Recalling the second equality in (5.5), we have

$$\varphi^{\psi_{\mathrm{opt}}}(x) \geqslant (\varphi^{\psi_{\mathrm{opt}}})^{**}(x) = \sup_{\zeta\in\mathbb{R}^d} \Big(\langle \zeta, x\rangle - \int \psi_{\mathrm{opt}}^*(\zeta + z) \, d\gamma(z)\Big), \tag{6.11}$$

and by (6.10), the right-hand side of (6.11) is greater than $-\infty$, for every $x \in \mathbb{R}^d$. In particular, we can apply Proposition 5.2 to the function ψ_{opt}. As a result, for every $x \in \operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})$, we obtain an equality in (6.11) and the right-hand side admits a unique maximizer $\zeta(x) \in \mathbb{R}^d$. Differentiating under the integral sign, which is justified by part (i) of Lemma 6.4 applied to the function $f = v = \psi_{\mathrm{opt}}^*$, the first order condition for the optimality of $\zeta(x)$ reads

$$(\nabla\psi_{\mathrm{opt}}^* * \gamma)\big(\zeta(x)\big) = x. \tag{6.12}$$

Now we use part (iii) of Lemma 6.4 and obtain the identity (6.6) from (6.12) above, for all $x \in \operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})$.

Next, we show that $\alpha = \zeta(\mu)$ and $\mu(\operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})) = 1$. Recalling from (6.1) that $(\nabla v * \gamma)(\alpha) = \mu$, and using (6.6), we get $\alpha = (\nabla v * \gamma)^{-1}(\mu) = \zeta(\mu)$. For the second claim, we again use (6.6) and obtain

$$1 = \alpha(\mathbb{R}^d) = \mu\big((\nabla v * \gamma)(\mathbb{R}^d)\big) = \mu(\operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})),$$

where the last equality follows from Lemma 6.5.

It remains to check that $\tilde{D}(\mu,\nu) = \mathcal{D}(\psi_{\mathrm{opt}})$ and $\mathrm{Law}(M_0, M_1) = \pi^{\mathrm{SBM}}$. For each $x \in \mathrm{int}(\mathrm{dom}\,\psi_{\mathrm{opt}})$, we define a probability measure $\pi^v_x := \nabla v(\zeta(x) + \cdot)(\gamma)$. By (6.9) and (6.12), π^v_x is an element of $\mathcal{P}_2^x(\mathbb{R}^d)$, for μ-a.e. $x \in \mathbb{R}^d$. Hence we can apply part (i) of Proposition 5.3, showing that the infimum

$$\varphi^{\psi_{\mathrm{opt}}}(x) = \inf_{p \in \mathcal{P}_2^x(\mathbb{R}^d)} \Big(\int \psi_{\mathrm{opt}}\, dp - \mathrm{MCov}(p,\gamma) \Big)$$

is attained by π^v_x.

Set $\pi^v(dx,dy) = \pi^v_x(dy)\mu(dx)$. Since M is a Bass martingale with associated convex function v, we have $\pi^v_x = \mathrm{Law}(M_1 | M_0 = x)$, for μ-a.e. $x \in \mathbb{R}^d$. Moreover, by Theorem 1.10 of [5], M is stretched Brownian motion; hence, by uniqueness of the primal optimizer,

$$\pi^v = \mathrm{Law}(M_0, M_1) = \pi^{\mathrm{SBM}}.$$

Therefore, from the representation (3.12) of the dual function $\mathcal{D}(\,\cdot\,)$ and the identity $\pi^v = \pi^{\mathrm{SBM}}$, we obtain

$$\mathcal{D}(\psi_{\mathrm{opt}}) = \int \Big(\int \psi_{\mathrm{opt}}(y)\, \pi^v_x(dy) - \varphi^{\psi_{\mathrm{opt}}}(x) \Big) \mu(dx) = \int \mathrm{MCov}(\pi^v_x, \gamma)\, \mu(dx), \tag{6.13}$$

and this is equal to $\leqslant P(\mu,\nu) = \tilde{D}(\mu,\nu)$. This completes the proof that ψ_{opt} is a dual optimizer. The identity $\mathrm{Law}(M_0, M_1) = \pi^{\mathrm{SBM}}$ has already been obtained above, and the equality $\pi^v_x = \mathrm{Law}(M_1 | M_0 = x)$ gives (6.4). □

We now give a one-dimensional example in which a dual optimizer ψ_{opt} in the sense of Definition 6.1 is not integrable with respect to ν. Example 6.7 below illustrates that the form (3.5) of the dual function $\mathcal{D}(\,\cdot\,)$ may fail to make sense when we allow ψ to range more generally than in $C_{\mathrm{q}}(\mathbb{R}^d)$. On the other hand, the dual function written in the form (3.12) makes perfect sense for general convex functions $\psi : \mathbb{R}^d \to (-\infty, +\infty]$, which are μ-a.s. finite (recall Lemma 3.7). This leads to the satisfactory characterization of dual attainment as given by Theorem 6.6.

Example 6.7. Let $\alpha := \sum_{n=1}^{\infty} 2^{-n} \delta_{z_n} \in \mathcal{P}(\mathbb{R})$, with the sequence $(z_n)_{n \geqslant 1} \subseteq \mathbb{R}$ satisfying $\lim_{n\to\infty} \frac{|z_n|}{2^n} = +\infty$. Consider the convex function $v(z) := z \arctan z - \frac{1}{2}\log(1+z^2)$, with the derivative $v'(z) = \arctan z$ being a strictly increasing, continuous and bounded function. Define $\mu := (v' * \gamma)(\alpha)$, so that $\mu = \sum_{n=1}^{\infty} 2^{-n} \delta_{x_n}$ for some bounded sequence $(x_n)_{n \geqslant 1} \subseteq \mathbb{R}$, and let $\nu := v'(\alpha * \gamma)$. Note that μ and ν are supported by $(-\frac{\pi}{2}, \frac{\pi}{2})$. Then, according to (6.1), the pair (v, α) defines a Bass martingale from μ to ν. Theorem 6.6 shows that the derivative of the dual optimizer is given by

$$\tfrac{d}{dy}\psi_{\mathrm{opt}}(y) = (v')^{-1}(y) = \tan y.$$

By definition of ν and α, this function is not ν-integrable, and neither is its antiderivative ψ_{opt}. ◇

We finish this section by deducing a trajectorial property of Bass martingales from Lemma 6.5 and Theorem 6.6, namely that a Bass martingale M with $M_1 \sim \nu$ can only reach the boundary of $\widehat{\mathrm{supp}}(\nu)$ at time 1. In this paper we only make use of this result as a technical step in the proof of Lemma 7.17. However, this is a natural property to examine from the point of view of stochastic analysis (e.g. Feller's explosion test is devoted to a related question), and it gives us some intuition about the behavior of Bass martingales.

COROLLARY 6.8. *Let $M = (M_t)_{0\leqslant t\leqslant 1}$ be a Bass martingale with $M_1 \sim \nu$. Define τ as the stopping time*

$$\tau \coloneqq \inf\left\{t \in [0,1] \colon M_t \notin \operatorname{ri}\widehat{\operatorname{supp}}(\nu)\right\} \wedge 1,$$

i.e., the minimum of 1 *and the first time that M reaches the boundary of $\widehat{\operatorname{supp}}(\nu)$. Then $\tau = 1$ a.s.*

PROOF. We will prove that

$$\mathbb{P}\big(t \in [0,1) \colon M_t \in \operatorname{ri}\widehat{\operatorname{supp}}(\nu)\big) = 1.$$

Thus, if M ever reaches the boundary of $\widehat{\operatorname{supp}}(\nu)$, this will happen at time $t = 1$, so $\tau = 1$. Otherwise, i.e., if M never reaches the boundary of $\widehat{\operatorname{supp}}(\nu)$, then by definition $\tau = 1$, too.

Recall that $\dim(\operatorname{supp}\nu) = d$ by Assumption 3.1, so we can replace the relative interior operator by the interior operator throughout the argument. By (6.2), the Bass martingale is given by $M_t = (\nabla v * \gamma^{1-t})(B_t)$, where $\nabla v(\gamma) = \nu$, for some finite-valued convex function v. We now define the proper, lower semicontinuous convex function $\psi \coloneqq v^*$. As in the proof of the second implication of Theorem 6.6, namely the part of the proof corresponding to "$\Leftarrow$", we see that $\operatorname{int}(\operatorname{dom} v^*) \neq \varnothing$, that $\varphi^{\psi}(x) > -\infty$ for all $x \in \mathbb{R}^d$, and that $\nabla\psi^* \in L^2(\gamma_{B_0}; \mathbb{R}^d)$ a.s. (for the latter, see (6.9) and recall that $B_0 \sim \alpha$). Therefore we can apply Lemma 6.5 and from (6.3) we obtain for all $t \in [0,1)$ the equality

$$\operatorname{int}(\operatorname{dom}\psi) = \operatorname{int}\big((\partial\psi^*)(\mathbb{R}^d)\big) = (\nabla\psi^* * \gamma^{1-t})(\mathbb{R}^d).$$

As we justify in (B.11) in Appendix B, we automatically have the equality

$$\operatorname{int}\big((\partial\psi^*)(\mathbb{R}^d)\big) = \operatorname{int}\operatorname{conv}\big((\partial\psi^*)(\mathbb{R}^d)\big),$$

so we conclude that

$$\operatorname{int}\operatorname{conv}\big((\partial\psi^*)(\mathbb{R}^d)\big) = (\nabla\psi^* * \gamma^{1-t})(\mathbb{R}^d).$$

On the other hand, we clearly have

$$\operatorname{int}\operatorname{conv}\big((\partial\psi^*)(\mathbb{R}^d)\big) = \operatorname{int}\overline{\operatorname{conv}}\big((\partial\psi^*)(\mathbb{R}^d)\big) = \operatorname{int}\widehat{\operatorname{supp}}(\nu),$$

so we obtain for all $t \in [0,1)$ that

$$\operatorname{int}\widehat{\operatorname{supp}}(\nu) = (\nabla\psi^* * \gamma^{1-t})(\mathbb{R}^d).$$

Since $M_t = (\nabla\psi^* * \gamma^{1-t})(B_t)$, we conclude as desired that M_t lies in $\operatorname{int}\widehat{\operatorname{supp}}(\nu)$ as long as $t \in [0,1)$. □

7. Irreducibility and existence of dual optimizers.

7.1. *Outline and objectives.* In this technically demanding section we connect our findings with the groundbreaking analysis presented in [27] by H. De March and N. Touzi. Their work establishes the existence of a unique maximal paving of $\mathbb{R}^d$ into relatively open, convex, invariant sets. Each martingale transport from μ to ν maps these relatively open sets into their closures. Our focus here is on the irreducible case, where this paving is trivial, i.e., it consists of a single cell. The general case is treated in the follow-up paper [56].

We show that the triviality of the De March–Touzi (DMT) paving (as formalized in Definition 7.2 below) is equivalent to the irreducibility condition in Definition 1.2. This equivalence, along with additional equivalent characterizations of irreducibility, is established in Theorem D.1 of Appendix D, which builds on the results of the present section.

Assuming the triviality of the DMT paving in Assumption 7.3 below, we fix a martingale transport $\pi^{\mathrm{DMT}} \in \mathsf{MT}(\mu,\nu)$ such that every transition kernel π^{DMT}_x has full convex support. The existence of such a martingale transport π^{DMT} was established in [27]. In Lemma 7.5, we prove a uniformity property of this full-support feature, valid for μ-a.e. $x \in \mathbb{R}^d$.

The main result of this section is Theorem 7.6, which links the triviality of the DMT paving to the existence of an optimizer for the dual problem (3.14). As a direct consequence of this theorem, we obtain that triviality of the DMT paving implies the existence of a Bass martingale M from μ to ν with $\mathrm{Law}(M_0, M_1) = \pi^{\mathrm{SBM}}$. Notably, this Bass martingale serves as a concrete example of a De March–Touzi transport since, for μ-a.e. $x \in \mathbb{R}^d$, the measures π^{DMT}_x are equivalent to ν.

The first step in proving Theorem 7.6 is to identify a limit $\psi_{\lim}$ of an arbitrary optimizing sequence of convex functions $(\psi_n)_{n\geqslant 1}$ for the dual problem (3.1). A crucial observation is that the value of the dual function (3.12) remains invariant under the addition of affine functions to ψ. This invariance allows us to choose and add affine functions that facilitate our analysis. Concrete examples of such choices of affine functions are provided in [56] (Examples 6.1 and 6.2). After appropriately selecting representatives of ψ_n (by adding affine functions), we establish in Lemma 7.12 (by applying Komlós' theorem) the existence of a sequence of Cesàro means of a subsequence that is bounded on compact subsets of $I = \mathrm{ri}\,\widehat{\mathrm{supp}}(\nu)$. Using arguments from convex analysis, we extract a further subsequence that converges uniformly on compact subsets $K \subseteq I$ to a convex function $\psi_{\lim}$.

The second step in proving Theorem 7.6 is to verify that the limiting function $\psi_{\lim}$ is indeed a dual optimizer. To this end, we introduce in Definition 7.14 the Brenier map $\nabla v^x_{\mathrm{SBM}}$ from γ to π^{SBM}_x. The next goal (see Lemma 7.18) is to show that, for μ-a.e. $x \in I$, the function $\psi_{\lim}$ equals $(v^x_{\mathrm{SBM}})^*$, modulo the addition of an affine function. Specifically, we introduce (see (7.6)) the set

$$A = \big\{x \in I \colon \psi_{\lim} \not\equiv (v^x_{\mathrm{SBM}})^* \bmod (\mathrm{aff})\big\},$$

and prove that $\mu(A) = 0$. This proof proceeds by contradiction. Assuming $\mu(A) > 0$, we construct, for $x \in A$, measures $\check{\pi}_x \in \mathcal{P}^2_x(\mathbb{R}^d)$ in Lemma 7.16, which are adjusted in Lemma 7.17 to have compact support. The measures $\check{\pi}_x$ then contradict the optimality of π^{SBM}_x, allowing us to conclude $\mu(A) = 0$ and thereby complete the proof of Lemma 7.18. With the help of Lemma 7.19 we can then finish the second step and thus also the proof of Theorem 7.6. Broadly speaking, Lemma 7.19 shows that $\psi_{\lim}$ is a dual optimizer if and only if $\mu(A) = 0$. The proof of this result is based on Proposition 5.3 and Theorem 6.6.

