Martingale Transports and Monge Maps

Marcel Nutz* Ruodu Wang† Zhenyuan Zhang‡

July 1, 2024

Abstract

It is well known that martingale transport plans between marginals $\mu \neq \nu$ are never given by Monge maps—with the understanding that the map is over the first marginal μ, or forward in time. Here, we change the perspective, with surprising results. We show that any distributions μ, ν in convex order with ν atomless admit a martingale coupling given by a Monge map over the *second* marginal ν. Namely, we construct a particular coupling called the barcode transport. Much more generally, we prove that such "backward Monge" martingale transports are dense in the set of all martingale couplings, paralleling the classical denseness result for Monge transports in the Kantorovich formulation of optimal transport. Various properties and applications are presented, including a refined version of Strassen's theorem and a mimicking theorem where the marginals of a given martingale are reproduced by a "backward deterministic" martingale, a remarkable type of process whose current state encodes its whole history.

Keywords martingale transport; backward Monge map; Strassen's theorem
AMS 2010 Subject Classification 60G42; 49N05; 60E15

1 Introduction

Martingale optimal transport was introduced by Beiglböck et al. (2013) in the discrete-time setting and Galichon et al. (2014) in continuous time. Since then, it has been an area of vigorous research thanks to its rich structures, connections with mathematical finance (see Hobson (2011) and Henry-Labordère (2017) for surveys) and the optimal Skorokhod embedding problem (see Beiglböck et al. (2017) and the literature thereafter), and analogies with classical transport theory (e.g., Beiglböck and Juillet (2016), Beiglböck et al. (2017)). Given probability measures μ, ν on $\mathbb{R}$, a transport plan (or transport, or coupling) is the joint distribution of a random vector (X, Y) with $X \overset{\text{law}}{\sim} \mu$ and $Y \overset{\text{law}}{\sim} \nu$. It is a martingale transport (MT) if in addition $\mathbb{E}[Y|X] = X$; that is, if (X, Y) is a one-period martingale. We denote the set of transports by $\Pi(\mu, \nu)$ and its subset of martingale

*Departments of Statistics and Mathematics, Columbia University, USA. Email: mnutz@columbia.edu. Research supported by NSF grants DMS-1812661, DMS-2106056.

†Department of Statistics and Actuarial Science, University of Waterloo, Canada. Email: wang@uwaterloo.ca. Research supported by NSERC grant RGPIN-2018-03823 and Canada Research Chairs CRC-2022-00141.

‡Department of Mathematics, Stanford University, USA. Email: zzy@stanford.edu

transports by $\mathcal{M}(\mu,\nu)$. Strassen's theorem states that $\mathcal{M}(\mu,\nu)$ is nonempty if and only if μ,ν are in convex order, denoted $\mu \leqslant_{\mathrm{cx}} \nu$. See Section 2 below for detailed definitions.

In classical transport theory (without the martingale constraint), much attention has been devoted to transport plans given by Monge maps (transport maps); i.e., transports (X,Y) where $Y = g(X)$ for some measurable function $g : \mathbb{R} \to \mathbb{R}$, or equivalently $\pi \in \Pi(\mu,\nu)$ of the form $\pi = (\mathrm{id}_{\mathbb{R}}, g)_{\#}\mu$ where $\#$ denotes pushforward. The existence of such Monge transports typically requires μ to be atomless (unless ν has atoms satisfying particular conditions). Under this natural requirement, it is known that the optimizers for numerous important optimal transport problems are indeed Monge, for instance, the quantile (or Fréchet–Hoeffding) coupling which minimizes the square-distance cost. Moreover, the set of all Monge transports is known to be weakly dense in $\Pi(\mu,\nu)$, which leads to the equivalence of the Kantorovich and Monge formulations of optimal transport: for any continuous and suitably integrable cost function c, the value $\inf_{\pi\in\Pi(\mu,\nu)} \int c \,\mathrm{d}\pi$ remains the same if the infimum is only taken over the subset of Monge transports. See for instance Ambrosio (2003, Theorem 9.3) and Pratelli (2007, Theorem B), as well as the monographs Villani (2003, 2009) and Santambrogio (2015) for further background and numerous references.

In the literature on martingale transport, Monge transports have been mentioned mostly[1] to state that they are uninteresting: because any deterministic martingale is constant, a martingale transport can only be of the form $(X, g(X))$ if g is the identity map. In that case, $\mu = \nu$, and (X,X) is the only martingale coupling. In the martingale setting, one may think automatically along the forward-in-time direction $\mu \to \nu$ that is natural for adapted stochastic processes. In this paper, we change the perspective and look backward in time: nothing obvious precludes the existence of non-trivial Monge maps over the second marginal; that is, martingales (X,Y) of the form $(f(Y),Y)$, or martingale laws $\pi = (f, \mathrm{id}_{\mathbb{R}})_{\#}\nu$. The name "backward Monge martingale transport" seems descriptive but lengthy, and as the "forward" version is uninteresting, we simply say *Monge martingale transport* (MMT). Their collection is denoted $\mathcal{M}_M(\mu,\nu)$.

This paper is dedicated to the theory of Monge martingale transports as well as their implications. Given marginals $\mu \leqslant_{\mathrm{cx}} \nu$, it is not obvious if an MMT exists—apart from the trivial fact that atoms in ν often preclude the existence of any Monge transport (martingale or not) from ν to μ. Of all the martingale couplings that have been described in the literature, we are not aware of one that is Monge for reasonably generic marginals. Assuming that ν is atomless, we prove in Theorem 2.1 that $\mathcal{M}_M(\mu,\nu)$ is never empty: we construct a particular MMT that we call the *barcode transport*, a name derived from its pictorial representation (see Figure 1 on page 6).[2] The basic idea is to decompose the marginals μ and ν into countably many pieces (the *bars* of the barcode) that can

[1] A notable exception, kindly pointed out to us by D. Kramkov, is the work of Kramkov and Xu (2022) on a Kyle-type equilibrium model of insider trading. There, a particular two-dimensional martingale (X,Y) is shown to be of the form $(X_1,X_2) = (f_1(Y_1,Y_2), f_2(Y_1,Y_2))$ and that property is crucial for the interpretation of (X_1,X_2) as the total order and price, respectively, of the equilibrium. In this problem, the law ν of Y is prescribed whereas the law μ of X is endogenous to the equilibrium. Remarkably, in our notation, $\mathcal{M}(\mu,\nu)$ is shown to be a singleton for that particular μ, which suggests that μ has quite distinct properties (cf. Theorem 2.5).

[2] Strictly speaking, the barcode transport is constructed using the *left-curtain transport*, whereas using the *right-curtain transport* would yield a different barcode transport. However, for notational convenience, we simply call it a barcode transport instead of a left-barcode transport.

be coupled by MMTs more easily, and then aggregate. As an auxiliary result, we provide a novel structural description (Proposition 2.2) of the *left-curtain transport* π_{lc} prominently introduced by Beiglböck and Juillet (2016); we show in particular that π_{lc} is Monge if the first marginal has more mass than the second marginal at any point of its support. While this condition is of course quite special, we can always construct a decomposition of the original marginals μ, ν such as to satisfy the condition on each "bar".

The aforementioned construction is rather particular and one may wonder whether the barcode transport is just an isolated curious example. Our main result (Theorem 2.3) states that the set $\mathcal{M}_M(\mu,\nu)$ of Monge martingale transports is weakly dense in the set $\mathcal{M}(\mu,\nu)$ of all martingale transports. This shows that there are many MMTs (for typical marginals) and, paralleling the aforementioned results in classical transport theory, that the value $\inf_{\pi\in\mathcal{M}(\mu,\nu)} \int c \,\mathrm{d}\pi$ of a martingale optimal transport problem remains the same if the infimum is only taken over the subset of Monge transports (Corollary 2.4), for any continuous and suitably integrable c. We mention that a quite different (and maybe less direct) parallel was established in the Skorokhod embedding problem: Beiglböck at al. (2021) show that the stopping times of the Brownian filtration that embed a given distribution are weakly dense in the set of randomized stopping times embedding the distribution.

While the above shows that standard optimal transport problems cannot distinguish $\mathcal{M}_M(\mu,\nu)$ from $\mathcal{M}(\mu,\nu)$, a natural characterization of $\mathcal{M}_M(\mu,\nu)$ within $\Pi(\mu,\nu)$ will be given in terms of generalized (or "weak") transport costs in the sense of Gozlan et al. (2017). These are cost functions depending not only on the origin and destination points of a transport but directly on the kernel (conditional distribution) of the coupling. We show in Proposition 3.8 that $\mathcal{M}_M(\mu,\nu)$ is the set of minimizers for a class of such problems, in particular (with obvious abuse of notation)

$$\mathcal{M}_M(\mu,\nu) = \operatorname*{arg\,min}_{(X,Y)\in\Pi(\mu,\nu)} \mathbb{E}\left[\mathbb{E}[Y|X]^2 - \mathbb{E}[X|Y]^2\right] - 2\mathbb{E}[XY].$$

We also discuss in detail the uniqueness of MMT (Theorem 2.5) which is equivalent to the uniqueness of MT, and happens only in very particular circumstances that we characterize in terms of so-called shadows. If both marginals μ, ν are atomless, the only case with uniqueness is $\mu = \nu$.

Several applications of MMTs are presented. The first is a refinement of Strassen's theorem on $\mathbb{R}$ (Theorem 3.1) saying that if random variables X and Y on an atomless probability space satisfy $X \leqslant_{\mathrm{cx}} Y$, then there exists a random variable $X' \overset{\mathrm{law}}{=} X$ on the same space such that $X' = \mathbb{E}[Y|X']$ is a martingale. Thus Y is preserved, whereas the usual Strassen's theorem only guarantees a martingale (X', Y') with the same marginal distributions but no particular relation to the original random variables (X, Y).

Going further in a similar direction, we develop a mimicking theorem (in the sense of Gyöngy (1986)) with a class of martingales that we call *backward deterministic.* These are processes $(X_n)_{n\in\mathbb{N}}$ where $(X_j)_{j=1}^n$ is $\sigma(X_n)$-measurable. We may see this as a strengthening of the Markov property where the current state X_n already encodes the whole history $(X_j)_{j=1}^n$. A non-recombining binary tree is a good illustration. Our mimicking theorem (Corollary 3.6) states that given a martingale $(Y_n)_{n\in\mathbb{N}}$ with atomless marginals, there exists a backward deterministic martingale $(X_n)_{n\in\mathbb{N}}$ such

that $X_n \overset{\text{law}}{=} Y_n$ for all n.

The remainder of this paper is organized as follows. Section 2 collects the main results on Monge martingale transports, as well as the result on the left-curtain transport to be used in the existence proof. In Section 3 we discuss the applications regarding Strassen's theorem, the mimicking theorem with backward deterministic martingales, and the characterization of $\mathcal{M}_M(\mu,\nu)$ via generalized optimal transport. Section 4 contains the proofs for the main results stated in Section 2. We conclude with some comments and open problems in Section 5.

2 Main results

Let $\mathcal{P}(\mathbb{R})$ denote the set of Borel probability measures on $\mathbb{R}$ with finite first moment. We say that $\mu,\nu \in \mathcal{P}(\mathbb{R})$ are in *convex order*, denoted $\mu \leqslant_{\text{cx}} \nu$, if $\int \phi \,\mathrm{d}\mu \leqslant \int \phi \,\mathrm{d}\nu$ for any convex function $\phi : \mathbb{R} \to \mathbb{R}$. This implies that μ,ν have the same mean. We use the same notation for unnormalized finite measures; in that case μ,ν must also have the same total mass. Occasionally we write $X \leqslant_{\text{cx}} Y$ for random variables X, Y to indicate that their laws are in convex order. Recall from the Introduction that $\Pi(\mu,\nu)$ denotes the set of couplings, $\mathcal{M}(\mu,\nu)$ the subset of martingale couplings, and $\mathcal{M}_M(\mu,\nu)$ the further subset of (backward) Monge martingale transports. We say that a measure π is supported on a set A if A^c is a π-nullset. The topological support (that is, the smallest such set A that is closed) may be different.

Our first result yields the existence of a Monge martingale transport when the second marginal ν is atomless. More generally, when ν has atoms, it establishes a martingale transport that is (backward) Monge outside the atoms—the Monge property on the atoms is typically not achievable even without the martingale constraint.

Theorem 2.1 (Existence)**.** *Let $\mu,\nu \in \mathcal{P}(\mathbb{R})$ satisfy $\mu \leqslant_{\text{cx}} \nu$. There exists $\pi \in \mathcal{M}(\mu,\nu)$ and a Borel function $h : \mathbb{R} \to \mathbb{R}$ such that $\pi(T_{\text{rg}} \cup T_{\text{atom}}) = 1$, where*

(i) $T_{\text{rg}} = \{(h(y), y) : y \in \mathbb{R}\}$;

(ii) $T_{\text{atom}} = \{(x,y) : \nu(\{y\}) > 0\}$.

In particular, if ν is atomless, π is a Monge martingale transport.

To prove Theorem 2.1, we will explicitly construct a coupling called the *barcode transport*. As mentioned in the Introduction, the basic idea is to decompose the marginals into countably many mutually singular parts—the bars of the barcode; cf. Figure 1 below—tailored such that the left-curtain transport π_{lc} for each part is Monge outside of the atoms of ν. We thus need criteria for π_{lc} to be Monge, and that is the purpose of the next result.

