Weak transport for non-convex costs and model-independence in a fixed-income market

ORIGINAL ARTICLE

MATHEMATICAL FINANCE An International Journal of Mathematics, Statistics and Financial Economics WILEY

Weak transport for non-convex costs and model-independence in a fixed-income market

Beatrice Acciaio[1] | **Mathias Beiglböck[2]** | **Gudmund Pammer[1]**

[1] ETH Zurich, Zurich, Switzerland

[2] University of Vienna, Vienna, Austria

Correspondence
Beatrice Acciaio, ETH Zurich, Zurich, Switzerland.
Email: beatrice.acciaio@math.ethz.ch

Abstract

We consider a model-independent pricing problem in a fixed-income market and show that it leads to a weak optimal transport problem as introduced by Gozlan et al. We use this to characterize the extremal models for the pricing of caplets on the spot rate and to establish a first robust super-replication result that is applicable to fixed-income markets. Notably, the weak transport problem exhibits a cost function which is non-convex and thus not covered by the standard assumptions of the theory. In an independent section, we establish that weak transport problems for general costs can be reduced to equivalent problems that do satisfy the convexity assumption, extending the scope of weak transport theory. This part could be of its own interest independent of our financial application, and is accessible to readers who are not familiar with mathematical finance notation.

KEYWORDS
fixed-income markets, robust pricing and hedging, weak transport problem

1 | INTRODUCTION

Prompted by Hobson's seminal paper (Hobson, 1998), the field of model-independent finance has seen a steep development. Typically the framework consists in a market where some assets are dynamically traded, while some derivatives are statically traded at time zero. Then,

without imposing any model or underlying probability measure, one seeks robust pricing bounds for other derivatives. Usually the payoffs are already expressed in discounted terms, and the pricing problem corresponds to looking for market-compatible martingale measures, that is, probability measures on the path space such that the discounted asset prices—taken to be the canonical processes—are martingales, and reproduce the observed marked prices.

A widely used assumption is the availability, for static trading, of call options with maturity T written on an asset S, for all strikes K, which determines the distribution of S_T under the pricing measure. However, this reasoning crucially relies on the existence of a deterministic numéraire for discounting. This point is crucial for the use of tools from the Optimal Transport and Skorokhod embedding, see Hobson (2011), Obłój (2004), Beiglböck et al. (2013), Galichon et al. (2014), Beiglböck and Juillet (2016), Dolinsky and Soner (2014), Campi et al. (2017), Beiglböck et al. (2017), Cox et al. (2019), Cheridito et al. (2021), among many others.

In the present paper, we do not assume the existence of a deterministic bank account, or of a bank account at all. Instead, we consider a fixed-income market, where bonds are dynamically traded, and some options on them are statically traded. A similar setting is considered by Acciaio and Grbac (2020), who assume finitely many call options with possibly different maturities being traded on a bond, and investigate market consistency with absence of arbitrage. This analysis follows the spirit of Davis and Hobson (2007), and displays in a simple market the different characteristics of a robust framework when stochastic discounting is allowed. On the other hand, the focus of the present paper is on robust pricing. We will use bonds as numéraires, thus the pricing measures will be the so-called forward measures, a notion that we recall in the next paragraph.

For $T > 0$, we refer to a (zero coupon) bond with maturity T as a T-bond, and denote its price at time $t \leq T$ by $p(t, T)$. A T-forward measure $\mathbb{Q}_T$ is a probability measure on $\mathcal{F}_T$ (where $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \leq T})$ is some abstract filtered space) such that every traded asset expressed in units of the T-bond is a martingale. The pricing formula, at time s, for a claim Φ with maturity t, for $s \leq t \leq T$, is then given by

$$\frac{V_s(\Phi)}{p(s,T)} = \mathbb{E}^{\mathbb{Q}_T}\left[\frac{\Phi}{p(t,T)} \middle| \mathcal{F}_s \right]. \tag{1}$$

In particular, having the prices at time 0 of all call options with maturity T on a given asset, identifies the distribution of that asset at time T under the T-forward measure.

1.1 | Robust pricing setting

Setup 1.1. *We consider maturities $T_0 = 0 < T_1 < T_2 < T_3$, and let T_2- and T_3-bonds be (dynamically) traded in the market.[1] We also let call options with maturity T_1 on the T_2-bond, and call options with maturity T_2 on the T_3-bond be (statically) traded at time zero, for every strike K, and denote the respective prices by $C(T_1, T_2, K)$ and $C(T_2, T_3, K)$.*

A main contribution of this paper is to provide a robust superreplication theorem that applies to general derivatives written on $p(T_1, T_2), p(T_1, T_3)$, that is, we will consider payoffs of the form

$$\Phi(p(T_1, T_2), p(T_1, T_3)). \tag{2}$$

Example 1.2. The following example fits the above setup well. Let us recall that, for $0 \leq S < T$, and a simple spot rate $F(S, T)$ prevailing at S, a caplet with reset date S, settlement date T, and

strike rate K, yields the following cash-flow at time T:

$$(T - S)(F(S,T) - K)^+.$$

Note that this corresponds to the following cash-flow at time S:

$$(1 + (T - S)K)\left(\frac{1}{1 + (T - S)K} - p(S,T)\right)^+,$$

that is, to holding $1 + (T - S)K$ puts with maturity S and strike $1/(1 + (T - S)K)$ written on the T-bond; see (Filipovic, 2009, Section 2.6.1).

We now invoke the fact that 6 month caplets (i.e., with $T - S = 6$ months) are much more liquid than the 12 month ones. As common in robust finance literature, we stretch this fact and actually assume that 6 month caplets are liquidly traded for every strike rate. In particular, for $T_{i+1} - T_i = 6$ months, $i = 1, 2$, this means availability of all caplets with reset date T_i and settlement date T_{i+1}, for $i = 1, 2$. From the above, this corresponds to knowing the prices of puts with maturity T_i written on the T_{i+1}-bond for every strike, for $i = 1, 2$. This is captured by Setup 1.1, and the analysis of the current paper allows to obtain lower and upper pricing bounds for options of the form (2), thus in particular of the (less liquid) 12 month caplets with reset date T_1 and settlement date T_3, that is, $\Phi = (1 + (T_3 - T_1)K)(\frac{1}{1 + (T_3 - T_1)K} - p(T_1, T_3))^+$. In fact, in this case we can also identify the extremal models, as well as the corresponding sub- and superreplication strategies. (see Section 2.2).

1.2 | Robust superreplication theorem

In our setup it is most convenient to take the T_2-bond as a numéraire. In particular, we consider the *discounted* processes

$$X_i := \frac{p(T_i, T_1)}{p(T_i, T_2)}, \quad i = 0, 1, \quad Y_i := \frac{p(T_i, T_3)}{p(T_i, T_2)}, \quad i = 0, 1, 2. \tag{3}$$

For the convenience of the reader, we detail in Section 2.1 below how to switch between undiscounted quantities and quantities expressed in T_2-bond units. In the present introductory section we will consistently refer to quantities expressed in T_2-bonds. The assumption in Setup 1.1 then asserts that options on X_1 and Y_2 are liquidly traded. In the spirit of Breeden and Litzenberger (1978), this amounts to the identification of probabilities μ, ν on $\mathbb{R}_+$ with finite first moments, such that for derivatives φ, ψ

$$\text{price}(\varphi(X_1)) = \int \varphi \, d\mu, \quad \text{price}(\psi(Y_2)) = \int \psi \, d\nu. \tag{4}$$

In view of the classical (in the sense of *non-robust*) theory of fixed-income markets, we expect pricing functionals to arise from T_2-forward measures. A generic T_2-forward measure is compatible with the information given by the market if (compare (1))

$$\text{law}_{\mathbb{Q}_{T_2}}(X_1) \sim \mu, \quad \text{law}_{\mathbb{Q}_{T_2}}(Y_2) \sim \nu. \tag{5}$$

We write $Q(\mu, \nu)$ for the class of all setups $(\Omega, \mathcal{F}, (\mathcal{F}_{T_i})_{i=0}^2, \mathbb{Q}_{T_2})$, $p(T_i, T_j), i \leq j \leq 3, i \leq 2$, satisfying the *marginal constraints* (5) and the *martingale constraints*

$$(X_i)_{i=0,1}, \quad (Y_i)_{i=0,1,2} \quad \text{are } (\mathcal{F}_{T_i})\text{-martingales.} \tag{6}$$

(Note that we make no further assumptions on the underlying stochastic basis.) We will slightly abuse notation in that we write $\mathbb{Q}_{T_2} \in Q(\mu, \nu)$ when we really mean that we run over all such setups. Writing $\hat{\Phi}(X_1, Y_2)$ for the payoff (2) in terms of T_2-bonds (see Section 2.1), we arrive at the primal optimization problem

$$\hat{P}_F := \sup_{\mathbb{Q}_{T_2} \in Q(\mu, \nu)} \mathbb{E}^{\mathbb{Q}_{T_2}}[\hat{\Phi}(X_1, Y_1)]. \tag{7}$$

We stress that problem (7) differs significantly from a usual martingale optimal transport problem in that our objective functional is written on Y_1, while the marginal constraint concerns Y_2. We also note that, in contrast to most problems considered in robust finance, it is delicate to pass from (7) to a problem on a canonical setup, cf. Proposition 1.5.

Switching to the dual problem of (7), one is allowed to trade statically in vanilla payoffs φ, ψ written on X_1, Y_2 and to trade dynamically in $(Y_i)_{i=1,2}$. Therefore we consider superhedges of the form[2]

$$\hat{\Phi}(X_1, Y_1) \leq \varphi(X_1) + \psi(Y_2) + \Delta(Y_1)(Y_2 - Y_1). \tag{8}$$

Since our goal is to obtain a robust superhedge, we shall require that (8) holds for arbitrary $(X_1, Y_1, Y_2) = (x_1, y_1, y_2) \in \mathbb{R}_+^3$. According to (4), the costs for this strategy (in discounted terms) amount to $\int \varphi \, d\mu + \int \psi \, d\nu$. Our main superreplication theorem is then:

Theorem 1.3. *Assume that μ, ν are probabilities on $\mathbb{R}_+$ with finite r-th moments, for some $r \geq 1$. Assume that $\hat{\Phi}(x, y)$ is upper semicontinuous and bounded from above by a multiple of $|x|^r + |y|^r$. Then we have*

$$\hat{P}_F = \inf \left\{ \int \varphi \, d\mu + \int \psi \, d\nu : \hat{\Phi}(x_1, y_1) \leq \varphi(x_1) + \psi(y_2) + \Delta(y_1)(y_2 - y_1) \text{ for all } x_1, y_1, y_2 \geq 0 \right\},$$

where the infimum is taken over continuous functions φ and ψ that are bounded in absolute terms by a multiple of $x \mapsto (1 + |x|)^r$ and Δ is an arbitrary function[3].

Remark 1.4. It is worth noticing that the T_3-bond plays no special role for our results, in the sense that, by replacing it with a generic asset S, we would get the same superreplication result. That is, if in Setup 1.1, instead of the T_3-bond, we assume that a generic asset S is dynamically traded and T_2-calls on it statically traded, then Theorem 1.3 holds true for any derivative of the form $\Phi(p(T_1, T_2), S_{T_1})$. In this case, from market prices we deduce the distributions under $\mathbb{Q}_{T_2}$ of $X_1 = 1/p(T_1, T_2)$ and of S_{T_2}.

Robust versions of the superhedging duality/FTAP have received particular attention in robust finance, see for example, Hobson and Neuberger (2012), Acciaio et al. (2016), Bouchard and Nutz (2015), Burzoni et al. (2017), Cheridito et al. (2017), among many others. On the other hand,

ambiguity in fixed-income markets has been recently studied by Hölzermann (2018), Fadina and Schmidt (2019), Hölzermann and Lin (2019), Hölzermann (2020), that assume specific models (Heath-Jarrow-Morton, Hull-White) and consider a non-dominated set of probability measures. On the other hand in the current paper, we consider a robust framework for fixed-income markets, without assuming any model nor any set of probability measures, and provide first superhedging duality results and characterization of extremal models.

1.3 | Weak transport formulation

Our proof of Theorem 1.3 is based on new results in the theory of weak optimal transport (WOT) (see Section 3). Specifically, we will show that WOT problems for general costs can be reduced to the much better understood case of WOT problems that satisfy convexity constraints. This reduction enables us to rewrite the pricing problem (7) as a weak transport problem for barycentric costs. To formulate this, we write $\Pi(\mu, \nu)$ for the set of all couplings of the probabilities μ, ν, and for $\pi \in \Pi(\mu, \nu)$ we denote its disintegration w.r.t. μ by $(\pi_x)_x$. For a measure p on $\mathbb{R}$ that has a finite first moment we write $b(p)$ for its barycenter. Using these notions, it is possible to rewrite (7) as an optimization problem over measures on $\mathbb{R}^2$.

