The Directional Optimal Transport[*]

Marcel Nutz[†]   Ruodu Wang[‡]

June 19, 2021

Abstract

We introduce a constrained optimal transport problem where origins x can only be transported to destinations $y \geq x$. Our statistical motivation is to describe the sharp upper bound for the variance of the treatment effect $Y - X$ given marginals when the effect is monotone, or $Y \geq X$. We thus focus on supermodular costs (or submodular rewards) and introduce a coupling P_* that is optimal for all such costs and yields the sharp bound. This coupling admits manifold characterizations—geometric, order-theoretic, as optimal transport, through the cdf, and via the transport kernel—that explain its structure and imply useful bounds. When the first marginal is atomless, P_* is concentrated on the graphs of two maps which can be described in terms of the marginals, the second map arising due to the binding constraint.

Keywords: Optimal Transport; Monotone Treatment Effect; Submodular Reward
AMS 2010 Subject Classification: 49N05; 62G10; 93E20

1 Introduction

We study a constrained Monge–Kantorovich optimal transport problem between marginal distributions μ and ν on the real line where the couplings are required to be "directional" in the sense that an origin x can only be transported to destinations y with $y \geq x$. While one can think of several natural transport or matching problems with such a constraint, our initial motivation comes from the statistical analysis of treatment effects. There, one compares a (treated) experiment group of patients with an (untreated) control group. A fundamental problem is that any potential outcome that

treated patients would have received without treatment is not observed, and vice versa. While the marginal distributions μ and ν of the performance evaluations X and Y of the two groups can be estimated from experiment data, the joint distribution cannot, as the two groups are non-overlapping by design—Neyman noted as early as 1923 (cf. [4]) that there are no unbiased or consistent estimators for the covariance. The improvement of the performance measure due to treatment, $Y - X$, is known as *treatment effect.* To test the hypothesis of substantial treatment effect, it is important to understand bounds on $\mathrm{Var}(Y - X)$ or more generally the joint distribution P of (X, Y). Crude (yet popular) bounds can be obtained by mapping one group to the extremes of the support of the other. The classical Fréchet–Hoeffding (or Hardy–Littlewood) mechanism gives better bounds and is often used in the literature (see, e.g., [4, 14], and [26, 27] for mathematical background). The lower bound for $\mathrm{Var}(Y - X)$ over all couplings is attained by the comonotone (or Fréchet–Hoeffding) coupling. The upper bound over all couplings leads to the antitone coupling, which may be unrealistic in the context of many treatment effects: this coupling corresponds to the idea that the healthiest untreated subject would have become the least healthy patient if treated, and vice versa, which seems exceedingly pessimistic, e.g., in a study on the impact of physical activity on obesity. As proposed in [22], this issue can be alleviated by the assumption of *monotone treatment effect* when suitable, postulating that the treatment effect is nonnegative: $Y \geq X$ means that an untreated individual's performance would not have been worsened by the treatment, and vice versa. Of course, this assumption is only made after verifying that ν stochastically dominates μ in the data. Under the assumption of monotone treatment effect, the sharp upper bound of $\mathrm{Var}(Y - X)$ corresponds to a coupling P_* that we call optimal directional coupling.[1] More generally, P_* yields the sharp upper bound for $E^P[g(X, Y)]$ whenever g is supermodular. The lower bound remains trivial in that it still corresponds to the comonotone coupling (which satisfies $Y \geq X$ in view of the necessary stochastic dominance), whence our focus on the upper bound.

In the next section we introduce P_* for general marginals μ, ν in stochastic order and provide manifold characterizations that resemble familiar properties of the antitone coupling while also taking into account the constraint. Globally, the geometry is significantly richer than in the classical antitone case. At a local level, the interaction between supermodularity and con-

---

[1] We prefer "directional" over "monotone" as the latter terminology often refers to the Fréchet–Hoeffding coupling in the transport literature.
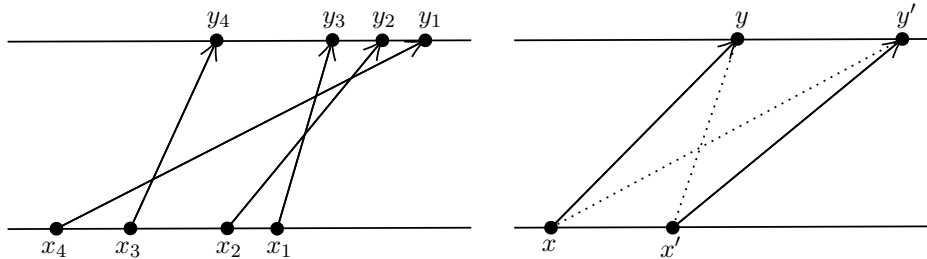
Figure 1: Left panel: An example of P_*, with the y-axis shown at the top. Right panel: An improvable pair which can be "improved" to the dotted pair.

straint is much more transparent, and each of our characterizations clarifies that interaction from a different angle.

The construction of P_* is best explained in the simple case $\mu = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$ and $\nu = \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}$ where both marginals consist of a common number of atoms of equal size at distinct locations, and moreover $x_1 > \cdots > x_n$ are numbered from *right to left*. The transport P_* processes these atoms x_i in that order, sending each origin to the minimal (left-most) destination $y = T(x_i)$ that is allowed by the constraint $y \geq T(x_i)$ and has not been filled yet (Figure 1). That is, starting with the set $S_1 = \{y_1, \ldots, y_n\}$ of all destinations, we iterate for $k = 1, \ldots, n$:

(i) $T(x_k) := \min\{y \in S_k : y \geq x_k\}$,

(ii) $S_{k+1} := S_k \setminus \{T(x_k)\}$.

A less formal description is to imagine a left parenthesis "(" at each location x_i and a right parenthesis ")" at each y_i. Then T agrees with to the usual rule of matching a left with its corresponding right parenthesis in a mathematical statement. The antitone coupling would be obtained omitting the inequality in (i) above, making apparent how the constraint creates the difference with the classical coupling at the local level.

Further properties provided in the next section include a geometric characterization through the support of the coupling and of course the optimality as transport for all supermodular costs (or submodular rewards, including variance of treatment effect); here the notion of cyclical monotonicity plays a key role. In particular, we provide sharp conditions under which P_* admits a Monge map. Finally, one can also describe P_* through its joint cdf.

The constraint is responsible for qualitative differences with the antitone coupling. Assuming that the first marginal is atomless, the latter coupling

3

always admits a Monge map, in other words, it is concentrated on a graph. By contrast, the constrained coupling is concentrated on two graphs. The two maps can be described in detail: one is the identity function and appears when the constraint is locally binding, the other admits a graphical interpretation and a semi-explicit formula based on the difference of the marginal cdf's. The appearance of the identity is clearly reminiscent of the unconstrained transport problem for costs like $c(x, y) = |y - x|^p$, $0 < p < 1$ that combine concavity away from the origin with convexity at the origin, and was first observed in [17] in that context. See also [29, Section 3.3.2] for a discussion. Another difference is the behavior under marginal transformations. The antitone coupling is invariant with respect to arbitrary monotone transformations of the coordinate axes; more precisely, the copula corresponding to the coupling is the same for all marginals. This is no longer true for the constrained version, the reason being that the underlying constraint $Y \geq X$ is not invariant. Instead, the copula depends on the marginals and an invariance property holds only when a common transformation is applied to both axes.

Several constrained optimal transport problems have been of lively interest in recent years. One related problem is the optimal transport with quadratic cost $c(x, y) = |y - x|^2$ in $\mathbb{R}^d$ studied in [19] (see also [10, 11]) under a convex constraint: transports have to satisfy $y - x \in C$ for a given convex set C. It is shown that this problem admits an optimal transport map (Monge map) in great generality. The specification $y - x \in C$ accommodates our constraint, but minimizing the quadratic cost (rather than maximizing) yields the comonotone coupling in our setting. Indeed, [19] details that the comonotone coupling is the optimal solution for general C in the scalar case—the constraint is not binding as soon as an admissible coupling exists. In our problem, the constraint is typically binding and the optimal coupling typically does not admit a Monge map but instead requires a randomization between two maps. (See also Section 6.3 for a generalization of P_* to cone constraints that may simplify the comparison with [19].)

A different constrained problem is the martingale optimal transport introduced in [6, 16, 33], corresponding to the constraint $E[Y|X] = X$ as motivated from financial mathematics (see [1, 7, 8, 9, 12, 18], among many others). In particular, the Left- and Right-Curtain couplings of [7] correspond to the constrained versions of the comonotone/antitone couplings. It is worth noting that these couplings are also concentrated on the graphs of two maps in typical cases, like P_*. (However, the appearance of a randomization is more obvious: only a constant martingale is deterministic.) The supermartingale constraint $E[Y|X] \leq X$ in [24] resembles the current situ-

4

ation in being an inequality constraint. Compared to all of these examples, the present case yields by far the most explicit and detailed results. In hindsight, the directional transport is arguably the most canonical and simplest nontrivial example of a constrained optimal transport problem. For general transport problems in Polish spaces, cyclical monotonicity and duality theory with constraints (or equivalently cost functions with infinite values) were studied by [2, 5, 13, 21, 31], among others.

The literature on copulas features several directly related results; these works seem to be mostly unaware of one another and of the results in the optimal transport literature. The earliest related contribution that we are aware of, [32], features a bound on the cdf of any directional coupling (see also Remark 4.4 below). It is not investigated if or when that bound corresponds to a coupling. Almost two decades later, [28] was interested in coupling random walks "fast" and determined a directional coupling which maximizes a cost of the form $\varphi(y - x)$ with φ strictly convex, nonnegative and decreasing. It is clear from Theorem 2.2 below that this coupling is P_*; the decrease of φ is irrelevant as convexity alone implies submodularity. In [28], the application to random walks is successful only when the difference of the marginal distributions is unimodular, and in that case, P_* has a trivial structure as the sum of an identity and an antitone coupling between disjoint intervals (see Example 4.5 below)—that may explain why [28] did not investigate the coupling further. The recent work [3] characterizes all directional dependence structures of marginals in stochastic order and derives several related bounds, in particular one on the cdf which gives exactly the cdf of P_*. (In fact, the same cdf was previously stated in [28], in a slightly more implicit form.) The structure of the coupling, and more generally the point of view of optimal transport, are not highlighted in these works.