To state the definition of π_{lc} given by Beiglböck and Juillet (2016), we write $\mu \leqslant_{\text{E}} \nu$ for finite measures μ,ν with finite first moment if $\int \phi \,\mathrm{d}\mu \leqslant \int \phi \,\mathrm{d}\nu$ for any nonnegative convex function $\phi : \mathbb{R} \to \mathbb{R}$. If μ and ν have the same total mass, this is equivalent to $\mu \leqslant_{\text{cx}} \nu$, but a quite different example is that $\mu \leqslant \nu$ (set-wise) implies $\mu \leqslant_{\text{E}} \nu$. Given $\mu \leqslant_{\text{E}} \nu$, the *shadow* $S^\nu(\mu)$ of μ in ν is defined

as $S^\nu(\mu) = \min\{\eta : \mu \leqslant_{\mathrm{cx}} \eta \leqslant \nu\}$, where the minimum is taken in the partial order $\leqslant_{\mathrm{cx}}$. Intuitively, the shadow is formed by mapping each μ-particle into ν while greedily dispersing its mass as little as possible. See Beiglböck and Juillet (2016, Lemma 4.6) for the wellposedness of $S^\nu(\mu)$.

Given $\mu \leqslant_{\mathrm{cx}} \nu$, the *left-curtain transport* $\pi_{\mathrm{lc}} \in \mathcal{M}(\mu,\nu)$ is uniquely defined by the property that it transports $\mu|_{(-\infty,x]}$ to its shadow $S^\nu(\mu|_{(-\infty,x]})$ for every $x \in \mathbb{R}$. It can be considered as the martingale analogue of the quantile coupling with respect to the convex order. The "forward" structure of π_{lc} has been analyzed in detail by Beiglböck and Juillet (2016) as well as Henry-Labordère and Touzi (2016) and Hobson and Norgilas (2019); see also Section 4.1. The following result describes the structure from the backward perspective and may be of independent interest. It states that in general, π_{lc} is supported on three sets: the reverse graph (or antigraph) S_{rg} of a function $h : \mathbb{R} \to \mathbb{R}$, the diagonal S_{diag}, and the atomic part S_{atom}. For the proof of Theorem 2.1, we will only use the second assertion, namely that if $\mathrm{d}\mu/\mathrm{d}(\mu+\nu) \geqslant 1/2$ μ-a.e., the reverse graph can also capture the mass on S_{diag}.

Proposition 2.2 (Structure of π_{lc})**.** *Let $\mu \leqslant_{\mathrm{cx}} \nu$. There exists a Borel function $h : \mathbb{R} \to \mathbb{R}$ such that the left-curtain transport π_{lc} satisfies $\pi_{\mathrm{lc}}(S_{\mathrm{rg}} \cup S_{\mathrm{diag}} \cup S_{\mathrm{atom}}) = 1$, where*

(i) $S_{\mathrm{rg}} = \{(h(y), y) : y \in \mathbb{R}\}$;

(ii) $S_{\mathrm{diag}} = \{(x,x) : x \in \mathbb{R}\}$;

(iii) $S_{\mathrm{atom}} = \{(x,y) : \nu(\{y\}) > 0\}$.

If $\mathrm{d}\mu/\mathrm{d}(\mu+\nu) \geqslant 1/2$ μ-a.e., then $\pi_{\mathrm{lc}}(S_{\mathrm{rg}} \cup S_{\mathrm{atom}}) = 1$ for some Borel h. In particular, if in addition ν is atomless, then $\pi_{\mathrm{lc}} \in \mathcal{M}_M(\mu,\nu)$.

The second assertion is not directly a consequence of the first part as the function h may need to be redefined. We refer to Section 4.1 for further comments on π_{lc}.

Figure 1 illustrates the barcode transport and the left-curtain transport for Gaussian marginals. We observe that the left-curtain transport is not Monge in this case, and this arises due to the mass on S_{diag} represented in light-gray over a subset of $\{\mathrm{d}\mu/\mathrm{d}(\mu+\nu) < 1/2\}$.

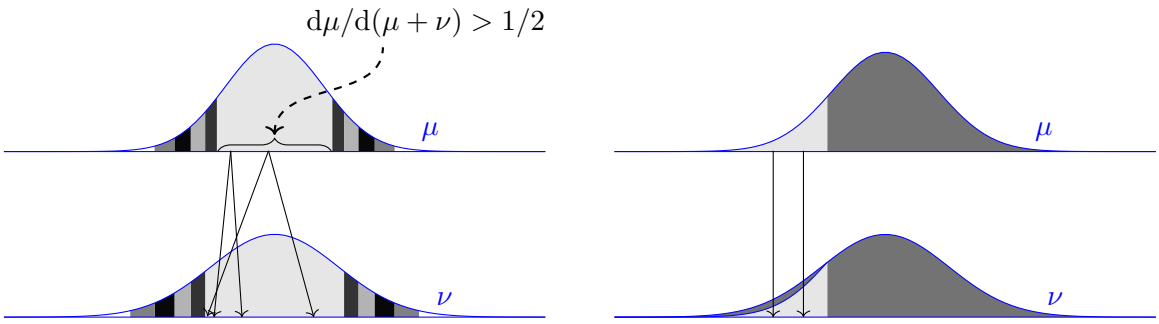
We continue with our main result, showing that the set $\mathcal{M}_M(\mu,\nu)$ of Monge martingale transports is surprisingly rich.

Theorem 2.3 (MMTs are dense)**.** *Let $\mu \leqslant_{\mathrm{cx}} \nu$ with ν atomless. Then $\mathcal{M}_M(\mu,\nu)$ is weakly dense in $\mathcal{M}(\mu,\nu)$. If μ is discrete, it is also dense for the ∞-Wasserstein topology.*

The proof is significantly more involved than the existence argument, hence we defer a sketch to Section 4.3. As a consequence of Theorem 2.3, we obtain the equivalence of the Kantorovich and (backward) Monge formulations for martingale optimal transport.

Corollary 2.4. *Let $\mu \leqslant_{\mathrm{cx}} \nu$ with ν atomless. If $c : \mathbb{R}^2 \to \mathbb{R}$ is continuous with $|c(x,y)| \leqslant a(x)+b(y)$ for some $a \in L^1(\mu)$ and $b \in L^1(\nu)$, then*

$$\inf_{\pi \in \mathcal{M}_M(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} c(x,y)\,\pi(\mathrm{d}x,\mathrm{d}y) = \inf_{\pi \in \mathcal{M}(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} c(x,y)\,\pi(\mathrm{d}x,\mathrm{d}y).$$



(a) The barcode transport (b) The left-curtain transport

Figure 1: Comparison of the barcode transport and the left-curtain transport for Gaussian marginals. (a) The barcode transport consists of a collection of left-curtain transports represented by different shades. The map h follows the reverse of the indicated arrows. (b) The left-curtain transport is the identity on the light-gray area and does not admit a (backward) Monge map there.

The final theorem of this section characterizes the uniqueness of MMT; that is, when $\mathcal{M}_M(\mu,\nu)$ is a singleton. We can already see from the denseness result in Theorem 2.3 that this is equivalent to $\mathcal{M}(\mu,\nu)$ being a singleton (a more direct proof will be given in Section 4). In terms of the marginals, uniqueness turns out to depend on the atoms of μ and their shadows.

Theorem 2.5 (Uniqueness)**.** *Let $\mu \leqslant_{\mathrm{cx}} \nu$ with ν atomless. The following are equivalent:*

(i) The MT from μ to ν is unique.

(ii) The MMT from μ to ν is unique.

(iii) Let $\mu_a := \sum_{j\in\mathbb{N}} a_j\delta_{x_j}$ be the atomic part of μ, where $\{x_j\}_{j\in\mathbb{N}}$ are distinct. Then the shadows $S^\nu(a_j\delta_{x_j})$, $j\in\mathbb{N}$ are mutually singular and $\mu-\mu_a = \nu - \sum_{j\in\mathbb{N}} S^\nu(a_j\delta_{x_j})$.

Remark 2.6. As kindly pointed out by an anonymous referee, a further equivalent statement for Theorem 2.5 can be formulated using the concept of irreducible components. For probability measures μ,ν on $\mathbb{R}$ satisfying $\mu \leqslant_{\mathrm{cx}} \nu$, we let $u_\mu : \mathbb{R}\to\mathbb{R}$, $x\mapsto \int_{\mathbb{R}} |y-x|\mu(\mathrm{d}y)$ be the potential function of μ, and similarly define u_ν. Let $(I_k)_{1\leqslant k\leqslant N}$ be the (open) components of $\{u_\mu < u_\nu\}$ where $N\in\mathbb{N}\cup\{\infty\}$, and let $I_0 = \mathbb{R}\setminus\bigcup_{k\geqslant 1} I_k$. Define $\mu_k = \mu|_{I_k}$, so that $\mu = \sum_{k\geqslant 0}\mu_k$; this is called the irreducible decomposition of μ (which depends on ν). By Theorem A.4 of Beiglböck and Juillet (2016), there exists a unique decomposition $\nu = \sum_{k\geqslant 0}\nu_k$ such that $\mu_0=\nu_0$ and $\mu_k \leqslant_{\mathrm{cx}} \nu_k$ for all k, and any $\pi\in\mathcal{M}(\mu,\nu)$ transports μ_k to ν_k for $k\in\mathbb{N}$ and μ_0 to ν_0 via the identity transport. Then we have the following equivalent condition for uniqueness of the MT:

(iv) Each μ_k, $k\in\mathbb{N}$ in the irreducible decomposition of μ is concentrated on a singleton.

Indeed, (iv) implies the MT on each irreducible component is unique, and hence (i); the structure (iii) implies that $\mu_k = a_k\delta_{x_k}$, $k\in\mathbb{N}$ and $\mu_0 = \nu-\mu_a$ define the irreducible decomposition of μ, implying (iv). The more general irreducible decomposition for probability measures on $\mathbb{R}^d$ instead of $\mathbb{R}$ will be discussed in Section 5.

As a special case of Theorem 2.5, if μ and ν are both atomless, uniqueness is equivalent to $\mu = \nu$. A nontrivial example with uniqueness is illustrated in Figure 2.

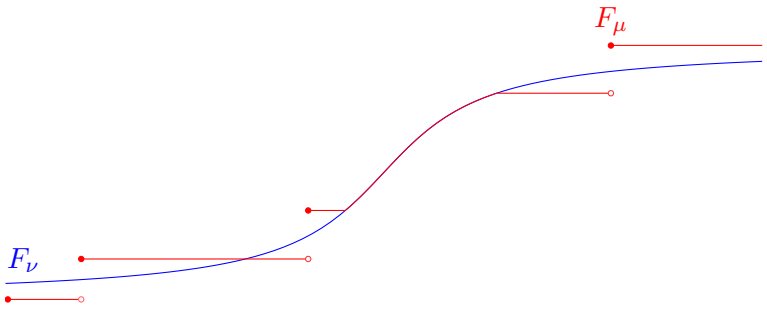


Figure 2: Distribution functions of μ, ν where the MMT (and MT) from μ to ν is unique

We conclude with simple examples illustrating subtleties that can arise when ν is not atomless.

Example 2.7 (MT exists; MMT does not)**.** Let μ and ν be two-point distributions satisfying $\mu \leqslant_{\mathrm{cx}} \nu$. Then there is a unique MT, as there is a unique distribution on two distinct points with a given mean. On the other hand, there is no MMT unless $\mu = \nu$. In general, if μ, ν are discrete and $\mathrm{card}(\cdot)$ denotes the cardinality of the support, the existence of an MMT implies $(2\,\mathrm{card}((\mu - \nu)_+)) \vee \mathrm{card}(\mu) \leqslant \mathrm{card}(\nu)$.

Example 2.8 (MMT is unique; MT is not)**.** Let μ be uniform on $\{2, 5\}$ and ν be uniform on $\{0, 3, 4, 7\}$. The unique MMT is given by transporting $\{2\}$ to $\{0, 4\}$ and $\{5\}$ to $\{3, 7\}$, while it is easy to see that there exist many MTs.

3 Applications and further properties

3.1 Refinement of Strassen's theorem

The celebrated Strassen's theorem (Strassen (1965, Theorem 8)) shows that if two random variables X and Y satisfy $X \leqslant_{\mathrm{cx}} Y$, then we can build $X' \overset{\text{law}}{=} X$ and $Y' \overset{\text{law}}{=} Y$ on another probability space such that $X' = \mathbb{E}[Y'|X']$. Theorem 2.1 gives rise to the following refinement where X' is built on the original space supporting Y and there is no need for an auxiliary random variable Y'.

Theorem 3.1 (Refinement of Strassen's theorem)**.** *Let $X \leqslant_{\mathrm{cx}} Y$ be real-valued random variables on an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$. There exists a random variable X' on $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying $X' \overset{\text{law}}{=} X$ and $X' = \mathbb{E}[Y|X']$.*

Proof. Let $\{y_n : n \in I\} \subseteq \mathbb{R}$ be the atoms of the distribution of Y, where I is a countable set. As $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless, we can construct for each $n \in I$ a uniform random variable U_{y_n} on $\{Y = y_n\}$ equipped with the restrictions of $\mathcal{F}$ and $\mathbb{P}$. It suffices to construct a random variable X' that is $\sigma(Y, U_{y_n}, n \in I)$-measurable such that $X' = \mathbb{E}[Y|X']$. By Theorem 2.1, there exists a coupling π of X, Y supported on the union of a reverse graph $\{(h(y), y) : y \in \mathbb{R}\}$ and $\bigcup_{n\in I}\{(x, y_n) : x \in \mathbb{R}\}$. Let

F_{y_n} be the cdf of the conditional distribution of π given $Y = y_n$ and let $F_{y_n}^{\leftarrow}$ denote its left-continuous inverse. We define

$$X'(\omega) := \begin{cases} h(Y(\omega)) & \text{if } \omega \notin \bigcup_{n\in I}\{Y = y_n\}; \\ F_{y_n}^{\leftarrow}(U_{y_n}(\omega)) & \text{if } \omega \in \{Y = y_n\} \text{ for some } n \in I. \end{cases}$$

Then X' is $\sigma(Y, U_{y_n}, n \in I)$-measurable and the joint distribution of (X', Y) is π. □

Remark 3.2. Theorem 2.1 implies the existence of an MMT when the second marginal ν is atomless. This statement can also be recovered from Theorem 3.1 by taking $\mathcal{F} = \sigma(Y)$, so that X' must be a function of Y.