At this point we encounter a remarkable and somewhat surprising feature. In Proposition 7.20 we use the wisdom of hindsight to show that in the arguments outlined above, there is no need of passing to a subsequence or of forming Cesàro means of the optimizing sequence $(\psi_n)_{n\geqslant 1}$. Rather already the original sequence $(\psi_n)_{n\geqslant 1}$ — modulo adding affine functions — converges to the optimizer $\psi_{\lim}$. Note that in Proposition 7.20 the pair (μ,ν) is not necessarily irreducible, rather the boundedness assumption (7.25) is made (cf. Lemma 7.12). This refinement plays a crucial role in the follow-up paper [56].

Towards the end of Section 7 we are finally in the position to prove Theorem 1.3 as well as the second part of Theorem 1.4, thereby completing the main body of the paper (apart from four appendices).

In the following subsections we now proceed with a formal and detailed treatment of the outlined program for Section 7.

7.2. *Irreducibility and dual attainment.* We fix $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$. The following notation from [27] will be used throughout this section.

Definition 7.1. We denote by $C := \widehat{\operatorname{supp}}(\nu)$ the closed convex hull of the support of ν and by $I := \operatorname{ri} C$ its relative interior.

Recall that $\dim(\operatorname{supp}\nu) = d$ by Assumption 3.1, so that I is open in $\mathbb{R}^d$.

Definition 7.2. We say that the pair (μ, ν) is *De March–Touzi irreducible*, if the irreducible convex paving of De March–Touzi [27] consists of the single irreducible component I. By [27, Theorem 2.1], this condition means that

(i) the set $I = \operatorname{ri} C$, with $C = \widehat{\operatorname{supp}}(\nu)$, satisfies $\mu(I) = 1$,
(ii) there exists some martingale transport $\pi^{\mathrm{DMT}} \in \mathsf{MT}(\mu, \nu)$ with the property that $C = \widehat{\operatorname{supp}}(\pi_x^{\mathrm{DMT}})$, for μ-a.e. $x \in \mathbb{R}^d$. We refer to π^{DMT} as a *De March–Touzi transport*.

The main assumption of this section (with the exception of Lemma 7.9, Lemma 7.19, Proposition 7.20 and Corollary 7.21) is the following.

Assumption 7.3. The pair (μ, ν) is De March–Touzi irreducible.

Assumption 7.3 is equivalent to the irreducibility of the pair (μ, ν) in the sense of Definition 1.2. For a proof of this result and for further equivalent characterizations of irreducibility we refer to Theorem D.1 in Appendix D.

In Lemma 7.5 below we give a more quantitative description of the defining property of a De March–Touzi transport $\pi^{\mathrm{DMT}} \in \mathsf{MT}(\mu, \nu)$ between an irreducible pair (μ, ν).

Definition 7.4. Let $K \subseteq I$ be compact. For an element y^* of the unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$ we define the slice S_{y^*} of C beyond K by

$$S_{y^*} := \big\{ y \in C \colon \langle y, y^* \rangle > \sup\{\langle \tilde{y}, y^* \rangle \colon \tilde{y} \in K\} \big\}. \tag{7.1}$$

Lemma 7.5. *Under Assumption 7.3, for μ-a.e. $x \in \mathbb{R}^d$ and for every compact set $K \subseteq I$, there exists a constant $\delta(K, x) > 0$ such that $\pi_x^{\mathrm{DMT}}(S_{y^*}) \geqslant \delta(K, x)$ for every $y^* \in \mathbb{S}^{d-1}$.*

Proof. We fix $x \in \mathbb{R}^d$ with $C = \widehat{\operatorname{supp}}(\pi_x^{\mathrm{DMT}})$ and note that the map $r_x \colon \mathbb{S}^{d-1} \to [0, 1]$ given by $r_x(y^*) := \pi_x^{\mathrm{DMT}}(S_{y^*})$, for $y^* \in \mathbb{S}^{d-1}$, is lower semicontinuous. As the set C equals the closed convex hull of the support of π_x^{DMT}, the map r_x is also strictly positive. Indeed, if there was some $y^* \in \mathbb{S}^{d-1}$ with $r_x(y^*) = 0$, then the closed convex set $C \setminus S_{y^*}$ would support π_x^{DMT}, which is in contradiction to $C \setminus S_{y^*} \subsetneq C = \widehat{\operatorname{supp}}(\pi_x^{\mathrm{DMT}})$. Since $\mathbb{S}^{d-1}$ is compact we conclude that r_x attains its minimum $\delta(K, x) := \min_{y^* \in \mathbb{S}^{d-1}} r_x(y^*) > 0$. □

The main result of this section is the following.

Theorem 7.6. *Under Assumption 7.3, there exists a dual optimizer ψ_{opt} in the sense of Definition 6.1.*

Together with Theorem 6.6, this has the following consequence for the primal optimizer $\pi^{\mathrm{SBM}} \in \mathsf{MT}(\mu, \nu)$ of the primal problem (1.5).

COROLLARY 7.7. *Under Assumption 7.3, there exists a Bass martingale $(M_t)_{0\leqslant t\leqslant 1}$ from μ to ν. Moreover,* $\mathrm{Law}(M_0, M_1) = \pi^{\mathrm{SBM}}$ *and, for μ-a.e. $x \in \mathbb{R}^d$, the measure π^{SBM}_x is equivalent to ν. In particular, π^{SBM} is a De March–Touzi transport in the sense of Definition 7.2, (ii).*

PROOF. Admitting Theorem 7.6, there exists a dual optimizer ψ_{opt}. Thus, by Theorem 6.6, there is a Bass martingale from μ to ν with the property that $\mathrm{Law}(M_0, M_1) = \pi^{\mathrm{SBM}}$. In particular, by (6.4), there is a measurable set A with $\mu(A) = 1$, such that π^{SBM}_x and $\pi^{\mathrm{SBM}}_{x'}$ are image measures of suitably translated Gaussians under the same function ∇v, for all $x, x' \in A$. Hence $\pi^{\mathrm{SBM}}_x \sim \pi^{\mathrm{SBM}}_{x'}$, for all $x, x' \in A$, and this in turn implies the equivalence of $\nu(dy) = \int \pi^{\mathrm{SBM}}_x(dy)\,\mu(dx)$ with $\pi^{\mathrm{SBM}}_{x'}(dy)$, for each $x' \in A$. In particular, $\mathrm{supp}(\pi^{\mathrm{SBM}}_x) = \mathrm{supp}(\nu)$, for μ-a.e. $x \in \mathbb{R}^d$, which implies Definition 7.2, (ii) for π^{SBM}. □

7.3. *Proof of Theorem 7.6.* The proof is split into two main steps.

7.3.1. *Step 1 of the proof of Theorem 7.6.* We first construct a *convergent* optimizing sequence of convex functions $(\psi_n)_{n\geqslant 1} \subseteq C^{\mathrm{aff}}_{\mathrm{q}}(\mathbb{R}^d)$ (see Remark 7.11 below) for the dual problem (3.1).

PROPOSITION 7.8. *Under Assumption 7.3, there is an optimizing sequence $(\psi_n)_{n\geqslant 1}$ of non-negative convex functions in $C^{\mathrm{aff}}_{\mathrm{q}}(\mathbb{R}^d)$ for the dual problem (3.1), which converges compactly on I (i.e., uniformly on compact subsets $K \subseteq I$) to some convex function $\psi_{\mathrm{lim}} \colon I \to [0, +\infty)$.*

Before we turn to the proof of Proposition 7.8, we still need some preparation. The following auxiliary result does not require the irreducibility Assumption 7.3, but solely relies on the finiteness of the value $\tilde{D}(\mu, \nu)$ of the dual problem (3.1).

LEMMA 7.9. *Let $(\psi_n)_{n\geqslant 1} \subseteq C_{\mathrm{q}}(\mathbb{R}^d)$ be an optimizing sequence for the dual problem (3.1) and take any $\pi \in \mathsf{MT}(\mu, \nu)$. Then*

$$\sup_{n\geqslant 1} \int \int \big(\psi_n(y) - \psi_n(x)\big)\,\pi_x(dy)\,\mu(dx) < +\infty. \tag{7.2}$$

PROOF. Recalling (3.2), we consider for $n \geqslant 1$ the functions

$$\mathbb{R}^d \ni x \longmapsto \varphi^{\psi_n}(x) = \inf_{p \in \mathcal{P}^x_2(\mathbb{R}^d)} \Big(\int \psi_n\, dp - \mathrm{MCov}(p, \gamma) \Big). \tag{7.3}$$

By taking $p = \delta_x$ in (7.3) we obtain the trivial estimate $\varphi^{\psi_n}(x) \leqslant \psi_n(x)$ and consequently

$$\int \int \big(\psi_n(y) - \psi_n(x)\big)\,\pi_x(dy)\,\mu(dx) \leqslant \int \int \big(\psi_n(y) - \varphi^{\psi_n}(x)\big)\,\pi_x(dy)\,\mu(dx).$$

Since $\psi_n \in C_{\mathrm{q}}(\mathbb{R}^d)$ and π has second marginal ν, the definition (3.5) gives

$$\mathcal{D}(\psi_n) = \int \Big(\int \psi_n(y)\,\pi_x(dy) - \varphi^{\psi_n}(x) \Big)\,\mu(dx),$$

and as the sequence of real numbers $(\mathcal{D}(\psi_n))_{n\geqslant 1}$ converges to the finite number $\tilde{D}(\mu, \nu)$, we conclude (7.2). □

DEFINITION 7.10. We fix an arbitrary optimizing sequence $(\psi_n)_{n\geqslant 1} \subseteq C_{\mathrm{q}}(\mathbb{R}^d)$ of *convex* functions for the dual problem (3.1), which is possible thanks to Proposition 3.5. For $x \in \mathbb{R}^d$, we define

$$\psi^x_n(\,\cdot\,) \coloneqq \psi_n(\,\cdot\,) - \psi_n(x) - \langle \partial\psi_n(x), \cdot - x \rangle, \tag{7.4}$$

so that $\psi_n^x(x) = 0$ and $\psi_n^x(\,\cdot\,)$ takes values in $[0, +\infty)$, for every $n \geqslant 1$. Here — by abuse of notation — $\partial\psi_n(x)$ denotes a subgradient of ψ_n at x, i.e., an arbitrary element of the subdifferential of ψ_n at x.

Remark 7.11. We observe that passing from $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$ to $\psi + \mathrm{aff}$, where $\mathrm{aff}\colon \mathbb{R}^d \to \mathbb{R}$ is an arbitrary affine function, does not change the value of the dual function (3.12), i.e., $\mathcal{D}(\psi) = \mathcal{D}(\psi + \mathrm{aff})$. In particular, we note that the sequence $(\psi_n^x)_{n\geqslant 1}$ defined in (7.4) is contained in the class of functions $C_{\mathrm{q}}^{\mathrm{aff}}(\mathbb{R}^d)$, as introduced in (2.1), and hence $\mathcal{D}(\psi_n^x) = \mathcal{D}(\psi_n)$.

Our goal is to show that — under the irreducibility Assumption 7.3 and after passing to Cesàro means of a suitable subsequence — the sequence $(\psi_n^x)_{n\geqslant 1}$ of (7.4) is bounded on compact subsets of I, for μ-a.e. $x \in I$. This is the content of the following lemma.

Lemma 7.12. *Under Assumption 7.3, for any given optimizing sequence $(\psi_n)_{n\geqslant 1} \subseteq C_{\mathrm{q}}(\mathbb{R}^d)$ of convex functions, there is a sequence of Cesàro means of a subsequence, still denoted by $(\psi_n)_{n\geqslant 1}$, such that, for μ-a.e. $x \in I$ and for every compact set $K \subseteq I$, we have that $\sup_{n\geqslant 1, y\in K} \psi_n^x(y) < +\infty$, for every choice of $(\psi_n^x)_{n\geqslant 1} \subseteq C_{\mathrm{q}}^{\mathrm{aff}}(\mathbb{R}^d)$ as in (7.4).*

Proof. We rely on [27] and fix a De March–Touzi transport $\pi^{\mathrm{DMT}} \in \mathsf{MT}(\mu, \nu)$ satisfying point (ii) of Definition 7.2, which exists by the irreducibility Assumption 7.3. Let $(\psi_n)_{n\geqslant 1}$ and $(\psi_n^x)_{n\geqslant 1}$, measurable selected for $x \in \mathbb{R}^d$, be as in Definition 7.10 above. By Lemma 7.9, the sequence

$$\Psi_n(x) \coloneqq \int \psi_n^x(y)\, \pi_x^{\mathrm{DMT}}(dy) = \int \big(\psi_n(y) - \psi_n(x)\big)\, \pi_x^{\mathrm{DMT}}(dy)$$

is bounded in $L^1(\mu)$. Applying Komlós' theorem [45, Theorem 1], we can find a subsequence $(n_k)_{k\geqslant 1} \subseteq \mathbb{N}$, such that the Cesàro means

$$\overline{\Psi}_k \coloneqq \frac{\Psi_{n_1} + \Psi_{n_2} + \ldots + \Psi_{n_k}}{k}$$

converge μ-a.s. to some random variable $\Psi \in L^1(\mu)$. If we define the Cesàro means

$$\bar{\psi}_k^x \coloneqq \frac{\psi_{n_1}^x + \psi_{n_2}^x + \ldots + \psi_{n_k}^x}{k},$$

we have

$$\overline{\Psi}_k(x) = \int \bar{\psi}_k^x(y)\, \pi_x^{\mathrm{DMT}}(dy), \qquad x \in \mathbb{R}^d.$$

Note that passing to a subsequence $(n_k)_{k\geqslant 1} \subseteq \mathbb{N}$ and forming Cesàro means $(\bar{\psi}_k^x)$ preserves the property of being an optimizing sequence of convex functions in $C_{\mathrm{q}}^{\mathrm{aff}}(\mathbb{R}^d)$ of the form (7.4). Therefore we can replace the original sequences (Ψ_n) and (ψ_n^x) by $(\overline{\Psi}_k)$ and $(\bar{\psi}_k^x)$, respectively, and may again relabel them as (Ψ_n) and (ψ_n^x), respectively. With that said, as a consequence of the μ-a.s. convergence to a finite limit, we have that

$$m(x) \coloneqq \sup_{n\geqslant 1} \Psi_n(x) = \sup_{n\geqslant 1} \int \psi_n^x(y)\, \pi_x^{\mathrm{DMT}}(dy) < +\infty, \tag{7.5}$$

for μ-a.e. $x \in \mathbb{R}^d$.