Proposition 1.5. *Assume that μ, ν are probabilities on $\mathbb{R}_+$ with finite r-th moment, for some $r \geq 1$, and that μ is continuous. Assume that $\hat{\Phi}(x, y)$ is continuous and has at most growth of order $|x|^r + |y|^r$. Then*

$$\hat{P}_F = \sup_{\pi \in \Pi(\mu, \nu)} \int \hat{\Phi}(x, b(\pi_x)) \, \mu(dx). \tag{9}$$

The basic idea behind the reformulation in (9) is to reinterpret the martingale property of (Y_1, Y_2) in terms of barycenters of the disintegration of a transport plan between the prescribed marginals.

The formulation of the robust pricing problem given in Proposition 1.5 plays a key role in our explicit solution of the weak transport problem in the case where Φ is a put or call option (see Section 2.2). (9) can also be used to formulate our superreplication theorem 1.3 in the language of optimal transport. We state here the corresponding theorem to highlight the connection with the classical Monge-Kantorovich duality:

Theorem 1.6. *Assume that μ is a continuous probability on a Polish space $\mathcal{X}$ and ν a probability on $\mathbb{R}^d$, $d \in \mathbb{N}$, with finite r-th moments, for some $r \geq 1$. Assume that $c : \mathcal{X} \times \mathbb{R}^d \to \mathbb{R}$ is continuous and has at most growth of order $d_{\mathcal{X}}(x, x_0)^r + |y|^r$ (where x_0 is an arbitrary fixed point in $\mathcal{X}$ and $d_{\mathcal{X}}$ is a compatible, complete metric on $\mathcal{X}$). Then we have*

$$\sup_{\pi \in \Pi(\mu, \nu)} \int c(x, b(\pi_x)) \, \mu(dx) \tag{10}$$

$$= \inf \left\{ \int \varphi \, d\mu + \int \psi \, d\nu : c(x, y_1) \leq \varphi(x) + \psi(y_2) + \Delta(y_1)(y_2 - y_1) \text{ for all } x, y_1, y_2 \right\} \tag{11}$$

$$= \inf \left\{ \int \varphi \, d\mu + \int \psi \, d\nu \; : \; c(x, y) \leq \varphi(x) + \psi(y) \; \textit{for all } x, y, \, \psi \textit{ convex} \right\}, \qquad (12)$$

where the infimum is taken over continuous functions φ, ψ *that are bounded in absolute terms by a multiple of* $x \mapsto (1 + d_{\mathcal{X}}(x, x_0))^r$, $y \mapsto (1 + |y|)^r$, *respectively, and* Δ *is measurable.*

We note that equality between (11) and (12) appears quite natural upon expressing φ in terms of an inf/sup convolution.

1.4 | Organization of the paper

In Section 2 we reformulate the robust pricing problem in discounted terms (with all quantities expressed in units of T_2-bonds) and illustrate in the case of caplets our results on superreplication duality and characterization of primal and dual optimizers. In Section 3 we formulate and prove our new results for the weak optimal transport problem, and from these we derive the results stated in the introduction.

2 | ROBUST PRICING PROBLEM

2.1 | On discounting

Throughout this section, we work in the framework described in Sections 1.1-1.2, thus in particular under Setup 1.1. Recall that we discount our financial instruments by expressing them in terms of T_2-bonds, see (3). In particular we consider

$$X_1 = \frac{1}{p(T_1, T_2)}, \quad Y_1 = \frac{p(T_1, T_3)}{p(T_1, T_2)}, \quad Y_2 = p(T_2, T_3). \qquad (13)$$

Then a generic T_2-forward measure is compatible with the information given by the market if for all $K \geq 0$ (cf. (1))

$$\frac{C(T_1, T_2, K)}{p(0, T_2)} = \mathbb{E}^{\mathbb{Q}_{T_2}} \left[\frac{(p(T_1, T_2) - K)^+}{p(T_1, T_2)} \right] = \mathbb{E}^{\mathbb{Q}_{T_2}} \left[(1 - KX_1)^+ \right],$$

$$\frac{C(T_2, T_3, K)}{p(0, T_2)} = \mathbb{E}^{\mathbb{Q}_{T_2}} \left[\frac{(p(T_2, T_3) - K)^+}{p(T_2, T_2)} \right] = \mathbb{E}^{\mathbb{Q}_{T_2}} \left[(Y_2 - K)^+ \right]. \qquad (14)$$

Therefore, by Breeden and Litzenberger (1978), from the observed call prices we can deduce the distribution under $\mathbb{Q}_{T_2}$ of X_1 and Y_2, which we denoted by μ and ν, respectively.

We are interested in a derivative of the form $\Phi(p(T_1, T_2), p(T_1, T_3))$, with maturity T_1. In terms of T_2-bonds, its payoff equals

$$\frac{\Phi(p(T_1, T_2), p(T_1, T_3))}{p(T_1, T_2)} = \Phi \left(\frac{1}{X_1}, \frac{Y_1}{X_1} \right) X_1.$$

This justifies our notation in Section 1.2, where we denoted the discounted payoff as a function of X_1, Y_1:

$$\hat{\Phi}(X_1, Y_1), \tag{15}$$

that is, formally we define

$$\hat{\Phi}(x_1, y_1) := \Phi\left(\frac{1}{x_1}, \frac{y_1}{x_1}\right)x_1. \tag{16}$$

According to a T_2-forward measure $\mathbb{Q}_{T_2}$, the discounted price for such a derivative is then

$$\mathbb{E}^{\mathbb{Q}_{T_2}}\left[\frac{\Phi(p(T_1, T_2), p(T_1, T_3))}{p(T_1, T_2)}\right] = \mathbb{E}^{\mathbb{Q}_{T_2}}[\hat{\Phi}(X_1, Y_1)].$$

Therefore, the robust price bounds are given by optimization problems (cf. (7))

$$\inf_{\mathbb{Q}_{T_2} \in \mathcal{Q}(\mu, \nu)} \mathbb{E}^{\mathbb{Q}_{T_2}}[\hat{\Phi}(X_1, Y_1)], \qquad \sup_{\mathbb{Q}_{T_2} \in \mathcal{Q}(\mu, \nu)} \mathbb{E}^{\mathbb{Q}_{T_2}}[\hat{\Phi}(X_1, Y_1)]. \tag{17}$$

On the other hand, the products available for trading in the market suggest the following semi-static hedging strategies, again expressed in discounted terms:

$$\varphi(X_1) + \psi(Y_2) + \Delta(X_1, Y_1)(Y_2 - Y_1). \tag{18}$$

In view of the definition of the discounted assets in (13), the static parts $\varphi(X_1)$, $\psi(Y_2)$ correspond to vanilla options written on $p(T_1, T_2)$ and $p(T_2, T_3)$, respectively. As usual these could be approximated using call/put options. The dynamic part $\Delta(X_1, Y_1)(Y_2 - Y_1)$ corresponds to rebalancing the portfolio in T_1 in a self-financing way, between T_2- and T_3-bonds, according to Δ. The quantity in (18) then amounts to the value of such a portfolio at time T_2. In fact, we find that it suffices to rebalance at time T_1 only depending on the value of Y_1, that is, it suffices to consider strategies of the form $\Delta(Y_1)$ rather than $\Delta(X_1, Y_1)$, see Theorem 1.3.

2.2 | Robust pricing of caplets: Primal and dual optimizers

In this section we consider vanilla derivatives of the type $\Phi = (p(T_1, T_3) - K)^+$, and comment on their primal and dual optimizers. W.l.o.g. we can consider $K = 1$. In discounted terms, this leads to considering the following function in (15):

$$\hat{\Phi}(x, y) = (y - x)^+.$$

By Proposition 1.5, the robust pricing bounds in (17) can be computed via

$$\inf_{\pi \in \Pi(\mu, \nu)} \int (b(\pi_x) - x)^+ \mu(dx), \tag{19}$$

$$\sup_{\pi \in \Pi(\mu,\nu)} \int (b(\pi_x) - x)^+ \, \mu(dx), \tag{20}$$

respectively. In Section 3.4 we discuss in detail the WOT for cost functions of the form $C(x, p) = \theta(b(p) - x)$, where θ is convex or concave. In particular we will determine the primal as well as dual optimizers and thus obtain the solutions to the problems (19) and (20):

- *Lower bound.* The optimizer for the lower bound (19) is determined by the so called *weak monotone rearrangement* of μ and ν (see Alfonsi et al. (2020) and Backhoff-Veraguas et al. (2020) for details): Briefly, by using $*$ to indicate a push-forward measure, and $\leq_c$ to denote convex order between measures[4], the family

$$\{T : \mathbb{R} \to \mathbb{R}, T \text{ is monotone, 1-Lip}, T_*(\mu) \leq_c \nu\}$$

has a unique element $\hat{T}$ for which the 1-Wasserstein distance $\mathcal{W}_1(\mu, \hat{T}_*(\mu))$ is minimized. (19) is then minimized if we put $Y_1 = \hat{T}(X_1)$ and couple Y_1 and Y_2 by an arbitrary martingale coupling of $T_*(\mu)$ and ν. In particular, in this market model, $X_1 = \frac{1}{p(T_1,T_2)}$ and $Y_1 = \frac{p(T_1,T_3)}{p(T_1,T_2)}$ are comonotone.

The corresponding sub-replication strategy is determined in Example 3.16, (68). It consists in the static position of being short a call on X_1 (that is, short puts on $p(T_1, T_2)$) and holding a call on Y_2 (that is, holding a call on $p(T_2, T_3)$), and on rearranging the portfolio at time T_1 (between the T_2 and the T_3 bonds) in order to hold one unit of the T_3-bond if Y_1 is above a certain threshold and else investing all in the T_2-bond.
- *Upper bound.* On the other hand, the upper bound (19) is obtained in a market where $Y_1 = Y_2$ a.s. and where X_1, Y_1 are coupled through the *anticomonotone coupling*. In particular,

$$p(T_1, T_3) = p(T_1, T_2)p(T_2, T_3) \quad \text{a.s.}$$

The corresponding superreplication strategy (Example 3.16, (69)) consists in a static position of holding a put on X_1 (that is, long calls on $p(T_1, T_2)$) and being short a put on Y_2 (that is, short a put on $p(T_2, T_3)$), and on rearranging the portfolio at time T_1 (between the T_2- and the T_3-bonds) in order to be one unit short in the T_3-bond if Y_1 is below a certain threshold and else investing all in the T_2-bond.

3 | WEAK OPTIMAL TRANSPORT PROBLEM

For the convenience of the reader who is potentially unfamiliar with the mathematical finance aspect of this work, this section can be read independently of the previous sections. The optimal transport problem for weak transport costs (WOT) was initiated by Gozlan et al. (2017) motivated by applications in geometric inequalities. Their contribution kicked off a vivid research activity by various groups: Gozlan et al. (2018), Alfonsi et al. (2019), Alfonsi et al. (2020), Gozlan and Juillet (2020), Alibert et al. (2019), Shu (2020), Backhoff-Veraguas et al. (2019), Backhoff-Veraguas et al. (2020), Backhoff-Veraguas and Pammer (2021b), Backhoff-Veraguas and Pammer (2021a), Fathi et al. (2020).

In this section we use techniques established in WOT to prove our main results (stated in the introduction). On the way, we prove some auxiliary results that might be of independent interest.

Specifically, we will show in Theorem 3.9 that, for μ, C continuous, the WOT transport problem can be reduced to the case where the cost function is convex in the second argument. We prove that

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}} C(x, \pi_x)\, \mu(dx) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}} C^{**}(x, \pi_x)\, \mu(dx), \tag{21}$$

where for each x, the function $C^{**}(x, \cdot)$ denotes the convex hull of the function $p \mapsto C(x, p)$. This reduction is relevant since virtually all results obtained in WOT so far assume convexity in the second argument.

In Subsection 3.3 we consider a particular class of WOT problems, where the cost function is barycentric. We draw a connection of barycentric costs to a specific class of OT problems where the second marginal is not fixed, but has to satisfy a convex order constraint. Moreover we derive duality results for semi-continuous costs of barycentric type that closely resemble the classical Monge-Kantorovich duality.

Remark 3.1. Following mathematical finance tradition, we focused on "super-hedging" rather than on "sub-hedging"-results in the introductory Sections 1 and 2. In contrast, optimal transport problems are usually phrased as minimization problems. In the case of the weak transport problem, this sign-convention also matters for the formulation of the convexity assumption on the cost function and for the dual problem that is based on a specific inf-convolution. We will thus switch sign in this section and hope that this does not cause confusion.