A different way of framing those relations is to introduce a partial order on $\mathcal{P}(\mathbb{R})$ via MMT. Noting that the convex order can be defined as $\mu \leqslant_{\text{cx}} \nu \Leftrightarrow \mathcal{M}(\mu, \nu) \neq \emptyset$, let us write $\mu \leqslant_{\text{MM}} \nu$ if $\mathcal{M}_M(\mu, \nu) \neq \emptyset$. This is indeed a partial order.

Lemma 3.3. *The binary relation $\leqslant_{\text{MM}}$ is a partial order on $\mathcal{P}(\mathbb{R})$. Moreover, $\leqslant_{\text{MM}}$ implies $\leqslant_{\text{cx}}$.*

Proof. Clearly $\leqslant_{\text{MM}}$ implies $\leqslant_{\text{cx}}$, hence reflexivity and antisymmetry of $\leqslant_{\text{MM}}$ follow from those of $\leqslant_{\text{cx}}$. To show transitivity, let $\eta \leqslant_{\text{MM}} \mu$ and $\mu \leqslant_{\text{MM}} \nu$. By definition, there exist functions g and f such that given $Y \overset{\text{law}}{\sim} \nu$ and $X \overset{\text{law}}{\sim} \mu$, we have $\mathbb{E}[Y|f(Y)] = f(Y) \overset{\text{law}}{\sim} \mu$ and $\mathbb{E}[X|g(X)] = g(X) \overset{\text{law}}{\sim} \eta$. In particular, setting $X := f(Y)$,

$$\mathbb{E}[Y|g \circ f(Y)] = \mathbb{E}\left[\mathbb{E}[Y|f(Y)]|g \circ f(Y)\right] = \mathbb{E}[X|g(X)] = g(X) = g \circ f(Y),$$

showing that $g \circ f$ is an MMT for (η, ν). □

Proposition 3.4. *Let $\nu \in \mathcal{P}(\mathbb{R})$ and $Y \overset{\text{law}}{\sim} \nu$. Then*

$$\{\mu \in \mathcal{P}(\mathbb{R}) : \mu \leqslant_{\text{MM}} \nu\} = \big\{\textit{law of } \mathbb{E}[Y|f(Y)] : \textit{ } f \textit{ measurable}\big\}.$$

If ν is atomless, then furthermore

$$\{\mu \in \mathcal{P}(\mathbb{R}) : \mu \leqslant_{\text{MM}} \nu\} = \big\{\textit{law of } \mathbb{E}[Y|X] : \textit{ } X \in L^0\big\} = \{\mu \in \mathcal{P}(\mathbb{R}) : \mu \leqslant_{\text{cx}} \nu\}$$

where L^0 is the set of random variables on the same space as Y.

Proof. The second part follows directly from Theorem 3.1. For the first part, the inclusion "$\subseteq$" is immediate from the definition of $\leqslant_{\text{MM}}$. To see "$\supseteq$", let $\mu \overset{\text{law}}{\sim} Z := \mathbb{E}[Y|f(Y)]$ for some measurable function f. As Z is $\sigma(f(Y))$-measurable, we can write $Z = h(Y)$ for some measurable function h. The tower property of conditional expectation gives $Z = \mathbb{E}[\mathbb{E}[Y|f(Y)]|Z] = \mathbb{E}[Y|Z]$. Therefore, $h(Y) = \mathbb{E}[Y|h(Y)]$, showing that h is the Monge map as required in the definition of $\mu \leqslant_{\text{MM}} \nu$. □

3.2 Backward deterministic martingales

Theorem 2.1 gives rise to the remarkable class of backward deterministic martingales.

Definition 3.5. A stochastic process $(X_n)_{n\in\mathbb{N}}$ is *backward deterministic* if $(X_j)_{j=1}^n$ is $\sigma(X_n)$-measurable for all $n \in \mathbb{N}$.

In that case, $(X_n)_{n\in\mathbb{N}}$ is indeed a "deterministic" process if we go backward in time: the path $\{X_j,\, 1 \leqslant j \leqslant n\}$ is deterministic given X_n. Equivalently, $\sigma(X_j)$ is non-decreasing in n. As a direct consequence, a backward deterministic process $(X_n)_{n\in\mathbb{N}}$ is Markovian; in fact, it has perfect memory in the sense that its time-n value records all its history up to time n. While this may seem to be a fairly rare property, the following consequence of Theorem 2.1 shows that the class of backward deterministic martingales is rich enough to mimic (in the sense of Gyöngy (1986)) any given martingale with continuous marginals.

Corollary 3.6. *Given any martingale $(Y_n)_{n\in\mathbb{N}}$ with atomless marginals, there exists a backward deterministic martingale $(X_n)_{n\in\mathbb{N}}$ such that $X_n \overset{\text{law}}{=} Y_n$ for all $n \in \mathbb{N}$.*

Proof. Let μ_n be the distribution of Y_n for $n \in \mathbb{N}$. Then $\mu_n \leqslant_{\text{cx}} \mu_{n+1}$, so that Theorem 2.1 provides a sequence $\pi_n \in \mathcal{M}_M(\mu_n, \mu_{n+1})$, $n \in \mathbb{N}$. Let U_n, $n \in \mathbb{N}$ be a sequence of iid random variables uniformly distributed on $[0, 1]$. We construct the sequence $(X_n)_{n\in\mathbb{N}}$ inductively as follows. First, let $X_1 = g(U_1)$ where g is the left quantile function of μ_1; then $X_1 \overset{\text{law}}{\sim} \mu_1$. For $n = 2, 3, \dots$, let X_n be such that $(X_{n-1}, X_n) \overset{\text{law}}{\sim} \pi_{n-1}$ and X_n is measurable with respect to $(U_1, \dots, U_n)$. Such a sequence can be constructed by the inverse Rosenblatt transform; see, e.g., Rüschendorf (2013, Theorem 1.10). Then $(X_n)_{n\in\mathbb{N}}$ is a martingale with the marginal distributions μ_n, $n \in \mathbb{N}$. Moreover, since $(X_{n-1}, X_n) \overset{\text{law}}{\sim} \pi_{n-1}$ and π_{n-1} is an MMT, X_{n-1} is a function of X_n for each $n \geqslant 2$. Applying this repeatedly, we see that X_j is a function of X_n for all $j = 1, \dots, n$. □

The celebrated mimicking theorem of Gyöngy (1986) shows that the marginals of a (possibly non-Markovian) Itô process can also be generated with a Markovian Itô process. Here, in discrete time, we provide a mimicking martingale that is even backward deterministic. Of course, the relevant input of Corollary 3.6 is a family of distributions increasing in convex order rather than the process (Y_n). In that sense, it is a result about "peacocks" in the sense of Hirsch et al. (2011). To the best of our knowledge, the class of backward deterministic martingales has not been discussed in the previous literature. A deeper investigation remains for future work; we limit ourselves to the following observation.

Remark 3.7. A backward deterministic martingale $(X_n)_{n\in\mathbb{N}}$ cannot be a Gaussian process, except for the trivial form $(c, \dots, c, Z, Z, \dots)$ for some $c \in \mathbb{R}$ and Gaussian random variable Z. Indeed, suppose that $(X_n)_{n\in\mathbb{N}}$ is a backward deterministic martingale and a centered Gaussian process. It is clear that the variance σ_n^2 of X_n is increasing in n. Moreover, for $k < n$, $\mathbb{E}[X_n X_k] = \mathbb{E}[X_k^2] = \sigma_k^2$ since $(X_n)_{n\in\mathbb{N}}$ is a martingale. As the centered Gaussian distribution with a given covariance is unique, we conclude that X_k cannot be a function of X_n unless $\sigma_k = \sigma_n$ or $\sigma_k = 0$. Hence, for some $k_0 \in \mathbb{N}$, it holds that $X_k = 0$ for $k < k_0$ and $X_k = X_{k_0}$ for $k \geqslant k_0$. At a higher level, the

joint distribution of a backward martingale is concentrated on a set of Hausdorff dimension one (contrasting that a positive definite Gaussian vector is supported on the entire space).

3.3 MMTs as minimizers of generalized optimal transport

In this section we characterize $\mathcal{M}_M(\mu,\nu)$ through a generalized optimal transport problem. Starting with Gozlan et al. (2017), transport costs involving conditional distributions have been studied under the name of *generalized* or *weak* optimal transport. Such problems have found manifold applications such as the geometric inequalities of Gozlan et al. (2017) and the Brenier–Strassen theorem of Gozlan and Juillet (2020), and have counterparts to classic concepts such as the Kantorovich duality and cyclical monotonicity established by Gozlan et al. (2017) and Backhoff-Veraguas et al. (2019). We refer to Backhoff-Veraguas and Pammer (2022) for a recent survey.

Fix $\mu \leqslant_{\mathrm{cx}} \nu$ with ν atomless. It will be convenient to use random vectors (X,Y) instead of joint distributions; e.g., we abuse notation and write $(X,Y) \in \Pi(\mu,\nu)$. We first note that $\mathcal{M}_M(\mu,\nu)$ naturally arises through a two-stage optimization problem. The primary optimization is to minimize $\mathbb{E}[\mathbb{E}[Y-X|X]^2]$ over $\Pi(\mu,\nu)$, and its arg min is given by $\mathcal{M}(\mu,\nu)$. The secondary optimization is to minimize $\mathbb{E}[\mathbb{E}[Y-X|Y]^2]$, or equivalently $\mathbb{E}[\mathrm{Var}[X|Y]]$, over $\mathcal{M}(\mu,\nu)$; here the arg min is $\mathcal{M}_M(\mu,\nu)$. This is a symmetric variant of the *barycentric optimal transport cost* introduced by Gozlan et al. (2017). Extending this idea, the following result represents $\mathcal{M}_M(\mu,\nu)$ as the arg min of a class of generalized optimal transport problems over $\Pi(\mu,\nu)$.

Proposition 3.8. *Consider $\mu \leqslant_{\mathrm{cx}} \nu$ with ν atomless. For any strictly convex $f,g:\mathbb{R}\to\mathbb{R}$,*

$$\mathcal{M}_M(\mu,\nu) = \operatorname*{arg\,min}_{(X,Y)\in\Pi(\mu,\nu)} \mathbb{E}\left[f(\mathbb{E}[Y|X]-X) - g(\mathbb{E}[X|Y])\right]. \tag{3.1}$$

Proof. Recall from Theorem 2.1 that $\mathcal{M}_M(\mu,\nu)\neq\emptyset$, let $f,g:\mathbb{R}\to\mathbb{R}$ be strictly convex and $(X,Y)\in\Pi(\mu,\nu)$. Using the conditional Jensen's inequality and recalling that $\mu\leqslant_{\mathrm{cx}}\nu$ implies $\mathbb{E}[X]=\mathbb{E}[Y]$,

$$\begin{aligned}\mathbb{E}\left[f(\mathbb{E}[Y|X]-X) - g(\mathbb{E}[X|Y])\right] &\geqslant f\left(\mathbb{E}\left[\mathbb{E}[Y|X]-X\right]\right) - \mathbb{E}\left[\mathbb{E}[g(X)|Y]\right]\\ &= f(\mathbb{E}[Y]-\mathbb{E}[X]) - \mathbb{E}[g(X)] = f(0) - \mathbb{E}[g(X)].\end{aligned}$$

Clearly, the right-hand side is independent of the coupling $(X,Y)\in\Pi(\mu,\nu)$. The above inequality is an equality if and only if $\mathbb{E}[Y-X|X]=0$ and X is $\sigma(Y)$-measurable, or equivalently $(X,Y)\in\mathcal{M}_M(\mu,\nu)$. □

Remark 3.9. For $f(x)=g(x)=x^2$, the generalized transport cost in (3.1) is equivalent to

$$\mathbb{E}\big[\mathbb{E}[Y|X]^2 - \mathbb{E}[X|Y]^2 - 2\mathbb{E}[XY]\big].$$

We note that this cost is not symmetric in X and Y, and moreover, the term $-2\mathbb{E}[XY]$ is essential: one can check that $\mathcal{M}_M(\mu,\nu)$ does *not* solve the problem of minimizing $\mathbb{E}[\mathbb{E}[Y|X]^2 - \mathbb{E}[X|Y]^2]$

unless X is a constant.

4 Proofs of the main results

4.1 Structure of the left-curtain transport π_{lc}

In this subsection, we prove Proposition 2.2. Fix $\mu, \nu \in \mathcal{P}(\mathbb{R})$ with $\mu \leqslant_{\mathrm{cx}} \nu$. We first recall two properties of the left-curtain transport π_{lc}. The first one is Theorem 1.5 of Beiglböck and Juillet (2016).

Lemma 4.1 (π_{lc} is left-monotone)**.** *The left-curtain transport $\pi_{\mathrm{lc}} \in \mathcal{M}(\mu, \nu)$ satisfies $\pi_{\mathrm{lc}}(\Gamma) = 1$, where $\Gamma \subseteq \mathbb{R} \times \mathbb{R}$ is a left-monotone set; that is, whenever $(x, y^-), (x, y^+), (x', y') \in \Gamma$, it* cannot *hold that*

$$x < x' \quad \text{and} \quad y^- < y' < y^+.$$

Moreover, $\pi_{\mathrm{lc}} \in \mathcal{M}(\mu, \nu)$ is uniquely characterized by that property.

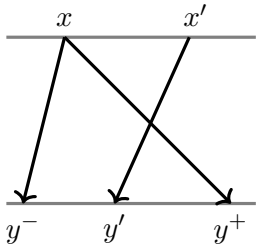


Figure 3: Forbidden configuration for left-monotonicity: the legs of a point x' cannot step into the legs of another point x to the left of x'.