While we hope that this paper is a fairly complete study of the scalar case with inequality constraint (or, more generally, one-dimensional cone constraint; cf. Section 6.3), we mention that the multidimensional case is wide open. To stick with the above motivation, consider a treatment which affects two (or more) separately measured qualities—e.g., the impact of physical exercise on blood pressure and body mass index. Control and experiment group now give rise to distributions in $\mathbb{R}^2$, and the assumption of monotone treatment effect for both performance measures corresponds to a *cone* constraint $y - x \in [0, \infty)^2$. It is worth noting that even if a scalar quantity is used to aggregate the two performances, the cone constraint is typically more stringent than what would be obtained by constraining the aggregated performances.

The remainder of the paper is organized as follows. Section 2 formalizes

the problem and presents the main results. The subsequent Sections 3–5 provide the proofs and some required tools, as well as examples and additional consequences. Section 6 gathers three discussions that we omitted in the main results: another decomposition of P_*, optimality properties in unconstrained transport problems, and an extension to general (random) cone constraints.

2 Main Results

Let μ and ν be probability measures on $\mathbb{R}$ and denote by $X(x, y) = x$, $Y(x, y) = y$ the coordinate projections on $\mathbb{R}^2$. A coupling, or transport, of μ and ν is a probability P on $\mathbb{R}^2$ with marginals $P \circ X^{-1} = \mu$ and $P \circ Y^{-1} = \nu$. We call a coupling P *directional* if it is concentrated on the closed halfplane above the diagonal,

$$\mathbb{H} = \{Y \geq X\} = \{(x, y) \in \mathbb{R}^2 : y \geq x\},$$

meaning that μ-almost every origin x is transported to a destination located to the right of x (or to x itself). Denoting by $\mathcal{D} = \mathcal{D}(\mu, \nu)$ the set of all directional couplings, we have $\mathcal{D} \neq \emptyset$ if and only if μ and ν are in stochastic order, denoted $\mu \preceq_{\mathrm{st}} \nu$, meaning that their cdf's satisfy $F_\mu \geq F_\nu$. Indeed, $\mu \preceq_{\mathrm{st}} \nu$ if and only if the comonotone coupling is directional. More generally, we indicate by $\theta_1 \preceq_{\mathrm{st}} \theta_2$ two subprobabilities with common mass $\theta_1(\mathbb{R}) = \theta_2(\mathbb{R})$ and $F_{\theta_1} \geq F_{\theta_2}$. The other notions also have obvious generalizations.

The following theorem corresponds to a general version of the discrete construction of P_* in the Introduction. We write $\theta \leq \nu$ for a subprobability θ with $\theta(A) \leq \nu(A)$ for all $A \in \mathcal{B}(\mathbb{R})$.

Theorem 2.1. *Let $\mu \preceq_{\mathrm{st}} \nu$. There exists a unique directional coupling $P_* = P_*(\mu, \nu)$ which couples $\mu|_{(x, \infty)}$ to ν_x for all $x \in \mathbb{R}$, where the subprobability ν_x is defined by its cdf*

$$F_{\nu_x} = \sup_{\theta \in S_x} F_\theta \quad for \quad S_x = \{\theta : \mu|_{(x, \infty)} \preceq_{\mathrm{st}} \theta \leq \nu\}.$$

The measure ν_x is the unique minimal element of S_x for the order $\preceq_{\mathrm{st}}$.

The coupling P_* differs from the antitone coupling except in the trivial case where all couplings are directional; that is, when $\mu((-\infty, x]) = \nu([x, \infty)) = 1$ for some $x \in \mathbb{R}$. Indeed, this is the only case where the antitone coupling is directional.

We make $\mu \preceq_{\mathrm{st}} \nu$ a *standing assumption* in all that follows. The above theorem is one of several equivalent characterizations of P_* that we detail next. The most important for our analysis is geometric, describing the support of P_* based on the idea that we would like any two trajectories of the transport to cross whenever that is allowed by the constraint. We say that the pair $((x,y),(x',y')) \in \mathbb{H}^2$ is *improvable* if $x < x' \leq y < y'$. This means that (x,y) and (x',y') do not cross, but they could be rearranged ("improved") into the configuration $((x,y'),(x',y))$ which forms a cross and remains $\mathbb{H}^2$ (Figure 1). A set $\Gamma \subseteq \mathbb{H}$ satisfies the *constrained crossing property* if it contains no improvable pairs. Stated differently, any two trajectories in Γ either cross, or they cannot be rearranged into a cross without exiting $\mathbb{H}$.

This property is closely related to a characterization of P_* through optimal transport with specific reward functions. A Borel function $g : \mathbb{H} \to \mathbb{R}$ is *submodular* (on $\mathbb{H}$) if

$$g(x,y) + g(x',y') \leq g(x,y') + g(x',y) \quad \text{for all} \quad x < x' \leq y < y' \qquad (2.1)$$

and *strictly submodular* if the inequality in (2.1) is strict; two examples are $g(x,y) = (x-y)^2$ and $g(x,y) = -\sqrt{|x-y|}$. If g is differentiable, the Spence–Mirrlees condition $-g_{xy} > 0$ is a sufficient condition. We say that g is (μ,ν)-*integrable* if $|g(x,y)| \leq \phi(x) + \psi(y)$ for some $\phi \in L^1(\mu)$ and $\psi \in L^1(\nu)$. This implies uniform bounds on $\int g\,dP$ for any coupling P and in particular that the optimal transport problem

$$\sup_{P \in \mathcal{D}} \int g\,dP \quad \left(\text{or equivalently,} \ \inf_{P \in \mathcal{D}} \int -g\,dP\right) \qquad (2.2)$$

is finite as soon as $\mathcal{D} \neq \emptyset$. Finally, $P \in \mathcal{D}$ is *optimal* for g if it attains the supremum. To see the connection with the constrained crossing property, observe that for any strictly submodular g,

$$g(x,y) + g(x',y') < g(x,y') + g(x',y) \quad \text{if} \quad ((x,y),(x',y')) \text{ is improvable.}$$

The following result also contains a third (straightforward) characterization in terms of the so-called concordance order in (i).

Theorem 2.2. *For a coupling $P \in \mathcal{D}(\mu,\nu)$, the following are equivalent.*

(i) *$F_P \leq F_Q$ on $\mathbb{R}^2$ for all $Q \in \mathcal{D}(\mu,\nu)$, where F_Q is the cdf of Q.*

(ii) *P is optimal for all (μ,ν)-integrable and submodular g.*

(iii) *P is optimal for some (μ,ν)-integrable and strictly submodular g.*

(iv) P is supported by a set $\Gamma \subseteq \mathbb{H}$ with the constrained crossing property.

(v) $P = P_*$.

The geometric characterization in Theorem 2.2 (iv) implies that the optimal coupling P_* is invariant with respect to common transformations of both coordinate axes as follows.

Corollary 2.3. *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a strictly increasing function. Then*

$$P_*(\mu, \nu) = P_*(\mu \circ \phi^{-1}, \nu \circ \phi^{-1}) \circ (\phi, \phi).$$

In particular, copulas of $P_*(\mu, \nu)$ are precisely those of $P_*(\mu \circ \phi^{-1}, \nu \circ \phi^{-1})$, and thus these copulas are invariant under common, strictly increasing transformations of the axes. The strict increase of ϕ is necessary to retain the constrained crossing property. Similarly, it is clear that the same transformation must be applied to both axes—in contrast to the unconstrained transport problem, as highlighted in the Introduction.

Theorem 2.2 (i) yields an implicit description of the optimal cdf which, by a result of [3], implies the following representation. A proof by direct computation will be sketched in Section 4, as well as resulting bounds.

Corollary 2.4. *The cdf of P_* is given by*

$$F_*(x, y) = \begin{cases} F_\nu(y) & \text{if } y \leq x, \\ F_\mu(x) - \inf_{z \in [x,y]}(F_\mu(z) - F_\nu(z)) & \text{if } y > x. \end{cases} \tag{2.3}$$

See also Figure 2 for a graphical representation. As a first consequence, we observe the continuity of P_* with respect to weak convergence ($\xrightarrow{\text{w}}$) of the marginals.

Corollary 2.5. *Consider marginals $\mu_n \preceq_{\text{st}} \nu_n$, $n \geq 1$ with $\mu_n \xrightarrow{\text{w}} \mu$ and $\nu_n \xrightarrow{\text{w}} \nu$, and suppose that μ and ν are atomless. Then $P_*(\mu_n, \nu_n) \xrightarrow{\text{w}} P_*(\mu, \nu)$.*
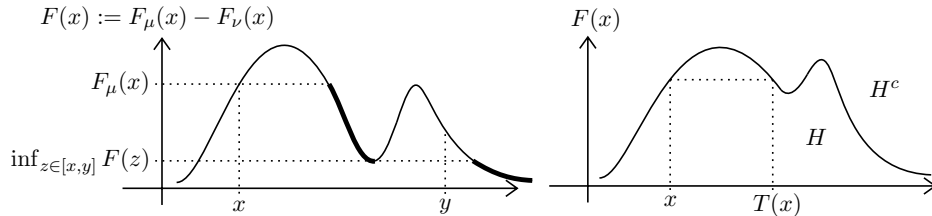

Figure 2: Left panel: On the formula for F_*. Right panel: Definition of T.

We will see in Example 4.2 that the continuity can fail in the presence of atoms.

The subsequent results describe the finer structure of the optimal transport. The *common part* $\mu \wedge \nu$ of μ and ν is the measure defined by

$$\frac{d(\mu \wedge \nu)}{d(\mu + \nu)} := \frac{d\mu}{d(\mu + \nu)} \wedge \frac{du}{d(\mu + \nu)}.$$

Alternately, $\mu \wedge \nu$ is the maximal measure θ satisfying $\theta \leq \mu$ and $\theta \leq \nu$, and we can note that μ, ν are mutually singular if and only if $\mu \wedge \nu = 0$. Importantly, P_* always transports $\mu \wedge \nu$ according to the identity coupling, similarly as in [17, Main Theorem 6.4] for unconstrained transport with cost $l(|y - x|)$ and l strictly concave (see Figure 3 for two simple examples).

Proposition 2.6. *The optimal coupling $P_*(\mu, \nu)$ satisfies*

$$P_*(\mu, \nu) = \mathrm{Id}(\mu \wedge \nu) + P_*(\mu', \nu')$$

where $\mathrm{Id}(\mu \wedge \nu) = (\mu \wedge \nu) \otimes_x \delta_x$ is the identical coupling of $\mu \wedge \nu$ with itself whereas $\mu' = \mu - \mu \wedge \nu$ and $\nu' = \nu - \mu \wedge \nu$ are the mutually singular parts of μ and ν.

