CANONICAL SUPERMARTINGALE COUPLINGS

MARCEL NUTZ AND FLORIAN STEBEGG

Abstract. Two probability distributions μ and ν in second stochastic order can be coupled by a supermartingale, and in fact by many. Is there a canonical choice? We construct and investigate two couplings which arise as optimizers for constrained Monge–Kantorovich optimal transport problems where only supermartingales are allowed as transports. Much like the Hoeffding–Fréchet coupling of classical transport and its symmetric counterpart, the antitone coupling, these can be characterized by order-theoretic minimality properties, as simultaneous optimal transports for certain classes of reward (or cost) functions, and through no-crossing conditions on their supports; however, our two couplings have asymmetric geometries. Remarkably, supermartingale optimal transport decomposes into classical and martingale transport in several ways.

1. Introduction

Let μ and ν be probability measures on the real line. A measure P on $\mathbb{R}^2$ whose first and second marginals are μ and ν, respectively, is called a coupling (or transport) of μ and ν, and the set of all such measures is denoted by $\Pi(\mu,\nu)$. We shall be interested in couplings that are supermartingales; that is, if (X,Y) denotes the identity on $\mathbb{R}^2$, then $E^P[Y|X] \leq X$ P-a.s. Thus, we assume throughout that μ and ν have a finite first moment, and denote by $\mathcal{S}(\mu,\nu)$ the set of supermartingale couplings. A classical result of Strassen (cf. Proposition 2.1) shows that $\mathcal{S}(\mu,\nu)$ is nonempty if and only if μ and ν are in convex-decreasing (or second stochastic) order, denoted $\mu \leq_{cd} \nu$ and defined by the requirement that $\mu(\phi) \leq \nu(\phi)$ for any convex and decreasing function ϕ, where $\mu(\phi) := \int \phi \, d\mu$. Given $\mu \leq_{cd} \nu$, there are typically infinitely many supermartingale couplings. Our question: are there some special, "canonical" choices? The aim of this paper is to introduce and describe two such couplings, called the *Increasing* and the *Decreasing Supermartingale Transport* and denoted $\vec{P}$ and $\overleftarrow{P}$, respectively. They have

Date: November 26, 2017.

2010 *Mathematics Subject Classification.* 60G42; 49N05.

Key words and phrases. Coupling; Optimal Transport; Spence–Mirrlees Condition.

M. Nutz is supported by an Alfred P. Sloan Fellowship and NSF Grant DMS-1512900. The authors would like to thank Mathias Beiglböck, Nicolas Juillet, Jan Obłój, Nizar Touzi and an anonymous referee for encouragement and advice.

remarkable properties that are, in several ways, analogous to the Hoeffding–Fréchet and antitone couplings which can be considered canonical choices in $\Pi(\mu,\nu)$ but typically are not supermartingales. The study undertaken is also a model problem of optimal transport under inequality constraints. We shall see that the supermartingale transport problem decomposes into unconstrained and equality (martingale) constrained transport, in multiple and sometimes unexpected ways.

1.1. **Synopsis.** The couplings $\vec{P}$ and $\overleftarrow{P}$ will be characterized in three different ways: an order-theoretic minimality property, optimality for a specific class of transport reward (or cost) functions, and a geometric property of the support stating that certain paths do or do not intersect.

Let us begin with the order-theoretic characterization. To explain the idea, suppose that μ consists of finitely many atoms at $x_1,\dots,x_n\in\mathbb{R}$, then a coupling of μ and ν can be defined by specifying a "destination" measure for each atom. We know from Strassen's result that the convex-decreasing order plays a special role, so it is natural to rank all possible destination measures for the first atom (as allowed by the given marginal ν and the supermartingale constraint) according to that order. A minimal element $S^\nu(\mu|_{x_1})$ called the *shadow* will be shown to exist; essentially, it maximizes the barycenter of the destination measure and minimizes the variance. The procedure can be iterated after subtracting $S^\nu(\mu|_{x_1})$ from ν, and that determines a supermartingale coupling of μ and ν. Depending on the order in which the atoms are processed, the coupling will have a very different structure. Two obvious choices are the increasing and the decreasing order of the x_k, and that gives rise to $\vec{P}$ and $\overleftarrow{P}$ (the arrows representing the order of processing). In the general, continuum version of the construction, we instead specify the destination of $\mu|_{(-\infty,x]}$ and $\mu|_{[x,\infty)}$ for each $x\in\mathbb{R}$. The following is taken from Theorem 6.6 in the body of the paper; the precise definition of the shadow can be found in Lemma 6.2.

Theorem 1.1. *There exists a unique measure $\vec{P}$ on $\mathbb{R}^2$ which transports $\mu|_{(-\infty,x]}$ to its shadow $S^\nu(\mu|_{(-\infty,x]})$ for all $x\in\mathbb{R}$. Similarly, there exists a unique measure $\overleftarrow{P}$ which transports $\mu|_{[x,\infty)}$ to $S^\nu(\mu|_{[x,\infty)})$ for all $x\in\mathbb{R}$. Moreover, these two measures are elements of $\mathcal{S}(\mu,\nu)$.*

While the shadow construction illuminates the local order-theoretic nature of the couplings, it does not reveal the global geometric structure that is apparent in Figures 1 and 2 (rendered on page 3). The figures show simulations of $\vec{P}$ and $\overleftarrow{P}$ for piecewise uniform marginals and discrete marginals; the mass is transported from the x-axis (top) to the y-axis (bottom).

The Monge–Kantorovich optimal transport problem is a framework that enables a geometric description for its optimal transports, and thus it is

Increasing Supermartingale Transport $\vec{P}$

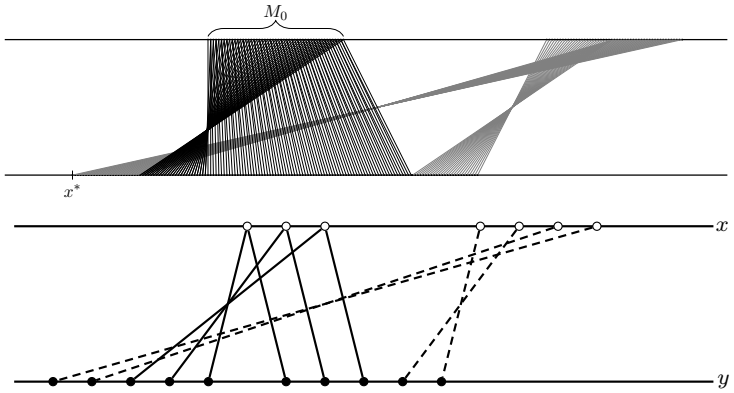

FIGURE 1. Simulations of the Increasing Supermartingale Transport. We observe an interval of Left-Curtain kernels (black/continuous) on the left and an interval of antitone kernels (gray/dashed) on the right. The destinations of the right interval are on both sides of the destinations of the left one. (The definitions of x^* and M are given in Sections 3 and 5, respectively.)

Decreasing Supermartingale Transport $\overleftarrow{P}$

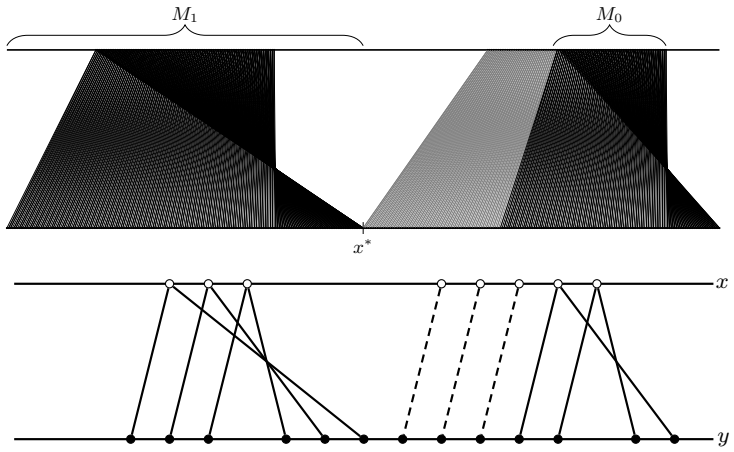

FIGURE 2. Simulations of the Decreasing Supermartingale Transport. We observe an interval of Right-Curtain kernels on the left, followed by an interval of Hoeffding–Fréchet kernels and another interval of Right-Curtain kernels. The destinations of these intervals preserve the original order.

desirable to represent $\vec{P}$ and $\overleftarrow{P}$ as corresponding solutions. More precisely, we shall introduce the *supermartingale optimal transport* problem

$$\sup_{P\in\mathcal{S}(\mu,\nu)} P(f) \tag{1.1}$$

where transports are required to be supermartingales, and then $\vec{P}, \overleftarrow{P}$ will be optimizers for reward functions f satisfying certain geometric properties. To make the connection with other texts on optimal transport, notice that $P(f) = E^P[f(X,Y)]$ in our notation, and that f can be seen as a cost function by a change of sign. We shall say that $f : \mathbb{R}^2 \to \mathbb{R}$ is *supermartingale Spence–Mirrlees* if

$$f(x_2,\cdot) - f(x_1,\cdot) \text{ is strictly decreasing and strictly convex for all } x_1 < x_2. \tag{1.2}$$

If f is smooth, this can be expressed through the cross-derivatives conditions $f_{xy} < 0$ and $f_{xyy} > 0$; the first one is the negative of the classical Spence–Mirrlees condition and the second is the so-called martingale Spence–Mirrlees condition. The following is a slightly simplified statement of Corollary 9.4.

Theorem 1.2. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be Borel, supermartingale Spence–Mirrlees and suppose that there exist $a \in L^1(\mu)$, $b \in L^1(\nu)$ such that*

$$|f(x,y)| \le a(x) + b(y), \quad x,y \in \mathbb{R}.$$

Then, $\vec{P}$ is the unique solution of the supermartingale optimal transport problem (1.1)*. Similarly, $\overleftarrow{P}$ is the unique solution of $\inf_{P\in\mathcal{S}(\mu,\nu)} P(f)$, or equivalently of* (1.1) *if instead $-f$ is supermartingale Spence–Mirrlees.*

Since $\vec{P}$ and $\overleftarrow{P}$ correspond to the combinations $f_{xy} < 0, f_{xyy} > 0$ and $f_{xy} > 0, f_{xyy} < 0$ of known conditions, it is natural to ask for the remaining two combinations, $f_{xy} > 0, f_{xyy} > 0$ and $f_{xy} < 0, f_{xyy} < 0$. While the associated optimal transports also have interesting features, they turn out to depend on the function f within that class and hence, *cannot* be called canonical; cf. Section 10.2.

The third characterization of $\vec{P}$ and $\overleftarrow{P}$ is through their supports. A point (x,y) in the support can be thought of as a path that the transport is using, and the conditions are expressed as crossing or no-crossing conditions between the paths of the transport. While this characterization is an incarnation of the *c-cyclical monotonicity* of classical transport, the supermartingale constraint requires a novel distinction of the origins x into a set M of "martingale points" and their complement. Intuitively, the supermartingale constraint is binding at points of M and absent elsewhere—this will be made precise later on (Corollary 5.3). Thus, we work with a Borel set $\Gamma \in \mathcal{B}(\mathbb{R}^2)$ that should be thought of as a support, and a second set

$M \in \mathcal{B}(\mathbb{R})$. Consider arbitrary paths $(x_1, y_1), (x_2, y_2) \in \Gamma$ with $x_1 < x_2$; then, we call the pair (Γ, M)

(i) first-order left-monotone if $y_1 \leq y_2$ whenever $x_2 \notin M$,
(ii) first-order right-monotone if $y_2 \leq y_1$ whenever $x_1 \notin M$.

We also need the following properties of Γ alone: considering three paths $(x, y_1), (x, y_2), (x', y') \in \Gamma$ with $y_1 < y_2$, the set Γ is second-order left-monotone (right-monotone) if $y' \notin (y_1, y_2)$ whenever $x < x'$ ($x > x'$). The latter two properties are taken from [9] where they are simply called left- and right-monotonicity, and all four properties are summarized in Figure 3.

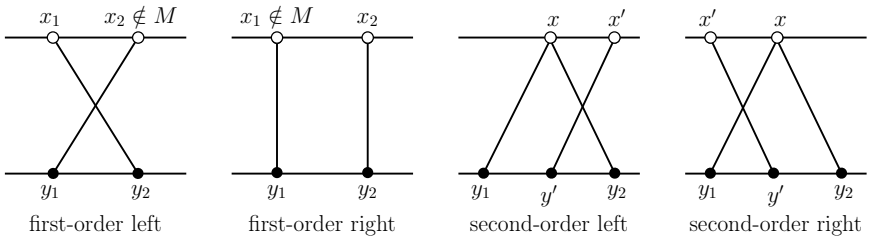

FIGURE 3. *Forbidden* configurations in the monotonicity properties

The following result is the summary of Theorem 8.1 and Corollary 9.5 in the body of the paper.

Theorem 1.3. *There exist nondegenerate*[1] $(\Gamma, M) \in \mathcal{B}(\mathbb{R}^2) \times \mathcal{B}(\mathbb{R})$ *which are first-order right-monotone and second-order left-monotone such that* $\vec{P}$ *is concentrated on* Γ *and* $\vec{P}|_{M\times\mathbb{R}}$ *is a martingale. Conversely, if* (Γ, M) *have those properties and* $P \in \mathcal{S}(\mu, \nu)$ *is a transport which is concentrated on* Γ *and* $P|_{M\times\mathbb{R}}$ *is a martingale, then* $P = \vec{P}$.

The analogous statement, interchanging left and right, holds for $\overleftarrow{P}$.

With some additional work, these theorems will allow us to explain the geometric features apparent in Figures 1 and 2. To that end, let us first recall two pairs of related couplings.

Our characterizations highlight the analogies between $\vec{P}, \overleftarrow{P}$ and the classical Hoeffding–Fréchet and antitone couplings $P_{HF}, P_{AT} \in \Pi(\mu, \nu)$; see, e.g., [55, Section 3.1]. Indeed, the latter have a minimality property similar to Theorem 1.1, but for the first stochastic order instead of the convex-decreasing one. Moreover, they are optimal transports for reward functions satisfying the classical Spence–Mirrlees condition $f_{xy} > 0$ and its reverse, and they are characterized by what we called the first-order left- and right-monotonicity of their supports Γ (with $M = \mathbb{R}$).

[1]This is a minor notion detailed in Definition 7.5.

The second pair of related couplings is given by the Left- and Right-Curtain couplings P_{LC}, P_{RC} introduced in [9] where martingale transport is studied; that is, the given marginals are in *convex* order and the transports are martingales. Indeed, these couplings are special cases of $\vec{P}$ and $\overleftarrow{P}$ that arise when the marginals $\mu \leq_{cd} \nu$ have the same barycenter—this corresponds to the fact that a supermartingale with constant mean is a martingale and vice versa. In that case, the first-order properties turn out to be irrelevant: in the shadow construction, the barycenter is constant and hence only the variance needs to be minimized; the condition for the reward functions is $f_{xyy} > 0$ (or < 0), and the second-order monotonicity property of Γ alone describes the support. As we shall see, it is the *interaction* between the first and second-order properties as well as the set M that generates the rich structure of $\vec{P}$ and $\overleftarrow{P}$.

Turning to $\vec{P}$ in Figure 1, the first observation is that there are only two types of transport kernels. On the left, $\vec{P}$ uses martingale kernels of the Left-Curtain type: each point on the x-axis is mapped to two points on the y-axis, and any two points x, x' satisfy the condition of second-order left-monotonicity. On the right, the transport is of Monge-type (each point x is mapped to a single point y) and has the structure of an antitone coupling: any two paths intersect, which is the first-order right-monotonicity property. On the strength of the same property, points x in the portion to the right (thus not in M) can further be divided into two groups—the left group is mapped to points y to the right of the destinations of the martingale points, and vice versa. These facts about $\vec{P}$ are true not only in our example, but for arbitrary atomless marginals $\mu \leq_{cd} \nu$; see Remark 9.6.

Let us now turn to $\overleftarrow{P}$ in Figure 2. Similarly as before, we observe two types of paths; the Right-Curtain and the Hoeffding–Fréchet kernels. However, the intervals of martingale and deterministic transport alternate twice—there is no longer a unique phase transition; in general, there can be countably many transitions. On the other hand, the order of the intervals is now preserved by the transport—this corresponds to the combination of the first- and second-order properties. These two differences show that the geometries of $\overleftarrow{P}$ and $\vec{P}$ differ fundamentally and suggest that one cannot obtain one coupling from the other by a transformation of the marginals. By contrast, it is well known that P_{AT} can be constructed from P_{HF} via the transformation $(x, y) \mapsto (x, -y)$, whereas P_{RC} can be obtained from P_{LC} via $(x, y) \mapsto (-x, -y)$.

One common feature of $\vec{P}$ and $\overleftarrow{P}$ is that each consists of an optimal martingale transport and an optimal (unconstrained) Monge–Kantorovich transport. That turns out to be a general fact: a result that we call Extremal Decomposition (Corollary 5.3) states that given an optimal supermartingale transport P for an arbitrary reward function f, the restriction of P to $M \times \mathbb{R}$

is an optimal martingale transport and the restriction to $M^c \times \mathbb{R}$ is an optimal Monge–Kantorovich transport between its own marginals. (These marginals, however, are not easily determined a priori.)

1.2. **Methodology and Literature.** Most of our results are based on the study of the optimal transport problem (1.1). We analyze this problem for general, Borel-measurable reward functions f, formulate a corresponding dual problem and establish strong duality; i.e., absence of a duality gap and existence of dual optimizers. A formal application of Lagrange duality suggests to consider triplets $\varphi \in L^1(\mu)$, $\psi \in L^1(\mu)$, $h : \mathbb{R} \to \mathbb{R}_+$ such that

$$\varphi(x) + \psi(y) + h(x)(y - x) \geq f(x, y), \quad (x, y) \in \mathbb{R}^2 \tag{1.3}$$

and define the dual value as $\inf\{\mu(\varphi)+\nu(\psi)\}$, where the infimum is taken over all triplets. Indeed, φ and ψ are Lagrange multipliers for the constraints μ and ν, whereas $h(x)(y - x)$ with $h \geq 0$ represents the supermartingale constraint $E^P[Y|X] \leq X$. We refer to [44, Section 5] for an intuitive discussion of the Lagrangian approach. While the corresponding duality for standard transport (without h) is valid by the celebrated result of [48], the dual problem for the supermartingale case needs to be relaxed in three ways to avoid a duality gap and ensure dual existence (Theorem 4.11). Namely, the range of h needs to be widened on parts of the state space, the integrability of φ and ψ needs to be loosened, and the inequality (1.3) needs to be relaxed on paths (x, y) that are not used by any transport (see Section 10.1 for pertinent counterexamples). In particular, it is important to classify all obstructions to supermartingale couplings; i.e., "barriers" that cannot be crossed (Proposition 3.2). Remarkably, there are no barriers beyond a specific point as soon as the barycenters of the marginals are not identical: the state space decomposes into one half-plane behaving as in the martingale case and another half-plane behaving as in classical transport.

For the martingale transport, a related duality theory was provided in [11]. In that case, the barycenters of the marginals agree and the compactness arguments underlying the duality focus on controlling the convexity of certain functions. While we shall greatly benefit from those ideas, the supermartingale case requires us to control simultaneously first and second order properties (slope and convexity) which gives rise to substantial differences on the technical side; in fact, it turns out that controlling the slope necessitates a nontrivial increment between the barycenters of μ and ν. The above-mentioned decomposition of the state space is instrumental here.

Strong duality results in a monotonicity principle (Theorem 5.2) along the lines of the c-cyclical monotonicity condition of classical transport (e.g., [2, Theorem 2.13]): a variational result linking the optimality of a transport to the pointwise properties of its support. This principle is our main tool to

study the couplings $\vec{P}$ and $\overleftarrow{P}$, parallel to the celebrated variational principle for the martingale case in [9] which has pioneered the idea that concepts similar to cyclical monotonicity can be developed beyond the classical transport setting. In the supermartingale transport problem, the monotonicity principle has a novel form describing a pair (Γ, M) of sets as in Theorem 1.3 rather than the support Γ alone. The set M enters the variational formulation by determining the class of competitors, much like it determines which paths are subject to the first-order monotonicity condition, and turns out to be fundamental in determining the geometries of $\vec{P}$ and $\overleftarrow{P}$.

As a variational result, the monotonicity principle necessitates knowing a priori that an optimal transport exists. We show that a supermartingale Spence–Mirrlees function f is automatically continuous (Proposition 9.2) in a tailor-made topology that is coarse enough to preserve weak compactness of $\mathcal{S}(\mu, \nu)$, and that yields the required existence. This result, together with the purely geometric formulation of the Spence–Mirrlees conditions (Definition 7.1), also improves the literature on martingale transport [9, 38, 45] where a range of assumptions is imposed on f both to ensure existence and to express the Spence–Mirrlees condition in terms of partial derivatives or a specific functional form; cf. Corollary 9.4. A second generalization is that Theorem 1.2 remains true if the Spence–Mirrlees condition (1.2) is satisfied in the non-strict sense, except that the optimizer need not be unique.

With the appropriate notions in place, the proofs of Theorems 1.2 and 1.3 use the monotonicity principle to analyze the interplay between the geometry of the set M and the first- and second-order monotonicity and Spence–Mirrlees conditions. The construction of $\vec{P}$ and $\overleftarrow{P}$ with the minimality property of Theorem 1.1 rests on the precise understanding of the shadow of a single atom (Lemma 6.3) and compactness arguments; a novel phenomenon is that the barycenter of the shadow needs to be found through an optimization rather than being known a priori as in the martingale case.