The second property is that, outside of μ-atoms, π_{lc} is supported on the graphs of two functions ("legs") over the first marginal (i.e., forward in time); cf. Corollary 1.6 of Beiglböck and Juillet (2016) and Theorem 1 of Hobson and Norgilas (2019).

Lemma 4.2 (Support of π_{lc})**.** *There exist two functions $T_{\mathrm{d}},\ T_{\mathrm{u}} : \mathbb{R} \to \mathbb{R}$ such that $\pi_{\mathrm{lc}}(R_{\mathrm{legs}} \cup R_{\mathrm{atom}}) = 1$, where*

(a) R_{legs} is the union of the graphs of $T_{\mathrm{d}},\ T_{\mathrm{u}}$ over the first marginal;

(b) $R_{\mathrm{atom}} = \{(x, y) : \mu(\{x\}) > 0\}$.

Define the densities

$$d_\mu := \frac{\mathrm{d}\mu}{\mathrm{d}(\mu + \nu)} \quad \text{and} \quad d_\nu := \frac{\mathrm{d}\nu}{\mathrm{d}(\mu + \nu)}, \tag{4.1}$$

and denote by $\kappa_x(\mathrm{d}y)$ the disintegration of π_{lc} by μ, or conditional distribution given the first marginal: $\pi_{\mathrm{lc}}(\mathrm{d}x, \mathrm{d}y) = \mu(\mathrm{d}x) \otimes \kappa_x(\mathrm{d}y)$.

Lemma 4.3. *We have $d_\mu \leqslant d_\nu$ μ-a.e. on $\{x \in \mathbb{R} : \kappa_x = \delta_x\}$.*

Proof. Define $A = \{x \in \mathbb{R} : \kappa_x = \delta_x\} \cap \{x \in \mathbb{R} : d_\mu(x) > d_\nu(x)\}$. Assuming $\mu(A) > 0$, we find

$$\mu(A) = \int_A d_\mu \,\mathrm{d}(\mu+\nu) > \int_A d_\nu \,\mathrm{d}(\mu+\nu) = \nu(A) = \int \kappa_x(A)\mu(\mathrm{d}x) = \int \delta_x(A)\mu(\mathrm{d}x) = \mu(A),$$

a contradiction. □

Proof of Proposition 2.2. We first detail the proof for the second assertion, namely that

$$\pi_{\mathrm{lc}} \text{ is supported on } S_{\mathrm{rg}} \cup S_{\mathrm{atom}} \text{ if } d_\mu \geqslant d_\nu \ \mu\text{-a.e.}$$

and $\pi_{\mathrm{lc}} \in \mathcal{M}_M(\mu,\nu)$ if in addition ν is atomless.

Step 1. We have the martingale property $\int_{\mathbb{R}} y\,\kappa_x(\mathrm{d}y) = x$ for μ-a.e. x. Then by Lemma 4.2, for μ-a.e. x with $\mu(\{x\}) = 0$, either $\kappa_x = \delta_x$ or κ_x is supported on two points $T_{\mathrm{d}}(x) < x < T_{\mathrm{u}}(x)$. Moreover, if $\mu(\{x\}) > 0$, then either $\kappa_x(\{x\}) = 0$ or x belongs to the set $A_\nu = \{y \in \mathbb{R} : \nu(\{y\}) > 0\}$ of atoms of ν. In view of Lemma 4.3 and $d_\mu \geqslant d_\nu$ μ-a.e., we conclude that

$$\{x \in \mathbb{R} : \kappa_x(\{x\}) > 0\} = \{x \in \mathbb{R} : \kappa_x = \delta_x\} \subseteq \{x \in \mathbb{R} : d_\mu(x) = d_\nu(x)\} \quad \mu\text{-a.e. outside } A_\nu.$$

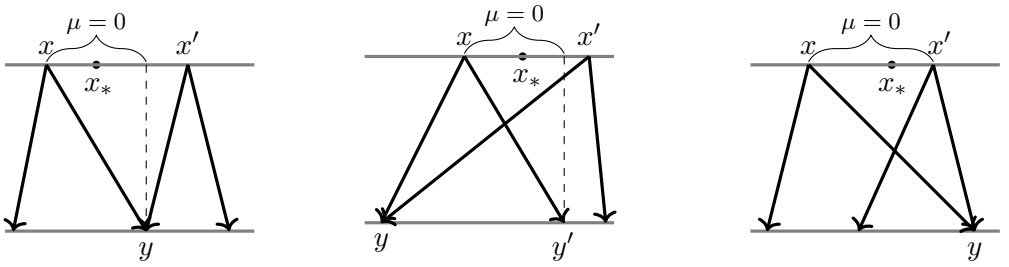
In summary, π_{lc} is the identity transport on $S := \{x \in \mathbb{R} : \kappa_x(\{x\}) > 0\} \setminus A_\nu$ and has the backward Monge property on S. Thus, we may without loss of generality "remove" $\mu|_S$ from the two marginals and assume that $\kappa_x(\{x\}) = 0$ μ-a.e. outside A_ν for the remainder of the proof.

Step 2. Let Γ be the left-monotone set provided by Lemma 4.1. By taking intersection, we may assume that $\Gamma \subseteq \mathrm{supp}(\mu) \times \mathrm{supp}(\nu)$ and $\Gamma \subseteq R_{\mathrm{legs}} \cup R_{\mathrm{atom}}$, where $\mathrm{supp}(\cdot)$ denotes topological support. By Step 1, we may further assume $(\Gamma \setminus S_{\mathrm{atom}}) \cap \{(x,x) : x \in \mathbb{R}\} = \emptyset$. Suppose that $x < x'$ are two points being transported to the same point $y \notin A_\nu$, or more precisely, that the pairs $(x,y), (x',y) \in \Gamma \setminus S_{\mathrm{atom}}$, and in particular $y \notin \{x, x'\}$. Then there are three possible cases (see Figure 4):

(a) If $x < y < x'$, then $\mu((x,y)) = 0$ (here, (x,y) refers to an interval instead of a pair). Indeed, if $x_* \in (x,y)$, then by Lemma 4.1, its right "leg" must lie on y because otherwise the left leg of x' "steps into" the legs of x_*. Since $\nu(\{y\}) = 0$, $\mu((x,y)) = 0$.

(b) If $y < x < x'$, denote by y' the right leg of x. Then by Lemma 4.1, the left leg of any $x_* \in (x, \min\{y', x'\}]$ cannot lie to the right of y, to avoid stepping into the legs of x, and not to the left of y because otherwise the left leg of x' steps into the legs of x_*. Thus the left leg of x_* must lie on y, implying that $\mu((x, \min\{y', x'\})) = 0$.

(c) If $x < x' < y$, consider $x_* \in (x, x')$. Then by Lemma 4.1, the right leg of x_* cannot lie to the left of y, to avoid stepping into the legs of x, and not to the right of y, because otherwise the right leg of x' steps into the legs of x_*. This shows that the right leg of x_* must lie on y, and thus $\mu((x,x')) = 0$.



(a) The case $x < y < x'$ (b) The case $y < x < x'$ (c) The case $x < x' < y$

Figure 4: Illustration of the three cases

As $\mathrm{supp}(\mu)$ is closed, its complement can be written as a countable disjoint union of open intervals. We have shown that each non-injective pair of $(x, y), (x', y) \in \Gamma \setminus S_{\mathrm{atom}}$ with $x \neq x'$ corresponds to an endpoint of one of the open intervals, and the map from the collection of all intervals to the collection of y values is at most one-to-two (since there are at most two legs). Thus, there are at most countably many such points y, and as ν is atomless outside A_ν, it follows that these points are ν-negligible. In summary, we have shown that π_{lc} is supported on the union of the (reverse) graph S_{rg} of a function $h : \mathbb{R} \to \mathbb{R}$ and S_{atom}.

It remains to see that h can be chosen to be measurable, and that $\pi_{\mathrm{lc}} = (h, \mathrm{id}_{\mathbb{R}})_{\#}\nu$ when ν is atomless. In the latter case, the mere fact that π_{lc} is concentrated on the graph of h already implies that h is ν-measurable and $\pi_{\mathrm{lc}} = (h, \mathrm{id}_{\mathbb{R}})_{\#}\nu$; see Ahmad et al. (2011, Lemma 3.1) for a detailed argument exploiting the inner regularity of Borel measures. Redefining h on a ν-nullset then gives the desired Borel measurable function. In the case with atoms, we can apply the same lemma to the restriction π' of π_{lc} to the Borel set $\mathbb{R}^2 \setminus S_{\mathrm{atom}}$. The lemma then yields that h is ν'-measurable where ν' is the second marginal of π', and we can again extract a Borel version. This completes the proof of the second assertion in Proposition 2.2.

The proof of the first assertion, namely that π_{lc} is supported on $S_{\mathrm{rg}} \cup S_{\mathrm{diag}} \cup S_{\mathrm{atom}}$, is similar to Step 2 above (but simpler): we now argue on the left-monotone set $\Gamma \setminus (S_{\mathrm{diag}} \cup S_{\mathrm{atom}})$. □

Remark 4.4 (When is π_{lc} Monge?)**.** While not directly required for our main results, it seems natural to ask when π_{lc} has the (reverse) Monge property. In the following discussion, we assume that ν is atomless. First of all, we note that the converse of Proposition 2.2 is false: $\pi_{\mathrm{lc}} \in \mathcal{M}_M(\mu, \nu)$ does not imply that $d_\mu \geqslant d_\nu$ μ-a.e. This can be seen by choosing the black density in Figure 5 small enough.

Recall that π_{lc} is supported on the union of the (forward) graphs of T_{d} and T_{u}. It follows from Proposition 2.2 that π_{lc} is Monge if and only if $d_\mu = d_\nu$ μ-a.e. on the set $\{x \in \mathbb{R} : T_{\mathrm{d}}(x) = T_{\mathrm{u}}(x)\}$ where the two legs of π_{lc} coincide. Under additional regularity assumptions, the main results of Henry-Labordère and Touzi (2016) imply (somewhat convoluted) equivalent conditions for this that can be stated in terms of the primitives μ and ν. To see the basic complication, consider $x \in \mathbb{R}$ with $d_\mu(x) \leqslant d_\nu(x)$. It is possible that $T_{\mathrm{d}}(x) = T_{\mathrm{u}}(x)$, i.e., the two legs coincide, while it is also possible that the ν-mass at x already lies in the shadow of $\mu|_{(-\infty, y]}$ for some $y < x$, making the legs separate

instead, as shown in Figure 5. In the proof of Theorem 2.1 below, we circumvent these issues by using the tractable sufficient condition $d_\mu \geqslant d_\nu$ and guaranteeing it through the decomposition into bars.

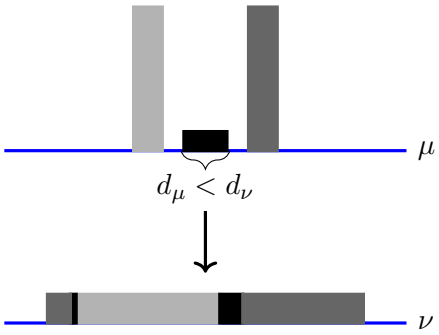


Figure 5: The left-curtain transport π_{lc} is not the identity on $\{x \in \mathbb{R} : d_\mu(x) < d_\nu(x)\}$

4.2 Proof of Theorem 2.1

We follow the notation of Section 4.1 but consider possibly unnormalized finite measures μ, ν on $\mathbb{R}$ as the following auxiliary results will be applied to sub-measures of the given marginals. We denote the barycenter by $\mathrm{bary}(\mu) := \int_{\mathbb{R}} x\, \mu(\mathrm{d}x)/\mu(\mathbb{R})$.

Lemma 4.5. *If $\mu(\mathbb{R}) = \nu(\mathbb{R}) > 0$, then $\mu(\{d_\mu \geqslant d_\nu\}) > 0$.*

Proof. Suppose $\mu(\{d_\mu \geqslant d_\nu\}) = 0$, then also $\nu(\{d_\mu \geqslant d_\nu\}) = 0$. Thus $\mu(\mathbb{R}) = \mu(\{d_\mu < d_\nu\}) < \nu(\{d_\mu < d_\nu\}) = \nu(\mathbb{R})$, contradicting our assumption. □

Two properties of shadows will be used repeatedly. The first is due to Beiglböck and Juillet (2016, Theorem 4.8).

Lemma 4.6 (Associativity of shadows)**.** *Suppose that $\mu = \mu_1 + \mu_2 \leqslant_{\mathrm{E}} \nu$. Then $\mu_2 \leqslant_{\mathrm{E}} \nu - S^\nu(\mu_1)$ and $S^\nu(\mu) = S^\nu(\mu_1) + S^{\nu - S^\nu(\mu_1)}(\mu_2)$.*

The second can be found in Beiglböck and Juillet (2016, Example 4.7).

Lemma 4.7. *When ν is atomless, the shadow of an atom of μ is ν restricted to an interval.*

The following significantly generalizes Lemma 4.7 by using Proposition 2.2.