A coupling P is of *Monge-type* if $P(Y|X) = T(X)$ is a deterministic function T of X which is then called a Monge map or transport map of P. Equivalently, the stochastic kernel κ in the decomposition $P = \mu \otimes \kappa$ has the form $\kappa(x, dy) = \delta_{T(x)}(dy)$ μ-a.s. Proposition 2.6 suggests that the constrained nature of our transport problem may render P_* randomized (i.e., not of Monge-type) even in the absence of atoms.

Example 2.7. Let $\mu = \mathrm{Unif}[0,1]$ and $\nu = \mathrm{Unif}[0,2]$. Then $\mu \preceq_{\mathrm{st}} \nu$ and there are no atoms, yet P_* has non-deterministic kernel $\kappa(x) = \frac{1}{2}(\delta_x + \delta_{2-x})$; cf. Figure 3. This can be seen, e.g., from the constrained crossing property.
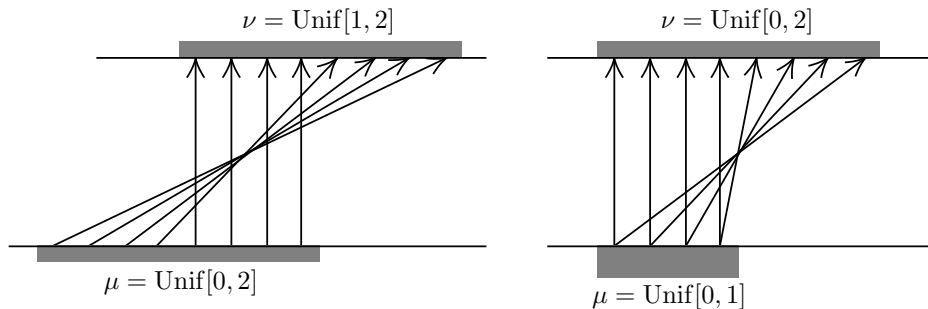

Figure 3: Illustration of Proposition 2.6 (left) and Example 2.7 (right).

The next results show that this example is representative: the "coin-flip" randomization into two maps is the only randomization in P_* when μ is atomless, and it occurs if and only if $\mu \wedge \nu$ and $\mu - \mu \wedge \nu$ are not mutually singular. The second transport map can also be analyzed in detail. To that end, suppose first that $\mu \wedge \nu = 0$, so that (μ, ν) is already in the reduced form (μ', ν') of Proposition 2.6. Moreover, suppose for the moment that the marginals are atomless—we discuss later how to reduce atoms to diffuse measures. With the convention $\inf \emptyset = \infty$, we have the following (see Figure 2 for the graphical interpretation).

Theorem 2.8. *Let μ, ν be atomless and $\mu \wedge \nu = 0$. Then P_* is of Monge-type with transport map T given by*

$$T(x) = \inf\{y \geq x : (y, F(x)) \notin H\}$$

for the function $F = F_\mu - F_\nu$ and its hypograph $H = \{(x, z) : z \leq F(x)\}$.

The proof proceeds by showing that T couples μ and ν and that the graph of T satisfies the constrained crossing property. Some of our considerations regarding the local regularity of F may be of independent interest. Combining the last two results and noting that $F_\mu - F_\nu = F_{\mu'} - F_{\nu'}$ in Proposition 2.6, we deduce the aforementioned assertion on the coin-flip.

Corollary 2.9. *Let μ, ν be atomless. Then*

$$P_*(\mu, \nu) = (\mu \wedge \nu) \otimes_x \delta_x + \mu' \otimes_x \delta_{T(x)}$$

where $\mu' = \mu - \mu \wedge \nu$. In particular, P_* is of Monge-type if and only if μ' and $\mu \wedge \nu$ are mutually singular.

This result immediately extends to the case where ν has atoms, essentially by "filling in" vertical lines in the graph of F where there are jumps (cf. Figure 2). Using a simple transformation detailed in Section 5.4, it also generalizes to atoms in both marginals, but then T is replaced by a (possibly randomized) coupling; see Theorem 5.5.

We remark that the invariance property in Corollary 2.3 translates immediately: if T is the map of $P_*(\mu, \nu)$, then $T^\phi := \phi \circ T \circ \phi^{-1}$ is that of $P_*(\mu \circ \phi^{-1}, \nu \circ \phi^{-1})$; in other words, T^ϕ transports $\phi(x)$ to $\phi(y)$ whenever T transports x to y.

While we consider the above the main results, three further considerations are presented in Section 6. We discuss when and how P_* can be decomposed as a sum of antitone couplings of sub-marginals, remark that P_* occurs as optimizer in specific unconstrained transport problems, and finally offer an extension to cone constraints more general than $Y \geq X$.

3 Equivalent Characterizations of P_*

In this section we prove Theorems 2.1–2.2 and Proposition 2.6, the latter being a consequence of the former. The first step is to show that ν_x in Theorem 2.1 is well-defined. We write $\mathcal{M}$ for the set of finite measures on $\mathbb{R}$ and recall that $\theta_1, \theta_2 \in \mathcal{M}$ satisfy $\theta_1 \preceq_{\mathrm{st}} \theta_2$ if $\theta_1(\mathbb{R}) = \theta_2(\mathbb{R})$ and $F_{\theta_1} \geq F_{\theta_2}$.

Lemma 3.1. *Let $\mu_0 \leq \mu$. The set $S = \{\theta \in \mathcal{M} : \mu_0 \preceq_{\mathrm{st}} \theta \leq \nu\}$ has a unique minimal element θ_*; that is, $\theta_* \in S$ and $\theta_* \preceq_{\mathrm{st}} \theta$ for all $\theta \in S$. The measure θ_* has cdf $\sup_{\theta \in S} F_\theta$ and we denote $\theta^\nu(\mu_0) := \theta_*$.*

Proof. We first show that $F := \sup_{\theta \in S} F_\theta$ is a cdf. Given $x < y$, we have $F_\theta(y) - F_\theta(x) \leq F_\nu(y) - F_\nu(x)$ for any $\theta \in S$ and hence

$$F(y) - F(x) \leq \sup_{\theta \in S} [F_\theta(y) - F_\theta(x)] \leq F_\nu(y) - F_\nu(x) \to 0 \quad \text{as} \quad y \downarrow x,$$

showing that F is right-continuous. As the remaining properties of a cdf are immediate, we can introduce θ_* as the measure associated to F. In view of $F = \sup_{\theta \in S} F_\theta$, we have that $\mu_0 \preceq_{\mathrm{st}} \theta_*$ and $\theta_* \preceq_{\mathrm{st}} \theta$ for every $\theta \in S$. It remains to see that $\theta_* \leq \nu$, or equivalently that $F_{\nu - \theta_*}$ is nondecreasing. Indeed, $F_{\nu - \theta_*} = F_\nu - \sup_{\theta \in S} F_\theta = \inf_{\theta \in S} F_{\nu - \theta}$, and $F_{\nu - \theta}$ is nondecreasing for every $\theta \in S$. $\qquad\square$

Next, we show that the map $\mu_0 \mapsto \theta^\nu(\mu_0)$ of Lemma 3.1 is "divisible", which is important for its iterated application: mapping $\mu_0 = \mu_1 + \mu_2$ into ν produces the same cumulative result as first mapping μ_1 and then mapping μ_2 into the remaining part of ν.

Lemma 3.2. *Let μ_1, μ_2 satisfy $\mu_1 + \mu_2 \leq \mu$. Then $\mu - \mu_1 \preceq_{\mathrm{st}} \nu - \theta^\nu(\mu_1)$ and*

$$\theta^\nu(\mu_1 + \mu_2) = \theta^\nu(\mu_1) + \theta^{\nu - \theta^\nu(\mu_1)}(\mu_2).$$

Proof. Let $Q = \mu \otimes \kappa \in \mathcal{D}(\mu, \nu)$ be arbitrary and let $Q(\mu_1)$ be its image of μ_1 (that is, the second marginal of $\mu_1 \otimes \kappa$). In view of $Q \in \mathcal{D}(\mu, \nu)$ we have $\mu_1 \preceq_{\mathrm{st}} Q(\mu_1) \leq \nu$ and $\mu - \mu_1 \preceq_{\mathrm{st}} \nu - Q(\mu_1)$. The minimality property of $\theta^\nu(\mu_1)$ then yields $\theta^\nu(\mu_1) \preceq_{\mathrm{st}} Q(\mu_1)$ and therefore

$$\mu_2 \leq \mu - \mu_1 \preceq_{\mathrm{st}} \nu - Q(\mu_1) \preceq_{\mathrm{st}} \nu - \theta^\nu(\mu_1).$$

In particular, the measure $\theta^{\nu - \theta^\nu(\mu_1)}(\mu_2)$ is well defined, and its definition entails $\theta^\nu(\mu_1) + \theta^{\nu - \theta^\nu(\mu_1)}(\mu_2) \leq \nu$. The minimality property of $\theta^\nu(\mu_1 + \mu_2)$ now shows that

$$\theta^\nu(\mu_1 + \mu_2) \preceq_{\mathrm{st}} \theta^\nu(\mu_1) + \theta^{\nu - \theta^\nu(\mu_1)}(\mu_2). \tag{3.1}$$

On the other hand, the minimality properties of $\theta^\nu(\mu_1)$ and $\theta^\nu(\mu_1 + \mu_2)$ and direct arguments (omitted for brevity) imply that $\theta^\nu(\mu_1) \leq \theta^\nu(\mu_1 + \mu_2)$. The minimality property of $\theta^\nu(\mu_1)$ then states in particular that $\theta^\nu(\mu_1)$ is minimal in stochastic order among all sub-measures of $\theta^\nu(\mu_1 + \mu_2)$ with mass $\mu_1(\mathbb{R})$. As a consequence, we see that

$$\mu_2 \preceq_{\mathrm{st}} \theta^\nu(\mu_1 + \mu_2) - \theta^\nu(\mu_1).$$

Clearly also $\theta^\nu(\mu_1 + \mu_2) - \theta^\nu(\mu_1) \leq \nu - \theta^\nu(\mu_1)$, and so the minimality property of $\theta^{\nu - \theta^\nu(\mu_1)}(\mu_2)$ implies

$$\theta^{\nu - \theta^\nu(\mu_1)}(\mu_2) \preceq_{\mathrm{st}} \theta^\nu(\mu_1 + \mu_2) - \theta^\nu(\mu_1).$$

In view of (3.1), the claim follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We can now construct P_*.