3.1 | Notation and setting

Fix $r \in [1, \infty)$. We denote by $\mathcal{X}$ and $\mathcal{Y}$ Polish spaces with compatible metrics $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$. We denote by $\mathcal{P}_r(\mathcal{X})$ the set of probability measures on $\mathcal{X}$ which finitely integrate $x \mapsto d_{\mathcal{X}}(x, x_0)^r$ (for some $x_0 \in \mathcal{X}$ and therefore also for any). The r-Wasserstein distance $\mathcal{W}_r(\mu, \nu)$ is defined for probabilities $\mu, \nu \in \mathcal{P}_r(\mathcal{X})$ by

$$\mathcal{W}_r(\mu, \nu)^r = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{X}} d_{\mathcal{X}}(x, x')^r\, \pi(dx, dx'),$$

and we equip $\mathcal{P}_r(\mathcal{X})$ with the topology induced by $\mathcal{W}_r$. The set of continuous functions which are absolutely bounded by a multiple of $1 + d_{\mathcal{X}}(x, x_0)^r$ is denoted by $C_r(\mathcal{X})$. For products $\mathcal{X} \times \mathcal{Y}$ of Polish spaces, we will consider the metric $d_r((x, x'), (y, y'))^r = d_{\mathcal{X}}(x, x')^r + d_{\mathcal{Y}}(y, y')^r$ and the corresponding Wasserstein space $\mathcal{P}_r(\mathcal{X} \times \mathcal{Y})$. We will often use the notation $\mu(f)$ to abbreviate $\int_{\mathcal{X}} f(x)\, \mu(dx)$.

3.1.1 | The weak optimal transport problem

Consider a measurable cost function $C : \mathcal{X} \times \mathcal{P}(\mathcal{Y}) \to (-\infty, \infty]$ satisfying the standing assumption that there exist a constant $K > 0$ and $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$ with

$$C(x, p) \geq -K\big(1 + d_{\mathcal{X}}(x, x_0)^r + \mathcal{W}_r(p, \delta_{y_0})^r\big). \tag{22}$$

Then the weak transport problem with cost C between $\mu \in \mathcal{P}_r(\mathcal{X})$ and $\nu \in \mathcal{P}_r(\mathcal{Y})$ is given by

$$\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X}} C(x, \pi_x) \, \mu(dx). \qquad \text{(WOT)}$$

Note that (22) guarantees that (WOT) is well-defined. Since the disintegration of a measure is measurable but not (weakly) continuous, we observe an inherent asymmetry in (WOT). To be able to obtain basic existence and duality results, different additional conditions have been imposed on the cost function C in the literature. In particular it is generically assumed that C is lower semicontinuity and satisfies

$$\forall x \in \mathcal{X} : p \mapsto C(x, p) \text{ is convex.} \qquad (23)$$

In Backhoff-Veraguas et al. (2019) it is established that these conditions are in fact sufficient to guarantee existence of a minimizer as well as a duality. The key observation in Backhoff-Veraguas et al. (2019) is that the spaces $\mathcal{X}$ and $\mathcal{Y}$ play different roles: $\mathcal{X}$ is akin to the state space of a stochastic process at time 1 and $\mathcal{Y}$ akin to the state space at time 2. This viewpoint leads to a technical convenient relaxation of (WOT) that we discuss in the next section.

3.1.2 | The relaxation of WOT

We start by defining an embedding of $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ into $\mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y}))$. To this end we write proj^i for the projection onto the i-th component and consider the disintegration $(\pi_x)_{x \in \mathcal{X}}$ of a measure $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ w.r.t. its first marginal $\text{proj}^1_*(\pi)$. Specifically we define

$$J(\pi) := (x \mapsto (x, \pi_x))_* \left(\text{proj}^1_*(\pi) \right) \in \mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y})).$$

The embedding J forms the basis for the relaxed version of (WOT). As we are interested in probabilities/processes with prescribed marginals μ, ν, we need to make sense of what it means for a measure in $\mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y}))$ to have ν as its "second marginal". To this end, consider the intensity maps $\hat{I} : \mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y})) \to \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and $I : \mathcal{P}(\mathcal{P}(\mathcal{Y})) \to \mathcal{P}(\mathcal{Y})$, where for $P \in \mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y}))$ and $Q \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))$ their intensities $\hat{I}(P)$ and $I(Q)$ are given by the unique measures satisfying

$$\hat{I}(P)(f) = \int_{\mathcal{X} \times \mathcal{P}(\mathcal{Y})} \int_{\mathcal{Y}} f(x, y) \, p(dy) \, P(dx, dp) \quad \forall f \in C_b(\mathcal{X} \times \mathcal{Y}), \qquad (24)$$

$$I(Q)(f) = \int_{\mathcal{P}(\mathcal{Y})} \int_{\mathcal{Y}} f(y) \, p(dy) \, Q(dp) \quad \forall f \in C_b(\mathcal{Y}), \qquad (25)$$

respectively. Note that $\hat{I}$ is the left inverse of J, that is, $\hat{I}(J(\pi)) = \pi$ for any $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. We denote by $\mathcal{P}(\mathcal{X} \rightsquigarrow \mathcal{Z})$ the set of probability measures on $\mathcal{X} \times \mathcal{Z}$ concentrated on the graph of a measurable function. Then J maps onto $\mathcal{P}(\mathcal{X} \rightsquigarrow \mathcal{P}(\mathcal{Y}))$, and $\hat{I}$ restricted to $\mathcal{P}(\mathcal{X} \rightsquigarrow \mathcal{P}(\mathcal{Y}))$ is its inverse.

We define $\Lambda(\mu, \nu) \subseteq \mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y}))$ by

$$\Lambda(\mu, \nu) := \{P \in \mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y})) : \hat{I}(P) \in \Pi(\mu, \nu)\}. \tag{26}$$

Since $\text{proj}^1_*(\hat{I}(P)) = \text{proj}^1_* P$ and $\text{proj}^2_*(\hat{I}(P)) = I(\text{proj}^2_* P)$, we obtain

$$\Lambda(\mu, \nu) = \left\{ P \in \mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y})) : \text{proj}^1_* P = \mu, I(\text{proj}^2_* P) = \nu \right\} = \bigcup_{\substack{\nu' \in \mathcal{P}(\mathcal{P}(\mathcal{Y})) \\ I(\nu') = \nu}} \Pi(\mu, \nu'). \tag{27}$$

Following Backhoff-Veraguas et al. (2019), we consider the following relaxation of (WOT)

$$\inf_{P \in \Lambda(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{P}(\mathcal{Y})} C(x, p) \, P(dx, dp). \tag{WOT'}$$

Clearly, we have that $J(\Pi(\mu, \nu)) \subseteq \Lambda(\mu, \nu)$ and hence (WOT') $\leq$ (WOT), but in many cases of interest more can be said. As long as μ is atomless, we have by the next lemma that $J(\Pi(\mu, \nu))$ is dense in $\Lambda(\mu, \nu)$ and hence the values of the two problems coincide under mild regularity assumptions.

Proposition 3.2. *Let $\mu \in \mathcal{P}(\mathcal{X})$ be continuous and $\nu \in \mathcal{P}(\mathcal{Y})$. Then the set $J(\Pi(\mu, \nu)) = \Lambda(\mu, \nu) \cap \mathcal{P}(\mathcal{X} \leadsto \mathcal{P}(\mathcal{Y}))$ is dense in $\Lambda(\mu, \nu)$. If in addition $\mu \in \mathcal{P}_r(\mathcal{X}), \nu \in \mathcal{P}_r(\mathcal{Y})$, and $C \in C_r(\mathcal{X} \times \mathcal{P}_r(\mathcal{Y}))$, the values of (WOT) and (WOT') coincide.*

Proof. It is well-known, see for example, Pratelli (2007), that, for a continuous probability measure μ on some Polish space $\mathcal{X}$ and an arbitrary probability measure ν' on some Polish space $\mathcal{Z}$, the set of Monge couplings, that is $\Pi(\mu, \nu') \cap \mathcal{P}(\mathcal{X} \leadsto \mathcal{Z})$, is dense in $\Pi(\mu, \nu')$. Applying this to $\mathcal{Z} = \mathcal{P}(\mathcal{Y})$ and $\nu' \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))$ yields that $\Pi(\mu, \nu') \cap \mathcal{P}(\mathcal{X} \leadsto \mathcal{P}(\mathcal{Y}))$ is dense in $\Pi(\mu, \nu')$. Hence, (27) yields that

$$\Lambda(\mu, \nu) \cap \mathcal{P}(\mathcal{X} \leadsto \mathcal{P}(\mathcal{Y})) = \bigcup_{\substack{\nu' \in \mathcal{P}(\mathcal{P}(\mathcal{Y})) \\ I(\nu') = \nu}} \Pi(\mu, \nu') \cap \mathcal{P}(\mathcal{X} \leadsto \mathcal{P}(\mathcal{Y}))$$

is dense in $\Lambda(\mu, \nu)$. Finally, since $\hat{I}$ restricted to $\mathcal{P}(\mathcal{X} \leadsto \mathcal{P}(\mathcal{Y}))$ is the inverse of J, we have that $J : \Pi(\mu, \nu) \to \mathcal{P}(\mathcal{X} \leadsto \mathcal{P}(\mathcal{Y})) \cap \Lambda(\mu, \nu)$ is a bijection and $J(\Pi(\mu, \nu)) = \mathcal{P}(\mathcal{X} \leadsto \mathcal{P}(\mathcal{Y})) \cap \Lambda(\mu, \nu)$. Now let $\mu \in \mathcal{P}_r(\mathcal{X}), \nu \in \mathcal{P}_r(\mathcal{Y})$, and $C \in C_r(\mathcal{X} \times \mathcal{P}_r(\mathcal{Y}))$. For $P \in \Lambda(\mu, \nu)$ we find a sequence $P^k \in \Lambda(\mu, \nu) \cap \mathcal{P}_r(\mathcal{X} \leadsto \mathcal{P}_r(\mathcal{Y}))$ that converges to P weakly. Since the marginals are fixed with finite r-th moments, Definition 6.8 in Villani (2009) yields that this convergence holds even in $\mathcal{W}_r$ on $\mathcal{P}_r(\mathcal{X} \times \mathcal{P}_r(\mathcal{Y}))$. Let $\pi^k \in \Pi(\mu, \nu)$ with $J(\pi^k) = P^k, k \in \mathbb{N}$. Then

$$\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X}} C(x, \pi_x) \, \mu(dx) \leq \liminf_{k \to \infty} \int_{\mathcal{X}} C(x, \pi_x^k) \, \mu(dx)$$

$$= \lim_{k \to \infty} \int_{\mathcal{X} \times \mathcal{P}_r(\mathcal{Y})} C(x, p) \, P^k(dx, dp) = \int_{\mathcal{X} \times \mathcal{P}_r(\mathcal{Y})} C(x, p) \, P(dx, dp).$$

As P was an arbitrary element of $\Lambda(\mu, \nu)$ and (WOT) $\geq$ (WOT'), this shows that the values of (WOT) and (WOT') coincide. $\square$

3.1.3 | On the process interpretation of the weak transport problem and its relaxation

A common interpretation of probability measures $\pi \in \Pi(\mu, \nu)$ in optimal transport is to view π as the law of a two step stochastic process $(X_1, X_2) = (X, Y)$ where $X \sim \mu$ and $Y \sim \nu$. Due to the inherent asymmetry of (WOT), this point of view seems even more natural in the framework of weak optimal transport. Specifically, the usual weak transport problem can be reformulated as an optimization problem over (2-period) stochastic processes as indicated by the following equality:

$$\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X}} C(x, \pi_x) \, \mu(dx) = \inf_{\substack{(\Omega, \mathbb{P}, X, Y) \\ X \sim \mu, Y \sim \nu}} \mathbb{E}_{\mathbb{P}}[C(X, \mathscr{L}(Y|X))], \tag{28}$$

where $\mathscr{L}$ denotes the law under $\mathbb{P}$. Notably, also the relaxed problem introduced in the previous subsection admits a natural formulation in terms of *filtered processes*, that is, processes together with a filtration. Indeed, relaxing the set $\Pi(\mu, \nu)$ on the left-hand side of (28) corresponds on the right-hand side of (28) to minimizing over adapted processes together with arbitrary filtrations, rather than allowing only for filtrations generated by the process itself:

Proposition 3.3. *Let* $C : \mathcal{X} \times \mathcal{P}(\mathcal{Y}) \to \mathbb{R}$ *be measurable. Then we have*

$$\inf_{P \in \Lambda(\mu, \nu)} \int_{\mathcal{X}} C(x, p) \, P(dx, dp) = \inf_{\substack{(\Omega, \mathbb{P}, (\mathcal{F}_t)_{t=1,2}, X, Y) \\ X \sim \mu, Y \sim \nu}} \mathbb{E}_{\mathbb{P}}[C(X, \mathscr{L}(Y|\mathcal{F}_1))], \tag{29}$$

in the sense that either side is defined if the other is and then the two are equal.

To obtain (29) we need to establish the appropriate correspondence between elements of $\Lambda(\mu, \nu)$ and stochastic processes with marginals μ, ν, which is what we do in the proof of the proposition.

