The Annals of Probability
2020, Vol. 48, No. 5, 2258–2289
https://doi.org/10.1214/20-AOP1422

MARTINGALE BENAMOU–BRENIER: A PROBABILISTIC PERSPECTIVE

BY JULIO BACKHOFF-VERAGUAS[1], MATHIAS BEIGLBÖCK[2], MARTIN HUESMANN[3] AND SIGRID KÄLLBLAD[4]

[1] *Department of Applied Mathematics, University of Twente, julio.backhoff@utwente.nl*
[2] *Department of Mathematics, University of Vienna, mathias.beiglboeck@univie.ac.at*
[3] *Institut für Mathematische Stochastik, Westälichen Wilhelms-Universität-Münster, martin.huesmann@uni-muenster.de*
[4] *Department of Mathematics, KTH Royal Institute of Technology, sigrid.kallblad@math.kth.se*

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.

We suggest a Benamou–Brenier type formulation of the martingale transport problem for given d-dimensional distributions μ, ν in convex order. The unique solution $M^* = (M_t^*)_{t\in[0,1]}$ of this problem turns out to be a Markov-martingale which has several notable properties: In a specific sense it mimics the movement of a Brownian particle as closely as possible subject to the conditions $M_0^* \sim \mu$, $M_1^* \sim \nu$. Similar to McCann's displacement-interpolation, M^* provides a time-consistent interpolation between μ and ν. For particular choices of the initial and terminal law, M^* recovers archetypical martingales such as Brownian motion, geometric Brownian motion, and the Bass martingale. Furthermore, it yields a natural approximation to the local vol model and a new approach to Kellerer's theorem.

This article is parallel to the work of Huesmann–Trevisan, who consider a related class of problems from a PDE-oriented perspective.

1. Introduction. The roots of optimal transport as a mathematical field go back to Monge [54] and Kantorovich [44] who established its modern formulation. Important triggers for its steep development in the last decades were the seminal results of Benamou, Brenier, and McCann [19, 22, 23, 53]. Today the field is famous for its striking applications in areas ranging from mathematical physics and PDE-theory to geometric and functional inequalities. We refer to [2, 60, 63, 64] for comprehensive accounts of the theory.

Recently, there has also been interest in optimal transport problems where the transport plan must satisfy additional martingale constraints. Such problems arise naturally in robust finance, but are also of independent mathematical interest, for example, they have important consequences for the study of martingale inequalities (see, e.g., [21, 38, 58]) and the Skorokhod embedding problem [10, 43]. Early papers to investigate such problems include [15, 24, 26, 28, 41, 62], and this topic is commonly referred to as martingale optimal transport.

In view of the central role taken by the seminal results of Benamou, Brenier, and McCann on optimal transport for squared Euclidean distance, the related continuous time transport problem and McCann's displacement interpolation, it is intriguing to search for similar concepts also in the martingale context. While [17, 39] propose a martingale version of Brenier's monotone transport mapping, our starting point is the Benamou–Brenier continuous time transport problem which we restate here for comparison with the martingale analogues that we will consider subsequently.

Received January 2019; revised September 2019.
MSC2020 subject classifications. Primary 60G42, 60G44; secondary 91G20.
Key words and phrases. Optimal transport, martingales, weak transport problems, Brenier's theorem, Benamou–Brenier, cyclical monotonicity, causal transport, Knothe Rosenblatt coupling, Schrödinger problem.

1.1. *Benamou–Brenier transport problem and McCann-interpolation in probabilistic terms*. In view of the probabilistic nature of the results, we present subsequently, it is convenient to recall some classical concepts and results of optimal transport in probabilistic language. Given probabilities μ, ν in the space $\mathcal{P}_2(\mathbb{R}^d)$ of d-dimensional distributions with finite second moment consider

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

Then by [22] we have:

THEOREM 1.1. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ and assume that μ is absolutely continuous with respect to Lebesgue. Then* (BB) *has a unique optimizer X^*.*

REMARK 1.2. In Theorem 1.1 (and similarly below), the solution to (BB) is unique in the sense that there exists a unique probability measure on the pathspace $C([0,1])$ such that the canonical/identity process optimizes (BB).

In probabilistic terms, McCann's displacement interpolation is defined by $[\mu,\nu]_t := law(X_t^*)$ where $t \in [0,1]$ and μ, ν, X^* as in Theorem 1.1.

THEOREM 1.3. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ and assume that μ is absolutely continuous with respect to Lebesgue measure. Let $s, t, \lambda \in [0,1]$, $s < t$. Then*

$$\big[[\mu,\nu]_s, [\mu,\nu]_t\big]_\lambda = [\mu,\nu]_{(1-\lambda)s+\lambda t}. \tag{1}$$

Moreover,

$$(t-s)T_2^{1/2}(\mu,\nu) = T_2^{1/2}\big([\mu,\nu]_s, [\mu,\nu]_t\big). \tag{2}$$

Finally, the optimizer of (BB) is given through the gradient of a convex function. More precisely, by [19], we have the following theorem.

THEOREM 1.4. *Assume that μ is absolutely continuous with respect to Lebesgue measure and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$. A candidate process X, $X_0 \sim \mu$, $X_1 \sim \nu$ is an optimizer if and only if $X_1 = f(X_0)$, where f is the gradient of a convex function $\phi : \mathbb{R}^n \to \mathbb{R}$ and all particles move with constant speed, that is, $X_t = tX_1 + (1-t)X_0 = X_0 + t(X_1 - X_0)$.*

1.2. *Martingale counterparts*. Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ be in convex order (denoted $\mu \preceq_c \nu$) and write B for Brownian motion on $\mathbb{R}^d$. We consider the optimization problem

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

see also (12) below. We have the following theorem.

THEOREM 1.5. *Assume that $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ satisfy $\mu \preceq_c \nu$. Then* (MBB) *has an optimizer M^* which is unique in law.*

At its face, the optimization problems (BB) and (MBB) look rather different. However, it is not hard to see that both problems are equivalent to optimization problems that are much more obviously related. In Section 6 below, we establish that

$$X^* = \operatorname*{argmin}_{X_0\sim\mu, X_1\sim\nu} W^2(X, \text{constant speed particle}), \tag{3}$$

$$M^* = \operatorname*{argmin}_{M_0\sim\mu, M_1\sim\nu} W_c^2(M, \text{constant volatility martingale}), \tag{4}$$

where W^2 denotes Wasserstein distance with respect to squared Cameron–Martin norm, while W_c^2 denotes an *adapted* or *causal* analogue[1] (in the terminology of Lassalle [49]), see Section 6 for details.

The reformulation in (4) allows for the following interpretation: M^* is the process whose evolution follows the movement of a Brownian particle as closely as possible subject to the marginal conditions $M_0 \sim \mu$, $M_1 \sim \nu$. This motivates the name in the following definition.

DEFINITION 1.6. Let μ, ν, M^* be as in Theorem 1.5. Then we call M^* the *stretched Brownian motion* (sBm) from μ to ν. We define the *martingale displacement interpolation* by

$$[\mu, \nu]_t^M := law\, M_t^*, \tag{5}$$

for $t \in [0, 1]$.

In analogy to Theorem 1.3, we have the following.

THEOREM 1.7. *Assume that $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ satisfy $\mu \preceq_c \nu$. Let $s, t, \lambda \in [0, 1]$, $s < t$. Then*

$$\big[[\mu, \nu]_s^M, [\mu, \nu]_t^M\big]_\lambda^M = [\mu, \nu]_{(1-\lambda)s+\lambda t}^M. \tag{6}$$

Moreover,

$$(t - s)MT^2(\mu, \nu) = MT^2\big([\mu, \nu]_s^M, [\mu, \nu]_t^M\big). \tag{7}$$

1.3. *Structure of stretched Brownian motion.* In the solution of the classical Benamou–Brenier transport problem, particles travel with constant speed along straight lines. In contrast, we will see that in the case of sBm the movement of individual particles mimic that of Brownian motion. Broadly speaking, the "direction" of these particles will be determined—similar to the classical case—by a mapping which is the gradient of a convex function.

For simplicity, we first consider the particular case where μ, ν, $\mu \preceq_c \nu$ are probabilities on the real line and μ is concentrated in a single point, that is, $\mu = \delta_m$ where m is the center of ν. It turns out that in this case sBm M^* is precisely the "Bass martingale" [8] (or "Brownian martingale") with terminal distribution ν. We briefly recall its construction: Pick $f : \mathbb{R} \to \mathbb{R}$ increasing such that $f(\gamma) = \nu$, where γ is the standard Gaussian distribution on $\mathbb{R}$. Then set for $t \in [0, 1]$

$$M_t := \mathbb{E}\big[f(B_1)|\mathcal{F}_t\big] = \mathbb{E}\big[f(B_1)|B_t\big] = f_t(B_t), \tag{8}$$

where $B = (B_t)_{t\in[0,1]}$ denotes Brownian motion started in $B_0 \sim \delta_0$, $(\mathcal{F}_t)_{t\in[0,1]}$ the Brownian filtration and $f_t(b) := \int f(b + y)\, d\gamma_{1-t}(y)$, $\gamma_s \sim N(0, s)$. Clearly M is a continuous Markov

[1] Causal transport plans generalize adapted processes in the same way as classical Kantorovich transport plans extend Monge maps.

martingale such that $M_0 \sim \delta_m$, $M_1 \sim \nu$. As a particular consequence of the results below we will see that M is a stretched Brownian motion.

To state our results for the general, multidimensional case we need to consider an extension of the Bass construction. Let $F : \mathbb{R}^d \to \mathbb{R}$ be a convex function and set

$$f_t(b) = \int \nabla F(b + y)\gamma^d_{1-t}(dy), \tag{9}$$

where γ^d_s denotes the centered d-dimensional Gaussian with covariance matrix $s\,\mathrm{Id}$. If B denotes d-dimensional Brownian motion started in $B_0 \sim \alpha$, we have

$$\mathbb{E}\big[\nabla F(B_1)|\mathcal{F}_t\big] = f_t(B_t), \quad t \in [0, 1]. \tag{10}$$

We denote by $*$ the convolution operator between measures. If f is a function and ρ a measure we write $f(\rho) := \rho \circ f^{-1}$ for the associated push-forward measure.

DEFINITION 1.8. A continuous $\mathbb{R}^d$-valued martingale M is a *standard stretched Brownian motion (s²Bm)* from μ to ν if there exist a probability measure α on $\mathbb{R}^d$ and a convex function $F : \mathbb{R}^d \to \mathbb{R}$ with $\nabla F(\alpha * \gamma^d) = \nu$, such that

$$M_t = E\big[\nabla F(B_1)|\mathcal{F}_t\big] \quad \text{and} \quad M_0 \sim \mu,$$

where B is a Brownian motion with $B_0 \sim \alpha$.

Note that, for $\alpha, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, there exists a convex function F with $\nabla F(\alpha * \gamma^d) = \nu$ and F is $\alpha * \gamma^d$-unique up to an additive constant. (This is a consequence of Brenier's theorem; see, e.g., Theorem 1.1 or [63], Theorem 2.12.)

REMARK 1.9. Both Brownian motion and geometric Brownian motion are examples of standard stretched Brownian motion.

We have the following results.

THEOREM 1.10. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$. If M is a standard stretched Brownian motion from μ to ν, then M is an optimizer of* (MBB), *that is, M is the stretched Brownian motion from μ to ν.*

THEOREM 1.11. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$. Let M^* be the stretched Brownian motion from μ to ν, that is, the optimizer of* (MBB). *Write $M^{*,x}$ for the martingale M conditioned on starting in $M_0 = x$. Then for μ-a.a. $x \in \mathbb{R}^d$ the martingale $M^{*,x}$ is a standard stretched Brownian motion.*

As a particular consequence of these results, the notions sBm and s²Bm coincide if μ is concentrated in a single point. However, the relation between sBm and s²Bm is more complicated in general: A notable intricacy of the martingale transport problem is caused by the fact that, loosely speaking, certain regions of the space do not communicate with each other.

Consider for a moment the particular case where μ, ν are distributions on the real line. In this instance, a martingale transport problem can be decomposed into countably many "minimal" components and on each of these components the behaviour of the problem is fairly similar to the classical transport problem. We refer the reader to Section 3.1 for the precise definition and only provide an illustrative example at this stage.

EXAMPLE 1.12. Let $\mu := 1/2(\lambda_{|[-3,-2]} + \lambda_{|[2,3]})$, $\nu := 1/6(\lambda_{|[-4,-1]} + \lambda_{|[1,4]})$. Then *any* martingale M, $M_0 \sim \mu$, $M_1 \sim \nu$ will satisfy the following: If $M_0 > 0$, then $M_1 > 0$ and if $M_0 \leq 0$ then $M_1 \leq 0$. That is, the positive and the negative halfline do not "communicate," and a problem of martingale transport should be considered on either of these parts of space separately.

If the pair (μ, ν) decomposes into more than one minimal component, as in the previous example, there exists no s^{2}Bm from μ to ν. However, for the one-dimensional case we will establish the following: A martingale is a sBm if and only if it behaves like a s^{2}Bm on each minimal component, see Theorem 3.1.

Notably, the challenges posed by noncommunicating regions appear much more intricate for dimension $d \geq 2$, see the deep contributions of Ghoussoub–Kim–Lim [30], DeMarch–Touzi [25] and Obłój–Siorpaes [57]. In particular, it is not yet fully understood how to break up a martingale transport problem into distinct pieces which mimic the behaviour of minimal components in the one dimensional case.

Below we will give special emphasis to the case $d = 2$ under the additional regularity assumption that ν is absolutely continuous. This instance seems of particular interest since it allows to recognize the geometric structure of the problem while avoiding the more intricate effects of nonminimality which are present in higher dimension. Based on the results of [25, 57] and a particular "monotonicity principle" we will be able to largely recover the main one-dimensional result (Theorem 3.1) in the two-dimensional case, see Sections 3.2-3.3 below and specifically Theorem 3.13 therein. We conjecture that a similar structural characterization of sBm can be established in general dimensions, pending future developments in the direction of [25, 57].

1.4. *Further remarks.*

1.4.1. *Discrete time version and monotonicity principle.* The classical Benamou–Brenier transport formulation immediately reduces to the familiar discrete time transport problem for squared distance costs. Similarly, the martingale version (MBB) can be reformulated as discrete time problem, more precisely, a weak transport problem in the sense of [34].

The discrete time reformulation of (MBB) plays an important role in the derivation of our main results. To analyze the discrete problem, we introduce a "monotonicity principle" for weak transport problems. The origin of this approach is the characterization of optimal transport plans in terms of c-cyclical monotonicity. In optimal transport, the potential of this concept has been recognized by Gangbo–McCann [29]. More recently, variants of this idea have proved to be useful in a number of related situations, see [10, 12, 14, 17, 36, 37, 48, 55, 65] among others. In view of this, it seems possible that the monotonicity principle for weak transport problems could also be of interest in its own right (cf. [5, 32] which appeared after we first posted this article).

1.4.2. *Schrödinger problem.* Our variational problem (MBB) is reminiscent of the celebrated Schrödinger problem, in which the idea is to minimize the relative entropy with respect to Wiener measure (or other Markov laws) over path-measures with fixed initial and final marginals. We refer to the survey [50] and the references therein. Among the similarities, let us mention that the solution to the Schrödinger problem is unique and is a Markov law, and furthermore this problem also has a transport-like discrete time reformulation which is fundamental to the dynamic path-space version. On the other hand, (MBB) and the Schrödinger problem are in particular sense at opposing ends of probabilistic variational problems, we optimize over "volatilities keeping the drift fixed" whereas the latter optimizes over "drifts keeping the volatility fixed."

1.4.3. *Bass-martingale and Skorokhod embedding.* The Bass-martingale (8) was used by Bass [8] to solve the Skorokhod embedding problem. Hobson asked whether there are natural optimality properties related to this construction and if one could give a version with a nontrivial starting law. (MBB) yields such an optimality property of the Bass construction and stretched Brownian motion gives rise to a version of the Bass embedding with non trivial starting law. Notably a characterization of the Bass martingale in terms of an optimality property was first obtained in [9], the variational problem considered in that article refers to measure valued martingales and appears rather different from the one considered in (MBB).

1.4.4. *Geometric Brownian motion.* From the above results, it is clear that Brownian motion is (up to an appropriate scaling of time) a s^{2}Bm between any of its marginals. In fact, the same holds for Brownian martingales $dM_t = \sigma\, dB_t$ for constant and time-independent σ. We find it notable that the same applies in the case of *geometric* Brownian motion.

1.4.5. *Kellerer's theorem and Lipschitz kernels.* Kellerer's theorem [47] states that if a family of distributions $(\mu_t)_{t\in[0,1]}$ on the real line satisfies $s \le t \Rightarrow \mu_s \preceq_c \mu_t$, there exists a Markovian martingale $(X_t)_{t\in\mathbb{R}_+}$ with $law(X_t) = \mu_t$ for every t. In contemporary terms (see [40]), $(\mu_t)_{t\in\mathbb{R}_+}$ is called a *peacock* and $(X_t)_{t\in\mathbb{R}_+}$ is a Markovian martingale associated to this peacock.