Proof of Theorem 2.1. Noting that $\nu - \theta^\nu(\mu|_{(x,\infty)})$ is a nonnegative measure for fixed x, the function

$$F(x, y) := \big(\nu - \theta^\nu(\mu|_{(x,\infty)})\big)(-\infty, y]$$

is clearly nondecreasing and right-continuous in y. Moreover, Lemma 3.2 implies that

$$\theta^\nu(\mu|_{(x_1,\infty)}) - \theta^\nu(\mu|_{(x_2,\infty)}) = \theta^{\nu - \theta^\nu(\mu|_{(x_2,\infty)})}(\mu|_{(x_1,x_2]}) \geq 0, \quad x_1 \leq x_2. \quad (3.2)$$

The total mass of the right-hand side equals $\mu(x_1, x_2]$ and thus converges to zero as $x_2 \downarrow x_1$, showing that $x \mapsto F(x, y)$ is right-continuous. Relation (3.2) also implies that F is supermodular (or nondecreasing on $\mathbb{R}^2$): for $x_1 \leq x_2$ and $y_1 \leq y_2$,

$$\begin{aligned}
[F(x_2, y_2) - F(x_2, y_1)] &- [F(x_1, y_2) - F(x_1, y_1)] \\
&= \theta^\nu(\mu|_{(x_1,\infty)})(y_1, y_2] - \theta^\nu(\mu|_{(x_2,\infty)})(y_1, y_2] \\
&= \theta^{\nu - \theta^\nu(\mu|_{(x_2,\infty)})}(\mu|_{(x_1,x_2]})(y_1, y_2] \geq 0.
\end{aligned}$$

As F has the proper normalization, we conclude (e.g., [20, p. 27]) that F induces a unique probability measure P_* on $\mathcal{B}(\mathbb{R}^2)$. It remains to observe that $P_* \in \mathcal{D}(\mu, \nu)$. Indeed, the second marginal of P_* is clearly ν. The first marginal is equal to μ as for each x,

$$\lim_{y \to \infty} F(x, y) = \nu(\mathbb{R}) - \theta^\nu(\mu|_{(x,\infty)})(\mathbb{R}) = 1 - \mu((x, \infty)) = \mu((-\infty, x]).$$

12

Finally, P_* is directional since

$$P_*\big((x,\infty)\times(-\infty,x]\big) = \theta^\nu(\mu|_{(x,\infty)})(-\infty,x] = 0, \quad x\in\mathbb{R}$$

due to the fact that $\mu|_{(x,\infty)} \preceq_{\mathrm{st}} \theta^\nu(\mu|_{(x,\infty)})$ by the definition of $\theta^\nu(\cdot)$. $\qquad\square$

Remark 3.3. While we have defined P_* as mapping $\mu|_{(x,\infty)}$ to $\theta^\nu(\mu|_{(x,\infty)})$, it equivalently maps $\mu|_{[x,\infty)}$ to $\theta^\nu(\mu|_{[x,\infty)})$ for all $x\in\mathbb{R}$. This follows from Theorem 2.1 and Lemma 3.2.

We now turn the the equivalent characterizations in Theorem 2.2; here the most important tool is the notion of cyclical monotonicity in optimal transport (e.g., [17, 34]).

Proof of Theorem 2.2. Given two probability measures P, Q on $\mathbb{R}^2$ with the same marginals, it is known that the concordance order $F_P \leq F_Q$ is equivalent to $\int g\,dP \geq \int g\,dQ$ for all (suitably integrable) supermodular g; cf. [23, Theorem 3.8.2, p. 108]. The implication (i)$\Rightarrow$(ii) is a direct consequence of that fact, and (ii)$\Rightarrow$(iii) is trivial.

(iii)$\Rightarrow$(iv): Let g be Borel and (μ,ν)-integrable. We consider the (unconstrained) Monge–Kantorovich optimal transport problem on $\mathbb{R}\times\mathbb{R}$ with marginals (μ,ν) and cost function

$$c(x,y) = \begin{cases} -g(x,y), & (x,y)\in\mathbb{H}, \\ \infty, & \text{otherwise.} \end{cases}$$

Noting that $c(x,y) \geq \phi(x) + \psi(y)$ for some $\phi\in L^1(\mu)$ and $\psi\in L^1(\nu)$, it follows from [5, Theorem 1(a)] that any optimal transport P is concentrated on a Borel set $\Gamma\subseteq\mathbb{R}^2$ that is c-cyclically monotone. As no transport with finite cost charges the complement $\mathbb{H}^c$, we may replace Γ with $\Gamma\cap\mathbb{H}$ to ensure that $\Gamma\subseteq\mathbb{H}$. Cyclical monotonicity then states in particular that[2]

$$g(x,y) + g(x',y') \geq g(x,y') + g(x',y) \quad \text{for all} \quad (x,y),(x',y')\in\Gamma.$$

Thus, if g is strictly submodular, Γ cannot contain improvable pairs.

(iv)$\Rightarrow$(v): Suppose for contradiction that $P\neq P_*$. In view of Theorem 2.1, there exists $x\in\mathbb{R}$ such that P maps $\mu|_{(x,\infty)}$ to a measure $\nu'_x\neq\nu_x$,

---

[2]More generally, the monotonicity holds for cycles of finite length n; that is, $\sum_{i=1}^n g(x_i,y_i) \geq \sum_{i=1}^n g(x_i,y_{\pi(i)})$ for all $(x_1,y_1),\ldots,(x_n,y_n)\in\Gamma$ and permutations π of $\{1,\ldots,n\}$. The stated property corresponds to $n=2$.

and $\nu_x \preceq_{\mathrm{st}} \nu'_x$ by the minimality property of ν_x. It follows from Lemma 3.4 below that there exist $z > y \geq x$ such that

$$\nu_x((x,y]) > \nu'_x((x,z)) \quad \text{and} \quad \nu_x([y,z)) > \nu'_x([y,z)). \qquad (3.3)$$

Using also that $\mu((x,y]) \geq \nu_x((x,y])$ due to $\mu|_{(x,\infty)} \preceq_{\mathrm{st}} \nu_x$, we deduce

$$P((x,y] \times [z,\infty)) \geq \mu((x,y]) - \nu'_x((x,z)) \geq \nu_x((x,y]) - \nu'_x((x,z)) > 0.$$

By the constrained crossing property, this implies $P((-\infty,x]) \times [y,z)) = 0$ and thus

$$\nu([y,z)) = P(\mathbb{R} \times [y,z)) = P((x,\infty) \times [y,z)) = \nu'_x([y,z)),$$

contradicting (3.3).

   (v)$\Rightarrow$(i): Let $x,y \in \mathbb{R}$; we show $F_{P_*}(x,y) \leq F_Q(x,y)$ for $Q \in \mathcal{D}(\mu,\nu)$. As P_* and Q have the same second marginal, this is equivalent to

$$P_*((x,\infty) \times (-\infty,y]) \geq Q((x,\infty) \times (-\infty,y]).$$

Recalling ν_x from Theorem 2.1 and denoting by θ the measure that $\mu|_{(x,\infty)}$ is transported to by Q, the above can be stated as $\nu_x((-\infty,y]) \geq \theta((-\infty,y])$, and that clearly follows from the formula for F_{ν_x} in Theorem 2.1. $\qquad\square$

   The following was used in the preceding proof of (iv)$\Rightarrow$(v).

Lemma 3.4. *Given $\mu_1, \mu_2 \in \mathcal{M}$ with $\mu_1 \preceq_{\mathrm{st}} \mu_2$ and $\mu_1 \neq \mu_2$, there exist $z > y$ such that*

$$\mu_1((-\infty,y]) > \mu_2((-\infty,z)) \quad \text{and} \quad \mu_1([y,z)) > \mu_2([y,z)). \qquad (3.4)$$

Proof. Define two real functions

$$\phi^+(y) = \mu_1((-\infty,y]) - \mu_2((-\infty,y]), \quad \phi^-(y) = \mu_1((-\infty,y)) - \mu_2((-\infty,y)).$$

Then ϕ^+ and ϕ^- are right- and left-continuous, respectively, both are non-negative, and $\phi^+(y) = \phi^-(y)$ whenever $\mu_1(\{y\}) = \mu_2(\{y\})$. If $y \in \mathbb{R}$ satisfies

$$\phi^+(y) > 0 \quad \text{and} \quad \sup_{z \in (y,y+\varepsilon)} \phi^-(z) > \phi^-(y) \text{ for each } \varepsilon > 0, \qquad (3.5)$$

then (3.4) holds by choosing $z > y$ close enough to y. We argue by contradiction and suppose that there is no $y \in \mathbb{R}$ satisfying (3.5). Thus, if $\phi^+(y) > 0$, there exists $\varepsilon > 0$ such that $\phi^-(z) \leq \phi^-(y)$ for $z \in (y, y+\varepsilon)$. This implies

that the function ϕ^- has no upward jumps; i.e., $\Delta\phi^- \leq 0$. As $\mu_1 \neq \mu_2$, there exists $y_0 \in \mathbb{R}$ such that $\phi^-(y_0) > 0$. Since $\phi^-(y) \to 0$ as $y \downarrow -\infty$ and there are no upward jumps, there exists $y_1 < y_0$ such that $0 < \phi^-(y_1) < \phi^-(y_0)$. Let $y = \inf\{z > y_1 : \phi^-(z) > \phi^-(y_1)\}$. Then the left-continuity of ϕ^- implies $y < y_0$ and the absence of upward jumps implies $\phi^-(y) = \phi^-(y_1)$ as well as that y cannot be the location of a downward jump. Therefore, $\mu_1(\{y\}) = \mu_2(\{y\})$ and $\phi^+(y) = \phi^-(y) = \phi^-(y_1) > 0$. Finally, given $\varepsilon > 0$, we have $\phi^-(z) > \phi^-(y_1) = \phi^-(y)$ for some $z \in (y, y + \varepsilon)$ by the definition of y, so that y satisfies (3.5) and we have reached a contradiction. $\qquad\square$

Remark 3.5. The integrability condition in Theorem 2.2 can we weakened to the positive part g^+ being (μ, ν)-integrable and the negative part satisfying $\int g^- \, dP < \infty$ for some $P \in \mathcal{D}$, so that the value function is not trivial.

The final task of this section is to deduce the decomposition in Proposition 2.6 from Theorem 2.2.