Arguing by contradiction to the statement of Lemma 7.12, we assume that there is a compact set $K_0 \subseteq I$ such that the set

$$A_1 \coloneqq \Big\{x \in \mathbb{R}^d \colon \sup_{n\geqslant 1, y\in K_0} \psi_n^x(y) = +\infty\Big\}$$

has positive μ-measure. Furthermore, by Lemma 7.5, the set

$$A_2 := \left\{x \in \mathbb{R}^d : \ \exists\, \delta(K_0, x) > 0 \text{ such that } \pi_x^{\mathrm{DMT}}(S_{y^*}) \geqslant \delta(K_0, x) \text{ for every } y^* \in \mathbb{S}^{d-1}\right\}$$

has full μ-measure. By (7.5) above, also the set

$$A_3 := \{x \in \mathbb{R}^d : m(x) < +\infty\}$$

has full μ-measure, so that the intersection $A := A_1 \cap A_2 \cap A_3$ has positive μ-measure. Pick a point $x_0 \in A$. As $x_0 \in A_1$, for arbitrarily large $M > m(x_0)/\delta(K_0, x_0)$, we can find $n_0 \geqslant 1$ and $y_0 \in K_0$ such that $\psi_{n_0}^{x_0}(y_0) \geqslant M$. The function

$$\ell(y) := \left\langle \partial\psi_{n_0}^{x_0}(y_0), y - y_0 \right\rangle + M, \qquad y \in \mathbb{R}^d,$$

satisfies $\ell \leqslant \psi_{n_0}^{x_0}$ and $\ell(y) \geqslant M$, for all $y \in S_{y^*}$, with $y^* := \frac{z}{|z|}$ and $z := \partial\psi_{n_0}^{x_0}(y_0)$. We conclude with

$$m(x_0) \geqslant \int \psi_n^x(y)\, \pi_{x_0}^{\mathrm{DMT}}(dy) \geqslant \int \ell^+(y)\, \pi_{x_0}^{\mathrm{DMT}}(dy) \geqslant M\, \pi_{x_0}^{\mathrm{DMT}}(S_{y^*}) \geqslant M\, \delta(K_0, x_0),$$

which is the desired contradiction, as M is arbitrarily large. □

Proof of Proposition 7.8. By Lemma 7.12 we can fix some $x_0 \in I$ such that the sequence of convex functions $(\psi_n^{x_0})_{n\geqslant 1}$ is bounded on all compact subsets $K \subseteq I$. In particular, the sequence $(\psi_n^{x_0})_{n\geqslant 1}$ is pointwise bounded on I. By [54, Theorem 10.9] we can select a subsequence, still denoted by $(\psi_n^{x_0})_{n\geqslant 1}$, which converges uniformly on compact subsets $K \subseteq I$ to some convex function $\psi_{\lim}^{x_0} : I \to [0, +\infty)$. Dropping the superscript x_0 to ease notation, we arrive at a sequence $(\psi_n)_{n\geqslant 1}$ with limit $\psi_{\lim}$ as required in the statement of Proposition 7.8. □

7.3.2. *Step 2 of the proof of Theorem 7.6.* We extend the function $\psi_{\lim} : I \to [0, +\infty)$ of Proposition 7.8 to a lower semicontinuous convex function $\psi_{\lim} : \mathbb{R}^d \to [0, +\infty]$ which is equal to $+\infty$ on $\mathbb{R}^d \setminus C$. Since $I \subseteq \operatorname{dom}\psi_{\lim}$, I is open and $\mu(I) = 1$, it follows that $\mu(\operatorname{int}(\operatorname{dom}\psi_{\lim})) = 1$.

Proposition 7.13. *Under Assumption 7.3, the convex function $\psi_{\lim} : \mathbb{R}^d \to [0, +\infty]$ is a dual optimizer in the sense of Definition 6.1, i.e., satisfies $\tilde{D}(\mu, \nu) = \mathcal{D}(\psi_{\lim})$.*

Definition 7.14. For μ-a.e. $x \in \mathbb{R}^d$, we denote by v_{SBM}^x the Brenier potential from γ to π_x^{SBM}, so that $\nabla v_{\mathrm{SBM}}^x(\gamma) = \pi_x^{\mathrm{SBM}}$ and

$$\mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) = \int \left\langle \nabla v_{\mathrm{SBM}}^x(z), z \right\rangle \gamma(dz);$$

we write $\psi_{\mathrm{SBM}}^x := (v_{\mathrm{SBM}}^x)^*$ for its convex conjugate.

Remark 7.15. As γ has full support, the convex function v_{SBM}^x is finite-valued and continuous everywhere on $\mathbb{R}^d$. In particular, v_{SBM}^x is unique, up to an additive constant. Therefore also its convex conjugate $\psi_{\mathrm{SBM}}^x : \mathbb{R}^d \to (-\infty, +\infty]$ is unique, up to an additive constant. Also note that ψ_{SBM}^x is finite-valued on I and takes the value $+\infty$ on $\mathbb{R}^d \setminus C$, with $C = \widehat{\mathrm{supp}}(\nu)$ and $I = \operatorname{int} C$.

Our goal is to show that $\psi_{\lim}\colon \mathbb{R}^d \to [0,+\infty]$ is a dual optimizer. For this purpose, we will prove that, for μ-a.e. $x \in I$, the function $\psi_{\lim}(\,\cdot\,)$ equals $\psi^x_{\mathrm{SBM}}(\,\cdot\,)$, modulo adding an affine function; in short, $\psi_{\lim} \equiv \psi^x_{\mathrm{SBM}}$ mod (aff), for μ-a.e. $x \in \mathbb{R}^d$. In Lemma 7.18 below we will show that the set

$$A \coloneqq \big\{x \in I\colon \psi_{\lim} \not\equiv \psi^x_{\mathrm{SBM}} \bmod (\mathrm{aff})\big\} \tag{7.6}$$

indeed has μ-measure zero. It will then follow from Lemma 7.19 that $\psi_{\lim}$ is actually a dual optimizer. First, we need some auxiliary results.

Lemma 7.16. *Under Assumption 7.3, for μ-a.e. $x \in A$, there exists a measure $\check{\pi}_x \in \mathcal{P}^x_2(\mathbb{R}^d)$ supported by C and there is a constant $\tilde{\beta}(x) > 0$ such that*

$$\mathrm{MCov}(\check{\pi}_x, \gamma) + \int \psi_{\lim}\, d(\pi^{\mathrm{SBM}}_x - \check{\pi}_x) \geqslant \mathrm{MCov}(\pi^{\mathrm{SBM}}_x, \gamma) + \tilde{\beta}(x). \tag{7.7}$$

Proof. Recalling the representation (3.12) of the dual function $\mathcal{D}(\,\cdot\,)$ and the definition (3.2) of the function $x \mapsto \varphi^{\psi}(x)$, we consider for μ-a.e. $x \in I$ the functions

$$\begin{aligned} \phi^{\psi^x_{\mathrm{SBM}}}(x) &\coloneqq \int \psi^x_{\mathrm{SBM}}(y)\, \pi^{\mathrm{SBM}}_x(dy) - \varphi^{\psi^x_{\mathrm{SBM}}}(x) \\ &= \sup_{p \in \mathcal{P}^x_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,\gamma) + \int \psi^x_{\mathrm{SBM}}\, d(\pi^{\mathrm{SBM}}_x - p)\Big) \end{aligned} \tag{7.8}$$

and

$$\begin{aligned} \phi^{\psi_{\lim}}(x) &\coloneqq \int \psi_{\lim}(y)\, \pi^{\mathrm{SBM}}_x(dy) - \varphi^{\psi_{\lim}}(x) \\ &= \sup_{p \in \mathcal{P}^x_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,\gamma) + \int \psi_{\lim}\, d(\pi^{\mathrm{SBM}}_x - p)\Big). \end{aligned} \tag{7.9}$$

Recalling Definition 7.14, we have $\nabla(\psi^x_{\mathrm{SBM}})^*(\gamma) = \pi^{\mathrm{SBM}}_x \in \mathcal{P}^x_2(\mathbb{R}^d)$. Therefore, we can apply part (i) of Proposition 5.3, which yields that the supremum in (7.8) is attained by π^{SBM}_x, so that $\phi^{\psi^x_{\mathrm{SBM}}}(x) = \mathrm{MCov}(\pi^{\mathrm{SBM}}_x, \gamma)$. By taking $p = \pi^{\mathrm{SBM}}_x$ in (7.9), we obtain the inequality $\phi^{\psi_{\lim}}(x) \geqslant \phi^{\psi^x_{\mathrm{SBM}}}(x)$, for μ-a.e. $x \in I$.

Now we define the sets

$$B \coloneqq \big\{x \in I\colon \phi^{\psi_{\lim}}(x) > \phi^{\psi^x_{\mathrm{SBM}}}(x)\big\}, \qquad X \coloneqq \big\{x \in I\colon \phi^{\psi^x_{\mathrm{SBM}}}(x) < +\infty\big\}$$

and claim that for the set A defined in (7.6) we have the relation $\tilde{A} \coloneqq A \cap X = B$. In other words, since $\mu(X) = 1$, the sets A and B are equal, up to a set of μ-measure zero.

In fact, if $x \notin \tilde{A}$, then $\psi_{\lim} \equiv \psi^x_{\mathrm{SBM}}$ mod (aff) or $\phi^{\psi^x_{\mathrm{SBM}}}(x) = +\infty$, so that in both cases $\phi^{\psi_{\lim}}(x) = \phi^{\psi^x_{\mathrm{SBM}}}(x)$. Conversely, if $x \notin B$, then $x \notin X$ or we have the equality $\phi^{\psi_{\lim}}(x) = \phi^{\psi^x_{\mathrm{SBM}}}(x)$ of real numbers and thus also the supremum in the definition (7.9) of $\phi^{\psi_{\lim}}(x)$ is attained by

$$\pi^{\mathrm{SBM}}_x = \nabla(\psi^x_{\mathrm{SBM}})^*(\gamma) \in \mathcal{P}^x_2(\mathbb{R}^d). \tag{7.10}$$

Hence we can apply part (ii) of Proposition 5.3 to the lower semicontinuous convex function $\psi_{\lim}\colon \mathbb{R}^d \to [0,+\infty]$ satisfying $\mu(\mathrm{int}(\mathrm{dom}\,\psi_{\lim})) = 1$, which tells us that the optimizer (7.10) of the supremum in (7.9) is equal to

$$\nabla\psi^*_{\lim}(\gamma_{\zeta(x)}),$$

for some $\zeta(x) \in \mathbb{R}^d$. We conclude the γ-a.s. equality

$$\nabla(\psi^x_{\mathrm{SBM}})^* = \nabla\psi^*_{\lim}\big(\,\cdot\,+\zeta(x)\big),$$

which is equivalent to the γ-a.s. equality

$$(\psi^x_{\mathrm{SBM}})^* = \psi^*_{\lim}\big(\,\cdot\,+\zeta(x)\big) + c, \tag{7.11}$$

for some constant $c \in \mathbb{R}$. Since the convex function $(\psi^x_{\mathrm{SBM}})^*$ is finite-valued and continuous everywhere on $\mathbb{R}^d$, the equality (7.11) has to hold everywhere on $\mathbb{R}^d$. This implies that $\psi_{\lim} \equiv \psi^x_{\mathrm{SBM}}$ mod (aff) and we deduce that $x \notin A$. Altogether, we have seen that $x \notin B$ implies that $x \notin \tilde{A}$.

As a consequence of $A = B$, up to a set of μ-measure zero, for μ-a.e. $x \in A$ we have $\phi^{\psi_{\lim}}(x) > \phi^{\psi^x_{\mathrm{SBM}}}(x)$, so that we can measurably select some $\check{\pi}_x \in \mathcal{P}^x_2(\mathbb{R}^d)$ with

$$\mathrm{MCov}(\check{\pi}_x, \gamma) + \int \psi_{\lim}\, d(\pi^{\mathrm{SBM}}_x - \check{\pi}_x) > \mathrm{MCov}(\pi^{\mathrm{SBM}}_x, \gamma), \tag{7.12}$$

which gives (7.7). As the right-hand side of (7.12) is finite, for μ-a.e. $x \in I$, and since $\psi_{\lim}(y) = +\infty$ for $y \in \mathbb{R}^d \setminus C$, we see that $\check{\pi}_x$ is supported by C. □

Our next step is to modify the measures $\{\check{\pi}_x\}_{x\in A}$ of Lemma 7.16, so that they have compact support and still satisfy (7.7) for some $\beta(x) > 0$ instead of $\tilde{\beta}(x)$. To this end, we choose an increasing sequence $(K_j)_{j\geqslant 1}$ of compact subsets of I such that $\bigcup_{j\geqslant 1} K_j = I$. Denoting by M^x the Bass martingale from δ_x to $\check{\pi}_x$, and by τ^x_j the first exit time of M^x from K_j (similarly as in Corollary 6.8), we define $\check{\pi}^j_x := \mathrm{Law}(M^x_{\tau^x_j \wedge 1})$. By optional sampling, $\check{\pi}^j_x \in \mathcal{P}^x_2(\mathbb{R}^d)$ and, by definition, $\check{\pi}^j_x$ is supported by the compact set K_j.

Lemma 7.17. *Under Assumption 7.3, for μ-a.e. $x \in A$, there exists $\check{\pi}^{j(x)}_x \in \mathcal{P}^x_2(\mathbb{R}^d)$ supported by $K_{j(x)}$ for some $j(x) \in \mathbb{N}$ and there is a constant $\beta(x) > 0$ such that*

$$\mathrm{MCov}(\check{\pi}^{j(x)}_x, \gamma) + \int \psi_{\lim}\, d(\pi^{\mathrm{SBM}}_x - \check{\pi}^{j(x)}_x) \geqslant \mathrm{MCov}(\pi^{\mathrm{SBM}}_x, \gamma) + \beta(x). \tag{7.13}$$

Proof. From Lemma 7.16 we already have the inequality (7.7). In order to derive (7.13), we have to show that

$$\lim_{j\to+\infty} \mathrm{MCov}(\check{\pi}^j_x, \gamma) = \mathrm{MCov}(\check{\pi}_x, \gamma) \tag{7.14}$$

and

$$\limsup_{j\to+\infty} \int \psi_{\lim}(y)\, d\check{\pi}^j_x(y) \leqslant \int \psi_{\lim}(y)\, d\check{\pi}_x(y). \tag{7.15}$$

We begin with the proof of (7.14). First, observe that $M^x_{\tau^x_j\wedge 1} \to M^x_{\tau^x\wedge 1}$ in L^2, where τ^x is the first exit time of M^x from I. Since $\tau^x \wedge 1 = 1$ a.s. by Corollary 6.8, we conclude that $M^x_{\tau^x_j\wedge 1} \to M^x_1$ in L^2. Consequently,

$$\mathcal{W}_2^2(\check{\pi}^j_x, \check{\pi}_x) := \inf_{q\in\mathsf{Cpl}(\check{\pi}^j_x,\check{\pi}_x)} \int |x_1 - x_2|^2\, q(dx_1, dx_2)$$

converges to zero as $j \to +\infty$, and the inequality

$$|\mathrm{MCov}(\check{\pi}^j_x, \gamma) - \mathrm{MCov}(\check{\pi}_x, \gamma)| \leqslant \mathcal{W}_2(\check{\pi}^j_x, \check{\pi}_x)\,\sqrt{d}$$

yields (7.14).