To the best of our knowledge, supermartingale couplings have not been specifically studied in the extant literature. However, as indicated above, martingale optimal transport has received considerable attention since it was introduced in [6] and [29]. In particular, [9, 38, 43, 44, 45] study optimal martingale transports between two marginals for specific cost functions; the martingale Spence–Mirrlees condition in the form $f_{xyy} > 0$ appears for the first time in [38], generalizing the functional form used in [9]. The non-strict condition, as well as the geometric definition, are novelties of this paper. We also remark that some of the technical developments in Sections 7–8 provide simplifications with respect to previous works, when specialized to the martingale case.

While martingale (equality) constraint and classical (unconstrained) case can occur as special cases of supermartingale transport, the more surprising discovery is that the latter can be "built" from these two ingredients: the supermartingale (inequality) constraint is decomposed into two extremal cases. This forms a common thread in this paper, starting with the analysis of the maximal barrier which splits the plane into half-planes behaving like in these two cases and thus allows us to apply the compactness result of Proposition 4.4. The variational principle decomposes the domain into points where the supermartingale constraint is felt as an equality constraint and points where it is not felt at all, and correspondingly, the Extremal Decomposition shows that any optimal supermartingale transport can be split into a martingale-optimal one and an optimizer of an unconstrained problem. Conversely, our study of optimal supermartingale transports for Spence–Mirrlees reward functions in Sections 7–8 shows how the geometric properties of Left-Curtain and Fréchet–Hoeffding couplings interact to create the patterns of the canonical supermartingale couplings.

Martingale optimal transport is motivated by considerations of model uncertainty in financial mathematics. If, in the financial context, dynamic hedging is restricted by a no-shorting constraint, the dual problem is supermartingale transport. Thus, it can be seen as a special case of the dual problem in [27] where general portfolio constraints are studied. For background on Monge–Kantorovich transport, we refer to [2, 55, 56, 60, 61]. Recently, a rich literature has emerged around martingale transport and model uncertainty; see [41, 53, 59] for surveys and, e.g., [1, 10, 15, 16, 17, 18, 23, 27, 30, 31, 32, 50, 52, 62] for models in discrete time, [4, 14, 20, 21, 24, 25, 26, 36, 37, 40, 49, 51, 57, 58] for continuous time, and [3, 7, 8, 19, 22, 33, 34, 35, 39, 42, 46, 54] for related Skorokhod embedding and mimicking problems.

The remainder of this paper is organized as follows. While Section 2 recalls basic facts related to the convex-decreasing order, Section 3 contains a complete description of the barriers to supermartingale couplings and more precisely, the structure of $\mathcal{S}(\mu,\nu)$-polar sets. After these preparations, Section 4 presents a complete duality theory for Borel reward functions, and Section 5 formulates the resulting monotonicity principle. Section 6 introduces the couplings $\vec{P}$ and $\overleftarrow{P}$ via the shadow construction. In Section 7, we propose the Spence–Mirrlees conditions for reward functions and show via the monotonicity principle that the associated optimal transports are supported on sets (Γ, M) satisfying corresponding monotonicity properties. Section 8 continues the analysis by showing that any coupling supported on such sets must coincide with $\vec{P}$ or $\overleftarrow{P}$, respectively. In Section 9, we close the circle: Spence–Mirrlees functions are shown to admit optimal transports and on the strength of the duality theory, that allows us to conclude the existence

of suitable sets (Γ, M). The main theorems stated in the Introduction then follow. The concluding Section 10 collects a number of counterexamples.

2. Preliminaries

It will be useful to consider finite measures, not necessarily normalized to be probabilities. Let μ, ν be finite measures on $\mathbb{R}$ with finite first moment. Extending the notation from the Introduction, we write $\Pi(\mu, \nu)$ for the set of all couplings; i.e., measures P on $\mathbb{R}^2$ such that $P \circ X^{-1} = \mu$ and $P \circ Y^{-1} = \nu$, where $(X, Y) : \mathbb{R}^2 \to \mathbb{R}^2$ is the identity. Moreover, $\mathcal{S}(\mu, \nu)$ is the subset of all $P \in \Pi(\mu, \nu)$ which are supermartingales; i.e., $\int Y \mathbf{1}_A(X)\, dP \leq \int X \mathbf{1}_A(X)\, dP$ for all $A \in \mathcal{B}(\mathbb{R})$, and finally $\mathcal{M}(\mu, \nu)$ consist of all $P \in \Pi(\mu, \nu)$ satisfying this condition with equality.

We say that μ and ν are in *convex-decreasing order*, or *second stochastic order*, denoted $\mu \leq_{cd} \nu$, if $\mu(\phi) \leq \nu(\phi)$ for any convex, nonincreasing function $\phi : \mathbb{R} \to \mathbb{R}$. It then follows that μ and ν have the same total mass; moreover, we shall use repeatedly that it suffices to check the inequality for functions ϕ of linear growth. An alternative characterization of this order refers to the put (price) function, defined by

$$p_\mu : \mathbb{R} \to \mathbb{R}, \quad p_\mu(t) := \int (t - s)^+ \, \mu(ds).$$

Writing $\mathrm{bary}(\mu) := (\int x \, d\mu)/\mu(\mathbb{R})$ for the barycenter (with $\mathrm{bary}(\mu) := 0$ if $\mu = 0$) and $\partial^\pm p_\mu$ for the right and left derivatives, the following properties are easily verified:

(i) p_μ is nonnegative, increasing[2], convex, and $\partial^+ p_\mu(t) - \partial^- p_\mu(t) = \mu(\{t\})$,
(ii) $\lim_{t \to -\infty} p_\mu(t) = 0$ and $\lim_{t \to \infty} p_\mu(t) = \infty \mathbf{1}_{\mu \neq 0}$,
(iii) $\lim_{t \to \infty} \{p_\mu(t) - \mu(\mathbb{R})[t - \mathrm{bary}(\mu)]\} = 0$.

In particular, we may extend p_μ continuously to $\overline{\mathbb{R}} = [-\infty, \infty]$. The following result is classical; see, e.g., [28, Theorem 2.58].

Proposition 2.1. *Let μ, ν be finite measures on $\mathbb{R}$ with finite first moment and $\mu(\mathbb{R}) = \nu(\mathbb{R})$. The following are equivalent:*

(i) $\mu \leq_{cd} \nu$,
(ii) $p_\mu \leq p_\nu$,
(iii) $\mathcal{S}(\mu, \nu) \neq \emptyset$,
(iv) there exists a stochastic kernel $\kappa(x, dy)$ with finite mean such that $\int y \, \kappa(x, dy) \leq x$ for all $x \in \mathbb{R}$ and $\nu = (\mu \otimes \kappa) \circ Y^{-1}$, where $\mu \otimes \kappa$ denotes the product.

In all that follows, the statement $\mu \leq_{cd} \nu$ implicitly means that μ, ν are finite measures on $\mathbb{R}$ with finite first moment. Moreover, such a pair and the corresponding supermartingale optimal transport problem will be called

[2]Throughout this paper, increasing means nondecreasing.

proper if the barycenters of μ and ν do *not* coincide. In the improper case, the problem degenerates to a martingale optimal transport problem because any supermartingale with constant mean is a martingale. Indeed, let us convene that μ and ν are in *convex order*, denoted $\mu \leq_c \nu$, if $\mu(\phi) \leq \nu(\phi)$ for any convex function $\phi : \mathbb{R} \to \mathbb{R}$, and introduce the symmetric potential function $u_\mu : \mathbb{R} \to \mathbb{R}$ by $u_\mu(t) := \int |t-s|\, \mu(ds)$. Given $\mu \leq_{cd} \nu$, the following are then equivalent: (a) $\mathrm{bary}(\mu) = \mathrm{bary}(\nu)$, (b) $\mu \leq_c \nu$, (c) $u_\mu \leq u_\nu$, (d) $\mathcal{M}(\mu,\nu) \neq \emptyset$, (e) the kernel κ in (iv) can be chosen with $\int y\, \kappa(x, dy) = x$ for all $x \in \mathbb{R}$.

3. Barriers and Polar Sets

We fix $\mu \leq_{cd} \nu$ throughout this section. Our first aim is to characterize all points $x \in \overline{\mathbb{R}}$ which cannot be crossed by any supermartingale transport $P \in \mathcal{S}(\mu,\nu)$.

Definition 3.1. A point $x \in \overline{\mathbb{R}}$ is called a *barrier* if $Y \leq x$ P-a.s. on $\{X \leq x\}$ and $Y \geq x$ P-a.s. on $\{X \geq x\}$, for all $P \in \mathcal{S}(\mu,\nu)$.

We may note that $\pm\infty$ are always barriers. The following result not only shows how barriers can be described as points where the put functions touch, but also introduces a particular barrier x^* which divides the real line into two parts: To the left of x^*, the supermartingale transport problem is in fact just a martingale transport problem. To the right of x^*, we have a proper supermartingale transport problem and there are no non-trivial barriers. For example, in Figure 1, the point x^* is the left boundary of the support of ν, whereas in Figure 2 it indeed splits the supports of μ and ν into two parts. The convention $\sup \emptyset = -\infty$ is used.

Proposition 3.2. *Define* $x^* := \sup\{x \in \mathbb{R} : p_\mu(x) = p_\nu(x)\} \in \overline{\mathbb{R}}$. *Then*

(i) x^* *is a barrier and* $p_\mu(x^*) = p_\nu(x^*)$,
(ii) a point $x \in [-\infty, x^*)$ *is a barrier if and only if* $p_\mu(x) = p_\nu(x)$,
(iii) if $x \in (x^*, \infty]$ *is a barrier then* $\mu(x,\infty) = \nu(x,\infty) = 0$.

Moreover, x^* *is the maximal barrier* $x \in \overline{\mathbb{R}}$ *such that* $P|_{\{X<x\}}$ *is a martingale transport for some (and then all)* $P \in \mathcal{S}(\mu,\nu)$.

The reverse implication in (iii) is almost true: a point x with $\mu(x,\infty) = \nu(x,\infty) = 0$ is not crossed by any transport. However, if μ has an atom at x, this mass may be transported to $(-\infty, x)$ and then x does not satisfy our definition of a barrier which is chosen so that any mass at the barrier remains invariant.

Before reporting the proof in Section 3.1, we use the above result to characterize the polar sets and the irreducible components.

Definition 3.3. The pair $\mu \leq_{cd} \nu$ is *irreducible* if the set $I = \{p_\mu < p_\nu\}$ is connected and $\mu(I) = \mu(\mathbb{R})$. In this situation, let J be the union of I and any endpoints of I that are atoms of ν; then (I, J) is the *domain* of (μ,ν).

This definition coincides with the notion of [9, 11] in the context of martingale transport. More precisely, for $x < x^*$, we have $p_\mu(x) = p_\nu(x)$ if and only if $u_\mu(x) = u_\nu(x)$.

In the general case, the supermartingale transport problem will be decomposed into at most countably many irreducible components. We recall that a set is called polar for a family $\mathcal{P}$ of measures if it is P-null for all $P \in \mathcal{P}$.

Proposition 3.4. *Let $\mu \leq_{cd} \nu$, let $I_0 = (x^*, \infty)$ and let $(I_k)_{1\leq k\leq N}$ be the (open) components of $\{p_\mu < p_\nu\} \cap (-\infty, x^*)$, where $N \in \{0, 1, \dots, \infty\}$.*

(i) Set $I_{-1} = \mathbb{R}\backslash \cup_{k\geq 0} I_k$ and $\mu_k = \mu|_{I_k}$ for $k \geq -1$, so that $\mu = \sum_{k\geq -1} \mu_k$. Then, there exists a unique decomposition $\nu = \sum_{k\geq -1} \nu_k$ such that

$$\mu_{-1} = \nu_{-1} \quad \text{and} \quad \mu_0 \leq_{cd} \nu_0 \quad \text{and} \quad \mu_k \leq_c \nu_k \quad \text{for all} \quad k \geq 1.$$

Moreover, this decomposition satisfies $I_k = \{p_{\mu_k} < p_{\nu_k}\}$ for all $k \geq 0$; i.e., each such pair (μ_k, ν_k) is irreducible. Finally, any $P \in \mathcal{S}(\mu, \nu)$ admits a unique decomposition $P = \sum_{k\geq -1} P_k$ such that $P_0 \in \mathcal{S}(\mu_0, \nu_0)$ and $P_k \in \mathcal{M}(\mu_k, \nu_k)$ for all $k \neq 0$.

(ii) Let $B \subseteq \mathbb{R}^2$ be a Borel set. Then B is $\mathcal{S}(\mu, \nu)$-polar if and only if there exist a μ-nullset N_μ and a ν-nullset N_ν such that

$$B \subseteq (N_\mu \times \mathbb{R}) \cup (\mathbb{R} \times N_\nu) \cup \left(\Delta \cup \bigcup_{k\geq 0} I_k \times J_k\right)^c,$$

where $\Delta = \{(x, x) \in \mathbb{R}^2 : x \in \mathbb{R}\}$ is the diagonal and J_k is constructed from I_k as in Definition 3.3.

3.1. **Proofs of Propositions 3.2 and 3.4.** We begin with the proof of Proposition 3.2, stated through a sequence of lemmas. We may assume that μ and ν are probability measures.

Lemma 3.5. *Let $x \in \overline{\mathbb{R}}$. If $p_\mu(x) = p_\nu(x)$, then x is a barrier and the equality $E^P[X\mathbf{1}_{X<x}] = E^P[Y\mathbf{1}_{X<x}]$ holds for all $P \in \mathcal{S}(\mu, \nu)$.*

Proof. Let $p_\mu(x) = p_\nu(x)$ and let $E[\,\cdot\,]$ be the expectation associated with an arbitrary $P \in \mathcal{S}(\mu, \nu)$. Using $E[Y|X] \leq X$ and Jensen's inequality, $(x-X)^+ \leq (x-E[Y|X])^+ \leq E[(x-Y)^+|X]$, and since $p_\mu(x) = p_\nu(x)$ means that $E[(x-X)^+] = E[(x-Y)^+]$, it follows that $(x-X)^+ = E[(x-Y)^+|X]$.

As a first consequence, $E[(x-Y)^+\mathbf{1}_{X\geq x}] = E[(x-X)^+\mathbf{1}_{X\geq x}] = 0$ and hence $Y \geq x$ P-a.s. on $\{X \geq x\}$. A second consequence is $E[(x-Y)\mathbf{1}_{X\leq x}] \leq E[(x-Y)^+\mathbf{1}_{X\leq x}] = E[(x-X)^+\mathbf{1}_{X\leq x}]$. Since $E[Y|X] \leq X$ implies that $E[(x-Y)\mathbf{1}_{X\leq x}] \geq E[(x-X)\mathbf{1}_{X\leq x}] = E[(x-X)^+\mathbf{1}_{X\leq x}]$, it follows that $E[(x-Y)\mathbf{1}_{X\leq x}] = E[(x-Y)^+\mathbf{1}_{X\leq x}]$ and thus $Y \leq x$ P-a.s. on $\{X \leq x\}$. This completes the proof of the barrier property.

The above inequalities also show that $E[(x-Y)\mathbf{1}_{X\leq x}] = E[(x-X)\mathbf{1}_{X\leq x}]$ and hence $E[Y\mathbf{1}_{X\leq x}] = E[X\mathbf{1}_{X\leq x}]$. To infer the second assertion, it remains to note that the barrier property implies that $Y = x$ P-a.s. on $\{X = x\}$. $\square$

Corollary 3.6. *We have $p_\mu(x^*) = p_\nu(x^*)$ and $E^P[X\mathbf{1}_{X<x^*}] = E^P[Y\mathbf{1}_{X<x^*}]$ for all $P \in \mathcal{S}(\mu,\nu)$.*

Proof. The claim is trivial if $x^* = -\infty$. Otherwise, the first claim follows from the fact that p_μ and p_ν are continuous, and then the second claim follows from Lemma 3.5. □

Lemma 3.7. *Let $x \in \overline{\mathbb{R}}$ be a barrier. The following are equivalent:*

(i) $E^P[X\mathbf{1}_{X<x}] = E^P[Y\mathbf{1}_{X<x}]$ for some (and then all) $P \in \mathcal{S}(\mu,\nu)$.
(ii) $P|_{\{X<x\}}$ is a martingale transport for some (and then all) $P \in \mathcal{S}(\mu,\nu)$.
(ii') $P|_{\{X\le x\}}$ is a martingale transport for some (and then all) $P \in \mathcal{S}(\mu,\nu)$.

Proof. If (i) holds for some $P \in \mathcal{S}(\mu,\nu)$, then (ii) holds for the same P since a supermartingale with constant mean is a martingale, and the converse holds as any martingale has constant mean. We complete the equivalence of (i) and (ii) by showing that if (i) holds for one $P \in \mathcal{S}(\mu,\nu)$, it necessarily holds for all elements of $\mathcal{S}(\mu,\nu)$. The cases $x = \pm\infty$ are clear, so let $x \in \mathbb{R}$. Let $P \in \mathcal{S}(\mu,\nu)$ and let ν' be the second marginal of $P' := P|_{\{X<x\}}$. As x is a barrier, we have $\nu' = \nu$ on $(-\infty, x)$. If $\bar{P} \in \mathcal{S}(\mu,\nu)$ is arbitrary and $\bar{P}', \bar{\nu}'$ are defined analogously, we have $\bar{\nu}' = \nu = \nu'$ on $(-\infty, x)$ by the same reasoning. But then also $\nu'(\{x\}) = \bar{\nu}'(\{x\})$, since this is the remaining mass transported from $(-\infty, x)$: we have $\nu'(\{x\}) = \mu(-\infty, x) - \nu'(-\infty, x) = \mu(-\infty, x) - \bar{\nu}'(-\infty, x) = \bar{\nu}'(\{x\})$. As a result, $\bar{\nu}' = \nu'$ on $(-\infty, x]$, and $\bar{P}'$ satisfies (i) whenever P does. Finally, (ii) implies (ii') because x is a barrier, and the reverse is clear. □

Lemma 3.8. *Let $x \in \overline{\mathbb{R}}$ be a barrier such that $P|_{\{X<x\}}$ is a martingale transport for some $P \in \mathcal{S}(\mu,\nu)$. Then $p_\mu(x) = p_\nu(x)$.*

Proof. The cases $x = \pm\infty$ are clear, so let $x \in \mathbb{R}$. The martingale property yields that $p_\mu(x) = E[(x - X)\mathbf{1}_{X<x}] = E[(x - Y)\mathbf{1}_{X<x}]$. Since $Y \le x$ P-a.s. on $\{X < x\}$ and $\{Y < x\} \subseteq \{X < x\}$ P-a.s., $E[(x - Y)\mathbf{1}_{X<x}] = E[(x - Y)^+\mathbf{1}_{X<x}] \ge E[(x - Y)^+\mathbf{1}_{Y<x}] = p_\nu(x)$. Thus, $p_\mu(x) \ge p_\nu(x)$. As the converse inequality is always true, we deduce that $p_\mu(x) = p_\nu(x)$. □

The following completes the proof of Proposition 3.2(ii).

Corollary 3.9. *Let $x \in [-\infty, x^*]$ be a barrier. Then $p_\mu(x) = p_\nu(x)$.*

Proof. We may assume that $x \in \mathbb{R}$ which entails that $x^* > -\infty$. Lemma 3.5, Corollary 3.6 and Lemma 3.7 show that the restriction of any $P \in \mathcal{S}(\mu,\nu)$ to $\{X < x^*\}$ is a martingale transport. As $x \le x^*$, the same holds for the restriction to $\{X < x\}$, and now Lemma 3.8 applies. □

Lemma 3.10. *If $\bar{x} \in (x^*, \infty]$ is a barrier, then $\mu(\bar{x}, \infty) = \nu(\bar{x}, \infty) = 0$.*

Proof. The case $\bar{x} = \infty$ is clear. Let $\bar{x} \in (x^*, \infty)$ be a barrier and suppose for contradiction that $\mu(\bar{x}, \infty) > 0$ or $\nu(\bar{x}, \infty) > 0$.

Case 1: $\nu(\bar{x},\infty)>0$. We contradict the barrier property with an element of $\mathcal{S}(\mu,\nu)$ transporting mass from $(-\infty,\bar{x})$ to $(\bar{x},\infty)$, and vice versa.

Let $P\in\mathcal{S}(\mu,\nu)$ be arbitrary and let $P=\mu\otimes\kappa$ be a disintegration such that for all $x<\bar{x}$, we have $\mathrm{bary}(\kappa(x))\leq x$ and $\kappa(x,dy)$ is concentrated on $(-\infty,\bar{x}]$ but not on $\{\bar{x}\}$; these choices are possible due to the barrier and the supermartingale property.

For each $x\in(-\infty,\bar{x})$, let $\varepsilon(x)\in[0,1]$ be the largest number such that

$$\kappa'(x):=(1-\varepsilon(x))\kappa(x)|_{(-\infty,\bar{x})}+\tilde{\varepsilon}(x)\nu|_{(\bar{x},\infty)}+\kappa(x)|_{\{\bar{x}\}}$$

satisfies $\mathrm{bary}(\kappa'(x))\leq x$; here $\tilde{\varepsilon}(x)$ is the unique constant such that $\kappa'(x)$ is a probability measure. This defines a stochastic kernel with the properties

$$\kappa'(x)\{\bar{x}\}=\kappa(x)\{\bar{x}\}\quad\text{for all } x,\quad \kappa'(x)[\bar{x},\infty)>\kappa(x)[\bar{x},\infty)\quad\text{if } \varepsilon(x)>0.$$

Moreover, $\varepsilon>0$ on a set of positive μ-measure, as otherwise $P|_{\{X<\bar{x}\}}$ is a martingale transport which would contradict $\bar{x}>x^*$ (Lemma 3.8). Let ν_2 be the restriction to $(\bar{x},\infty)$ of the second marginal of $\mu|_{(-\infty,\bar{x})}\otimes\kappa'$. By truncating the above function $\varepsilon(\cdot)$ at some positive constant $\bar{\varepsilon}$, we may assume that $\nu_2\leq\nu$ while retaining the other properties. Thus, we can define a measure $\mu_2\leq\mu$ by taking the preimage of ν_2 under P (obtained by disintegrating $P=\nu(dy)\otimes\hat{\kappa}(y,dx)$ and taking μ_2 to be the first marginal of $\nu_2(dy)\otimes\hat{\kappa}(y,dx)$). Moreover, let ν_1 be the restriction to $(-\infty,\bar{x})$ of the second marginal of $\mu|_{(-\infty,\bar{x})}\otimes\kappa\;-\;\mu|_{(-\infty,\bar{x})}\otimes\kappa'$. Then $c:=\nu_1(\mathbb{R})=\mu_2(\mathbb{R})$ and by construction,

$$\mu|_{(-\infty,\bar{x})}\otimes\kappa'\;\;+\;\;(\mu|_{[\bar{x},\infty)}-\mu_2)\otimes\kappa\;\;+\;\;c^{-1}\mu_2\otimes\nu_1$$

is an element of $\mathcal{S}(\mu,\nu)$. Since $\mu(-\infty,\bar{x})>0$ and $\nu(\bar{x},\infty)>0$, it transports mass across $\bar{x}$, contradicting that $\bar{x}$ is a barrier.