Proof of Proposition 2.6. By Theorem 2.2, the optimal coupling $P_*(\mu', \nu')$ of μ', ν' is supported by a set Γ' with the constrained crossing property. Define $\Gamma = \Gamma' \cup \Delta$ where $\Delta = \{(x, x) : x \in \mathbb{R}\}$ is the diagonal in $\mathbb{R}^2$, then Γ again has the constrained crossing property. Set $P = \mathrm{Id}(\mu \wedge \nu) + P_*(\mu', \nu')$ and note $P \in \mathcal{D}(\mu, \nu)$. As Δ supports the identical coupling, P is supported by Γ and (iv)$\Rightarrow$(v) of Theorem 2.2 shows that $P = P_*(\mu, \nu)$. $\qquad\square$

4 Joint Distribution Function

As mentioned in Section 2, the formula for the joint distribution function F_* of P_* in Corollary 2.4 can be deduced from Theorem 2.2 (i) and [3, Theorem 6] which uses arguments from copula theory. Below, we sketch a direct derivation and some consequences.

Proof of Corollary 2.4. As P_* is directional, $y \leq x$ implies

$$F_*(x, y) = P_*((-\infty, x] \times (-\infty, y]) = \nu((-\infty, y]) = F_\nu(y),$$

so we can focus on $y > x$. Denote $c = \inf_{z \in [x,y]}(F_\mu(z) - F_\nu(z))$ and recall that X, Y are the coordinate projections. We first consider an arbitrary

15

$P \in \mathcal{D}(\mu, \nu)$. Then as $X \leq Y$ P-a.s., we have for $z \in [x, y]$ that

$$
\begin{aligned}
P(X \leq z, X > x) &\geq P(Y \leq z, X > x) \\
&= P(Y \leq z) - P(Y \leq z, X \leq x) \\
&= P(Y \leq z) - P(X \leq x) + P(Y > z, X \leq x) \\
&\geq P(Y \leq z) - P(X \leq x) + P(Y > y, X \leq x);
\end{aligned}
$$

that is, $F_\mu(z) - F_\mu(x) \geq F_\nu(z) - F_\mu(x) + P(Y > y, X \leq x)$. This shows $P(Y > y, X \leq x) \leq \inf_{z \in [x,y]}(F_\mu(z) - F_\nu(z)) = c$ and we conclude that

$$
F_P(x, y) = P(X \leq x) - P(Y > y, X \leq x) \geq F_\mu(x) - c. \qquad (4.1)
$$

In view of Theorem 2.2 we have $F_*(x, y) = \inf_{P \in \mathcal{D}(\mu, \nu)} F_P(x, y)$. Thus, to complete the proof, it suffices to show that some $P \in \mathcal{D}(\mu, \nu)$ attains equality in the above inequality.

Let $a = F_\mu(x)$ and $b = F_\nu(y)$; note that $0 \leq c \leq a \leq b + c \leq 1$. Let $U \sim \mathrm{Unif}[0, 1]$ and define a random variable V as

$$
V = \begin{cases}
U + b + c - a, & a - c < U \leq a, \\
U - c, & a < U \leq b + c, \\
U, & \text{otherwise.}
\end{cases}
$$

Then $V \sim \mathrm{Unif}[0, 1]$ like U, and thus $P := \mathrm{Law}(F_\mu^{-1}(U), F_\nu^{-1}(V))$ has marginals μ and ν, respectively. One checks by direct arguments that P is directional. Finally, if $U \in (a - c, a]$, then $F_\mu^{-1}(U) \leq F_\mu^{-1}(a) \leq x$ and $F_\nu^{-1}(V) \geq F_\nu^{-1}(b+)$, so that $P(X \leq x, Y > y) \geq P(U \in (a - c, a]) = c$. This shows that P attains equality in (4.1). $\qquad \square$

Remark 4.1. One can give a yet another proof of Corollary 2.4 based on Theorem 2.8 below, as may be intuitive given Figure 2.

Corollary 2.4 implies that P_* is continuous with respect to the marginals as stated in Corollary 2.5. The next example shows that this assertion may fail if the limiting marginals have atoms, a phenomenon caused by the directional constraint.

Example 4.2 (Discontinuity wrt. Marginals)**.** For $n \in \mathbb{N}$, let μ_n and ν_n be such that $\mu_n\{0\} = \mu_n\{1\} = 1/2$ and $\nu_n\{1 - 1/n\} = \nu_n\{2\} = 1/2$. Then $\mu_n \preceq_{\mathrm{st}} \nu_n$ and $\nu_n \overset{\mathrm{w}}{\to} \nu$ with $\nu\{1\} = \nu\{2\} = 1/2$, and $\mu_n \equiv \mu$ is constant. We see that $P_*(\mu_n, \nu_n)$ is the comonotone coupling, $P_*(\mu, \nu)$ is the antitone coupling, and $P_*(\mu_n, \nu_n) \overset{\mathrm{w}}{\nrightarrow} P_*(\mu, \nu)$.

Another consequence are simple bounds on F_*. A right-continuous function on $\mathbb{R}$ is *unimodal* if it is nondecreasing on $(-\infty, x_0)$ and nonincreasing on $[x_0, \infty)$ for some $x_0 \in \mathbb{R}$.

Corollary 4.3. *We have $H^\wedge \leq F_* \leq H^\vee$ for*

$$H^\wedge(x,y) = F_\nu(y) - [(F_\mu(y) - F_\mu(x)) \wedge (F_\nu(y) - F_\nu(x))]_+,$$
$$H^\vee(x,y) = F_\mu(x) \wedge F_\nu(y).$$

(i) $F_* = H^\wedge$ *if and only if $F = F_\mu - F_\nu$ is unimodal.*

(ii) $F_* = H^\vee$ *if and only if $\mathcal{D}(\mu, \nu)$ is a singleton. If, in addition, F is continuous, these conditions are further equivalent to $\mu = \nu$.*

Proof. The lower bound follows by considering $z \in \{x, y\}$ in (2.3). The upper bound follows directly from (2.3); alternately, it can also be obtained by noting that $H^\vee$ is the cdf of the comonotone coupling.

To see (i), note that by (2.3), $F_* = H^\wedge$ if and only if $\min_{z \in [x,y]} F(z) = F(x) \wedge F(y)$ for all $x < y$. This is equivalent to F being unimodal. Turning to (ii), we first recall from Theorem 2.2 (i) that P_* has the minimal cdf in $\mathcal{D}(\mu, \nu)$. On the other hand, $H^\vee$ is the cdf of the comonotone coupling, which is the maximal cdf among all couplings and in particular in $\mathcal{D}(\mu, \nu)$. Thus, $F_* = H^\vee$ if and only if all directional couplings have the same cdf, showing the first claim. Now let F be continuous and suppose for contradiction that $\mu \neq \nu$. In view of Proposition 2.6, we may assume that $\mu \wedge \nu = 0$. By Lemma 5.1, $\mu(I) > 0$ for the set I of strict increase of F. In particular, there exists $x \in I$, which implies that $F_\mu(x) > F_\nu(x)$ and $\mu((x, z]) > 0$ for any $z > x$. As P_* is the comonotone coupling, $\mu|_{(x,\infty)}$ is transported to $\nu|_{(y,\infty)}$ for some $y > x$. On the other hand, $\nu((x, y]) > 0$ due to $\mu((x, \infty)) = \nu((y, \infty)) < \nu((x, \infty))$, which by minimality implies that ν_x charges $(x, y]$, contradicting $\nu_x = \nu|_{(y,\infty)}$. Conversely, $\mu = \nu$ clearly implies that the identity is the only directional coupling. $\square$

Remark 4.4. (a) In view of Theorem 2.2 (i), the lower bound $F_* \geq H^\wedge$ is equivalent to the statement that $F_Q \geq H^\wedge$ for all $Q \in \mathcal{D}(\mu, \nu)$. The latter result was first obtained in [32]. See also [30] for a lower bound on a different coupling in a similar spirit. Both upper and lower bound were noted in [3], where it was also observed that the lower bound holds in the case of unimodality. The sharpness conditions are novel, to the best of our knowledge.

(b) The continuity assumption in (ii) is clearly important for the last conclusion: if μ is a Dirac mass, all couplings of μ and ν coincide and in particular $F_* = H^\vee$, but of course μ and ν need not be equal.

The following is a standard example satisfying the condition in Corollary 4.3 (i) and covering, for instance, two normal or exponential marginals in stochastic order. The appearance of an antitone coupling is a particular case of a phenomenon that will be discussed in detail in Section 6.1.

Example 4.5 (Single-crossing Densities). Suppose that μ and ν have densities f_μ and f_ν which cross exactly once; that is, there exists a point $x_0 \in \mathbb{R}$ such that $f_\mu(x) \geq f_\nu(x)$ for $x \leq x_0$ and $f_\mu(x) \leq f_\nu(x)$ for $x \geq x_0$. Then F is unimodal and hence $F_* = H^\wedge$. By Proposition 2.6 and the fact that the measures μ' and ν' (defined therein) are supported on disjoint sets, we see that $P_*(\mu, \nu)$ is the sum of an identity coupling $\mathrm{Id}(\mu \wedge \nu)$ and an antitone coupling $P_*(\mu', \nu')$.

5 The Transport Map

The aim of this section is to prove Theorem 2.8 on the optimal transport map T. The analysis rests on a specific Hahn decomposition that holds for arbitrary signed, diffuse measures on $\mathbb{R}$ and is provided in the first subsection. We then return to our transport problem, showing in Sections 5.2–5.3 that T induces a coupling with the constrained crossing property, and thus is optimal. Section 5.4 explains how marginals with atoms can be reduced to the continuous case by a simple transformation.

5.1 Sets of Increase and Decrease

Let $F : \mathbb{R} \to \mathbb{R}$ be a continuous function of bounded variation. We recall that the signed measure ρ associated to F admits a unique Jordan decomposition $\rho = \mu - \nu$ into mutually singular nonnegative measures, and then $\tau = \mu + \nu$ is the total variation measure of ρ. (In this section, μ and ν are arbitrary finite measures—not necessarily of the same mass or even $\mu \preceq_{\mathrm{st}} \nu$.) Similarly to ρ, the function F can be uniquely decomposed as $F = F_\mu - F_\nu$ into continuous nondecreasing functions that are mutually singular; that is, $V := F_\mu + F_\nu$ is the total variation of F. Disjoint Borel sets B_μ, B_ν form a *Hahn decomposition* for ρ (or F) if $\mu(B_\mu^c) = \nu(B_\nu^c) = 0$ and $\mu(B_\nu) = \nu(B_\mu) = 0$. In particular, τ is then carried by $B_\mu \cup B_\nu$.

If F is of class C^1, the sets $\{\partial F > 0\}$ and $\{\partial F < 0\}$ clearly form a Hahn decomposition. Moreover, the two sets are countable unions of intervals where F is monotone. Our purpose is to provide a similar Hahn decomposition for bounded variation functions—here the sets will merely be Borel, as

it is well known that a function can be absolutely continuous without being monotone on any interval (e.g., [15, p. 109, Exercise 41]).