Finally, we show (7.15). Note that by optional sampling, $\check{\pi}_x^j \in \mathcal{P}_2^x(\mathbb{R}^d)$. Since a martingale composed with a convex function is a submartingale, it follows that

$$\forall j \geqslant 1 \colon \int \psi_{\lim}\, d\check{\pi}_x^j \leqslant \int \psi_{\lim}\, d\check{\pi}_x$$

and we obtain (7.15). □

Lemma 7.18. *Under Assumption 7.3, the set $A \subseteq \mathbb{R}^d$ defined in (7.6) has μ-measure zero.*

Proof. We assume for contradiction that $\mu(A) > 0$. Thanks to Lemma 7.17, the set A is μ-a.s. equal to the union

$$\bigcup_{\substack{j\in\mathbb{N},\\ \beta\in\mathbb{Q}_+\setminus\{0\}}} \left\{x \in A \colon \mathrm{MCov}(\check{\pi}_x^j, \gamma) + \int \psi_{\lim}\, d(\pi_x^{\mathrm{SBM}} - \check{\pi}_x^j) \geqslant \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) + \beta\right\}.$$

Hence we can find a subset $B \subseteq A$ with $\mu(B) > 0$, such that (7.13) holds with uniform constants $j \in \mathbb{N}$ and $\beta > 0$ for all $x \in B$, i.e.,

$$\exists j\ \exists \beta\ \forall x \in B \colon \quad \mathrm{MCov}(\check{\pi}_x^j, \gamma) + \int \psi_{\lim}\, d(\pi_x^{\mathrm{SBM}} - \check{\pi}_x^j) \geqslant \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) + \beta. \tag{7.16}$$

Now we define a measurable collection of probability measures $\{\tilde{\pi}_x\}_{x\in\mathbb{R}^d} \subseteq \mathcal{P}_2(\mathbb{R}^d)$ with $\mathrm{bary}(\tilde{\pi}_x) = x$ by

$$\tilde{\pi}_x \coloneqq \begin{cases} \check{\pi}_x^j, & x \in B, \\ \pi_x^{\mathrm{SBM}}, & x \in \mathbb{R}^d \setminus B. \end{cases} \tag{7.17}$$

From the inequality (7.16) and the definition (7.17) we deduce that

$$\int \int \psi_{\lim}\, d(\pi_x^{\mathrm{SBM}} - \tilde{\pi}_x)\, d\mu(x) + \int \mathrm{MCov}(\tilde{\pi}_x, \gamma)\, d\mu(x) \tag{7.18}$$

$$\geqslant \int \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma)\, d\mu(x) + \beta\, \mu(B) \tag{7.19}$$

$$= P(\mu, \nu) + \beta\, \mu(B). \tag{7.20}$$

Recall from Proposition 7.8 that there is an optimizing sequence $(\psi_n)_{n\geqslant 1}$ of convex functions $\psi_n \colon \mathbb{R}^d \to [0, +\infty)$ for the dual problem (3.1), which converges uniformly on compact subsets $K \subseteq I$ to $\psi_{\lim}$. To find the desired contradiction, we want to replace $\psi_{\lim}$ in (7.18) by $\lim_{n\to\infty} \psi_n$ and write the limit outside of the integral. This is clearly not a problem if $x \notin B$. For $x \in B$, note that the measure $\tilde{\pi}_x = \check{\pi}_x^j$ is supported by the compact set K_j. As $(\psi_n)_{n\geqslant 1}$ converges to $\psi_{\lim}$ uniformly on K_j, we have

$$\lim_{n\to\infty} \int_B \int \psi_n(y)\, d\tilde{\pi}_x(y)\, d\mu(x) = \int_B \int \psi_{\lim}(y)\, d\tilde{\pi}_x(y)\, d\mu(x). \tag{7.21}$$

On the other hand, Fatou's lemma gives

$$\liminf_{n\to\infty} \int_B \int \psi_n(y)\, d\pi_x^{\mathrm{SBM}}(y)\, d\mu(x) \geqslant \int_B \int \psi_{\lim}(y)\, d\pi_x^{\mathrm{SBM}}(y)\, d\mu(x). \tag{7.22}$$

Now combining (7.18) – (7.20) with (7.21), (7.22) yields the inequality

$$\liminf_{n\to\infty} \int \int \psi_n\, d(\pi_x^{\mathrm{SBM}} - \tilde{\pi}_x)\, d\mu(x) + \int \mathrm{MCov}(\tilde{\pi}_x, \gamma)\, d\mu(x) \geqslant P(\mu, \nu) + \beta\, \mu(B).$$

Recalling the definition (3.2) of the function $x \mapsto \varphi^{\psi}(x)$ and the dual function (3.12), we observe that the left-hand side of this inequality is less than or equal to

$$\liminf_{n\to\infty} \mathcal{D}(\psi_n) = \liminf_{n\to\infty} \int \Big(\int \psi_n(y)\, \pi_x^{\mathrm{SBM}}(dy) - \varphi^{\psi_n}(x) \Big)\, \mu(dx).$$

Since $(\psi_n)_{n\geqslant 1}$ is an optimizing sequence for the dual problem, it follows that

$$\tilde{D}(\mu,\nu) = \lim_{n\to\infty} \mathcal{D}(\psi_n) = \liminf_{n\to\infty} \mathcal{D}(\psi_n) \geqslant P(\mu,\nu) + \beta\,\mu(B) > P(\mu,\nu).$$

But this is a contradiction to the fact that there is no duality gap by Theorem 3.3. □

Lemma 7.19. *A lower semicontinuous convex function $\psi_{\mathrm{opt}}\colon \mathbb{R}^d \to (-\infty,+\infty]$ satisfying $\mu(\mathrm{int}(\mathrm{dom}\,\psi_{\mathrm{opt}})) = 1$ is a dual optimizer in the sense of Definition 6.1 if and only if*

$$\mu\big(\big\{x \in \mathbb{R}^d\colon \psi_{\mathrm{opt}} \equiv \psi^x_{\mathrm{SBM}} \bmod (\mathrm{aff})\big\}\big) = 1. \tag{7.23}$$

Proof. "⇐": As we have seen in the proof of Lemma 7.16, for μ-a.e. $x \in \mathbb{R}^d$, the supremum in (7.8) is attained by $p = \pi_x^{\mathrm{SBM}}$. Hence integrating with respect to $\mu(dx)$ yields

$$\int \mathrm{MCov}(\pi_x^{\mathrm{SBM}},\gamma)\,\mu(dx) = \int \Big(\int \psi^x_{\mathrm{SBM}}(y)\, \pi_x^{\mathrm{SBM}}(dy) - \varphi^{\psi^x_{\mathrm{SBM}}}(x) \Big)\, \mu(dx).$$

By assumption (7.23) it follows that

$$\int \mathrm{MCov}(\pi_x^{\mathrm{SBM}},\gamma)\,\mu(dx) = \int \Big(\int \psi_{\mathrm{opt}}(y)\, \pi_x^{\mathrm{SBM}}(dy) - \varphi^{\psi_{\mathrm{opt}}}(x) \Big)\, \mu(dx). \tag{7.24}$$

Recalling the representation (3.12) of the dual function $\mathcal{D}(\,\cdot\,)$, we see that the expression on the right-hand side of (7.24) equals $\mathcal{D}(\psi_{\mathrm{opt}})$. On the other hand, the left-hand side is equal to $P(\mu,\nu) = \tilde{D}(\mu,\nu)$ by Theorem 3.3, so that ψ_{opt} is a dual optimizer.

"⇒": Let ψ_{opt} be a dual optimizer. As in the proof of the implication "⇒" in Theorem 6.6 we conclude that

$$\pi_x^{\mathrm{SBM}} = \nabla \psi^*_{\mathrm{opt}}\big(\zeta(x) + \,\cdot\,\big)(\gamma),$$

for μ-a.e. $x \in \mathbb{R}^d$, with $\zeta(x) \in \mathbb{R}^d$ as in (6.5). On the other hand, by Definition 7.14 we have $\pi_x^{\mathrm{SBM}} = \nabla(\psi^x_{\mathrm{SBM}})^*(\gamma)$, for μ-a.e. $x \in \mathbb{R}^d$. Arguing as in the proof of Lemma 7.16, we obtain (7.23). □

Let us summarize why the above arguments complete the proof of Proposition 7.13 and thus also of Theorem 7.6.

Proof of Proposition 7.13. According to Lemma 7.18, the function $\psi_{\mathrm{lim}}\colon \mathbb{R}^d \to [0,+\infty]$ equals ψ^x_{SBM} mod (aff), for μ-a.e. $x \in I$. By Lemma 7.19, this implies that ψ_{lim} is a dual optimizer. □

Proof of Theorem 7.6. The existence of a dual optimizer follows from Proposition 7.8 and Proposition 7.13. □

Let us have one more look at the structure of the proof of Theorem 7.6 and Corollary 7.7 above. We started with an arbitrary optimizing sequence $(\psi_n)_{n\geqslant 1} \subseteq C_{\mathrm{q}}(\mathbb{R}^d)$ of convex functions for the dual problem (3.1), which we normalized to obtain $(\psi^x_n)_{n\geqslant 1}$ as in (7.4). Then we showed in Proposition 7.8 that under the irreducibility Assumption 7.3, which guarantees the existence of a De March–Touzi transport $\pi^{\mathrm{DMT}} \in \mathsf{MT}(\mu,\nu)$, we could find

a limiting function $\psi_{\lim}$. However, this was only possible *after passing to a subsequence, forming Cesàro means, and then choosing a further subsequence of* $(\psi_n^x)_{n\geqslant 1}$. In Proposition 7.13 we argued that $\psi_{\lim}$ is a dual optimizer and hence, by Theorem 6.6, there is a Bass martingale from μ to ν.

In the follow-up paper [56], the general case of a pair (μ,ν) which is not necessarily irreducible — and thus does not necessarily satisfy Assumption 7.3 — is treated. In this case, a variant of this line of reasoning is needed. Suppose we *already know* that an optimizing sequence $(\psi_n)_{n\geqslant 1}$ is pointwise bounded on a relatively open convex set $I\subseteq\mathbb{R}^d$ with $\mu(I)=1$ and such that $\widehat{\operatorname{supp}}(\nu)$ is contained in the closure of I. Under this assumption — *but without imposing irreducibility on the pair* (μ,ν) — we will prove in Proposition 7.20 below that there is no need of passing to a subsequence or of forming convex combinations of the optimizing sequence $(\psi_n)_{n\geqslant 1}$. Rather already the original sequence $(\psi_n)_{n\geqslant 1}$ — modulo adding affine functions — converges.

Proposition 7.20. *Let $\mu,\nu\in\mathcal{P}_2(\mathbb{R}^d)$ with $\mu\preceq_{\mathrm{c}}\nu$. Let $I\subseteq\mathbb{R}^d$ be a relatively open convex set with $\mu(I)=1$. Denote by C the closure of I and assume that $\widehat{\operatorname{supp}}(\nu)\subseteq C$. Let $(\psi_n)_{n\geqslant 1}$ be an optimizing sequence of non-negative convex functions in $C_{\mathrm{q}}^{\mathrm{aff}}(\mathbb{R}^d)$ for the dual problem (3.1) such that*

$$\forall y\in I\colon\ \sup_{n\geqslant 1}\psi_n(y)<+\infty. \tag{7.25}$$

Then there is a lower semicontinuous convex function $\psi_{\lim}\colon\mathbb{R}^d\to[0,+\infty]$ and a sequence $(\tilde\psi_n)_{n\geqslant 1}$ such that $\psi_n\equiv\tilde\psi_n$ mod (aff), for each $n\geqslant 1$, and

$$\forall y\in I\colon\ \psi_{\lim}(y)=\lim_{n\to\infty}\tilde\psi_n(y)<+\infty, \tag{7.26}$$

$$\forall y\in\mathbb{R}^d\setminus C\colon\ \psi_{\lim}(y)=\lim_{n\to\infty}\tilde\psi_n(y)=+\infty. \tag{7.27}$$

The convergence in (7.26) is uniform on compact subsets of I. Moreover, we have that $C=\widehat{\operatorname{supp}}(\nu)$ and $\psi_{\lim}$ is a dual optimizer, which is unique modulo adding affine functions.

The proof of Proposition 7.20 is delayed until Appendix C.

Corollary 7.21. *Under the assumptions of Proposition 7.20, there exists a Bass martingale $(M_t)_{0\leqslant t\leqslant 1}$ from μ to ν. Moreover, $\operatorname{Law}(M_0,M_1)=\pi^{\mathrm{SBM}}$ and, for μ-a.e. $x\in I$, the measure π_x^{SBM} is equivalent to ν. In particular, π^{SBM} is a De March–Touzi transport and the pair (μ,ν) is irreducible.*

Proof. By Proposition 7.20, the limiting function $\psi_{\lim}$ is a dual optimizer. Therefore, by analogy with the deduction of Corollary 7.7 from Theorem 7.6, we deduce Corollary 7.21 from Proposition 7.20. To see that the pair (μ,ν) is irreducible we refer to Theorem D.1. □

7.4. *Proof of Theorem 1.3.* We are now in the position to prove our first main result of the introduction.

Proof of Theorem 1.3. The implication "(2) ⇒ (1)" is Theorem 1.10 of [5]. For the proof of "(1) ⇒ (2)" we apply Corollary 7.7 and obtain the existence of a Bass martingale from μ to ν. By the uniqueness results of Theorem 2.2 in [5], this Bass martingale has to agree with the given stretched Brownian motion. □

7.5. *Proof of the second part of Theorem 1.4.* Finally, we complete the proof of Theorem 1.4. Recalling the results of Section 4 and the definition (4.3) of $\mathcal{E}(\,\cdot\,)$, we can formulate the second part of Theorem 1.4 equivalently as follows.

PROPOSITION 7.22. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. The value*

$$D_{\mathrm{rel}}(\mu,\nu) = \inf_{\substack{\mu(\operatorname{dom}\psi)=1,\\ \psi \text{ convex}}} \mathcal{E}(\psi) \tag{7.28}$$

is attained by a lower semicontinuous convex function $\psi_{\mathrm{opt}}\colon \mathbb{R}^d \to (-\infty,+\infty]$ satisfying $\mu(\mathrm{ri}(\operatorname{dom}\psi_{\mathrm{opt}})) = 1$ if and only if (μ,ν) is irreducible. In this case the (unique) optimizer to (MBB) *is given by the Bass martingale*

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

where $v = \psi_{\mathrm{opt}}^*$ and $B_0 \sim \nabla(\psi_{\mathrm{opt}}^* * \gamma)^*(\mu)$.