Lemma 4.8. *Consider $\mu \leqslant_{\mathrm{E}} \nu$ with $d_\mu \geqslant d_\nu$ μ-a.e. Then $S^\nu(\mu)$ and $\nu - S^\nu(\mu)$ are mutually singular outside of $\{y \in \mathbb{R} : \nu(\{y\}) > 0\}$.*

Proof. In case $\mu(\mathbb{R}) = \nu(\mathbb{R})$, it must hold that $S^\nu(\mu) = \nu$ and the conclusion is vacuously true. Thus we may assume $\mu(\mathbb{R}) < \nu(\mathbb{R})$. Since $\mu \leqslant_{\mathrm{E}} \nu$, we may add to μ a Dirac mass to get a measure dominated by ν in convex order: taking $\lambda = \nu(\mathbb{R}) - \mu(\mathbb{R})$ and $m = \lambda^{-1}(\nu(\mathbb{R})\,\mathrm{bary}(\nu) - \mu(\mathbb{R})\,\mathrm{bary}(\mu))$ yields that $\mu + \lambda\delta_m \leqslant_{\mathrm{cx}} \nu$. Applying Proposition 2.2 to the measures $\mu + \lambda\delta_m$ and ν yields that

the left-curtain transport from $\mu + \lambda\delta_m$ to ν is Monge outside the set $A := \{y \in \mathbb{R} : \nu(\{y\}) > 0\}$ of atoms of ν. Since the left-curtain transport sends $\mu|_{(-\infty,m]}$ to its shadow $S^\nu(\mu|_{(-\infty,m]})$, we deduce that $S^\nu(\mu|_{(-\infty,m]})$ and $\nu' := \nu - S^\nu(\mu|_{(-\infty,m]})$ are mutually singular outside of A. Note that

$$\frac{\mathrm{d}(\mu|_{(m,\infty)} + \lambda\delta_m)}{\mathrm{d}(\mu|_{(m,\infty)} + \lambda\delta_m + \nu')} \geqslant \frac{1}{2}, \quad (\mu|_{(m,\infty)} + \lambda\delta_m)\text{-a.e.}$$

By a symmetrical argument using Proposition 2.2, the right-curtain transport from $\mu|_{(m,\infty)} + \lambda\delta_m$ to ν' is backward Monge outside of A and sends $\mu|_{(m,\infty)}$ to $S^{\nu'}(\mu|_{(m,\infty)})$, and thus $S^{\nu'}(\mu|_{(m,\infty)})$ and $\nu' - S^{\nu'}(\mu|_{(m,\infty)})$ are mutually singular. By Lemma 4.6, it holds that $S^\nu(\mu) = S^\nu(\mu|_{(-\infty,m]}) + S^{\nu'}(\mu|_{(m,\infty)})$. Therefore, $S^\nu(\mu)$ and $\nu - S^\nu(\mu)$ are mutually singular outside of A. □

We can now construct the barcode transport.

Proof of Theorem 2.1. Given $\mu, \nu \in \mathcal{P}(\mathbb{R})$ with $\mu \leqslant_{\mathrm{cx}} \nu$, we let $(d_\mu^{(0)}, d_\nu^{(0)}) := (d_\mu, d_\nu)$ be defined as in (4.1). Consider $A_0 := \{d_\mu^{(0)} \geqslant d_\nu^{(0)}\}$. We transport $\mu|_{A_0}$ to $S^\nu(\mu|_{A_0})$ using the left-curtain coupling, which is Monge outside the set $\{y \in \mathbb{R} : \nu(\{y\}) > 0\}$ of atoms of ν by Proposition 2.2. (In Figure 1 (a), this corresponds to the light-gray area in the center.) Define the remaining measures

$$\mu_1 := \mu - \mu|_{A_0}, \qquad \nu_1 := \nu - S^\nu(\mu|_{A_0}),$$

so that $\mu_1 \leqslant_{\mathrm{cx}} \nu_1$. We continue recursively: given $n \in \mathbb{N}$ and measures $\mu_n \leqslant_{\mathrm{cx}} \nu_n$, we define the densities $d_\mu^{(n)}, d_\nu^{(n)}$ of μ_n, ν_n with respect to $(\mu + \nu)$ and $A_n := \{d_\mu^{(n)} \geqslant d_\nu^{(n)}\}$, as well as

$$\mu_{n+1} := \mu_n - \mu_n|_{A_n}, \qquad \nu_{n+1} := \nu_n - S^{\nu_n}(\mu_n|_{A_n}).$$

Let also $\pi_n \in \mathcal{M}(\mu_n|_{A_n}, S^{\nu_n}(\mu_n|_{A_n}))$ be the left-curtain transport, which is again Monge outside the atoms of ν by Proposition 2.2. By construction, the measures $\{\mu_n - \mu_{n+1}\}$ are mutually singular, and by Lemma 4.8, the measures $\{\nu_n - \nu_{n+1}\}$ are mutually singular outside of $\{y \in \mathbb{R} : \nu(\{y\}) > 0\}$.

Again by construction, we have that $d_\mu^{(n)}, d_\nu^{(n)}$ are decreasing sequences of functions $(\mu+\nu)$-a.e. Denote their limits $d_\mu^{(\infty)}, d_\nu^{(\infty)}$ respectively. Let $x \in \mathbb{R}$ belong to the $(\mu + \nu)$-a.e. set where $d_\mu^{(n)}, d_\nu^{(n)}$ are decreasing and such that $d_\mu^{(\infty)}(x) \geqslant d_\nu^{(\infty)}(x)$. Then by mutual singularity of $\{\nu_n - \nu_{n+1}\}$, we have $d_\nu^{(n)}(x) \in \{d_\nu^{(0)}(x), 0\}$ for all n. There are two possible cases:

(a) Suppose that there is a finite n such that $d_\nu^{(n)}(x) = 0$. Then $d_\nu^{(\infty)}(x) = 0$ and $d_\mu^{(n)}(x) \geqslant d_\nu^{(n)}(x)$. This means that the μ-mass at x must be transported at step $n + 1$ or earlier, giving that $d_\mu^{(n+1)}(x) = 0$.

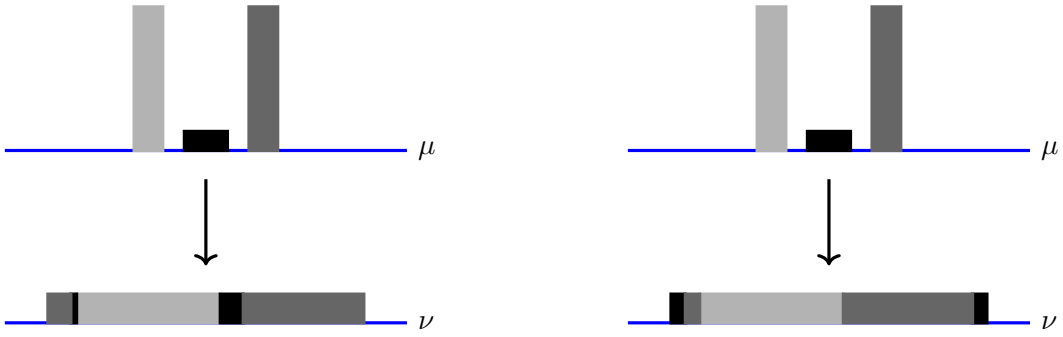
(b) Suppose that $d_\nu^{(n)}(x) = d_\nu^{(0)}(x)$ for all n. Then $d_\mu^{(0)}(x) \geqslant d_\mu^{(\infty)}(x) \geqslant d_\nu^{(\infty)}(x) = d_\nu^{(0)}(x)$. By construction, the μ-mass at x must be transported in the first step, so that $d_\mu^{(1)}(x) = 0$.

It follows that $d_\mu^{(\infty)}(x) = 0$. Therefore, μ_∞ is the zero measure by Lemma 4.5, and so is ν_∞ since $\mu_n(\mathbb{R}) = \nu_n(\mathbb{R})$ by construction. Since outside of $\{y \in \mathbb{R} : \nu(\{y\}) > 0\}$, each transport

$\pi_n \in \mathcal{M}(\mu_n - \mu_{n+1}, \nu_n - \nu_{n+1})$ is Monge and the measures $\{\nu_n - \nu_{n+1}\}$ are mutually singular, aggregating these transports yields a transport from μ to ν that is Monge outside that set. □

We remark that, by construction, the barcode transport belongs to the broad class of shadow couplings introduced by Beiglböck and Juillet (2021). While our construction uses the left-curtain transport for its relatively simple behavior, this is certainly not the only possible choice.

Remark 4.9. Even if the left-curtain transport is an MMT for two given marginals, our construction may result in a different transport; see Figure 6 for an example.



(a) Left-curtain transport (b) Barcode transport

Figure 6: Left-curtain and barcode transport are MMTs, yet do not coincide

4.3 Proof of Theorem 2.3

Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ where ν is atomless. For $p \in [1, \infty]$, we denote by W_p the p-Wasserstein distance of measures on either $\mathbb{R}$ or $\mathbb{R}^2$ equipped with the Euclidean metric. While the two assertions of Theorem 2.3 will be proved independently, the proof for discrete μ is presented first as it is much simpler yet contains some of the basic ideas for both cases.

Lemma 4.10. *Let $\nu \in \mathcal{P}(\mathbb{R})$ be atomless. Given any decomposition $\nu = \sum_{i=1}^{\infty} \nu_i$ of ν, there exist mutually singular $\tilde{\nu}_i$, $i \in \mathbb{N}$ such that $\nu = \sum_{i=1}^{\infty} \tilde{\nu}_i$ and $\nu_1 \leqslant_{\mathrm{cx}} \tilde{\nu}_1$ and $\mathrm{bary}(\nu_i) = \mathrm{bary}(\tilde{\nu}_i)$ for $i \geqslant 2$.*

Proof. Define $\tilde{\mu}_i = \nu_i(\mathbb{R})\delta_{\mathrm{bary}(\nu_i)}$, $i \in \mathbb{N}$. Note that $\sum_{i=1}^{\infty} \tilde{\mu}_i \leqslant_{\mathrm{cx}} \nu_1 + \sum_{i=2}^{\infty} \tilde{\mu}_i \leqslant_{\mathrm{cx}} \sum_{i=1}^{\infty} \nu_i = \nu$. We consider the shadow $\nu_0 := S^{\nu}(\sum_{i=2}^{\infty} \tilde{\mu}_i)$ and set $\tilde{\nu}_1 = \nu - \nu_0$. By Lemma 4.6 and Lemma 4.7, $\nu_1 \leqslant_{\mathrm{cx}} \tilde{\nu}_1$ and $\tilde{\nu}_1$ is mutually singular with ν_0. Roughly speaking, $\tilde{\nu}_1$ is the *largest* possible image of ν_1 under a martingale transport, in the sense of the convex order.

Next, we apply a shadow coupling from $\sum_{i=2}^{\infty} \tilde{\mu}_i$ to ν_0, processing these atoms in the order $i = 2, 3, \ldots$. More precisely, we let $\tilde{\nu}_2 := S^{\nu_0}(\tilde{\mu}_2)$ and $\tilde{\nu}_i := S^{\nu_0 - \sum_{j=2}^{i} \tilde{\mu}_j}(\tilde{\mu}_i)$ for $i \geqslant 3$. By construction and Lemma 4.7, these shadows $\tilde{\nu}_i$, $i \geqslant 2$, are mutually singular. As sub-measures of ν_0, they are also mutually singular with $\tilde{\nu}_1$. The other assertions are clear. □

Proof of Theorem 2.3 for discrete μ. Fix $\pi \in \mathcal{M}(\mu, \nu)$ and $\varepsilon > 0$; we construct $\pi_\varepsilon \in \mathcal{M}_M(\mu, \nu)$ with $W_\infty(\pi, \pi_\varepsilon) \leqslant \varepsilon$. Partition $\mathbb{R}$ into intervals $\{I_\ell\}_{\ell \in \mathbb{N}}$ of length ε and write $\nu = \sum_{\ell=1}^{\infty} \nu|_{I_\ell}$.

Decompose the discrete measure μ into its atoms, $\mu = \sum_{k=1}^{\infty} \mu_k$. Then, decompose $\nu|_{I_\ell} = \sum_{k=1}^{\infty} \nu_{k,\ell}$ where $\nu_{k,\ell}$ is the image of μ_k under π restricted to I_ℓ. For each ℓ, apply Lemma 4.10 to the decomposition $\nu|_{I_\ell} = \sum_{k=1}^{\infty} \nu_{k,\ell}$, yielding measures $\{\mu_{k,\ell}\}_{k,\ell\in\mathbb{N}}$ and $\{\tilde{\nu}_{k,\ell}\}_{k,\ell\in\mathbb{N}}$ such that $\mu_k = \sum_{\ell=1}^{\infty} \mu_{k,\ell}$ and $\nu|_{I_\ell} = \sum_{k=1}^{\infty} \tilde{\nu}_{k,\ell}$ and $\mu_{k,\ell} \leqslant_{\mathrm{cx}} \tilde{\nu}_{k,\ell}$ and $\{\tilde{\nu}_{k,\ell}\}_{k,\ell\in\mathbb{N}}$ are mutually singular. Moreover, $W_\infty(\tilde{\nu}_{k,\ell}, \nu_{k,\ell}) \leqslant \varepsilon$ for all k, ℓ by construction. Consider the transport $\pi_\varepsilon \in \mathcal{M}(\mu, \nu)$ that sends each atom $\mu_{k,\ell}$ to $\tilde{\nu}_{k,\ell}$. Then $\pi_\varepsilon \in \mathcal{M}_M(\mu, \nu)$ since $\{\tilde{\nu}_{k,\ell}\}_{k,\ell}$ are mutually singular, and $W_\infty(\pi, \pi_\varepsilon) \leqslant \varepsilon$ since $W_\infty(\tilde{\nu}_{k,\ell}, \nu_{k,\ell}) \leqslant \varepsilon$. □

Before entering the technical details of the proof of Theorem 2.3 for general μ, let us try to sketch the main ideas. Similarly as in the discrete case above, we want to partition the supports of μ and ν into small enough intervals $\{J_k\}, \{I_\ell\}$ and define ν_k to be the image of $\mu|_{J_k}$ under the given transport $\pi_0 \in \mathcal{M}(\mu, \nu)$ to be approximated. Using barcodes, we would then approximate the measures $\nu_k|_{I_\ell}$ within the set I_ℓ for each ℓ, meaning that we find mutually singular $\{\hat{\nu}_{k,\ell}\}$ such that $\sum_k \nu_k|_{I_\ell} = \sum_k \hat{\nu}_{k,\ell}$ for each ℓ. This idea does not carry through directly, because these rearrangements may destroy vital convex order properties. Instead, we perform yet another approximation to create some "wiggle room" in the convex order. Rather than directly approximating the given coupling π_0, we approximate $\tilde{\pi}_0 = (1-\lambda)\pi_0 + \lambda\pi_3$ for small λ and a particular martingale transport $\pi_3 \in \mathcal{M}(\mu, \nu)$ with a tailored transport kernel based on a carefully chosen Rademacher noise. Roughly speaking, adding the noise yields a locally uniform lower bound on the dispersion of the transport kernels.