Consider a function $F : \mathbb{R} \to \mathbb{R}$ and $x \in \mathbb{R}$. We call x a *point of strict increase* if there is a neighborhood of x in which $x_0 < x < x_1$ implies $F(x_0) < F(x) < F(x_1)$. The set of all such points is called the *set of strict increase* of F and denoted I_F. Points of strict decrease are defined analogously, and their set is denoted D_F.

Proposition 5.1. *Let* $F : \mathbb{R} \to \mathbb{R}$ *be a continuous function of bounded variation. The sets* I_F, D_F *of strict increase and decrease form a Hahn decomposition for* F.

Proof. Step 1. Let μ, ν, τ and F_μ, F_ν, V be as introduced above. Clearly μ, ν admit densities f_μ, f_ν with respect to τ, and these can be chosen to be indicator functions of complementary sets by the Hahn decomposition theorem. That is, $f_\mu(x), f_\nu(x) \in \{0, 1\}$ and $f_\mu(x) + f_\nu(x) = 1$ for all $x \in \mathbb{R}$.

Next, we claim that (with $z/0 := 0$, say) the limit

$$f(x) := \lim_{\varepsilon \to 0} \frac{\mu([x, x + \varepsilon])}{\tau([x, x + \varepsilon])}$$

exists for τ-a.e. $x \in \mathbb{R}$ and defines a version of the Radon–Nikodym derivative $d\mu/d\tau$—existence meaning particular that the limit is the same along any sequence $0 \neq \varepsilon_n \to 0$. Let V^{-1} be the right-continuous inverse of V. Then $F_\mu \circ V^{-1}$ is nondecreasing and $F_\mu \ll V$ implies that $\mu_{F_\mu \circ V^{-1}} \ll \lambda$, where $\mu_{F_\mu \circ V^{-1}}$ is the Lebesgue–Stieltjes measure of $F_\mu \circ V^{-1}$ and λ is the Lebesgue measure. By Lebesgue's differentiation theorem [15, Theorem 3.21, p. 98], $F_\mu \circ V^{-1}$ is λ-a.e. differentiable and the derivative ϕ defines a density $d\mu_{F_\mu \circ V^{-1}}/d\lambda$. (In fact, $F_\mu \circ V^{-1}$ is even Lipschitz.) That is, there exists a Lebesgue-nullset N_λ such that for $y \notin N_\lambda$ and $y' \to y$,

$$\phi(y) = \lim_{y' \to y} \frac{F_\mu(V^{-1}(y')) - F_\mu(V^{-1}(y))}{y' - y}$$

exists. Let $N = V^{-1}(N_\lambda)$; then $\tau(N) = 0$ as $\tau = \lambda \circ V$. For $x \notin N$ we have $y := V(x) \notin N_\lambda$. As V is continuous and $F_\mu = F_\mu \circ V^{-1} \circ V$, using the above with $y' = V(x')$ yields that

$$f(x) = \lim_{x' \to x} \frac{F_\mu(x') - F_\mu(x)}{V(x') - V(x)} = \lim_{y' \to y} \frac{F_\mu(V^{-1}(y')) - F_\mu(V^{-1}(y))}{y' - y}$$

exists and satisfies $f(x) = \phi(V(x))$. By the change-of-variable formula we see that f is a density of μ with respect to τ. It now follows that $f = f_\mu$

19

τ-a.e. As a result, for all x outside a τ-nullset and any sequence $\varepsilon_n \to 0$,

$$f(x) = \lim_n \frac{F_\mu(x + \varepsilon_n) - F_\mu(x)}{F_\mu(x + \varepsilon_n) - F_\mu(x) + F_\nu(x + \varepsilon_n) - F_\nu(x)} \in \{0, 1\}.$$

Step 2. Let $I = I_F$, $D = D_F$. The set $(I \cup D)^c$ consists of three types of points. First, the strict local minimum and maximum points; this subset is countable and hence a τ-nullset as V is continuous. Second, the points which are contained in an interval of constancy of F. There are countably many such intervals and each one is clearly a τ-nullset. Third, the points of oscillation: If $x \in (I \cup D)^c$ is not in an interval of constancy of F and if $0 \neq \varepsilon_n \to 0$, then for all n large we have either $F_\mu(x + \varepsilon_n) - F_\mu(x) \neq 0$ or $F_\nu(x + \varepsilon_n) - F_\nu(x) \neq 0$. If, in addition, x is not a strict local extremum, continuity implies that there exist $0 \neq \varepsilon_n \to 0$ such that $F(x) = F(x + \varepsilon_n)$; that is, $F_\mu(x + \varepsilon_n) - F_\mu(x) = F_\nu(x + \varepsilon_n) - F_\nu(x)$. Combining these two properties,

$$F_\mu(x + \varepsilon_n) - F_\mu(x) = F_\nu(x + \varepsilon_n) - F_\nu(x) \neq 0$$

for all n large. In particular,

$$\frac{F_\mu(x + \varepsilon_n) - F_\mu(x)}{F_\mu(x + \varepsilon_n) - F_\mu(x) + F_\nu(x + \varepsilon_n) - F_\nu(x)} \to \frac{1}{2}.$$

In view of Step 1, the set of all such x must be a τ-nullset. This completes the proof that $(I \cup D)^c$ is τ-null. It is easy to see that I and D are disjoint Borel sets. Noting also that $\{f = 1\} \subseteq I$ and $\{f = 0\} \subseteq D$, it follows that I, D form a Hahn decomposition. $\qquad \square$

5.2 Basic Properties of T

We return to our setting with given marginals $\mu \preceq_{\mathrm{st}} \nu$. Throughout this section we assume that $\mu \wedge \nu = 0$, or equivalently, that μ and ν are mutually singular. For simplicity of exposition, we first focus on the case of diffuse marginals μ and ν; the extension to measures with atoms is then simple and carried out in Section 5.4.

We consider $F = F_\mu - F_\nu$, a nonnegative continuous function of bounded variation with $F(-\infty) = F(\infty) = 0$, its graph G and its hypograph H,

$$G = \{(x, z) : z = F(x)\}, \qquad H = \{(x, z) : z \leq F(x)\}.$$

Recall from Theorem 2.8 that

$$T(x) = \inf\{y \geq x : (y, F(x)) \notin H\} \tag{5.1}$$

for $x \in \mathbb{R}$, with the convention $\inf \emptyset = \infty$. Let $I = I_F$ and $D = D_F$ be the sets of strict increase and decrease of F, respectively (see Section 5.1).

Lemma 5.2. *We have $\mu(I) = \nu(D) = 1$. The function T is upper semicontinuous and bimeasurable, it satisfies $(T(x), F(x)) \in G$ whenever $T(x) < \infty$, and $T(x) = \infty$ if and only if $F(x) = 0$.*

Proof. The statement $\mu(I) = \nu(D) = 1$ follows directly from Proposition 5.1 since I, D form a Hahn decomposition for F and $\mu \wedge \nu = 0$. As H is closed, T is upper semicontinuous. In view of $G = \partial H$, we also have $(T(x), F(x)) \in G$ whenever $T(x) < \infty$. Finally, $F(\infty) = 0$ implies that $T(x) = \infty$ if and only if $F(x) = 0$. To see that T is bimeasurable—i.e., also satisfies $T(\mathcal{B}(\mathbb{R})) \subseteq \mathcal{B}(\mathbb{R})$—it suffices to show that there are at most countably many points y whose preimage $T^{-1}(y)$ is uncountable; see for instance [25, Main Theorem]. Let y be such that $T^{-1}(y)$ contains more than one point. The construction of T shows that all elements $x \in T^{-1}(y)$, except possibly one, are local minima of F, and they have the common value $F(x) = F(T^{-1}(y))$. Any real function f only has countably many local minimum values $f(x)$ (because each local minimum is minimal within a rational interval, yielding an injection of the minimum values into $\mathbb{Q}^2$), so it suffices to show that for fixed y, $T^{-1}(y)$ contains at most countably many points x which also have the property that $T^{-1}(x)$ has several elements. If $x_0 < x$ is such that $T(x_0) = x$, it follows that $T^{-1}(x') = \emptyset$ for all $x' \in (x_0, x)$ with $F(x') = F(x)$. Thus we can associate with x an interval of positive length in which it is unique with the property in question, and that implies the claim. □

5.3 Marginals and Geometry of T

Lemma 5.3. *The map T transports μ to ν.*

Proof. We show that $\mu\{T \leq y\} = \nu((-\infty, y])$ for $y \in \mathbb{R}$. Define the continuous function

$$M(x) = F(x) - \min_{z \in [x, y]} F(z) \geq 0, \quad x \in (-\infty, y].$$

For $x \in I$ with $x \leq y$, $M(x) > 0$ is equivalent to the existence of $z \in (x, y]$ such that $F(z) < F(x)$, thus equivalent to $T(x) \leq y$. As μ is concentrated on I and T is directional, it follows that

$$\mu\{T \leq y\} = \mu\{x \in (-\infty, y] : T(x) \leq y\} = \mu((-\infty, y] \cap \{M > 0\}).$$

On the other hand, $M > 0$ on $D \cap (-\infty, y)$ and ν is concentrated on D, hence $\nu((-\infty, y]) = \nu((-\infty, y] \cap \{M > 0\})$ and it suffices to show that

$$(\mu - \nu)((-\infty, y] \cap \{M > 0\}) = 0.$$

Noting that $M(-\infty) = M(y) = 0$, we see that the set $(-\infty, y] \cap \{M > 0\}$ is open and thus is the union of countably many open intervals of the form $J = (a, b)$ with $M(a) = M(b) = 0$ and $M > 0$ on J. The last two facts and the definition of M imply that

$$F(a) = \min_{z \in [a, y]} F(z) = \min_{z \in [b, y]} F(z) = F(b)$$

and hence $(\mu - \nu)(J) = F(b) - F(a) = 0$, completing the proof. $\qquad\square$

Lemma 5.4. *If $x', x \in \mathbb{R}$ satisfy $x' < x \le T(x')$, then $T(x') \ge T(x)$. In particular, the graph of T has the constrained crossing property.*

Proof. Let $x', x \in \mathbb{R}$ satisfy $x' < x \le T(x')$. Note that $F(x) < F(x')$ would imply $(x, F(x')) \notin H$ and hence $T(x') < x$, a contradiction. Thus, $F(x') \le F(x)$. The semi-infinite rectangle $R = \{(a, b) : x \le a \le T(x), \ b \le F(x)\}$ is contained in the hypograph H, and similarly for the rectangle R' defined with x' instead of x (cf. Figure 4). To see that $T(x') \ge T(x)$, it suffices to
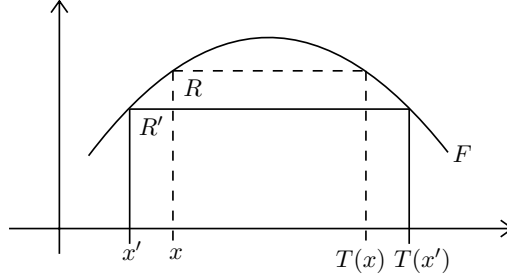

Figure 4: On the proof of Lemma 5.4

check that the segment $[x', T(x)] \times \{F(x')\}$ is contained in H. In view of $T(x') \ge x$, the first $(x - x')$-long part of the segment has that property, and the rest of the segment is contained in R and thus in H. $\qquad\square$

We now have all the ingredients for the main result on T.