The technically most involved part in establishing Kellerer's theorem is to prove that for $\mu \preceq_c \nu$ there exists a martingale transition kernel P having the following *Lipschitz-property*: A kernel $P : x \mapsto \pi_x$, $\nu(dy) = \int \mu(dx)\pi_x(dy)$ is called Lipschitz (or more precisely 1-Lipschitz) if $\mathcal{W}_1(\pi_x, \pi_{x'}) \le |x - x'|$ for all x, x'. Kellerer's proof of the existence of Lipschitz-kernels is not constructive and employs Choquet's theorem. Other proofs are based on solutions to the Skorokhod problem for nontrivial starting law, see [16, 51].

Stretched Brownian motion yields a new construction of a Lipschitz-kernel: Given probabilities $\mu, \nu, \mu \preceq_c \nu$ on the real line and writing M^* for sBm from μ to ν, then $law(M_1^*|M_0^*)$ is a Lipschitz kernel. We provide the argument in Corollary 3.2 below.

The question whether Kellerer's theorem can be extended to the case of marginal measures on $\mathbb{R}^d, d \ge 2$ remains open. While all previously known constructions of kernels used for the proof of Kellerer's theorem were inherently limited to dimension $d = 1$, the approach sketched above seems more susceptible to generalization. We intend to pursue this question further in future work.

1.4.6. *Almost continuous diffusions / local volatility model.* Assume that $(\mu_t)_{t\in[0,1]}$ (where $\mu_t, t \in [0, 1]$ are probabilities on the real line) is a peacock such that $t \mapsto \mu_t$ is continuous in the weak topology. Lowther [51] establishes that an appropriate continuity condition makes the Markov martingale appearing in Kellerer's theorem unique. In his terms, there is a unique "almost continuous" martingale diffusion M^{ac} such that $M_t^{ac} \sim \mu_t$, $t \in [0, 1]$. Under further regularity conditions, M^{ac} is precisely Dupire's local volatility model.

Stretched Brownian motion yields a simple approximation scheme to M^{ac}. Write M^n for the Markov martingale satisfying that for each $k \in \{1, \dots, n\}$, $(M_t^n)_{t\in[(k-1)/n,k/n]}$ is (modulo the obvious affine time-change) stretched Brownian motion between $\mu_{(k-1)/n}$ and $\mu_{k/n}$. M^n is then a continuous diffusion and based on Lowther's [51, 52] it is straightforward that

$$M^{ac} = \lim_{n\to\infty} M^n, \tag{11}$$

where the limit is in the sense of convergence of finite dimensional distribution (cf. [16]).

1.4.7. *Lévy processes.* Many arguments in this article rely only on the independence and stationarity of increments of Brownian motion. Therefore a problem similar to (MBB), but based on a reference Lévy process instead, should conceivably exhibit similar properties as we find in the Brownian case. In this direction, it could be an interesting question to identify the outcome of the approximation procedure described in (11).

1.4.8. *Dual problem, related work.* Optimization problems akin to (MBB) were first studied from a general perspective by Tan and Touzi [62], in particular establishing a duality theory for these type of problems. The dual viewpoint is also emphasized in [42], which is parallel to the present work. Among other results, [42] derives a PDE that yields a sufficient condition for a flow of measures to optimize (MBB) or related cost criteria.

1.5. *Outline of the article.* In Section 2, we introduce the discrete-time variant of our optimization problem. We also prove some of the multidimensional results stated in the Introduction and provide further properties of sBm (dynamic programming principle for (MBB), the Markov property of sBm). In Section 3, we state our main results regarding the structure of sBm in dimensions one and two. In Section 4, we present a monotonicity principle for weak transport problems, which is crucial for our analysis in dimension two, but may also be of independent interest. In Section 5, we conclude the proofs of our main results. Finally in Section 6, we present further optimality properties of sBm and s^{2}Bm in terms of a (causal) optimal transport problem between martingale laws.

1.6. *Notation.* The set of probability measures on a set X will be denoted by $\mathcal{P}(\mathsf{X})$. For $\rho_1, \rho_2 \in \mathcal{P}(\mathsf{X})$ we write $\Pi(\rho_1, \rho_2)$ for the set of all couplings of ρ_1 and ρ_2, that is, all measures on the product space with marginals ρ_1 and ρ_2 resp. Two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ are said to be in convex order, short $\mu \preceq_c \nu$ iff for all convex real valued functions ϕ it holds that $\int \phi \, d\mu \le \int \phi \, d\nu$.

In this article, we fix $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, assume that $\mu \preceq_c \nu$ and that both measures have finite second moment.

We denote by $\mathsf{M}(\mu, \nu)$ the set of all martingale couplings with marginals μ and ν (which is nonempty by Strassen's theorem [61]), that is,

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

For a generic measure π on $\mathbb{R}^d \times \mathbb{R}^d$ we denote by $(\pi_x)_{x \in \mathbb{R}^d}$ the conditional transition kernel given the first coordinate or equivalently its disintegration w.r.t. the first marginal. For $\rho \in \mathcal{P}(\mathsf{X})$ and a measurable map $f : \mathsf{X} \to \mathsf{Y}$, we write $f(\rho) = \rho \circ f^{-1}$ for the pushforward of ρ under f.

For a set $A \subset \mathbb{R}^d$ we denote by $\mathrm{aff}(A)$ the smallest affine vector space containing it, $\dim(A)$ the dimension of $\mathrm{aff}(A)$, $\mathrm{ri}(A)$ the relative interior of A (i.e., interior of A with respect to the relative topology of $\mathrm{aff}(A)$ as inherited from the usual topology in $\mathbb{R}^d$), and $\partial A := \overline{A} \setminus \mathrm{ri}(A)$ the relative boundary. By $\mathrm{co}(A)$ and $\overline{\mathrm{co}}(A)$, we denote the convex hull and the closed convex hull of A, respectively. The relative face of A at a is defined by $\mathrm{rf}_a(A) = \{y \in A : (a - \varepsilon(y - a), y + \varepsilon(y - a)) \subset A, \text{some } \varepsilon > 0\}$. For a set $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$, we denote $\Gamma_x := \{y : (x, y) \in \Gamma\}$ and $\mathrm{proj}_1(\Gamma)$ the projection of Γ onto the first coordinate. Given $\pi \in \mathsf{M}(\mu, \nu)$ we say that $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ is a martingale support for π if $\pi(\Gamma) = 1$ and $x \in \mathrm{ri}(\Gamma_x)$ for μ-a.e. x.

Finally, we denote by $\lambda^d, \gamma^d, \gamma_t^d$ resp. the Lebesgue, standard Gaussian, and the Gaussian measure with covariance matrix $t\,\mathrm{Id}$ in $\mathbb{R}^d$, and reserve the symbol $*$ for convolution.

2. Refined and auxiliary results in arbitrary dimensions. We start by restating our main optimization problem in (slightly) more precise form.

$$MT := MT(\mu, \nu) := \sup_{M_t = M_0 + \int_0^t \sigma_s \, dB_s, M_0 \sim \mu, M_1 \sim \nu} \mathbb{E}\left[\int_0^1 \operatorname{tr}(\sigma_t) \, dt\right]. \tag{12}$$

Here the supremum is taken over the class of 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. In fact, as a particular consequence of Theorem 2.2, the choice of the underlying probability space is not relevant, provided that $(\Omega, \mathcal{F}, \mathbb{P})$ is rich enough to support a $\mathcal{F}_0$-measurable random variable with continuous distribution.

By Doob's martingale representation theorem (see, e.g., [45], Theorem 4.2), the supremum above is the same if we optimized over all continuous d-dimensional local martingales from μ to ν with absolutely continuous cross variation matrix (one then replaces the cost by the trace of the root of the Radon Nikodym density of said matrix).

We will be also interested in a "static" version of the above problem, just as the Benamou–Brenier formula is associated to the static optimal transport problem with quadratic cost

$$WT := WT(\mu, \nu) := \sup_{\substack{\{\pi_x\}_x, \operatorname{mean}(\pi_x) = x \\ \int \mu(dx)\pi_x(dy) = \nu(dy)}} \int \mu(dx) \sup_{q \in \Pi(\pi_x, \gamma^d)} \int q(dm, db) \, m \cdot b. \tag{WOT}$$

The tag (WOT) reflects the fact that this is a weak optimal transport problem (the cost function is nonlinear in the optimization variable).

REMARK 2.1. Completing the square in (WOT) yields

$$1 + \int |y|^2 \, d\nu - 2WT = \inf_{\substack{\{\pi_x\}_x, \operatorname{mean}(\pi_x) = x \\ \int \mu(dx)\pi_x(dy) = \nu(dy)}} \int \mu(dx) \mathcal{W}_2(\pi_x, \gamma^d)^2, \tag{13}$$

where $\mathcal{W}_2$ is the usual L^2 Wasserstein distance on $\mathcal{P}(\mathbb{R}^d)$. The r.h.s. of (13) is clearly a weak transport problem in the setting of Gozlan et al. [33, 34].

We start by establishing the link between the static and dynamic problems introduced so far, and moreover, establish the uniqueness of optimizers in either case. As a corollary, this yields Theorem 1.5 stated in the Introduction.

THEOREM 2.2. *The static and the dynamic problems* (WOT) *and* (MBB) *are equivalent. More precisely*:

1. $WT = MT < \infty$,
2. (WOT) *has a unique optimizer* π^*;
3. (MBB) *has a unique-in-law optimizer* M^*;
4. $\pi^* = law(M_0^*, M_1^*)$ *and* $M^* = G(\pi^*)$ *for some function* G, *that is*, M^* *can be explicitly constructed from* π^* (*see Remark* 2.3 *for details*).

PROOF. Let M be feasible for (MBB). By Itô's formula and the martingale property of M, we have

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

Letting $q_x = law(M_1, B_1 - B_0|M_0 = x)$, we find $q_x \in \Pi(\pi_x, \gamma^d)$ for $\pi_x = law(M_1|M_0 = x)$ and

$$\mathbb{E}\left[\int_0^1 \mathrm{tr}(\sigma_t)\,dt\right] = \int \mu(dx) \int q_x(dm, db)\, m \cdot b.$$

From this, we easily conclude $WT \geq MT$.

Now let π be feasible for (*WOT*). For each x we can find $F^x(\cdot)$ convex such that $\nabla F^x(\gamma^d) = \pi_x$. We now define $M_t^x := E[\nabla F^x(B_1)|\mathcal{F}_t^B]$ for a given standard Brownian motion on $\mathbb{R}^d$ with Brownian filtration $\mathcal{F}^B$. Potentially enlarging our probability space we can assume the existence of a random variable X independent of the Brownian motion B with $X \sim \mu$. We denote the filtration (on the potentially bigger probability space) by $\mathcal{F}$. Since $M_0^x = \int y \pi_x(dy) = x$ and $\int \mu(dx)\pi_x(dy) = \nu(dy)$ we conclude[2] that $\{M_t^X\}_{t\in[0,1]}$ is a continuous martingale from μ to ν. By construction

$$\begin{aligned}\int \mu(dx) \sup_{q\in\Pi(\pi_x,\gamma^d)} \int q(dm, db)\, m \cdot b &= \int \mu(dx) \int \gamma^d(db) b \cdot \nabla F^x(b) \\ &= \mathbb{E}[\mathbb{E}[B_1 \cdot M_1^X|X]],\end{aligned}$$

and the last term equals $\mathbb{E}[\int_0^1 \mathrm{tr}(\sigma_t)\,dt]$ as before (σ can easily be computed from ∇F^x). This proves $WT \leq MT$ and hence $WT = MT$. The finiteness $\infty > WT$ follows from $m \cdot b \leq |m|^2 + |b|^2$ and ν and γ having finite second moment; see (*WOT*).

To show that (*WOT*) is attained,[3] denote by $(\pi^n)_{n\in\mathbb{N}}$ (where $\pi^n(dx, dy) = \pi_x^n(dy)\mu(dy)$) an optimizing sequence. The set $\Pi(\mu, \nu)$ is weakly compact in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. Moreover, the convex subset $\mathsf{M}(\mu, \nu)$ is weakly closed (hence weakly compact), for example, [63], Theorem 7.12(iv). By [7], Theorem 3.7, we obtain the existence of a measurable kernel $x \mapsto \pi_x \in \mathcal{P}(\mathbb{R}^d)$ and a subsequence, still denoted by $(\pi^n)_n$, such that on a μ-full set

$$\frac{1}{N} \sum_{n\leq N} \pi_x^n(dy) \to \pi_x(dy),$$

with respect to weak convergence in $\mathcal{P}(\mathbb{R}^d)$. In particular $\frac{1}{N}\sum_{n\leq N} \pi^n \to \pi$ in the weak topology in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$, where $\pi(dx, dy) := \mu(dx)\pi_x(dy)$. Since $\mathsf{M}(\mu, \nu)$ is closed, we have that $\pi \in \mathsf{M}(\mu, \nu)$. Finally,

$$\begin{aligned}WT &= \lim_n \int \mu(dx) \sup_{q\in\Pi(\pi_x^n,\gamma^d)} \int q(dm, db)\, m \cdot b \\ &= \lim_N \int \mu(dx) \frac{1}{N} \sum_{n\leq N} \sup_{q\in\Pi(\pi_x^n,\gamma^d)} \int q(dm, db)\, m \cdot b \\ &\leq \lim_N \int \mu(dx) \sup_{q\in\Pi(\frac{1}{N}\sum_{n\leq N}\pi_x^n,\gamma^d)} \int q(dm, db)\, m \cdot b\end{aligned}$$

[2]A technical point here is to observe that $(x, y) \mapsto \nabla F^x(y)$ admits a jointly measurable version, as follows by [27], Theorem 1.1. Hence $(x, y) \mapsto f_t^x(y) := \int \nabla F^x(b+y)\gamma_{1-t}^d(db)$ is jointly measurable, and as $M_t^x = f_t^x(B_t)$, we see that M_t^X does define an (X, B)-adapted process.

[3]Note added in revision: Since we first posted this article the general theory of weak optimal transport has been further developed and in particular there are now abstract existence results which also cover (*WOT*), see [5]. We keep the short argument for the convenience of the reader.

$$\leq \int \mu(dx) \limsup_N \sup_{q\in\Pi(\frac{1}{N}\sum_{n\leq N}\pi_x^n,\gamma^d)} \int q(dm,db)\, m\cdot b$$

$$\leq \int \mu(dx) \sup_{q\in\Pi(\pi_x,\gamma^d)} \int q(dm,db)\, m\cdot b \leq WT.$$

The first inequality is a consequence of the concavity of $\eta \mapsto H(\eta) := \sup_{q\in\Pi(\eta,\gamma^d)} \int q(dm, db)\, m\cdot b$ w.r.t. convex combinations of measures. The second inequality is Fatou's lemma, noticing that the integrand is bounded in $L^1(\mu)$ (the bound equals the sum of the second moments of μ and γ). The third inequality follows by weak convergence of the averaged kernel on a μ-full set and upper semicontinuity of $H(\cdot)$ (see [64], Lemma 4.3 and Remark 6.12). For uniqueness, it suffices to notice that $H(\cdot)$ is actually strictly concave, which is an easy consequence of Brenier's theorem. Hence, (*WOT*) is attained and we denote the unique optimizer by π^*.

Taking π^* we may build an optimizer M^* for (MBB) as in the first part of the proof (as the value of both problems agree).

We finally establish the uniqueness of optimizers for (MBB). Let $\tilde{M}$ be any such optimizer. From the previous considerations, we deduce that the law of $(\tilde{M}_0, \tilde{M}_1)$ is the unique optimizer π^* of (*WOT*). Conditioning on $\{\tilde{M}_0 = x\}$ we thus have that $\tilde{M}$ connects δ_x to π_x^*. It follows that $\mu(dx)$-a.s. $\tilde{M}$ conditioned on $\{\tilde{M}_0 = x\}$ is optimal between these marginals. Indeed,

$$\sup_{N_t = x+\int_0^t \sigma_s\, dB_s,\, N_1\sim\pi_x^*} \mathbb{E}\left[\int_0^1 \mathrm{tr}(\sigma_t)\, dt\right] = \sup_{q\in\Pi(\pi_x^*,\gamma^d)} \int q(dm,db)\, m\cdot b, \tag{14}$$

by the results obtained so far, since if $\tilde{M}$ conditioned on $\{\tilde{M}_0 = x\}$ was not optimal for the l.h.s. it could not deliver the equality $MT = WT$. So it suffices to show that the l.h.s. of (14) is uniquely attained. But any candidate martingale N with volatility σ satisfies $\mathbb{E}[\int_0^1 \mathrm{tr}(\sigma_t)\, dt] = \mathbb{E}[N_1 B_1]$ (since here we can assume $B_0 = 0$). Hence, Brenier's theorem implies that $\tilde{M}_1 = \nabla F^x(B_1)$ on $\{\tilde{M}_0 = x\}$, for a convex function F^x. Since the optimal transport map ∇F^x is unique, and the martingale property determines uniquely the law of $\tilde{M}$, we finally get $\tilde{M} = M^*$ in law. □

REMARK 2.3. The proof of Theorem 2.2 shows how to build the optimizer for (MBB) via the following procedure, making the statement $M^* = G(\pi^*)$ in Theorem 2.2 (2) precise:

1. Find the unique optimizer π^* of (*WOT*).
2. Find convex functions F^x such that $\nabla F^x(\gamma^d) = \pi_x^*$.
3. Define $M_t^x := \mathbb{E}[\nabla F^x(B_1)|B_t] = \int \nabla F^x(y + B_t)\gamma_{1-t}^d(dy)$.
4. Take $X \sim \mu$ independent of B and let $M_t := M_t^X$.