It will be important to quantify how far two marginals are separated from one another in the convex order—specifically, how large a perturbation (in W_∞) can be applied without violating the order. To that end, the characterization of the convex order by potential functions is useful. The potential function $u_\mu : \mathbb{R} \to \mathbb{R}$ of μ is defined as $x \mapsto \int_{\mathbb{R}} |y - x| \mathrm{d}\mu(y)$. This function is convex and Lipschitz. If μ and ν have the same mass and barycenter, then $u_\mu \leqslant u_\nu$ if and only if $\mu \leqslant_{\mathrm{cx}} \nu$; see Shaked and Shanthikumar (2007, Theorem 3.A.2). The difference $u_\nu(x) - u_\mu(x)$ will be used as a local measure of separation between the marginals.

Lemma 4.11. *Without loss of generality, we may assume that $I := \{u_\mu < u_\nu\}$ is an (open) interval and that $\mu(I) = \nu(I) = 1$. In particular, $\mu(\{u_\mu = u_\nu\}) = 0$.*

Proof. Consider the decomposition $\mu = \sum_{i\geqslant 0} \mu_i$ and $\nu = \sum_{i\geqslant 0} \nu_i$ of (μ, ν) into the so-called irreducible components; cf. Beiglböck and Juillet (2016, Theorem A.4). Here (μ_i, ν_i) are in convex order and any $\pi \in \mathcal{M}(\mu, \nu)$ transports μ_i to ν_i. Moreover, $\mu_0 = \nu_0$ are such that any $\pi \in \mathcal{M}(\mu, \nu)$ transports μ_0 to ν_0 via the identity transport. Finally, $(\mu_i)_{i\geqslant 1}$ are supported on the disjoint intervals $\{u_{\mu_i} < u_{\nu_i}\}$ and μ_0 is supported on the complement of their union. The same holds for $(\nu_i)_{i\geqslant 0}$, as follows from Beiglböck and Juillet (2016, Lemma A.6): while in general ν_i may place mass at the endpoints of its interval, that is not the case here as ν is atomless. It follows that any $\pi \in \mathcal{M}(\mu, \nu)$ is Monge on the complement of the intervals (since the only transport there is the identity), and if the denseness result of Theorem 2.3 holds for each (μ_i, ν_i) with $i \geqslant 1$, then aggregating yields the desired theorem for (μ, ν). □

In the remainder of the proof, we assume that the condition of Lemma 4.11 holds.

Lemma 4.12. *We have* $\lim_{\delta\downarrow 0}\mu(A_\delta)=0$ *for* $A_\delta := [-\delta,\delta]+\{x\in\mathbb{R}: 0\leqslant u_\nu(x)-u_\mu(x)<\delta\}$.

Proof. The sets A_δ are decreasing and $\cap_{\delta>0}A_\delta=\{u_\mu=u_\nu\}$ which is μ-null by our assumption. □

The next lemma quantifies how much "wiggle room" of convex order the Rademacher noise introduces into a distribution. We denote by Rade the Rademacher distribution, or uniform on $\{-1,+1\}$.

Lemma 4.13. *Fix* $x_0\in\mathbb{R}$, $\lambda\in(0,1]$, *and* $\varepsilon>0$. *Let* μ_1 *be a probability measure with mean* x_0 *such that* $\mu_1([x_0-\lambda\varepsilon/6, x_0+\lambda\varepsilon/6])=1$, *and* μ_2 *be the distribution of* $X_1+\varepsilon B\xi$ *where* $X_1\overset{\text{law}}{\sim}\mu_1$, $B\overset{\text{law}}{\sim}\text{Bernoulli}(\lambda)$ *and* $\xi\overset{\text{law}}{\sim}\text{Rade}$ *are independent. Suppose that* μ_3 *and* μ_4 *are probability measures with the same mean* x_0 *such that* $\mu_2\leqslant_{\text{cx}}\mu_3$ *and* $W_\infty(\mu_3,\mu_4)\leqslant\lambda\varepsilon/6$. *Then* $\mu_1\leqslant_{\text{cx}}\mu_4$.

Proof. We first claim that there exists μ_3' with mean x_0 such that $W_\infty(\mu_2,\mu_3')<\lambda\varepsilon/6$ and $\mu_3'\leqslant_{\text{cx}}\mu_4$. Using the disintegration theorem, we may write kernels $\kappa_x^{(2)}$ and $\kappa_x^{(3)}$ that transport μ_2 to μ_3 and μ_3 to μ_4 respectively, such that the mean of $\kappa_x^{(2)}$ is x (i.e., $\kappa_x^{(2)}$ is an MT) and $\kappa_x^{(3)}$ is concentrated in $[x-\lambda\varepsilon/6, x+\lambda\varepsilon/6]$ for each $x\in\mathbb{R}$. Denote by x^* the mean of the measure $\kappa^{(3)}\circ\kappa_x^{(2)}$. Let $\mu_3'=(T_3)_{\#}\mu_2$ where $T_3: x\mapsto x^*$. Since by assumption the mean of $\kappa_x^{(3)}$ lies in $[x-\lambda\varepsilon/6, x+\lambda\varepsilon/6]$, we must have $|x-x^*|\leqslant\lambda\varepsilon/6$. Therefore, $W_\infty(\mu_2,\mu_3')<\lambda\varepsilon/6$. Consider the map $x^*\mapsto\mathbb{E}[\kappa^{(3)}\circ\kappa_{X_2}^{(2)}|T_3(X_2)=x^*]$ that aggregates $\kappa^{(3)}\circ\kappa_x^{(2)}$ among all sources x such that $T_3(x)=x^*$, where $X_2\overset{\text{law}}{\sim}\mu_2$. Since such a map forms a martingale transport from μ_3' to μ_4, it follows that $\mu_3'\leqslant_{\text{cx}}\mu_4$.

It now suffices to prove $\mu_1\leqslant_{\text{cx}}\mu_3'$. Consider a coupling (X_1,X_2,X_3) such that $X_i\overset{\text{law}}{\sim}\mu_i$ for $i=1,2$ and $X_3\overset{\text{law}}{\sim}\mu_3'$, $X_2=X_1+\varepsilon B\xi$, and $|X_2-X_3|<\lambda\varepsilon/6$. Let $a\in\mathbb{R}$; we will show that $\mathbb{E}[(X_1-a)_+]\leqslant\mathbb{E}[(X_3-a)_+]$. The case $a>x_0+\lambda\varepsilon/6$ is obvious. If $a\in[x_0,x_0+\lambda\varepsilon/6]$, we have using $|X_1-x_0|\leqslant\lambda\varepsilon/6$ that

$$\mathbb{E}[(X_1-a)_+]\leqslant\frac{\lambda\varepsilon}{6}\leqslant\frac{\lambda}{2}\left(\varepsilon-\frac{3\lambda\varepsilon}{6}\right)\leqslant\mathbb{E}[(X_3-a)_+].$$

The other cases are symmetric using our assumption $\mathbb{E}[X_1]=x_0=\mathbb{E}[X_3]$. □

Proof of Theorem 2.3 for general μ. Let $\mu\leqslant_{\text{cx}}\nu$ with ν atomless, $\pi_0\in\mathcal{M}(\mu,\nu)$ and $\varepsilon>0$. Consider quantities $\delta,\lambda\in(0,1)$ small enough (to be determined below) depending on ε. Define

$$A_\delta=[-\delta,\delta]+\{x\in\mathbb{R}: 0\leqslant u_\nu(x)-u_\mu(x)<\delta\}$$

which appears in Lemma 4.12, as well as $A_\delta'=\{x\in\mathbb{R}: 0\leqslant u_\nu(x)-u_\mu(x)<\delta\}$. We divide the rest of the proof into three steps.

Step I: inserting Rademacher noise. Let $X\overset{\text{law}}{\sim}\mu$ and $\xi\overset{\text{law}}{\sim}\text{Rade}$ be independent. Denote by $\tilde{\mu}$ the distribution of $X_\delta:=X+\delta\xi\mathbb{1}_{\{X\notin A_\delta\}}$. We have $\mu\leqslant_{\text{cx}}\tilde{\mu}$. Observe that for $x\notin A_\delta'$,

$u_\nu(x) \geqslant u_\mu(x) + \delta$, so that $u_\nu(x) \geqslant u_{\tilde{\mu}}(x)$ by the triangle inequality. For $x \in A'_\delta$, we have

$$\begin{aligned} u_{\tilde{\mu}}(x) &= \mathbb{E}[|X - x| \mathbb{1}_{\{X \in A_\delta\}}] + \mathbb{E}[|X + \delta\xi - x| \mathbb{1}_{\{X \notin A_\delta\}}] \\ &= \mathbb{E}[|X - x| \mathbb{1}_{\{X \in A_\delta\}}] + \mathbb{E}[|X - x| \mathbb{1}_{\{X \notin A_\delta\}}] = u_\mu(x) \leqslant u_\nu(x), \end{aligned}$$

where in the second equality we used that $|X - x| \geqslant \delta$ on the set $\{X \notin A_\delta\}$, by definition of A_δ, A'_δ. As a result, $\mu \leqslant_{\mathrm{cx}} \tilde{\mu} \leqslant_{\mathrm{cx}} \nu$.

Let π_1 be any martingale transport between $\tilde{\mu}$ and ν, and π_2 be the martingale transport given by (X, X_δ). Note that the kernel of π_2 has support $\{-\delta, \delta\}$ on $\mathbb{R} \setminus A_\delta$ and is the identity kernel on A_δ. Composing π_2 and π_1 we get a coupling from μ to ν, denoted π_3. Let $\tilde{\pi}_0 = (1 - \lambda)\pi_0 + \lambda\pi_3$. It then suffices to approximate $\tilde{\pi}_0$ instead of π_0, i.e., to show that $\tilde{\pi}_0$ belongs to the weak closure of $\mathcal{M}_M(\mu, \nu)$. Once that is shown, it will follow by taking $\lambda \to 0$ that π_0 is also in the closure.

Step II: decomposition of the measures. Partition $\mathbb{R}$ into intervals $\{I_\ell\}_{\ell \in \mathbb{N}}$ such that $|I_\ell| \leqslant \lambda\varepsilon/6$, where $|I|$ denotes the length of an interval I. Let us discard all I_ℓ with $\nu(I_\ell) = 0$. We also partition $\mathbb{R} \setminus A_\delta$ into intervals $\{J_k\}_{k \in \mathbb{N}}$ such that $|J_k| \leqslant \lambda\varepsilon/6$, and define $J_0 = A_\delta$. Note that this is possible since A_δ is the union of some intervals. Again, let us discard all J_k with $\mu(J_k) = 0$.

Next, focus on one interval I_ℓ. Let $\mathbb{N}_0$ denote the set of nonnegative integers. For $k \in \mathbb{N}_0$, consider the image of $\mu|_{J_k}$ under $\tilde{\pi}_0$ which we denote by $\tilde{\nu}_k$. Moreover, let $\tilde{\nu}_{k,\ell} = \tilde{\nu}_k|_{I_\ell}$ for $k \in \mathbb{N}_0$. Note that $\{\tilde{\nu}_{k,\ell}\}_{k \in \mathbb{N}_0}$ forms a decomposition of $\nu|_{I_\ell}$. Applying Lemma 4.10 to this decomposition, we obtain mutually singular $\{\hat{\nu}_{k,\ell}\}_{k \in \mathbb{N}_0}$ such that $\nu|_{I_\ell} = \sum_{k=1}^\infty \hat{\nu}_{k,\ell}$, $\mathrm{bary}(\tilde{\nu}_{k,\ell}) = \mathrm{bary}(\hat{\nu}_{k,\ell})$ for $k \in \mathbb{N}$, $W_\infty(\tilde{\nu}_{k,\ell}, \hat{\nu}_{k,\ell}) \leqslant |I_\ell|$, and $\tilde{\nu}_{0,\ell} \leqslant_{\mathrm{cx}} \hat{\nu}_{0,\ell}$; see Figure 7 below for an illustration. Recall the definitions of π_3 and $\tilde{\pi}_0$.

(a) Applying Lemma 4.13 with $\mu_1 = \mu|_{J_k}$, μ_2 the image of μ_1 under the transport $(1 - \lambda)\mathrm{id} + \lambda\pi_2$, $\mu_3 = \tilde{\nu}_k = \sum_{\ell \in \mathbb{N}} \tilde{\nu}_{k,\ell}$, and $\mu_4 = \sum_{\ell \in \mathbb{N}} \hat{\nu}_{k,\ell}$ while noting that

$$W_\infty(\mu_3, \mu_4) \leqslant \sup_{\ell \in \mathbb{N}} W_\infty(\tilde{\nu}_{k,\ell}, \hat{\nu}_{k,\ell}) \leqslant \sup_{\ell \in \mathbb{N}} \tilde{\nu}_{k,\ell}(\mathbb{R})|I_\ell| \leqslant \frac{\lambda\varepsilon}{6},$$

we conclude that $\mu|_{J_k} \leqslant_{\mathrm{cx}} \sum_{\ell \in \mathbb{N}} \hat{\nu}_{k,\ell}$ for $k \in \mathbb{N}$.