PROOF. We call a function ψ_{opt} as in the statement of the proposition a dual optimizer of (7.28) and first show that this notion is equivalent to a dual optimizer in the sense of Definition 6.1. Indeed, if ψ_{opt} is a dual optimizer of (7.28), it follows from the inequality $(\psi^* * \gamma)^* = (\varphi^\psi)^{**} \leqslant \varphi^\psi$ that ψ_{opt} is also a dual optimizer according to Definition 6.1. Conversely, suppose that ψ_{opt} is an optimizer in the latter sense. Then it follows from Lemma 7.19 that (7.23) is satisfied. Note that by the classical Kantorovich duality we have

$$\mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) = \int \psi_{\mathrm{SBM}}^x \, d\pi_x^{\mathrm{SBM}} + \int (\psi_{\mathrm{SBM}}^x)^* \, d\gamma,$$

for μ-a.e. $x \in \mathbb{R}^d$. Thus, by (7.23), for μ-a.e. $x \in \mathbb{R}^d$ there exists $\zeta(x) \in \mathbb{R}^d$ such that

$$\mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) = \int \big(\psi_{\mathrm{opt}}(\,\cdot\,) - \langle \zeta(x), \cdot\,\rangle\big)\, d\pi_x^{\mathrm{SBM}} + \int \big(\psi_{\mathrm{opt}}(\,\cdot\,) - \langle \zeta(x), \cdot\,\rangle\big)^*\, d\gamma.$$

Reading the proof of Proposition 4.2 backwards we conclude that

$$\int \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma)\, d\mu(x) = \int \Big(\int \psi_{\mathrm{opt}}(y)\, \pi_x^{\mathrm{SBM}}(dy) - (\psi_{\mathrm{opt}}^* * \gamma)^*(x) \Big) \mu(dx),$$

i.e., $D_{\mathrm{rel}}(\mu,\nu) = \mathcal{E}(\psi_{\mathrm{opt}})$. This shows that ψ_{opt} is also a dual optimizer of (7.28).

Now that we know that both definitions of a dual optimizer are equivalent, we can conclude the assertions of Proposition 7.22 from the results we have already established. Indeed, by Theorem 6.6, the existence of a dual optimizer is equivalent to the existence of a Bass martingale $(M_t)_{0\leqslant t\leqslant 1}$ in the given form. Finally, by Theorem 1.3 and Theorem D.1, the existence of a Bass martingale is equivalent to the irreducibility of (μ,ν). □

APPENDIX A: PROOF OF THEOREM 3.3

PROOF OF THEOREM 3.3. Existence and uniqueness of the optimizer $\pi^{\mathrm{SBM}} \in \mathsf{MT}(\mu,\nu)$ of the primal problem (1.5), as well as finiteness of the primal value $P(\mu,\nu)$ in (3.3), were proved in [5, Theorem 2.2]. In order to show that there is no duality gap, we apply [6, Theorem 1.3] with the cost function

$$C(x,p) := \begin{cases} -\mathrm{MCov}(p,\gamma) + \frac{1}{2}\int |y|^2\, dp, & \text{if } \mathrm{bary}(p) = x,\\ +\infty, & \text{if } \mathrm{bary}(p) \neq x, \end{cases}$$

for $x \in \mathbb{R}^d$ and $p \in \mathcal{P}_2(\mathbb{R}^d)$. This function is bounded from below and convex in the second argument. If we equip $\mathcal{P}_2(\mathbb{R}^d)$ with the topology induced by the quadratic Wasserstein

distance one verifies that $C(x,p)$ is also jointly lower semicontinuous with respect to the product topology on $\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$. We introduce the space of continuous functions which are bounded from below and have at most quadratic growth

$$C_{\mathrm{b,q}}(\mathbb{R}^d) := \big\{\tilde{\psi} : \mathbb{R}^d \to \mathbb{R} \text{ continuous s.t. } \exists\, a,k,\ell \in \mathbb{R} \text{ with } \ell \leqslant \tilde{\psi}(\,\cdot\,) \leqslant a + k|\,\cdot\,|^2\big\}.$$

Then by [6, Theorem 1.3] the value $P(\mu,\nu)$ of the primal problem (1.5) equals

$$\tilde{D}_{\mathrm{b,q}}(\mu,\nu) := \inf_{\tilde{\psi}\in C_{\mathrm{b,q}}(\mathbb{R}^d)} \Big(\int \big(\tilde{\psi}(\,\cdot\,) + \tfrac{|\,\cdot\,|^2}{2}\big)\, d\nu - \int \tilde{\varphi}^{\tilde{\psi}}\, d\mu \Big),$$

where

$$\tilde{\varphi}^{\tilde{\psi}}(x) := \inf_{p\in\mathcal{P}_2^x(\mathbb{R}^d)} \Big(\int \big(\tilde{\psi}(\,\cdot\,) + \tfrac{|\,\cdot\,|^2}{2}\big)\, dp - \mathrm{MCov}(p,\gamma) \Big).$$

Finally, passing from the functions $\tilde{\psi} \in C_{\mathrm{b,q}}(\mathbb{R}^d)$ to $\psi(\,\cdot\,) := \tilde{\psi}(\,\cdot\,) + \frac{|\,\cdot\,|^2}{2} \in C_{\mathrm{q}}(\mathbb{R}^d)$, we see that $\tilde{D}_{\mathrm{b,q}}(\mu,\nu) = \tilde{D}(\mu,\nu)$. □

APPENDIX B: PROOFS OF LEMMAS 4.4, 5.1, 6.2, 6.4 AND 6.5

Proof of Lemma 4.4. Using probabilistic notation, we rewrite the supremum in (4.5) as

$$\varrho^{\psi} = \sup \mathbb{E}\big[\langle Y, Z\rangle - \psi(Y)\big], \tag{B.1}$$

where the supremum in (B.1) is taken over all probability spaces such that $Z \sim \gamma$ and Y is an $\mathbb{R}^d$-valued random variable with finite second moment. Replacing Y by $\mathbb{E}[Y \,|\, Z]$, we observe that the maximization in (B.1) can be restricted to random variables Y which are measurable functions of Z, and we obtain

$$\varrho^{\psi} = \sup_{Y\in L^2(\gamma;\mathbb{R}^d)} \int \Big(\big\langle Y(z), z\big\rangle - \psi\big(Y(z)\big)\Big)\, \gamma(dz), \tag{B.2}$$

where $L^2(\gamma;\mathbb{R}^d)$ denotes the space of $\mathbb{R}^d$-valued Borel measurable functions on $\mathbb{R}^d$, which are square-integrable under γ. Clearly, for any $Y \in L^2(\gamma;\mathbb{R}^d)$ we have

$$\int \Big(\big\langle Y(z), z\big\rangle - \psi\big(Y(z)\big)\Big)\, \gamma(dz) \leqslant \int \sup_{y\in\mathbb{R}^d} \big(\langle y, z\rangle - \psi(y)\big)\, \gamma(dz) = \int \psi^*(z)\, \gamma(dz),$$

which shows the inequality $\varrho^{\psi} \leqslant \int \psi^*\, d\gamma$. In order to see the reverse inequality, we define the auxiliary problem

$$\varrho^{\psi}_{\infty} := \sup_{Y\in L^\infty(\gamma;\mathbb{R}^d)} \int \Big(\big\langle Y(z), z\big\rangle - \psi\big(Y(z)\big)\Big)\, \gamma(dz), \tag{B.3}$$

where $L^\infty(\gamma;\mathbb{R}^d)$ denotes the space of $\mathbb{R}^d$-valued Borel measurable functions on $\mathbb{R}^d$, which are bounded γ-a.e. Comparing (B.2) with (B.3), we obviously have $\varrho^{\psi} \geqslant \varrho^{\psi}_{\infty}$. Now we claim that

$$\varrho^{\psi}_{\infty} \geqslant \int \sup_{y\in\mathbb{R}^d} \big(\langle y, z\rangle - \psi(y)\big)\, \gamma(dz) = \int \psi^*(z)\, \gamma(dz), \tag{B.4}$$

which will finish the proof of (4.7). To see this, we first write

$$\varrho^{\psi}_{\infty} = \lim_{N\to\infty} \sup_{\substack{Y\in L^\infty(\gamma;\mathbb{R}^d),\\ |Y|\leqslant N}} \int \Big(\big\langle Y(z), z\big\rangle - \psi\big(Y(z)\big)\Big)\, \gamma(dz).$$

Using a measurable selection argument, we obtain

$$\sup_{\substack{Y\in L^\infty(\gamma;\mathbb{R}^d),\\ |Y|\leqslant N}} \int \Big(\big\langle Y(z),z\big\rangle - \psi\big(Y(z)\big)\Big)\,\gamma(dz) \geqslant \int \sup_{\substack{y\in\mathbb{R}^d,\\ |y|\leqslant N}} \big(\langle y,z\rangle - \psi(y)\big)\,\gamma(dz).$$

Since ψ is proper we can choose $y_0 \in \operatorname{dom}\psi \neq \varnothing$. Then for N large enough we have

$$\sup_{\substack{y\in\mathbb{R}^d,\\ |y|\leqslant N}} \big(\langle y,z\rangle - \psi(y)\big) \geqslant \langle y_0,z\rangle - \psi(y_0),$$

with the right-hand side being integrable with respect to $\gamma(dz)$. Hence we can apply the monotone convergence theorem and deduce that

$$\lim_{N\to\infty} \int \sup_{\substack{y\in\mathbb{R}^d,\\ |y|\leqslant N}} \big(\langle y,z\rangle - \psi(y)\big)\,\gamma(dz) = \int \sup_{y\in\mathbb{R}^d} \big(\langle y,z\rangle - \psi(y)\big)\,\gamma(dz),$$

which completes the proof of (B.4). □

Proof of Lemma 5.1. To see that φ^ψ is convex on $\operatorname{dom}\psi$, we let $x \coloneqq cx_1 + (1-c)x_2$ for $x_1, x_2 \in \operatorname{dom}\psi$ and $c \in (0,1)$. If $\varphi^\psi(x_1) = -\infty$, then there are $p_{x_1}^{(n)} \in \mathcal{P}_2^{x_1}(\mathbb{R}^d)$ with $\int \psi\, dp_{x_1}^{(n)} - \mathrm{MCov}(p_{x_1}^{(n)},\gamma) \to -\infty$. We observe that $p_x^{(n)} \coloneqq cp_{x_1}^{(n)} + (1-c)\delta_{x_2} \in \mathcal{P}_2^x(\mathbb{R}^d)$, and consequently

$$\begin{aligned}\varphi^\psi(x) &\leqslant c\int \psi\, dp_{x_1}^{(n)} + (1-c)\psi(x_2) - \mathrm{MCov}(p_x^{(n)},\gamma)\\ &\leqslant c\Big(\int \psi\, dp_{x_1}^{(n)} - \mathrm{MCov}(p_{x_1}^{(n)},\gamma)\Big) + (1-c)\psi(x_2),\end{aligned}$$

where we have used the convexity of $\mathcal{P}_2(\mathbb{R}^d) \ni p \mapsto -\mathrm{MCov}(p,\gamma)$. We conclude that $\varphi^\psi(x) = -\infty$. The case $\varphi^\psi(x_2) = -\infty$ is treated similarly. If, on the other hand, both $\varphi^\psi(x_1) > -\infty$ and $\varphi^\psi(x_2) > -\infty$, then $\varphi^\psi(x) \leqslant c\varphi^\psi(x_1) + (1-c)\varphi^\psi(x_2)$ follows by standard arguments. All in all we see that φ^ψ is convex on $\operatorname{dom}\psi$.

If $\varphi^\psi(x) > -\infty$ for one $x \in \operatorname{int}(\operatorname{dom}\psi)$, then $\varphi^\psi(\tilde{x}) > -\infty$ for all $\tilde{x} \in \operatorname{int}(\operatorname{dom}\psi)$, as can be seen directly by convexity.

Finally, given some $x \in \operatorname{int}(\operatorname{dom}\psi)$ with $\varphi^\psi(x) > -\infty$, we show that ψ is co-finite. Without loss of generality, we assume that $x = 0$ and $\psi(0) = 0$. Using probabilistic notation, we rewrite the supremum in (4.4) as

$$-\varphi^\psi(0) = \sup \mathbb{E}\big[\langle Y,Z\rangle - \psi(Y)\big] < +\infty, \tag{B.5}$$

where the supremum in (B.5) is taken over all probability spaces such that $Z \sim \gamma$ and Y is an $\mathbb{R}^d$-valued random variable with finite second moment and $\mathbb{E}[Y] = 0$. By contradiction, suppose there is $e \in \mathbb{R}^d \setminus \{0\}$ such that

$$\lim_{t\to+\infty} \frac{\psi(te)}{t} < +\infty. \tag{B.6}$$

We define the convex function $\bar{\psi}\colon \mathbb{R} \to (-\infty,+\infty]$ by $\bar{\psi}(t) \coloneqq \psi(te)$, for $t \in \mathbb{R}$. By assumption we have $0 \in \operatorname{int}(\operatorname{dom}\psi)$, thus also $0 \in \operatorname{int}(\operatorname{dom}\bar{\psi})$. In particular, $\bar{\psi}$ is continuous in a neighbourhood of 0. By (B.6), there is a constant K such that $\bar{\psi}(t) \leqslant Kt$, for all t large enough. As

$\bar\psi(0)=0$ and $\bar\psi$ is convex, we conclude that $\bar\psi(t)\leqslant Kt$, for all $t\geqslant 0$. Next we introduce four parameters, $A<0$ and $\delta, M, C>0$, and define the $\mathbb{R}^d$-valued function

$$\mathbb{R}^d\ni z\longmapsto f^{M,C}_{\delta,A}(z)\coloneqq\begin{cases} Me, & \text{if } \langle e,z\rangle\geqslant C,\\ -\delta e, & \text{if } \langle e,z\rangle\in(A,0),\\ 0, & \text{else.}\end{cases}$$