In particular, this proves Theorem 1.11 in the Introduction.

We now establish further properties of the optimizer M^* of (MBB), which hold likewise in any number of dimensions. The first two of them will be important for the proofs of the results yet to come, namely that (MBB) obeys a dynamic programming principle and that M^* is a strong Markov martingale. The final property, that M^* is an "optimal constant-speed" interpolation between its marginals, is crucial for the interpretation of our martingale as an analogue of displacement interpolation in classical transport and in particular proves Theorem 1.7 in the Introduction.

Assume that $(B_t)_{t\geq 0}$ is a Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_t)$ and assume that Ω supports a continuously distributed RV independent of $(\mathcal{F}_t)_t$. For $(\mathcal{F}_t)_t$-stopping times $0 \leq \tau \leq T$ we define

$$V(\tau, T, \mu, \nu) := \sup_{\substack{M_r = X + \int_\tau^r \sigma_u \, dB_u, \tau \leq r \leq T \\ X \sim \mu, M_T \sim \nu}} \mathbb{E}\left[\int_\tau^T \operatorname{tr}(\sigma_u) \, du\right], \tag{15}$$

so that $MT = V(0, 1, \mu, \nu)$. Recall that for the supremum in (15) it is irrelevant whether one considers an extension of the original probability space (cf. Theorem 2.2).

LEMMA 2.4 (Dynamic programming principle). *For every stopping time* $0 \leq \tau \leq 1$

$$V(0, 1, \mu, \nu) = \sup_{\substack{M_s = M_0 + \int_0^s \sigma_r \, dB_r \\ 0 \leq s \leq \tau, M_0 \sim \mu}} \left\{\mathbb{E}\left[\int_0^\tau \operatorname{tr}(\sigma_r) \, dr\right] + V\big(\tau, 1, law(M_\tau), \nu\big)\right\}, \tag{16}$$

with the convention that $\sup \varnothing = -\infty$. *In particular if* M^* *is optimal for* $V(0, 1, \mu, \nu)$, *then*:

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

PROOF. Obviously the l.h.s. of (16) is smaller than the r.h.s. of (16). Take now M^1 feasible for the r.h.s. (so that M^1 is adapted to a filtration $\{\mathcal{F}^1_{s\wedge\tau}\}_{s\geq 0}$, B is a Brownian motion on $[0, \tau]$ adapted to it, and $dM^1 = \sigma^{(1)} \, dB$). Let M^2 be optimal for $V(\tau, 1, law(M^1_\tau), \nu)$. By Remark 2.3 we may build M^2 from the starting distribution M^1_τ and the filtration $\mathcal{F}^2$ of this random variable and a Brownian motion W independent of $\mathcal{F}^1$ (and so of M^1_τ), so $dM^2 = \sigma^{(2)} \, dW$. We then build a continuous martingale M on $[0, 1]$ by setting it to M^1 on $[0, \tau]$ and M^2 on $(\tau, 1]$, obtaining easily that

$$\mathbb{E}\left[\int_0^\tau \operatorname{tr}\big(\sigma^{(1)}_r\big) \, dr\right] + V\big(\tau, law(M^1_\tau), \nu\big) = \mathbb{E}\left[\int_0^\tau \operatorname{tr}\big(\sigma^{(1)}_r\big) \, dr + \int_\tau^1 \operatorname{tr}\big(\sigma^{(2)}_r\big) \, dr\right].$$

Observing that $\tilde{B}_s = 1_{[0,\tau]}(s) B_s + 1_{(\tau,1]}(s)[B_\tau + W_s - W_\tau]$ is a Brownian motion for the concatenation of filtrations $\mathcal{F}^1$ and $\mathcal{F}^2$, and $dM = (1_{[0,\tau]}(s)\sigma^{(1)}_s + 1_{(\tau,1]}(s)\sigma^{(2)}_s) \, d\tilde{B}$, then the r.h.s. above is the cost of M as a martingale starting at μ and ending at ν, and so is smaller than $V(0, 1, \mu, \nu)$.

Let M^* be optimal for $V(0, 1, \mu, \nu)$. Using (16) it is trivial to show Points (1)-(2). But from this follows that $M^*|_{[0,\tau]}$ is optimal for the r.h.s. of (16). This, Point (1), and the arguments in the previous paragraph show how to stitch together $M^*|_{[0,\tau]}$ and $M^*|_{[\tau,1]}$ to produce an optimizer M for $V(0, 1, \mu, \nu)$. But this must then coincide with M^*, by uniqueness. On the other hand M_1 is defined via M^*_τ and a Brownian motion independent of $\{M^*_s : s \leq \tau\}$, so $law(M_1 | M^*_s, s \leq \tau) = law(M_1 | M^*_\tau)$ and we conclude. □

Point (3) of Lemma 2.4 immediately implies the following corollary.

COROLLARY 2.5. *The unique optimizer* M^* *of* (MBB) *has the strong Markov property.*

PROPOSITION 2.6. *Let* M^* *be the optimizer of* (MBB) *and set*

$$[\mu, \nu]^M_t := law(M^*_t).$$

Then law(M_0^*, M_t^*) *is optimal for* (WOT) *between the marginals* μ *and* $[\mu,\nu]_t^M$. *Similarly, the optimizer of* (MBB) *between these marginals is the time-changed martingale* $s \in [0,1] \mapsto M_{st}^*$. *Finally, for* $0 \le r \le t \le 1$, *holds*

$$WT([\mu,\nu]_r^M, [\mu,\nu]_t^M) = MT([\mu,\nu]_r^M, [\mu,\nu]_t^M) = \sqrt{t-r}MT(\mu,\nu) = \sqrt{t-r}WT(\mu,\nu). \tag{17}$$

PROOF. We use the notation in Remark 2.3 and write $M_t^* = M_t^X = f_t^X(B_t)$ where

$$f_t^x(\cdot) := \int \nabla F^x(b+\cdot)\gamma_{1-t}^d(db).$$

Since $[\mu,\nu]_t^M = f_t^X(\sqrt{t}B_1)$, it is not difficult to see that

$$N_s^* := \mathbb{E}[f_t^X(\sqrt{t}B_1)|\mathcal{F}_s^B] = f_{st}^X(\sqrt{t}B_s),$$

is the optimizer of (MBB) from μ at $s=0$ to $[\mu,\nu]_t^M$ at $s=1$. Of course N^* coincides (in law) with the time-changed martingale $s \mapsto M_{st}^*$, and by Theorem 2.2 we get the optimality of *law*(M_0^*, M_t^*). We next remark that $J(f_s^X)(B_s)$ is a matrix-valued martingale, where J stands for Jacobian, as can be easily seen from the convolution structure or PDE arguments. Thus, $\mathbb{E}[J(f_s^X)(B_s)] = \mathbb{E}[J(f_{st}^X)(\sqrt{t}B_s)]$. To recognize the "$\sigma$" of N^* and M^* we observe that

$$dN_s^* = \sqrt{t}J(f_{st}^X)(\sqrt{t}B_s)\,dB_s,$$
$$dM_s^* = J(f_s^X)(B_s)\,dB_s,$$

by Itô's formula. Putting all together we find

$$\begin{aligned}\mathbb{E}\left[\int_0^1 \sqrt{t}J(f_{st}^X)(\sqrt{t}B_s)\,ds\right] &= \sqrt{t}\int_0^1 \mathbb{E}[J(f_{st}^X)(\sqrt{t}B_s)]\,ds \\ &= \sqrt{t}\int_0^1 \mathbb{E}[J(f_s^X)(B_s)]\,ds \\ &= \sqrt{t}\mathbb{E}\left[\int_0^1 J(f_s^X)(B_s)\,ds\right],\end{aligned}$$

and again by Theorem 2.2 we get

$$MT([\mu,\nu]_0^M, [\mu,\nu]_t^M) = \sqrt{t}MT(\mu,\nu).$$

The general case of (17) follows similarly. □

REMARK 2.7. The identities (17), at least for the continuous-time problems, have been obtained in [42], Remark 4.1, in a more general setting, via a scaling argument. The interpretation of (17) is clear: Our optimal martingale is a constant-speed geodesic when distance is measured wrt the square of our cost functional.

3. Main results in dimensions one and two. In this section, we study finer structural properties of the unique optimizer of (MBB) established in the previous section. We get a full description in dimension one, in dimension two under the additional Assumption 3.8 and a partial description in general dimensions.

3.1. *The one-dimensional case.* Let $\mu \preceq_c \nu$ be probability measures on the line with finite second moment. For a measure α on $\mathbb{R}$ and $x \in \mathbb{R}$, we write $u_\alpha(x) := \int |x-y|\, d\alpha(y)$. The convex order relation $\mu \preceq_c \nu$ is equivalent to $u_\mu \le u_\nu$.

We recall from [17], Appendix A.1, that the "irreducible components of (μ, ν)" are determined by the (unique) family of open disjoint intervals $\{I_k\}_{k\in\mathbb{N}}$ whose union equals the open set

$$\{u_\mu < u_\nu\} := \left\{x \in \mathbb{R} : \int |x-y|\, d(\mu-\nu)(y) \neq 0\right\}.$$

One can then decompose

$$\mu = \eta + \sum_k \mu_k \quad \text{and} \quad \nu = \eta + \sum_k \nu_k,$$

where $\mu_k = \mu|_{I_k}$, with $I_k = \{u_{\mu_k} < u_{\nu_k}\}$ and $\nu_k(\overline{I_k}) = \mu_k(I_k)$, whereas η is concentrated on $\mathbb{R} \setminus \bigcup_k I_k$. A useful straightforward result is that every martingale coupling from μ to ν (i.e., $\pi \in \mathsf{M}(\mu, \nu)$) is fully characterized by how it looks on the sets $I_k \times \overline{I_k}$. The restrictions $\pi_k := \pi_{|I_k \times \overline{I_k}} = \pi_{|I_k \times \mathbb{R}}$ are still martingale couplings (in the sense that their respective disintegrations satisfy $\int y (\pi_k)_x(y) = x$ for μ_k-a.a. x) but with total mass $\mu_k(I_k)$ and marginals μ_k, ν_k.

We can now state our main result for $d = 1$, characterizing the structure of stretched Brownian motion.

THEOREM 3.1. *Let $\mu \preceq_c \nu$ be probability measures on the line with finite second moment. A candidate martingale is an optimizer of* (MBB) *if and only if it is a* standard stretched Brownian motion *on each irreducible component* (μ_k, ν_k) *of* (μ, ν). *Hence*, stretched Brownian motion (*sBm*) *is a* standard stretched Brownian motion (s^2*Bm*) *in each irreducible component*.

Let us explain the terminology used here. Saying that M is s^{2}Bm on the irreducible components of (μ, ν) concretely means that, conditionally on $M_0 \in I_k$, M is a s^{2}Bm from $\frac{1}{\mu_k(I_k)}\mu_k$ to $\frac{1}{\mu_k(I_k)}\nu_k$. We stress that in the present $1-d$ case Theorem 3.1 is significantly stronger than Theorem 1.11. Theorem 3.1 is an easy consequence of the two-dimensional considerations, see the end of Section 5.

We now prove the fact, first mentioned in the Introduction, that in $1-d$ the transition kernel of stretched Brownian motion is Lipschitz:

COROLLARY 3.2. *Let $\mu \preceq_c \nu$ be probability measures on the line with finite second moment, and M^* the unique stretched Brownian motion from μ to ν. Then the kernel*

$$x \mapsto \pi^*_x := law\big(M^*_1 | M^*_0 = x\big),$$

has the Lipschitz property: $\mathcal{W}_1(\pi^*_x, \pi^*_{x'}) \le |x - x'|$.

PROOF. By Theorem 3.1, M^* is a (s^{2}Bm) in each irreducible component. Assume first that $x < x'$ and that they belong to the same component. Conditioning to starting in this component, we can write $M^*_t = \mathbb{E}[f(B_1)|B_t]$, with f increasing and B a Brownian motion with some starting law. Choose y, y' such that

$$\mathbb{E}^y\big[f(B_1)\big] = x < x' = \mathbb{E}^{y'}\big[f(B_1)\big]$$

and observe that this implies $y < y'$ since f is increasing. (The notation $\mathbb{E}^y$ means we are conditioning B to start at y.) This in turn implies that for a Brownian motion $\tilde{B}$ starting in zero the random vector

$$\big(f(\tilde{B}_1 + y), f(\tilde{B}_1 + y')\big),$$

is ordered and has marginals π^*_x and $\pi^*_{x'}$. Indeed, the latter property follows since $M^*_0 = x \iff B_0 = y$, so $\pi^*_x = \mathrm{Law}(M^*_1 | B_0 = y) = f(\tilde{B}_1 + y)$, and likewise for x' and y'. Hence,

$$\begin{aligned}\mathcal{W}_1(\pi^*_x, \pi^*_{x'}) &\le \mathbb{E}\big[\big|f(\tilde{B}_1 + y') - f(\tilde{B}_1 + y)\big|\big] \\ &= \mathbb{E}^{y'}\big[f(B_1)\big] - \mathbb{E}^{y}\big[f(B_1)\big] = x' - x.\end{aligned}$$

On the other hand, if x, x' are not in the same component, we let f and g denote the increasing functions associated to the representations in terms of (s^{2}Bm)'s. If $x < x'$, then the range of f lies below the range of g, and we conclude much as in the above display. □

We now proceed towards the subtler extension of Theorem 3.1 for $d = 2$.

3.2. *Preliminaries.* We briefly discuss some of the aspects related to the decomposition of martingale couplings in arbitrary dimensions. Later this will be mostly used in dimension two. After this, we also provide an analytical result of much importance for the next sections.

DEFINITION 3.3. A convex paving $\mathcal{C}$ is a collection of disjoint relatively open convex sets from $\mathbb{R}^d$. Denoting $\bigcup \mathcal{C} := \bigcup_{C \in \mathcal{C}} C$, we will always assume $\mu(\bigcup \mathcal{C}) = 1$ for such objects. For $x \in \bigcup \mathcal{C} \subset \mathbb{R}^d$ we denote by $C(x)$ the unique element of $\mathcal{C}$ which contains x. We say that $\mathcal{C}$ is measurable (resp. μ-measurable, universally measurable) if the function $x \mapsto \overline{C(x)}$ is measurable as a map from $\mathbb{R}^d$ to the Polish space of all closed (convex) subsets of $\mathbb{R}^d$ equipped with the *Wijsman* topology.[4]

DEFINITION 3.4. Let $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ and $\pi \in \mathsf{M}(\mu, \nu)$. We say that a convex paving $\mathcal{C}$ is:

- π-invariant if $\pi_x(\overline{C(x)}) = 1$ for μ-a.e. x,
- Γ-invariant if $\mathrm{ri}(\Gamma_x) \subset C(x)$ for all $x \in \mathrm{proj}_1(\Gamma)$.

Note that a natural order between convex pavings $\mathcal{C}, \mathcal{C}'$ is given by

$$\mathcal{C} \le_\mu \mathcal{C}' \quad \iff \quad C(x) \subset C'(x) \quad \text{for } \mu\text{-a.e. } x,$$

in which case we say that $\mathcal{C}$ is finer than $\mathcal{C}'$ (and the latter is coarser than the former). The following two theorems are shown in [25, 31, 57].

THEOREM 3.5 (Ghoussoub–Kim–Lim [31]). *Given $\pi \in \mathsf{M}(\mu, \nu)$ and $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ a martingale support for π, there is a finest Γ-invariant convex paving. We denote it by $\mathcal{C}_{\pi,\Gamma}$.*

THEOREM 3.6 (De March–Touzi [25], Obłój–Siorpaes [57]). *There is a finest convex paving, denoted $\mathcal{C}_{\mu,\nu}$, which is π-invariant for all $\pi \in \mathsf{M}(\mu, \nu)$ simultaneously. Writing $\mathcal{C}_{\mu,\nu} = \{C_{\mu,\nu}(x)\}_{x \in \mathbb{R}^d}$, the function $x \mapsto C_{\mu,\nu}(x)$ is universally measurable.*

We will actually use another convex paving which incorporates ideas / properties from the above two.

[4]The Wijsman topology on the collection of all closed subsets of a metric space (X, d) is the weak topology generated by $\{\mathrm{dist}(x, \cdot) : x \in \mathsf{X}\}$, cf. [25].