(b) Similarly, it follows that $\mu|_{J_0} = \mu|_{A_\delta} \leqslant_{\mathrm{cx}} \tilde{\nu}_0 = \sum_{\ell \in \mathbb{N}} \tilde{\nu}_{0,\ell} \leqslant_{\mathrm{cx}} \sum_{\ell \in \mathbb{N}} \hat{\nu}_{0,\ell}$.

Step III: approximating $\tilde{\pi}_0$ by MMT. We can now construct an approximation $\hat{\pi}$ of $\tilde{\pi}_0$ as follows. Note that since ν is atomless, so is $\hat{\nu}_{k,\ell}$ for all $k \in \mathbb{N}_0$, $\ell \in \mathbb{N}$.

(a) For each $k \in \mathbb{N}$, applying Theorem 2.1 to $\mu|_{J_k}$ and $\sum_{\ell \in \mathbb{N}} \hat{\nu}_{k,\ell}$ yields a coupling $\hat{\pi}^k$ which is an MMT between $\mu|_{J_k}$ and $\sum_{\ell \in \mathbb{N}} \hat{\nu}_{k,\ell}$. Denote by $\tilde{\pi}^k$ the original coupling between $\mu|_{J_k}$ and $\tilde{\nu}_k$

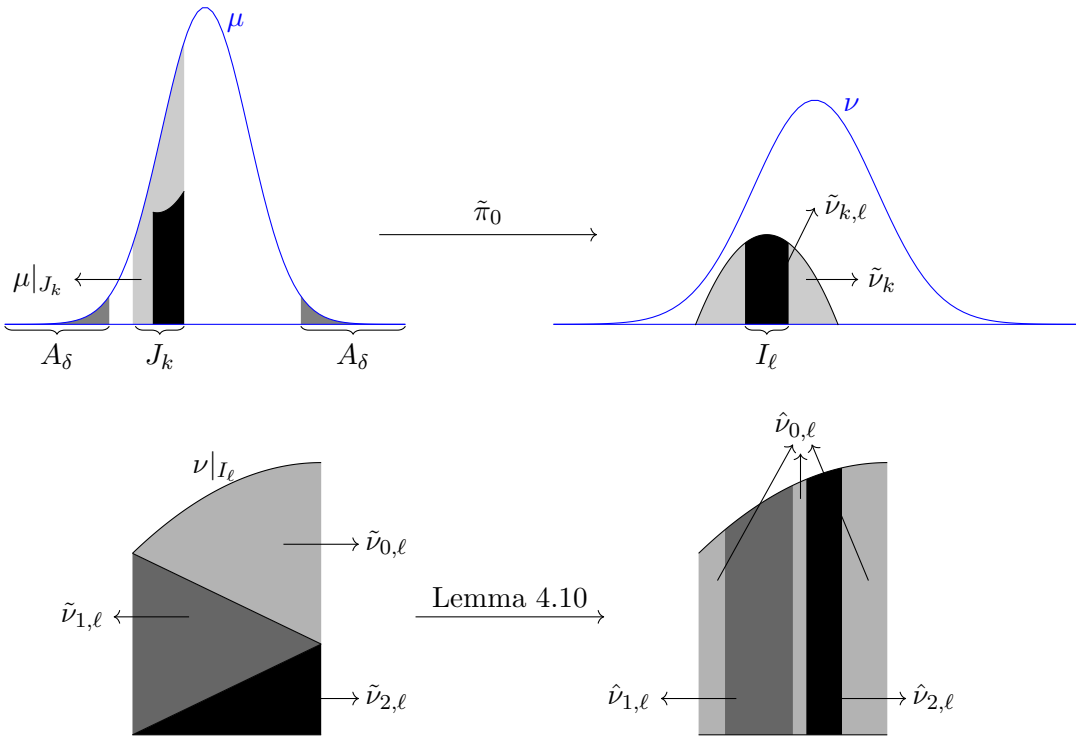


Figure 7: Illustrating the transport $\tilde{\pi}_0$ and Lemma 4.10

induced by $\tilde{\pi}_0$. It follows that

$$W_1\left(\sum_{k\in\mathbb{N}}\tilde{\pi}^k,\sum_{k\in\mathbb{N}}\hat{\pi}^k\right)\leqslant W_\infty\left(\sum_{k\in\mathbb{N}}\tilde{\pi}^k,\sum_{k\in\mathbb{N}}\hat{\pi}^k\right)$$
$$\leqslant\sup_{k\in\mathbb{N}}W_\infty(\tilde{\pi}^k,\hat{\pi}^k)\leqslant\sup_{k\in\mathbb{N}}\mu(J_k)\left(|J_k|+\max_{\ell\in\mathbb{N}}|I_\ell|\right)\leqslant\lambda\varepsilon.$$

(b) We apply Theorem 2.1 to $\mu|_{A_\delta}$ and $\sum_{\ell\in\mathbb{N}}\hat{\nu}_{0,\ell}$, and get another MMT, denoted $\hat{\pi}^0$. Denote by $\tilde{\pi}^0$ the original coupling between $\mu|_{A_\delta}$ and $\tilde{\nu}_0$ induced by $\tilde{\pi}_0$. By Lemma 4.12, $\mu(A_\delta)\to 0$ as $\delta\to 0$, so that $W_1(\tilde{\pi}^0,\hat{\pi}^0)\to 0$.

Since $\{\hat{\nu}_{k,\ell}\}_{k\in\mathbb{N}_0,\ell\in\mathbb{N}}$ are mutually singular as noted above, it follows that $\hat{\pi}:=\sum_{k=0}^\infty\hat{\pi}^k$ is an MMT. The first marginal of $\hat{\pi}$ is $\mu|_{A_\delta}+\sum_{k=1}^\infty\mu|_{J_k}=\mu$ and the second marginal of $\hat{\pi}$ is $\sum_{k=0}^\infty\sum_{\ell\in\mathbb{N}}\hat{\nu}_{k,\ell}=\sum_{\ell\in\mathbb{N}}\nu|_{I_\ell}=\nu$. Therefore, $\hat{\pi}\in\mathcal{M}_M(\mu,\nu)$. Note that

$$W_1(\tilde{\pi}_0,\hat{\pi})\leqslant W_1(\tilde{\pi}^0,\hat{\pi}^0)+W_1\left(\sum_{k\in\mathbb{N}}\tilde{\pi}^k,\sum_{k\in\mathbb{N}}\hat{\pi}^k\right).$$

As shown above, both terms tend to 0. Since W_1 convergence implies weak convergence, we conclude that $\mathcal{M}_M(\mu,\nu)$ is weakly dense in $\mathcal{M}(\mu,\nu)$. □

4.4 General results on the uniqueness of MT and MMT

In this subsection, we characterize the uniqueness of martingale transports and Monge martingale transports using shadow measures, for general marginals $\mu, \nu \in \mathcal{P}(\mathbb{R})$ with $\mu \leqslant_{\mathrm{cx}} \nu$ (possibly with atoms). To the best of our knowledge, the uniqueness of MT has not been completely characterized, except for a few simple examples mentioned in De March (2018) and Obłój and Siorpaes (2017). The first result states that $\mathcal{M}(\mu, \nu)$ is a singleton if and only if the shadows of any decomposition of μ do not affect each other.

Proposition 4.14. *The MT between μ and ν is unique if and only if $\nu = \sum_{j=1}^{n} S^{\nu}(\mu_j)$ for any $n \in \mathbb{N}$ and mutually singular $\mu_1, \dots, \mu_n \leqslant \mu$ satisfying $\sum_{j=1}^{n} \mu_j = \mu$.*

Proof. We first show the "if" statement. Suppose that $\nu = \sum_{j=1}^{n} S^{\nu}(\mu_j)$. We claim that the only possible MT is to transport μ_i to $S^{\nu}(\mu_i)$ for each i. Suppose otherwise, and let ν_i be the image of μ_i under a different MT. Then, by the minimality property of the shadow, there exist i and a convex function ϕ such that $\int \phi \, \mathrm{d}\nu_i > \int \phi \, \mathrm{d}S^{\nu}(\mu_i)$. As $\sum_{j=1}^{n} \int \phi \, \mathrm{d}\nu_j = \int \phi \, \mathrm{d}\nu = \sum_{j=1}^{n} \int \phi \, \mathrm{d}S^{\nu}(\mu_j)$, it follows that there exists j with $\int \phi \, \mathrm{d}\nu_j < \int \phi \, \mathrm{d}S^{\nu}(\mu_j)$, violating the definition of the shadow.

To show the "only if" statement, suppose that $\nu \neq \sum_{j=1}^{n} S^{\nu}(\mu_j)$ for some mutually singular $\mu_1, \dots, \mu_n$ adding up to μ. Note that necessarily $n \geqslant 2$ and fix $j \in \{1, \dots, n\}$. We define $\pi_j \in \mathcal{M}(\mu, \nu)$ by first transporting μ_j to $S^{\nu}(\mu_j)$, then removing $S^{\nu}(\mu_j)$ from ν, and continuing in the same way for $\mu_{j+1}, \dots, \mu_n, \mu_1, \dots, \mu_{j-1}$. If $\pi_1, \dots, \pi_n$ all coincide, then as the image of μ_j under π_j is $S^{\nu}(\mu_j)$, we have $\nu = \sum_{j=1}^{n} S^{\nu}(\mu_j)$, a contradiction. □

The second result further characterizes when the singleton $\mathcal{M}(\mu, \nu)$ consists of an MMT.

Proposition 4.15. *The MT between μ and ν is unique and is an MMT if and only if for any $n \in \mathbb{N}$ and mutually singular $\mu_1, \dots, \mu_n \leqslant \mu$, the shadows $S^{\nu}(\mu_1), \dots, S^{\nu}(\mu_n)$ are mutually singular.*

Proof. We first show the "if" statement. Suppose that $\mu_1, \dots, \mu_n \leqslant \mu$ are mutually singular and satisfy $\sum_{j=1}^{n} \mu_j = \mu$. If $S^{\nu}(\mu_1), \dots, S^{\nu}(\mu_n)$ are mutually singular, then $\nu = \sum_{j=1}^{n} S^{\nu}(\mu_j)$ and Proposition 4.14 shows that the MT is unique. Next, we show that this MT is an MMT. As seen in the proof of Proposition 4.14, the MT transports any $\mu' \leqslant \mu$ to $S^{\nu}(\mu')$. For $N \in \mathbb{N}$, we divide $\mathbb{R}$ into countably many disjoint subsets A_i^N, $i \in \mathbb{N}$, each of length $1/N$. The mutual singularity assumption ensures that the set B^N of points y which transport (in the $\nu \to \mu$ direction) to at least two different subsets in $\{A_i^N : i \in \mathbb{N}\}$ is ν-negligible. Thus, $\nu(\bigcup_{N \in \mathbb{N}} B^N) = 0$, showing that the set of points y that map to a single x has ν-measure 1. In other words, the MT is an MMT.

To see the "only if" statement, let $\mu_1, \dots, \mu_n \leqslant \mu$ be mutually singular. We may assume that $\sum_{j=1}^{n} \mu_j = \mu$. Suppose that the MT is unique, then $\nu = \sum_{j=1}^{n} S^{\nu}(\mu_j)$ by Proposition 4.14. If $S^{\nu}(\mu_1)$ and $S^{\nu}(\mu_2)$ are not mutually singular, then points in their common part must be transported to two disjoint sets supporting μ_1 and μ_2, so that this MT is not an MMT. □

As seen in Example 2.8, uniqueness of MMT does not imply uniqueness of MT. Therefore, uniqueness of MMT is not sufficient for the conditions in Proposition 4.14 or Proposition 4.15.

4.5 Proof of Theorem 2.5

Continuing the study of uniqueness, we now aim to characterize the uniqueness of MMT and MT more explicitly for $\mu \leqslant_{\mathrm{cx}} \nu$ with ν atomless.

Lemma 4.16. *Suppose that ν is atomless and there is a unique MMT. For any $\gamma_1 = a_1\delta_{x_1}, \gamma_2 = a_2\delta_{x_2}$ with $x_1 \neq x_2$ and $\gamma_1 + \gamma_2 \leqslant \mu$, we have $S^{\nu - S^\nu(\gamma_1)}(\gamma_2) = S^\nu(\gamma_2)$. In particular, $S^\nu(\gamma_1)$ and $S^\nu(\gamma_2)$ are restrictions of ν to disjoint intervals.*

Proof. Recall that shadows are associative (Lemma 4.6). As $\gamma_1 \leqslant_{\mathrm{cx}} S^\nu(\gamma_1)$ and $\gamma_2 \leqslant_{\mathrm{cx}} S^{\nu - S^\nu(\gamma_1)}(\gamma_2)$, by Theorem 2.1 we obtain two MMTs, say π_1 from γ_1 to $S^\nu(\gamma_1)$ and π_2 from γ_2 to $S^{\nu - S^\nu(\gamma_1)}(\gamma_2)$. Moreover, $\mu - \gamma_1 - \gamma_2 \leqslant_{\mathrm{cx}} \nu - S^\nu(\gamma_1) - S^{\nu - S^\nu(\gamma_1)}(\gamma_2) = \nu - S^\nu(\gamma_1 + \gamma_2)$, yielding another MMT π_3 from $\mu - \gamma_1 - \gamma_2$ to $\nu - S^\nu(\gamma_1 + \gamma_2)$. By Lemma 4.8, the measures $S^\nu(\gamma_1)$, $S^{\nu - S^\nu(\gamma_1)}(\gamma_2)$ and $\nu - S^\nu(\gamma_1 + \gamma_2)$ are mutually singular. Thus, we may aggregate π_i, $i = 1, 2, 3$ to get an MMT π from μ to ν.