Proof of Proposition 3.3. To show '$\leq$' assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is some probability space and that (X, Y), $X \sim \mu, Y \sim \nu$, is a process adapted to a filtration $(\mathcal{F}_t)_{t=1,2}$. Then this process induces the required candidate in $\Lambda(\mu, \nu)$ by setting

$$P := \mathscr{L}(X, \mathscr{L}(Y|\mathcal{F}_1)). \tag{30}$$

To show the other inequality '$\geq$', fix $P \in \Lambda(\mu, \nu)$. Define a filtered probability space by $\mathcal{X} \times \mathcal{P}(\mathcal{Y}) \times \mathcal{Y}$ with σ-algebra $\mathcal{F} = \mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{P}(\mathcal{Y})) \otimes \mathcal{B}(\mathcal{Y})$, probability measure $\mathbb{P}(dx, dp, dy) = P(dx, dp) p(dy)$, and filtration $\mathcal{F}_1 = \sigma((x, p, y) \mapsto (x, p))$, $\mathcal{F}_2 = \mathcal{F}$. Then the process $(X(x, p, y) = x, Y(x, p, y) = y)$ is as required on the r.h.s. of (29). □

Remark 3.4. We briefly describe the wider context of the relation between $\mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y}))$ and the filtered processes which is the topic of Bartl et al. (2021). There it is detailed that the space FP of filtered (stochastic) processes is naturally equipped with an adapted variant of the Wasserstein distance denoted by $\mathcal{AW}$. Two processes have distance 0 if and only if they have the same probabilistic properties in a specific sense. Identifying such processes, the space $(FP, \mathcal{AW})$ is a complete metric space, isometric to a classical Wasserstein space $(Z, \mathcal{W}_Z)$. Specifically, in the case of

two periods,

$$(\mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y})), \mathcal{W}) \quad \text{is isometric to} \quad (FP, \mathcal{A}\mathcal{W})$$

and the corresponding isometry (and its inverse, respectively) are the operations given in the proof of Proposition 3.3. Stochastic processes where the filtration is generated by the process itself correspond to elements of $J(\Lambda(\mu, \nu))$. Proposition 3.2 then asserts precisely that such processes with given marginals are dense in the set of general processes with these marginals.

3.2 | Weak transport duality for non-convex cost

The main result of this section is Theorem 3.9, which can be seen as a justification for the convexity assumption on the cost C (in the sense of (23)) that is typically used in the WOT-literature. First we show, in Proposition 3.8, that the relaxed problem (WOT') for a cost C coincides with a weak transport problem (WOT) for its convexification C^{**}. Then in Theorem 3.9 we obtain that the values of (WOT) for cost C and the convexification C^{**} coincide, provided that the cost is continuous and the first marginal is continuous as well. Hence, when one is solely interested in the value of (WOT) for continuous cost, it is of no harm to convexify C and thus work in the framework of classical WOT theory.

For the sake of completeness we include the following (slightly more general) version of the first assertion of (Backhoff-Veraguas et al., 2019, Theorem 3.1), where we weaken the assumption that C is bounded from below to (22). So far the weak transport duality was usually stated in the form of (31) using an inf-convolution instead of pairs of dual functions, c.f. (Gozlan et al., 2017, Theorem 9.6), (Alibert et al., 2019, Theorem 4.2) and (Backhoff-Veraguas et al., 2019, Theorem 3.1). A minor contribution of the next theorem is the harmonisation of the weak transport duality and the Kantorovich duality (known from classical optimal transport) in terms of dual pairs instead of utilizing an inf-convolution.

Theorem 3.5 (Duality). *Let $(\mu, \nu) \in \mathcal{P}_r(\mathcal{X}) \times \mathcal{P}_r(\mathcal{Y})$ and $C : \mathcal{X} \times \mathcal{P}_r(\mathcal{Y}) \to (-\infty, \infty]$ be lower semicontinuous such that Then*

$$\inf_{P \in \Lambda(\mu, \nu)} \int C(x, p) P(dx, dp) = \sup_{\psi \in C_r(\mathcal{Y})} \nu(\psi) + \mu(R_C \psi) \tag{31}$$

$$= \sup_{\substack{\varphi \in C_r(\mathcal{X}), \psi \in C_r(\mathcal{Y}) \\ \varphi(x) + p(\psi) \leq C(x, p)}} \mu(\varphi) + \nu(\psi), \tag{32}$$

where $R_C \psi(x) := \inf\{-p(\psi) + C(x, p) : p \in \mathcal{P}_r(\mathcal{Y})\}$ and $\mu(R_C \psi) := -\infty$ if the integral of $R_C \psi$ w.r.t. μ is not well-defined.

Proof. Define the auxiliary cost function $C^K : \mathcal{X} \times \mathcal{P}(\mathcal{Y}) \to [0, \infty)$ by

$$C^K(x, p) := C(x, p) + K\left(1 + d_{\mathcal{X}}(x, x_0)^r + \mathcal{W}_r(p, \delta_{y_0})^r\right),$$

where $(x, p) \in \mathcal{X} \times \mathcal{P}(\mathcal{Y})$. For $P \in \Lambda(\mu, \nu)$ we have

$$\int_{\mathcal{X} \times \mathcal{P}(\mathcal{Y})} 1 + d_{\mathcal{X}}(x, x_0)^r + \mathcal{W}_r(p, \delta_{y_0})^r \, P(dx, dp)$$

$$= 1 + \int_{\mathcal{X}} d_{\mathcal{X}}(x, x_0)^r \, \mu(dx) + \int_{\mathcal{X} \times \mathcal{P}(\mathcal{Y})} \int_{\mathcal{Y}} d_{\mathcal{Y}}(y, y_0)^r \, p(dy) \, P(dx, dp)$$

$$= 1 + \int_{\mathcal{X}} d_{\mathcal{X}}(x, x_0)^r \, \mu(dx) + \int_{\mathcal{Y}} d_{\mathcal{Y}}(y, y_0)^r \, \nu(dy),$$

and therefore

$$\int_{\mathcal{X} \times \mathcal{P}(\mathcal{Y})} C^K(x, p) \, P(dx, dp) = \int_{\mathcal{X} \times \mathcal{P}(\mathcal{Y})} C(x, p) \, P(dx, dp)$$

$$+ K \left(1 + \int_{\mathcal{X}} d_{\mathcal{X}}(x, x_0)^r \, \mu(dx) + \int_{\mathcal{Y}} d_{\mathcal{Y}}(y, y_0)^r \, \nu(dy) \right).$$

For $\psi \in C_r(\mathcal{Y})$ we write

$$\psi_K(y) := \psi(y) - K \left(1 + d_{\mathcal{Y}}(y, y_0)^r \right) \in C_r(\mathcal{Y})$$

and obtain the following relation between the inf-convolutions of C and C^K:

$$R_{C^K} \psi(x) = \inf_{p \in \mathcal{P}_r(\mathcal{Y})} -p(\psi) + C^K(x, p)$$

$$= \inf_{p \in \mathcal{P}_r(\mathcal{Y})} -p(\psi) + C(x, p) + K \left(1 + d_{\mathcal{X}}(x, x_0)^r + \int_{\mathcal{Y}} d_{\mathcal{Y}}(y, y_0)^r \, p(dy) \right)$$

$$= \inf_{p \in \mathcal{P}_r(\mathcal{Y})} -p(\psi_K) + C(x, p) + K d_{\mathcal{X}}(x, x_0)^r$$

$$= R_C \psi_K(x) + K d_{\mathcal{X}}(x, x_0)^r.$$

In particular, the μ-integral of $R_{C^K} \psi$ is well-defined if and only if the μ-integral of $R_C \psi_K$ is. Since C^K is bounded from below and lower semicontinuous, (Backhoff-Veraguas et al., 2019, Theorem 3.1) yields[5]

$$\inf_{P \in \Lambda(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{P}(\mathcal{Y})} C^K(x, p) \, P(dx, dp) = \sup_{\psi \in C_r(\mathcal{Y})} \nu(\psi) + \mu(R_{C^K} \psi)$$

$$= \sup_{\psi \in C_r(\mathcal{Y})} \nu(\psi) + \int_{\mathcal{X}} R_C \psi_K(x) + K d_{\mathcal{X}}(x, x_0)^r \, \mu(dx)$$

$$= K \left(1 + \int_{\mathcal{X}} d_{\mathcal{X}}(x, x_0)^r \, \mu(dx) + \int_{\mathcal{Y}} d_{\mathcal{Y}}(y, y_0)^r \, \nu(dy) \right)$$

$$+ \sup_{\psi \in C_r(\mathcal{Y})} \nu(\psi_K) + \mu(R_C \psi_K).$$

Rearranging the terms and relabeling ψ_K as ψ yields (31).

The next goal is to show (32), which resembles more closely the classical Kantorovich duality. By the first part of the proof, we may assume w.l.o.g. that C is nonnegative. It is well-known that for any lower semicontinuous, nonnegative function f (on a metric space), there exists a sequence of nonnegative functions $(f_k)_{k\in\mathbb{N}}$ such that f_k is k-Lipschitz continuous and absolutely bounded by k, and $f_k \nearrow f$.

Let $(C^K)_{k\in\mathbb{N}}$ be such a sequence for C. Monotone convergence and existence of minimizers, see (Backhoff-Veraguas et al., 2019, Theorem 2.9), yield

$$\sup_k \inf_{P\in\Lambda(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})} C^K(x,p)\,P(dx,dp) = \inf_{P\in\Lambda(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})} C(x,p)\,P(dx,dp).$$

Assume for a moment that (32) holds for all $k \in \mathbb{N}$, then

$$\inf_{P\in\Lambda(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})} C(x,p)\,P(dx,dp) = \sup_k \sup_{\substack{\varphi\in C_r(\mathcal{X}),\psi\in C_r(\mathcal{Y})\\ \varphi(x)+p(\psi)\leq C^K(x,p)}} \nu(\psi)+\mu(\varphi)$$

$$\leq \sup_{\substack{\varphi\in C_r(\mathcal{X}),\psi\in C_r(\mathcal{Y})\\ \varphi(x)+p(\psi)\leq C(x,p)}} \nu(\psi)+\mu(\varphi)$$

$$\leq \inf_{P\in\Lambda(\mu,\nu)} \int C(x,p)\,P(dx,dp).$$

Therefore, the inequalities are in fact equalities and (32) holds for C as in . the statement of the theorem.

It is sufficient to show (32) for k-Lipschitz weak transport costs C, $k \in \mathbb{N}$. When C is k-Lipschitz, its inf-convolution $R_C\psi$ is also k-Lipschitz as the next computation shows:

$$|R_C\psi(x) - R_C\psi(x')| \leq \sup_{p\in\mathcal{P}_r(\mathcal{Y})} |C(x,p) - C(x',p)| \leq kd_{\mathcal{X}}(x,x').$$

As a consequence we have that $R_C\psi \in C_r(\mathcal{X})$, and

$$\sup_{\psi\in C_r(\mathcal{Y})} \nu(\psi)+\mu(R_C\psi) \geq \sup_{\substack{\varphi\in C_r(\mathcal{X}),\psi\in C_r(\mathcal{Y})\\ \varphi(x)+p(\psi)\leq C(x,p)}} \nu(\psi)+\mu(\varphi).$$

On the other hand, the reverse inequality is easy to show: Let $\varphi \in C_r(\mathcal{X})$ and $\psi \in C_r(\mathcal{Y})$ such that $\varphi(x) + p(\psi) \leq C(x,p)$. Then $\varphi(x) \leq -p(\psi) + C(x,p)$ for all $(x,p) \in \mathcal{X} \times \mathcal{P}_r(\mathcal{Y})$, hence,

$$\varphi(x) \leq \inf_{p\in\mathcal{P}_r(\mathcal{Y})} -p(\psi) + C(x,p) = R_C\psi(x) \quad \forall x \in \mathcal{X}.$$

Therefore we have

$$\nu(\psi) + \mu(\varphi) \leq \nu(\psi) + \mu(R_C\psi),$$

which readily shows the reverse inequality and completes the proof. □

The following result represents a version of Jensen's inequality that proves to be useful in our context.

Lemma 3.6. *Let* $F : \mathcal{P}_r(\mathcal{Y}) \to (-\infty, \infty]$ *be convex. Assume that either of the following holds:*

(a) F is lower semicontinuous;
(b) F is upper semicontinuous and upper bounded by a multiple of $p \mapsto \mathcal{W}_r(p, p_0)^r$ *for some* $p_0 \in \mathcal{P}_r(\mathcal{Y})$.