LEMMA 3.7. *Given $\pi \in \mathsf{M}(\mu, \nu)$ there is a finest measurable π-invariant convex paving, which we denote $\mathcal{C}_\pi$.*

If we knew that $\mathcal{C}_{\mu,\nu} = \mathcal{C}_\pi$ for arbitrary $\pi \in \mathsf{M}(\mu, \nu)$, this would streamline some of our proofs. For the case $d = 1$ this is indeed the case: Here convex pavings correspond precisely to the supporting intervals of the individual irreducible components. Specifically, if (μ, ν) is irreducible, then C_π has just one cell for each $\pi \in \mathsf{M}(\mu, \nu)$ (see Appendix A and in particular Lemma A.6 in [17]). Note, however, that already for $d = 2$ this can fail.

Lemma 3.7 can be established by a close reading of [25], and adapting the arguments therein (of course [25] achieves much more!). We give a self-contained, shorter argument under the following additional hypothesis, which will also appear in Section 3.3.

ASSUMPTION 3.8. For all $\pi \in \mathsf{M}(\mu, \nu)$ and $\mathcal{C}$ convex paving we have

$$\pi_x\big(\overline{C(x)}\big) = 1 \quad \mu\text{-a.s.} \quad \Rightarrow \quad \pi_x\big(C(x)\big) = 1 \quad \mu\text{-a.s.}$$

In particular, for such $\mathcal{C}$ and π, $\mathcal{C}$ is π-invariant iff $\pi_x(C(x)) = 1\mu$-a.s.

PROOF OF LEMMMA 3.7 UNDER ASSUMPTION 3.8. Inspired by [25], we introduce the optimization problem

$$\inf\left\{\int \mu(dx)G\big(C(x)\big) : \mathcal{C} \text{ is a } \pi\text{-invariant measurable convex paving}\right\},$$

where $G(C) := \dim(C) + g_C(C)$ and g_C is the standard Gaussian measure on $\mathrm{aff}(C)$, that is, as obtained from the $\dim(C)$-dimensional Lebesgue measure on $\mathrm{aff}(C)$. Let $\mathcal{C}^n$ be an optimizing sequence of π-invariant convex pavings and let Ω be a set of μ-full measure on which we have $\pi_x(\overline{C^n(x)}) = 1$ for all n (here $C^n(x)$ denotes an element of $\mathcal{C}^n$). Introduce for $x \in \Omega$ the relatively open convex sets $C_\pi(x) := \mathrm{rf}_x(\bigcap C^n(x))$. We have[5] $x \in C_\pi(x)$ since $x \in \bigcap C^n(x)$. Moreover, we have that $\mathcal{C}_\pi := \{C_\pi(x) : x \in \Omega\}$ forms a partition since already $\{\bigcap C^n(x) : x \in \Omega\}$ is a partition. Let us establish that $\pi_x(\overline{C_\pi(x)}) = 1$.

We start assuming

$$\forall K \text{convex:} \quad \mathrm{ri}\,\overline{\mathrm{co}}\,\mathrm{supp}\,\pi_x \subset K \Rightarrow \pi_x(\overline{\mathrm{rf}_x K}) = 1. \tag{18}$$

Let us take $K := \bigcap C^n(x)$. Since $\overline{C^n(x)}$ is closed, convex and satisfies $\pi_x(\overline{C^n(x)}) = 1$ we have $\overline{\mathrm{co}}\,\mathrm{supp}\,\pi_x \subset \overline{C^n(x)}$. On the other hand, $\overline{\mathrm{co}}\,\mathrm{supp}\,\pi_x$ cannot be contained in $\partial C^n(x)$ since by Assumption 3.8 we have $\pi_x(\partial C^n(x)) = 0$. By [59], Corollary 6.5.2, we must then have $\mathrm{ri}\,\overline{\mathrm{co}}\,\mathrm{supp}\,\pi_x \subset \mathrm{ri}\,C^n(x) = C^n(x)$ for all n, so $\mathrm{ri}\,\overline{\mathrm{co}}\,\mathrm{supp}\,\pi_x \subset \bigcap C^n(x) = K$. By (18) we get $\pi_x(\overline{\mathrm{rf}_x K}) = \pi_x(\overline{C_\pi(x)}) = 1$ as desired. All in all $\mathcal{C}_\pi$ is a π-invariant convex paving, and since $C_\pi(x) \subset C^n(x)$ we find $\int \mu(dx)G(C_\pi(x)) \leq \int \mu(dx)G(C^n(x))$ from which we get the optimality of $\mathcal{C}_\pi$.

To finish, let us establish (18). By the martingale property we easily see[6] that $x \in \mathrm{ri}\,\overline{\mathrm{co}}\,\mathrm{supp}\,\pi_x$. From this, $\mathrm{ri}\,\overline{\mathrm{co}}\,\mathrm{supp}\,\pi_x = \mathrm{rf}_x(\mathrm{ri}\,\overline{\mathrm{co}}\,\mathrm{supp}\,\pi_x) \subset \mathrm{rf}_x K$. Hence, $\overline{\mathrm{ri}\,\overline{\mathrm{co}}\,\mathrm{supp}\,\pi_x} \subset \overline{\mathrm{rf}_x K}$, whose l.h.s. equals $\overline{\mathrm{co}}\,\mathrm{supp}\,\pi_x$ by [59], Theorem 6.3, so (18) follows. □

REMARK 3.9. The same proof, modulo obvious changes, proves the existence of a finest measurable convex paving invariant for all $\pi \in \mathsf{M}(\mu, \nu)$ simultaneously. This, however, does not establish the existence of a maximally spreading martingale coupling as in [25].

[5]Recall that $A \subset A' \Rightarrow \mathrm{rf}_a A \subset \mathrm{rf}_a A'$, that $a \in A \iff a \in \mathrm{rf}_a(A)$ and that $\mathrm{rf}_a(A) = \mathrm{ri}\,A \iff a \in \mathrm{ri}\,A$.

[6]Let $m = \dim(\overline{\mathrm{co}}\,\mathrm{supp}\,\pi_x)$ and suppose $x \in \partial(\overline{\mathrm{co}}\,\mathrm{supp}\,\pi_x)$. We can then find an $(m-1)$-dimensional hyperplane supporting x and having $\overline{\mathrm{co}}\,\mathrm{supp}\,\pi_x$ contained in one associated half-space. By the martingale property one obtains that necessarily $\mathrm{supp}\,\pi_x$, and then $\overline{\mathrm{co}}\,\mathrm{supp}\,\pi_x$ too, must be actually contained in the hyperplane itself. Thus $\dim(\overline{\mathrm{co}}\,\mathrm{supp}\,\pi_x) \leq m - 1$ yielding a contradiction.

Here is a sufficient criterion for Assumption 3.8 to hold.

LEMMA 3.10. *Assumption 3.8 is satisfied if $d \in \{1,2\}$ and $\nu \ll \lambda^d$.*

PROOF. This follows by similar arguments as in [31], Lemma C.1. We omit the details. □

A direct consequence of Theorem 3.6 and Assumption 3.8 is the decomposition of a martingale into irreducible components. Notice the resemblance to the one-dimensional case explained in Section 3.1.

PROPOSITION 3.11. *Let $\mathcal{C}_{\mu,\nu} = \{C_{\mu,\nu}(x)\}_{x\in\mathbb{R}^d}$ be the convex paving of Theorem 3.6 and assume Assumption 3.8. Then*:

(i) *we may decompose*

$$\mu = \int \mu(\cdot|K)\, dC_{\mu,\nu}(\mu)(K) \quad \text{and} \quad \nu = \int \nu(\cdot|K)\, dC_{\mu,\nu}(\mu)(K),$$

with $\mu(\cdot|K) \preceq_c \nu(\cdot|K)$ for $C_{\mu,\nu}(\mu)$-a.e. K;

(ii) *for any martingale coupling $\pi \in \mathsf{M}(\mu,\nu)$ we have that*

$$\pi(\cdot|K\times K) = \pi\big(\cdot|K\times\mathbb{R}^d\big) \quad \text{for } C_{\mu,\nu}(\mu)\text{-a.e. } K,$$

and this common measure has first and second marginals equal to $\mu(\cdot|K)$ and $\nu(\cdot|K)$ respectively;

(iii) *any martingale coupling $\pi \in \mathsf{M}(\mu,\nu)$ can be uniquely decomposed as*

$$\pi = \int \pi(\cdot|K\times K)\, dC_{\mu,\nu}(\mu)(K).$$

The proof is just as in [17], Appendix A.1, but simpler, thanks to the fact that under Assumption 3.8 we have that *martingales started on two neighbouring cells will not go on to reach the intersection of the boundaries of the cells*. We thus omit the proof.

We finally present a technical lemma which will be extremely useful in the proofs of the main results in dimension two.

LEMMA 3.12. *Let η be a probability measure in $\mathbb{R}^d$ with finite second moment, and $F:\mathbb{R}^d\to\mathbb{R}$ convex such that $\nabla F(\gamma^d)=\eta$. Denote $V := \mathrm{aff}(\mathrm{supp}(\eta))$ and let P be the orthogonal projection onto V. Then, there exists a convex function $\tilde F : V\to\mathbb{R}$ such that γ^d-a.s. $\nabla F = \nabla\tilde F\circ P$. For all $s>0$, the function*

$$\mathbb{R}^d \ni b \mapsto f_s(b) := \int \nabla F(b+y)\gamma_s^d(dy) = \int \nabla\tilde F(Pb+z)P\big(\gamma_s^d\big)(dz) \in \mathbb{R}^d,$$

has the following properties:

1. *It is infinitely continuously differentiable.*
2. *Restricted to V, it is one-to-one.*
3. $\overline{f_s(\mathbb{R}^d)} = \overline{\mathrm{co}}\,\nabla F(\mathbb{R}^d)$.
4. *$f_s(\gamma^d)$ is equivalent to the m-dimensional Lebesgue measure on V restricted to $\mathrm{co}\,\nabla F(\mathbb{R}^d)$, where $m=\dim(V)$.*
5. *$\mathrm{supp}(f_s(\gamma_t^d)) = \overline{\mathrm{co}}\,\nabla F(\mathbb{R}^d)$ is convex and does not depend on $s>0$ nor $t>0$.*

PROOF. The γ^d-a.s. equality $\nabla F = \nabla \tilde{F} \circ P$, follows from Brenier's Theorem by taking $\nabla \tilde{F}$ mapping $P(\gamma^d)$ into η and observing that $\nabla(\tilde{F} \circ P) = P((\nabla \tilde{F}) \circ P) = (\nabla \tilde{F}) \circ P$. Point (1) follows by change of variables and differentiation under the integral sign. Alternatively, one can argue with the classical backwards heat equation. Points (2), (3) and (5) follow by the full-support property of γ^d in $\mathbb{R}^d$ and $P(\gamma^d)$ in V.

Point 4 is trivially true if η is a Dirac delta (then $m = 0$). Otherwise it suffices to consider the smooth function $V \ni v \mapsto \tilde{f}_s(v) := \int \nabla \tilde{F}(v + z)\tilde{\gamma}(dz)$, with $\tilde{\gamma} = P(\gamma_s^d)$, and to prove that $\tilde{f}_s(\tilde{\gamma}) \sim \lambda_V|_{\mathrm{co}\nabla F(\mathbb{R}^d)}$, where the latter denotes m-dimensional Lebesgue on V restricted to $\mathrm{co}\nabla F(\mathbb{R}^d)$. Since $\tilde{\gamma} \sim \lambda_V$, we have by [63], Theorem 4.8(i), that λ_V-a.e. the Jacobian of $\tilde{f}_s$ is invertible. By the change of variables formula, it is easy to obtain that $\tilde{f}_s(\tilde{\gamma}) \ll \lambda_V$, and the previous observation with the Monge–Ampère equation [63], Theorem 4.8(iii), yield

$$\lambda_V\text{-a.e. } r: \quad \frac{d\tilde{f}_s(\tilde{\gamma})}{d\lambda_V}(r) = \left|\det\left((J\tilde{f}_s)^{-1}(r)\right)\right| \frac{d\tilde{\gamma}}{d\lambda_V}\left((\tilde{f}_s)^{-1}(r)\right)\mathbf{1}_{\tilde{f}_s(V)}(r). \tag{19}$$

By Point 3, $\tilde{f}_s(V) = \mathrm{co}\nabla F(\mathbb{R}^d)$, and so we conclude $\tilde{f}_s(\tilde{\gamma}) \sim \lambda_V|_{\mathrm{co}\nabla F(\mathbb{R}^d)}$ since under the latter measure the density $\frac{d\tilde{f}_s(\tilde{\gamma})}{d\lambda_V|_{\mathrm{co}\nabla F(\mathbb{R}^d)}}$ is a.e. nonvanishing. □

3.3. *The two-dimensional case.* Our first main result for $d = 2$ is a characterization of the structure of sBm, providing a significantly strengthened version of Theorem 1.11 in the Introduction.

THEOREM 3.13. *Let $\mu \preceq_c \nu$ be probability measures in $\mathbb{R}^2$ with finite second moments. Suppose $\nu \ll \lambda^2$, and let M^* be the unique optimizer for* (MBB). *Set $\pi^t = law(M_0^*, M_t^*)$ for $0 < t < 1$. Then the stretched Brownian motion M^* is a standard stretched Brownian motion on each cell of $\mathcal{C}_{\pi^t}$.*

The second main result of this part is the optimality of s^2Bm whenever we are able to build them with respect to the coarser $\mathcal{C}_{\mu,\nu}$ convex paving. Our proof of such a result relies on the simplifying Assumption 3.8, which as seen in Lemma 3.10 is *verified in dimension two under the further requirement that ν be absolutely continuous*. We therefore place this result here, although in principle it is a result valid in arbitrary dimensions.

THEOREM 3.14. *Under Assumption 3.8, if M is a standard stretched Brownian motion on each cell[7] of the convex paving $\mathcal{C}_{\mu,\nu}$, then it is optimal for* (MBB) (*i.e., it is a sBm*).

REMARK 3.15. The difference between Theorem 1.10 and Theorem 3.14 is as follows: the first result says that standard stretched Brownian motion is optimal in its own, whereas the second statement allows for more freedom in that we are allowed to choose the convex function in the definition of stretched Brownian motion dependent on the cells of $\mathcal{C}_{\mu,\nu}$. Therefore, this result is a strengthened version of Theorem 1.10.

REMARK 3.16. For dimension one ($d = 1$), Theorem 3.1 establishes the existence of standard stretched Brownian motion, and characterizes it as the sole optimizer. Both existence and optimality are understood with respect to the same (countable) convex paving. For two dimensions ($d = 2$), Theorems 3.13 and 3.14 and Lemma 3.10 establish, under the assumption that $\nu \ll \lambda^2$, the existence and optimality characterization of standard stretched Brownian motion. In this case, however, existence and optimality are understood with respect to potentially different convex pavings.

[7]This means that for $C_{\mu,\nu}(\mu)$-a.e. K, the conditioning of M^* to $M_0^* \in K$ is a stretched Brownian motion between the marginals $\mu(\cdot|K)$ and $\nu(\cdot|K)$ introduced in Proposition 3.11

The proofs of these results are deferred to Section 5. Theorem 3.13 relies crucially on a monotonicity principle which we now establish and which seems of independent interest.

4. A monotonicity principle for weak optimal transport problems. For this part only, we adopt a more general setting. Let X, Y be Polish spaces and $C : \mathsf{X} \times \mathcal{P}(\mathsf{Y}) \to \mathbb{R} \cup \{+\infty\}$ Borel measurable. Consider for $\mu \in \mathcal{P}(\mathsf{X})$, $\nu \in \mathcal{P}(\mathsf{Y})$ the optimization problem

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_X \mu(dx) C(x, \pi_x). \tag{20}$$

This is a weak (i.e., nonlinear) transport problem in the sense of [33, 34] and the references therein. We now obtain a *monotonicity principle* for this problem, that is, a finitistic zeroth-order necessary optimality condition. The role model for this result is a basic fundamental fact from optimal transport theory: transport plans are optimal iff their support set is *cyclically monotone*; see [64], Section 5, for a general version of this result as well as bibliographical remarks. Recently a number of similar results have been obtained in related areas, see [11, 14, 17, 18, 35, 37, 48, 55, 56, 65]. The term "monotonicity principle" for this type of results has been suggested by Jiri Cerny and was first used in [10]. Since the present article was first posted, variants of Proposition 4.1 have been obtained in [5, 32]. We refer to [6] for a discussion of the relation of Proposition 4.1 and the concept of cyclical monotonicity in classical optimal transport as well as the connection to martingale optimal transport.