Repeat the above construction switching the roles of γ_1, γ_2. The resulting MMT π' transports γ_2 to $S^\nu(\gamma_2)$. As π transports γ_2 to $S^{\nu - S^\nu(\gamma_1)}(\gamma_2)$ and $\pi = \pi'$ by the assumed uniqueness, we conclude $S^{\nu - S^\nu(\gamma_1)}(\gamma_2) = S^\nu(\gamma_2)$. The last statement then follows from Lemma 4.8. □

Proof of Theorem 2.5. Clearly (i) implies (ii). To see that (ii) implies (iii), suppose that the MMT from μ to ν is unique. Consider the atomic part $\mu_a := \sum_{j \in \mathbb{N}} a_j \delta_{x_j}$ of μ where the x_j are distinct. Applying Lemma 4.16 with $\gamma_1 = a_j\delta_{x_j}$ and $\gamma_2 = a_{j'}\delta_{x_{j'}}$ yields that the shadows $S^\nu(a_j\delta_{x_j})$ are restrictions of ν to disjoint intervals. Removing μ_a and its shadow, we may thus assume that μ is atomless and prove $\mu = \nu$. Suppose that $\mu \neq \nu$. There exists an interval $[a, b]$ such that $\mu([a, b]) > \nu([a, b])$. More precisely, we can find $a < b$ and $\varepsilon_1, \varepsilon_2 > 0$ such that

$$0 < \nu([a - \varepsilon_1, a]), \nu([b, b + \varepsilon_2]) < \frac{\mu([a, b]) - \nu([a, b])}{2} \quad \text{and} \quad \mu([a - \varepsilon_1, a]), \mu([b, b + \varepsilon_2]) > 0.$$

The minimality property of the shadow implies that either (a) $\nu|_{[a-\varepsilon_1, a]} \leqslant S^\nu(\mu|_{[a,b]})$ or (b) $\nu|_{[b, b+\varepsilon_2]} \leqslant S^\nu(\mu|_{[a,b]})$. Suppose that (a) holds. Similarly as in the proof of Lemma 4.16, taking shadow first on $\mu|_{[a,b]}$ and then on $\mu|_{[a-\varepsilon_1, a]}$, or vice versa, yields different MMTs, a contradiction. Case (b) is analogous, thus (ii) implies (iii).

Suppose that (iii) holds and consider mutually singular $\mu_1, \ldots, \mu_n \leqslant \mu$ satisfying $\sum_{j=1}^n \mu_j = \mu$. Decompose them into an atomic part μ_j^a and a continuous part μ_j^c. Then by (iii) and Lemma 4.7, $S^\nu(\mu_j^a)$ and $S^\nu(\mu_j^c) = \mu_j^c$ are mutually singular, and these are mutually singular for distinct j's because $(\mu_j^a)_{1 \leqslant j \leqslant n}$ are mutually singular. This implies $\nu = \sum_{j=1}^n S^\nu(\mu_j)$. Thus Proposition 4.14 shows that (i) holds, completing the proof. □

5 Concluding remarks

In this section, we briefly discuss some open problems.

MMT in higher dimensions. The present paper focuses on martingale transport on $\mathbb{R}$. Starting with Ghoussoub et al. (2019), Obłój and Siorpaes (2017), and De March and Touzi (2019), martingale transport in $\mathbb{R}^d$ has been actively studied in the recent literature, but is well known to be intricate. See, e.g., Wiesel and Zhang (2022) for further references. We continue to use $\mathcal{M}(\mu,\nu)$ (resp. $\mathcal{M}_M(\mu,\nu)$) for the set of all martingale (resp. Monge martingale) transports between μ and ν.

A crucial ingredient in analyzing martingale transport in higher dimensions is the irreducible decomposition, which disintegrates the martingale transport problem into irreducible components. Following De March and Touzi (2019), let $\hat{\mathcal{K}}$ be the set of all convex closed subsets of $\mathbb{R}^d$. For probability measures μ,ν on $\mathbb{R}^d$, the irreducible components map $I:\mathbb{R}^d\to\hat{\mathcal{K}}$ is the (μ-a.e. unique) map such that for some $\hat{\mathbb{P}}\in\mathcal{M}(\mu,\nu)$, $\operatorname{ri}\operatorname{conv}\operatorname{supp}\mathbb{P}_X\subseteq I(X)=\operatorname{ri}\operatorname{conv}\operatorname{supp}\hat{\mathbb{P}}_X$ holds μ-a.e. (where $X\overset{\text{law}}{\sim}\mu$ and $\{\mathbb{P}_x\}_{x\in\mathbb{R}^d}$ is the disintegration of $\mathbb{P}$), for all $\mathbb{P}\in\mathcal{M}(\mu,\nu)$. Moreover, $\{I(x):x\in\mathbb{R}^d\}$ forms a partition of $\mathbb{R}^d$. We may further disintegrate ν into $\{\nu_x:x\in\mathbb{R}^d\}$ along such a partition.

Conjecture 5.1. Let μ,ν be probability measures on $\mathbb{R}^d$ satisfying $\mu\leqslant_{\text{cx}}\nu$. Suppose that ν_x is atomless for μ-a.e. $x\in\mathbb{R}^d$. Then $\mathcal{M}_M(\mu,\nu)$ is weakly dense in $\mathcal{M}(\mu,\nu)$. If μ is discrete, it is also dense for the ∞-Wasserstein topology.

In particular, an analogue of the existence result in Theorem 2.1 may pave the path to a denseness result along the lines of Theorem 2.3 with similar proof ideas. The main difficulty in proving Conjecture 5.1 lies in constructing a suitable analogue of the left-curtain coupling in higher dimensions. Note also that in dimension $d=1$, the irreducible decomposition (cf. Remark 2.6) is countable, so assuming non-atomicity before the irreducible decomposition is sufficient. The following remark shows that the absence of atoms (before the irreducible decomposition) is not sufficient for existence in dimensions $d>1$.

Remark 5.2. Naïve analogues of Theorem 2.1 and Theorem 3.1 in $\mathbb{R}^d$, assuming only that the marginals are atomless, are false. Let μ be uniform on $[0,1]\times\{0,\pm1\}$ and ν uniform on $[0,1]\times\{\pm2\}$. Let $(X,Y)=((X_1,X_2),(Y_1,Y_2))$ be a martingale transport; then (X_1,Y_1) is a martingale with both marginals $\text{Unif}[0,1]$, so that $X_1=Y_1$. Moreover, (X_2,Y_2) is the unique (in law) martingale from $\text{Unif}\{0,\pm1\}$ to $\text{Unif}\{\pm2\}$. We see that $\mathcal{M}(\mu,\nu)$ is a singleton, and this martingale transport is clearly not (backward) Monge. In the language of De March and Touzi (2019), the irreducible decomposition corresponds to disintegration along the first coordinate; cf. Example 2.2 of Obłój and Siorpaes (2017). As seen in Remark 3.2, this non-existence of an MMT also precludes the assertion of Theorem 3.1.

Denseness results under different constraints. Going back to transports on $\mathbb{R}$, let us turn to a different generalization, namely the constraint. We have seen that martingale transports typically do not admit Monge maps in the forward direction and that the left-curtain transport is supported on the union of two graphs. These facts are due to the martingale constraint. Similar phenomena arise for other constraints, in particular the supermartingale constraint $\mathbb{E}[Y|X]\leqslant X$ of Nutz and Stebegg (2018), Bayraktar at al. (2021, 2022) and the directional constraint $X\leqslant Y$ of Nutz and

Wang (2022). A supermartingale coupling between μ and ν exists if and only if $\mu \leqslant_{\mathrm{cd}} \nu$ (meaning that $\int \phi \, \mathrm{d}\mu \leqslant \int \phi \, \mathrm{d}\nu$ for all convex decreasing ϕ), and a coupling (X, Y) of μ and ν satisfying the directional constraint $X \leqslant Y$ exists if and only if $\mu \leqslant_{\mathrm{st}} \nu$ (meaning that their cdfs satisfy $F_\mu \geqslant F_\nu$). We speculate that, in analogy with Theorem 2.3, the set of constrained (backward) Monge transports is dense also in those settings, and possibly for other constraints.

Conjecture 5.3. Let $\mu \leqslant_{\mathrm{cd}} \nu$ with ν atomless. Then the set of (backward) Monge supermartingale couplings is weakly dense in the set of supermartingale couplings between μ and ν.

Conjecture 5.4. Let $\mu \leqslant_{\mathrm{st}} \nu$ with ν atomless. Then the set of (backward) Monge couplings (X, Y) satisfying $X \leqslant Y$ is weakly dense in the set of all couplings (X, Y) satisfying $X \leqslant Y$ between μ and ν.

Acknowledgements

We thank two anonymous referees for their careful reading of a previous version of this paper and for pointing out Remark 2.6.

References

Ahmad, N., Kim, H. K., and McCann, R. J. (2011). Optimal transportation, topology and uniqueness. *Bulletin of Mathematical Sciences*, **1**(1), 13–32.

Ambrosio, L. (2003). Lecture notes on optimal transport problems. In *Mathematical Aspects of Evolving Interfaces* (pp. 1–52). Springer, Berlin, Heidelberg.

Backhoff-Veraguas, J. Beiglböck, M., and Pammer, G. (2019). Existence, duality, and cyclical monotonicity for weak transport costs. *Calculus of Variations and Partial Differential Equations*, **58**(6), 1–28.

Backhoff-Veraguas, J. and Pammer, G. (2022). Applications of weak transport theory. *Bernoulli*, **28**(1), 370–394.

Bayraktar, E., Deng, S., and Norgilas, D. (2021). A potential-based construction of the increasing supermartingale coupling. *The Annals of Applied Probability*, **33**(5), 3803–3834.

Bayraktar, E., Deng, S., and Norgilas, D. (2022). Supermartingale shadow couplings: the decreasing case. *Bernoulli*, **30**(1), 143–169.

Beiglböck, M., Cox, A. M., and Huesmann, M. (2017). Optimal transport and Skorokhod embedding. *Inventiones Mathematicae*, **208**(2), 327–400.

Beiglböck, M., Henry-Labordère, P., and Penkner, F. (2013). Model-independent bounds for option prices—a mass transport approach. *Finance and Stochastics*, **17**(3), 477–501.

Beiglböck, M. and Juillet, N. (2016). On a problem of optimal transport under marginal martingale constraints. *The Annals of Probability*, **44**(1), 42–106.

Beiglböck, M. and Juillet, N. (2021). Shadow couplings. *Transactions of the American Mathematical Society*, **374**(7), 4973–5002.

Beiglböck, M., Nutz, M., and Touzi, N. (2017). Complete duality for martingale optimal transport on the line. *The Annals of Probability*, **45**(5), 3038–3074.

Beiglböck, M., Nutz, M., and Stebegg, F. (2021). Fine properties of the optimal Skorokhod embedding problem. *Journal of the European Mathematical Society*, **24**(4), 1389–1429.

De March, H. (2018). Quasi-sure duality for multi-dimensional martingale optimal transport. *ArXiv preprint arXiv:1805.01757*.

De March, H. and Touzi, N. (2019). Irreducible convex paving for decomposition of multidimensional martingale transport plans. *The Annals of Probability*, **47**(3), 1726–1774, 2019.

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. *The Annals of Applied Probability*, **24**(1), 312–336.

Ghoussoub, N., Kim, Y. H., and Lim, T. (2019). Structure of optimal martingale transport plans in general dimensions. *The Annals of Probability*, **47**(1), 109–164.

Gozlan, N. and Juillet, N. (2020). On a mixture of Brenier and Strassen theorems. *Proceedings of the London Mathematical Society*, **120**(3), 434–463.

Gozlan, N., Roberto, C., Samson, P. M., and Tetali, P. (2017). Kantorovich duality for general transport costs and applications. *Journal of Functional Analysis*, **273**(11), 3327–3405.

Gyöngy, I. (1986). Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probability Theory and Related Fields*, **71**(4), 501–516.

Henry-Labordère, P. (2017). *Model-free Hedging: A Martingale Optimal Transport Viewpoint*. Chapman and Hall/CRC.

Henry-Labordère, P. and Touzi, N. (2016). An explicit martingale version of the one-dimensional Brenier theorem. *Finance and Stochastics*, **20**(3), 635–668.

Hirsch, F., Profeta, C., Roynette, B. and Yor, M. (2011). *Peacocks and Associated Martingales, with Explicit Constructions*. Springer Science and Business Media.

Hobson, D. (2011). The Skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance 2010* (pp. 267–318). Springer, Berlin, Heidelberg.

Hobson, D. and Norgilas, D. (2019). The left-curtain martingale coupling in the presence of atoms. *The Annals of Applied Probability*, **29**(3), 1904–1928.

Kramkov, D. and Xu, Y. (2022). An optimal transport problem with backward martingale constraints motivated by insider trading. *The Annals of Applied Probability*, **32**(1), 294–326.

Nutz, M. and Stebegg, F. (2018). Canonical supermartingale couplings. *The Annals of Probability*, **46**(6), 3351–3398.

Obłój, J. and Siorpaes, P. (2017). Structure of martingale transports in finite dimensions. *ArXiv preprint arXiv:1702.08433*.

Nutz, M. and Wang, R. (2022). The directional optimal transport. *The Annals of Applied Probability*, **32**(2), 1400–1420.

Pratelli, A. (2007). On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*,

43(1), 1–13.

Rüschendorf, L. (2013). *Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer, Heidelberg.

Santambrogio, F. (2015). *Optimal Transport for Applied Mathematicians.* Springer, New York.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer, New York.

Strassen, V. (1965). The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, **36**(2), 423–439.

Villani, C. (2003). *Topics in Optimal Transportation*. GSM 58, American Mathematical Society.

Villani, C. (2009). *Optimal Transport: Old and New*. Springer-Verlag, Berlin.

Wiesel, J. and Zhang, E. (2022). An optimal transport-based characterization of convex order. *Dependence Modeling*, **11**(1).