Given $Z\sim\gamma$, this induces a random variable $Y^{M,C}_{\delta,A}\coloneqq f^{M,C}_{\delta,A}(Z)$. We impose the relation

$$M\,\mathbb{P}\big[\langle e,Z\rangle\geqslant C\big]=\delta\,\mathbb{P}\big[\langle e,Z\rangle\in(A,0)\big] \tag{B.7}$$

between M and δ, so that $\mathbb{E}[Y^{M,C}_{\delta,A}]=0$. Now we fix $A<0$ and some $\varepsilon>0$. Then we choose $\delta>0$ small enough such that

$$|\bar\psi(-\delta)|\vee\delta\,\mathbb{E}\big[|\langle e,Z\rangle|\big]\leqslant\tfrac{\varepsilon}{2},$$

which we achieve by continuity of $\bar\psi$ and the fact that $\bar\psi(0)=0$. We leave C as a free parameter, which fixes M via (B.7). We compute

$$\begin{aligned}
&\mathbb{E}\big[\langle Y^{M,C}_{\delta,A},Z\rangle-\psi(Y^{M,C}_{\delta,A})\big]\\
&=\mathbb{E}\big[\mathbf{1}_{\langle e,Z\rangle\geqslant C}\big(M\langle e,Z\rangle-\bar\psi(M)\big)+\mathbf{1}_{\langle e,Z\rangle\in(A,0)}\big(-\delta\langle e,Z\rangle-\bar\psi(-\delta)\big)\big]\\
&\geqslant M\,\mathbb{E}\big[\mathbf{1}_{\langle e,Z\rangle\geqslant C}\big(\langle e,Z\rangle-K\big)\big]-\delta\,\mathbb{E}\big[\mathbf{1}_{\langle e,Z\rangle\in(A,0)}\langle e,Z\rangle\big]-\bar\psi(-\delta)\,\mathbb{P}\big[\langle e,Z\rangle\in(A,0)\big]\\
&\geqslant M(C-K)\,\mathbb{P}\big[\langle e,Z\rangle\geqslant C\big]-\varepsilon\\
&=\delta(C-K)\,\mathbb{P}\big[\langle e,Z\rangle\in(A,0)\big]-\varepsilon.
\end{aligned}$$

Now taking $C\nearrow+\infty$ we conclude that $-\varphi^{\psi}(0)=+\infty$, which is a contradiction to (B.5). □

Proof of Lemma 6.2. By contradiction, we suppose that

$$\dim(\operatorname{dom}\psi_{\mathrm{opt}})<\dim(\operatorname{supp}\nu)=d.$$

Defining $B_\ell\coloneqq\{x\in\mathbb{R}^d\colon\psi_{\mathrm{opt}}(x)\leqslant\ell\}$ for $\ell\geqslant 1$, we clearly have $\mu(B_\ell)\nearrow 1$ and furthermore

$$\sup_{\ell\geqslant 1}\pi^{\mathrm{SBM}}\big(B_\ell\times(\mathbb{R}^d\setminus\operatorname{dom}\psi_{\mathrm{opt}})\big)>0.$$

Indeed, otherwise ν would be concentrated on $\operatorname{dom}\psi_{\mathrm{opt}}$, contradicting the assumption that $\dim(\operatorname{dom}\psi_{\mathrm{opt}})<\dim(\operatorname{supp}\nu)$. As a consequence,

$$\int_{B_\ell\times\mathbb{R}^d}\psi_{\mathrm{opt}}(y)\,\pi^{\mathrm{SBM}}(dx,dy)=+\infty$$

for all $\ell\geqslant 1$ large enough. We now show that ψ_{opt} could not have been optimal, by establishing that

$$\mathcal{D}(\psi_{\mathrm{opt}})=\int\left(\sup_{p\in\mathcal{P}_2^x(\mathbb{R}^d)}\left(\mathrm{MCov}(p,\gamma)-\int\psi_{\mathrm{opt}}\,dp\right)+\int\psi_{\mathrm{opt}}(y)\,\pi^{\mathrm{SBM}}_x(dy)\right)\mu(dx)=+\infty.$$

To wit, selecting $p_x=\pi^{\mathrm{SBM}}_x$ for $x\in\mathbb{R}^d\setminus B_\ell$ and $p_x=\delta_x$ for $x\in B_\ell$ we find

$$\begin{aligned}
\mathcal{D}(\psi_{\mathrm{opt}})&\geqslant\int_{\mathbb{R}^d\setminus B_\ell}\mathrm{MCov}(\pi^{\mathrm{SBM}}_x,\gamma)\,\mu(dx)-\int_{B_\ell}\left(\psi_{\mathrm{opt}}(x)-\int\psi_{\mathrm{opt}}(y)\,\pi^{\mathrm{SBM}}_x(dy)\right)\mu(dx)\\
&\geqslant c-\ell+\int_{B_\ell\times\mathbb{R}^d}\psi_{\mathrm{opt}}(y)\,\pi^{\mathrm{SBM}}(dx,dy),
\end{aligned}$$

for a finite constant c. We conclude by taking $\ell\geqslant 1$ large enough. □

Proof of Lemma 6.4. We first prove the identity (i). We denote $F \coloneqq \nabla f$, which is well defined Lebesgue-a.e. For $y, \eta \in \mathbb{R}^d$ we have to compute

$$\frac{(f * \gamma)(y + h\eta) - (f * \gamma)(y)}{h} = \int \frac{f(y + h\eta + z) - f(y + z)}{h}\, d\gamma(z),$$

as $h \to 0$. For all y, η and for γ-a.e. $z \in \mathbb{R}^d$, we have

$$\lim_{h \to 0} \frac{f(y + h\eta + z) - f(y + z)}{h} = \langle F(y + z), \eta \rangle,$$

so we only need to justify the exchange of limit and integral. To this end, it suffices to show the uniform integrability, with respect to the reference measure γ, of the family

$$\mathscr{U} \coloneqq \left\{ \frac{f(y + h\eta + \cdot) - f(y + \cdot)}{h} : 0 \leqslant h \leqslant 1 \right\}. \tag{B.8}$$

Using twice the above-tangent characterization of convexity, we obtain

$$\langle F(y + z), \eta \rangle \leqslant \frac{f(y + h\eta + z) - f(y + z)}{h} \leqslant \langle F(y + h\eta + z), \eta \rangle. \tag{B.9}$$

We take $r \in (1, 2)$, $p \coloneqq 2/r$, and $q \coloneqq p/(p - 1)$ its Hölder conjugate, and compute

$$\begin{aligned}
\int |F(y + h\eta + z)|^r \, d\gamma(z) &= \int |F(\zeta + z)|^r \, d\gamma_{y + h\eta - \zeta}(z) \\
&= \int |F(\zeta + z)|^r \, \exp\big(- |y - \zeta + h\eta|^2/2 + \langle z, y - \zeta + h\eta \rangle \big)\, d\gamma(z) \\
&\leqslant \sqrt[p]{\|F\|^2_{L^2(\gamma_\zeta; \mathbb{R}^d)}} \sqrt[q]{\int \exp\big(- q|y - \zeta + h\eta|^2/2 + q\langle z, y - \zeta + h\eta \rangle \big)\, d\gamma(z)} \\
&\leqslant \sqrt[p]{\|F\|^2_{L^2(\gamma_\zeta; \mathbb{R}^d)}} \sqrt[q]{\int \exp\big(q\langle z, y - \zeta + h\eta \rangle \big)\, d\gamma(z)} \\
&= \sqrt[p]{\|F\|^2_{L^2(\gamma_\zeta; \mathbb{R}^d)}} \sqrt[q]{\exp\big(q^2 |y - \zeta + h\eta|^2/2 \big)}.
\end{aligned} \tag{B.10}$$

We see that the first factor in (B.10) is finite (and independent of h) by assumption, while the second factor is uniformly bounded for $0 \leqslant h \leqslant 1$. Hence both the upper and lower bounds in (B.9) are uniformly integrable for $0 \leqslant h \leqslant 1$, implying the same property for the family $\mathscr{U}$ of (B.8).

We now turn to (ii) and (iii). The arguments given at the end of the proof of Proposition 5.2 show that, under the assumptions made here (but with f instead of ψ^*), the function $f * \gamma$ is strictly convex. We already know that this function is differentiable, and that the interior of its domain is $\mathbb{R}^d$. Hence by [54, Theorem 26.5] we deduce the stated properties. The only thing that merits an explanation is the inclusion $\operatorname{int}(\operatorname{dom}(f * \gamma)^*) \supseteq \operatorname{int}(\operatorname{dom} f^*)$. To this end, observe that

$$\begin{aligned}
(f * \gamma)^*(x) &= \sup_{\zeta \in \mathbb{R}^d} \left\{ \langle \zeta, x \rangle - \int f(\zeta + z)\, d\gamma(z) \right\} \leqslant \int \sup_{\zeta \in \mathbb{R}^d} \{ \langle \zeta, x \rangle - f(\zeta + z) \}\, d\gamma(z) \\
&= \int \sup_{\zeta \in \mathbb{R}^d} \big\{ \langle \zeta - z, x \rangle - f(\zeta) \big\}\, d\gamma(z) = \int \big(f^*(x) - \langle z, x \rangle \big)\, d\gamma(z) = f^*(x),
\end{aligned}$$

so $\operatorname{dom}(f * \gamma)^* \supseteq \operatorname{dom} f^*$. □

Proof of Lemma 6.5. The first equality in (6.3) follows from

$$\operatorname{int}(\operatorname{dom}\psi) \subseteq \operatorname{dom}\partial\psi \subseteq \operatorname{dom}\psi$$

and $\operatorname{dom}\partial\psi = (\partial\psi^*)(\mathbb{R}^d)$. We also note that

$$\operatorname{int}\big((\partial\psi^*)(\mathbb{R}^d)\big) = \operatorname{int}\operatorname{conv}\big((\partial\psi^*)(\mathbb{R}^d)\big), \tag{B.11}$$

where conv denotes the convex hull. Indeed, as a consequence of a result by Mirsky [50], we have that $\operatorname{int}\operatorname{conv}((\partial\psi^*)(\mathbb{R}^d)) \subseteq (\partial\psi^*)(\mathbb{R}^d)$, since the subdifferential (multi-valued) mapping of any lower semicontinuous proper convex function is also a maximal monotone mapping by [54, Corollary 31.5.2].

Let us turn to the proof of the second equality in (6.3). Without loss of generality we can assume that $\eta = 0$ and $t = 1$. Since we already verified the first equality and by virtue of (B.11), it suffices to show the inclusions

$$(\nabla\psi^* * \gamma)(\mathbb{R}^d) \subseteq \operatorname{int}\operatorname{conv}((\partial\psi^*)(\mathbb{R}^d)) \tag{B.12}$$

and

$$\operatorname{int}(\operatorname{dom}\psi) \subseteq (\nabla\psi^* * \gamma)(\mathbb{R}^d). \tag{B.13}$$

Note that we have $(\nabla\psi^* * \gamma)(\mathbb{R}^d) \subseteq \overline{\operatorname{conv}}((\partial\psi^*)(\mathbb{R}^d))$ and therefore

$$\operatorname{int}(\nabla\psi^* * \gamma)(\mathbb{R}^d) \subseteq \operatorname{int}\overline{\operatorname{conv}}((\partial\psi^*)(\mathbb{R}^d)) = \operatorname{int}\operatorname{conv}((\partial\psi^*)(\mathbb{R}^d)). \tag{B.14}$$

Under the assumptions of Lemma 6.5, the function ψ is co-finite by Lemma 5.1 and therefore we can apply Lemma 6.4 to its convex conjugate $f = \psi^*$. In particular, by (iii), the set $(\nabla\psi^* * \gamma)(\mathbb{R}^d)$ is open and hence the inclusion (B.12) follows from (B.14).

In order to prove (B.13), we first observe that $\operatorname{int}(\operatorname{dom}\psi) \subseteq \operatorname{int}(\operatorname{dom}(\psi^* * \gamma)^*)$ by Lemma 6.4, (ii). Since $\operatorname{int}(\operatorname{dom}(\psi^* * \gamma)^*) = (\nabla\psi^* * \gamma)(\mathbb{R}^d)$ by Lemma 6.4, (iii), the inclusion (B.13) follows. □

APPENDIX C: PROOF OF PROPOSITION 7.20

Lemma C.1. *For every sequence $(\psi_n)_{n\geqslant 1}$ of lower semicontinuous convex functions $\psi_n\colon \mathbb{R}^d \to [0,+\infty)$ there is a closed convex set $C_{\lim} \subseteq \mathbb{R}^d$ with relative interior $I_{\lim}$ and a subsequence $(\psi_{n_k})_{k\geqslant 1}$ such that the limits*

$$\psi_{\lim}(y) := \lim_{k\to\infty} \psi_{n_k}(y) < +\infty, \qquad y \in I_{\lim} \tag{C.1}$$

$$\psi_{\lim}(y) := \lim_{k\to\infty} \psi_{n_k}(y) = +\infty, \qquad y \in \mathbb{R}^d \setminus C_{\lim} \tag{C.2}$$

exist. Moreover, the convergence in (C.1) is uniform on compact subsets of $I_{\lim}$.

Proof. *Step 1.* We denote by $\mathcal{N}$ the collection of increasing subsequences $\mathcal{N} = (n_k)_{k\geqslant 1}$ of $\mathbb{N}$. For two subsequences $\mathcal{N}, \mathcal{N}' \in \mathcal{N}$ we call $\mathcal{N}'$ *finer* than $\mathcal{N}$, denoted by $\mathcal{N}' \succcurlyeq \mathcal{N}$, if $\mathcal{N}' \subseteq \mathcal{N}$, up to finitely many elements.

We write $\mathcal{I}$ for the collection of relatively open convex sets $I \subseteq \mathbb{R}^d$. For two pairs $(I,\mathcal{N}), (I',\mathcal{N}') \in \mathcal{I} \times \mathcal{N}$ we say that $(I',\mathcal{N}')$ is *finer* than $(I,\mathcal{N})$, again denoted by $(I',\mathcal{N}') \succcurlyeq (I,\mathcal{N})$, if $I' \supsetneq I$ and $\mathcal{N}' \succcurlyeq \mathcal{N}$.

We call a pair $(I,\mathcal{N}) \in \mathcal{I} \times \mathcal{N}$ *admissible* if it satisfies

$$\forall y \in I\colon \quad \sup_{n_k\in\mathcal{N}} \psi_{n_k}(y) < +\infty. \tag{C.3}$$

Step 2. Let $(I, \mathcal{N})$ be admissible. Write $\mathcal{N} = (n_k)_{k\geqslant 1} \in \mathcal{N}$ and denote by C the closure of I. Then one of the following two alternatives hold: either

$$\lim_{k\to\infty} \psi_{n_k}(y) = +\infty, \qquad y \in \mathbb{R}^d \setminus C, \tag{C.4}$$

in which case we call $(I, \mathcal{N})$ *maximal*; or there is some $y_0 \in \mathbb{R}^d \setminus C$ and a subsequence $\mathcal{N}' = (n'_k)_{k\geqslant 1} \in \mathcal{N}$ finer than $\mathcal{N}$ such that

$$\sup_{n'_k \in \mathcal{N}'} \psi_{n'_k}(y_0) < +\infty.$$

Now we define the relatively open convex set $I' := \operatorname{ri}\operatorname{conv}(y_0, I)$ and note that $I' \supsetneq I$. Then the pair $(I', \mathcal{N}')$ is admissible and finer than $(I, \mathcal{N})$. In this second case we say that $(I, \mathcal{N})$ is *refinable*.