PROPOSITION 4.1. *Suppose that*:

- *Problem* (20) *is finite with optimizer* π;
- C *is jointly measurable*;
- $\mu(dx)$*-a.e. the function* $C(x, \cdot)$ *is convex and lower semicontinuous.*

Then there exists a Borel set $\Gamma \subset X$ *with* $\mu(\Gamma) = 1$ *and the following property*

$$\text{if } x, x' \in \Gamma \text{ and } m_x, m_{x'} \in \mathcal{P}(Y) \text{ satisfy } m_x + m_{x'} = \pi_x + \pi_{x'}, \text{ then}$$
$$C(x, \pi_x) + C(x', \pi_{x'}) \le C(x, m_x) + C(x', m_{x'}).$$

PROOF. Let

$$\mathcal{D} := \left\{ ((x,x'),(m_1,m_2)) \in X^2 \times \mathcal{P}(Y)^2 : \begin{array}{r} m_1 + m_2 = \pi_x + \pi_{x'}, \text{ and} \\ C(x,\pi_x) + C(x',\pi_{x'}) \\ > C(x,m_1) + C(x',m_2) \end{array} \right\},$$

which is an analytic set. By the Jankov-von Neumann uniformization theorem there is [46], Theorem 18.1, an analytically measurable function

$$D := \operatorname{proj}_{X^2}(\mathcal{D}) \ni (x,x') \mapsto \big(m_1^{(x,x')}, m_2^{(x,x')}\big) \in \mathcal{P}(Y)^2,$$

so that $(x, x', m_1^{(x,x')}, m_2^{(x,x')}) \in \mathcal{D}$. Since clearly $(x,x',m_1,m_2) \in \mathcal{D} \iff (x',x,m_2,m_1) \in \mathcal{D}$, it is possible to prove that we may actually assume that

$$\big(m_1^{(x',x)}, m_2^{(x',x)}\big) = \big(m_2^{(x,x')}, m_1^{(x,x')}\big). \tag{21}$$

Of course the set D is likewise analytic. Thus extending $(m_1^{(\cdot,\cdot)}, m_2^{(\cdot,\cdot)})$ to $(x,x') \notin D$ by setting it to $(\pi_x, \pi_{x'})$, analytic-measurability and the symmetry property (21) are preserved.

Assume that there exists $Q \in \Pi(\mu,\mu)$ such that $Q(D) > 0$. We now show that this is in conflict with the optimality of π. By considering $\frac{Q+e(Q)}{2}$, where $e(x,x') := (x',x)$, we may assume that Q is symmetric. We first define

$$\tilde{\pi}(dx,dy) := \mu(dx) \int_{x'} Q_x(dx') m_1^{(x,x')}(dy), \tag{22}$$

which is legitimate owing to the measurability precautions we have taken. We will prove:

1. $\tilde{\pi} \in \Pi(\mu, \nu)$,
2. $\int \mu(dx) C(x, \pi_x) > \int \mu(dx) C(x, \tilde{\pi}_x)$.

For (1): Evidently the first marginal of $\tilde{\pi}$ is μ. On the other hand

$$\int_x \mu(dx) \tilde{\pi}_x(dy) = \int_x \mu(dx) \int_{x'} Q_x(dx') m_1^{(x,x')}(dy) = \int_{x,x'} Q(dx, dx') m_1^{(x,x')}(dy).$$

The last quantity is equal to $\int_{x,x'} Q(dx, dx') m_2^{(x,x')}(dy)$ by symmetry of Q and (21). So

$$\begin{aligned}\int_x \mu(dx) \tilde{\pi}_x(dy) &= \int_{x,x'} Q(dx, dx') \frac{m_1^{(x,x')} + m_2^{(x,x')}}{2}(dy) \\ &= \int_{x,x'} Q(dx, dx') \frac{\pi_{x'} + \pi_x}{2}(dy) = \nu(dy),\end{aligned}$$

by definition of $m_i^{(x,x')}$ and Q. Thus, $\tilde{\pi}$ has second marginal ν.

For (2): By convexity of $C(x, \cdot)$, the symmetry of Q and (21), and by the assumption that on the Q-non negligible set D we have $C(x, \pi_x) + C(x', \pi_{x'}) > C(x, m_1^{(x,x')}) + C(x', m_2^{(x,x')})$, we obtain

$$\begin{aligned}\int_x \mu(dx) C(x, \tilde{\pi}_x) &= \int_x \mu(dx) C\Big(x, \int_{x'} Q_x(dx') m_1^{(x,x')}\Big) \\ &\le \int_x \mu(dx) \int_{x'} Q_x(dx') C(x, m_1^{(x,x')}) \\ &= \int_{x,x'} Q(dx, dx') C(x, m_1^{(x,x')}) \\ &= \int_{x,x'} Q(dx, dx') \frac{C(x, m_1^{(x,x')}) + C(x, m_2^{(x,x')})}{2} \\ &< \int_{x,x'} Q(dx, dx') \frac{C(x, \pi_x) + C(x', \pi_{x'})}{2} \\ &= \int_x \mu(dx) C(x, \pi_x).\end{aligned}$$

As expected, we have contradicted the optimality of π.

We conclude that no measure Q with the stated properties exists. By "Kellerer's lemma" [13], Proposition 2.1, which is also true for analytic sets, we obtain that D is contained in a set of the form $N \times N$ where $\mu(N) = 0$. Letting $\Gamma := N^c$, so $\Gamma \times \Gamma \subset D^c$, we easily conclude. □

We now go back to the main framework in this article. The monotonicity principle will be crucially used, under the following guise, in order to prove the results in Section 3.3. For a kernel $\pi_x(dy)$ and $\tilde{\mu}(d\tilde{x}) = \frac{1}{2}(\delta_x(d\tilde{x}) + \delta_{x'}(d\tilde{x}))$, we write $\pi_{\tilde{x}}(dy)\tilde{\mu}(d\tilde{x}) = \frac{1}{2}(\delta_x \pi_x + \delta_{x'} \pi_{x'})$.

COROLLARY 4.2. *Let π be optimal for (WOT). Then there exists $\Gamma \subset \mathbb{R}^d$ with $\mu(\Gamma) = 1$ such that*

$$\text{if } x, x' \in \Gamma, \text{ then the measure } \frac{\delta_x \pi_x + \delta_{x'} \pi_{x'}}{2} \text{ is optimal for}$$

$$\inf_{\substack{\mathrm{mean}(m_{x'}) = x', \mathrm{mean}(m_x) = x \\ (m_x + m_{x'})/2 = (\pi_x + \pi_{x'})/2}} \{\mathcal{W}_2(m_x, \gamma^d)^2 + \mathcal{W}_2(m_{x'}, \gamma^d)^2\}. \tag{23}$$

PROOF. Consider Proposition 4.1, taking $X = Y = \mathbb{R}^d$ and setting

$$C(x,m) = \mathcal{W}_2\big(m,\gamma^d\big)^2,$$

if $\mathrm{mean}(m) = x$ and $+\infty$ otherwise. It is immediate that $C(x,\cdot)$ is convex and lower semicontinuous. Taking Γ to be the μ-full set given by Proposition 4.1, the result follows. □

Observe that Problem (23) is of the same kind as (*WOT*), with initial marginal $\frac{\delta_x+\delta_{x'}}{2}$ and terminal marginal $\frac{\pi_x+\pi_{x'}}{2}$. It follows as in Theorems 2.2 and Lemma 2.4 that (23) has a continuous-time analogue, which enjoys the dynamic programming principle, and whose optimizer is a strong Markov martingale. This fact will be repeatedly used in the next part.

5. Pending proofs.

5.1. *Proofs of Theorems* 1.10 *and* 3.14.

PROOF OF THEOREM 1.10. Let $A : \mathbb{R}^d \to \mathbb{R}^d$ be in $L^2(\mu)$ and $\phi, \psi : \mathbb{R}^d \to \mathbb{R}$ be conjugate convex functions. We start by proving that

$$WT \le \int \phi\, d\nu - \int x \cdot A(x)\, d\mu + \int \mu(dx) \int \gamma^{(A(x))}(db)\psi(b), \tag{24}$$

where $\gamma^{(a)} := \delta_a * \gamma^d$. First, observe that

$$\sup_{q\in\Pi(\pi,\gamma)} \int q(dm,db)\, m\cdot b = \sup_{q\in\Pi(\pi,\gamma^{(a)})} \int q(dm,db)\, m \cdot [b-a].$$

Let us write $\Sigma := \{\{\pi_x\}_x : \mathrm{mean}(\pi_x) = x \text{ and } \int \mu(dx)\pi_x(dy) = \nu(dy)\}$. From here,

$$\begin{aligned}
WT &= \sup_{\{\pi_x\}_x\in\Sigma} \int \mu(dx) \sup_{q\in\Pi(\pi_x,\gamma^{(A(x))})} \int q(dm,db)\, m\cdot\big[b - A(x)\big] \\
&= \sup_{\{\pi_x\}_x\in\Sigma} \int \mu(dx) \Big[-\int \pi_x(dm)\, m\cdot A(x) + \sup_{q\in\Pi(\pi_x,\gamma^{(A(x))})} \int q(dm,db)\, m\cdot b \Big] \\
&= \sup_{\{\pi_x\}_x\in\Sigma} \int \mu(dx) \Big[-x\cdot A(x) + \sup_{q\in\Pi(\pi_x,\gamma^{(A(x))})} \int q(dm,db)\, m\cdot b \Big] \\
&\le -\int x\cdot A(x)\, d\mu + \sup_{\{\pi_x\}_x\in\Sigma} \int \mu(dx) \sup_{q\in\Pi(\pi_x,\gamma^{(A(x))})} \int q(dm,db)\big[\phi(m) + \psi(b)\big] \\
&= -\int x\cdot A(x)\, d\mu + \int \phi\, d\nu + \int \mu(dx) \int \gamma^{(A(x))}(db)\psi(b)
\end{aligned}$$

by the conjugacy relationship $m \cdot b \le \phi(m) + \psi(b)$ and the defining property of Σ. Hence, (24) follows.

Let now M be standard stretched Brownian motion from μ to ν in the notation of Definition 1.8 and Equation (9). By classical convex analysis arguments, or optimal transport theory, there exists ϕ, ψ convex conjugate functions such that λ^d-a.e. (γ^d-a.e.)

$$\nabla F(b)\cdot b = \phi\big(\nabla F(b)\big) + \psi(b).$$

We choose $A(x) = f_0^{-1}(x)$, which is well defined on $\mathrm{supp}(\mu)$ by Lemma 3.12.

By definition $\mu \sim M_0 = f_0(B_0) \sim f_0(\alpha)$, so

$$\int x\cdot A(x)\, d\mu(x) = \int x\cdot A(x)\, df_0(\alpha)(x) = \int f_0(x)\cdot x\, d\alpha(x) = \mathbb{E}\big[f_0(B_0)B_0\big] = \mathbb{E}[M_0\cdot B_0].$$

On the other hand,

$$\begin{aligned}\int \phi\, d\nu + \int \mu(dx) \int \gamma^{(A(x))}(db)\psi(b) &= \mathbb{E}[\phi(M_1)] + \int A(\mu)(dx) \int \gamma^{(x)}(db)\psi(b) \\ &= \mathbb{E}[\phi(M_1)] + \int A(\mu)(dx)\mathbb{E}[\psi(B_1)|B_0 = x] \\ &= \mathbb{E}[\phi(M_1)] + \mathbb{E}[\mathbb{E}[\psi(B_1)|B_0]], \\ &= \mathbb{E}[\phi(M_1) + \psi(B_1)],\end{aligned}$$

since $A(\mu) = \alpha = law(B_0)$. So the r.h.s. of (24) becomes in this case

$$\begin{aligned}\mathbb{E}[\phi(M_1) + \psi(B_1) - M_0 \cdot B_0] &= \mathbb{E}[\phi(\nabla F(B_1)) + \psi(B_1) - M_0 \cdot B_0] \\ &= \mathbb{E}[\nabla F(B_1) \cdot B_1 - M_0 \cdot B_0] \\ &= \mathbb{E}[M_1 \cdot B_1 - M_0 \cdot B_0] = \mathbb{E}\left[\int_0^1 \mathrm{tr}(\sigma_t)\, dt\right].\end{aligned}$$

Hence, Theorem 2.2 implies the optimality of M. □

We now work under Assumption 3.8, still in arbitrary dimension d.

PROOF OF THEOREM 3.14. We observe from Proposition 3.11 that the optimization problem (*WOT*) can be decomposed/disintegrated along the cells of $\mathcal{C}_{\mu,\nu}$. Therefore, optimality must only hold for $C_{\mu,\nu}(\mu)$-a.e. K for the corresponding transport problems with first and second marginals $\mu(\cdot|K)$ and $\nu(\cdot|K)$, respectively. This reduces the argument to the previous case of Theorem 1.10, and we conclude. □

5.2. *Proof of Theorem* 3.13. Although this is eventually a two-dimensional result, for the arguments we do not fix the dimension d to two unless we explicitly say so.

Let M be the unique optimizer of (MBB), where we drop the superscript $*$ for simplicity. By Theorem 2.2 this continuous-time martingale is associated to the unique two-step martingale π optimizing (*WOT*). Let ∇F^x be the optimal transport map pushing γ^d to π_x.

By Remark 2.3, we know that conditioning on $M_0 = x$ the martingale M is given by

$$M_t^x := f_t^x(B_t) \quad \text{where } f_t^x(\cdot) := \int \nabla F^x(b + \cdot)\gamma_{1-t}^d(db). \tag{25}$$

We fix $0 < t < 1$ throughout. By Lemma 3.12, we find $B_t = (f_t^x)^{-1}(M_t^x)$. We denote

$$\pi_{x,y} := law(M_1|M_0 = x, M_t = y) = \nabla F^x(\delta_{(f_t^x)^{-1}(y)} * \gamma_{1-t}^d). \tag{26}$$

Important convention: For the rest of this section, we make the convention that x, y, z denote possible values of the random variables M_0, M_t, M_1, respectively.

LEMMA 5.1. *Let g be the unique gradient of a convex function such that $g(\gamma_{1-t}^d) = \pi_{x,y}$. Then $\nabla F^x(\cdot) = g(-(f_t^x)^{-1}(y) + \cdot)$. In particular ∇F^x is uniquely determined by the family of translates of g, which we denote by*

$$\mathrm{type}(\pi_{x,y}) := \{a \mapsto g(a - r) : r \in \mathbb{R}^d\}.$$

PROOF. For $r \in \mathbb{R}^d$ write $g_r(\cdot) = g(\cdot - r)$. Then, we have $\pi_{x,y} = g(\gamma_{1-t}^d) = g_{(f_t^x)^{-1}(y)}(\delta_{(f_t^x)^{-1}(y)} * \gamma_{1-t}^d)$. Hence, both ∇F^x and $g_{(f_t^x)^{-1}(y)}$ push forward $\delta_{(f_t^x)^{-1}(y)} * \gamma_{1-t}^d$ into $\pi_{x,y}$, and both are gradients of convex functions. By the uniqueness result in Brenier's theorem, it follows that they are equal. Thus, knowing ∇F^x determines g modulo translation.

Conversely, knowing type$(\pi_{x,y})$ (i.e., the translations of g) determines ∇F^x upon finding the vector r such that $\int g(r+a)\gamma_{1-t}^d(da) = y$. □

Let

$$\pi^t := law(M_0, M_t) \tag{27}$$

and consider

$$\mathcal{C} := \{C(x)\}_x := \mathcal{C}_{\pi^t},$$

the minimal π^t-invariant measurable convex paving of Lemma 3.7. We need to show that on each cell of $\mathcal{C}$, M is a standard stretched Brownian motion, that is, on each cell $C(x)$, we need to find a convex function $F = F_{C(x)}$ such that

$$M_1^x = \nabla F\big((f_0)^{-1}(x) + B_1\big), \tag{28}$$

where f_0 and F are related as in (25). To this end, we introduce

$$A(x) := \text{type}(\pi_x) = \{a \mapsto \nabla F^x(a-r) : r \in \mathbb{R}^d\}$$

and we need to show that on each cell $A(x)$ is constant. We start by establishing a few preliminary results.

LEMMA 5.2. *If $A(x)$ is constant in each cell of $\mathcal{C}$, then M is a standard stretched Brownian motion on each of these cells.*

PROOF. Let $x \in K$ for $K \in \mathcal{C}$. For $x', x'' \in K$, as $A(x) = A(x') = A(x'')$, by definition there must exist a vectors r', r'' (possibly depending on x' and x'' respectively) such that $\nabla F^{x'}(\cdot) = \nabla F^x(\cdot - r')$ and $\nabla F^{x''}(\cdot) = \nabla F^x(\cdot - r'')$. We then set $\nabla F := \nabla F^x$. Doing this for each $K \in \mathcal{C}$, proves the claim. Indeed, decomposing $\mu = \int \mu(\cdot|K)\, dC(\mu)(K)$ as in Proposition 3.11, shows that M is a standard stretched Brownian motion (cf. Definition 1.8) on each such K. □

To make use of the previous lemma, we shall study the behaviour of the martingale M for times in $[0,t]$ and $[t,1]$. Let

$$\tilde{\pi}(dy, dz) = law(M_t, M_1).$$

For π^t from (27), we denote its disintegration w.r.t. the second marginal by $(\pi_y^t)_y$. Finally, recall $\pi_{x,y}$ from (26).