Case 2: $\mu(\bar{x},\infty)>0$ *and* $\nu(\bar{x},\infty)=0$. Note that in this case, $\nu|_{[\bar{x},\infty)}$ is concentrated at $\bar{x}$ and the entire mass $\mu(\bar{x},\infty)>0$ is transported to that atom by any $P\in\mathcal{S}(\mu,\nu)$, in addition to any mass coming from $(-\infty,\bar{x}]$. We shall contradict the barrier property by constructing an element of $\mathcal{S}(\mu,\nu)$ which transports mass from $(\bar{x},\infty)$ to $(-\infty,\bar{x})$; this will be balanced by moving appropriate mass from $(-\infty,\bar{x})$ to $\{\bar{x}\}$.

Let $P\in\mathcal{S}(\mu,\nu)$ be arbitrary and let κ be as above. For each $x\in(-\infty,\bar{x})$, let $\varepsilon(x)\in[0,1]$ be the largest number such that

$$\kappa'(x):=(1-\varepsilon(x))\kappa(x)|_{(-\infty,\bar{x})}+\tilde{\varepsilon}(x)\nu|_{\{\bar{x}\}}$$

satisfies $\mathrm{bary}(\kappa'(x))\leq x$; again, $\tilde{\varepsilon}(x)$ is the unique constant such that $\kappa'(x)$ is a probability measure. This defines a stochastic kernel with

$$\kappa'(x)\{\bar{x}\}\geq\kappa(x)\{\bar{x}\}\quad\text{for all } x,\quad \kappa'(x)\{\bar{x}\}>\kappa(x)\{\bar{x}\}\quad\text{if } \varepsilon(x)>0,$$

and again, $\varepsilon>0$ on a set of positive μ-measure. Let ν_2 be the restriction to $\{\bar{x}\}$ of the second marginal of $\mu|_{(-\infty,\bar{x})}\otimes\kappa'-\mu|_{(-\infty,\bar{x})}\otimes\kappa$. After truncating $\varepsilon(\cdot)$ we again have $\nu_2\leq\nu$; recall that P transports the mass $\mu(\bar{x},\infty)>0$

to $\bar{x}$. Continuing the construction as above, the latter property shows that $\mu_2(\bar{x}, \infty) > 0$, and the barrier property is again contradicted. □

Proof of Proposition 3.2. Proposition 3.2 is now a consequence of Lemma 3.5, Corollary 3.6, Corollary 3.9 and Lemma 3.10. □

Proof of Proposition 3.4(i). According to Proposition 3.2, we face a pure martingale transport problem on $(-\infty, x^*]$; in particular, we may apply the decomposition result of [9, Theorem 8.4] on this part of the state space to obtain ν_k and P_k for $k \geq 1$. Since x^* is itself a barrier by Proposition 3.2, the only possible choice for ν_0 is $\nu_0 = \nu|_{(x^*,\infty)} + [\mu(x^*, \infty) - \nu(x^*, \infty)]\delta_{x^*}$, and this measure satisfies $\mu_0 \leq_{cd} \nu_0$. □

We proceed towards the proof of the second part of Proposition 3.4.

Lemma 3.11. *If $\mu \leq_{cd} \nu$ is irreducible, the $\Pi(\mu, \nu)$-polar sets and the $\mathcal{S}(\mu, \nu)$-polar sets coincide.*

Proof. If μ and ν have the same barycenter, then $\mathcal{S}(\mu, \nu) = \mathcal{M}(\mu, \nu)$ and this is the result of [11, Corollary 3.4]. Thus, we may assume that (μ, ν) is proper. By Proposition 3.2, the associated domain (I, J) satisfies $I = (x^*, \infty)$ for some $x^* \in [-\infty, \infty)$, while $J = I$ if $\nu(\{x^*\}) = 0$ (including the case $x^* = -\infty$) and $J = [x^*, \infty)$ if $\nu(\{x^*\}) > 0$.

Since $\mathcal{S}(\mu, \nu) \subseteq \Pi(\mu, \nu)$, it suffices to show that for any $\pi \in \Pi(\mu, \nu)$ there exists $P \in \mathcal{S}(\mu, \nu)$ such that $P \gg \pi$. Let us show more generally that

$$\text{for any measure } \pi \text{ on } \mathbb{R}^2 \text{ with marginals } \pi_1 \leq \mu \text{ and } \pi_2 \leq \nu$$
$$\text{there exists } P \in \mathcal{S}(\mu, \nu) \text{ such that } P \gg \pi.$$

While π is necessarily supported by $I \times J$, we prove the claim under the additional condition that π is concentrated on a compact rectangle $K \times L \subset I \times J$. This entails no loss of generality: a general π may be decomposed into a sum $\pi = \sum_n \pi^n$ of measures satisfying this condition, and if P^n are the corresponding supermartingales, $P = \sum_n 2^{-n} P^n$ satisfies the claim.

The definition of (I, J) implies that ν assigns positive mass to any neighborhood of the lower endpoint x^* of J. More precisely, we can find a compact set $B \subset J$, located entirely to the left of $K \subset I$, such that $\nu(B) > 0$. (If $\nu(\{x^*\}) > 0$ we can simply take $B = \{x^*\}$.) Consider a disintegration $\pi = \pi_1 \otimes \kappa$ where $\kappa(x, dy)$ is concentrated on L for all $x \in K$. We introduce another stochastic kernel κ' of the form

$$\kappa'(x, dy) = \frac{\kappa(x, dy) + \varepsilon(x)\nu(dy)|_B}{c(x)}.$$

Here $c(x) \geq 1$ is the normalizing constant such that $\kappa'(x, dy)$ is a stochastic kernel. Moreover, $\varepsilon(x) := 0$ for x such that $\text{bary}(\kappa(x)) \leq x$, whereas for x with $\text{bary}(\kappa(x)) > x$ we let $\varepsilon(x)$ be the unique positive number such that

$\mathrm{bary}(\kappa'(x)) = x$—this number exists by the intermediate value theorem; note that B is located to the left of $x \in K$. By construction,

$$\pi' := \nu(B)\pi_1 \otimes \kappa'$$

is a supermartingale with $\pi' \gg \pi$ and its marginals satisfy $\pi'_1 \leq \pi_1 \leq \mu$ as well as $\pi'_2 \leq \nu$; the latter is due to $\pi_1(\mathbb{R}) \leq \mu(\mathbb{R}) = 1$ and $\kappa'(x) \leq \nu(B)^{-1}\,\nu|_B + \kappa(x)$ and $\kappa(x)$ being concentrated on B^c. We also note that

$$\pi' \text{ is concentrated on a quadrant of the form } [k,\infty)^2 \tag{3.1}$$

with $[k,\infty) \subseteq J$; here $k \in \mathbb{R}$ is determined by the lower bound of the set B. We shall complete the proof by constructing $P \in \mathcal{S}(\mu,\nu)$ such that $P \gg \pi'$.

(i) We first consider the case where $\nu(\{x^*\}) = 0$ and hence $I = J = (x^*,\infty)$ and $k > x^*$. Since $p_\nu - p_\mu$ is continuous, strictly positive on I and $\lim_{t\to\infty} p_\nu(t) - p_\mu(t) = \mu(\mathbb{R})[\mathrm{bary}(\nu) - \mathrm{bary}(\mu)] > 0$, we see that $p_\nu - p_\mu$ is uniformly bounded away from zero on $[k,\infty)$. On the other hand, $p_{\pi'_2} - p_{\pi'_1}$ is uniformly bounded on $[k,\infty)$ since

$$\lim_{t\to\infty} p_{\pi'_2}(t) - p_{\pi'_1}(t) = \pi'_1(\mathbb{R})[\mathrm{bary}(\pi'_1) - \mathrm{bary}(\pi'_2)] < \infty.$$

As a result, there exists $\epsilon > 0$ such that $p_\mu - \epsilon p_{\pi'_1} \leq p_\nu - \epsilon p_{\pi'_2}$ on $[k,\infty)$, but then also on $\mathbb{R}$ because $p_{\pi'_1} = p_{\pi'_2} = 0$ outside of $[k,\infty)$ due to (3.1). Noting that this inequality may also be stated as $p_{\mu-\epsilon\pi'_1} \leq p_{\nu-\epsilon\pi'_2}$, Proposition 2.1 shows that there exists some $P' \in \mathcal{S}(\mu - \epsilon\pi'_1, \nu - \epsilon\pi'_2)$, and we complete the proof by setting $P := P' + \epsilon\pi'_1(\mathbb{R})^{-1}\pi'$.

(ii) In the case $\nu(\{x^*\}) > 0$ we need to argue differently that there exists $\epsilon > 0$ such that $p_\mu - \epsilon p_{\pi'_1} \leq p_\nu - \epsilon p_{\pi'_2}$ on $[k,\infty)$. By enlarging $[k,\infty)$, we may assume that $k = x^*$ is the left endpoint of J. As $\mu(I) = \mu(\mathbb{R}) = \nu(J)$, we have $\partial^+ p_\mu(x^*) = \partial^+ p_\mu(x^*) - \partial^- p_\mu(x^*) = \mu(\{x^*\}) = 0$ and similarly

$$\partial^+ p_{\pi'_1}(x^*) = 0, \quad \partial^+ p_{\pi'_2}(x^*) = \pi'_2(\{x^*\}), \quad \partial^+ p_\nu(x^*) = \nu(\{x^*\}) > 0.$$

Since $\nu(\{x^*\}) \geq \pi'_2(\{x^*\})$, it then follows that $0 \neq \partial^+(p_\nu - p_\mu)(x^*) \geq \partial^+(p_{\pi'_2} - p_{\pi'_1})(x^*)$. The existence of the desired $\epsilon > 0$ follows and the rest of the argument is as in (i). □

Proof of Proposition 3.4(ii). By the decomposition in Proposition 3.4(i) and Lemma 3.11, a Borel set $B \subseteq \mathbb{R}^2$ is $\mathcal{S}(\mu,\nu)$-polar if and only if $B \cap (I_k \times J_k)$ is $\Pi(\mu_k,\nu_k)$-polar for all $k \geq 0$ and $B \cap \Delta$ is P_{-1}-null. It remains to apply the result of [5, Proposition 2.1] for each $k \geq 0$: a Borel set B_k is $\Pi(\mu_k,\nu_k)$-polar if and only if $B_k \subseteq (N_{\mu_k} \times \mathbb{R}) \cup (\mathbb{R} \times N_{\nu_k})$ for nullsets N_{μ_k} and N_{ν_k}. □

4. Duality Theory

In this section, we introduce and analyze a dual problem for supermartingale optimal transport. We shall prove that this problem admits an optimizer and that there is no duality gap.

4.1. **Integration on a Proper Irreducible Component.** We first introduce the notion of integrability that will be used for the dual elements. Let $\mu \leq_{cd} \nu$ be proper and irreducible with domain (I, J), and let $\chi : J \to \mathbb{R}$ be a concave increasing function. Since χ^+ has linear growth, $\mu(\chi)$ and $\nu(\chi)$ are well defined in $[-\infty, \infty)$. In what follows, we give a meaningful definition of the difference $\mu(\chi) - \nu(\chi)$ in cases where both terms are infinite. We write χ' for the *left* derivative of χ, with the convention that $\chi'(\infty) := \lim_{t\to\infty} \chi'(t) = \inf_{t\in I} \chi'(t)$, and $-\chi''$ for the second derivative measure of the convex function $-\chi$ on I. Finally, recall that $I = (x^*, \infty)$. If ν has an atom at x^*, then χ may have a jump at x^* and we denote its magnitude by $\Delta\chi(x^*) := \chi(x^*+) - \chi(x^*) \in \mathbb{R}_+$.

Lemma 4.1. *Let $\mu \leq_{cd} \nu$ be proper and irreducible with domain (I, J), let $\chi : J \to \mathbb{R}$ be a concave increasing function, and let $P = \mu \otimes \kappa$ be an arbitrary element of $\mathcal{S}(\mu, \nu)$. Then*

$$\begin{aligned}(\mu - \nu)(\chi) &:= \int_I \left[\chi(x) - \int_J \chi(y)\kappa(x, dy)\right] \mu(dx) \\ &= \chi'(\infty)[\mathrm{bary}(\mu) - \mathrm{bary}(\nu)] + \int_I (p_\mu - p_\nu)\, d\chi'' + \Delta\chi(x^*)\nu(\{x^*\}).\end{aligned}$$

In particular, the definition of $(\mu - \nu)(\chi) \in [0, \infty]$ does not depend on P.

The proof follows the lines of [11, Lemma 4.1] and is omitted. Our next aim is to define expressions of the form $\mu(\varphi) + \nu(\psi)$ in a situation where the individual integrals are not necessarily finite. We continue to assume that $\mu \leq_{cd} \nu$ is proper and irreducible with domain (I, J).

Definition 4.2. Let $\varphi : I \to \overline{\mathbb{R}}$ and $\psi : J \to \overline{\mathbb{R}}$ be Borel functions. If there exists a concave increasing function $\chi : J \to \mathbb{R}$ such that $\varphi - \chi \in L^1(\mu)$ and $\psi + \chi \in L^1(\nu)$, we say that χ is a *moderator* for (φ, ψ) and set

$$\mu(\varphi) + \nu(\psi) := \mu(\varphi - \chi) + \nu(\psi + \chi) + (\mu - \nu)(\chi) \in (-\infty, \infty].$$

(This value is independent of the choice of χ by an argument similar to [11, Remark 4.8].) We denote by $L^{ci}(\mu, \nu)$ the space of all pairs (φ, ψ) which admit a moderator χ such that $(\mu - \nu)(\chi) < \infty$.

4.2. **Closedness on a Proper Irreducible Component.** In this section, we introduce the dual problem for a proper and irreducible pair $\mu \leq_{cd} \nu$ with domain (I, J). It will be convenient to work with a nonnegative reward function f and alleviate this restriction later on (Remark 4.12).

Definition 4.3. Let $f : I \times J \to [0, \infty]$. We denote by $\mathcal{D}^{ci,pw}_{\mu,\nu}(f)$ the set of all Borel functions $(\varphi, \psi, h) : \mathbb{R} \to \overline{\mathbb{R}} \times \overline{\mathbb{R}} \times \mathbb{R}_+$ with $(\varphi, \psi) \in L^{ci}(\mu, \nu)$ and

$$\varphi(x) + \psi(y) + h(x)(y - x) \geq f(x, y), \quad (x, y) \in I \times J.$$

We emphasize that the above inequality is stated in the pointwise ("pw") sense. The following is the key result of this section. We stress that its assertion fails if (μ,ν) is not proper; cf. Section 10.1 for a counterexample.

Proposition 4.4. *Given $f, f_n : I \times J \to [0,\infty]$ such that $f_n \to f$ pointwise and $(\varphi_n, \psi_n, h_n) \in \mathcal{D}^{ci,pw}_{\mu,\nu}(f_n)$ satisfying $\sup_n \mu(\varphi_n)+\nu(\psi_n) < \infty$, there exist*

$$(\varphi,\psi,h) \in \mathcal{D}^{ci,pw}_{\mu,\nu}(f) \quad \textit{such that} \quad \mu(\varphi)+\nu(\psi) \le \liminf_n \mu(\varphi_n)+\nu(\psi_n).$$

For the course of the proof, we abbreviate $\mathcal{D}^{ci}(f) := \mathcal{D}^{ci,pw}_{\mu,\nu}(f)$.

Lemma 4.5. *Let $(\varphi,\psi,h) \in \mathcal{D}^{ci}(0)$. There exists a moderator $\chi : J \to \mathbb{R}$ for (φ,ψ) such that $\chi \le \varphi$ on I and $-\chi \le \psi$ on J. In particular, we have $(\mu-\nu)(\chi) \le \mu(\varphi)+\nu(\psi)$.*

Proof. Let $P = \mu \otimes \kappa$ be a disintegration of some $P \in \mathcal{S}(\mu,\nu)$ and let $(\varphi,\psi,h) \in \mathcal{D}^{ci}(0)$. With a careful application of Fubini's theorem, one can verify that

$$\iint h(x)(y-x)\,\kappa(x,dy)\,\mu(dx) = \int h(x)(\mathrm{bary}(\kappa(x))-x)\,\mu(dx) > -\infty. \tag{4.1}$$

(This is quite different from the property that $h(X)(Y-X) \in L^1(P)$ which may fail.) A second application of Fubini's theorem then yields that

$$\begin{aligned} &P[\varphi(X)+\psi(Y)+h(X)(Y-X)] \\ &\qquad = \mu(\varphi)+\nu(\psi)+\iint h(x)(y-x)\,\kappa(x,dy)\,\mu(dx) \in \mathbb{R}. \end{aligned} \tag{4.2}$$

The concave and increasing function

$$\chi(y) := \inf_{x\in I}\,[\varphi(x)+h(x)(y-x)], \quad y \in J$$

satisfies $\chi \le \varphi$ on I and $-\chi \le \psi$ on J, and one can check that χ is finite-valued as a consequence of $(\varphi,\psi) \in L^{ci}(\mu,\nu)$. Set $\bar\varphi := \varphi - \chi \ge 0$ and $\bar\psi := \psi + \chi \ge 0$. By the first part of the proof, the iterated integral of $\varphi(x)+\psi(y)+h(x)(y-x)$ with respect to κ and μ is finite. The function

$$\bar\varphi(x)+\bar\psi(y)+[\chi(x)-\chi(y)]+h(x)(y-x) \tag{4.3}$$

is identical to the former; therefore, the iterated integral of (4.3) is again finite. For fixed $x \in I$, all four terms in (4.3) are bounded from below by linearly growing functions. It follows that for μ-a.e. $x \in I$, the integral with respect to $\kappa(x,dy)$ can be computed term-by-term, which yields

$$\bar\varphi(x)+\int \bar\psi(y)\,\kappa(x,dy)+\int[\chi(x)-\chi(y)]\,\kappa(x,dy)+h(x)(\mathrm{bary}(\kappa(x))-x).$$

The first three terms are nonnegative, and the last term is known to be μ-integrable by the first part of the proof. Thus, we may again integrate term-by-term with respect to μ. In conclusion, the iterated integral of (4.3), which

was already determined to be finite, may also be computed term-by-term. In particular, we deduce that $\mu(\bar{\varphi}) < \infty$, $\nu(\bar{\psi}) < \infty$ and $(\mu - \nu)(\chi) < \infty$, showing that $(\bar{\varphi}, \bar{\psi}) \in L^{ci}(\bar{\mu}, \bar{\nu})$ with concave moderator χ, and $\mu(\varphi)+\nu(\psi) = \mu(\bar{\varphi}) + \nu(\bar{\psi}) + (\mu - \nu)(\chi) \geq (\mu - \nu)(\chi)$ as desired. □

Our last tool for the proof of Proposition 4.4 is a compactness principle for concave increasing functions. We mention that the conclusion fails if the pair $\mu \leq_{cd} \nu$ is not proper (see also Section 10.1): a nontrivial difference between the barycenters is crucial to control the first derivatives.