Proof of Theorem 2.8. In view of Lemma 5.3 and $T(x) \ge x$, we have that $P := \mu \otimes \delta_T \in \mathcal{D}(\mu, \nu)$. Lemma 5.4 shows that P is supported on a set with the constrained crossing property and then Theorem 2.2 yields $P = P_*$. $\qquad\square$

5.4 Reduction of Atoms

Let $\mu \preceq_{\mathrm{st}} \nu$ satisfy $\mu \wedge \nu = 0$ as before, but consider the case where μ and ν may have atoms. We still write $F = F_\mu - F_\nu$, now this function is right-continuous rather than continuous. The idea is to reduce to the atomless case by a transformation which inserts an interval at the location of each atom, with its length corresponding to the atom's mass. The atom is then replaced by a uniform density (cf. Figure 5).
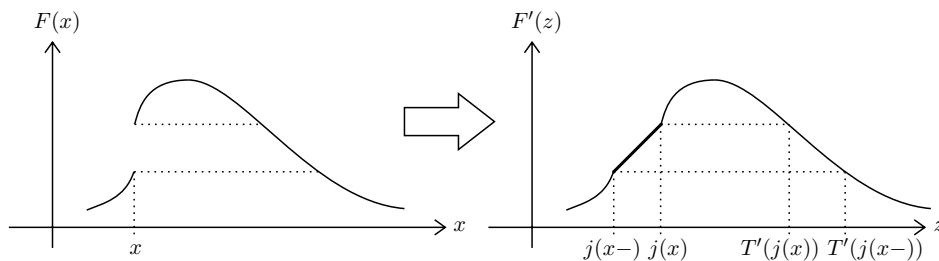

Figure 5: Transformation of an atom in μ at x.

Let $\tau = \mu + \nu$ be the total variation and let

$$j(x) = x + \sum_{y \le x} |F(y) - F(y-)|, \quad x \in \mathbb{R}$$

be the sum of the identity function and the cdf of the jump part of τ. Clearly j is strictly increasing and right-continuous; we denote its right-continuous inverse function by $j^{-1} : j(\mathbb{R}) \to \mathbb{R}$. Moreover, let

$$J_x = [j(x-), j(x)]$$

be the interval representing the jump of j at x. In particular, J_x is an interval of length $\tau(\{x\})$ and a singleton $\{j(x)\}$ if x is not an atom of μ or ν.

Define an auxiliary measure μ' on $\mathbb{R}$ through its cdf as follows: for $z \in j(\mathbb{R})$ we set $F_{\mu'}(z) = F_\mu(j^{-1}(z))$, whereas on the complement of $j(\mathbb{R})$ we define $F_{\mu'}(z)$ by linearly interpolating from its values on $j(\mathbb{R})$. In other words, μ' is defined by the two properties that $F_{\mu'}(j(x)) = F_\mu(x)$ for $x \in \mathbb{R}$ and if τ has an atom at x, then μ' is uniform on the interval J_x with total mass $\mu'(J_x) = \mu(\{x\})$. It follows that j is measure-preserving in the sense that $\mu'(j(B)) = \mu(B)$ for any $B \in \mathcal{B}(\mathbb{R})$. A second measure ν' is defined analogously from ν.

The construction implies that $\mu' \preceq_{\mathrm{st}} \nu'$ if and only if $\mu \preceq_{\mathrm{st}} \nu$, and $\mu' \wedge \nu' = 0$ if and only if $\mu \wedge \nu = 0$. Moreover, μ' and ν' are atomless.

23

Thus, Theorem 2.8 applies to $F' = F_{\mu'} - F_{\nu'}$ and yields a Monge map $T' := T(\mu', \nu')$. Reversing the transformation j, this map describes the desired coupling $P_*(\mu, \nu)$ as follows. (Of course, we can further apply Proposition 2.6 to produce a statement analogous to Corollary 2.9, covering the case of arbitrary marginals $\mu \preceq_{\mathrm{st}} \nu$ without imposing the condition $\mu \wedge \nu = 0$.)

Theorem 5.5. *Let $\mu \wedge \nu = 0$ and define $T' = T(\mu', \nu')$ as above. Then $P_*(\mu, \nu) = \mu \otimes \kappa$ for the stochastic kernel*

$$\kappa(x) = \begin{cases} \frac{1}{\mu(\{x\})} \nu(\,\cdot\, \cap j^{-1}(T'(J_x))) & \text{if } \mu(\{x\}) > 0, \\ \delta_{j^{-1}(T'(j(x)))} & \text{if } \mu(\{x\}) = 0. \end{cases}$$

In particular, κ is of Monge-type with transport map $T(x) = j^{-1}(T'(j(x)))$ whenever μ is atomless.

Proof. If $\mu(\{x\}) > 0$, then $\kappa(x)$ is well defined by Lemma 5.2 and has the proper normalization as $\mu(\{x\}) = \mu'(J_x) = \nu'(T'(J_x))$. Among the points x with $\mu(\{x\}) = 0$, it suffices to consider those with $j(x) \in I'$, the set of points of strict increase of F'—indeed, as j is measure-preserving, it follows from Lemma 5.2 that the complementary set is μ-null. For $j(x) \in I'$, Lemma 5.2 shows that $\kappa(x) = \delta_{j^{-1}(T'(j(x)))}$ is well defined. As T' defines a coupling in $\mathcal{D}(\mu', \nu')$ and j is strictly monotone and measure-preserving, it follows that κ defines a coupling in $\mathcal{D}(\mu, \nu)$. Moreover, we know that the graph Γ' of T' has the constrained crossing property (Lemma 5.4). The strictly monotone transform j does not invalidate that property (Corollary 2.3), hence $\Gamma := j^{-1}(\Gamma')$ has the same property, and Γ carries $\mu \otimes \kappa$, as noted above. We conclude by Theorem 2.2. $\square$

We note that P_* can still be of Monge-type when μ has atoms: by Theorem 5.5, that happens precisely if $j^{-1}(T'(J_x))$ is a singleton whenever $\mu(\{x\}) > 0$. This requires very specific atoms in ν, as κ must transport each upward jump point of F to a downward jump point, and moreover the downward jump must have at least the same size as the upward jump. One example of such a match-up is given in (a) below.

Example 5.6 (Empirical Distributions). Consider marginals $\mu = \frac{1}{n_\mu} \sum_{i=1}^{n_\mu} \delta_{x_i}$ and $\nu = \frac{1}{n_\nu} \sum_{i=1}^{n_\nu} \delta_{y_i}$ in stochastic order.

(a) If the x_i are distinct and $n_\mu = n_\nu =: n$, then P_* is Monge and the transport map T is as constructed in the introduction: considering the destinations $S_1 = \{y_1, \ldots, y_n\}$ as a multi-set (i.e., distinguishing the y_i even if they have the same value), we iterate for $k = 1, \ldots, n$:

(i) $T(x_k) := \min\{y \in S_k : y \geq x_k\}$,

(ii) $S_{k+1} = S_k \setminus \{T(x_k)\}$.

(b) The case $n_\mu \neq n_\nu$ is natural when μ and ν are empirical distributions of observed data—in the study of treatment effects, data are often not observed in pairs and hence the two marginals may not have the same number of observations; see Section 1. The above algorithm immediately extends to the case where $n_\mu = mn_\nu$ for an integer m, by redefining the y_i. If n_μ and n_ν are arbitrary, and/or the atoms have possibly different, rational weights, we can still write the marginals in the form $\mu = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ and $\nu = \frac{1}{n}\sum_{i=1}^n \delta_{y_i}$ after by choosing a suitable n, now with the x_i not necessarily distinct. The principle of the above algorithm to find P_* still applies, but when several x_i are at the same location, it will typically deliver a randomized coupling since an atoms of μ may be mapped into multiple atoms of ν.

6 Further Properties

6.1 Antitone Decomposition

As seen in Example 4.5, P_* is the sum of an identity coupling and an antitone coupling when the marginal densities satisfy a single-crossing condition. In this section, we analyze to which extent such a decomposition generalizes to other marginals. The first result (together with Proposition 2.6) shows that P_* is always the sum of an identity coupling and countably many antitone couplings. We will see that in certain cases, the marginal measures for those antitone coupling are simply restrictions of μ and ν to specific intervals, as in the aforementioned example. In general, however, the decomposition remains more implicit as the marginal measures do not admit such a simple description.

Proposition 6.1. *Let $\mu \preceq_{\mathrm{st}} \nu$ satisfy $\mu \wedge \nu = 0$. Then P_* is the sum of countably many antitone couplings.*

Proof. In view of Theorem 5.5, we may assume that μ, ν are atomless. For any continuous, nonnegative, nonconstant function G of finite variation with $G(-\infty) = G(\infty) = 0$, we define $x_G = \min(\arg\max G)$ as the smallest global maximum point and set

$$G'(x) = \min_{y \in [x, x_G]} G(y)\mathbf{1}_{\{x \leq x_G\}} + \min_{y \in [x_G, x]} G(y)\mathbf{1}_{\{x > x_G\}},$$

whereas if $G \equiv 0$, we use $x_G := -\infty$ instead. Note that G' is continuous, increasing on $(-\infty, x_G]$ and decreasing on $[x_G, \infty)$, with $0 \leq G' \leq G$ and $\max G' = \max G$. Thus G' can be decomposed as $G' = F_{\mu'} - F_{\nu'}$ where the singular measures μ' and ν' can be coupled by a directional antitone coupling. This coupling, while equal to $P_*(\mu', \nu')$, will be denoted by $P(G)$ for brevity. Moreover, $\mu' \leq \mu$ and $\nu' \leq \nu$. Finally, the total variation $V(G') = (\mu' + \nu')(\mathbb{R})$ satisfies $V(G') \geq 2 \max G' = 2 \max G$.