LEMMA 5.3. *For $law(M_t)$-a.e. y and π_y^t-a.e. x, we have*

$$\pi_{x,y}(dz) = \tilde{\pi}_y(dz).$$

PROOF. We have $\tilde{\pi}_y = law(M_1|M_t = y)$. On the other hand, $\pi_{x,y} = law(M_1|M_t = y, M_0 = x)$ so by Lemma 2.4 we get $\pi_{x,y}(dz) = \tilde{\pi}_y(dz)$ for $law(M_t)$-a.e. y and π_y^t-a.e. x. □

The previous lemma shows that the type of $\pi_{x,y}$ in fact only depends on y. Indeed, the same applies also for x.

LEMMA 5.4. *For μ-a.e. x and π_x^t-a.e. y we have*

$$\text{type}(\pi_{x,y}) = A(x).$$

PROOF. By Lemma 5.1, if $g \in \operatorname{type}(\pi_{x,y})$ then ∇F^x is a translate of g (the translation may depend on x, y). But this means conversely that g is a translate of ∇F^x, that is, $g \in A(x)$. Reversing the steps gives the equality. □

We finalize the proof of Theorem 3.13. In a nutshell, the key is to deal with the null sets in Lemmas 5.3-5.4. *Only from now until the end of Section 5 we assume that $d = 2$.*

PROOF OF THEOREM 3.13. Lemma 5.4 proves that for π^t-a.e. (x, y), we have $\operatorname{type}(\pi_{x,y}) = A(x)$. On the other hand, Lemma 5.3 implies that for π^t-a.e. (x, y), $\operatorname{type}(\pi_{x,y}) = D(y) := \operatorname{type}(\tilde{\pi}_y)$. By Fubini, we have

$$\pi^t\big(\{(x, y) : A(x) = D(y)\}\big) = 1.$$

We want to use this to show that $A(\cdot)$ is constant on the cells of $\mathcal{C}_{\pi^t}$. We first prove this for

$$\mathcal{C}^t := \{\operatorname{ri}\operatorname{supp}(\pi^t_x)\}_{x\in\mathbb{R}^d}.$$

By (25) and Lemma 3.12(v) we have that $\operatorname{supp}(\pi^t_x) = \operatorname{supp}(law(M_t|M_0 = x)) = \overline{\operatorname{co}}\,\nabla F^x(\mathbb{R}^d)$ is convex. As in the final part of the proof of Lemma 3.7, the martingale property implies that μ-almost surely $x \in \operatorname{ri}\overline{\operatorname{co}}\operatorname{supp}\pi^t_x = \operatorname{ri}\operatorname{supp}\pi^t_x$, and by [59], Theorem 6.3, we know $\overline{\operatorname{ri}\operatorname{supp}\pi^t_x} = \operatorname{supp}\pi^t_x$. Hence, to show that $\mathcal{C}^t$ is a candidate π^t-invariant convex paving, it remains to show that the cells of $\mathcal{C}^t$ are pairwise disjoint or equal.

By Proposition 5.5 below, there is a μ-full set of initial positions with the property that, if x, x' satisfy $\operatorname{ri}\operatorname{supp}\pi^t_x \cap \operatorname{ri}\operatorname{supp}\pi^t_{x'} \neq \varnothing$, then $A(x) = A(x')$, that is, the types of π_x and $\pi_{x'}$ coincide. This means that ∇F^x and $\nabla F^{x'}$ are translates of each other, implying that $\overline{\operatorname{co}}\,\nabla F^x(\mathbb{R}^d) = \overline{\operatorname{co}}\,\nabla F^{x'}(\mathbb{R}^d)$. From the previous paragraph, this shows $\operatorname{supp}(\pi^t_x) = \operatorname{supp}(\pi^t_{x'})$ and in particular $\operatorname{ri}\operatorname{supp}\pi^t_x = \operatorname{ri}\operatorname{supp}\pi^t_{x'}$. In one stroke, this proves that $\mathcal{C}^t$ is a (π^t-invariant) convex paving and that $A(\cdot)$ is constant on its cells.

Since $\mathcal{C}_{\pi^t}$ is finer than $\mathcal{C}^t$, this proves that $A(\cdot)$ is constant in the cells of $\mathcal{C}_{\pi^t}$ as well, and we conclude the proof by Lemma 5.2 □

The crucial Proposition 5.5 below relies on the "monotonicity principle" Proposition 4.1, and more specifically Corollary 4.2.

For the rest of this section, Γ is the μ-full set of Corollary 4.2.

PROPOSITION 5.5. *There is a μ-full set $S \subset \Gamma$ with the following property: If $x, x' \in S$ satisfy* $\operatorname{ri}\operatorname{supp}\pi^t_x \cap \operatorname{ri}\operatorname{supp}\pi^t_{x'} \neq \varnothing$, *then $A(x) = A(x')$.*

PROOF. *Step* 1: By Lemma 5.6 below, for $x, x' \in \Gamma$ we have

$$\begin{aligned}
&\dim\operatorname{ri}\operatorname{supp}\pi^t_x = \dim\operatorname{ri}\operatorname{supp}\pi^t_{x'} = \dim\big(\operatorname{ri}\operatorname{supp}\pi^t_x \cap \operatorname{ri}\operatorname{supp}\pi^t_{x'}\big)\\
&\quad\Rightarrow\quad 0 = \pi^t_x\big(\operatorname{ri}\operatorname{supp}\pi^t_x \cap \operatorname{ri}\operatorname{supp}\pi^t_{x'} \cap \{y : \pi_{x,y} \neq \pi_{x',y}\}\big)\\
&\qquad\qquad = \pi^t_{x'}\big(\operatorname{ri}\operatorname{supp}\pi^t_x \cap \operatorname{ri}\operatorname{supp}\pi^t_{x'} \cap \{y : \pi_{x,y} \neq \pi_{x',y}\}\big).
\end{aligned}\tag{29}$$

The goal is now to prove that for pairs $x, x' \in \Gamma$ the l.h.s. of (29) holds. As we will see in the final step of the proof, the r.h.s. of (29) is a strengthening of the dynamic programming principle that allows to deal with the null sets in Lemmas 5.3 and 5.4 more effectively.

Step 2: By Lemma 5.7, we know that if $x, x' \in \Gamma$, then

$$\begin{aligned}
&\dim\operatorname{ri}\operatorname{supp}\pi^t_x = \dim\operatorname{ri}\operatorname{supp}\pi^t_{x'} = 1 \text{ and } \operatorname{ri}\operatorname{supp}\pi^t_x \cap \operatorname{ri}\operatorname{supp}\pi^t_{x'} \neq \varnothing\\
&\quad\Rightarrow\quad \dim\big(\operatorname{ri}\operatorname{supp}\pi^t_x \cap \operatorname{ri}\operatorname{supp}\pi^t_{x'}\big) = 1.
\end{aligned}$$

Step 3: By Lemma 5.8, we have for $x, x' \in \Gamma$

$$\dim \operatorname{ri} \operatorname{supp} \pi_x^t = 1 \quad \text{and} \quad \operatorname{ri} \operatorname{supp} \pi_{x'}^t = \{x'\} \quad \Rightarrow \quad x' \notin \operatorname{ri} \operatorname{supp} \pi_x^t.$$

Step 4: By Lemma 5.9, we have for $x, x' \in \Gamma$

$$\dim \operatorname{ri} \operatorname{supp} \pi_x^t = 2 \quad \text{and} \quad \dim \operatorname{ri} \operatorname{supp} \pi_{x'}^t = 1 \quad \Rightarrow \quad \operatorname{ri} \operatorname{supp} \pi_x^t \cap \operatorname{ri} \operatorname{supp} \pi_{x'}^t = \varnothing.$$

Step 5: By Lemma 5.10, we have that the family $\mathcal{C}_2^t := \{\operatorname{ri} \operatorname{supp} \pi_x^t : x \in \Gamma, \dim \operatorname{ri} \operatorname{supp} \pi_x^t = 2\}$ consists of pairwise either disjoint or equal sets. As these are open sets, there can only be countable many different such sets (i.e. $|\mathcal{C}_2^t| = |\mathbb{N}|$). If $C \in \mathcal{C}_2^t$ is such that $\mu(\{x : \operatorname{ri} \operatorname{supp} \pi_x^t = C\}) = 0$, then we discard this set C from our convex paving. So we may assume, for $C \in \mathcal{C}_2^t$ that $\mu(\{x : \operatorname{ri} \operatorname{supp} \pi_x^t = C\}) > 0$. By Lemma 5.10 the set $\{x' \in C : \operatorname{ri} \operatorname{supp} \pi_{x'}^t = \{x'\}\}$ is μ-null under the assumption $\nu \ll \lambda^2$. Hence, for each of the countable many $C \in \mathcal{C}_2^t$ we can discard a μ-null set such that on a possibly smaller but still μ-full subset of Γ, which we keep calling Γ for simplicity, we have

$$x, x' \in \Gamma, \quad \dim \operatorname{ri} \operatorname{supp} \pi_x^t = 2 \quad \text{and} \quad \{x'\} = \operatorname{ri} \operatorname{supp} \pi_{x'}^t \quad \Rightarrow \quad x' \notin \operatorname{ri} \operatorname{supp} \pi_x^t.$$

Final Step: By Steps 3, 4 and 5, we may assume that, for x, x' in a μ-full set, we have

$$\operatorname{ri} \operatorname{supp} \pi_x^t \cap \operatorname{ri} \operatorname{supp} \pi_{x'}^t \neq \varnothing \quad \Rightarrow \quad \dim \operatorname{ri} \operatorname{supp} \pi_x^t = \dim \operatorname{ri} \operatorname{supp} \pi_{x'}^t.$$

In this situation, if the common dimension in the r.h.s. is equal to one, by Step 2 also the dimension of the intersection in the l.h.s. is equal to one. On the other hand, if the common dimension in the r.h.s. is two, then automatically the dimension of the intersection is two (as an open convex set in $\mathbb{R}^2$). In any case, call $d^{(x,x')}$ this common dimension.[8] We find ourselves in the setting of (29), so by Step 1 we must have with $I := \operatorname{ri} \operatorname{supp} \pi_x^t \cap \operatorname{ri} \operatorname{supp} \pi_{x'}^t$

$$\pi_x^t\big(I \cap \{y : \pi_{x,y} \neq \pi_{x',y}\}\big) = \pi_{x'}^t\big(I \cap \{y : \pi_{x,y} \neq \pi_{x',y}\}\big) = 0. \tag{30}$$

Let E be the μ-full set of Lemma 5.4 and call $S := E \cap \Gamma$. For suitable sets Y, Y' with $\pi_x^t(\mathsf{Y}) = \pi_{x'}^t(\mathsf{Y}') = 1$ it then holds for $x, x' \in S$ that

$$\operatorname{type}(\pi_{x,y}) = A(x) \quad \forall y \in \mathsf{Y},$$
$$\operatorname{type}(\pi_{x',y}) = A(x') \quad \forall y \in \mathsf{Y}'.$$

By Lemma 3.12, π_x^t is equivalent to $d^{(x,x')}$-dimensional Lebesgue measure on the $d^{(x,x')}$-dimensional open set $\operatorname{ri} \operatorname{supp} \pi_x^t$. Since $\operatorname{ri} \operatorname{supp} \pi_x^t \cap \operatorname{ri} \operatorname{supp} \pi_{x'}^t$ is a $d^{(x,x')}$-dimensional open subset it is also of positive $d^{(x,x')}$-Lebesgue measure. Then it is also of positive π_x^t-measure. Thus $\operatorname{ri} \operatorname{supp} \pi_x^t \cap \operatorname{ri} \operatorname{supp} \pi_{x'}^t \cap \mathsf{Y}$ has positive π_x^t-measure, and positive $d^{(x,x')}$-Lebesgue measure. But then again by Lemma 3.12 this same set must have positive $\pi_{x'}^t$-measure. We conclude that $I \cap \mathsf{Y} \cap \mathsf{Y}'$ has likewise positive $\pi_{x'}^t$-measure. The symmetric argument shows that the same set has positive π_x^t-measure. But by (30) the set $\{y : \pi_{x,y} = \pi_{x',y}\}$ is π_x^t-full in I. It follows that

$$I \cap \mathsf{Y} \cap \mathsf{Y}' \cap \{y : \pi_{x,y} = \pi_{x',y}\},$$

has positive π_x^t-measure, and by the same token it has positive $\pi_{x'}^t$-measure. In particular,

$$\mathsf{Y} \cap \{y : \pi_{x,y} = \pi_{x',y}\} \cap \mathsf{Y}' \neq \varnothing,$$

and taking y in this intersection we find

$$A(x) = \operatorname{type}(\pi_{x,y}) = \operatorname{type}(\pi_{x',y}) = A(x'). \qquad \square$$

[8] Actually the case $d^{(x,x')} = 2$ is settled by Lemma 5.10, but we prefer to give a general argument.

LEMMA 5.6. *$\Gamma \times \Gamma$ is disjoint with the set*

$$\left\{ (x,x') : \begin{array}{l} \dim \operatorname{ri} \operatorname{supp} \pi_x^t = \dim \operatorname{ri} \operatorname{supp} \pi_{x'}^t = \dim(\operatorname{ri} \operatorname{supp} \pi_x^t \cap \operatorname{ri} \operatorname{supp} \pi_{x'}^t), \\ \textit{and either } \pi_x^t(\operatorname{ri} \operatorname{supp} \pi_x^t \cap \operatorname{ri} \operatorname{supp} \pi_{x'}^t \cap \{\pi_{x,y} \neq \pi_{x',y}\}) > 0, \\ \textit{or } \pi_{x'}^t(\operatorname{ri} \operatorname{supp} \pi_x^t \cap \operatorname{ri} \operatorname{supp} \pi_{x'}^t \cap \{\pi_{x,y} \neq \pi_{x',y}\}) > 0 \end{array} \right\}.$$

PROOF. Take $x, x' \in \Gamma$. By Corollary 4.2, the 2-step martingale $\frac{\delta_x \pi_x + \delta_{x'} \pi_{x'}}{2}$ is optimal for (23). Consider its continuous-time analogue, that is, the martingale which started at x equals M^x and started at x' equals $M^{x'}$ (cf. (25)) and both starting points have equal probability. We denote this continuous time martingale by $M^{(x,x')}$. By construction, $law(M_t^{(x,x')} | M_0^{(x,x')} = x) = \pi_x^t$ and likewise for x'. Similarly $law(M_1^{(x,x')} | M_0^{(x,x')} = x, M_t^{(x,x')} = y) = \pi_{x,y}$ and the same holds for x' instead of x. By optimality of $\frac{\delta_x \pi_x + \delta_{x'} \pi_{x'}}{2}$ also $M^{(x,x')}$ is optimal for the continuous-time analogue of (23), then by dynamic programming (Lemma 2.4), we obtain sets Y, Y' such that

$$\pi_{x,y} = law(M_1^{(x,x')} | M_t^{(x,x')} = y) \quad \text{for } y \in \mathsf{Y} \text{ with } \pi_x^t(\mathsf{Y}) = 1,$$

$$\pi_{x',y} = law(M_1^{(x,x')} | M_t^{(x,x')} = y) \quad \text{for } y \in \mathsf{Y}' \text{ with } \pi_{x'}^t(\mathsf{Y}') = 1.$$

The important point is that this is "pointwise" in $M_0^{(x,x')} \in \{x, x'\}$.

Now assume further that

$$\dim \operatorname{ri} \operatorname{supp} \pi_x^t = \dim \operatorname{ri} \operatorname{supp} \pi_{x'}^t = \dim(\operatorname{ri} \operatorname{supp} \pi_x^t \cap \operatorname{ri} \operatorname{supp} \pi_{x'}^t),$$

and call $d^{(x,x')}$ this common dimension. By Lemma 3.12, we have that π_x^t and $\pi_{x'}^t$ restricted to $\operatorname{ri} \operatorname{supp} \pi_x^t \cap \operatorname{ri} \operatorname{supp} \pi_{x'}^t$ are equivalent to $d^{(x,x')}$-dimensional Lebesgue measure restricted to this same set. We write

$$I := \operatorname{ri} \operatorname{supp} \pi_x^t \cap \operatorname{ri} \operatorname{supp} \pi_{x'}^t.$$

Necessarily $\mathsf{Y} \cap I$ is π_x^t-full in I, and therefore also $\pi_{x'}^t$-full in I. But then $\mathsf{Y} \cap I \cap \mathsf{Y}'$ is $\pi_{x'}^t$-full in I too. Inverting the roles of x, x' this set must also be π_x^t-full in I. We conclude

$$\pi_x^t(I \setminus (\mathsf{Y} \cap \mathsf{Y}')) = \pi_{x'}^t(I \setminus (\mathsf{Y} \cap \mathsf{Y}')) = 0.$$

But on $\mathsf{Y} \cap \mathsf{Y}'$ we have $\pi_{x,y} = law(M_1^{(x,x')} | M_t^{(x,x')} = y) = \pi_{x',y}$, so

$$\pi_x^t(I \cap \{\pi_{x,y} \neq \pi_{x',y}\}) = \pi_{x'}^t(I \cap \{\pi_{x,y} \neq \pi_{x',y}\}) = 0.$$

This concludes the proof. □

LEMMA 5.7. *Call $\mathcal{V}$ the set of pairs (x, x') such that $\operatorname{ri} \operatorname{supp} \pi_x^t$ and $\operatorname{ri} \operatorname{supp} \pi_{x'}^t$ have dimension 1 and intersect in a singleton. Then,*

$$\mathcal{V} \cap (\Gamma \times \Gamma) = \varnothing.$$

PROOF. We use the same notation as in the proof of Lemma 5.6, and assume $(x, x') \in \mathcal{V}$.