Proposition 4.6. *Let $a = \mathrm{bary}(\mu)$ and let $\chi_n : J \to \mathbb{R}$ be concave increasing functions such that $\chi_n(a) = 0$ and $\sup_{n\geq 1}(\mu - \nu)(\chi_n) < \infty$. There exists a subsequence χ_{n_k} which converges pointwise on J to a concave increasing function $\chi : J \to \mathbb{R}$ such that $(\mu - \nu)(\chi) \leq \liminf_k (\mu - \nu)(\chi_{n_k})$.*

Proof. By our assumption, $(\mu - \nu)(\chi_n)$ is bounded uniformly in n. Since $\mathrm{bary}(\mu) > \mathrm{bary}(\nu)$, Lemma 4.1 yields a constant C such that $0 \leq \chi_n'(\infty) \leq C$ and $0 \leq \int_I (p_\mu - p_\nu)\, d\chi_n'' \leq C$, as well as $0 \leq \Delta\chi_n(x^*) \leq C$ in the case where $\nu(\{x^*\}) > 0$. For a suitable subsequence χ_{n_k}, we have

$$\lim_k \chi_{n_k}'(\infty) = \liminf_n \chi_n'(\infty), \tag{4.4}$$

$$\text{and similarly} \quad \lim_k \Delta\chi_{n_k}(x^*) = \liminf_n \Delta\chi_n(x^*) \quad \text{if } \nu(\{x^*\}) > 0. \tag{4.5}$$

Without loss of generality we assume that $n_k = k$. Given $y_0 \in I$, we recall from the proof of Lemma 3.11 that $p_\mu - p_\nu$ is strictly negative and uniformly bounded away from zero on $[y_0, \infty) \subseteq (x^*, \infty) = I$, and deduce that $0 \leq -\chi_n''[y_0, \infty) \leq C'$ for a constant C'. Since the (left) derivative χ_n' is decreasing, it follows that $\chi_n'(y) = -\chi_n''[y, \infty) + \chi_n'(\infty) \leq C' + C$ for all $y \in [y_0, \infty)$. Thus, the Lipschitz constant of χ_n is bounded on compact subsets of I, uniformly in n. Recalling that $\chi_n(a) = 0$, the Arzela–Ascoli theorem then yields a function $\chi : I \to \mathbb{R}$ such that $\chi_n \to \chi$ locally uniformly, after passing to a subsequence. Clearly χ is concave and increasing, and integration by parts shows that $-\chi_n''$ converges to the second derivative measure $-\chi''$ associated with χ, in the sense of weak convergence relative to the compactly supported continuous functions on I. Approximating $p_\mu - p_\nu$ from above with compactly supported continuous functions g_n, we then see that $\int_I (p_\mu - p_\nu)\, d\chi'' = \lim_m \lim_n \int_I g_m\, d\chi_n'' \leq \liminf_{n\to\infty} \int_I (p_\mu - p_\nu)\, d\chi_n''$. Using also (4.4), (4.5) and the representation in Lemma 4.1, we obtain $(\mu - \nu)(\chi) \leq \liminf_{n\to\infty}(\mu - \nu)(\chi_n)$ as desired. □

Proof of Proposition 4.4. We may assume that $\liminf_n \mu(\varphi_n) + \nu(\psi_n) = \limsup_n \mu(\varphi_n) + \nu(\psi_n)$, by passing to a subsequence. Since $(\varphi_n, \psi_n, h_n) \in \mathcal{D}^{ci}(f_n)$ and $f_n \geq 0$, we can introduce the associated moderators χ_n as in Lemma 4.5. We may assume that $\chi_n(a) = 0$, where $a := \mathrm{bary}(\mu) \in I$, by

translating φ_n and ψ_n appropriately. After passing to a subsequence, Proposition 4.6 then yields a pointwise limit $\chi : J \to \mathbb{R}$. Now, $(\varphi, \psi, h) \in \mathcal{D}^{ci}(f)$ can be constructed using Komlos' lemma and concave-increasing envelopes, following the ideas in the proof of [11, Proposition 5.2]. □

4.3. **Duality on a Proper Irreducible Component.** Recall that the pair $\mu \leq_{cd} \nu$ is proper and irreducible. Next, we define the primal and dual values.

Definition 4.7. Let $f : \mathbb{R}^2 \to [0, \infty]$ and write $P(f)$ for the outer integral. The *primal* and *dual* problems are respectively given by

$$\mathbf{S}_{\mu,\nu}(f) := \sup_{P \in \mathcal{S}(\mu,\nu)} P(f), \quad \mathbf{I}^{pw}_{\mu,\nu}(f) := \inf_{(\varphi,\psi,h) \in \mathcal{D}^{ci,pw}_{\mu,\nu}(f)} \mu(\varphi) + \nu(\psi).$$

A function $f : \mathbb{R}^2 \to [0, \infty]$ is *upper semianalytic* if the sets $\{f \geq c\}$ are analytic for all $c \in \mathbb{R}$, where a subset of $\mathbb{R}^2$ is called analytic if it is the (forward) image of a Borel subset of a Polish space under a Borel mapping. Any Borel function is upper semianalytic; we refer to [13] for further background.

Proposition 4.8. *Let $\mu \leq_{cd} \nu$ be proper and irreducible, $f : \mathbb{R}^2 \to [0, \infty]$.*

(i) If f is upper semianalytic, then $\mathbf{S}_{\mu,\nu}(f) = \mathbf{I}^{pw}_{\mu,\nu}(f) \in [0, \infty]$.

(ii) If $\mathbf{I}^{pw}_{\mu,\nu}(f) < \infty$, there exists a dual optimizer $(\varphi, \psi, h) \in \mathcal{D}^{ci,pw}_{\mu,\nu}(f)$.

Proof. The inequality "$\leq$" in (i) follows from (4.1) and (4.2). The converse inequality as well as (ii) follow from Proposition 4.4, using the Hahn–Banach and Choquet theorems along the lines of [11, Theorem 6.2]. □

4.4. **Global Duality.** In this section, we formulate a global duality result. We shall be brief since it is little more than the combination of the preceding results for the proper irreducible case and the known martingale case; however, it requires some notation.

Let $\mu \leq_{cd} \nu$ be probability measures and let $f : \mathbb{R}^2 \to [0, \infty]$ be Borel. As in the irreducible case, the primal problem is $\mathbf{S}_{\mu,\nu}(f) := \sup_{P \in \mathcal{S}(\mu,\nu)} P(f)$. For the dual problem, we first recall from Proposition 3.4 the decompositions $\mu = \sum_{k \geq -1} \mu_k$ and $\nu = \sum_{k \geq -1} \nu_k$, where $\mu_k \leq_{cd} \nu_k$ is irreducible with domain (I_k, J_k) for $k \geq 0$ and $\mu_{-1} = \nu_{-1}$; moreover, P_{-1} is the unique element of $\mathcal{S}(\mu_{-1}, \mu_{-1})$. So far, we have focused on a proper pair (μ_0, ν_0) and its dual problem. The pairs (μ_k, ν_k) for $k \geq 1$ are in convex order (μ_k and ν_k have the same barycenter) and the corresponding martingale optimal transport has an analogous duality theory. While the arguments are different, the preceding results hold true if "convex-increasing" is replaced by "convex" and the function h is allowed to take values in $\mathbb{R}$ instead of $\mathbb{R}_+$; we refer to [11] for the proofs. The spaces corresponding to $L^{ci}(\mu, \nu)$ and $\mathcal{D}^{ci}_{\mu,\nu}(f)$ are denoted $L^c(\mu, \nu)$ and $\mathcal{D}^c_{\mu,\nu}(f)$, respectively.

Let $(\varphi, \psi, h) : \mathbb{R} \to \overline{\mathbb{R}} \times \overline{\mathbb{R}} \times \mathbb{R}$ be Borel. Since P_{-1} is concentrated on the diagonal Δ, the dual problem associated to (μ_{-1}, ν_{-1}) is trivially

solved, for instance, by setting $\varphi(x) = f(x,x)$ and $\psi = h = 0$. To simplify the notation below, we set $L^c_{\mu_{-1},\nu_{-1}} := \{(\varphi,\psi) : \varphi + \psi \in L^1(\mu_{-1})\}$ and $\mu_{-1}(\varphi)+\nu_{-1}(\psi) := \mu_{-1}(\varphi+\psi)$ for $(\varphi,\psi) \in L^c_{\mu_{-1},\nu_{-1}}$. Moreover, $\mathcal{D}^{c,pw}_{\mu_{-1},\nu_{-1}}(f)$ is the set of all (φ,ψ,h) with $(\varphi,\psi) \in L^c_{\mu_{-1},\nu_{-1}}$ and $\varphi(x) + \psi(x) \geq f(x,x)$ for all $x \in I_{-1}$. Finally, we set $\mathbf{S}_{\mu_{-1},\nu_{-1}}(f) := P_{-1}(f) \equiv \mu_{-1}(f(X,X))$.

We can now introduce the domain for the global dual problem which will be stated in the quasi-sure sense. A property is said to hold $\mathcal{S}(\mu,\nu)$-quasi surely, or $\mathcal{S}(\mu,\nu)$-q.s. for short, if it holds P-a.s. for all $P \in \mathcal{S}(\mu,\nu)$.

Definition 4.9. Let $L(\mu,\nu)$ be the set of all Borel functions $\varphi,\psi : \mathbb{R} \to \overline{\mathbb{R}}$ such that $(\varphi,\psi) \in L^{ci}(\mu_0,\nu_0)$ and $(\varphi,\psi) \in L^c(\mu_k,\nu_k)$ for all $k \neq 0$ and $\sum_{k\geq -1} |\mu_k(\varphi) + \nu_k(\psi)| < \infty$. For $(\varphi,\psi) \in L(\mu,\nu)$, we define

$$\mu(\varphi) + \nu(\psi) := \sum_{k\geq -1} \mu_k(\varphi) + \nu_k(\psi) < \infty,$$

and $\mathcal{D}_{\mu,\nu}(f)$ is the set of all Borel functions $(\varphi,\psi,h) : \mathbb{R} \to \overline{\mathbb{R}} \times \overline{\mathbb{R}} \times \mathbb{R}$ such that $(\varphi,\psi) \in L(\mu,\nu)$, $h = 0$ on I_{-1}, $h \geq 0$ on I_0 and

$$\varphi(X) + \psi(Y) + h(X)(Y - X) \geq f(X,Y) \quad \mathcal{S}(\mu,\nu)\text{-q.s.}$$

Finally, $\mathbf{I}_{\mu,\nu}(f) := \inf_{(\varphi,\psi,h)\in\mathcal{D}_{\mu,\nu}(f)} \mu(\varphi) + \nu(\psi) \ \in [0,\infty]$.

We emphasize that h is required to be nonnegative on I_0 but can take arbitrary real values outside of I_0 and I_{-1}. It is shown in Section 10.1 that nonnegativity cannot be enforced everywhere.

Before making precise the correspondence between this quasi-sure formulation and the individual components, recall that the intervals J_k may overlap at their endpoints, so we have to avoid counting certain things twice. Indeed, let $(\varphi_k,\psi_k,h_k) \in \mathcal{D}^{ci,pw}_{\mu_k,\nu_k}(f)$ for $k = 0$ and $(\varphi_k,\psi_k,h_k) \in \mathcal{D}^{c,pw}_{\mu_k,\nu_k}(f)$ for $k \geq 1$; we claim that ψ_k can be normalized such that

$$\psi_k = 0 \quad \text{on} \quad J_k \setminus I_k. \tag{4.6}$$

Indeed, if J_k contains one of its endpoints, it is an atom of ν and hence ψ_k is finite on $J_k \setminus I_k$. If $k \geq 1$, we can translate ψ_k by an affine function and translate φ_k and h_k accordingly. In the supermartingale case $k = 0$, we recall from Proposition 3.4 that $I_0 = (x^*,\infty)$, so that J_0 can have at most one endpoint. As a result, we may obtain the normalization by translating ψ_0 with a constant, which can be compensated by translating φ_0 alone and thus respecting the requirement that $h_0 \geq 0$ on I_0.

Lemma 4.10. *Let $f : \mathbb{R}^2 \to [0,\infty]$ be Borel, let $\mu \leq_{cd} \nu$ and let μ_k,ν_k be as in Proposition 3.4.*

(i) Let $(\varphi_0,\psi_0,h_0) \in \mathcal{D}^{ci,pw}_{\mu_0,\nu_0}(f)$ and $(\varphi_k,\psi_k,h_k) \in \mathcal{D}^{c,pw}_{\mu_k,\nu_k}(f)$ for $k \geq 1$ be normalized as in (4.6), and let $\varphi_{-1}(x) = f(x,x)$ and $\psi_{-1} = 0$. If

$\sum_{k\geq -1} \mu(\varphi_k) + \nu(\psi_k) < \infty$, *then*

$$\varphi := \sum_{k\geq -1} \varphi_k \mathbf{1}_{I_k}, \quad \psi := \sum_{k\geq 0} \psi_k \mathbf{1}_{J_k}, \quad h := \sum_{k\geq 0} h_k \mathbf{1}_{I_k}$$

satisfies $(\varphi, \psi, h) \in \mathcal{D}_{\mu,\nu}(f)$ *and* $\mu(\varphi) + \nu(\psi) = \sum_{k\geq -1} \mu_k(\varphi_k) + \nu_k(\psi_k)$.

(ii) Conversely, let $(\varphi, \psi, h) \in \mathcal{D}_{\mu,\nu}(f)$. *After changing* φ *on a* μ*-nullset and* ψ *on a* ν*-nullset, we have* $(\varphi, \psi, h) \in \mathcal{D}^{ci,pw}_{\mu_0,\nu_0}(f)$ *and* $(\varphi, \psi, h) \in \mathcal{D}^{c,pw}_{\mu_k,\nu_k}(f)$ *for* $k \neq 0$, *and* $\sum_{k\geq -1} \mu_k(\varphi) + \nu_k(\psi) = \mu(\varphi) + \nu(\psi) < \infty$.

This follows from Proposition 3.4; the details of the proof are analogous to [11, Lemma 7.2]. We can now state the global duality result.

Theorem 4.11. *Let* $f : \mathbb{R}^2 \to [0, \infty]$ *be Borel and let* $\mu \leq_{cd} \nu$. *Then*

$$\mathbf{S}_{\mu,\nu}(f) = \mathbf{I}_{\mu,\nu}(f) \in [0, \infty].$$

If $\mathbf{I}_{\mu,\nu}(f) < \infty$, *there exists an optimizer* $(\varphi, \psi, h) \in \mathcal{D}_{\mu,\nu}(f)$ *for* $\mathbf{I}_{\mu,\nu}(f)$.

This is a consequence of Proposition 4.8 and the corresponding result in the martingale case; the arguments are as in [11, Theorem 7.4].

Remark 4.12. The lower bound on f in Theorem 4.11 can easily be relaxed. Indeed, let $f : \mathbb{R}^2 \to \overline{\mathbb{R}}$ be Borel and suppose that there exist $a \in L^1(\mu)$, $b \in L^1(\nu)$ such that $f(x, y) \geq a(x) + b(y)$ for all $x, y \in \mathbb{R}$. Then, we may apply Theorem 4.11 to $\bar{f} := [f(X, Y) - a(X) - b(Y)]^+$ and deduce the duality result for f as well.

5. Monotonicity Principle

An important consequence of the duality theorem is a monotonicity principle describing the support of optimal transports; it can be seen as a substitute for the cyclical monotonicity from classical transport theory. The following notion will be useful for our study of the canonical couplings.

Definition 5.1. Let π be a finite measure on $\mathbb{R}^2$ with finite first moment and let $M_0, M_1 \subseteq \mathbb{R}$ be Borel. Denote by π_1 its first marginal and by $\pi = \pi_1 \otimes \kappa$ a disintegration. A measure π' is an (M_0, M_1)*-competitor* of π if it has the same marginals and if its disintegration $\pi' = \pi_1 \otimes \kappa'$ satisfies

$$\begin{aligned} \mathrm{bary}(\kappa'(x)) &\leq \mathrm{bary}(\kappa(x)) \quad \text{for } \pi_1\text{-a.e. } x \in M_0, \\ \mathrm{bary}(\kappa'(x)) &= \mathrm{bary}(\kappa(x)) \quad \text{for } \pi_1\text{-a.e. } x \in M_1. \end{aligned}$$

This definition extends a concept of [9] where the barycenters are required to be equal on the whole real line. In our context, we need to distinguish three regimes for the applications in the subsequent sections: equality, inequality, and no constraint on the barycenters. One consequence of the inequality is that the notion of competitors is no longer symmetric.

Given $\mu \leq_{cd} \nu$, we recall from Proposition 3.4 the sets I_k, J_k, where the labels $k \geq 1$ correspond to the martingale components, $k = 0$ is the supermartingale component, and $k = -1$ is the complement (where any transport from μ to ν is the identity). Moreover, any element of $\mathcal{S}(\mu, \nu)$ is necessarily supported by the set

$$\Sigma := \Delta \cup \bigcup_{k \geq 0} I_k \times J_k. \tag{5.1}$$

Theorem 5.2 (Monotonicity Principle)**.** *Let $f : \mathbb{R}^2 \to [0, \infty]$ be Borel, let $\mu \leq_{cd} \nu$ be probability measures and suppose that $\mathbf{S}_{\mu,\nu}(f) < \infty$. There exist Borel sets $\Gamma \subseteq \mathbb{R}^2$ and $M \subseteq \mathbb{R}$ with the following properties.*

(i) A measure $P \in \mathcal{S}(\mu, \nu)$ is optimal for $\mathbf{S}_{\mu,\nu}(f)$ if and only if it is concentrated on Γ and $P|_{M \times \mathbb{R}}$ is a martingale.

(ii) Let $\bar{\mu} \leq_{cd} \bar{\nu}$ be probabilities on $\mathbb{R}$. If $\bar{P} \in \mathcal{S}(\bar{\mu}, \bar{\nu})$ is concentrated on Γ and $\bar{P}|_{M \times \mathbb{R}}$ is a martingale, then $\bar{P}$ is optimal for $\mathbf{S}_{\bar{\mu},\bar{\nu}}(f)$.

(iii) Let $M_0 = M \cap I_0$ and $M_1 = M \setminus M_0$, and let π be a finitely supported probability on $\mathbb{R}^2$ which is concentrated on Γ. Then $\pi(f) \geq \pi'(f)$ for any (M_0, M_1)-competitor π' of π that is concentrated on Σ.

If $(\varphi, \psi, h) \in \mathcal{D}_{\mu,\nu}(f)$ is a suitable[3] version of the optimizer from Theorem 4.11, then we can take

$$M := (I_0 \cap \{h > 0\}) \cup (\cup_{k \neq 0} I_k),$$
$$\Gamma := \big\{(x, y) \in \mathbb{R}^2 : \varphi(x) + \psi(y) + h(x)(y - x) = f(x, y)\big\} \cap \Sigma.$$

Moreover, the assertion in (iii) remains true if π is not finitely supported, as long as $(\varphi, \psi) \in L(\pi_1, \pi_2)$, where π_1 and π_2 are the marginals of π.

Before proving the theorem, let us draw a corollary stating that the supermartingale optimal transport can be decomposed as follows. On M, an optimizer $P \in \mathcal{S}(\mu, \nu)$ is also an optimizer of a martingale optimal transport problem. Thus, we think of M as the set where the supermartingale constraint is "binding," and in fact it acts like the seemingly stronger martingale constraint (thus M as in *martingale*). Whereas on $N := \mathbb{R} \setminus M$, the measure P is also an optimizer of a (Monge–Kantorovich) optimal transport problem with no constraint at all on the dynamics (N as in *no* constraint).

Corollary 5.3 (Extremal Decomposition)**.** *Let $f : \mathbb{R}^2 \to [0, \infty]$ be Borel and let $\mu \leq_{cd} \nu$ be probability measures such that $\mathbf{S}_{\mu,\nu}(f) < \infty$. There exists a Borel set $M \subseteq \mathbb{R}$ with the following property.*

Given an optimizer $P \in \mathcal{S}(\mu, \nu)$ for $\mathbf{S}_{\mu,\nu}(f)$, let $\mu_M = \mu|_M$ and let ν_M be the image[4] of μ_M under P. Moreover, let $\mu_N = \mu|_{\mathbb{R} \setminus M}$ and let ν_N be the image of μ_N under P. Then for the same reward function f,

[3] chosen as in Lemma 4.10 (ii)

[4] If $P = \mu \otimes \kappa$, the image of μ_M under P is defined as the second marginal of $\mu_M \otimes \kappa$.

(i) $P|_{M\times\mathbb{R}}$ *is an optimal* martingale *transport from* μ_M *to* ν_M*,*
(ii) $P|_{N\times\mathbb{R}}$ *is an optimal* Monge–Kantorovich *transport from* μ_N *to* ν_N*.*

A word of caution is in order: while the set M is defined without reference to P, the second marginals ν_M, ν_N in the extremal problems do depend on P. In that sense, the decomposition is non-unique—which, however, is quite natural given that the optimizer P is non-unique as well, for general f.

Remark 5.4. The lower bound on f in Theorem 5.2 and Corollary 5.3 can be relaxed as follows. Instead of f being nonnegative, suppose that there exist real functions $a \in L^1(\mu)$, $b \in L^1(\nu)$ such that $f(x,y) \geq a(x) + b(y)$ for all $x, y \in \mathbb{R}$. Then, Theorem 5.2 (i), (iii) as well as Corollary 5.3 hold as above, using Remark 4.12 but otherwise the same proofs. Moreover, Theorem 5.2 (ii) as well as the last statement in Theorem 5.2 hold under the condition that a, b are integrable for $\bar{\mu}, \bar{\nu}$ and π_1, π_2, respectively.

Example 5.5. In the context of Corollary 5.3, suppose that μ has no atoms and that f is smooth, of linear growth, and satisfies the Spence–Mirrlees condition $f_{xy} > 0$ and the martingale Spence–Mirrlees condition $f_{xyy} > 0$ (this is *not* one of the canonical cases studied later). Then an optimizer P exists and the corollary implies that $P|_{M\times\mathbb{R}}$ is the Left-Curtain coupling [9] between its marginals and $P|_{N\times\mathbb{R}}$ is the Hoeffding–Fréchet coupling [55, Section 3.1] between its marginals. In particular, writing $P = \mu \otimes \kappa$ and using the results of the indicated references, we immediately deduce the possible forms of the kernel: at almost every x, $\kappa(x)$ is either deterministic (the Hoeffding–Fréchet kernel) or a martingale kernel concentrated at two points (the Left–Curtain kernel). In particular, $\kappa(x)$ is never what might seem to be the typical case—a truly random process with downward drift.

We mention that the coupling P is nevertheless not canonical in the sense of the Introduction: even if uniqueness holds, the optimal coupling may change substantially if we replace f by a different function satisfying the same Spence–Mirrlees conditions; cf. Section 10.2 for a counterexample.

Proof of Corollary 5.3. Let M and (φ, ψ, h) be as in Theorem 5.2, and note that $P(f) < \infty$.

(i) We have $P_M := P|_{M\times\mathbb{R}} \in \mathcal{M}(\mu_M, \nu_M)$ by (i) of the theorem. Moreover, setting $P_N := P|_{N\times\mathbb{R}}$, any $\bar{P}_M \in \mathcal{M}(\mu_M, \nu_M)$ induces an element of $\mathcal{S}(\mu, \nu)$ via $\bar{P} := \bar{P}_M + P_N$. Thus, $\bar{P}_M(f) \leq P_M(f)$ by the optimality of P.