Step 3. Assume that $(I, \mathcal{N})$ is maximal, so that both (C.3) and (C.4) are satisfied. In particular, the sequence $(\psi_{n_k})_{k\geqslant 1}$, with $\mathcal{N} = (n_k)_{k\geqslant 1}$, is pointwise bounded on I. By [54, Theorem 10.9] we can select a further subsequence, still denoted by $(\psi_{n_k})_{k\geqslant 1}$, such that (C.1) holds for all $y \in I_{\lim} := I$ and the convergence is uniform on compact subsets of $I_{\lim}$. Furthermore, condition (C.2) is still satisfied for this further subsequence thanks to (C.4).

Step 4. In light of Step 3, in order to prove Lemma C.1, it suffices to show the existence of a maximal pair $(I, \mathcal{N})$, which we will denote by $(I_{\lim}, \mathcal{N}_{\lim})$. The construction of the maximal pair $(I_{\lim}, \mathcal{N}_{\lim})$ is done via transfinite recursion along the countable ordinal numbers $\alpha < \omega_1$, where ω_1 is the first uncountable ordinal.

We start the construction as follows. If $\sup_{n\in\mathbb{N}} \psi_n(y) = +\infty$ holds for all $y \in \mathbb{R}^d$, the pair $(I_0, \mathcal{N}_0) = (\varnothing, \mathbb{N})$ is admissible. Otherwise, we can find some $y_0 \in \mathbb{R}^d$ such that $\sup_{n\in\mathbb{N}} \psi_n(y_0) < +\infty$ and then the pair $(I_0, \mathcal{N}_0) = (\{y_0\}, \mathbb{N})$ is admissible.

Now let $\alpha < \omega_1$ be an ordinal number and suppose that the transfinite family $(I_\beta, \mathcal{N}_\beta)_{\beta<\alpha}$ in $\mathcal{I} \times \mathcal{N}$ is such that

(i) the pair $(I_\beta, \mathcal{N}_\beta)$ is admissible, for all $\beta < \alpha$,
(ii) the pair $(I_{\beta_2}, \mathcal{N}_{\beta_2})$ is finer than $(I_{\beta_1}, \mathcal{N}_{\beta_1})$, for all $\beta_1 < \beta_2 < \alpha$.

We first assume that α is a successor ordinal, i.e., there is an ordinal β such that $\alpha = \beta + 1$. Since $(I_\beta, \mathcal{N}_\beta)$ is admissible, by Step 2 only two cases are possible. If $(I_\beta, \mathcal{N}_\beta)$ is maximal, we set $(I_{\lim}, \mathcal{N}_{\lim}) := (I_\beta, \mathcal{N}_\beta)$ and finish the construction. If $(I_\beta, \mathcal{N}_\beta)$ is refinable, we can find a pair $(I_\alpha, \mathcal{N}_\alpha)$ which is finer than $(I_\beta, \mathcal{N}_\beta)$.

Suppose that α is a limit ordinal, i.e., it can be written as $\alpha = \{\beta : \beta < \alpha\}$. Since $\alpha < \omega_1$, the ordinal α is countable and can be enumerated as $\alpha = \{\alpha_k : k < \omega\}$, where ω is the first infinite ordinal. Letting $\beta_0 := 0$ and defining inductively $\beta_{m+1} := \alpha_\ell$, where $\ell := \min\{k : \beta_m < \alpha_k\}$, we obtain a strictly increasing cofinal sequence $(\beta_m)_{m<\omega}$ in α. Now we define $I_\alpha := \bigcup_{m<\omega} I_{\beta_m}$. By (ii), $(I_{\beta_m})_{m<\omega}$ is a strictly increasing sequence of relatively open convex sets, therefore also $I_\alpha \in \mathcal{I}$. Again by (ii), the sequence $(\mathcal{N}_{\beta_m})_{m<\omega}$ satisfies $\mathcal{N}_{\beta_{m_1}} \succcurlyeq \mathcal{N}_{\beta_{m_2}}$, for all $m_1 < m_2 < \omega$. Thus we can find a diagonal subsequence $\mathcal{N}_\alpha \in \mathcal{N}$ of $(\mathcal{N}_{\beta_m})_{m<\omega}$, meaning that $\mathcal{N}_\alpha \succcurlyeq \mathcal{N}_{\beta_m}$, for all $m < \omega$. Since by (i) the pair $(I_{\beta_m}, \mathcal{N}_{\beta_m})$ is admissible, for all $m < \omega$, by construction also $(I_\alpha, \mathcal{N}_\alpha)$ is admissible. Furthermore, $(I_\alpha, \mathcal{N}_\alpha)$ satisfies $(I_\alpha, \mathcal{N}_\alpha) \succcurlyeq (I_{\beta_m}, \mathcal{N}_{\beta_m})$, for all $m < \omega$. Since $(\beta_m)_{m<\omega}$ is cofinal, it follows that $(I_\alpha, \mathcal{N}_\alpha) \succcurlyeq (I_\beta, \mathcal{N}_\beta)$, for all $\beta < \alpha$.

After countably many steps, this transfinite recursion must stop, which happens when $\alpha < \omega_1$ is such that $(I_\alpha, \mathcal{N}_\alpha)$ is maximal and we can then set $(I_{\lim}, \mathcal{N}_{\lim}) := (I_\alpha, \mathcal{N}_\alpha)$. Indeed, otherwise we would have a transfinite sequence $(I_\beta)_{\beta<\omega_1}$ of strictly increasing relatively open convex sets, indexed by the countable ordinals, which is impossible. □

Proof of Proposition 7.20:
Step 1. By Lemma C.1, there is a closed convex set $C_{\lim} \subseteq \mathbb{R}^d$ with relative interior $I_{\lim}$ and a subsequence $(\psi_{n_k})_{k\geqslant 1}$ such that the limits

$$\psi_{\lim}(y) := \lim_{k\to\infty} \psi_{n_k}(y) < +\infty, \qquad y \in I_{\lim} \tag{C.5}$$

$$\psi_{\lim}(y) := \lim_{k\to\infty} \psi_{n_k}(y) = +\infty, \qquad y \in \mathbb{R}^d \setminus C_{\lim} \tag{C.6}$$

exist, where the convergence in (C.5) is uniform on compact subsets of $I_{\lim}$. Via (C.5) and (C.6) we define a lower semicontinuous convex function $\psi_{\lim}\colon \mathbb{R}^d \to [0,+\infty]$ on all of $\mathbb{R}^d$.

Step 2. By the hypothesis (7.25) and the construction of $I_{\lim}$, we clearly have $I \subseteq I_{\lim}$. Our goal is to show that $I_{\lim} = I$. To this end, we define the set $A \subseteq \mathbb{R}^d$ as in (7.6), now with I replaced by $I_{\lim}$, i.e.,

$$A = \big\{x \in I_{\lim}\colon \psi_{\lim} \not\equiv \psi^x_{\mathrm{SBM}} \bmod (\mathrm{aff})\big\}.$$

Again, we will show that $\mu(A) = 0$. Repeating the reasoning of Lemma 7.16 we obtain, for μ-a.e. $x \in A$, measures $\check{\pi}_x \in \mathcal{P}^x_2(\mathbb{R}^d)$ satisfying (7.7), which now are supported by $C_{\lim}$. Then we choose an increasing sequence $(K_j)_{j\geqslant 1}$ of relatively compact subsets of $I_{\lim}$ such that $\bigcup_{j\geqslant 1} K_j = I_{\lim}$. As in Lemma 7.17, for μ-a.e. $x \in A$, we find measures $\check{\pi}^{j(x)}_x \in \mathcal{P}^x_2(\mathbb{R}^d)$ satisfying (7.13) and supported by $K_{j(x)}$, for some $j(x) \in \mathbb{N}$. Arguing as in Lemma 7.18 we conclude that $\mu(A) = 0$, i.e., $\psi_{\lim} \equiv \psi^x_{\mathrm{SBM}} \bmod (\mathrm{aff})$, for μ-a.e. $x \in I_{\lim}$. In particular, since $I_{\lim} = \mathrm{ri}(\mathrm{dom}\,\psi_{\lim})$ and $\mathrm{ri}(\mathrm{dom}\,\psi^x_{\mathrm{SBM}}) = \mathrm{ri}\,\widehat{\mathrm{supp}}(\nu)$, we conclude from the assumption $\widehat{\mathrm{supp}}(\nu) \subseteq C$ that $I_{\lim} = I$ and consequently $C = \widehat{\mathrm{supp}}(\nu)$. From Lemma 7.19, we conclude that the limiting function $\psi_{\lim}$ is a dual optimizer. Together with Step 1, this proves the statement of Proposition 7.20 for the subsequence $(\psi_{n_k})_{k\geqslant 1}$.

Step 3. Applying Step 1 and Step 2 to an arbitrary subsequence $(\psi_{m_\ell})_{\ell\geqslant 1}$ instead of the original sequence $(\psi_n)_{n\geqslant 1}$, we obtain a further subsequence $(\psi_{m_{\ell_j}})_{j\geqslant 1}$ and a lower semicontinuous convex function $\tilde{\psi}_{\lim}\colon \mathbb{R}^d \to [0,+\infty]$ such that

$$\forall y \in I\colon\ \tilde{\psi}_{\lim}(y) = \lim_{j\to\infty} \psi_{m_{\ell_j}}(y) < +\infty,$$

$$\forall y \in \mathbb{R}^d \setminus C\colon\ \tilde{\psi}_{\lim}(y) = \lim_{j\to\infty} \psi_{m_{\ell_j}}(y) = +\infty,$$

and $\tilde{\psi}_{\lim} \equiv \psi^x_{\mathrm{SBM}} \bmod (\mathrm{aff})$, for μ-a.e. $x \in I$. We conclude that there is an affine function aff depending on the subsequence $(m_{\ell_j})_{j\geqslant 1}$ such that $\psi_{\lim} = \tilde{\psi}_{\lim} + \mathrm{aff}$. In particular,

$$\forall y \in I \cup (\mathbb{R}^d \setminus C)\colon\ \lim_{j\to\infty} \big(\psi_{m_{\ell_j}}(y) + \mathrm{aff}(y)\big) = \psi_{\lim}(y). \tag{C.7}$$

Step 4. Let $x_0, \dots, x_d$ be affinely independent points in $I \cup (\mathbb{R}^d \setminus C)$. For each $n \geqslant 1$ there is a unique affine function aff_n such that

$$\forall i \in \{0,\dots,d\}\colon\ \psi_n(x_i) + \mathrm{aff}_n(x_i) = \psi_{\lim}(x_i). \tag{C.8}$$

We claim that

$$\forall y \in I \cup (\mathbb{R}^d \setminus C)\colon\ \lim_{n\to\infty} \big(\psi_n(y) + \mathrm{aff}_n(y)\big) = \psi_{\lim}(y). \tag{C.9}$$

Indeed, if this were not the case, there would be some $y_0 \in I \cup (\mathbb{R}^d \setminus C)$ and a subsequence $(m_\ell)_{\ell\geqslant 1}$ such that

$$\lim_{\ell\to\infty} \big(\psi_{m_\ell}(y_0) + \mathrm{aff}_{m_\ell}(y_0)\big) \neq \psi_{\lim}(y_0). \tag{C.10}$$

By Step 3 there is a further subsequence $(m_{\ell_j})_{j\geqslant 1}$ and an affine function aff such that (C.7) holds. In particular,

$$\forall i \in \{0,\dots,d\}\colon \quad \lim_{j\to\infty} \big(\psi_{m_{\ell_j}}(x_i) + \mathrm{aff}(x_i)\big) = \psi_{\lim}(x_i), \tag{C.11}$$

$$\lim_{j\to\infty} \big(\psi_{m_{\ell_j}}(y_0) + \mathrm{aff}(y_0)\big) = \psi_{\lim}(y_0). \tag{C.12}$$

From (C.8) and (C.11) we obtain that

$$\forall i \in \{0,\dots,d\}\colon \quad \lim_{j\to\infty} \mathrm{aff}_{m_{\ell_j}}(x_i) = \mathrm{aff}(x_i). \tag{C.13}$$

Since the points $x_0,\dots,x_d$ are affinely independent, the convergence in (C.13) holds for every $y \in I \cup (\mathbb{R}^d \setminus C)$. Then (C.12) leads to a contradiction to (C.10), which proves claim (C.9). Defining $\tilde{\psi}_n := \psi_n + \mathrm{aff}_n$, for $n \geqslant 1$, we derive from (C.9) the limiting assertions (7.26) and (7.27). Finally, by [54, Theorem 10.8] the convergence in (7.26) is uniform on compact subsets of I and we have already seen in Step 2 that $\psi_{\lim} \equiv \psi^x_{\mathrm{SBM}}$ mod (aff), for μ-a.e. $x \in I$. □

APPENDIX D: CHARACTERIZATION OF IRREDUCIBILITY

THEOREM D.1. *Let μ,ν in $\mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. Then the following are equivalent.*

(1) *The pair (μ,ν) is De March–Touzi irreducible in the sense of Definition 7.2.*
(2) *There exists $\pi \in \mathsf{MT}(\mu,\nu)$ such that $\pi_x \sim \nu$, for μ-a.e. $x \in \mathbb{R}^d$.*
(3) *The pair (μ,ν) is irreducible in the sense of Definition 1.2, i.e., for all Borel sets $A,B \subseteq \mathbb{R}^d$ with $\mu(A),\nu(B) > 0$ there is a martingale $(X_t)_{0\leqslant t\leqslant 1}$ with $X_0 \sim \mu$, $X_1 \sim \nu$ such that $\mathbb{P}(X_0 \in A, X_1 \in B) > 0$.*
(4) *For all Borel sets $A,B \subseteq \mathbb{R}^d$ with $\mu(A),\nu(B) > 0$ there exists $\pi \in \mathsf{MT}(\mu,\nu)$ such that $\pi(A\times B) > 0$.*
(5) *For all compact sets $A \subseteq \mathbb{R}^d$ and open halfspaces H with $\mu(A),\nu(H) > 0$ there exists $\pi \in \mathsf{MT}(\mu,\nu)$ such that $\pi(A\times H) > 0$.*

PROOF. For the proof of "(1) ⇒ (2)" we can take $\pi = \pi^{\mathrm{SBM}}$ in (2) by Corollary 7.7. We turn to the proof of the implication "(2) ⇒ (3)": Clearly condition (2) implies (4). Thus, in order to show (3), it suffices to construct a continuous-time martingale $(X_t)_{0\leqslant t\leqslant 1}$ with $\mathrm{Law}(X_0,X_1) = \pi$. This can be achieved as in the proof of [5, Theorem 2.2]. For the proof of "(3) ⇒ (4)" we can take $\pi = \mathrm{Law}(X_0,X_1)$ in (4). The implication "(4) ⇒ (5)" is trivial. It remains to show that "(5) ⇒ (1)": We fix an open halfspace H satisfying $\nu(H) > 0$ and set

$$m := \sup_{\pi\in\mathsf{MT}(\mu,\nu)} \Big\{\mu\big(\{x \in \mathbb{R}^d\colon \pi_x(H) > 0\}\big)\Big\}.$$

Considering countable convex combinations of elements of $\mathsf{MT}(\mu,\nu)$ it follows that the supremum is attained by some $\bar{\pi} \in \mathsf{MT}(\mu,\nu)$. If the set

$$\{x \in \mathbb{R}^d\colon \bar{\pi}_x(H) = 0\}$$

had positive μ-measure, it would contain a compact set A with positive μ-measure in contradiction to condition (5). Hence $m = 1$.