By construction, $law(M_t^{(x,x')} | M_0^{(x,x')} = x) = \pi_x^t$ and likewise for x'. So $M_t^{(x,x')}$ conditioned to start at x, or at x', live respectively in line segments exactly intersecting in a single point $p \in \mathbb{R}^2$. By Lemma 3.12, the paths of these martingales (restricted to times in $[0, t]$) evolve in different "space-time" strips that only intersect along the line $L := \{(p, s) : s \geq 0\}$. Let $\tau := \inf\{s : (M_s^{(x,x')}, s) \in L\}$. It follows that $0 < \tau < t$ on a nonnegligible set. The law of τ conditioned on the starting point of $M^{(x,x')}$ is equivalent to Lebesgue measure on (0, 1). The reason is that this is true for 1-dimensional Brownian motion, and thanks to Lemma 3.12

the martingale $M^{(x,x')}$ conditioned to start say in x, is a one-dimensional Brownian motion after a continuous strictly increasing time-change. Hence, for any set $E \subset (0,1)$ of positive Lebesgue measure we have $\mathbb{P}(\tau \in E \cap (0,t) | M_0^{(x,x')} = x) > 0$ and $\mathbb{P}(\tau \in E \cap (0,t) | M_0^{(x,x')} = x') > 0$. Thus, we observe that the law of $M_t^{(x,x')}$ given $\{M_s^{(x,x')} : s \le \tau \wedge t\}$ is different from the law of $M_t^{(x,x')}$ given $M_{\tau\wedge t}^{(x,x')}$. Indeed, when $\tau < t$ (equivalently when $M_{\tau\wedge t}^{(x,x')} = p$) and $\tau \in E$, one cannot for sure say in which of the aforementioned strips the martingale will continue to evolve. On the contrary, by observing $\{M_s^{(x,x')} : s \le \tau \wedge t\}$ and on $\{\tau < t\} \cap \{\tau \in E\}$, such a strip is completely determined. Therefore $M^{(x,x')}$ fails to have the strong Markov property. But then it cannot be optimal between its marginals, by Corollary 2.5, and so neither can be $\frac{\delta_x \pi_x + \delta_{x'}\pi_{x'}}{2}$ optimal for (23). We conclude by Corollary 4.2 that $(x,x') \notin \Gamma \times \Gamma$. □

LEMMA 5.8. *We have*

$$\big\{(x,x') : \dim \operatorname{ri} \operatorname{supp} \pi_x^t = 1, \operatorname{ri} \operatorname{supp} \pi_{x'}^t = x', x' \in \operatorname{ri} \operatorname{supp} \pi_x^t\big\} \cap (\Gamma \times \Gamma) = \varnothing. \tag{31}$$

PROOF. The proof is very similar to that of Lemma 5.7. Let (x,x') belong to the leftmost set in (31). Using the same notation, $M^{(x,x')}$ is a martingale which evolves in a space-time strip if started at x, and otherwise is a constant equal to x'. We denote τ the first hitting time of $\{(x',s) : s \ge 0\}$. Since the martingale lives in a strip, we have that $\tau < t$ has probability strictly greater than $1/2$. The strong Markov property of $M^{(x,x')}$ is destroyed at $\tau \wedge t$, since the knowledge of the past up to $\tau \wedge t$ reveals whether the martingale is constant or not thereafter. As before, by Corollary 2.5 and Corollary 4.2, $M^{(x,x')}$ cannot be optimal and $(x,x') \notin \Gamma \times \Gamma$. □

LEMMA 5.9. *The set* $\Gamma \times \Gamma$ *is disjoint with*

$$\big\{(x,x') : \dim \operatorname{ri} \operatorname{supp} \pi_x^t = 2, \dim \operatorname{ri} \operatorname{supp} \pi_{x'}^t = 1, \operatorname{ri} \operatorname{supp} \pi_x^t \cap \operatorname{ri} \operatorname{supp} \pi_{x'}^t \neq \varnothing\big\}.$$

PROOF. As in the previous proofs, with $M^{(x,x')}$ we associate $\tau = \inf\{s : M_s^{(x,x')} \in \operatorname{ri} \operatorname{supp} \pi_{x'}^t\}$. Taking (x,x') in the leftmost set, it is tedious but not difficult to see that

$$law\big((M_\tau^{(x,x')}, \tau) | \tau \le t, M_0^{(x,x')} = x\big) \quad \text{and} \quad law\big((M_U^{(x,x')}, U) | M_0^{(x,x')} = x'\big),$$

are equivalent to Lebesgue measure on $\operatorname{ri} \operatorname{supp} \pi_{x'}^t \times [0,t]$, where U is uniformly distributed on $[0,t]$ and independent of everything. The point is that there is a common "space-time" set E charged by the two aforementioned laws. But the behaviour of $M_t^{(x,x')}$ conditioned on its past up to $\tau \wedge t$ is drastically depending on its starting position (see, e.g., whether it will evolve in a one- or two- dimensional set), whereas if, for example, we knew $(M_\tau^{(x,x')}, \tau) \in E$ then this does not reveal the dimension of the set where the martingale will continue to evolve. This contradicts the strong Markov property and we conclude as before. □

LEMMA 5.10. *The family*

$$\mathcal{C}_2^t := \big\{\operatorname{ri} \operatorname{supp} \pi_x^t : x \in \Gamma, \dim \operatorname{ri} \operatorname{supp} \pi_x^t = 2\big\},$$

consists of open sets which are pairwise disjoint or equal. Assuming $\nu \ll \lambda^2$, *we have*

$$C \in \mathcal{C}_2^t \quad \text{and} \quad \mu\big(\{x : \operatorname{ri} \operatorname{supp} \pi_x^t = C\}\big) > 0$$
$$\Longrightarrow \mu\big(\{x' \in C : \operatorname{ri} \operatorname{supp} \pi_{x'}^t = \{x'\}\}\big) = 0.$$

PROOF. Let Λ consist of all (x,x') such that:

- $\dim \operatorname{ri} \operatorname{supp} \pi_x^t = 2 = \dim \operatorname{ri} \operatorname{supp} \pi_{x'}^t$,
- $\operatorname{ri} \operatorname{supp} \pi_x^t \neq \operatorname{ri} \operatorname{supp} \pi_{x'}^t$, and
- $\operatorname{ri} \operatorname{supp} \pi_x^t \cap \operatorname{ri} \operatorname{supp} \pi_{x'}^t \neq \varnothing$.

As before we can show that Λ cannot intersect $\Gamma \times \Gamma$. We do not give the argument, to avoid repetition, but mention that instead of contradicting the strong Markov property it suffices to contradict the regular Markov property. We conclude the first assertion.

Now let $C \in \mathcal{C}_2^t$ such that $\mu(\{x : \operatorname{ri} \operatorname{supp} \pi_x^t = C\}) > 0$, $K := \{x' \in C : \operatorname{ri} \operatorname{supp} \pi_{x'}^t = \{x'\}\}$, and suppose $\mu(K) > 0$. We think of K as a nonnegligible cloud of dots where the martingale M stays frozen. Since $M_0 \in K \Rightarrow M_1 \in K$, we have $\upsilon(K) > 0$ and by assumption $\lambda^2(K) > 0$. It follows that $\{M_t \in K\}$ is nonnegligible, no matter if M has started on K or on $\{x : \operatorname{ri} \operatorname{supp} \pi_x^t = C\}$ at time zero (in the latter case, by Lemma 3.12). Since both sets of initial conditions are nonnegligible, we contradict the regular Markov property of M. Indeed, on $\{M_t \in K\}$ the behaviour of M after t is drastically different depending on the starting condition at time zero being in K or $\{x : \operatorname{ri} \operatorname{supp} \pi_x^t = C\}$. This contradicts the optimality of M, and we conclude $\mu(K) = 0$. □

PROOF OF THEOREM 3.1. A martingale is an optimizer of (MBB) between the marginals μ, ν iff it is an optimizer on each individual component. Hence we may assume wlog that (μ, ν) is irreducible. If M is a s^{2}Bm, then it is an optimizer by Theorem 1.10. It remains to establish that an optimizer M / π is a s^{2}Bm. As discussed after Lemma 3.7, $\mathcal{C}_\pi$ consists precisely of one irreducible component which is given by $I := \operatorname{ri} \operatorname{co} \operatorname{supp} \nu$. We can then establish the conclusion of Proposition 5.5 using essentially the same arguments with the simplification that Steps 4, 5 are irrelevant and hence no further assumptions on ν are required. We then conclude as in the proof of Theorem 3.13. □

6. Further optimality properties.

Let $\mathbb{T} := \{0 = t_0 \le t_1 \le \cdots \le t_{n-1} \le t_n = 1\} \subset [0, 1]$ be a finite subgrid. Suppose M is a standard stretched Brownian motion from μ to ν, so $M_t = f_t(B_t)$ for a Brownian motion starting with some distribution α; see (9). Then

$$M_{t_0} = f_{t_0}(B_{t_0}), \qquad M_{t_1} = f_{t_1}(B_{t_1}), \qquad \ldots, \qquad M_{t_n} = f_{t_n}(B_{t_n}).$$

Denote $\nu^{\mathbb{T}} := \mathit{law}(M_{t_0}, M_{t_1}, \ldots, M_{t_n})$, the projection of ν onto the time indices in $\mathbb{T}$, and $\gamma^{\mathbb{T}} := \mathit{law}(B_{t_0}, B_{t_1}, \ldots, B_{t_n})$. Finally, consider the *adapted map*

$$[\mathbb{R}^d]^{n+1} \ni (b_0, \ldots, b_n) \mapsto f^{\mathbb{T}}(b_0, \ldots, b_n) := (f_{t_0}(b_0), f_{t_1}(b_1), \ldots, f_{t_n}(b_n)) \in [\mathbb{R}^d]^{n+1}.$$

It follows that

$$f^{\mathbb{T}}(\gamma^{\mathbb{T}}) = \nu^{\mathbb{T}}.$$

Each component of $f^{\mathbb{T}}$ is *increasing* in the sense that it is the gradient of a convex function. Such a map is an example of an "increasing triangular transformation," as in [20]. It can also be understood in terms of increasingness w.r.t. lexicographical order in case $\nu^{\mathbb{T}}$ has a density. In a sense properly explained in [4], $f^{\mathbb{T}}$ sends $\gamma^{\mathbb{T}}$ into $\nu^{\mathbb{T}}$ in a canonical respect. optimal way: see respect. Proposition 5.6 and Corollary 2.10 therein.

Since this is true no matter the subgrid $\mathbb{T}$, we are entitled to think of M as an *adapted increasing rearrangement* of the Brownian motion into a martingale with given initial an final laws. Also, the aforementioned canonical/optimal character of such rearrangements should translate into the optimality of M as obtained in the previous section, and vice-versa. We now make this heuristics rigorous.

Problem (MBB) is equivalent to

$$\inf_{M_t = M_0 + \int_0^t \sigma_s \, dB_s, M_0 \sim \mu, M_1 \sim \nu} \mathbb{E}[\operatorname{tr}\langle M - B \rangle_1], \tag{32}$$

since $\mathbb{E}[\mathrm{tr}\langle M-B\rangle_1]=\mathbb{E}[\mathrm{tr}\langle M\rangle_1]+\mathbb{E}[\mathrm{tr}\langle B\rangle_1]-2\mathbb{E}[\int_0^1 \mathrm{tr}(\sigma_t)\,dt]$, and the first two quantities in the r.h.s. do not depend on the concrete coupling (M,B). This also proves that for (32) it is irrelevant where B is started. We now want to formulate a transport problem between laws of stochastic processes which is compatible with (32). For ease of notation, we denote

$$\Omega := C([0,1];\mathbb{R}^d).$$

DEFINITION 6.1. A causal coupling between $\mathbb{P}$ and $\mathbb{Q}$ is a probability measure π on $\Omega\times\Omega$ with first and second marginals $\mathbb{P}$ and $\mathbb{Q}$ respectively, and satisfying the additional property

$$\forall t, \forall A\in\mathcal{F}_t:\quad (\Omega\ni x\mapsto \pi_x(A)\in[0,1]) \text{ is } (\mathbb{P},\mathcal{F}_t)\text{-measurable,} \tag{33}$$

where $\mathcal{F}$ is the $\mathbb{P}$-completed canonical filtration and π_x is a regular conditional probability of π w.r.t. the first marginal. We denote $\Pi_c(\mathbb{P},\mathbb{Q})$ the set of all such π. We also denote $\Pi_{bc}(\mathbb{P},\mathbb{Q})=\{\pi\in\Pi_c(\mathbb{P},\mathbb{Q}):e(\pi)\in\Pi_c(\mathbb{Q},\mathbb{P})\}$ for $e(x,y)=(y,x)$, the set of bicausal couplings.

We refer to [1, 3, 4, 49] for more on this definition. In what follows, we write $(\omega,\bar\omega)$ for a generic element in $\Omega\times\Omega$.

LEMMA 6.2. *Let $\mathbb{P}$ and $\mathbb{Q}$ be martingale laws, and $\pi\in\Pi_{bc}(\mathbb{P},\mathbb{Q})$. Then the canonical process on $\Omega\times\Omega$ is a π-martingale in its own filtration.*

PROOF. One can easily see that under π we have:

- $\{\bar\omega_s:0\le s\le t\}$ is π-conditionally independent from $\{\omega_s:0\le s\le 1\}$ given $\{\omega_s:0\le s\le t\}$,
- $\{\omega_s:0\le s\le t\}$ is π-conditionally independent from $\{\bar\omega_s:0\le s\le 1\}$ given $\{\bar\omega_s:0\le s\le t\}$,

by bicausality. The first property above, for $T>t$, implies

$$\mathbb{E}^\pi[\omega_T|\{\omega_s,\bar\omega_s,s\le t\}]=\mathbb{E}^\pi[\omega_T|\{\omega_s,s\le t\}]=\mathbb{E}^{\mathbb{P}}[\omega_T|\{\omega_s,s\le t\}]=\omega_t.$$

The second property implies similarly $\mathbb{E}^\pi[\bar\omega_T|\{\omega_s,\bar\omega_s,s\le t\}]=\bar\omega_t$, so we conclude. □

Let us denote by $\mathbb{W}$ Wiener measure (started at zero) on Ω. The next lemma establishes the crucial connection between standard stretched Brownian motion and the present *causal transport* setting.

LEMMA 6.3. *Let M be standard stretched Brownian motion from μ to ν, with $M_t=M_0+\int_0^t\sigma_s\,dB_s$. Then*

$$law(B-B_0,M)\in\Pi_{bc}(\mathbb{W},law(M)).$$

More generally, if M is stretched Brownian motion and B is as in Remark 2.3, the same conclusion holds.

PROOF. Let M be standard stretched Brownian motion from μ to ν. By Lemma 3.12 there is an orthogonal projection P such that $M_t=\tilde f_t(\tilde B_t)$, where $\tilde B_t=PB_t$. By the same result, the filtrations of M and $\tilde B$ coincide. This shows that the coupling $law(B-B_0,M)$ is causal from $\mathbb{W}$ to $law(M)$. For the reverse causality, it suffices to observe that $\{\tilde B_{t+h}-\tilde B_t: h\ge 0\}$ is independent from $\{B_s:s\le t\}$, so in particular given $\{M_s:s\le t\}$ we have that $\{B_s-B_0:s\le t\}$ and $\{M_s:s\le 1\}$ are independent. The case of M a stretched Brownian motion is similar, taking B independent of M_0 and upon conditioning on the latter random variable. □

We can now put the pieces together to obtain optimality of (standard) stretched Brownian motion in the sense of trajectorial laws. Let us fix a refining sequence of partition P_n of $[0, 1]$ in order to define the quadratic variation $\langle\cdot\rangle$ pathwise on $C([0, 1]; \mathbb{R}^d)$ in the usual manner, namely

$$\omega \mapsto \langle\omega\rangle_1^{i,j} := \lim_{n\to\infty} \sum_{t_m \in P_n} (\omega^i_{t_{m+1}} - \omega^i_{t_m})(\omega^j_{t_{m+1}} - \omega^j_{t_m}),$$

when the limit exist, and otherwise $+\infty$. We then consider

$$\inf_{\substack{\mathbb{Q}\in \mathsf{M}^c(\mu,\nu)\\ \pi\in\Pi_{bc}(\mathbb{W},\mathbb{Q})}} \mathbb{E}^\pi\big[\mathrm{tr}\langle\omega - \bar\omega\rangle_1\big], \tag{34}$$

where $\mathsf{M}^c(\mu, \nu)$ denotes the set of laws of continuous martingales indexed by $[0, 1]$ starting in μ and terminating in ν.