(ii) This part is less direct because elements of $\Pi(\mu_N, \nu_N)$ are not supermartingales in general; we shall invoke Theorem 5.2(iii) with $\pi := P$. By (i) of the theorem, π is concentrated on Γ, and of course $(\varphi, \psi) \in L(\mu, \nu) = L(\pi_1, \pi_2)$. Moreover, for any $\pi'_N \in \Pi(\mu_N, \nu_N)$, the measure $\pi' = P_M + \pi'_N$ is an (M_0, M_1)-competitor of $\pi = P_M + P_N$ which is concentrated on Σ as P_N is concentrated on $I_0 \times J_0 \subseteq \Sigma$; note that $N \subseteq I_0$. Now the extension of (iii) at the end of the theorem yields $\pi(f) \geq \pi'(f)$ and hence $P_N(f) \geq \pi'_N(f)$. □

Proof of Theorem 5.2. As $\mathbf{I}_{\mu,\nu}(f) = \mathbf{S}_{\mu,\nu}(f) < \infty$, Theorem 4.11 yields a dual optimizer $(\varphi, \psi, h) \in \mathcal{D}_{\mu,\nu}(f)$ and we can define Γ and M as stated.

(i) Let $P \in \mathcal{S}(\mu, \nu)$ and let $P = \mu \otimes \kappa$ be a disintegration. Recalling (4.1) and (4.2) and the analogous facts for the martingale case [11], we have

$$P(f) \leq P[\varphi(X) + \psi(Y) + h(X)(Y - X)] \leq \mu(\varphi) + \nu(\psi).$$

Since $\mathbf{S}_{\mu,\nu}(f) = \mu(\varphi) + \nu(\psi)$, P is optimal if and only if both inequalities are equalities. As $P(f) < \infty$, the first inequality is an equality if and only if P is concentrated on Γ. Moreover, the second inequality is an equality if and only if $\int (y - x)\, \kappa(x, dy) = 0$ μ-a.e. on $\{h > 0\}$; note that the condition on κ holds automatically on the martingale components I_k, $k \geq 1$. In particular, this is equivalent to $P|_{M \times \mathbb{R}}$ being a martingale.

(ii) We choose a version of $(\varphi, \psi, h) \in \mathcal{D}_{\mu,\nu}(f)$ as in Lemma 4.10 (ii); moreover, we may assume that $\bar{P}(f) < \infty$. We need to show that $(\varphi, \psi, h) \in \mathcal{D}_{\bar{\mu},\bar{\nu}}(f)$; once this is established, optimality can be argued as in (i) above.

(a) On the one hand, we need to show that

$$\varphi(X) + \psi(Y) + h(X)(Y - X) \geq f(X, Y) \quad \mathcal{S}(\bar{\mu}, \bar{\nu})\text{-q.s.} \tag{5.2}$$

For this, it suffices to prove that the domains of the irreducible components of $\bar{\mu} \leq_{cd} \bar{\nu}$ are subsets of the ones of $\mu \leq_{cd} \nu$; i.e., that $p_\mu(x) = p_\nu(x)$ implies $p_{\bar{\mu}}(x) = p_{\bar{\nu}}(x)$, for any $x \in \mathbb{R}$. Indeed, let $p_\mu(x) = p_\nu(x)$. Since $\bar{P}$ is concentrated on $\Gamma \subseteq \Sigma$, we know that $Y \leq x$ $\bar{P}$-a.s. on $\{X \leq x\}$ and $Y \geq x$ $\bar{P}$-a.s. on $\{X \geq x\}$. Writing $E[\,\cdot\,]$ for the expectation under $\bar{P}$, it follows that $p_{\bar{\nu}}(x) = E[(x - Y)^+] = E[(x - Y)\mathbf{1}_{X \leq x}]$. Note that $p_\mu(x) = p_\nu(x)$ implies $x \leq x^*$, cf. Proposition 3.2. Recalling that $(-\infty, x^*) \subseteq M$, our assumption on $\bar{P}$ then yields that $\bar{P}|_{\{X < x\}}$ is a martingale. Thus, $E[(x - Y)\mathbf{1}_{X \leq x}] = E[(x - X)\mathbf{1}_{X \leq x}] = E[(x - X)^+] = p_{\bar{\mu}}(x)$ and part (a) is complete.

(b) On the other hand, we need to show that $(\varphi, \psi) \in L(\bar{\mu}, \bar{\nu})$. By reducing to the components, we may assume without loss of generality that $(\bar{\mu}, \bar{\nu})$ is irreducible with domain (I, J). Moreover, the argument for the martingale case is contained in the proof of [11, Corollary 7.8], so we shall assume that $(\bar{\mu}, \bar{\nu})$ is proper. Let $\chi(y) := \inf_{x \in I} [\varphi(x) + h(x)(y - x)]$. As $(\varphi, \psi, h) \in \mathcal{D}^{ci,pw}_{\mu,\nu}(f)$, the arguments below (4.2) yield that $\chi : J \to \mathbb{R}$ is concave and increasing, that $\bar{\varphi} := \varphi - \chi \geq 0$ and $\bar{\psi} := \psi + \chi \geq 0$, and that the expectation $\bar{P}[\varphi(X) + \psi(Y) + h(X)(Y - X)]$ can be computed as the $\bar{\mu}(dx)$-integral of

$$\bar{\varphi}(x) + \int \bar{\psi}(y)\, \kappa(x, dy) + \left[\chi(x) - \int \chi(y) \kappa(x, dy)\right] + h(x)(\mathrm{bary}(\kappa(x)) - x),$$

where $\bar{P} = \bar{\mu} \otimes \kappa$ for some kernel κ (not necessarily the same as in (i)). By the assumption that $\bar{P}|_{M \times \mathbb{R}}$ is a martingale and $\mathbb{R} \setminus M = \{h \leq 0\} \cap I_0 \subseteq \{h = 0\}$, either $h(x) = 0$ or $\mathrm{bary}(\kappa(x)) = x$, for $\bar{\mu}$-a.e. $x \in \mathbb{R}$. Using also that $\bar{P}$ is concentrated on Γ, we deduce that $\bar{P}(f) = \bar{P}[\varphi(X) + \psi(Y) + h(X)(Y - X)] = \bar{\mu}(\bar{\varphi}) + \bar{\nu}(\bar{\psi}) + (\bar{\mu} - \bar{\nu})(\chi)$, where the last step is justified by the

nonnegativity of the integrands. As $\bar{P}(f) < \infty$, we conclude that the three (nonnegative) terms on the right-hand side are finite; that is, $(\varphi, \psi) \in L(\bar{\mu}, \bar{\nu})$ with moderator χ.

(iii) Again, we may assume that $\pi(f) < \infty$. Let π' be an (M_0, M_1)-competitor of π, let $\bar{\mu}, \bar{\nu}$ be the common first and second marginals of π, π' and let $\pi = \bar{\mu} \otimes \kappa$, $\pi' = \bar{\mu} \otimes \kappa'$. If $(\varphi, \psi) \in L(\bar{\mu}, \bar{\nu})$, using $h \geq 0$ on $M_0 \subseteq I_0$ and $\mathbb{R} \setminus M \subseteq \{h = 0\}$ and the definition of the competitor yields $\pi(f) = \bar{\mu}(\varphi) + \bar{\nu}(\psi) + \int_M h(x)(\mathrm{bary}(\kappa(x)) - x)\, \bar{\mu}(dx) \geq \bar{\mu}(\varphi) + \bar{\nu}(\psi) + \int_M h(x)(\mathrm{bary}(\kappa'(x)) - x)\, \bar{\mu}(dx) \geq \pi'(f)$. Of course, $(\varphi, \psi) \in L(\bar{\mu}, \bar{\nu})$ holds in particular if π is finitely supported. □

6. Shadow Construction

In this section, we introduce the Increasing and Decreasing Supermartingale Transports via an order-theoretic construction. Let $\mathfrak{M}^1(\mathbb{R})$ be the set of all finite measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ which have a finite first moment, endowed with the weak convergence induced by the continuous functions of linear growth. We shall mainly use the restriction of this topology to subsets of measures of equal mass, and then it is equivalent to the Kantorovich or 1-Wasserstein distance $W(\nu, \nu') = \sup_f (\nu - \nu')(f)$, where f ranges over all 1-Lipschitz functions.

Definition 6.1. Let $\mu, \nu \in \mathfrak{M}^1(\mathbb{R})$. We say that μ, ν are in *positive-convex-decreasing* order, denoted $\mu \leq_{pcd} \nu$, if $\mu(\phi) \leq \nu(\phi)$ for all nonnegative, convex, decreasing functions $\phi : \mathbb{R} \to \mathbb{R}$.

We note that $\mu \leq_{pcd} \nu$ necessarily satisfy $\mu(\mathbb{R}) \leq \nu(\mathbb{R})$. In fact, the case of strict inequality is the one of interest: if $\mu(\mathbb{R}) = \nu(\mathbb{R})$, then $\mu \leq_{pcd} \nu$ is equivalent to $\mu \leq_{cd} \nu$.

Lemma 6.2. *Let $\mu, \nu \in \mathfrak{M}^1(\mathbb{R})$ satisfy $\mu \leq_{pcd} \nu$. Then the set*[5]

$$[\![\mu, \nu]\!] := \{\theta \in \mathfrak{M}^1(\mathbb{R}) : \mu \leq_{cd} \theta \leq \nu\}$$

is nonempty and contains a unique least element $S^\nu(\mu)$ for the convex-decreasing order: $S^\nu(\mu) \leq_{cd} \theta$ for all $\theta \in [\![\mu, \nu]\!]$. The measure $S^\nu(\mu)$ is called the shadow *of μ in ν.*

Proof. Without loss of generality, ν is a probability measure.

(i) We first show that $[\![\mu, \nu]\!]$ contains some element θ. Let λ be the Lebesgue measure on $\mathbb{R}$ and let G_ν be the quantile function of ν; that is, the left-continuous inverse of the c.d.f. of ν. We define

$$\theta := \lambda|_{[0,k]} \circ G_\nu^{-1} \quad \text{where} \quad k := \mu(\mathbb{R}) \in [0, 1].$$

[5]We think of the elements of $[\![\mu, \nu]\!]$ as lying between μ and ν, as the notation suggests. However, we caution the reader that $\mu, \nu \notin [\![\mu, \nu]\!]$ in general.

This implies that $\theta \in \mathfrak{M}^1(\mathbb{R})$, that $\theta(\mathbb{R}) = k$, and that $\theta \leq \nu$. Intuitively speaking, θ is the "left-most" measure $\theta \leq \nu$ of mass k on $\mathbb{R}$; in particular, if ν admits a density f_ν, the density of θ is $f_\theta = f_\nu \mathbf{1}_{(-\infty, G_\nu(k)]}$.

Let ϕ be a convex, decreasing function; we need to show that $\mu(\phi) \leq \theta(\phi)$. To this end, we may assume that $\phi(G_\nu(k)) = 0$ by translating ϕ, and then $\mu(\phi) \leq \mu(\phi^+) \leq \nu(\phi^+) = \theta(\phi^+) = \theta(\phi)$ since $\phi^+ = \phi$ on $G_\nu([0,k])$. As a result, $\theta \in [\![\mu, \nu]\!] \neq \emptyset$.

(ii) Next, we show that $[\![\mu, \nu]\!]$ is directed; i.e., given $\theta_i \in [\![\mu, \nu]\!]$, $i = 1, 2$ there exists $\theta \in [\![\mu, \nu]\!]$ such that $\theta \leq_{cd} \theta_i$. Indeed, let $p : \mathbb{R} \to \mathbb{R}$ be defined as the convex hull of the minimum of p_{θ_1} and p_{θ_2}. Then p is convex, and p is increasing like p_{θ_i}. Since the asymptotic slope of the functions p_{θ_i} is given by $\theta_i(\mathbb{R}) = \mu(\mathbb{R})$, the same is true for p, and finally, $p_{\theta_i} \geq p_\mu$ yields $p \geq p_\mu$. These facts imply that p is the put function associated with a measure θ satisfying $\mu \leq_{cd} \theta \leq_{cd} \theta_i$. It remains to show that $\theta \leq \nu$, which is equivalent to $p_\nu - p$ being convex. Indeed, the fact that $p_\nu - p_{\theta_i}$ is convex for $i = 1, 2$ implies this property; cf. the proof of [9, Lemma 4.6] for a detailed argument.

(iii) The set $[\![\mu, \nu]\!] \subseteq \mathfrak{M}^1(\mathbb{R})$ consists of measures with common total mass $\mu(\mathbb{R})$; we show that it is compact. Indeed, closedness is readily established. Moreover, any $\theta \in [\![\mu, \nu]\!]$ satisfies $\theta \leq \nu$. By Prokhorov's theorem, this immediately yields tightness in the weak topology induced by bounded continuous functions, and then using $\int |x| \, \nu(dx) < \infty$ yields relative compactness.

(iv) It follows from (iii) that for any convex, decreasing function ϕ of linear growth, the continuous functional $\theta \mapsto \theta(\phi)$ has a nonempty compact set $\Theta_\phi \subseteq [\![\mu, \nu]\!]$ of minimizers. The directedness of $[\![\mu, \nu]\!]$ from (ii) implies that a finite intersection $\Theta_{\phi_1} \cap \cdots \cap \Theta_{\phi_n}$ is still nonempty, and then compactness shows that $\theta \mapsto \theta(\phi)$ has a common minimizer $S^\nu(\mu)$ for all ϕ. Uniqueness of the minimizer holds since $\theta_1 \leq_{cd} \theta_2$ and $\theta_2 \leq_{cd} \theta_1$ imply $\theta_1 = \theta_2$. □

Lemma 6.3. *Let $\mu, \nu \in \mathfrak{M}^1(\mathbb{R})$ satisfy $\mu \leq_{pcd} \nu$ and suppose that μ is concentrated at a single point $x \in \mathbb{R}$. Then, the shadow $S^\nu(\mu)$ is of the form*

$$S^\nu(\mu) = \nu|_{(a,b)} + k_a \delta_a + k_b \delta_b.$$

Among all measures $\theta \leq \nu$ with mass $\mu(\mathbb{R})$ of this form, $S^\nu(\mu)$ is determined by maximizing $\mathrm{bary}(\theta)$ *subject to the constraint* $\mathrm{bary}(\theta) \leq x$. *Moreover, $S^\nu(\mu)$ has minimal variance among all measures $\theta \leq \nu$ with mass $\mu(\mathbb{R})$ and* $\mathrm{bary}(\theta) = \mathrm{bary}(S^\nu(\mu))$. *Finally, a and b can be chosen such that $a \leq x \leq b$.*

The map $\nu \mapsto S^\nu(\mu)$ is continuous when restricted to a set of measures $\nu \in \mathfrak{M}^1(\mathbb{R})$ of equal total mass satisfying $\mu \leq_{pcd} \nu$.

Proof. We may assume that $\nu(\mathbb{R})$=1. Then, $\mu = k\delta_x$ for some $k \in [0, 1]$, and we may focus on $k \in (0, 1)$. Consider the family

$$\theta_s = \lambda|_{[s,s+k]} \circ G_\nu^{-1}, \quad s \in [0, 1-k].$$

Similarly as in the proof of Lemma 6.2, we have $\theta_s \leq \nu$ for all s, whereas $\mu = k\delta_x \leq_{cd} \theta_s$ if and only if $\mathrm{bary}(\theta_s) \leq x$. As $\mu \leq_{pcd} \nu$, this inequality holds true in particular for $s = 0$. The function

$$s \mapsto \mathrm{bary}(\theta_s) = \frac{1}{k}\int_0^k G_\nu(s+t)\,\lambda(dt) = \frac{1}{k}\int_0^k G_\nu(s+t+)\,\lambda(dt) \tag{6.1}$$

is increasing and continuous: $G_\nu(s)$ and its right limit $G_\nu(s+)$ differ only on a Lebesgue nullset, the fist representation shows left-continuity and the second shows right-continuity. Thus, we may define s^* as the largest value in $[0, 1-k]$ for which $\mathrm{bary}(\theta_s) \leq x$, and then $\theta^* := \theta_{s^*}$ is in $[\![\mu, \nu]\!]$. We claim that θ^* is the least element in $[\![\mu, \nu]\!]$.

To show this, let $(a, b) = (G_\nu(s^*), G_\nu(s^* + k))$; then $\theta^*|_{(a,b)} = \nu|_{(a,b)}$ and θ^* is concentrated on $[a, b]$. Now let $\theta \in [\![\mu, \nu]\!]$ be arbitrary. As $\theta \leq \nu$, we see that $\theta - (\theta^* \wedge \theta)$ is concentrated on $(a, b)^c$, whereas $\theta^* - (\theta^* \wedge \theta)$ is concentrated on $[a, b]$. Moreover, we must have $\mathrm{bary}(\theta) \leq \mathrm{bary}(\theta^*)$. Indeed, this is clear if $\mathrm{bary}(\theta^*) = x$. If not, the definition of s^* implies that $\nu(b, \infty) = 0$ and then θ^* clearly has the largest barycenter among all measures $\theta \leq \nu$ with mass $\mu(\mathbb{R})$. Thus, Lemma 6.4 below implies that $\theta^* \leq_{cd} \theta$ and as a result, θ^* is the least element in $[\![\mu, \nu]\!]$; i.e., $S^\nu(\mu) = \theta^*$.

As $\mathrm{bary}(\theta^*) \leq x$, it is clear that $a \leq x$. With the above choice of b, it may happen that $b < x$. However, by the definition of s^*, this is possible only if $\theta^*(\{b\}) = \nu(\{b\})$ and $\nu(b, \infty) = 0$. In that case, we may redefine $b := x$ without invalidating the other assertions of the lemma, and then we have $a \leq x \leq b$ as required.

Finally, the continuity of $\nu \mapsto S^\nu(\mu)$ can be shown by using (6.1) and the property $W(\nu, \nu') = \int_0^1 |G_\nu(t) - G_{\nu'}(t)|\,\lambda(dt)$ of the 1-Wasserstein distance; we omit the details. □

Lemma 6.4. *Let $\mu, \nu \in \mathfrak{M}^1(\mathbb{R})$ satisfy $\mu(\mathbb{R}) = \nu(\mathbb{R})$ and $\mathrm{bary}(\mu) \geq \mathrm{bary}(\nu)$. If there exists an interval $I = (a, b)$ such that μ is concentrated on $\bar{I} := [a, b] \cap \mathbb{R}$ and ν is concentrated on I^c, then $\mu \leq_{cd} \nu$. The same is true if there exists an interval I such that $\mu - (\mu \wedge \nu)$ is concentrated on $\bar{I}$ and $\nu - (\mu \wedge \nu)$ is concentrated on I^c.*

Proof. The first claim implies the second, so we may focus on the former. We need to show that $\mu(\phi) \leq \nu(\phi)$ for any convex decreasing function ϕ. To this end, we may assume that the left endpoint a of the interval is finite and strictly smaller than the right endpoint b, as otherwise we must have $\mu = \nu = 0$; moreover, we may reduce to the case $\phi(a) = 0$. If b is finite as well, we define $\psi(x) := \phi(x) - \frac{\phi(b)}{b-a}(x-a)$ for $x \in \mathbb{R}$, whereas $\psi := \phi$ if $b = \infty$. Then $\psi \leq 0$ on $\bar{I}$ and $\psi \geq 0$ on I^c, which yields $\mu(\phi) \leq \mu(\psi^+) + \frac{\phi(b)}{b-a}[\mathrm{bary}(\mu) - a] \leq \nu(\psi^+) + \frac{\phi(b)}{b-a}[\mathrm{bary}(\nu) - a] = \nu(\phi)$ as desired. □

The following result is important to apply the shadow in an iterative fashion. The first assertion intuitively follows from the minimality of the shadow: if we transport part of a measure $\mu \leq_{pcd} \nu$ to its shadow in ν, the remaining part μ_2 of μ is still dominated by the remaining part of ν. Moreover, if we then transport μ_2 to its shadow in the remainder, the cumulative result is the same as the shadow of μ in ν.

Proposition 6.5. *Let $\mu_1, \mu_2, \nu \in \mathfrak{M}^1(\mathbb{R})$ satisfy $\mu_1 + \mu_2 \leq_{pcd} \nu$. Then $\mu_2 \leq_{pcd} \nu - S^\nu(\mu_1)$ and $S^\nu(\mu_1 + \mu_2) = S^\nu(\mu_1) + S^{\nu - S^\nu(\mu_1)}(\mu_2)$.*

Proof. Using the result of Lemma 6.3, one can first establish the claim when μ_1 is a single atom. Then, one can extend to the general case along the lines of [9, Theorem 4.8]; we omit further details. □

Next, we shall use the shadow mapping to construct specific supermartingale transports. Let $\mu \leq_{cd} \nu$ and suppose first that $\mu = \sum_{i=1}^n k_i \delta_{x_i}$ is finitely supported. We may transport μ to ν by first mapping $k_1\delta_{x_1}$ to its shadow in ν, continue by mapping $k_2\delta_{x_2}$ to its shadow in the "remainder" $\nu - S^\nu(k_1\delta_{x_1})$ of ν, and so on. Proceeding until $i = n$, this constructs the kernel κ corresponding to a supermartingale transport $\mu \otimes \kappa \in \mathcal{S}(\mu, \nu)$. In fact, this recipe leads to a whole family of transports—the labeling of the atoms was arbitrary, and a different order in their processing will typically give rise to a different transport. There are two choices that seem canonical: left-to-right (increasing) and right-to-left (decreasing). We shall show in the subsequent sections that the corresponding transports $\vec{P}$ and $\overleftarrow{P}$ are indeed canonical in several ways.

Theorem 6.6. *Let $\mu \leq_{cd} \nu$.*

(i) There exists a unique measure $\vec{P}$ on $\mathbb{R} \times \mathbb{R}$ which transports $\mu|_{(-\infty,x]}$ to its shadow $S^\nu(\mu|_{(-\infty,x]})$ for all $x \in \mathbb{R}$; that is, the first marginal of $\vec{P}$ equals μ and $\vec{P}((-\infty, x] \times A) = S^\nu(\mu|_{(-\infty,x]})(A)$ for $A \in \mathcal{B}(\mathbb{R})$.