Then we have for $Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))$

$$F(I(Q)) \le \int_{\mathcal{P}_r(\mathcal{Y})} F(q) \, Q(dq).$$

Proof. Fix $Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))$. If (a) is satisfied, for $p \in \mathcal{P}_r(\mathcal{Y})$ by Fenchel's duality theorem we have

$$F(p) = \sup_{\psi \in C_r(\mathcal{Y})} p(\psi) - \sup_{q \in \mathcal{P}_r(\mathcal{Y})} q(\psi) - F(q).$$

Therefore, for $Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))$ with $I(Q) = p$, by interchanging supremum with integration, we have that

$$F(p) = \sup_{\psi \in C_r(\mathcal{Y})} \int_{\mathcal{P}_r(\mathcal{Y})} q'(\psi) \, Q(dq') - \sup_{q \in \mathcal{P}_r(\mathcal{Y})} q(\psi) - F(q) \le \int_{\mathcal{P}_r(\mathcal{Y})} F(q') \, Q(dq').$$

Now assume that (b) holds. Since $\mathcal{P}_r(\mathcal{Y})$ is Polish, there exists by (Kallenberg, 2002, Lemma 3.22) a map $h : (0, 1) \to \mathcal{P}_r(\mathcal{Y})$ such that $Q = h_* \lambda$, where λ denotes the Lebesgue measure on $(0,1)$. We define, for $1 \le k \le n \in \mathbb{N}$,

$$Q_k^n(dq) := n \int_{\frac{k-1}{n}}^{\frac{k}{n}} \delta_{h(t)} \, dt.$$

It is immediate that $\frac{1}{n} \sum_{k=1}^n Q_k^n = Q$, and

$$\mathcal{W}_r\left(\frac{1}{n} \sum_{k=1}^n \delta_{I(Q_k^n)}, Q \right)^r = \mathcal{W}_r\left(\frac{1}{n} \sum_{k=1}^n \delta_{n \int_{\frac{k-1}{n}}^{\frac{k}{n}} h(t) \, dt}, \int_0^1 \delta_{h(t)} \, dt \right)^r$$

$$\le \frac{1}{n} \sum_{k=1}^n \mathcal{W}_r\left(\delta_{n \int_{\frac{k-1}{n}}^{\frac{k}{n}} h(t) \, dt}, n \int_{\frac{k-1}{n}}^{\frac{k}{n}} \delta_{h(t)} \, dt \right)^r$$

$$\le \sum_{k=1}^n \int_{\frac{k-1}{n}}^{\frac{k}{n}} \mathcal{W}_r\left(n \int_{\frac{k-1}{n}}^{\frac{k}{n}} h(t) \, dt, h(s) \right)^r ds$$

$$\leq n \sum_{k=1}^{n} \int_{\frac{k-1}{n}}^{\frac{k}{n}} \int_{\frac{k-1}{n}}^{\frac{k}{n}} \mathcal{W}_r(h(t), h(s))^r \, dt \, ds$$

$$\leq \sup \left\{ \int_{(0,1)^2} \mathcal{W}_r(h(t), h(s))^r \chi(dt, ds) : \chi \in \Pi(\lambda, \lambda) \right.$$

$$\left. \int_{(0,1)^2} |s - t|^r \chi(dt, ds) \leq \frac{1}{n^r} \right\},$$

where we applied multiple times Jensen's inequality (which is possible in this settings thanks to the first part of this lemma). The right-hand side vanishes for $n \to \infty$ by (Eder, 2019, Lemma 2.7). Due to convexity of F, we have

$$F(I(Q)) = F\left(\frac{1}{n} \sum_{k=1}^{n} I(Q_k^n) \right) \leq \frac{1}{n} \sum_{k=1}^{n} F(I(Q_k^n)).$$

We conclude by taking the limit superior for $n \to \infty$, which yields by upper semicontinuity

$$F(I(Q)) \leq \limsup_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} F(I(Q_k^n)) \leq \int_{\mathcal{P}_r(\mathcal{Y})} F(q) \, Q(dq).$$

$\square$

Lemma 3.7. *Assume that $C : \mathcal{X} \times \mathcal{P}_r(\mathcal{Y}) \to (-\infty, \infty]$ is measurable and bounded from below as in (22). For $x \in \mathcal{X}$ we denote by $C^{**}(x, \cdot)$ the lower semicontinuous convex envelope of $C(x, \cdot)$. For $x \in \mathcal{X}, p \in \mathcal{P}_r(\mathcal{Y})$ we write $\tilde{C}(x, p) := \inf_{Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y})), I(Q)=p} \int C(x, q) \, Q(dq)$. Let $Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))$ with $I(Q) = p$, then*

$$C^{**}(x, p) \leq \tilde{C}(x, p) \leq \int_{\mathcal{P}_r(\mathcal{Y})} \tilde{C}(x, q) \, Q(dq). \tag{33}$$

Moreover, if $C(x, \cdot)$ is lower semicontinuous there is equality, that is, $C^{**}(x, p) = \tilde{C}(x, p)$.

Proof. By Lemma 3.6 we have for $Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))$ with $I(Q) = p$ that

$$C^{**}(x, p) \leq \int_{\mathcal{P}_r(\mathcal{Y})} C^{**}(x, q) \, Q(dq) \leq \int_{\mathcal{P}_r(\mathcal{Y})} C(x, q) \, Q(dq).$$

Taking the infimum over all $Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))$ with $I(Q) = p$ shows the first inequality in (33).

To see the second inequality in (33), we observe that the map $q \mapsto \tilde{C}(x, q)$ is by Proposition 7.47 in Bertsekas and Shreve (1978) lower semianalytic, and that $D = \{(p, Q) \in \mathcal{P}_r(\mathcal{Y}) \times \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y})) : I(Q) = p\}$ is closed. Therefore, by Proposition 7.50 in Bertsekas and Shreve (1978), for each $\varepsilon > 0$ there is an analytically measurable function $q \mapsto Q^q$ with $I(Q^q) = q$ and

$$\int_{\mathcal{P}_r(\mathcal{Y})} C(x, q') \, Q^q(dq') \leq \tilde{C}(x, q) + \varepsilon \quad \forall q \in \mathcal{P}_r(\mathcal{Y}). \tag{34}$$

We now compute the intensity of $\hat{Q}(dq) := \int_{\mathcal{P}_r(\mathcal{Y})} Q^{q'}(dq) Q(dq')$:

$$I(\hat{Q}) = \int_{\mathcal{P}_r(\mathcal{Y})} I(Q^{q'}) Q(dq') = \int_{\mathcal{P}_r(\mathcal{Y})} q' Q(dq') = I(Q) = p.$$

Hence,

$$\tilde{C}(x, p) \le \int_{\mathcal{P}_r(\mathcal{Y})} C(x, q) \hat{Q}(dq) \le \int_{\mathcal{P}_r(\mathcal{Y})} \tilde{C}(x, q) Q(dq) + \varepsilon,$$

which proves the second inequality in (33) as ε is arbitrary.

If $C(x, \cdot)$ is lower semicontinuous, we have by Theorem 3.5

$$\sup_{\psi \in C_r(\mathcal{Y})} p(\psi) + R_C \psi(x) = \inf_{P \in \Lambda(\delta_x, p)} \int C(x', p) P(dx', dp) = \tilde{C}(x, p),$$

whence $C^{**}(x, p) = \tilde{C}(x, p)$ for all $p \in \mathcal{P}_r(\mathcal{Y})$. $\qquad\square$

Proposition 3.8. *Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}_r(\mathcal{Y})$. Under the assumptions of Lemma 3.7, we have*

$$\inf_{P \in \Lambda(\mu, \nu)} \int C(x, p) P(dx, dp) = \inf_{P \in \Lambda(\mu, \nu)} \int \tilde{C}(x, p) P(dx, dp) \qquad (35)$$

$$= \inf_{\pi \in \Pi(\mu, \nu)} \int \tilde{C}(x, \pi_x) \mu(dx) \qquad (36)$$

$$\ge \inf_{P \in \Lambda(\mu, \nu)} \int C^{**}(x, p) P(dx, dp) \qquad (37)$$

$$= \inf_{\pi \in \Pi(\mu, \nu)} \int C^{**}(x, \pi_x) \mu(dx). \qquad (38)$$

Moreover, if $C(x, \cdot)$ is lower semicontinuous for all $x \in \mathcal{X}$, there is equality.

Proof. Let $P \in \Lambda(\mu, \nu)$ and write $\hat{I}(P) =: \pi \in \Pi(\mu, \nu)$. For disintegrations $(P_x)_{x \in \mathcal{X}}$ and $(\pi_x)_{x \in \mathcal{X}}$ of P and π w.r.t. μ, respectively, we have by definition of the intensity maps, c.f. (24) and (25), the relation

$$I(P_x) = \pi_x \quad \mu\text{-almost surely.} \qquad (39)$$

Due to Lemma 3.7 and (39), we have

$$\int_{\mathcal{X} \times \mathcal{P}_r(\mathcal{Y})} C(x, p) P(dx, dp) \ge \int_{\mathcal{X} \times \mathcal{P}_r(\mathcal{Y})} \tilde{C}(x, p) P(dx, dp) = \int_{\mathcal{X}} \int_{\mathcal{P}_r(\mathcal{Y})} \tilde{C}(x, p) P_x(dp) \mu(dx)$$

$$\geq \int_{\mathcal{X}} \tilde{C}(x, \pi_x)\,\mu(dx) \geq \int_{\mathcal{X}} C^{**}(x, \pi_x)\,\mu(dx).$$

Clearly, (37) is dominated by (38), and at the same time we have by Lemma 3.6

$$\int_{\mathcal{X}\times\mathcal{P}_r(\mathcal{Y})} C^{**}(x, p)\,P(dx, dp) = \int_{\mathcal{X}}\int_{\mathcal{P}_r(\mathcal{Y})} C^{**}(x, p)\,P_x(dp)\,\mu(dx) \geq \int_{\mathcal{X}} C^{**}(x, I(P_x))\,\mu(dx),$$

which proves that (37) and (38) coincide.

It remains to show that the left-hand side of (35) is dominated by (36). Pick $\pi \in \Pi(\mu, \nu)$. By the measurable selection argument presented in the proof of Lemma 3.7, we find for $\varepsilon > 0$ an analytically measurable map $q \mapsto Q^q$ with $I(Q^q) = q$ and such that (34) holds. The intensity of $P(dx, dq) := \mu(dx)\,Q^{\pi_x}(dq)$ is given by

$$\hat{I}(P)(dx, dy) = \mu(dx)\,I(Q^{\pi_x})(dy) = \mu(dx)\,\pi_x(dy) = \pi(dx, dy).$$

Thus $P(dx, dp) \in \Lambda(\mu, \nu)$ and

$$\int_{\mathcal{X}\times\mathcal{P}_r(\mathcal{Y})} C(x, p)\,P(dx, dp) = \int_{\mathcal{X}}\int_{\mathcal{P}_r(\mathcal{Y})} C(x, p)\,Q^{\pi_x}(dp)\,\mu(dx) \leq \int_{\mathcal{X}} \tilde{C}(x, \pi_x)\,\mu(dx) + \varepsilon.$$

We see that (36) dominates the left-hand side of (35) by recalling that $\varepsilon > 0$ was arbitrary. Finally, the last statement follows from Lemma 3.7. $\qquad\square$

Theorem 3.9. *Assume that $\mu \in \mathcal{P}_r(\mathcal{X})$ is continuous, $\nu \in \mathcal{P}_r(\mathcal{Y})$, and $C \in C_r(\mathcal{X} \times \mathcal{P}_r(\mathcal{Y}))$. Then*

$$\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}} C(x, \pi_x)\,\mu(dx) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}} C^{**}(x, \pi_x)\,\mu(dx). \tag{40}$$

In particular, the minimization problems (35)-(38) yield the same value.

Proof. Since μ is continuous, by Proposition 3.2 the values of (WOT) and (WOT') coincide. Then, by Proposition 3.8 we have that the right-hand side of (40) and (WOT') yield the same value. Hence we conclude that there holds equality in (40). The last statement follows by Proposition 3.8, since $C \geq \tilde{C} \geq C^{**}$. $\qquad\square$

Proof of Proposition 1.5. First, we note that (7) can be recast as a (WOT')-problem. Indeed, consider $C : \mathbb{R} \times \mathcal{P}_r(\mathbb{R}) \to \mathbb{R}, C(x, p) := \hat{\Phi}(x, b(p))$. Then, we have by Proposition 3.3 that

$$\sup_{\mathbb{Q}\in\mathcal{Q}(\mu,\nu)} \mathbb{E}_{\mathbb{Q}}[\hat{\Phi}(X_1, Y_1)] = \sup_{P\in\Lambda(\mu,\nu)} \int C(x, p)\,P(dx, dp) = -\inf_{P\in\Lambda(\mu,\nu)} \int -C(x, p)\,P(dx, dp), \tag{41}$$

where in the last equality we switch from a sup- to an inf-formulation to match (WOT'). Since $\hat{\Phi}$, and thus C, is continuous, we obtain from Proposition 3.8 that

$$\inf_{P\in\Lambda(\mu,\nu)} \int -C(x, p)\,P(dx, dp) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}} (-C)^{**}(x, \pi_x)\,\mu(dx).$$

Finally, as μ is continuous, we can apply Theorem 3.9 and obtain (9). $\qquad\square$