PROPOSITION 6.4. *Problems (32) and (34) are equivalent. In particular, let M^* be the optimizer of the former, that is, stretched Brownian motion. Then $\mathbb{Q}^* := law(M^*)$ is optimal for the latter.*

PROOF. Let $\mathbb{Q}$, π be feasible for (34). Since

$$\begin{aligned}\mathbb{E}^\pi\big[\mathrm{tr}\langle\omega - \bar\omega\rangle_1\big] &= \mathbb{E}^{\mathbb{W}}\big[\mathrm{tr}\langle\omega\rangle_1\big] + \mathbb{E}^{\mathbb{Q}}\big[\mathrm{tr}\langle\bar\omega\rangle_1\big] - 2\mathbb{E}^\pi\big[\mathrm{tr}\langle\omega,\bar\omega\rangle_1\big]\\ &= \mathbb{E}^{\mathbb{W}}\big[|\omega_1|^2 - |\omega_0|^2\big] + \mathbb{E}^{\mathbb{Q}}\big[|\bar\omega_1|^2 - |\bar\omega_0|^2\big] - 2\mathbb{E}^\pi\big[\mathrm{tr}\langle\omega,\bar\omega\rangle_1\big],\end{aligned}$$

we can equivalently maximize $\mathbb{E}^\pi[\mathrm{tr}\langle\omega,\bar\omega\rangle_1]$ in (34), rather than minimizing $\mathbb{E}^\pi[\mathrm{tr}\langle\omega - \bar\omega\rangle_1]$. However by Lemma 6.2 the canonical process is a π-martingale so

$$\mathbb{E}^\pi\big[\mathrm{tr}\langle\omega,\bar\omega\rangle_1\big] = \mathbb{E}^\pi[\omega_1\cdot\bar\omega_1] = \mathbb{E}^\pi\big[\mathbb{E}^\pi[\omega_1\cdot\bar\omega_1|\bar\omega_0]\big],$$

by the product formula and as $\omega_0 = 0$ under π. Denoting $\pi_x = law_{\mathbb{Q}}(\bar\omega_1|\bar\omega_0 = x)$ and $q_x = law_\pi((\bar\omega_1, \omega_1)|\bar\omega_0 = x)$ we have that the first marginal of q_x is π_x and the second one is γ^d. Indeed, by bicausality $\pi - law(\omega_1|\omega_0, \bar\omega_0) = \pi - law(\omega_1|\omega_0) = \gamma^d$, so in particular $\pi - law(\omega_1|\bar\omega_0) = \gamma^d$. Therefore

$$\begin{aligned}\mathbb{E}^\pi\big[\mathrm{tr}\langle\omega,\bar\omega\rangle_1\big] &= \int \mu(dx)\int q^x(dm, db)\, m\cdot b\\ &\le \int \mu(dx) \sup_{q\in\Pi(\pi_x,\gamma^d)} \int q(dm, db)\, m\cdot b.\end{aligned} \tag{35}$$

By Theorem 2.2, we conclude that the value of (32) is greater or equal than that of (34). Let M^* be the optimizer of (32) (equiv. of (MBB)). By Remark 2.3 M^* is precisely built via attaining the r.h.s. of (35) when maximizing over kernels π^x. By the final part of Lemma 6.3, we may build a bicausal coupling π so that in (35) we have equality. This proves that Problems (32) and (34) have the same value and that $law(M^*)$ is optimal for the latter. □

REMARK 6.5. The discrete-time version of Problem (34) would have shown, in light of [4], that the optimal way to send a Gaussian random walk into a martingale is through the Knothe–Rosenblatt rearrangement (the unique increasing bicausal triangular transformation between its marginals). This is in tandem with the first paragraphs of the present part (once we switched to increments $b_i - b_{i-1}$). Via Proposition 6.4 we know that stretched Brownian motion attains Problem (34). Hence, one can arguably describe stretched Brownian motion as the canonical/optimal Knothe–Rosenblatt rearrangement of Brownian motion with prescribed initial and final marginals.

Acknowledgments. We thank Dario Trevisan for stimulating discussions at the outset of this project. Likewise we thank two anonymous referees for their comments.

The second author was supported by FWF-grant Y00782.

The third author was partially supported by the Deutsche Forschungsgemeinschaft through the CRC 1060 *The Mathematics of Emergent Effects* and the *Hausdorff Center for Mathematics* and by the Vienna Science and Technology Fund (WWTF) through project VRG17-005.

REFERENCES

[1] ACCIAIO, B., BACKHOFF-VERAGUAS, J. and ZALASHKO, A. (2020). Causal optimal transport and its links to enlargement of filtrations and continuous-time stochastic optimization. *Stochastic Process. Appl.* **130** 2918–2953. MR4080734 https://doi.org/10.1016/j.spa.2019.08.009

[2] 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. MR3050280 https://doi.org/10.1007/978-3-642-32160-3_1

[3] BACKHOFF-VERAGUAS, J., BEIGLBÖCK, M., EDER, M. and PICHLER, A. (2017). Fundamental properties of process distances. arXiv e-prints.

[4] BACKHOFF-VERAGUAS, J., BEIGLBÖCK, M., LIN, Y. and ZALASHKO, A. (2017). Causal transport in discrete time and applications. *SIAM J. Optim.* **27** 2528–2562. MR3738324 https://doi.org/10.1137/16M1080197

[5] BACKHOFF-VERAGUAS, J., BEIGLBÖCK, M. and PAMMER, G. (2019). Existence, duality, and cyclical monotonicity for weak transport costs. *Calc. Var. Partial Differential Equations* **58** Paper No. 203, 28. MR4029731 https://doi.org/10.1007/s00526-019-1624-y

[6] BACKHOFF-VERAGUAS, J. and PAMMER, G. (2019). Stability of martingale optimal transport and weak optimal transport. Preprint. Available at arXiv:1904.04171.

[7] BALDER, E. J. (2000). Lectures on Young measure theory and its applications in economics. *Rend. Istit. Mat. Univ. Trieste* **31** 1–69. MR1798830

[8] BASS, R. F. (1983). Skorokhod imbedding via stochastic integrals. In *Seminar on Probability, XVII*. *Lecture Notes in Math.* **986** 221–224. Springer, Berlin. MR0770414 https://doi.org/10.1007/BFb0068318

[9] BEIGLBÖCK, M., COX, A., HUESMANN, M. and KALLBLAD, S. (2017). Measure-valued martingales and optimality of bass-type solutions to the Skorokhod embedding problem. Preprint. Available at arXiv:1708.07071.

[10] BEIGLBÖCK, M., COX, A. M. G. and HUESMANN, M. (2017). Optimal transport and Skorokhod embedding. *Invent. Math.* **208** 327–400. MR3639595 https://doi.org/10.1007/s00222-016-0692-2

[11] BEIGLBÖCK, M., COX, A. M. G. and HUESMANN, M. (2020). The geometry of multi-marginal Skorokhod embedding. *Probab. Theory Related Fields* **176** 1045–1096. MR4087489 https://doi.org/10.1007/s00440-019-00935-z

[12] BEIGLBÖCK, M., EDER, M., ELGERT, C. and SCHMOCK, U. (2018). Geometry of distribution-constrained optimal stopping problems. *Probab. Theory Related Fields* **172** 71–101. MR3851830 https://doi.org/10.1007/s00440-017-0805-x

[13] BEIGLBÖCK, M., GOLDSTERN, M., MARESCH, G. and SCHACHERMAYER, W. (2009). Optimal and better transport plans. *J. Funct. Anal.* **256** 1907–1927. MR2498564 https://doi.org/10.1016/j.jfa.2009.01.013

[14] BEIGLBÖCK, M. and GRIESSLER, C. (2019). A land of monotone plenty. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **19** 109–127. MR3923841

[15] BEIGLBÖCK, M., HENRY-LABORDÈRE, P. and PENKNER, F. (2013). Model-independent bounds for option prices—a mass transport approach. *Finance Stoch.* **17** 477–501. MR3066985 https://doi.org/10.1007/s00780-013-0205-8

[16] BEIGLBÖCK, M., HUESMANN, M. and STEBEGG, F. (2016). Root to Kellerer. In *Séminaire de Probabilités XLVIII*. *Lecture Notes in Math.* **2168** 1–12. Springer, Cham. MR3618124

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

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

[19] BENAMOU, J.-D. 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 https://doi.org/10.1090/conm/226/03232

[20] BOGACHEV, V. I., KOLESNIKOV, A. V. and MEDVEDEV, K. V. (2005). Triangular transformations of measures. *Mat. Sb.* **196** 3–30. MR2144274 https://doi.org/10.1070/SM2005v196n03ABEH000882

[21] BOUCHARD, B. and NUTZ, M. (2015). Arbitrage and duality in nondominated discrete-time models. *Ann. Appl. Probab.* **25** 823–859. MR3313756 https://doi.org/10.1214/14-AAP1011

[22] BRENIER, Y. (1987). Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.* **305** 805–808. MR0923203

[23] BRENIER, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** 375–417. MR1100809 https://doi.org/10.1002/cpa.3160440402

[24] CAMPI, L., LAACHIR, I. and MARTINI, C. (2017). Change of numeraire in the two-marginals martingale transport problem. *Finance Stoch.* **21** 471–486. MR3626622 https://doi.org/10.1007/s00780-016-0322-2

[25] DE MARCH, H. and TOUZI, N. (2019). Irreducible convex paving for decomposition of multidimensional martingale transport plans. *Ann. Probab.* **47** 1726–1774. MR3945758 https://doi.org/10.1214/18-AOP1295

[26] DOLINSKY, Y. and SONER, H. M. (2014). Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Related Fields* **160** 391–427. MR3256817 https://doi.org/10.1007/s00440-013-0531-y

[27] FONTBONA, J., GUÉRIN, H. and MÉLÉARD, S. (2010). Measurability of optimal transportation and strong coupling of martingale measures. *Electron. Commun. Probab.* **15** 124–133. MR2643592 https://doi.org/10.1214/ECP.v15-1534

[28] 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. MR3161649 https://doi.org/10.1214/13-AAP925

[29] GANGBO, W. and MCCANN, R. J. (1996). The geometry of optimal transportation. *Acta Math.* **177** 113–161. MR1440931 https://doi.org/10.1007/BF02392620

[30] GHOUSSOUB, N., KIM, Y.-H. and LIM, T. (2016). Optimal skorokhod embedding for radially symmetric marginals in general dimension. Preprint.

[31] GHOUSSOUB, N., KIM, Y.-H. and LIM, T. (2019). Structure of optimal martingale transport plans in general dimensions. *Ann. Probab.* **47** 109–164. MR3909967 https://doi.org/10.1214/18-AOP1258

[32] GOZLAN, N. and JUILLET, N. (2020). On a mixture of Brenier and Strassen theorems. *Proc. Lond. Math. Soc.* (3) **120** 434–463. MR4008375 https://doi.org/10.1112/plms.12302

[33] GOZLAN, N., ROBERTO, C., SAMSON, P.-M., SHU, Y. and TETALI, P. (2018). Characterization of a class of weak transport-entropy inequalities on the line. *Ann. Inst. Henri Poincaré Probab. Stat.* **54** 1667–1693. MR3825894 https://doi.org/10.1214/17-AIHP851

[34] 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. MR3706606 https://doi.org/10.1016/j.jfa.2017.08.015

[35] GRIESSLER, C. (2013). A variational principle for martingale optimal transport. Preprint.

[36] GRIESSLER, C. (2018). C-cyclical monotonicity as a sufficient criterion for optimality in the multimarginal Monge–Kantorovich problem. *Proc. Amer. Math. Soc.* **146** 4735–4740. MR3856141 https://doi.org/10.1090/proc/14129

[37] GUO, G., TAN, X. and TOUZI, N. (2016). On the monotonicity principle of optimal Skorokhod embedding problem. *SIAM J. Control Optim.* **54** 2478–2489. MR3549873 https://doi.org/10.1137/15M1025268

[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. MR3449312 https://doi.org/10.1214/14-AAP1084

[39] HENRY-LABORDÈRE, P. and TOUZI, N. (2016). An explicit martingale version of the one-dimensional Brenier theorem. *Finance Stoch.* **20** 635–668. MR3519164 https://doi.org/10.1007/s00780-016-0299-x

[40] HIRSCH, F., PROFETA, C., ROYNETTE, B. and YOR, M. (2011). *Peacocks and Associated Martingales, with Explicit Constructions. Bocconi & Springer Series* **3**. Springer, Milan. MR2808243 https://doi.org/10.1007/978-88-470-1908-9

[41] HOBSON, D. and NEUBERGER, A. (2012). Robust bounds for forward start options. *Math. Finance* **22** 31–56. MR2881879 https://doi.org/10.1111/j.1467-9965.2010.00473.x

[42] HUESMANN, M. and TREVISAN, D. (2019). A Benamou–Brenier formulation of martingale optimal transport. *Bernoulli* **25** 2729–2757. MR4003563 https://doi.org/10.3150/18-BEJ1069

[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. MR3655851 https://doi.org/10.1214/16-AAP1191

[44] KANTOROVITCH, L. (1942). On the translocation of masses. *C. R. (Dokl.) Acad. Sci. URSS* **37** 199–201. MR0009619

[45] KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. *Graduate Texts in Mathematics* **113**. Springer, New York. MR1121940 https://doi.org/10.1007/978-1-4612-0949-2

[46] KECHRIS, A. S. (1995). *Classical Descriptive Set Theory*. *Graduate Texts in Mathematics* **156**. Springer, New York. MR1321597 https://doi.org/10.1007/978-1-4612-4190-4

[47] KELLERER, H. G. (1973). Integraldarstellung von Dilationen. In *Transactions of the Sixth Prague Conference on Information Theory*, *Statistical Decision Functions*, *Random Processes* (*Tech*. *Univ*., *Prague*, 1971; *Dedicated to the Memory of Antonín Špaček*) 341–374. MR0356249

[48] KIM, Y.-H. and PASS, B. (2014). A general condition for Monge solutions in the multi-marginal optimal transport problem. *SIAM J*. *Math*. *Anal*. **46** 1538–1550. MR3190751 https://doi.org/10.1137/130930443

[49] LACKER, D. (2018). Dense sets of joint distributions appearing in filtration enlargements, stochastic control, and causal optimal transport. arXiv e-prints.

[50] LÉONARD, C. (2014). A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin*. *Dyn*. *Syst*. **34** 1533–1574. MR3121631 https://doi.org/10.3934/dcds.2014.34.1533

[51] LOWTHER, G. (2008). Fitting martingales to given marginals. arXiv e-prints.

[52] LOWTHER, G. (2009). Limits of one-dimensional diffusions. *Ann*. *Probab*. **37** 78–106. MR2489160 https://doi.org/10.1214/08-AOP397

[53] MCCANN, R. J. (1994). *A Convexity Theory for Interacting Gases and Equilibrium Crystals*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Princeton University. MR2691540

[54] MONGE, G. (1781). Memoire sur la theorie des deblais et des remblais. *Histoire de L'académie Royale des Sciences de Paris*.

[55] NUTZ, M. and STEBEGG, F. (2018). Canonical supermartingale couplings. *Ann*. *Probab*. **46** 3351–3398. MR3857858 https://doi.org/10.1214/17-AOP1249

[56] NUTZ, M., STEBEGG, F. and TAN, X. (2020). Multiperiod martingale transport. *Stochastic Process*. *Appl*. **130** 1568–1615. MR4058283 https://doi.org/10.1016/j.spa.2019.05.010

[57] OBŁÓJ, J. and SIORPAES, P. (2017). Structure of martingale transports in finite dimensions. Preprint. Available at arXiv:1702.08433.

[58] 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*. *Lecture Notes in Math*. **2137** 227–247. Springer, Cham. MR3444301 https://doi.org/10.1007/978-3-319-18585-9_11

[59] ROCKAFELLAR, R. T. (1997). *Convex Analysis*. *Princeton Landmarks in Mathematics*. Princeton Univ. Press, Princeton, NJ. MR1451876

[60] SANTAMBROGIO, F. (2015). *Optimal Transport for Applied Mathematicians*: *Calculus of Variations*, *PDEs*, *and Modeling*. *Progress in Nonlinear Differential Equations and Their Applications* **87**. Birkhäuser/Springer, Cham. MR3409718 https://doi.org/10.1007/978-3-319-20828-2

[61] STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann*. *Math*. *Stat*. **36** 423–439. MR0177430 https://doi.org/10.1214/aoms/1177700153

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

[63] VILLANI, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483 https://doi.org/10.1090/gsm/058

[64] VILLANI, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454 https://doi.org/10.1007/978-3-540-71050-9

[65] ZAEV, D. A. (2015). On the Monge–Kantorovich problem with additional linear constraints. *Mat*. *Zametki* **98** 664–683. MR3438523 https://doi.org/10.4213/mzm10896