(ii) Similarly, there exists a unique measure $\overleftarrow{P}$ on $\mathbb{R} \times \mathbb{R}$ which transports $\mu|_{[x,\infty)}$ to its shadow $S^\nu(\mu|_{[x,\infty)})$ for all $x \in \mathbb{R}$.

Moreover, those two measures are elements of $\mathcal{S}(\mu, \nu)$. We call $\vec{P}$ and $\overleftarrow{P}$ the Increasing and the Decreasing Supermartingale Transport, respectively.

Proof. The function $F(x, y) := S^\nu(\mu|_{(-\infty,x]})(-\infty, y]$ is clearly increasing and right-continuous in y. Moreover, Proposition 6.5 implies that

$$S^\nu(\mu|_{(-\infty,x_2]}) - S^\nu(\mu|_{(-\infty,x_1]}) = S^{\nu - S^\nu(\mu|_{(-\infty,x_1]})}(\mu|_{(x_1,x_2]}) \geq 0, \quad x_1 \leq x_2$$

which yields the same properties for the variable x; note that the total mass of the right-hand side equals $\mu(x_1, x_2]$. Noting also that F has the proper normalization for a c.d.f., we conclude that F induces a unique measure $\vec{P}$

on $\mathcal{B}(\mathbb{R}\times\mathbb{R})$. It is clear that μ is the first marginal of $\vec{P}$. The second marginal is $S^\nu(\mu)\leq\nu$, and this is in fact an equality because both measures have the same mass. To conclude that $\vec{P}\in\mathcal{S}(\mu,\nu)$, it suffices to show that $\vec{P}[Y\phi(X)]\leq\vec{P}[X\phi(X)]$ for all $\phi=\mathbf{1}_{(x_1,x_2]}$ with $x_1<x_2$. Indeed, Proposition 6.5 implies that $\vec{P}[Y\phi(X)]=\int y\,[S^\nu(\mu|_{(-\infty,x_2]})-S^\nu(\mu|_{(-\infty,x_1]})](dy)=\mathrm{bary}(S^{\nu-S^\nu(\mu|_{(-\infty,x_1]})}(\mu|_{(x_1,x_2]}))\leq\mathrm{bary}(\mu|_{(x_1,x_2]})=\vec{P}[X\phi(X)]$. The arguments for (ii) are analogous. □

A different construction of $\vec{P}$ and $\overleftarrow{P}$ could proceed through an approximation of the marginals by discrete measures, for which the couplings can be defined explicitly by iterating Lemma 6.3, and a subsequent passage to the limit. We refer to [45, Remark 2.18] for a sketch of such a construction in the martingale case.

7. Spence–Mirrlees Functions and Geometry of their Optimal Transports

In this section, we relate monotonicity properties of the reward function f to the geometry of the supports of the corresponding optimal supermartingale transports, where the support will be described by a pair (Γ,M) as in Theorem 5.2. We first introduce the relevant properties of f.

Definition 7.1. A function $f:\mathbb{R}^2\to\mathbb{R}$ is *first-order Spence–Mirrlees* if

$$f(x_2,\cdot)-f(x_1,\cdot)\quad\text{is strictly increasing for all } x_1<x_2.$$

Moreover, f is *second-order Spence–Mirrlees* if

$$f(x_2,\cdot)-f(x_1,\cdot)\quad\text{is strictly convex for all } x_1<x_2,$$

and f is *supermartingale Spence–Mirrlees* if f is second-order Spence–Mirrlees and $-f$ is first-order Spence–Mirrlees.

We note that if f is smooth, the first and second order Spence–Mirrlees properties are equivalent to the classical cross-derivative conditions $f_{xy}>0$ and $f_{xyy}>0$, respectively. The latter is also called martingale Spence–Mirrlees condition in the literature on martingale optimal transport—the above terminology will be more convenient in what follows.

Remark 7.2. There exist smooth, linearly growing supermartingale Spence–Mirrlees functions on $\mathbb{R}^2$. Indeed, let φ be a smooth, bounded, strictly increasing function on $\mathbb{R}$; e.g., $\varphi(x)=\tanh(x)$. Let ψ be a smooth, linearly growing, strictly decreasing, strictly convex function on $\mathbb{R}$; e.g., $\psi(y)=(1+y^2)^{1/2}-y$. Then, $g(x,y):=\varphi(x)\psi(y)$ satisfies $g_{xy}<0$ and $g_{xyy}>0$, while $|g(x,y)|\leq C(1+|y|)$ for some $C>0$.

Next, we introduce the relevant geometric properties of the support.

Definition 7.3. Let $(\Gamma, M) \subseteq \mathbb{R}^2 \times \mathbb{R}$ and consider $(x_1, y_1), (x_2, y_2) \in \Gamma$ with $x_1 < x_2$. The pair (Γ, M) is

(i) *first-order left-monotone* if $y_1 \leq y_2$ whenever $x_2 \notin M$,
(ii) *first-order right-monotone* if $y_2 \leq y_1$ whenever $x_1 \notin M$.

We will also need the following properties of Γ; they are taken from [9] where they are simply called left- and right-monotonicity.

Definition 7.4. Let $\Gamma \subseteq \mathbb{R}^2$ and consider $(x, y_1), (x, y_2), (x', y') \in \Gamma$ with $y_1 < y_2$. Then Γ is

(i) *second-order left-monotone* if $y' \notin (y_1, y_2)$ whenever $x < x'$,
(ii) *second-order right-monotone* if $y' \notin (y_1, y_2)$ whenever $x' < x$.

For convenience, we shall use the same terminology for a pair (Γ, M) even though only Γ is relevant for the second-order properties. Yet another notion will be useful; we write $\Gamma^1 = \{x \in \mathbb{R} : (x, y) \in \Gamma \text{ for some } y \in \mathbb{R}\}$ for the projection of Γ onto the first coordinate.

Definition 7.5. A pair $(\Gamma, M) \subseteq \mathbb{R}^2 \times \mathbb{R}$ is *nondegenerate* if

(i) for all $x \in \Gamma^1$ such that $(x, y) \in \Gamma$ for some $y > x$, there exists $y' < x$ such that $(x, y') \in \Gamma$,
(ii) for all $x \in \Gamma^1 \cap M$ such that $(x, y) \in \Gamma$ for some $y < x$, there exists $y' > x$ such that $(x, y') \in \Gamma$.

These two conditions imply that

(i') for all $x \in \Gamma^1$ there exists $y \leq x$ such that $(x, y) \in \Gamma$,
(ii') for all $x \in \Gamma^1 \cap M$ there exists $y \geq x$ such that $(x, y) \in \Gamma$.

Essentially, nondegeneracy postulates that there is a down-path at every $x \in \Gamma^1$, and also an up-path if $x \in M$. Thus, it is a natural requirement if we intend to consider supermartingales supported by Γ which are martingales on $M \times \mathbb{R}$. For later use, let us record that nondegeneracy can be assumed without loss of generality in our context.

Remark 7.6. Let $(\Gamma, M) \in \mathcal{B}(\mathbb{R}^2) \times \mathcal{B}(\mathbb{R})$, let $\mu \leq_{cd} \nu$ be probability measures and suppose there is $P \in \mathcal{S}(\mu, \nu)$ with $P(\Gamma) = 1$ such that $P|_{M \times \mathbb{R}}$ is a martingale. Then, there exists a Borel subset $\Gamma' \subseteq \Gamma$ with $P(\Gamma') = 1$ such that (Γ', M) is nondegenerate.

Proof. Let N_1' be the set of all $x \in \Gamma^1$ such that Definition 7.5 (i) fails. Then N_1' is universally measurable and thus we can find a Borel set $N_1 \supseteq N_1'$ such that $N_1 \setminus N_1'$ is μ-null. The fact that P is a supermartingale implies that $\Gamma_1 := \Gamma \cap \{Y > X\} \cap (N_1 \times \mathbb{R})$ is P-null. After defining similarly a set N_2 for Definition 7.5 (ii), the martingale property of P on $M \times \mathbb{R}$ shows that $\Gamma_2 := \Gamma \cap \{Y < X\} \cap (N_2 \times \mathbb{R})$ is P-null as well, and then we can set $\Gamma' := \Gamma \setminus (\Gamma_1 \cup \Gamma_2)$. $\square$

The first-order properties turn out to be highly asymmetric when combined with nondegeneracy. The following observation will have far-reaching consequences regarding the geometry of the coupling $\vec{P}$ and has no analogue in the left-monotone case.

Remark 7.7. Let (Γ, M) be first-order right-monotone and nondegenerate. Then, M is a half-line unbounded to the left within Γ^1; that is,

$$\text{if } x_1, x_2 \in \Gamma^1 \text{ satisfy } x_1 < x_2 \text{ and } x_2 \in M, \text{ then } x_1 \in M.$$

Indeed, let x_1, x_2 be as stated; then nondegeneracy yields y_1, y_2 such that $y_1 \leq x_1 < x_2 \leq y_2$ and $(x_i, y_i) \in \Gamma$. If we had $x_1 \notin M$, this would contradict first-order right-monotonicity.

With these definitions in place, we can use the monotonicity principle of Theorem 5.2 to infer the geometry of (Γ, M) from the properties of f.

Proposition 7.8. *Let*[6] $\mu \leq_{cd} \nu$ *and recall the corresponding intervals* I_k, J_k *of Proposition 3.4 and the set* Σ *of* (5.1). *Let* $(\Gamma, M) \in \mathcal{B}(\mathbb{R}^2) \times \mathcal{B}(\mathbb{R})$ *be nondegenerate, where* $\Gamma \subseteq \Sigma$ *and* $M = M_0 \cup M_1$ *with Borel sets* $M_0 \subseteq I_0$ *and* $M_1 = \cup_{k \neq 0} I_k$, *and let* $f : \mathbb{R}^2 \to \mathbb{R}$. *Suppose that the assertion of Theorem 5.2 (iii) holds; that is, if* π *is a finitely supported probability which is concentrated on* Γ, *then* $\pi(f) \geq \pi'(f)$ *for any* (M_0, M_1)*-competitor* π' *of* π *that is concentrated on* Σ.

(i) If f *is first-order Spence–Mirrlees,* (Γ, M) *is first-order left-monotone.*
(ii) If $-f$ *is first-order Spence–Mirrlees,* (Γ, M) *is first-order right-monotone.*
(iii) If f *is second-order Spence–Mirrlees,* Γ *is second-order left-monotone.*
(iv) If $-f$ *is second-order Spence–Mirrlees,* Γ *is second-order right-monotone.*

Proof. (i) Consider $(x_1, y_1), (x_2, y_2) \in \Gamma$ with $x_1 < x_2$ and suppose for contradiction that $y_2 < y_1$. The measures

$$\pi := \tfrac{1}{2}\delta_{(x_1,y_1)} + \tfrac{1}{2}\delta_{(x_2,y_2)}, \quad \pi' := \tfrac{1}{2}\delta_{(x_1,y_2)} + \tfrac{1}{2}\delta_{(x_2,y_1)}$$

have the same first marginal $\pi_1 = \frac{1}{2}\delta_{x_1} + \frac{1}{2}\delta_{x_2}$. Let $\pi = \pi_1 \otimes \kappa$ and $\pi' = \pi_1 \otimes \kappa'$, then $\text{bary}(\kappa'(x_1)) < \text{bary}(\kappa(x_1))$ and $\text{bary}(\kappa'(x_2)) > \text{bary}(\kappa(x_2))$. Suppose that $x_1 \notin M_1$ and $x_2 \notin M$. Then, π' is an (M_0, M_1)-competitor of π. Moreover, $x_i \notin M_1$ implies that $x_i \in I_0$ and thus $y_i \in J_0$, $i = 1, 2$ which shows that π' is supported on Σ. Thus, we must have $\pi(f) \geq \pi'(f)$. However, $2(\pi(f) - \pi'(f)) = (f(x_2, y_2) - f(x_1, y_2)) - (f(x_2, y_1) - f(x_1, y_1)) < 0$ as f is first-order Spence–Mirrlees, so we have reached the desired contradiction.

Let $x_1 \in M_1$ and $x_2 \notin M$. Recalling that $M_1 = \cup_{k \neq 0} I_k = (-\infty, x^*]$, we have $y_1 \in J_k$ for some $k \neq 0$, whereas $x_2 \notin M$ implies $y_2 \in J_0$. Since J_0 is located to the right of J_k for $k \neq 0$, we must have $y_1 \leq y_2$.

(ii) Consider $(x_1, y_1), (x_2, y_2) \in \Gamma$ with $x_1 < x_2$ and suppose for contradiction that $y_1 < y_2$. We define π, π' as in (i); then $\text{bary}(\kappa'(x_1)) > \text{bary}(\kappa(x_1))$

[6] In fact, this result merely uses the general shape of Σ, not the specific marginals.

and $\mathrm{bary}(\kappa'(x_2)) < \mathrm{bary}(\kappa(x_2))$. Let $x_1 \notin M$. Then, $x_1 \in I_0 = (x^*, \infty)$ and thus $x_2 > x_1$ is in I_0 as well. In particular, $x_2 \notin M_1$ and $y_1, y_2 \in J_0$. Thus π' is an (M_0, M_1)-competitor of π that is concentrated on Σ and we reach a contradiction to $-f$ being first-order Spence–Mirrlees, similarly as in (i).

(iii) Let $(x, y_1), (x, y_2), (x', y') \in \Gamma$ satisfy $x < x'$ and assume for contradiction that $y_1 < y' < y_2$. Define $\lambda = \frac{y'-y_1}{y_2-y_1}$ and

$$\pi = \tfrac{\lambda}{2}\delta_{(x,y_1)} + \tfrac{1-\lambda}{2}\delta_{(x,y_2)} + \tfrac{1}{2}\delta_{(x',y')}, \quad \pi' = \tfrac{\lambda}{2}\delta_{(x',y_1)} + \tfrac{1-\lambda}{2}\delta_{(x',y_2)} + \tfrac{1}{2}\delta_{(x,y')}.$$

Then, π and π' have the same first marginal π_1 and if $\pi = \pi_1 \otimes \kappa$ and $\pi' = \pi_1 \otimes \kappa'$, then $\kappa(x)$, $\kappa'(x)$, $\kappa(x')$, $\kappa'(x')$ all have barycenter y'. Hence, π' is an (M_0, M_1)-competitor of π, and since the shape of $\Gamma \subseteq \Sigma$ shows that π' is concentrated on Σ, we deduce that $\pi(f) \geq \pi'(f)$. However, f being second-order Spence–Mirrlees implies that $\pi(f) < \pi'(f)$.

(iv) The proof is symmetric to (iii). □

8. Geometric Characterization of the Canonical Supermartingale Transports

In this section, we consider fixed probability measures $\mu \leq_{cd} \nu$ and show that the associated Increasing and Decreasing Supermartingale Transports $\vec{P}, \overleftarrow{P}$ (cf. Theorem 6.6) are characterized by geometric properties of their supports.

Theorem 8.1. *Let $(\Gamma, M) \in \mathcal{B}(\mathbb{R}^2) \times \mathcal{B}(\mathbb{R})$ be nondegenerate and let $P \in \mathcal{S}(\mu, \nu)$ be such that P is concentrated on Γ and $P|_{M\times\mathbb{R}}$ is a martingale.*

(i) If (Γ, M) is first-order right-monotone and second-order left-monotone, then P is the Increasing Supermartingale Transport $\vec{P}$.

(ii) If (Γ, M) is first-order left-monotone and second-order right-monotone, then P is the Decreasing Supermartingale Transport $\overleftarrow{P}$.

Before stating the proof, we record two auxiliary lemmas. The first one follows directly from the fact that $\mathcal{S}(\mu, \nu) \neq \emptyset$ by Proposition 2.1.

Lemma 8.2. *Let $a \in \mathbb{R}$ and $\mu \leq_{cd} \nu$. If ν is concentrated on $[a, \infty)$, then so is μ, and moreover $\nu(\{a\}) \geq \mu(\{a\})$. If $\mu \leq_c \nu$, the same holds for $(-\infty, a]$.*

Lemma 8.3 ([9, Lemma 5.4])**.** *Let σ be a nontrivial signed measure on $\mathbb{R}$ with $\sigma(\mathbb{R}) = 0$ and let $\sigma = \sigma^+ - \sigma^-$ be its Hahn decomposition. There exist $a \in \mathrm{supp}(\sigma^+)$ and $b > a$ such that $\int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)\, d\sigma(y) > 0$.*

Proof of Theorem 8.1 (i). Given $x \in \mathbb{R}$, we set $\mu_x := \mu|_{(-\infty,x]}$ and denote by ν_x^P the second marginal of $P|_{(-\infty,x]\times\mathbb{R}}$; that is, the image of μ_x under the transport P. Since P is concentrated on Γ and has the same mass as $\vec{P}$, it suffices to show that

$$\nu_x^P = \nu_x^{\vec{P}} \tag{8.1}$$

for all $x \in \Gamma^1$. In a first step we will show that (8.1) holds for all $x \in \Gamma^1 \cap M$. In view of Remark 7.7 it then follows that

$$P|_{M\times\mathbb{R}} = \vec{P}|_{M\times\mathbb{R}}. \tag{8.2}$$

After that we will show that (8.1) holds for all $x \in \Gamma^1$ if $M = \emptyset$, and then the latter assumption will be removed in a final step.

Let us first establish an auxiliary result that will be used in Steps 1 and 2. If (8.1) is violated for some $x \in \Gamma^1$, then the signed measure

$$\sigma := \nu_x^{\vec{P}} - \nu_x^P$$

is nontrivial and we can find $a \in \operatorname{supp}(\sigma^+)$ and $b > a$ as in Lemma 8.3. Note that $\sigma^+ \leq \nu - \nu_x^P$ and that $\nu - \nu_x^P$ is the image of $\mu|_{(x,\infty)}$ under P. Hence, $a \in \operatorname{supp}(\nu - \nu_x^P)$ and as $P(\Gamma) = 1$, there exists a sequence of points

$$(x_n, a_n) \in \Gamma \quad \text{with } x < x_n \text{ and } a_n \to a. \tag{8.3}$$

Step 1: Equality of the martingale parts. We argue by contradiction and assume that there exists $x \in \Gamma^1 \cap M$ violating (8.1). We first establish that

$$\nu_x^{\vec{P}} \leq_c \nu_x^P \quad \text{and in particular} \quad \operatorname{bary}(\nu_x^{\vec{P}}) = \operatorname{bary}(\nu_x^P). \tag{8.4}$$

Indeed, in view of $x \in M$, Remark 7.7 shows that $(-\infty, x] \cap \Gamma^1 \subseteq M$ and thus $P|_{(-\infty,x]\times\mathbb{R}}$ is a martingale. Therefore, $\operatorname{bary}(\nu_x^P) = \operatorname{bary}(\mu_x)$, and moreover $\operatorname{bary}(\mu_x) \geq \operatorname{bary}(\nu_x^{\vec{P}})$ since $\vec{P}$ is a supermartingale. Thus, $\operatorname{bary}(\nu_x^P) \geq \operatorname{bary}(\nu_x^{\vec{P}})$. On the other hand, $P \in \mathcal{S}(\mu, \nu)$ implies $\nu_x^P \in [\![\mu_x, \nu]\!]$ and hence $\nu_x^{\vec{P}} \leq_{cd} \nu_x^P$ by the minimality property defining $\vec{P}$; cf. Theorem 6.6. In view of Proposition 2.1, these two facts imply (8.4). Next, we show that

$$\Gamma_t \cap (a, \infty) = \emptyset, \quad t \leq a \wedge x, \quad \text{where} \quad \Gamma_t := \{y \in \mathbb{R} : (t, y) \in \Gamma\}. \tag{8.5}$$

Indeed, let $t \leq a \wedge x$ and suppose that $\Gamma_t \cap (a, \infty) \neq \emptyset$. Then in particular $\Gamma_t \cap (t, \infty) \neq \emptyset$ and thus nondegeneracy, more precisely Definition 7.5 (i), yields that $\Gamma_t \cap (-\infty, t) \neq \emptyset$ and hence $\Gamma_t \cap (-\infty, a) \neq \emptyset$. But now we obtain a contradiction to the second-order left-monotonicity of Γ by using (x_n, a_n) from (8.3) for (x', y') and t for x in Definition 7.4, for some large enough n.

Case (a): $x \in M$ and $x \leq a$. As $x \leq a$, (8.5) applies to all $t \leq x$ and hence $P(\Gamma) = 1$ implies that ν_x^P is concentrated on $(-\infty, a]$. In view of (8.4) and Lemma 8.2, it follows that $\nu_x^{\vec{P}}$ is concentrated on $(-\infty, a]$ as well, and $\nu_x^P(\{a\}) \geq \nu_x^{\vec{P}}(\{a\})$. These three facts imply $\int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)\, \nu_x^{\vec{P}}(dy) = (b-a)\nu_x^{\vec{P}}(\{a\})$ is dominated by $(b-a)\nu_x^P(\{a\}) = \int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)\, \nu_x^P(dy)$; that is, $\int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)\, \sigma(dy) \leq 0$. This contradicts the choice of a and b; cf. Lemma 8.3.