Observe that there is a countable family of open halfspaces $\{H_n\}_{n\geqslant 1}$ with $\nu(H_n) > 0$ such that, for every open halfspace H with $\nu(H) > 0$, there is some $n \geqslant 1$ satisfying $H_n \subseteq H$.

Next, for each $n \geqslant 1$, we pick $\pi^{(n)} \in \mathsf{MT}(\mu,\nu)$ such that $\pi^{(n)}_x(H_n) > 0$, for μ-a.e. $x \in \mathbb{R}^d$. Set $\hat{\pi} := \sum_{n\geqslant 1} 2^{-n}\pi^{(n)} \in \mathsf{MT}(\mu,\nu)$. Then by Lemma D.2 below it follows that $\widetilde{\mathrm{supp}}(\nu) \subseteq \widetilde{\mathrm{supp}}(\hat{\pi}_x)$, for μ-a.e. $x \in \mathbb{R}^d$. □

LEMMA D.2. *Let $\rho \in \mathcal{P}(\mathbb{R}^d)$. Then $y \in \widehat{\text{supp}}(\rho)$ if and only if $\rho(H) > 0$ for every open halfspace H such that $y \in H$.*

PROOF. This is a simple consequence of Hahn–Banach. □

REMARK D.3. We note that irreducibility is not only a sufficient assumption for the existence of a Bass martingale from μ to ν, but in fact also necessary. Indeed, as in the proof of Corollary 7.7 one sees that the Bass martingale does the job of connecting any two sets which are charged by μ and ν, implying the irreducibility of (μ, ν) by Theorem D.1.

Acknowledgments. The authors thank Ben Robinson and Gudmund Pammer for their valuable feedback and comments during the preparation of this paper.

Funding. This research was funded by the Austrian Science Fund (FWF) [Grant DOIs: 10.55776/P35197 and 10.55776/P35519]. For open access purposes, the authors have applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission.

REFERENCES

[1] ACCIAIO, B., MARINI, A. and PAMMER, G. (2023). Calibration of the Bass Local Volatility model. arXiv:2311.14567.

[2] ACCIAIO, B. and PAMMER, G. (2022). A short proof of the characterisation of convex order using the 2-Wasserstein distance. arXiv:2207.02096.

[3] AMBROSIO, L. and GIGLI, N. (2013). A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*. *Lecture Notes in Math.* **2062** 1–155. Springer, Heidelberg. https://doi.org/10.1007/978-3-642-32160-3_1 MR3050280

[4] BACKHOFF-VERAGUAS, J., BARTL, D., BEIGLBÖCK, M. and EDER, M. (2020). Adapted Wasserstein distances and stability in mathematical finance. *Finance Stoch.* **24** 601–632. https://doi.org/10.1007/s00780-020-00426-3 MR4118990

[5] BACKHOFF-VERAGUAS, J., BEIGLBÖCK, M., HUESMANN, M. and KÄLLBLAD, S. (2020). Martingale Benamou-Brenier: A probabilistic perspective. *Ann. Probab.* **48** 2258–2289. https://doi.org/10.1214/20-AOP1422 MR4152642

[6] BACKHOFF-VERAGUAS, J., BEIGLBÖCK, M. and PAMMER, G. (2019). Existence, duality, and cyclical monotonicity for weak transport costs. *Calculus of Variations and Partial Differential Equations* **58** 1–28. https://doi.org/doi:10.1137/18M1196479

[7] BASS, R. (1983). Skorokhod embedding via stochastic integrals. In *Séminaire de Probabilités XVII 1981/82*, (J. Azéma and M. Yor, eds.). *Lecture Notes in Mathematics* **986** 221–224. Springer.

[8] BAUSCHKE, H. H. and COMBETTES, P. L. (2017). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, second ed. *CMS Books in Mathematics*. Springer.

[9] BEIGLBÖCK, M., COX, A. and HUESMANN, M. (2017). Optimal Transport and Skorokhod Embedding. *Invent. Math.* **208** 327–400.

[10] BEIGLBÖCK, M., HENRY-LABORDÈRE, P. and PENKNER, F. (2013). Model-independent Bounds for Option Prices: A Mass Transport Approach. *Finance Stoch.* **17** 477–501.

[11] BEIGLBÖCK, M., HENRY-LABORDÈRE, P. and TOUZI, N. (2017). Monotone martingale transport plans and Skorokhod embedding. *Stochastic Process. Appl.* **127** 3005–3013. https://doi.org/10.1016/j.spa.2017.01.004 MR3682121

[12] BEIGLBÖCK, M. and JUILLET, N. (2016). On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.* **44** 42–106. https://doi.org/10.1214/14-AOP966 MR3456332

[13] BEIGLBÖCK, M., LIM, T. and OBŁÓJ, J. (2019). Dual attainment for the martingale transport problem. *Bernoulli* **25** 1640–1658. https://doi.org/10.3150/17-BEJ1015

[14] BEIGLBÖCK, M., NUTZ, M. and STEBEGG, F. (2022). Fine properties of the optimal Skorokhod embedding problem. *J. Eur. Math. Soc. (JEMS)* **24** 1389–1429. https://doi.org/10.4171/JEMS/1122 MR4397044

[15] BEIGLBÖCK, M., NUTZ, M. and TOUZI, N. (2017). Complete duality for martingale optimal transport on the line. *Ann. Probab.* **45** 3038–3074. https://doi.org/10.1214/16-AOP1131 MR3706738

[16] BEIGLBÖCK, M., PAMMER, G., RIESS, L. and SCHROTT, S. (2025). The Fundamental Theorem of Weak Optimal Transport. arXiv:2501.16316.

[17] Benamou, J. and Brenier, Y. (1999). 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). Contemp. Math.* **226** 1–11. Amer. Math. Soc., Providence, RI. MR1660739 (99j:65151)
[18] Bouchard, B. and Nutz, M. (2015). Arbitrage and duality in nondominated discrete-time models. *Ann. Appl. Probab.* **25** 823–859.
[19] Brenier, Y. (1987). Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.* **305** 805–808. MR923203 (89b:58226)
[20] Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Commun. Pure Appl. Math.* **44** 375–417. https://doi.org/10.1002/cpa.3160440402 MR1100809
[21] Campi, L., Laachir, I. and Martini, C. (2017). Change of numeraire in the two-marginals martingale transport problem. *Finance Stoch.* **21** 471–486.
[22] Cheridito, P., Kiiski, M., Prömel, D. J. and Soner, H. M. (2021). Martingale optimal transport duality. *Math. Ann.* **379** 1685–1712. https://doi.org/10.1007/s00208-019-01952-y MR4238277
[23] Ciosmak, K. (2024). Localisation for constrained transports I: theory. arXiv:2312.12281.
[24] Ciosmak, K. (2024). Localisation for constrained transports II: applications. arXiv:2312.13167.
[25] Conze, A. and Henry-Labordere, P. (2021). Bass Construction with Multi-Marginals: Lightspeed Computation in a New Local Volatility Model. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3853085
[26] Cox, A. M. G., Obłój, J. and Touzi, N. (2019). The Root solution to the multi-marginal embedding problem: an optimal stopping and time-reversal approach. *Probab. Theory Related Fields* **173** 211–259. https://doi.org/10.1007/s00440-018-0833-1 MR3916107
[27] De March, H. and Touzi, N. (2019). Irreducible convex paving for decomposition of multidimensional martingale transport plans. *Ann. Probab.* **47** 1726–1774. https://doi.org/10.1214/18-AOP1295 MR3945758
[28] Dolinsky, Y. and Soner, H. M. (2014). Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Relat. Fields* **160** 391–427. https://doi.org/10.1007/s00440-013-0531-y
[29] Föllmer, H. (2022). Optimal couplings on Wiener space and an extension of Talagrand's transport inequality. In *Stochastic analysis, filtering, and stochastic optimization* 147–175. Springer, Cham. https://doi.org/10.1007/978-3-030-98519-6_7 MR4433813
[30] Galichon, A., Henry-Labordère, P. and Touzi, N. (2014). A stochastic control approach to no-arbitrage bounds given marginals, with an application to Lookback options. *Ann. Appl. Probab.* **24** 312–336.
[31] Ghoussoub, N., Kim, Y.-H. and Lim, T. (2019). Structure of optimal martingale transport plans in general dimensions. *Ann. Probab.* **47** 109–164. https://doi.org/10.1214/18-AOP1258 MR3909967
[32] Ghoussoub, N., Kim, Y.-H. and Lim, T. (2020). Optimal Brownian stopping when the source and target are radially symmetric distributions. *SIAM J. Control Optim.* **58** 2765–2789. https://doi.org/10.1137/19M1270513 MR4149541
[33] Ghoussoub, N., Kim, Y.-H. and Palmer, A. Z. (2019). PDE methods for optimal Skorokhod embeddings. *Calc. Var. Partial Differential Equations* **58** Paper No. 113, 31. https://doi.org/10.1007/s00526-019-1563-7 MR3959935
[34] Gozlan, N. and Juillet, N. (2020). On a mixture of Brenier and Strassen theorems. *Proceedings of the London Mathematical Society* **120** 434–463.
[35] Gozlan, N., Roberto, C., Samson, P.-M. and Tetali, P. (2017). Kantorovich duality for general transport costs and applications. *J. Funct. Anal.* **273** 3327–3405. https://doi.org/10.1016/j.jfa.2017.08.015 MR3706606
[36] Guo, I. and Loeper, G. (2021). Path dependent optimal transport and model calibration on exotic derivatives. *The Annals of Applied Probability* **31** 1232 – 1263. https://doi.org/10.1214/20-AAP1617
[37] Guo, I., Loeper, G. and Wang, S. (2019). Local volatility calibration by optimal transport. In *2017 MATRIX Annals* 51–64. Springer.
[38] Henry-Labordère, P., Obłój, J., Spoida, P. and Touzi, N. (2016). The maximum maximum of a martingale with given n marginals. *Ann. Appl. Probab.* **26** 1–44. https://doi.org/10.1214/14-AAP1084 MR3449312
[39] Henry-Labordère, P. and Touzi, N. (2016). An explicit martingale version of the one-dimensional Brenier theorem. *Finance Stoch.* **20** 635–668. https://doi.org/10.1007/s00780-016-0299-x MR3519164
[40] Hobson, D. and Neuberger, A. (2012). Robust bounds for forward start options. *Math. Finance* **22** 31–56.
[41] Huesmann, M. and Trevisan, D. (2019). A Benamou-Brenier formulation of martingale optimal transport. *Bernoulli* **25** 2729–2757. https://doi.org/10.3150/18-BEJ1069 MR4003563
[42] Joseph, B., Loeper, G. and Obloj, J. (2024). The Measure Preserving Martingale Sinkhorn Algorithm. arXiv:2310.13797.

[43] Källblad, S., Tan, X. and Touzi, N. (2017). Optimal Skorokhod embedding given full marginals and Azéma-Yor peacocks. *Ann. Appl. Probab.* **27** 686–719. https://doi.org/10.1214/16-AAP1191 MR3655851

[44] Kantorovich, L. (1942). On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)* **37** 199–201. MR0009619 (5,174d)

[45] Komlós, J. (1967). A generalization of a problem of Steinhaus. *Acta Math. Acad. Sci. Hungar.* **18** 217–229. MR0210177 (35 ##1071)

[46] Loeper, G. (2018). Option pricing with linear market impact and nonlinear Black-Scholes equations. *Ann. Appl. Probab.* **28** 2664–2726. https://doi.org/10.1214/17-AAP1367 MR3847970

[47] McCann, R. (1994). A Convexity theory for interacting gases and equilibrum crystals. *PhD thesis, Princeton University*.

[48] McCann, R. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **80** 309–323. https://doi.org/10.1215/S0012-7094-95-08013-2 MR1369395 (97d:49045)

[49] McCann, R. J. (2001). Polar factorization of maps on Riemannian manifolds. *Geometric & Functional Analysis GAFA* **11** 589–608. https://doi.org/10.1007/PL00001679

[50] Mirsky, L. (1961). Majorization of vectors and inequalities for convex functions. *Monatsh. Math.* **65** 159–169. https://doi.org/10.1007/BF01307025 MR123661

[51] Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Histoire de l'académie Royale des Sciences de Paris*.

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

[53] Obłój, J., Spoida, P. and Touzi, N. (2015). Martingale inequalities for the maximum via pathwise arguments. In *In Memoriam Marc Yor-Séminaire de Probabilités XLVII* 227–247. Springer.

[54] Rockafellar, R. T. (1997). *Convex analysis*. *Princeton Landmarks in Mathematics*. Princeton University Press, Princeton, NJ Reprint of the 1970 original, Princeton Paperbacks. MR1451876

[55] Santambrogio, F. (2015). *Optimal Transport for Applied Mathematicians*. *Progress in Nonlinear Differential Equations and their Applications* **87**. Birkhäuser Cham Calculus of variations, PDEs, and modeling. https://doi.org/10.1007/978-3-319-20828-2 MR3409718

[56] Schachermayer, W. and Tschiderer, B. (2024). The decomposition of stretched Brownian motion into Bass martingales. arXiv:2406.10656.

[57] Tan, X. and Touzi, N. (2013). Optimal transportation under controlled stochastic dynamics. *Ann. Probab.* **41** 3201–3240. https://doi.org/10.1214/12-AOP797 MR3127880

[58] Villani, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. American Mathematical Society, Providence, RI. MR1964483 (2004e:90003)

[59] Villani, C. (2009). *Optimal Transport, Old and New*. *Grundlehren der mathematischen Wissenschaften* **338**. Springer.

[60] Wiesel, J. and Zhang, E. (2023). An optimal transport-based characterization of convex order. *Depend. Model.* **11** 20230102.