Remark 3.10. In this remark we show that under certain assumptions (WOT) for concave cost C reduces to a classical optimal transport problem. Recall the definition of $\tilde{C}(x, \cdot)$ from Lemma 3.7. Assume that $C(x, \cdot)$ is concave in the following sense:

$$\int_{\mathcal{P}_r(\mathcal{Y})} C(x, q) Q(dq) \leq C(x, I(Q)) \quad \forall Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y})), \tag{42}$$

which is satisfied as soon as $C(x, \cdot)$ is concave, lower/upper semicontinuous and sufficiently bounded from above, by Lemma 3.6. It turns out that then $\tilde{C}(x, \cdot)$ is in fact linear. Indeed, let $p \in \mathcal{P}_r(\mathcal{Y})$. The pushforward measure $Q^* := (y \mapsto \delta_y)_* p$ clearly constitutes an element of $\mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))$ such that $I(Q^*) = p$. Consequentially we deduce

$$\tilde{C}(x, p) = \inf_{\substack{Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y})) \\ I(Q)=p}} \int_{\mathcal{P}_r(\mathcal{Y})} C(x, q) Q(dq) \leq \int_{\mathcal{P}_r(\mathcal{Y})} C(x, q) Q^*(dq) = \int_{\mathcal{Y}} C(x, \delta_y) p(dy). \tag{43}$$

On the other hand, due to (42) we have

$$\int_{\mathcal{Y}} C(x, \delta_y) p(dy) = \int_{\mathcal{P}_r(\mathcal{Y})} \int_{\mathcal{Y}} C(x, \delta_y) q(dy) Q(dq) \leq \int_{\mathcal{P}_r(\mathcal{Y})} C(x, q) Q(dq), \tag{44}$$

for all $Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))$ with $I(Q) = p$. Thus the left-hand side of (44) is dominated by $\tilde{C}(x, p)$. By combining this with (43), we derive

$$\tilde{C}(x, p) = \int_{\mathcal{Y}} C(x, \delta_y) p(dy).$$

Therefore, (35) reduces to a classical optimal transport problem with cost $c(x, y) := C(x, \delta_y)$:

$$\inf_{P \in \Lambda(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{P}_r(\mathcal{Y})} C(x, p) P(dx, dp) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \pi(dx, dy).$$

3.3 | Barycentric cost functions

In this section we derive some properties of weak transport problem for 'barycentric' cost functions and use them to prove the main results stated in the introduction. We call a cost function $C : \mathcal{X} \times \mathcal{P}_1(\mathbb{R}^d) \to (-\infty, \infty]$ barycentric if, for any $x \in \mathcal{X}$, $p, q \in \mathcal{P}_r(\mathbb{R}^d)$, we have the implication

$$b(p) = b(q) \Rightarrow C(x, p) = C(x, q), \tag{45}$$

where we recall that $b(p)$ denotes the barycenter of p. Equivalently, C is barycentric if there exists $c : \mathcal{X} \times \mathbb{R}^d \to (-\infty, \infty]$ such that

$$C(x, p) = c(x, b(p)), \quad (x, p) \in \mathcal{X} \times \mathcal{P}_1(\mathbb{R}^d).$$

Weak optimal transport problems where the cost function C is of barycentric-type have been recently investigated by various authors in different contexts, see Gozlan et al. (2018), Gozlan et al. (2017), Gozlan and Juillet (2020), Shu (2020), Alfonsi et al. (2020), Daskalakis et al. (2017), Backhoff-Veraguas et al. (2019), Backhoff-Veraguas et al. (2020), Backhoff-Veraguas and Pammer (2021a). Typically in these papers, $\mathcal{X}$ was also given as $\mathbb{R}^d$ and $C(x, p) = \theta(b(p) - x)$ for some convex $\theta : \mathbb{R}^d \to \mathbb{R}$. Motivated by functional inequalities, this particular problem was explored by Gozlan et al. (2017), Gozlan et al. (2018), Shu (2020), Gozlan and Juillet (2020). On the other hand, in Fathi et al. (2020) it was used to give a new proof to Cafarelli's contraction theorem. In Alfonsi et al. (2020) the authors are mainly motivated by applications in robust mathematical finance. They use the barycentric WOT problem to construct a sampling technique preserving the convex order, which can then be used to approximate martingale optimal transport problems. Finally, in Daskalakis et al. (2017) the barycentric WOT problem appears in the context of revenue maximization and mechanism design in economics. For a clearer connection of this topic to WOT we refer to Backhoff-Veraguas and Pammer (2021a).

The next proposition fleshes out the intrinsic connection of (WOT') with barycentric C and the convex order (and thereby convex functions). Recall that we denote by $\leq_c$ the convex order between probabilities. For the special case when the cost $C(x, p)$ is given by $\theta(b(p) - x)$ the results of the next proposition can be (partially) found in Gozlan et al. (2017) and Backhoff-Veraguas et al. (2019).

Proposition 3.11. *Let $\mu \in \mathcal{P}_r(\mathcal{X})$, $\nu \in \mathcal{P}_r(\mathbb{R}^d)$, and C be measurable and barycentric such $\int C(x, p) P(dx, dp)$ is well-defined for any $P \in \Lambda(\mu, \nu)$. Then*

(a)

$$\inf_{P \in \Lambda(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{P}_r(\mathbb{R}^d)} C(x, p) P(dx, dp) = \inf_{\nu' \leq_c \nu} \inf_{\pi \in \Pi(\mu, \nu')} \int_{\mathcal{X} \times \mathbb{R}^d} C(x, \delta_y) \pi(dx, dy).$$

(b) In addition, if C is lower semicontinuous and satisfies (22), then

$$\inf_{P \in \Lambda(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{P}_r(\mathbb{R}^d)} C(x, p) P(dx, dp)$$

$$= \sup \left\{ \nu(\psi) + \mu(R_C \psi) : \psi \in C_r(\mathbb{R}^d) \text{ and concave} \right\}$$

$$= \sup \left\{ \mu(\varphi) + \nu(\psi) : \varphi \in C_r(\mathcal{X}), \text{concave } \psi \in C_r(\mathbb{R}^d), \right.$$

$$\left. \varphi(x) + \psi(y) \leq C(x, \delta_y) \, \forall (x, y) \in \mathcal{X} \times \mathbb{R}^d \right\}$$

$$= \sup \left\{ \mu(\varphi) + \nu(\psi) : \varphi \in C_r(\mathcal{X}), \psi \in C_r(\mathbb{R}^d), \Delta : \mathbb{R}^d \to \mathbb{R}^d, \right.$$

$$\left. \varphi(x) + \psi(z) + \Delta(y)(z - y) \leq C(x, \delta_y) \, \forall (x, y, z) \in \mathcal{X} \times \mathbb{R}^d \times \mathbb{R}^d \right\},$$

where $R_C \psi$ is defined as in Theorem 3.5.

Proof of Proposition 3.11. We first show (a). Let $\pi \in \Pi(\mu, \nu')$ with $\nu' \leq_c \nu$. By Strassen's theorem there is a martingale coupling $\pi^M \in \Pi_M(\nu', \nu)$, where $\Pi_M(\nu', \nu) := \{\pi' \in \Pi(\nu', \nu) : b(\pi'_x) =$

x μ-a.s.}. Define

$$P(dx, dp) := \mu(dx) \int_{\mathbb{R}^d} \delta_{\pi_y^M}(dp)\, \pi_x(dy).$$

Evidently, the first marginal of P is μ. Then the next line of computations establishes $P \in \Lambda(\mu, \nu)$ thanks to (27):

$$\text{proj}_*^2(\hat{I}(P)) = \int_{\mathcal{X} \times \mathbb{R}^d} \pi_y^M\, \pi(dx, dy) = \int_{\mathbb{R}^d} \pi_y^M\, \nu'(dy) = \text{proj}_*^2(\pi^M) = \nu.$$

Since C is barycentric, we find

$$\int_{\mathcal{X} \times \mathcal{P}_1(\mathbb{R}^d)} C(x, p)\, P(dx, dp) = \int_{\mathcal{X} \times \mathcal{P}_1(\mathbb{R}^d)} C(x, \pi_y^M)\, \pi(dx, dy) = \int_{\mathcal{X} \times \mathbb{R}^d} C(x, \delta_y)\, \pi(dx, dy). \quad (46)$$

We readily derive from (46) that the r.h.s. in item (a) dominates the l.h.s.

To derive the reverse inequality, we note that any $P \in \Lambda(\mu, \nu)$ induces a measure $\pi \in \mathcal{P}(\mathcal{X} \times \mathbb{R}^d)$ with second marginal $\nu' \leq_c \nu$, given by $\pi := ((x, p) \mapsto (x, b(p)))_* P$, so that

$$\int_{\mathcal{X} \times \mathcal{P}_1(\mathbb{R}^d)} C(x, p)\, P(dx, dp) = \int_{\mathcal{X} \times \mathcal{P}_1(\mathbb{R}^d)} C\left(x, \delta_{b(p)}\right) P(dx, dp)$$

$$= \int_{\mathcal{X} \times \mathbb{R}^d} C(x, \delta_y)\, \pi(dx, dy). \quad (47)$$

To see that $\text{proj}_*^2 \pi = \nu'$ is in convex order with ν, we pick any convex $\theta : \mathbb{R}^d \to \mathbb{R}$ and find that

$$\nu'(\theta) = \int_{\mathcal{X} \times \mathcal{P}_1(\mathbb{R}^d)} \theta(b(p))\, P(dx, dp) \leq \int_{\mathcal{X} \times \mathcal{P}_1(\mathbb{R}^d)} p(\theta)\, P(dx, dp) = \nu(\theta),$$

by Jensen's inequality, which means that $\nu' \leq_c \nu$. Hence, we obtain by (47) that the l.h.s. in item (a) dominates the r.h.s., and therefore we have equality.

Now, we prove (b). To see the first equality, note that by Theorem 3.5 it suffices to verify that the supremum in the r.h.s. of (31) can be restricted to concave $\psi \in C_r(\mathbb{R}^d)$. To this end, let $\psi \in C_r(\mathbb{R}^d)$ and ψ^{**} be its convex envelope. For any $y \in \mathbb{R}^d$, we have by Jensen's inequality

$$\psi^{**}(y) = \inf_{p \in \mathcal{P}_r(\mathbb{R}^d),\, b(p)=y} p(\psi^{**}) \leq \inf_{p \in \mathcal{P}_r(\mathbb{R}^d),\, b(p)=y} p(\psi).$$

On the other hand, the right-hand side is convex and dominated by ψ, whence the inequality is actually an equality. Therefore,

$$R_C \psi(x) = \inf_{p \in \mathcal{P}_r(\mathbb{R}^d)} -p(\psi) + C(x, p) = \inf_{y \in \mathbb{R}^d} C(x, \delta_y) + \inf_{p \in \mathcal{P}_r(\mathbb{R}^d),\, b(p)=y} -p(\psi)$$

$$= \inf_{y \in \mathbb{R}^d} C(x, \delta_y) + (-\psi)^{**}(y) = R_C(-(-\psi)^{**})(x). \quad (48)$$

Since $-(-\psi)^{**}$ is concave and dominates $\psi \in C_r(\mathbb{R}^d)$, we either have that $-(-\psi)^{**} \in C_r(\mathbb{R}^d)$ or $-(-\psi)^{**} = \infty$. If we are in the latter case, we have by (48) that $R_C\psi = -\infty$, thus, $\nu(\psi) + \mu(R_C\psi) = -\infty$. Hence, we may assume w.l.o.g. that $-(-\psi)^{**} \in C_r(\mathbb{R}^d)$ which yields

$$\nu(\psi) + \mu(R_C\psi) \leq \nu(-(-\psi)^{**}) + \mu(-(-\psi)^{**}),$$

and conclude that we may restrict to concave functions in the r.h.s. of (31).

The second equality follows from the first equality with the same line of argument as in the proof of Theorem 3.5.

For the third equality, we will show that any admissible pair in the second supremum of item (b) admits a better admissible pair in the third supremum, and vice versa. To this end, consider any pair $(\varphi, \psi) \in C_r(\mathcal{X}) \times C_r(\mathbb{R}^d)$. If ψ is concave, then there exists a measurable selection $\Delta : \mathbb{R}^d \to \mathbb{R}^d$ of the subgradient of $-\psi$, that is,

$$\varphi(x) + \psi(z) + \Delta(y)(z - y) \leq \varphi(x) + \psi(y) \quad \forall (x, y, z) \in \mathcal{X} \times \mathbb{R}^d \times \mathbb{R}^d. \tag{49}$$

We derive from (49) that (φ, ψ, Δ) is admissible in the last supremum whenever (φ, ψ) is admissible in the second supremum.

On the other hand, if there is a measurable $\Delta : \mathbb{R}^d \to \mathbb{R}^d$ with

$$\varphi(x) + \psi(z) + \Delta(y)(z - y) \leq C(x, \delta_y) \quad \forall (x, y, z) \in \mathcal{X} \times \mathbb{R}^d \times \mathbb{R}^d,$$

we can define the concave function $\tilde{\psi}(z) := \inf\{C(x, \delta_y) - \varphi(x) - \Delta(y)(z - y) : (x, y) \in \mathcal{X} \times \mathbb{R}^d\}$. Since $\tilde{\psi}$ is finitely valued, dominates ψ and is concave, we find that $\tilde{\psi} \in C_r(\mathbb{R}^d)$. In particular, $(\varphi, \tilde{\psi})$ is admissible for the second supremum which concludes the proof. $\square$

Proof of Theorem 1.3 *and Theorem* 1.6. Theorem 1.6 follows directly from Theorem 3.9 and Proposition 3.11.