Case (b): $x \in M$ and $a < x$. Since $a < x$, we can argue exactly as below (8.4) to obtain that

$$\nu_a^{\vec{P}} \leq_c \nu_a^P \quad \text{and in particular} \quad \mathrm{bary}(\nu_a^{\vec{P}}) = \mathrm{bary}(\nu_a^P). \tag{8.6}$$

Moreover, (8.5) and $P(\Gamma) = 1$ now imply that ν_a^P is concentrated on $(-\infty, a]$, and then Lemma 8.2 shows that

$$\nu_a^P,\ \nu_a^{\vec{P}} \text{ are concentrated on } (-\infty, a] \quad \text{and} \quad \nu_a^{\vec{P}}(\{a\}) \leq \nu_a^P(\{a\}). \tag{8.7}$$

Next, we establish that $\nu_x^P - \nu_a^P$ is concentrated on $[a, \infty)$. Let $a < t \leq x$ be such that $\Gamma_t \neq \emptyset$. Since $x \in M$, Remark 7.7 yields that $t \in M$ and now nondegeneracy, cf. Definition 7.5 (ii'), shows that $\Gamma_t \cap [t, \infty) \neq \emptyset$. Then, using (8.3) and the second-order left-monotonicity of Γ yield that $\Gamma_t \cap (-\infty, a) = \emptyset$, and therefore, $\nu_x^P - \nu_a^P$ is indeed concentrated on $[a, \infty)$. We shall prove below that

$$\nu_x^{\vec{P}} - \nu_a^{\vec{P}} \leq_{cd} \nu_x^P - \nu_a^P \tag{8.8}$$

and thus Lemma 8.2 shows that $\nu_x^{\vec{P}} - \nu_a^{\vec{P}}$ is concentrated on $[a, \infty)$ as well. Using these facts, (8.7) and that $y \mapsto (b-y)^+ \mathbf{1}_{[a,\infty)}(y)$ is convex decreasing on $[a, \infty)$, yields that $\int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)\, \nu_x^{\vec{P}}(dy) = \int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)\, (\nu_x^{\vec{P}} - \nu_a^{\vec{P}})(dy) + (b-a)\nu_a^{\vec{P}}(\{a\})$ is dominated by $\int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)\, (\nu_x^P - \nu_a^P)(dy) + (b-a)\nu_a^P(\{a\}) = \int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)\, \nu_x^P(dy)$. This again contradicts the choice of a and b; cf. Lemma 8.3.

It remains to show (8.8). Indeed, using again that $\nu_x^P - \nu_a^P$ is concentrated on $[a, \infty)$ as well as (8.7), we have

$$\nu_x^P - \nu_a^P = (\nu_x^P - \nu_a^P)|_{[a,\infty)} \leq (\nu - \nu_a^P)|_{[a,\infty)} \leq (\nu - \nu_a^{\vec{P}})|_{[a,\infty)} \leq \nu - \nu_a^{\vec{P}}.$$

On the other hand, we have $\mu|_{(a,x]} \leq_{cd} \nu_x^P - \nu_a^P$ by the supermartingale property, and thus $\nu_x^P - \nu_a^P \in [\![\mu|_{(a,x]}, \nu - \nu_a^{\vec{P}}]\!]$. Since $\nu_x^{\vec{P}} - \nu_a^{\vec{P}} = S^{\nu - \nu_a^{\vec{P}}}(\mu|_{(a,x]})$ is the minimal element of the above set by the definition of $\vec{P}$ and the additivity of the shadow (Proposition 6.5), we conclude that (8.8) holds, and that completes the proof of Step 1.

Step 2: $M = \emptyset$. Again, suppose there exists $x \in \Gamma^1$ such that (8.1) is violated. Define $y_x := \inf \Gamma_x$. If $(x', y) \in \Gamma$ and $x' < x$, first-order right-monotonicity implies that $y \geq y_x$ (since $M = \emptyset$), and the latter holds trivially for $x' = x$. Conversely, if $(x', y) \in \Gamma$ and $x < x'$, first-order right-monotonicity implies that $y \leq y_x$. As a result, P is concentrated on the set $(-\infty, x] \times [y_x, \infty) \ \cup\ (x, \infty) \times (-\infty, y_x]$ and as $P \in \mathcal{S}(\mu, \nu)$, this implies that $\nu_x^P = \nu|_{(y_x,\infty)} + k\delta_{y_x}$ where $k := \mu((-\infty, x]) - \nu((y_x, \infty))$. This is the minimal element of $[\![\mu|_{(-\infty,x]}, \nu]\!]$ by Lemma 6.4, and thus $\nu_x^P = \nu_x^{\vec{P}}$.

Step 3: $M \neq \emptyset$. In the general case, let $\mu_M = \mu|_M$ and let ν_M^P denote the second marginal of $P|_{M\times\mathbb{R}}$. We note that $x \notin M$ yields $M \subseteq (-\infty, x]$ by Remark 7.7 and hence $\mu_M \leq \mu_x$.

We may apply the result proved in Step 2 to $\Gamma' = \Gamma \cap (M^c \times \mathbb{R})$, $M' = \emptyset$ and the marginals $\mu' = \mu - \mu_M$, $\nu' = \nu - \nu_M$ to deduce that $P|_{M^c\times\mathbb{R}}$ is the Increasing Supermartingale Transport from μ' to ν'. In particular,

$$S^{\nu-\nu_M^P}(\mu_x - \mu_M) = \nu_x^P - \nu_M^P. \tag{8.9}$$

Observing that (8.2) implies $\nu_M^P = \nu_M^{\vec{P}} = S^\nu(\mu_M)$, the additivity of the shadow (Proposition 6.5) shows that $\nu_x^{\vec{P}} = S^\nu(\mu_x) = \nu_M^P + S^{\nu-\nu_M^P}(\mu_x - \mu_M)$ which equals ν_x^P by (8.9). As $x \notin M$ was arbitrary, the proof is complete. □

Proof of Theorem 8.1 (ii). It will be convenient to reverse the notation with respect to the preceding proof: given $x \in \mathbb{R}$, we set $\mu_x := \mu|_{[x,\infty)}$ and let ν_x^P be the second marginal of $P|_{[x,\infty)\times\mathbb{R}}$. Again, we assume for contradiction that there exists $x \in \Gamma^1$ such that $\nu_x^P \neq \nu_x^{\overleftarrow{P}}$, so that the signed measure $\sigma := \nu_x^{\overleftarrow{P}} - \nu_x^P$ is nontrivial and we can find $a \in \operatorname{supp}(\sigma^+)$ and $a < b$ as in Lemma 8.3. Similarly as in (8.3), there exist

$$(x_n, a_n) \in \Gamma \quad \text{with } x_n < x \text{ and } a_n \to a. \tag{8.10}$$

Moreover, $P \in \mathcal{S}(\mu,\nu)$ implies that $\nu_x^P \in [\![\mu_x, \nu]\!]$ and hence, by minimality,

$$\nu_x^{\overleftarrow{P}} \leq_{cd} \nu_x^P. \tag{8.11}$$

Case 1a: $x \in M$ and $a \leq x$. We first show that

$$\nu_x^P \quad \text{is concentrated on } [a,\infty). \tag{8.12}$$

Indeed, let $t \in \Gamma^1$ be such that $t > x$. Suppose that $\Gamma_t \cap (-\infty, a) \neq \emptyset$, where $\Gamma_t := \{y \in \mathbb{R} : (t,y) \in \Gamma\}$. If $t \in M$, nondegeneracy yields that $\Gamma_t \cap [t,\infty) \neq \emptyset$ and since $a \leq x < t$, (8.10) contradicts the second-order right-monotonicity of Γ. Hence, $t \notin M$. Since $x \in M$, nondegeneracy also yields that $\Gamma_x \cap [x,\infty) \neq \emptyset$. But now $\Gamma_t \cap (-\infty, a) \neq \emptyset$ and $a \leq x$ contradict first-order left-monotonicity as $t \notin M$. As a result, $\Gamma_t \cap (-\infty, a) = \emptyset$. To extend this to $t = x$, note that in this case we have $t \in M$. Thus, if $\Gamma_t \cap (-\infty, a) \neq \emptyset$, the nondegeneracy of Definition 7.5 (ii) and (8.10) contradict second-order right-monotonicity. We have shown that $\Gamma_t \cap (-\infty, a) = \emptyset$ for all $t \geq x$, and (8.12) follows. In view of (8.11) and Lemma 8.2, we conclude that

$$\nu_x^{\overleftarrow{P}} \text{ is concentrated on } [a,\infty) \quad \text{and} \quad \nu_x^{\overleftarrow{P}}(\{a\}) \leq \nu_x^P(\{a\}). \tag{8.13}$$

Since $(b-y)^+$ is convex and decreasing, (8.11), (8.12) and (8.13) then yield $\int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)\, \nu_x^{\overleftarrow{P}}(dy) \leq \int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)\, \nu_x^P(dy)$ which contradicts the choice of a and b; cf. Lemma 8.3.

Case 1b: $x \in M$ and $x < a$. Let $t \geq a$ and suppose that $\Gamma_t \cap (-\infty, a) \neq \emptyset$. If $t \in M$, nondegeneracy yields that $\Gamma_t \cap (t, \infty) \neq \emptyset$ and since $x < a \leq t$, (8.10) contradicts the second-order right-monotonicity of Γ. Hence, $t \notin M$, but then $\Gamma_t \cap (-\infty, a) \neq \emptyset$ and (8.10) contradict first-order left-monotonicity. As a result, $\Gamma_t \cap (-\infty, a) = \emptyset$ for all $t \geq a$ and hence ν_a^P is concentrated on $[a, \infty)$. Since

$$\nu_a^{\overleftarrow{P}} \leq_{cd} \nu_a^P \tag{8.14}$$

can be argued as in (8.11), Lemma 8.2 then yields that

$$\nu_a^P,\ \nu_a^{\overleftarrow{P}} \text{ are concentrated on } [a, \infty) \quad \text{and} \quad \nu_a^{\overleftarrow{P}}(\{a\}) \leq \nu_a^P(\{a\}). \tag{8.15}$$

Next, we show that symmetrically,

$$\nu_x^P - \nu_a^P,\ \nu_x^{\overleftarrow{P}} - \nu_a^{\overleftarrow{P}} \text{ are concentrated on } (-\infty, a] \tag{8.16}$$

$$\text{and} \quad (\nu_x^{\overleftarrow{P}} - \nu_a^{\overleftarrow{P}})(\{a\}) \leq (\nu_x^P - \nu_x^P)(\{a\}). \tag{8.17}$$

Indeed, let $t \in \Gamma^1$ be such that $x \leq t < a$ and suppose that $\Gamma_t \cap (a, \infty) \neq \emptyset$. Since $\Gamma_t \cap (-\infty, t] \neq \emptyset$ by nondegeneracy, (8.10) contradicts second-order right-monotonicity. Thus, $\Gamma_t \cap (a, \infty) = \emptyset$ and $\nu_x^P - \nu_a^P$ is concentrated on $(-\infty, a]$. In order to conclude (8.16) and (8.17) via Lemma 8.2, it remains to show that $\nu_x^{\overleftarrow{P}} - \nu_a^{\overleftarrow{P}} \leq_c \nu_x^P - \nu_a^P$. Indeed, let again $t \in \Gamma^1$ be such that $x \leq t < a$. If $t \notin M$, then $\Gamma_t \cap (-\infty, t] \neq \emptyset$ and (8.10) contradict first-order left-monotonicity; thus $t \in M$. As a result, $P|_{[x,a) \times \mathbb{R}}$ is a martingale and $\mathrm{bary}(\nu_x^P - \nu_a^P) = \mathrm{bary}(\mu_x - \mu_a)$. Hence, we only have to show that

$$\nu_x^{\overleftarrow{P}} - \nu_a^{\overleftarrow{P}} \leq_{cd} \nu_x^P - \nu_a^P. \tag{8.18}$$

Using that $\nu_x^P - \nu_a^P$ is concentrated on $(-\infty, a]$ as well as (8.15), we have $\nu_x^P - \nu_a^P = (\nu_x^P - \nu_a^P)|_{(-\infty,a]} \leq (\nu - \nu_a^P)|_{(-\infty,a]} \leq (\nu - \nu_a^{\overleftarrow{P}})|_{[a,\infty)} \leq \nu - \nu_a^{\overleftarrow{P}}$. On the other hand, $\mu|_{[x,a)} \leq_{cd} \nu_x^P - \nu_a^P$ by the supermartingale property of P, and thus (8.18) follows from the minimality of $\overleftarrow{P}$ and Proposition 6.5. This completes the proof of (8.16) and (8.17).

Finally, we apply (8.14)–(8.17) to find that $\int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)\, \nu_x^{\overleftarrow{P}}(dy) = \int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)\, (\nu_x^{\overleftarrow{P}} - \nu_a^{\overleftarrow{P}})(dy) + \int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)\, \nu_a^{\overleftarrow{P}}(dy)$ is equal to $(b-a)(\nu_x^{\overleftarrow{P}} - \nu_a^{\overleftarrow{P}})(\{a\}) + \int (b-y)^+ \nu_a^{\overleftarrow{P}}(dy) \leq (b-a)(\nu_x^P - \nu_a^P)(\{a\}) + \int (b-y)^+ \nu_a^P(dy) = \int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)\, \nu_x^P(dy)$ which again contradicts the choice of a and b.

Case 2: $x \notin M$. Define again $y_x = \inf \Gamma_x$; note that $y_x \leq x$ by nondegeneracy. Let $t \in \Gamma^1$ be such that $t < x$. If $\Gamma_t \cap (y_x, \infty) \neq \emptyset$, then as $x \notin M$, the definition of y_x yields a contradiction to first-order left-monotonicity. On the other hand, let $x < t$ and assume that $\Gamma_t \cap (-\infty, y_x) \neq \emptyset$. If $t \notin M$,

the construction of y_x again contradicts first-order left-monotonicity; thus $t \in M$. But then nondegeneracy shows that $\Gamma_t \cap [t, \infty) \neq \emptyset$ and the definition of y_x yields a contradiction to second-order right-monotonicity. Clearly, $\Gamma_x \subseteq [y_x, \infty)$, and we have established that P must be concentrated on the union of $(-\infty, x) \times (-\infty, y_x]$ and $[x, \infty) \times [y_x, \infty)$. Since $P \in \mathcal{S}(\mu, \nu)$, this implies that $\nu_x^P = \nu|_{(y_x,\infty)} + k\delta_{y_x}$, where $k := \mu([x, \infty)) - \nu((y_x, \infty))$. This is the minimal element of $[\![\mu|_{[x,\infty)}, \nu]\!]$, and thus $\nu_x^P = \nu_x^{\overleftarrow{P}}$, completing the proof. We remark that the proof is shorter than in (i) due to the asymmetry of the supermartingale constraint. □

9. Regularity of Spence–Mirrlees Functions

A supermartingale Spence–Mirrlees function f need not be (semi)continuous. For instance, if $f(x, y) = \varphi(x)\psi(y)$ for a strictly increasing function φ and a strictly convex and decreasing function ψ, then f is supermartingale Spence–Mirrlees but clearly φ need not be upper or lower semicontinuous. In general, f may have a continuum of various types of discontinuities.

However, we show in Proposition 9.2 below that a measurable second-order Spence–Mirrlees function is automatically continuous for a finer topology on $\mathbb{R}^2$, and this topology will be coarse enough to preserve the weak compactness of $\mathcal{S}(\mu, \nu)$. Thus, we can still deduce the existence of optimal transports (Lemma 9.3) for upper semicontinuous reward functions f, and in particular for supermartingale Spence–Mirrlees functions. That will allow us to apply the monotonicity principle of Theorem 5.2.

Before stating these results, we introduce a relaxed version of the Spence–Mirrlees conditions of Definition 7.1, where increase and convexity are required in the non-strict sense—we have reserved the shorter name for the object that appears more frequently.

Definition 9.1. We call $f : \mathbb{R}^2 \to \mathbb{R}$ *relaxed first-order Spence–Mirrlees* if

$$f(x_2, \cdot) - f(x_1, \cdot) \quad \text{is increasing for all } x_1 < x_2,$$

relaxed second-order Spence–Mirrlees if

$$f(x_2, \cdot) - f(x_1, \cdot) \quad \text{is convex for all } x_1 < x_2,$$

and *relaxed supermartingale Spence–Mirrlees* if f is relaxed second-order Spence–Mirrlees and $-f$ is relaxed first-order Spence–Mirrlees.

Proposition 9.2. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be Borel and relaxed second-order Spence–Mirrlees. There is a Polish topology τ on $\mathbb{R}$ such that f is $\tau \otimes \tau$-continuous. Moreover, τ refines the Euclidean topology and induces the same Borel sets.*

Proof. We begin by constructing the functions f_n; the topology will be defined in the last step.

Step 1: Regularity in y. We first suppose that f vanishes along the y-axis,

$$f(0,y) = 0, \quad y \in \mathbb{R}. \tag{9.1}$$

Under this hypothesis, the second-order Spence–Mirrlees condition implies

$$\begin{cases} f(x,\cdot) \text{ is convex}, & x \geq 0, \\ f(x,\cdot) \text{ is concave}, & x \leq 0. \end{cases} \tag{9.2}$$

Therefore, $y \mapsto f(x,y)$ admits a finite left derivative $\partial_y f(x,0)$ at $y = 0$. We impose the further hypothesis that

$$\partial_y f(x,0) = 0, \quad x \in \mathbb{R}. \tag{9.3}$$

Since $y \mapsto f(x,y)$ is convex or concave, its restriction to a compact interval $K_m = [-m,m]$ is Lipschitz continuous with some optimal Lipschitz constant $\mathrm{Lip}(f(x,\cdot)|_{K_m}) < \infty$. More precisely, (9.2) and (9.3) imply that the optimal constant is the supremum of the absolute slopes of the tangents at the endpoints $y = \pm m$. The second-order Spence–Mirrlees condition implies that the absolute slopes are increasing in $|x|$; in particular,

$$\sup_{x \in K_m} \mathrm{Lip}(f(x,\cdot)|_{K_m}) = \sup_{x=\pm m} \mathrm{Lip}(f(x,\cdot)|_{K_m}) < \infty. \tag{9.4}$$

Step 2: Approximation. Fix $n \in \mathbb{N}$, let $y_k^n = 2^{-n}k$ for $k \in \mathbb{Z}$ and let $f_n(x,\cdot)$ be the continuous, piecewise affine approximation to $f(x,\cdot)$ along this grid; that is, for $y_k^n \leq y < y_{k+1}^n$ we define

$$f_n(x,y) = \lambda f(x,y_k^n) + (1-\lambda) f(x,y_{k+1}^n), \quad \lambda := 2^n(y_{k+1}^n - y). \tag{9.5}$$

We then have $|f_n(x,y) - f(x,y)| \leq 2^{-n}L$ for all $y \in K_m$ if L is a Lipschitz constant for $f(x,\cdot)$ on K_m. In view of (9.4), this shows that

$$f_n \to f \quad \text{uniformly on } K_m \times K_m, \quad \text{for all } m \in \mathbb{N}. \tag{9.6}$$

Step 3: Refining the Topology. Next, we introduce the topology τ. The basic idea here is that if φ is a real function with a single discontinuity at $y_0 \in \mathbb{R}$, we can change the topology on $\mathbb{R}$ by declaring y_0 an isolated point and then φ becomes continuous. More generally, [47, Theorem 13.11, Lemma 13.3] show that given a countable family of Borel functions on $\mathbb{R}$, there exists a Polish topology $\tau \subseteq \mathcal{B}(\mathbb{R})$ which renders these functions continuous and refines the Euclidean topology. In particular, we can find τ such that $f(\cdot, y_k^n)$ is τ-continuous for all n,k. As τ refines the Euclidean topology, it readily follows that the functions f_n defined in (9.5) are $\tau \otimes \tau$-continuous. But now (9.6) yields that f is continuous as well (note that since τ refines the Euclidean topology, any $\tau \otimes \tau$-neighborhood contains a bounded one).

It remains to remove the hypotheses (9.1) and (9.3). The above shows that the claim holds for $\tilde{f}(x,y) := f(x,y) - f(0,y) - (\partial_y[f(x,y) - f(0,y)]_{y=0})y$; note that $\tilde{f}$ is again relaxed second-order Spence–Mirrlees. We can further

refine τ such that the two Borel functions subtracted on the right-hand side are τ-continuous, and then the result for f follows. □

The preceding result leads to the existence of optimal transports.

Lemma 9.3. *Let $\mu \leq_{cd} \nu$ and let τ be a Polish topology on $\mathbb{R}$ which refines the Euclidean topology and induces the same Borel sets. Moreover, let $f : \mathbb{R}^2 \to \overline{\mathbb{R}}$ be upper semicontinuous for the product topology $\tau \otimes \tau$ and suppose that f^+ is $\mathcal{S}(\mu,\nu)$-uniformly integrable; i.e.,*

$$\lim_{N\to\infty} \sup_{P\in\mathcal{S}(\mu,\nu)} P(f^+ \mathbf{1}_{f^+>N}) = 0. \tag{9.7}$$

Then, $\mathbf{S}_{\mu,\nu}(f) < \infty$ and there exists an optimal $P \in \mathcal{S}(\mu,\nu)$ for $\mathbf{S}_{\mu,\nu}(f)$.

Condition (9.7) is satisfied in particular if $f(x,y) \leq a(x) + b(y)$ for some functions $a \in L^1(\mu)$ and $b \in L^1(\nu)$.

Proof. Standard arguments show that $\mathcal{S}(\mu,\nu)$ is compact in the usual topology of weak convergence as induced by the Euclidean metric. However, the weak topology on $\mathcal{S}(\mu,\nu)$ induced by $\tau \otimes \tau$ does not depend on the choice of the Polish topology τ as long as $\sigma(\tau) = \mathcal{B}(\mathbb{R})$; this follows from [12, Lemma 2.3]. Thus, $\mathcal{S}(\mu,\nu)$ is still weakly compact relative to $\tau \otimes \tau$.

Under the additional condition that f is bounded from above, the mapping $P \mapsto P(f)$ is upper semicontinuous by [61, Lemma 4.3]. Applying this result to $f \wedge N$ and using (9.7), the same extends to f as in the lemma, and the claim follows. □

We remark that compactness of $\mathcal{S}(\mu,\nu)$ may fail if non-product topologies are considered on $\mathbb{R}^2$, so that the use of $\tau \otimes \tau$ is crucial. The following corollary also improves the existing results in the martingale transport case that occurs when $\mu \leq_c \nu$, so we state that case separately.