Define $F_1 := F$ and

$$F_{k+1} := F_k - F_k', \quad k \geq 1.$$

Using the above notation, $P(F_k)$ is the directional antitone coupling between the singular measures μ_k', ν_k' forming a decomposition for F_k'.

To see that $F = \sum_k F_k'$, note that $V(F_k') \to 0$ as $\sum_k V(F_k') \leq V(F) = 2$. On the other hand, $V(F_k') \geq 2 \max F_k$, so that $\max F_k \to 0$; that is, F_k uniformly decreases to zero and in particular $F = \sum_k F_k'$. This shows that $\sum_k P(F_k)$ is a coupling of μ and ν. Clearly this coupling is directional, and thus equal to $P_*(\mu, \nu)$ by Theorem 2.2 if it satisfies the constrained crossing property. To verify the latter, let x be a point of strict increase of F_k and suppose that the transport map T_k of $P(F_k)$ maps x to y. Then $F_k(x) = F_k(y)$ and $F_k(z) \geq F_k(x) > 0$ for all $z \in [x, y]$. It follows for any $j < n$ that $F_j'(z) < F_j(z)$ for all $z \in [x, y]$, which in turn implies that F_j' is constant over the interval $[x, y]$. In other words, the couplings $P(F_j)$ for $j < k$ cannot transport any mass into the interval or out of the interval. This shows the constrained crossing property, and in addition that the marginals μ_j' (resp. ν_j') of $P(F_j)$, $j \leq k$ are supported on disjoint sets which are finite unions of intervals. $\qquad\square$

In particular cases, we can obtain the antitone couplings in P_* explicitly as antitone couplings between disjoint intervals.

Example 6.2 (Multiple-crossing Densities)**.** Assume that μ and ν are atomless and that $F = F_\mu - F_\nu$ is piecewise monotone (with finitely many pieces). Then by inspecting the proof of Proposition 6.1, we see that P_* is the sum of the identical coupling of $\mu \wedge \nu$ and finitely many antitone couplings between pairs of *disjoint* intervals.

As an important special case extending Example 4.5, suppose that μ and ν have continuous densities that cross finitely many times. Then $F = F_\mu - F_\nu = F_{\mu - \mu \wedge \nu} - F_{\nu - \mu \wedge \nu}$ is piecewise monotone and the optimal coupling between $\mu - \mu \wedge \nu$ and $\nu - \mu \wedge \nu$ is the sum of finitely many antitone couplings between disjoint intervals.

In contrast to the above example, the following shows that a decomposition into antitone couplings between intervals is not possible in general.

Example 6.3 (Absence of Antitone Intervals)**.** Let μ be the Cantor distribution on $[0,1]$ and ν be uniform on $[0,2]$. Clearly $\mu \wedge \nu = 0$. We first verify that $\mu \preceq_{\mathrm{st}} \nu$, or equivalently $\mathcal{D}(\mu,\nu) \neq \emptyset$. Each element $x \in C$ can be represented in base 3 as $x = 2\sum_{n=1}^{\infty} x_n 3^{-n}$ where $x_n \in \{0,1\}$. The comonotone transport T_C given by $T_C(x) = 2\sum_{n=1}^{\infty} x_n 2^{-n}$ is directional and transports μ to ν. Hence, $\mu \preceq_{\mathrm{st}} \nu$.

Next, we show that $P_* \in \mathcal{D}(\mu,\nu)$ does not contain any antitone couplings between intervals. Assume for contradiction that there exists an interval $[a,b] \subseteq [0,1]$ such that $\mu([a,b]) > 0$ and $T|_{[a,b]}$ is the antitone mapping between $\mu|_{[a,b]}$ and its image. This implies that there exists c such that $\mu([a,c]) > 0$ and T transports $\mu|_{[a,c]}$ to a distribution supported by (c,∞). However, by Theorem 2.1, T transports $\mu|_{(a,\infty)}$ to a distribution ν_a whose minimality property together with $\nu([a,c]) > 0$ imply that ν_a charges $[a,c]$, a contradiction.

6.2 Optimality as Unconstrained Transport

The optimal directional coupling P_* is also the optimizer for certain classical transport problems (unconstrained and with finite cost function) where the constraint is "not binding," although only for specific marginals. We confine ourselves to giving one example. Consider $\mu \preceq_{\mathrm{st}} \nu$ and the transport problem

$$\inf_P \int c(|y-x|)\, P(dx,dy) \tag{6.1}$$

over all couplings P of μ and ν. Suppose that $c : \mathbb{R} \to \mathbb{R}_+$ is increasing and concave, so that $c(|y-x|)$ is supermodular on $\mathbb{H}$ but (typically) not on $\mathbb{R}^2$.

Proposition 6.4. *If $F = F_\mu - F_\nu$ is unimodal, then $P_*(\mu,\nu)$ is an optimal coupling for the unconstrained problem* (6.1)*. If c is strictly concave, the optimizer is unique.*

This follows from the general results stated in [17, Part II]. A direct argument is sketched below.

Proof. We know from Theorem 2.2 that P_* is optimal among all directional couplings. To rule out that a non-directional coupling has a smaller cost, the key observation is that if P is an optimizer, it is concentrated on a c-cyclically monotone set Γ, which implies that Γ cannot contain pairs $(x,y), (x',y')$ with

27

$y < x$ and either (i) $x' \in [y, x)$ and $y' \geq y$ or (ii) $y' \in [x, y)$ and $x' \leq x$. Together with the unimodality condition, this can be seen to imply the result. We omit the details in the interest of brevity. □

The unimodality condition in Proposition 6.4 is crucial; e.g., the assertion fails for $\mu = \frac{1}{2}(\delta_0 + \delta_{13})$ and $\nu = \frac{1}{2}(\delta_{12} + \delta_{25})$ with cost function $\sqrt{|y - x|}$.

6.3 Other Constraints

The directional constraint $Y \geq X$ naturally generalizes to $Y \geq X + D$ for a measurable function $D : \mathbb{R} \to \mathbb{R}$ such that $x \mapsto x + D(x)$ is strictly increasing. For instance, if $D \equiv d$ is constant, this means that the transport must travel as least a distance d to the right (or at most distance $|d|$ to the left, if $d < 0$). While $Y \geq X$ is equivalent to $P(\mathbb{H}) = 1$, the generalized constraint is expressed as $P(\mathbb{D}) = 1$ for the epigraph $\mathbb{D}$ of $x \mapsto x + D(x)$. We denote by $\mathcal{D}_D(\mu, \nu)$ the set of all such couplings P of μ, ν.

The construction of P_* naturally extends to this constraint. Indeed, let $Z(x) = x + D(x)$ and consider arbitrary distributions μ and ν on $\mathbb{R}$. We define the transformed marginal $\mu' = \mu \circ Z^{-1}$ and define $\mu \preceq_D \nu$ to mean that $\mu' \preceq_{\text{st}} \nu$. Then $\mu \preceq_D \nu$ if and only if $\mathcal{D}_D(\mu, \nu) \neq \emptyset$, and more generally, the transformation Z induces a bijection between $\mathcal{D}_D(\mu, \nu)$ and the set $\mathcal{D}(\mu', \nu)$ of directional couplings between μ' and ν. If we define the analogues of the constrained crossing property, constrained submodularity, etc., for $\mathbb{D}$, this bijection preserves the crossing/optimality properties and we find that

$$P_*^D(\mu, \nu) := P_*(\mu', \nu) \circ (Z, \text{Id})$$

has the properties analogous to the optimal directional coupling for the constraint $\mathbb{D}$. We omit the details in the interest of brevity.

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 A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation. In *Optimal transportation and applications (Martina Franca, 2001)*, volume 1813 of *Lecture Notes in Math.*, pages 123–160. Springer, Berlin, 2003.

[3] S. Arnold, I. Molchanov, and J. F. Ziegel. Bivariate distributions with ordered marginals. *J. Multivariate Anal.*, 177:104585, 2020.

[4] P. M. Aronow, D. P. Green, and D. K. K. Lee. Sharp bounds on the variance in randomized experiments. *Ann. Statist.*, 42(3):850–871, 2014.

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

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

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

[10] G. Carlier, L. De Pascale, and F. Santambrogio. A strategy for non-strictly convex transport costs and the example of $\|X - Y\|^p$ in $\mathbb{R}^2$. *Commun. Math. Sci.*, 8(4):931–941, 2010.

[11] P. Chen, F. Jiang, and X. Yang. Two dimensional optimal transportation problem for a distance cost with a convex constraint. *ESAIM Control Optim. Calc. Var.*, 19(4):1064–1075, 2013.

[12] 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.

[13] I. Ekren and H. M. Soner. Constrained optimal transport. *Arch. Ration. Mech. Anal.*, 227(3):929–965, 2018.

[14] Y. Fan and S. S. Park. Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, 26(3):931–951, 2010.

[15] G. B. Folland. *Real analysis*. Pure and Applied Mathematics. John Wiley & Sons, New York, second edition, 1999.

[16] 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.

[17] W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

[18] 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.

[19] C. Jimenez and F. Santambrogio. Optimal transportation for a quadratic cost with convex constraints and applications. *J. Math. Pures Appl. (9)*, 98(1):103–113, 2012.

[20] H. Joe. *Dependence modeling with copulas*, volume 134 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL, 2015.

[21] H. G. Kellerer. Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete*, 67(4):399–432, 1984.

[22] C. F. Manski. Monotone treatment response. *Econometrica*, 65(6):1311–1334, 1997.

[23] A. Müller and D. Stoyan. *Comparison methods for stochastic models and risks*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester,

2002.

[24] M. Nutz and F. Stebegg. Canonical supermartingale couplings. *Ann. Probab*, 46(6):3351–3398, 2018.

[25] R. Purves. Bimeasurable functions. *Fund. Math.*, 58:149–157, 1966.

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

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

[28] L. C. G. Rogers. Fastest coupling of random walks. *J. London Math. Soc. (2)*, 60(2):630–640, 1999.

[29] F. Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015.

[30] S. K. Sarkar and W. Smith. Probability inequalities for ordered MTP_2 random variables. *Sankhyā Ser. A*, 48(2):119–135, 1986.

[31] W. Schachermayer and J. Teichmann. Characterization of optimal transport plans for the Monge-Kantorovich problem. *Proc. Amer. Math. Soc.*, 137(2):519–529, 2009.

[32] W. Smith. Inequalities for bivariate distributions with $X \leq Y$ and marginals given. *Comm. Statist. A—Theory Methods*, 12(12):1371–1379, 1983.

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

[34] C. Villani. *Optimal transport, old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2009.