As already noticed in the proof of Proposition 1.5, the robust optimization problem in (7) can be reformulated as in (41). Finally Theorem 1.3 is a consequence of Proposition 3.3 and Proposition 3.11. $\square$

3.4 | Primal and dual optimizers for convex barycentric costs

In this section we characterize primal and dual optimizers in the case of vanilla derivatives as announced in Section 2.2. Hence we consider costs of the form

$$C(x, p) = \theta(b(p) - x), \tag{50}$$

with $\theta : \mathbb{R} \to \mathbb{R}$ convex, and investigate primal and dual optimizers for the problems

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}} \theta(b(\pi_x) - x) \, \mu(dx), \tag{51}$$

$$\sup_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}} \theta(b(\pi_x) - x)\, \mu(dx), \tag{52}$$

where $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$.

Corollary 3.12. *The optimization problems (51) and (52) satisfy*

$$\inf_{\pi \in \Pi(\mu,\nu)} \int \theta(b(\pi_x) - x)\, \mu(dx) = \sup_{\psi \in C_1(\mathbb{R}),\, concave} \nu(\psi) + \mu(R_C \psi), \tag{53}$$

$$\sup_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}} \theta(b(\pi_x) - x)\, \mu(dx) = \sup_{\pi \in \Pi(\mu,\nu)} \int \theta(y - x)\, \pi(dx, dy) = \inf_{\psi \in C_1(\mathbb{R}),\, convex} \nu(\psi) + \mu(\overline{R}_C \psi), \tag{54}$$

where $\overline{R}_C \psi(x) := \sup_{y \in \mathbb{R}} \theta(y - x) - \psi(y)$ and $R_C \psi(x) = \inf_{y \in \mathbb{R}} \theta(y - x) - \psi(y)$.

Proof. For $C(x, p) = \theta(x - b(p))$ where θ is convex, we have by Theorem 2.9 in Backhoff-Veraguas et al. (2019) that (WOT) = (WOT'). Hence, (53) is readily deduced from item (b) of Proposition 3.11.

To see (54), we note that by Remark 3.10

$$\sup_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}} \theta(b(\pi_x) - x)\, \mu(dx) = \sup_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}} \theta(y - x)\, \pi(dx, dy), \tag{55}$$

that is, the problem reduces to a classical optimal transport problem. Then (54) follows again from item (b) of Proposition 3.11. □

Let us now consider primal optimizers of (51) and (52). By Proposition 3.11 we have

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}} \theta(b(\pi_x) - x)\, \mu(dx) = \inf_{\eta \leq_c \nu} \inf_{\pi \in \Pi(\mu,\eta)} \int_{\mathbb{R}} \theta(y - x)\, \pi(dx, dy). \tag{56}$$

A minimal measure on $\mathbb{R}$ for the right-hand side, denoted by $\nu^* \leq_c \nu$, is given by the image of μ under the weak monotone rearrangement $\underline{T}$ of μ and ν, see Backhoff-Veraguas et al. (2020). Note that

$$\underline{T} = F_{\nu^*}^{-1} \circ F_\mu \quad \mu\text{-a.s.,}$$

where F_p and F_p^{-1} denote the cumulative distribution function and the generalized inverse distribution function, resp. of the probability $p \in \mathcal{P}(\mathbb{R})$.

Proposition 3.13. *A primal optimizer of (51) is given by the coupling $\mu(dx)\, \pi^M_{\underline{T}(x)}(dy)$ where $\pi^M \in \Pi_M(\mu, \underline{T}(\mu))$ and $\underline{T}$ is the weak monotone rearrangement of μ and ν. A primal optimizer of (52) is given by the anticomonotone coupling.*

Proof. The first statement can be found in Theorem 3.1 of Backhoff-Veraguas et al. (2020) and the second in Theorem 3.1.2 of Rachev and Rüschendorf (1998). □

Before turning to the description of dual optimizers of (51) and (52), we define potential candidates in the subsequent lemma. We remark that the construction in Lemma 3.14 is not novel, but is included for the sake of completeness as the authors are unaware of a fitting reference. A dual optimizer of the corresponding weak transport problem is constructed in Shu (2016) under the assumption that T is strictly increasing. The author remarks that the assumption is simply to avoid technicalities.

Lemma 3.14. *Let $\theta : \mathbb{R} \to \mathbb{R}$ be convex, $T : I \to \mathbb{R}$ a map where $I \subseteq \mathbb{R}$, $[a, b] = \overline{co(T(I))}$, and $y_0 \in (a, b)$.*

(a) Assume that T is nondecreasing and define

$$\underline{S}(y) := \inf\{x \in I : T(x) \geq y\}, \quad y \in (a, b), \tag{57}$$

$$\underline{\psi}(y) := \int_{y_0}^{y} \partial_{-}\theta(z - \underline{S}(z)) \, dz, \quad y \in (a, b). \tag{58}$$

Then $\underline{\psi}$ is continuous on (a, b) and satisfies for all $x \in I$, $y' \in [T(x-), T(x)] \cap [a, b]$ that

$$-\underline{\psi}(y') + \theta(y' - x) = \inf_{y \in [a,b]} -\underline{\psi}(y) + \theta(y - x). \tag{59}$$

(b) Assume that T is nonincreasing and define

$$\overline{S}(y) := \sup\{x \in I : T(x) \geq y\}, \quad y \in (a, b), \tag{60}$$

$$\overline{\psi}(y) := \int_{y_0}^{y} \partial_{-}\theta(z - \overline{S}(z)) \, dz, \quad y \in (a, b). \tag{61}$$

Then $\overline{\psi}$ is continuous on (a, b) and satisfies for all $x \in I$, $y' \in [T(x+), T(x)] \cap [a, b]$ that

$$-\overline{\psi}(y') + \theta(y' - x) = \sup_{y \in [a,b]} \overline{\psi}(y) + \theta(y - x). \tag{62}$$

Proof. Since θ is convex, it is locally absolutely continuous, thus

$$\theta(y') - \theta(y) = \int_{y}^{y'} \partial_{-}\theta(z) \, dz,$$

where we write $\partial_-\theta$ for the left derivative of θ, which is nondecreasing. For fixed $x \in I$ we introduce

$$\underline{f}(z) := -\underline{\psi}(z) + \theta(z - x) \quad \text{and} \quad \overline{f}(z) := -\overline{\psi}(z) + \theta(z - x),$$

where $z \in (a, b)$. Using that convex functions are almost surely differentiable, we have dz-almost surely on (a, b)

$$\frac{d}{dz}\underline{f}(z) = -\partial_-\theta(z - \underline{S}(z)) + \partial_-\theta(z - x)$$

$$\frac{d}{dz}\overline{f}(z) = -\partial_-\theta(z - \overline{S}(z)) + \partial_-\theta(z - x).$$

$$(63)$$

The maps $\underline{S}$ and $\overline{S}$ are nonincreasing and nondecreasing, respectively. If T is monotone, then for any $z \in (a, b)$ there are $\underline{x}, \overline{x} \in I$ such that

$$T \text{ nondecreasing}: \quad x \in (-\infty, \underline{x}] \cap I \Rightarrow T(x) < z; \quad \text{and} \quad x \in [\overline{x}, \infty) \cap I \Rightarrow T(x) > z;$$

$$T \text{ nonincreasing}: \quad x \in (-\infty, \underline{x}] \cap I \Rightarrow T(x) > z; \quad \text{and} \quad x \in [\overline{x}, \infty) \cap I \Rightarrow T(x) < z.$$

Hence, $\underline{S}$ and $\overline{S}$ are real-valued, which also shows that $\underline{\psi}$ and $\overline{\psi}$ are well-defined and continuous on (a, b). From now on, we implicitly assume that T is nondecreasing when talking about $\underline{f}$ and $\underline{S}$, whereas we assume that T is nonincreasing when talking about $\overline{f}$ and $\overline{S}$. We have

$$z \in (T(x-), T(x)) \Rightarrow x \le \underline{S}(z) \le \underline{S}(T(x)) \le x \iff x = \underline{S}(z),$$

$$z \in (T(x+), T(x)) \Rightarrow x \ge \overline{S}(z) \ge \overline{S}(T(x)) \ge x \iff x = \overline{S}(z).$$

Therefore, we obtain

$$\frac{d}{dz}\underline{f}|_{(T(x-),T(x))} = 0 \Rightarrow \underline{f}|_{[T(x-),T(x)]} = \underline{f}(T(x)),$$

$$\frac{d}{dz}\overline{f}|_{(T(x+),T(x))} = 0 \Rightarrow \overline{f}|_{[T(x+),T(x)]} = \overline{f}(T(x)).$$

It is now sufficient to show that

$$\frac{d}{dz}\underline{f}|_{(a,T(x)]} \le 0 \quad \text{and} \quad \frac{d}{dz}\underline{f}|_{[(T(x),b)} \ge 0; \quad \frac{d}{dz}\overline{f}|_{(a,T(x)]} \ge 0 \quad \text{and} \quad \frac{d}{dz}\overline{f}|_{[T(x),b)} \le 0. \quad (64)$$

Due to monotonicity and by definition of $\underline{S}$ and $\overline{S}$, we find:

$$z \in (a, T(x)) \Rightarrow \begin{cases} \underline{S}(z) \le \underline{S}(T(x)) \le x \Rightarrow & z - \underline{S}(z) \ge z - x, \\ x \le \overline{S}(z) & \Rightarrow \quad z - \overline{S}(z) \le z - x, \end{cases}$$

$$z \in (T(x), b) \Rightarrow \begin{cases} x \le \underline{S}(z) & \Rightarrow \quad z - x \ge z - \underline{S}(z), \\ \overline{S}(z) \le \overline{S}(T(x)) \le x \Rightarrow & z - x \le z - \overline{S}(z). \end{cases}$$

$$(65)$$

As the left derivative of a convex function is nondecreasing, (64) follows from (63) and (65). $\qquad\square$

In what follows we provide a (semi-)explicit representation of the dual optimizers.

Proposition 3.15. *Assume there exist $f \in L^1(\mu)$, $g \in L^1(\nu)$ such that $\theta(y - x) \le f(x) + g(y)$. Then the right-hand side of (53) is attained by $\underline{\psi}$, given in (58), for $T = \underline{T}$ the weak monotone rearrangement of μ and ν.*

Assume there exist $\tilde{f} \in L^1(\mu)$, $\tilde{g} \in L^1(\nu)$ such that $\theta(y - x) \ge \tilde{f}(x) + \tilde{g}(y)$. Then the right-hand side of (54) is attained by $\overline{\psi}$, given in (61), for $T(x) = \overline{T}(x) := F_\nu^{-1}(1 - F_\mu(x))$.

Proof. Let π^* be a minimizer of (51) given as in Proposition 3.13 by

$$\pi^*(dx, dy) := \mu(dx)\,\pi^M_{\underline{T}(x)}(dy),$$

where π^M is an arbitrary martingale coupling with marginals μ and ν^*, and $\underline{T}$ is nondecreasing and 1-Lipschitz. From (Backhoff-Veraguas et al., 2020, Theorem 1.3) we have that

$$\pi^M_{\underline{T}(x)}\big(\{y \in \mathbb{R} : \underline{T}(x) - x = \underline{T}(y) - y\}\big) = 1 \quad \mu\text{-almost surely.} \tag{66}$$

Since $\underline{T}$ is nondecreasing and 1-Lipschitz, the map $x \mapsto \underline{T}(x) - x$ is nonincreasing and

$$\{y \in \mathbb{R} : \underline{T}(x) - x = \underline{T}(y) - y\}$$

is a closed interval for every $x \in \mathbb{R}$. For z, z' in the interior of $\underline{T}(\mathbb{R})$ there are minimal $x, x' \in \mathbb{R}$ with $\underline{T}(x) = z$ and $\underline{T}(x') = z'$. Therefore, $\underline{S}(z) = x$, $\underline{S}(z') = x'$ (where $\underline{S}$ is as in Lemma 3.14), and

$$z < z' \Rightarrow z - \underline{S}(z) = \underline{T}(x) - x \ge \underline{T}(x') - x' = z' - \underline{S}(z'),$$

that is, $z \mapsto z - \underline{S}(z)$ is nonincreasing, from which we deduce concavity of $\underline{\psi}$. Furthermore, $\underline{\psi}$ restricted to $\{y \in \mathbb{R} : \underline{T}(y) - y = \underline{T}(x) - x\}$ is linear, whereby we find by (66) that μ-almost surely

$$\int_{\mathbb{R}} \underline{\psi}(y)\,\pi^M_{\underline{T}(x)}(dy) = \underline{\psi}\Big(b\big(\pi^M_{\underline{T}(x)}\big)\Big) = \underline{\psi}(\underline{T}(x)).$$

Therefore, by Lemma 3.14 (a),

$$\int_{\mathbb{R}} \underline{\psi}(y)\,\pi^M_{\underline{T}(x)}(dy) + R_C\underline{\psi}(x) = \theta(\underline{T}(x) - x). \tag{67}$$

If $\theta(y - x) \le f(x) + g(y)$ for $f \in L^1(\mu)$ and $g \in L^1(\nu)$, then the μ-integral of $R_C\underline{\psi}$ and the ν-integral of $\underline{\psi}$ are well-defined in $[-\infty, \infty)$. Due to concavity of $\underline{\psi}$ and as ν^* is in convex order smaller than ν, we observe that the ν^*-integral of $\underline{\psi}$ is well-defined in $[-\infty, \infty)$. The function $R_C\underline{\psi}$ is convex (as the infimum over a jointly convex function) yielding $R_C\underline{\psi} \in L^1(\mu)$. Therefore, by (67)

$$\int_{\mathbb{R}\times\mathbb{R}} \theta(y - x)\,\pi^*(dx, dy) = \int_{\mathbb{R}} \theta(\underline{T}(x) - x)\,\mu(dx) = \int_{\mathbb{R}} \underline{\psi}(\underline{T}(x)) + R_C\underline{\psi}(x)\,\mu(dx)$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} \psi(y) \, \pi^M_{\underline{T}(x)}(dy) + R_C \underline{\psi}(x) \, \mu(dx) = \nu(\underline{\psi}) + \mu(R_C \underline{\psi}),$$

hence $\underline{\psi}$ is a dual optimizer.