Corollary 9.4. *Let $\mu \leq_{cd} \nu$ be probability measures and let $f : \mathbb{R}^2 \to \mathbb{R}$ be Borel and relaxed supermartingale Spence–Mirrlees. Suppose that there exist $a \in L^1(\mu)$, $b \in L^1(\nu)$ such that $f(x,y) \geq a(x) + b(y)$ for all $x, y \in \mathbb{R}$ and that f^+ is $\mathcal{S}(\mu,\nu)$-uniformly integrable; cf. (9.7). Then, $\mathbf{S}_{\mu,\nu}(f) < \infty$ and $\vec{P} \in \mathcal{S}(\mu,\nu)$ is an optimizer. If f is supermartingale Spence–Mirrlees, the optimizer is unique. The analogue holds for $\overleftarrow{P}$ if instead $-f$ is (relaxed) supermartingale Spence–Mirrlees.*

If $\mu \leq_c \nu$, the same result holds with supermartingale Spence–Mirrlees replaced by second-order Spence–Mirrlees, and then $\vec{P}$ (resp. $\overleftarrow{P}$) coincides with the Left-Curtain (Right-Curtain) coupling of [9].

Proof. Let f be supermartingale Spence–Mirrlees (in the strict sense). Under the stated integrability condition, Proposition 9.2 and Lemma 9.3 show that $\mathbf{S}_{\mu,\nu}(f) < \infty$ and that an optimizer $P \in \mathcal{S}(\mu,\nu)$ exists. Now, the monotonicity principle of Theorem 5.2 and Remark 5.4 provide sets $(\Gamma, M) \in$

$\mathcal{B}(\mathbb{R}^2) \times \mathcal{B}(\mathbb{R})$ such that P is concentrated on Γ, $P|_{M\times\mathbb{R}}$ is a martingale and the assertion of Theorem 5.2 (iii) holds. In view of Remark 7.6, we may assume that Γ is nondegenerate by passing to a subset of full P-measure. Proposition 7.8 implies that (Γ, M) is first-order right monotone and second-order left-monotone, and then Theorem 8.1 yields that $P = \vec{P}$.

If f is relaxed supermartingale Spence–Mirrlees, let g be as in Remark 7.2 and note that for each $n \in \mathbb{N}$, the function $f_n = f + (1/n)g$ is supermartingale Spence–Mirrlees in the strict sense. Since f_n satisfies the stated integrability conditions, the above shows that $\vec{P}$ is the unique optimizer for f_n. Suppose that there exists $P_* \in \mathcal{S}(\mu, \nu)$ such that $P_*(f) > \vec{P}(f)$. Then, as monotone convergence yields $P_*(f_n) \to P_*(f)$ and $\vec{P}(f_n) \to \vec{P}(f)$, it follows that $P_*(f_n) > \vec{P}(f_n)$ for n large enough, contradicting the optimality of $\vec{P}$.

The argument for $\overleftarrow{P}$ is similar. The proofs for the martingale case are the same: when $M = \mathbb{R}$, the first-order monotonicity condition is vacuous. □

Finally, we also have the converse of Theorem 8.1 which completes the proofs for the main results as stated in the Introduction.

Corollary 9.5. *Let $\mu \leq_{cd} \nu$ be probability measures and let $\vec{P}$ be the associated Increasing Supermartingale Transport. There exists a nondegenerate pair $(\Gamma, M) \in \mathcal{B}(\mathbb{R}^2) \times \mathcal{B}(\mathbb{R})$ which is first-order right-monotone and second-order left-monotone such that $\vec{P}$ is concentrated on Γ and $\vec{P}|_{M\times\mathbb{R}}$ is a martingale. The analogue, exchanging left and right, holds for $\overleftarrow{P}$.*

Proof. Let g be a supermartingale Spence–Mirrlees function as in Remark 7.2. We know from Corollary 9.4 that $\vec{P}$ is the unique optimal transport for g, and the existence of (Γ, M) follows as in the proof of Corollary 9.4. □

Remark 9.6. Corollary 9.5 shows, in particular, that the no-crossing properties of $\vec{P}$ and $\overleftarrow{P}$ as stated in the Introduction are true for general marginals. The preservation of order mentioned in Figure 2 follows from the two monotonicity properties and nondegeneracy, and together with Remark 7.7, the corollary also yields that $\vec{P}$ has at most one transition from martingale kernels to proper supermartingale ones.

A martingale transport with second-order left-monotone support is the Left-Curtain coupling of its marginals and if the first marginal has no atoms, each kernel of this transport is concentrated on two points [9]. Moreover, an arbitrary transport with first-order right-monotone support is the antitone coupling and if the first marginal has no atoms, its kernels are deterministic [55, Section 3.1]. As a result, if μ is diffuse,

(i) $\vec{P}|_{M^c\times\mathbb{R}}$ is of Monge-type,
(ii) $\vec{P}|_{M\times\mathbb{R}}$ is concentrated on the union of two graphs.

The analogue holds for $\overleftarrow{P}$, with the Right-Curtain and Hoeffding–Fréchet couplings.

10. Counterexamples

10.1. **Duality Theory.** In the Introduction and the body of the text, we have claimed that certain relaxations cannot be avoided. In [11], we have already stated several examples related to the duality theory for the case of martingale transport. Bearing in mind that this is a special case of the supermartingale transport problem at hand, these examples still apply: If the inequality defining the dual elements is stated in the classical sense as $\varphi(x)+\psi(y)+h(x)(y-x) \geq f(x,y)$ for all $(x,y) \in \mathbb{R}^2$ rather than the quasi-sure sense, a duality gap may occur; cf. [11, Example 8.1]. A duality gap may also occur if integrability of dual elements is required in the usual sense; i.e., $\varphi \in L^1(\mu)$, or if f has no lower bound, see [11, Examples 8.4, 8.5].

Next, let us substantiate two claims made in the body of the text. Recall that the set $\mathcal{D}^{ci,pw}_{\mu,\nu}(f)$ was defined with nonnegative functions h, whereas for $\mathcal{D}_{\mu,\nu}(f)$ nonnegativity is required only on the proper portion of the state space (Definitions 4.3 and 4.9). We shall show below that this is necessary.

(i) The requirement that the dual elements (φ,ψ,h) satisfy $h \geq 0$ would preclude existence of dual optimizers.

Second, we have claimed that the restriction to proper pairs $\mu \leq_{cd} \nu$ in Section 4.2 is necessary. While we have already seen that the proof of Proposition 4.4 crucially uses a nontrivial difference of the barycenters of μ and ν in order to control the slope of χ, we still owe an argument that this is indeed unavoidable.

(ii) The closedness property of $\mathcal{D}^{ci,pw}_{\mu,\nu}(f)$ asserted in Proposition 4.4 fails if the (irreducible) pair $\mu \leq_{cd} \nu$ is not proper,

and this remains true even if, in view of (i), we were to alleviate the requirement that $h \geq 0$.

Indeed, let $c_i = i^{-3}C$, $i \geq 1$, where $C > 0$ is such that $\sum c_i = 1$, and define $\mu := \sum_{i\geq 1} c_i\delta_i$ and $\nu := \frac{1}{3}\sum_{i\geq 1} c_i(\delta_{i-1}+\delta_i+\delta_{i+1})$. Moreover, let $f(x,y) = \mathbf{1}_{x\neq y}$. Following [11, Examples 8.4, 8.5] we find that $\mu \leq_{cd} \nu$ is irreducible and

$$P := \sum_{i\geq 1} c_i\,\delta_i \otimes \frac{1}{3}(\delta_{i-1}+\delta_i+\delta_{i+1}) \in \mathcal{S}(\mu,\nu)$$

is a primal optimizer. Clearly, $\mathrm{bary}(\mu) = \mathrm{bary}(\nu)$; i.e., the pair is not proper. Let (φ,ψ,h) be a dual optimizer. Even if we are flexible about the precise definition of the dual domain, a minimal requirement to avoid a duality gap is that $\varphi(x)+\psi(y)+h(x)(y-x) = f(x,y)$ P-a.s. and hence

$$\varphi(x)+\psi(y)+h(x)(y-x) = f(x,y), \quad (x,y) \in \mathbb{N}\times\mathbb{N}_0,\ y \in \{x-1,x,x+1\}.$$

It follows that $\varphi(x) + \psi(x-1) - h(x) = 1$ and $\varphi(x) + \psi(x+1) + h(x) = 1$ and $\varphi(x) + \psi(x) = 0$ for $x \in \mathbb{N}$, and all solutions of this system satisfy

$$\varphi(x) = -x^2 + bx + c, \quad \psi(x) = x^2 - bx - c, \quad h(x) = -2x + b$$

for $x \in \mathbb{N}$, where $b, c \in \mathbb{R}$ are arbitrary constants. While any such triplet defines a dual optimizer in the sense of the body of the paper, we see that there is no solution satisfying $h \geq 0$, which was our claim in (i).

To argue (ii), suppose for contradiction that the closedness property of $\mathcal{D}^{ci,pw}_{\mu,\nu}(f)$ asserted in Proposition 4.4 were true even though $\mu \leq_{cd} \nu$ is not proper. Then, following the proofs in the body of the paper shows that the analogue of Proposition 4.8 would hold as well; i.e., there is no duality gap and there exists a dual optimizer in $\mathcal{D}^{ci,pw}_{\mu,\nu}(f)$. We have seen that this is not the case with the requirement that $h \geq 0$, but it fails even if this is dropped. Indeed, consider again a triplet (φ, ψ, h) satisfying the above system of equations. If $(\varphi, \psi, h) \in \mathcal{D}^{ci,pw}_{\mu,\nu}(f)$, then in particular there exists a concave and increasing moderator χ such that $\varphi - \chi \in L^1(\mu)$. Noting that μ has an infinite second moment and that $\varphi^-(x)$ has quadratic growth as $x \to \infty$ along the integers, it follows that $\chi^-(x)$ must have superlinear growth as $x \to \infty$. But then χ can certainly not be increasing, and we have reached the desired contradiction.

10.2. **Two Couplings that are not Canonical.** As mentioned in the Introduction, it is natural to ask if reward functions f that are first- and second-order Spence-Mirrlees are also maximized by a common supermartingale transport—i.e., $f_{xy} > 0, f_{xyy} > 0$ if f is smooth, rather than the mixed signs that were considered in the preceding sections (see also Example 5.5). However, it turns out that two functions f^1, f^2 satisfying these Spence–Mirrlees conditions may have different optimizers, even if the optimizer is unique for each f^i. This is shown in Example 10.1. The same is true when $-f^i$ are first- and second-order Spence-Mirrlees, as shown by Example 10.2; we confine ourselves to numerical counterexamples.

Example 10.1. Let μ and ν be uniformly distributed on $\{-1, 0, 1\}$ and $\{-4, -2.5, 2\}$, respectively; then $\mu \leq_{cd} \nu$. We consider the reward functions $f^1(x, y) = e^x e^y$ and $f^2(x, y) = e^x e^y + 4xy$ which satisfy $f^i_{xy} > 0$ and $f^i_{xyy} > 0$. The corresponding optimal transports can be obtained with an LP-solver; they are unique and given by

$$\pi^1 = \tfrac{5}{18}\delta_{(-1,-4)} + \tfrac{1}{18}\delta_{(-1,-2.5)} + \tfrac{5}{18}\delta_{(0,-2.5)} + \tfrac{1}{18}\delta_{(0,2)} + \tfrac{1}{18}\delta_{(1,-4)} + \tfrac{5}{18}\delta_{(1,2)},$$
$$\pi^2 = \tfrac{1}{3}\delta_{(-1,-4)} + \tfrac{7}{27}\delta_{(0,-2.5)} + \tfrac{2}{27}\delta_{(0,2)} + \tfrac{2}{27}\delta_{(1,-2.5)} + \tfrac{7}{27}\delta_{(1,2)}.$$

Their supports are shown in Figure 4. The transports are first- and second-order left-monotone with $M = \{1\}$, but the supports are different.

FIGURE 4. The optimal transports from Example 10.1

Example 10.2. Let μ and ν be uniformly distributed on $\{-1, 0, 1\}$ and $\{-4, -2.5, 0.5\}$, respectively; then again $\mu \leq_{cd} \nu$. We consider the reward functions $f^1(x, y) = -e^x e^y$ and $f^2(x, y) = -e^x e^y - 4xy$ which are the negatives of the functions in Example 10.1; they satisfy $f^i_{xy} < 0$ and $f^i_{xyy} < 0$. The corresponding (unique) optimal transports are given by

$$\pi^1 = \tfrac{1}{9}\delta_{(-1,-4)} + \tfrac{2}{9}\delta_{(-1,0.5)} + \tfrac{2}{9}\delta_{(0,-2.5)} + \tfrac{1}{9}\delta_{(0,0.5)} + \tfrac{2}{9}\delta_{(1,-4)} + \tfrac{1}{9}\delta_{(1,-2.5)},$$
$$\pi^2 = \tfrac{1}{6}\delta_{(-1,-2.5)} + \tfrac{1}{6}\delta_{(-1,0.5)} + \tfrac{1}{6}\delta_{(0,-2.5)} + \tfrac{1}{6}\delta_{(0,0.5)} + \tfrac{1}{3}\delta_{(1,-4)}.$$

Their supports are shown in Figure 5. The transports are first- and second-order right-monotone with $M = \{-1\}$, but the supports are different.

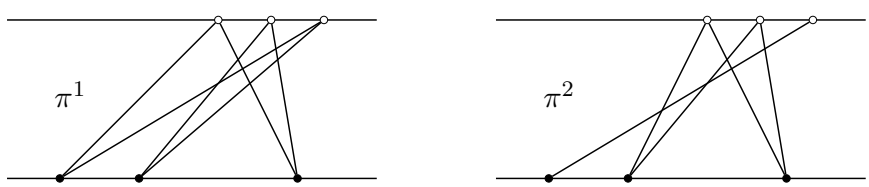

FIGURE 5. The optimal transports from Example 10.2

REFERENCES

[1] B. Acciaio, M. Beiglböck, F. Penkner, and W. Schachermayer. A model-free version of the fundamental theorem of asset pricing and the super-replication theorem. *Math. Finance*, 26(2):233–251, 2016.

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

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

[4] M. Beiglböck, A. M. G. Cox, M. Huesmann, N. Perkowski, and D. J. Prömel. Pathwise super-replication via Vovk's outer measure. *To appear in Finance Stoch.*, 2015.

[5] M. Beiglböck, M. Goldstern, G. Maresch, and W. Schachermayer. Optimal and better transport plans. *J. Funct. Anal.*, 256(6):1907–1927, 2009.

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

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

[8] M. Beiglböck, M. Huesmann, and F. Stebegg. Root to Kellerer. In *Séminaire de probabilités XLVIII*, volume 2168 of *Lecture Notes in Math.*, pages 1–12. Springer, Berlin, 2016.

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

[10] M. Beiglböck and M. Nutz. Martingale inequalities and deterministic counterparts. *Electron. J. Probab.*, 19(95):1–15, 2014.

[11] M. Beiglböck, M. Nutz, and N. Touzi. Complete duality for martingale optimal transport on the line. *Ann. Probab.*, 45(5):3038–3074, 2017.

[12] M. Beiglböck and A. Pratelli. Duality for rectified cost functions. *Calc. Var. Partial Differential Equations*, 45(1-2):27–41, 2012.

[13] D. P. Bertsekas and S. E. Shreve. *Stochastic Optimal Control. The Discrete-Time Case.* Academic Press, New York, 1978.

[14] S. Biagini, B. Bouchard, C. Kardaras, and M. Nutz. Robust fundamental theorem for continuous processes. *Math. Finance*, 27(4):963–987, 2017.

[15] B. Bouchard and M. Nutz. Arbitrage and duality in nondominated discrete-time models. *Ann. Appl. Probab.*, 25(2):823–859, 2015.

[16] M. Burzoni, M. Frittelli, and M. Maggis. Model-free superhedging duality. *Ann. Appl. Probab.*, 27(3):1452–1477, 2017.

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

[18] P. Cheridito, M. Kupper, and L. Tangpi. Representation of increasing convex functionals with countably additive measures. *Preprint arXiv:1502.05763v1*, 2015.

[19] A. M. G. Cox. Extending Chacon-Walsh: minimality and generalised starting distributions. In *Séminaire de probabilités XLI*, volume 1934 of *Lecture Notes in Math.*, pages 233–264. Springer, Berlin, 2008.

[20] A. M. G. Cox, Z. Hou, and J. Obłój. Robust pricing and hedging under trading restrictions and the emergence of local martingale models. *Finance Stoch.*, 20(3):669–704, 2016.

[21] A. M. G. Cox and J. Obłój. Robust pricing and hedging of double no-touch options. *Finance Stoch.*, 15(3):573–605, 2011.

[22] A. M. G. Cox, J. Obłój, and N. Touzi. The Root solution to the multi-marginal embedding problem: an optimal stopping and time-reversal approach. *Preprint arXiv:1505.03169v1*, 2015.

[23] S. De Marco and P. Henry-Labordère. Linking vanillas and VIX options: a constrained martingale optimal transport problem. *SIAM J. Financial Math.*, 6(1):1171–1194, 2015.

[24] Y. Dolinsky and A. Neufeld. Super-replication in extremely incomplete markets. *To appear in Math. Finance*, 2015.

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

[26] Y. Dolinsky and H. M. Soner. Martingale optimal transport in the Skorokhod space. *Stochastic Process. Appl.*, 125(10):3893–3931, 2015.

[27] A. Fahim and Y.-J. Huang. Model-independent superhedging under portfolio constraints. *Finance Stoch.*, 20(1):51–81, 2016.

[28] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time.* W. de Gruyter, Berlin, 3rd edition, 2011.

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

[30] N. Ghoussoub, Y.-H. Kim, and T. Lim. Structure of optimal martingale transport in general dimensions. *Preprint arXiv:1508.01806v1*, 2015.
[31] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *J. Funct. Anal.*, 273(11):3327–3405, 2017.
[32] C. Griessler. An extended footnote on finitely minimal martingale measures. *Preprint arXiv:1606.03106v1*, 2016.
[33] G. Guo, X. Tan, and N. Touzi. On the monotonicity principle of optimal Skorokhod embedding problem. *SIAM J. Control Optim.*, 54(5):2478–2489, 2016.
[34] G. Guo, X. Tan, and N. Touzi. Optimal Skorokhod embedding under finitely many marginal constraints. *SIAM J. Control Optim.*, 54(4):2174–2201, 2016.
[35] G. Guo, X. Tan, and N. Touzi. Tightness and duality of martingale transport on the Skorokhod space. *Stochastic Process. Appl.*, 127(3):927–956, 2017.
[36] P. Henry-Labordère, J. Obłój, P. Spoida, and N. Touzi. Maximum maximum of martingales given marginals. *Ann. Appl. Probab.*, 26(1):1–44, 2016.
[37] P. Henry-Labordère, X. Tan, and N. Touzi. An explicit version of the one-dimensional Brenier's theorem with full marginals constraint. *Stochastic Process. Appl.*, 126(9):2800–2834, 2016.
[38] P. Henry-Labordère and N. Touzi. An explicit martingale version of the one-dimensional Brenier theorem. *Finance Stoch.*, 20(3):635–668, 2016.
[39] F. Hirsch, C. Profeta, B. Roynette, and M. Yor. *Peacocks and Associated Martingales, with Explicit Constructions*. Springer, Milan, 2011.
[40] D. Hobson. Robust hedging of the lookback option. *Finance Stoch.*, 2(4):329–347, 1998.
[41] D. Hobson. The Skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance 2010*, volume 2003 of *Lecture Notes in Math.*, pages 267–318. Springer, Berlin, 2011.
[42] D. Hobson. Mimicking martingales. *Ann. Appl. Probab.*, 26(4):2273–2303, 2016.
[43] D. Hobson and M. Klimmek. Robust price bounds for the forward starting straddle. *Finance Stoch.*, 19(1):189–214, 2015.
[44] D. Hobson and A. Neuberger. Robust bounds for forward start options. *Math. Finance*, 22(1):31–56, 2012.
[45] N. Juillet. Stability of the shadow projection and the left-curtain coupling. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(4):1823–1843, 2016.
[46] S. Källblad, X. Tan, and N. Touzi. Optimal Skorokhod embedding given full marginals and Azéma–Yor peacocks. *Ann. Appl. Probab.*, 27(2):686–719, 2017.
[47] A. S. Kechris. *Classical Descriptive Set Theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.
[48] H. G. Kellerer. Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete*, 67(4):399–432, 1984.
[49] A. Neufeld and M. Nutz. Superreplication under volatility uncertainty for measurable claims. *Electron. J. Probab.*, 18(48):1–14, 2013.
[50] M. Nutz. Superreplication under model uncertainty in discrete time. *Finance Stoch.*, 18(4):791–803, 2014.
[51] M. Nutz. Robust superhedging with jumps and diffusion. *Stochastic Process. Appl.*, 125(12):4543–4555, 2015.
[52] M. Nutz, F. Stebegg, and X. Tan. Multiperiod martingale transport. *Preprint arXiv:1703.10588v1*, 2017.
[53] J. Obłój. The Skorokhod embedding problem and its offspring. *Probab. Surv.*, 1:321–390, 2004.

[54] J. Obłój and P. Spoida. An iterated Azéma–Yor type embedding for finitely many marginals. *Ann. Probab.*, 45(4):2210–2247, 2017.

[55] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Theory.

[56] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. II.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Applications.

[57] F. Stebegg. Model-independent pricing of Asian options via optimal martingale transport. *Preprint arXiv:1412.1429v1*, 2014.

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

[59] N. Touzi. Martingale inequalities, optimal martingale transport, and robust superhedging. In *Congrès SMAI 2013*, volume 45 of *ESAIM Proc. Surveys*, pages 32–47. EDP Sci., Les Ulis, 2014.

[60] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[61] C. Villani. *Optimal Transport, Old and New*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2009.

[62] D. Zaev. On the Monge–Kantorovich problem with additional linear constraints. *Math. Notes*, 98(5):725–741, 2015.