To see the second assertion, note that the primal optimizer of the upper bound, (52), is given by the anticomonotone coupling $\pi^a \in \Pi(\mu, \nu)$. We have

$$\pi^a_x([\overline{T}(x+), \overline{T}(x)]) = 1 \, \mu\text{-almost surely.}$$

Recalling the definition of $\overline{R}_C \psi$ given in Corollary 3.12, we find $\overline{\psi}(y) + \overline{R}_C \psi(y) = \theta(y - x) \, \pi^a$-almost surely. Moreover, $\overline{\psi}$ is convex on its domain, since $z \mapsto z - \overline{S}(z)$ (where $\overline{S}$ is as in Lemma 3.14) and $\partial_- \theta(z)$ are nondecreasing. If $\theta(y - x) \geq \tilde{f}(x) + \tilde{g}(y)$ for $\tilde{f} \in L^1(\mu)$ and $\tilde{g} \in L^1(\nu)$, then the integral of $\overline{\psi}$ w.r.t. ν and the integral of $\overline{R}_C \psi$ w.r.t. μ are well-defined in $(-\infty, \infty]$. Therefore, $\overline{\psi}$ is a dual optimizer, since we find by Lemma 3.14 (b) that

$$\int_{\mathbb{R} \times \mathbb{R}} \theta(y - x) \, \pi^a(dx, dy) = \int_{\mathbb{R} \times \mathbb{R}} \overline{\psi}(y) + \overline{R}_C \psi(x) \, \pi^a(dx, dy) = \nu(\overline{\psi}) + \mu(\overline{R}_C \psi).$$

$\square$

Example 3.16. In the setting of Corollary 3.12, let θ be given by $y \mapsto y^+$, and assume that μ and ν are equivalent to the Lebesgue measure restricted to some interval I. Then the left derivative of θ satisfies

$$\partial_- \theta(y - x) = \begin{cases} 0 & y \leq x, \\ 1 & \text{else.} \end{cases}$$

We have already seen in the proof of Corollary 3.12 that $y \mapsto y - \underline{S}(y)$ and $y \mapsto y - \overline{S}(y)$ are nonincreasing and nondecreasing, respectively. Therefore, there are uniquely determined points $\underline{a}, \overline{a} \in [-\infty, \infty]$ such that

$$y - \underline{S}(y) > 0 \quad \forall y < \underline{a} \quad \text{and} \quad y - \overline{S}(y) > 0 \quad \forall y > \overline{a}.$$

If $\underline{a} \in \mathbb{R}$, then

$$\underline{\psi}(y) = \int_{y_0}^y 1_{(-\infty, \underline{a})}(z) \, dz = (\underline{a} - y_0)^+ - (\underline{a} - y)^+,$$

$$R_C(x) = \inf_{y \in \mathbb{R}} (\underline{a} - y)^+ + (y - x)^+ - (\underline{a} - y_0)^+ = (\underline{a} - x)^+ - (\underline{a} - y_0)^+,$$

whereas if $\overline{a} \in \mathbb{R}$, then

$$\overline{\psi}(y) = \int_{y_0}^y 1_{(\overline{a}, \infty)}(z) \, dz = (y - \overline{a})^+ - (y_0 - \overline{a})^+,$$

$$\overline{R}_C(x) = \sup_{y \in \mathbb{R}} -(y - \overline{a})^+ + (y_0 - \overline{a})^+ + (y - x)^+ = -(x - \overline{a})^+ + (y_0 - \overline{a})^+.$$

Note that the constants are canceling out. Therefore, removing the constants and differentiating yields the triplets

$$(\underline{\varphi}(x_1), \underline{\psi}(y_2), \underline{\Delta}(y_1)) := ((\underline{a} - x_1)^+, -(\underline{a} - y_2)^+, -\partial_-\theta(\underline{a} - y_1)), \quad (68)$$

$$(\overline{\varphi}(x_1), \overline{\psi}(y_2), \overline{\Delta}(y_1)) := (-(x_1 - \overline{a})^+, (y_2 - \overline{a})^+, \partial_-\theta(y_1 - \overline{a})), \quad (69)$$

for $(x_1, y_1, y_2) \in \mathbb{R}^3$, which are optimizers of

$$\sup\{\mu(\varphi) + \nu(\psi) : \varphi(x_1) + \psi(y_2) + \Delta(y_1)(y_2 - y_1) \leq (y_1 - x_1)^+ \text{ for all } x_1, y_1, y_2\},$$

$$\inf\{\mu(\varphi) + \nu(\psi) : \varphi(x_1) + \psi(y_2) + \Delta(y_1)(y_2 - y_1) \geq (y_1 - x_1)^+ \text{ for all } x_1, y_1, y_2\},$$

respectively.

No data was used for this research.

ORCID

Beatrice Acciaio https://orcid.org/0000-0002-8138-2578
Mathias Beiglböck https://orcid.org/0000-0003-3787-2155
Gudmund Pammer https://orcid.org/0000-0003-2494-8739

ENDNOTES

[1] All our results are still valid if T_1-bonds are also dynamically traded.

[2] In principle we could allow for the hedge Δ to depend also on the value of X_1, but as part of our results we will obtain that this does not change the actual value of the superhedging problem.

[3] In fact it suffices to consider functions Δ which are increasing.

[4] Two measures $\nu', \nu \in \mathcal{P}_1(\mathbb{R}^d)$ are in convex order, denoted by $\nu' \leq_c \nu$, if $\int \theta d\nu' \leq \int \theta d\nu$ for all convex $\theta : \mathbb{R}^d \to \mathbb{R}$.

[5] The supremum in the duality of (Backhoff-Veraguas et al., 2019, Theorem 3.1) is taken over all functions $\psi \in C_r(\mathcal{Y})$ which are bounded from above. Therefore, $R_{C^K}\psi$ is bounded from below and its μ-integral is well-defined. We circumvent this by defining the μ-integral of $R_{C^K}\psi$ appropriately, that is $-\infty$, whenever it is not well-defined.

REFERENCES

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

Acciaio, B., & Grbac, Z. (2020). Consistency of traded option prices and absence of arbitrage in the presence of stochastic interest rates. https://people.math.ethz.ch/~beacciaio/pdfs/DH_BZ.pdf.

Alfonsi, A., Corbetta, J., & Jourdain, B. (2019). Sampling of one-dimensional probability measures in the convex order and computation of robust option price bounds. *International Journal of Theoretical and Applied Finance*, 22(3).

Alfonsi, A., Corbetta, J., & Jourdain, B. (2020). Sampling of probability measures in the convex order by Wasserstein projection. *Annales de l'Institut Henri Poincaré B, Probability and Statistics*, 56(3), 1706–1729.

Alibert, J.-J., Bouchitté, G., & Champion, T. (2019). A new class of costs for optimal transport planning. *European Journal of Applied Mathematics*, 30(6), 1229–1263.

Backhoff-Veraguas, J., Beiglböck, M., & Pammer, G. (2019). Existence, duality, and cyclical monotonicity for weak transport costs. *Calculus of Variations and Partial Differential Equations*, 58(6), 203.

Backhoff-Veraguas, J., Beiglböck, M., & Pammer, G. (2020). Weak monotone rearrangement on the line. *Electronic Communications in Probability*, *25*.

Backhoff-Veraguas, J. D., & Pammer, G. (2021a). Applications of weak transport theory. *Bernoulli*, to appear.

Backhoff-Veraguas, J. D., & Pammer, G. (2021b). Stability of martingale optimal transport and weak optimal transport. *The Annals of Applied Probability*, to appear.

Bartl, D., Beiglböck, M., & Pammer, G. (2021). The Wasserstein space of stochastic processes. ArXiv e-prints.

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

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

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

Bertsekas, D. P., & Shreve, S. E. (1978). *Stochastic optimal control*, volume 139 of *Mathematics in Science and Engineering*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London. The discrete time case.

Bouchard, B., & Nutz, M. (2015). Arbitrage and duality in nondominated discrete-time models. *The Annals of Applied Probability*, *25*(2), 823–859.

Breeden, D. T., & Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, 621–651.

Burzoni, M., Frittelli, M., & Maggis, M. (2017). Model-free superhedging duality. *The Annals of Applied Probability*, *27*(3), 1452–1477.

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

Cheridito, P., Kiiski, M., Prömel, D. J., & Soner, H. M. (2021). Martingale optimal transport duality. *Mathematische Annalen*, *379*(3), 1685–1712.

Cheridito, P., Kupper, M., & Tangpi, L. (2017). Duality formulas for robust pricing and hedging in discrete time. *SIAM Journal on Financial Mathematics*, *8*(1), 738–765.

Cox, A. M. G., Obłój, J., & Touzi, N. (2019). The Root solution to the multi-marginal embedding problem: An optimal stopping and time-reversal approach. *Probability Theory and Related Fields*, *173*(1-2), 211–259.

Daskalakis, C., Deckelbaum, A., & Tzamos, C. (2017). Strong duality for a multiple-good monopolist. *Econometrica*, *85*(3), 735–767.

Davis, M. H., & Hobson, D. G. (2007). The range of traded option prices. *Mathematical Finance*, *17*(1), 1–14.

Dolinsky, Y., & Soner, H. M. (2014). Martingale optimal transport and robust hedging in continuous time. *Probability Theory and Related Fields*, *160*(1-2), 391–427.

Eder, M. (2019). Compactness in adapted weak topologies. arXiv e-prints, arXiv:1905.00856.

Fadina, T., & Schmidt, T. (2019). Default ambiguity. *Risks*, *7*(2), 64.

Fathi, M., Gozlan, N., & Prodhomme, M. (2020). A proof of the Caffarelli contraction theorem via entropic regularization. *Calculus of Variations and Partial Differential Equations*, *59*(3), 96.

Filipovic, D. (2009). *Term-Structure Models. A Graduate Course*. Springer.

Galichon, A., Henry-Labordère, P., & Touzi, N. (2014). A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Annals of Applied Probability*, *24*(1), 312–336.

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

Gozlan, N., Roberto, C., Samson, P.-M., Shu, Y., & Tetali, P. (2018). Characterization of a class of weak transport-entropy inequalities on the line. *Annales de l Institut Henri Poincaré Probabilités et Statistiques*, *54*(3), 1667–1693.

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

Hobson, D. (1998). Robust hedging of the lookback option. *Finance and Stochastics*, *2*, 329–347.

Hobson, D. (2011). 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*. (pp. 267–318). Berlin: Springer.

Hobson, D., & Neuberger, A. (2012). Robust bounds for forward start options. *Mathematical Finance*, *22*(1), 31–56.

Hölzermann, J. (2018). The hull-white model under knightian uncertainty about the volatility. arXiv preprint arXiv:1808.03463.

Hölzermann, J. (2020). Pricing interest rate derivatives under volatility uncertainty. arXiv preprint arXiv:2003.04606.

Hölzermann, J., & Lin, Q. (2019). Term structure modeling under volatility uncertainty. arXiv preprint arXiv:1904.02930.

Kallenberg, O. (2002). Foundations of modern probability . *Probability and its Applications (New York)*. (2nd ed.). New York: Springer-Verlag.

Obłój, J. (2004). The Skorokhod embedding problem and its offspring. *Probability Surveys*, *1*, 321–390.

Pratelli, A. (2007). On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation. *Annales de l'I.H.P. Probabilités et Statistiques*, *43*(1), 1–13.

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

Shu, Y. (2016). From Hopf-Lax formula to optimal weak transfer plan. arXiv e-prints arXiv:1609.03405v1.

Shu, Y. (2020). From Hopf–Lax formula to optimal weak transfer plan. *SIAM Journal on Mathematical Analysis*, *52*(3), 3052–3072.

Villani, C. (2009). *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer.