Lipschitz continuity of the Wasserstein projections in the convex order on the line

Benjamin Jourdain[*]   William Margheriti[†]   Gudmund Pammer[‡]

Abstract

Wasserstein projections in the convex order were first considered in the framework of weak optimal transport, and found applications in various problems such as concentration inequalities and martingale optimal transport. In dimension one, it is well-known that the set of probability measures with a given mean is a lattice w.r.t. the convex order. Our main result is that, contrary to the minimum and maximum in the convex order, the Wasserstein projections are Lipschitz continuity w.r.t. the Wasserstein distance in dimension one. Moreover, we provide examples that show sharpness of the obtained bounds for the 1-Wasserstein distance.

Keywords: optimal transport; weak optimal transport; projection; convex order.
MSC2020 subject classifications: 49Q22.
Submitted to ECP on August 23, 2022, final version accepted on April 2, 2023.

1 Introduction and main result

In mathematical finance, the risk neutral distributions at times T_1 and T_2 with $T_1 < T_2$ of the vector of discounted prices of d assets are in convex order. For an exotic option with payoff depending on the vectors of prices at times T_1 and T_2, robust price bounds are obtained by solving martingale optimal transport (MOT) problems [8, 14, 10]. Even when $d = 1$, the distributions of the asset price are only imperfectly known, since one has to recover them from prices (up to a bid-ask spread) of finitely many Call or Put options. Furthermore, to numerically compute the robust price bounds the two distributions are approximated by finitely supported measures which permits to reformulate the MOT problem as a standard finite linear programming problem (LPP). The convex order may be violated in these steps, which, in view of Strassen's theorem, turns the corresponding LPP infeasible. This necessitates the restoration of the convex order which motivates the study of Wasserstein projections in the convex order, see [2].

For $p \geq 1$, we denote the celebrated p-Wasserstein distance between μ and ν in the set $\mathcal{P}_p(\mathbb{R}^d)$ of probability measures on $\mathbb{R}^d$ with finite p-th moment by

$$\mathcal{W}_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx, dy) \right)^{1/p}, \tag{1.1}$$

---
[*]CERMICS, Ecole des Ponts, INRIA, Marne-la-Vallée, France. E-mail: benjamin.jourdain@enpc.fr
[†]E-mail: william.margheriti@gmail.com
[‡]ETH Zürich, Switzerland. E-mail: gudmund.pammer@math.ethz.ch

where we write $\Pi(\mu, \nu)$ for the set of couplings with marginals μ and ν. We say that μ is smaller than ν in the convex order and denote $\mu \leq_c \nu$ if

$$\forall f \colon \mathbb{R}^d \to \mathbb{R} \text{ convex}, \ \int_{\mathbb{R}^d} f(x) \, \mu(dx) \leq \int_{\mathbb{R}^d} f(x) \, \nu(dy). \tag{1.2}$$

Then metric projections w.r.t. $\mathcal{W}_p$ of μ (resp. ν) onto $\{\eta \in \mathcal{P}_p(\mathbb{R}^d) : \eta \leq_c \nu\}$ (resp. $\{\eta \in \mathcal{P}_p(\mathbb{R}^d) : \mu \leq_c \eta\}$) are called Wasserstein projections in the convex order.

When $d = 1$, explicit formulas for the quantile functions of the Wasserstein projection of μ (resp. ν) on the set of probability measures smaller than ν (resp. greater than μ) in the convex order are derived in [2]. We denote the associated measure by $\mathcal{I}(\mu, \nu)$ (resp. $\mathcal{J}(\mu, \nu)$) and recall that the quantile function F_μ^{-1} of a probability measure μ on $\mathbb{R}$ is the left-continuous pseudo-inverse of its cumulative distribution function F_μ. Then [2, Theorem 2.6 and Proposition 4.2] state, for $u \in (0, 1)$,

$$F_{\mathcal{I}(\mu,\nu)}^{-1}(u) = F_\mu^{-1}(u) - \partial_- \mathrm{co}(G)(u) \text{ and } F_{\mathcal{J}(\mu,\nu)}^{-1}(u) = F_\nu^{-1}(u) + \partial_- \mathrm{co}(G)(u), \tag{1.3}$$

where $G(u) := \int_0^u (F_\mu^{-1} - F_\nu^{-1})(v) \, dv$, co denotes the convex hull, and ∂_- the left-hand derivative. More specifically, we have $\mathcal{I}(\mu, \nu) \leq_c \nu$, $\mu \leq_c \mathcal{J}(\mu, \nu)$, and if $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$

$$\mathcal{W}_p(\mu, \mathcal{I}(\mu, \nu)) = \inf_{\eta \leq_c \nu} \mathcal{W}_p(\mu, \eta) \text{ and } \mathcal{W}_p(\mathcal{J}(\mu, \nu), \nu) = \inf_{\mu \leq_c \eta} \mathcal{W}_p(\eta, \nu).$$

When $p > 1$, $\mathcal{I}(\mu, \nu)$ and $\mathcal{J}(\mu, \nu)$ are the unique respective metric projections but uniqueness may fail when $p = 1$, c.f. [2, Remark 2.3]. Our main result reads as follows.

Theorem 1.1 (Lipschitz continuity)**.** *When $d = 1$ and $p \in [1, \infty)$, the Wasserstein projections $\mathcal{I}, \mathcal{J}$ are Lipschitz continuous. For $\mu, \nu, \mu', \nu' \in \mathcal{P}_p(\mathbb{R})$, we have*

$$\mathcal{W}_p(\mathcal{I}(\mu, \nu), \mathcal{I}(\mu', \nu')) \leq 2\mathcal{W}_p(\mu, \mu') + \mathcal{W}_p(\nu, \nu'), \tag{1.4}$$

$$\mathcal{W}_p(\mathcal{J}(\mu, \nu), \mathcal{J}(\mu', \nu')) \leq \mathcal{W}_p(\mu, \mu') + 2\mathcal{W}_p(\nu, \nu'). \tag{1.5}$$

1.1 Discussion on Wasserstein projections in dimension d and related problems

Wasserstein projections in the convex order have also been considered in general dimension $d \geq 1$. For $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ with $p > 1$, there exists by [2, Theorem 2.1] a unique $\mathcal{I}_p(\mu, \nu) \leq_c \nu$ such that $\mathcal{W}_p(\mu, \mathcal{I}_p(\mu, \nu)) = \inf_{\eta \leq_c \nu} \mathcal{W}_p(\mu, \eta)$ but, unless $d = 1$, $\mathcal{I}_p(\mu, \nu)$ may depend on p according to [2, Example 2.5]. Similarly, by [2, Theorem 4.1] there also exists a $\mathcal{W}_p$-projection $\mathcal{J}_p(\mu, \nu)$ of ν onto $\{\eta \in \mathcal{P}_p(\mathbb{R}^d) : \nu \leq_c \eta\}$, which is unique under the additional assumption that ν is absolutely continuous w.r.t. the Lebesgue measure. When $d = 1$, Theorem 1.1 is a generalization of [2, Propositions 3.1 and 4.3]. These propositions state that for probability measures $\mu, \nu, \mu', \nu' \in \mathcal{P}_p(\mathbb{R}^d)$ such that $\mu \leq_c \nu$, and $\mathcal{I}_p(\mu, \nu) = \mu$ and $\mathcal{J}_p(\mu, \nu) = \nu$, then (1.4) and (1.5) hold true when $\mathcal{I}$ and $\mathcal{J}$ are replaced with $\mathcal{I}_p$ and $\mathcal{J}_p$, respectively. Hence, by Theorem 1.1, it is possible for $d = 1$ to drop the convex ordering constraint $\mu \leq_c \nu$. The extension of Theorem 1.1 to dimensions $d > 1$ is to the authors' understanding an interesting open question.

Gozlan, Roberto, Samson, and Tetali [17] introduced a generalization of optimal transport, the weak optimal transport, in order to study measure concentration inequalities. The following barycentric weak optimal transport problem received in recent years special attention, see for example [17, 16, 15, 1, 2, 4, 5]: for $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, consider

$$\mathcal{V}_p^p(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d} \left| x - \int_{\mathbb{R}^d} y \, \pi_x(dy) \right|^p \mu(dx), \tag{1.6}$$

where we write $(\pi_x)_{x \in \mathbb{R}^d}$ for a disintegration kernel of π w.r.t. its μ-marginal: $\pi(dx, dy) = \mu(dx)\pi_x(dy)$. This problem has an intrinsic connection with the problem of finding Wasserstein projection. Indeed, we have that the values of $\mathcal{V}_p(\mu, \nu)$ and $\mathcal{W}_p(\mu, \mathcal{I}_p(\mu, \nu))$ coincide,

see [2, 4]. Moreover, if π^* is an optimizer of (1.6) then the image of the first marginal μ under the map $x \mapsto \int_{\mathbb{R}^d} \pi_x^*(y)\,dy$ is a minimizer of $\inf_{\eta \leq_c \nu} \mathcal{W}_p(\mu, \eta)$ and coincides with $\mathcal{I}_p(\mu, \nu)$ when $p > 1$. Therefore, when $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ are finitely supported, (1.6) can be used to compute the Wasserstein projection. In particular, $\mathcal{I}_2(\mu, \nu)$ can be computed by solving a quadratic optimization problem with linear constraints. We refer to [17, 3] for dual formulations of weak optimal transport problems with additional martingale constraints, to [4] for the existence of optimal couplings and necessary and sufficient optimality conditions, to [7] for continuity of their value function in terms of the marginal distributions μ and ν, and to [6] for applications of such problems. We point out the connection of Wasserstein projections to Cafarelli's contraction theorem that was discovered in [13]. Note that dual formulations of the minimization problems defining $\mathcal{I}_p(\mu, \nu)$ and $\mathcal{J}_p(\mu, \nu)$ have recently been studied by Kim and Ruan [20].

1.2 Wasserstein projections in dimension one

In dimension $d = 1$, Wasserstein distance and convex order both can be characterized in terms of quantile functions, which gives intuition why they appear in (1.3). It is well-known that the comonotonous coupling is an optimizer in (1.1):

$$\forall \mu, \nu \in \mathcal{P}_p(\mathbb{R}), \ \mathcal{W}_p(\mu, \nu) = \left(\int_0^1 |F_\mu^{-1}(u) - F_\nu^{-1}(u)|^p du \right)^{1/p} =: \|F_\mu^{-1} - F_\nu^{-1}\|_p. \qquad (1.7)$$

Moreover, by [21, Theorem 3.A.5], for $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ that share the same barycenter,

$$\mu \leq_c \nu \iff \forall u \in [0, 1], \ \int_0^u F_\mu^{-1}(v)\,dv \geq \int_0^u F_\nu^{-1}(v)\,dv. \qquad (1.8)$$

A complete geometric characterization of $\mathcal{I}(\mu, \nu)$ and $\mathcal{J}(\mu, \nu)$ is given in [5]. Note that (1.3) implies

$$\mathcal{W}_p(\mathcal{I}(\mu, \nu), \mu) = \mathcal{W}_p(\mathcal{J}(\mu, \nu), \nu) \text{ and } \mathcal{W}_p(\mathcal{I}(\mu, \nu), \nu) = \mathcal{W}_p(\mathcal{J}(\mu, \nu), \mu), \qquad (1.9)$$

where, according to [2, Corollary 4.4], the first equality still holds for $d \geq 2$ when $\mathcal{I}$ and $\mathcal{J}$ are replaced by $\mathcal{I}_p$ and $\mathcal{J}_p$ respectively. The next examples show that the constants in (1.4) and (1.5) are sharp for $p = 1$.

Example 1.2. Let $\mu \in \mathcal{P}_p(\mathbb{R})$ and ν be a Dirac measure. As ν is the only measure dominated by itself in the convex order, $\mathcal{I}(\mu, \nu) = \nu$ and, as a consequence of (1.9), $\mathcal{J}(\mu, \nu) = \mu$. When ν' is also a dirac mass, we deduce for any $\mu' \in \mathcal{P}_p(\mathbb{R})$ that

$$\mathcal{W}_p(\mathcal{I}(\mu, \nu), \mathcal{I}(\mu, \nu')) = \mathcal{W}_p(\nu, \nu') \quad \text{and} \quad \mathcal{W}_p(\mathcal{J}(\mu, \nu), \mathcal{J}(\mu', \nu)) = \mathcal{W}_p(\mu, \mu').$$

Hence the factor 1 multiplying $\mathcal{W}_p(\nu, \nu')$ in the right-hand side of (1.4) and multiplying $\mathcal{W}_p(\mu, \mu')$ in the right-hand side of (1.5) is optimal. $\diamond$

Example 1.3. We fix $\mu := \delta_0$ and define, for $\alpha \in (0, 1)$,

$$\nu^\alpha := (1 - \alpha)\delta_{-\alpha^2} + \alpha\delta_1.$$

We have $\mathcal{I}(\nu^\alpha, \nu^\alpha) = \nu^\alpha$ and $\mathcal{I}(\mu, \nu^\alpha) = \delta_{\alpha(1-\alpha(1-\alpha))}$, so that

$$\mathcal{W}_1(\mathcal{I}(\nu^\alpha, \nu^\alpha), \mathcal{I}(\mu, \nu^\alpha)) = 2 \left(\alpha + \alpha^2(\alpha(1 - \alpha) - 1) \right),$$
$$\mathcal{W}_1(\mu, \nu^\alpha) = \alpha + \alpha^2(1 - \alpha).$$

Then, an application of the de l'Hôpital rule yields $\lim_{\alpha \searrow 0} \frac{\mathcal{W}_1(\mathcal{I}(\nu^\alpha, \nu^\alpha), \mathcal{I}(\mu, \nu^\alpha))}{\mathcal{W}_1(\mu, \nu^\alpha)} = 2$. Hence, the factor 2 in (1.4) is optimal when $p = 1$. Since $\mathcal{J}(\delta_{\alpha(1-\alpha(1-\alpha))}, \nu^\alpha) = \nu^\alpha$ and $\mathcal{J}(\delta_{\alpha(1-\alpha(1-\alpha))}, \delta_0) = \delta_{\alpha(1-\alpha(1-\alpha))}$, we find in the same way that the factor 2 is also optimal in (1.5) when $p = 1$. $\diamond$

Example 1.4. Let μ, μ', ν, ν' be the probability measures with quantile functions:

$$F_\mu^{-1}(u) = u\mathbb{1}_{(0,\frac{1}{2}]}(u) + \tfrac{1+u}{2}\mathbb{1}_{(\frac{1}{2},1)}(u), \quad F_\nu^{-1}(u) = \tfrac{u}{2},$$

$$F_{\mu'}^{-1}(u) = u\mathbb{1}_{(0,\frac{1}{2}]}(u) + \tfrac{12+5u}{18}\mathbb{1}_{(\frac{1}{2},1)}(u), \quad F_{\nu'}^{-1}(u) = \tfrac{u}{3}\mathbb{1}_{(0,\frac{1}{2}]}(u) + \tfrac{u}{2}\mathbb{1}_{(\frac{1}{2},1)}(u).$$

We check that $F_{\mathcal{I}(\mu,\nu)}^{-1}(u) = \tfrac{u}{2}$ and $F_{\mathcal{I}(\mu',\nu')}^{-1}(u) = \tfrac{u}{3}\mathbb{1}_{(0,\frac{1}{2}]} + \tfrac{3+5u}{18}\mathbb{1}_{(\frac{1}{2},1)}(u)$, whence,

$$\mathcal{W}_p^p(\mathcal{I}(\mu,\nu), \mathcal{I}(\mu',\nu')) = \mathcal{W}_p^p(\mu,\mu') + \mathcal{W}_p^p(\nu,\nu')$$

with two positive summands. $\diamond$

1.3 On the convex-order lattice in dimension one

In dimension one, when μ and ν share the same barycenter, it is possible to restore convex ordering by using that $\mathcal{P}_1(\mathbb{R})$ is a complete lattice for the increasing and decreasing convex orders (see [19]). Both orders coincide with the convex order on the subset $\mathcal{P}_p^{x_0}(\mathbb{R})$ of $\mathcal{P}_p(\mathbb{R})$ consisting in probability measures with barycenter $x_0 \in \mathbb{R}$. On $\mathcal{P}_1^{x_0}(\mathbb{R})$, the minimum $\wedge_c$ and maximum $\vee_c$ can be expressed in terms of potential functions: the potential function of $\mu \in \mathcal{P}_1(\mathbb{R})$ is defined by

$$u_\mu(x) := \int_{\mathbb{R}} |x - y|\, \mu(dy).$$

For $\mu, \nu \in \mathcal{P}_1^{x_0}(\mathbb{R})$, $\mu \wedge_c \nu$ and $\mu \vee_c \nu$ are uniquely determined by

$$u_{\mu\wedge_c\nu} = \mathrm{co}(u_\mu \wedge u_\nu) \quad \text{and} \quad u_{\mu\vee_c\nu} = u_\mu \vee u_\nu.$$

On the domain $\mathcal{P}_p^{x_0}(\mathbb{R}) \times \mathcal{P}_p^{x_0}(\mathbb{R})$,

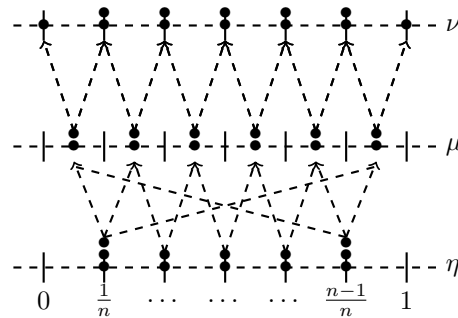$$(\mu,\nu) \mapsto \mu \wedge_c \nu \quad \text{and} \quad (\mu,\nu) \mapsto \mu \vee_c \nu$$

are continuous mappings into $\mathcal{P}_p^{x_0}(\mathbb{R})$: [9, Lemma 4.1] provides continuity for $p = 1$ and [2, Lemma 4.3], which ensures uniform integrability, permits to deduce continuity for general $p \geq 1$. However, unlike $\mathcal{I}$ and $\mathcal{J}$, the minimum and maximum in the convex order are not Lipschitz continuous.

Example 1.5. Consider for $n \geq 3$ the measures in $\mathcal{P}_p(\mathbb{R})$:

$$\nu := \frac{1}{2n}\delta_0 + \frac{1}{n}\sum_{i=1}^{n-1}\delta_{\frac{i}{n}} + \frac{1}{2n}\delta_1,$$

$$\mu := \frac{1}{n}\sum_{i=1}^{n}\delta_{\frac{2i-1}{2n}},$$

$$\eta := \frac{3}{2n}\delta_{\frac{1}{n}} + \frac{1}{n}\sum_{i=2}^{n-2}\delta_{\frac{i}{n}} + \frac{3}{2n}\delta_{\frac{n-1}{n}}.$$

Observe that, for the martingale kernel $K(x,dy) := \frac{1}{2}\delta_{x-\frac{1}{2n}} + \frac{1}{2}\delta_{x+\frac{1}{2n}}$, $\mu(dx)\,K(x,dy) \in \Pi(\mu,\nu)$. By Strassen's theorem [22] we find that $\mu \leq_c \nu$, $\eta \leq_c \tilde{\mu}$ where $\tilde{\mu}(dy) := \int_{x \in \mathbb{R}} K(x,dy)\,\eta(dx)$ is such that $\mu - \tilde{\mu} = \frac{1}{4n}\delta_{\frac{1}{2n}} + \frac{1}{4n}\delta_{\frac{2n-1}{2n}} - \frac{1}{4n}\delta_{\frac{3}{2n}} - \frac{1}{4n}\delta_{\frac{2n-3}{2n}}$ so that $\tilde{\mu} \leq_c \mu$. Hence $\eta \leq_c \mu \leq_c \nu$ and $\mu \vee_c \nu = \nu$, $\mu \vee_c \eta = \mu$, $\mu \wedge_c \nu = \mu$, $\mu \wedge_c \eta = \eta$. We compute

$$\mathcal{W}_p(\mu,\nu) = \mathcal{W}_p(\eta,\mu) = \frac{1}{2n} \quad \text{and} \quad \mathcal{W}_p(\eta,\nu) = \frac{1}{n^{1+1/p}},$$

from where we conclude that

$$\frac{\mathcal{W}_p(\mu\vee_c\nu, \mu\vee_c\eta)}{\mathcal{W}_p(\eta,\nu)} = \frac{\mathcal{W}_p(\mu\wedge_c\nu, \mu\wedge_c\eta)}{\mathcal{W}_p(\eta,\nu)} = \frac{n^{1/p}}{2}. \tag{1.10}$$

Consequently, $\wedge_c$ and $\vee_c$ are not Lipschitz continuous. $\diamond$

In particular, this example shows that for probability measures with the same barycenter, in general, the Wasserstein projections do not coincide with the minimum and maximum in the convex order. Beyond that, they satisfy the following order relation.

Proposition 1.6. *Let $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$ have the same barycenter. Then*

$$\mathcal{I}(\mu, \nu) \leq_c \mu \wedge_c \nu \quad \text{and} \quad \mathcal{J}(\mu, \nu) \geq_c \mu \vee_c \nu. \tag{1.11}$$

Proof. The map $u \mapsto F_{\mathcal{I}(\mu,\nu)}^{-1}(u) - F_\mu^{-1}(u)$ is non-increasing due to (1.3). Therefore, monotonicity of the integrand yields, for $u \in (0, 1)$ that

$$\tfrac{1}{u} \int_0^u (F_{\mathcal{I}(\mu,\nu)}^{-1}(v) - F_\mu^{-1}(v))\, dv \geq \int_0^1 (F_{\mathcal{I}(\mu,\nu)}^{-1}(v) - F_\mu^{-1}(v))\, dv = \int_\mathbb{R} y\,\nu(dy) - \int_\mathbb{R} x\,\mu(dx),$$

where the last equality comes from the inverse transform sampling and the fact that $\mathcal{I}(\mu, \nu)$ and ν share the same barycenter, which is a consequence of $\mathcal{I}(\mu, \nu) \leq_c \nu$. If μ and ν have the same barycenter, then $\mathcal{I}(\mu, \nu)$ shares this barycenter and we deduce by (1.8) that $\mathcal{I}(\mu, \nu) \leq_c \mu$, thus, $\mathcal{I}(\mu, \nu) \leq_c \mu \wedge_c \nu$. Analogously, we have $\mu \leq_c \mathcal{J}(\mu, \nu)$, and if μ and ν share the same barycenter, $\nu \leq_c \mathcal{J}(\mu, \nu)$, hence, $\mu \vee_c \nu \leq \mathcal{J}(\mu, \nu)$. $\qquad\square$

2 Proof of Theorem 1.1

The proof of Theorem 1.1 relies on the next two results whose proofs are postponed.

Lemma 2.1. *For $p \geq 1$, $\mathcal{I}$ and $\mathcal{J}$ are continuous maps on $\mathcal{P}_p(\mathbb{R}) \times \mathcal{P}_p(\mathbb{R})$ to $\mathcal{P}_p(\mathbb{R})$.*

Proposition 2.2. *Let f and g be real-valued càdlàg functions on $[0, 1]$ with respective antiderivatives F and G. We have, for $p \geq 1$,*

$$\|\partial_+ (\mathrm{co}(F) - \mathrm{co}(G))\|_p \leq \|f - g\|_p. \tag{2.1}$$

Proof of Theorem 1.1. Let $\mu, \mu', \nu, \nu' \in \mathcal{P}_p(\mathbb{R})$. Assume for a moment that (1.4) and (1.5) hold for probability measures with bounded support. Since such measures are dense in $\mathcal{P}_p(\mathbb{R})$, there exist $\mu_n, \mu'_n, \nu_n, \nu'_n \in \mathcal{P}_p(\mathbb{R})$, $n \in \mathbb{N}$ with bounded support such that

$$\lim_{n \to +\infty} \mathcal{W}_p(\mu, \mu_n) + \mathcal{W}_p(\mu', \mu'_n) + \mathcal{W}_p(\nu, \nu_n) + \mathcal{W}_p(\nu', \nu'_n) = 0. \tag{2.2}$$

We have by Lemma 2.1 that $(\mathcal{J}(\mu_n, \nu_n))_{n \in \mathbb{N}}$ and $(\mathcal{J}(\mu'_n, \nu'_n))_{n \in \mathbb{N}}$ converge to $\mathcal{J}(\mu, \nu)$ and $\mathcal{J}(\mu', \nu')$ resp. in $\mathcal{P}_p(\mathbb{R})$. Therefore,

$$\begin{aligned}
\mathcal{W}_p(\mathcal{J}(\mu, \nu), \mathcal{J}(\mu', \nu')) &= \lim_{n \to +\infty} \mathcal{W}_p(\mathcal{J}(\mu_n, \nu_n), \mathcal{J}(\mu'_n, \nu'_n)) \\
&\leq \lim_{n \to +\infty} 2\mathcal{W}_p(\mu_n, \mu'_n) + \mathcal{W}_p(\nu_n, \nu'_n) \\
&= 2\mathcal{W}_p(\mu, \mu') + \mathcal{W}_p(\nu, \nu').
\end{aligned}$$

Hence, we may assume that μ, ν, μ', ν' have bounded supports. This implies that the associated quantile functions are bounded on $(0, 1)$ and since they are non-decreasing and have at most countably many jumps, coincide λ-a.s. with càdlàg functions on $[0, 1]$, where λ denotes the Lebesgue measure on $[0, 1]$. Therefore,

$$G : v \mapsto \int_0^v (F_\mu^{-1} - F_\nu^{-1})(u)\, du \quad \text{and} \quad G' : v \mapsto \int_0^v (F_{\mu'}^{-1} - F_{\nu'}^{-1})(u)\, du,$$

are the antiderivatives of càdlàg functions on $[0, 1]$. By Proposition 2.2 and (1.7),

$$\begin{aligned}
\|\partial_+ (\mathrm{co}(G) - \mathrm{co}(G'))\|_p &\leq \|\partial_+ (G - G')\|_p = \|F_\mu^{-1} - F_\nu^{-1} - F_{\mu'}^{-1} + F_{\nu'}^{-1}\|_p \\
&\leq \|F_\mu^{-1} - F_{\mu'}^{-1}\|_p + \|F_\nu^{-1} - F_{\nu'}^{-1}\|_p = \mathcal{W}_p(\mu, \mu') + \mathcal{W}_p(\nu, \nu').
\end{aligned}$$

By (1.3), we have for λ-almost every $u \in (0,1)$ that

$$F_{\mathcal{I}(\mu,\nu)}^{-1}(u) = F_\mu^{-1}(u) - \partial_+ \operatorname{co}(G)(u) \text{ and } F_{\mathcal{I}(\mu',\nu')}^{-1}(u) = F_{\mu'}^{-1}(u) - \partial_+ \operatorname{co}(G')(u),$$

$$F_{\mathcal{J}(\mu,\nu)}^{-1}(u) = F_\nu^{-1}(u) + \partial_+ \operatorname{co}(G)(u) \text{ and } F_{\mathcal{J}(\mu',\nu')}^{-1}(u) = F_{\nu'}^{-1}(u) + \partial_+ \operatorname{co}(G')(u).$$

Therefore, using (1.7), we obtain

$$\mathcal{W}_p(\mathcal{I}(\mu,\nu), \mathcal{I}(\mu',\nu')) = \left\| F_{\mathcal{I}(\mu,\nu)}^{-1} - F_{\mathcal{I}(\mu',\nu')}^{-1} \right\|_p = \left\| F_\mu^{-1} - \partial_+ \operatorname{co}(G) - F_{\mu'}^{-1} + \partial_+ \operatorname{co}(G') \right\|_p$$

$$\leq \left\| F_\mu^{-1} - F_{\mu'}^{-1} \right\|_p + \left\| \partial_+ \left(\operatorname{co}(G) - \operatorname{co}(G') \right) \right\|_p \leq 2\mathcal{W}_p(\mu,\mu') + \mathcal{W}_p(\nu,\nu').$$

In the same way,

$$\mathcal{W}_p(\mathcal{J}(\mu,\nu), \mathcal{J}(\mu',\nu')) = \left\| F_\nu^{-1} + \partial_+ \operatorname{co}(G) - F_{\nu'}^{-1} - \partial_+ \operatorname{co}(G') \right\|_p$$

$$\leq \mathcal{W}_p(\mu,\mu') + 2\mathcal{W}_p(\nu,\nu'). \qquad \square$$

Proof of Lemma 2.1. In the following we will only show continuity of $\mathcal{J}$ and remark that continuity of $\mathcal{I}$ follows mutatis mutandis (and can be even shown with a simpler line of argument, since $\mathcal{I}(\mu,\nu) \leq_c \nu$). Let $(\mu_n)_{n \in \mathbb{N}}$, $(\nu_n)_{n \in \mathbb{N}}$ be sequences that converge to μ_∞, ν_∞ resp. in $\mathcal{P}_p(\mathbb{R})$.

Step 1. We show that $(\mathcal{J}(\mu_n,\nu_n))_{n \in \mathbb{N}}$ is a precompact subset of $\mathcal{P}_p(\mathbb{R})$. As a consequence of the de la Vallée-Poussin theorem, see for example [12, Theorem 4.5.9 and proof], there exists a continuous, increasing and strictly convex function $\theta \colon \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{r \to \infty} \frac{r}{\theta(r)} = 0$ and

$$\sup_{n \in \mathbb{N}} \int \theta(|x|^p)\, \mu_n(dx) < \infty, \quad \sup_{n \in \mathbb{N}} \int \theta(|y|^p)\, \nu_n(dy) < \infty.$$

Note that, when $p > 1$, $\theta \circ (|\cdot|^p/2^p)$ is also strictly convex as the composition of a convex, increasing function with a strictly convex function. On the other hand, when $p = 1$, $x \neq y$ and $\alpha \in (0,1)$, either $|x| \neq |y|$ or the inequality $|\alpha x + (1-\alpha)y| \leq \alpha |x| + (1-\alpha)|y|$ is strict, so that, since θ is increasing and strictly convex, $\theta \circ (|\cdot|/2)$ is again strictly convex. We conclude that in any case $\theta \circ (|\cdot|^p/2^p)$ is strictly convex.

Consider the transport problem $\mathcal{W}_\theta$ given by

$$\mathcal{W}_\theta(\eta,\nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int \theta\left(\frac{|x-y|^p}{2^p} \right) \pi(dx,dy), \tag{2.3}$$

for $\eta, \nu \in \mathcal{P}_p(\mathbb{R})$, and observe that

$$\mathcal{W}_\theta(\eta,\nu) \leq \int \theta\left(\frac{|x|^p + |y|^p}{2} \right) \eta \otimes \nu(dx,dy) \leq \frac{1}{2} \left(\int \theta(|x|^p)\, \eta(dx) + \int \theta(|y|^p)\, \nu(dy) \right).$$

We have by [5, Theorem 1.4] that

$$\mathcal{V}_\theta(\mu,\nu) := \inf_{\mu \leq_c \eta} \mathcal{W}_\theta(\eta,\nu) = \mathcal{W}_\theta(\mathcal{J}(\mu,\nu),\nu) \leq \mathcal{W}_\theta(\mu,\nu),$$

from which we deduce that $(\mathcal{V}_\theta(\mu_n,\nu_n))_{n \in \mathbb{N}}$ is a bounded sequence. For $n \in \mathbb{N}$, let π_n be an optimizer of $\mathcal{W}_\theta(\mathcal{J}(\mu_n,\nu_n),\nu_n)$ in $\Pi(\mathcal{J}(\mu_n,\nu_n),\nu_n)$. We then find by monotonicity of θ combined with $\frac{|x|^p}{2^{p-1}} \leq |x-y|^p + |y|^p$ and convexity of θ,

$$\int \theta\left(\frac{|x|^p}{4^p} \right) \mathcal{J}(\mu_n,\nu_n)(dx) \leq \int \theta\left(\frac{|x-y|^p + |y|^p}{2 \times 2^p} \right) \pi_n(dx,dy)$$

$$\leq \frac{1}{2} \left(\mathcal{V}_\theta(\mu_n,\nu_n) + \int \theta\left(\frac{|y|^p}{2^p} \right) \nu_n(dy) \right)$$

$$\leq \frac{1}{2} \left(\mathcal{V}_\theta(\mu_n,\nu_n) + \int \theta\left(|y|^p \right) \nu_n(dy) \right).$$

Therefore, the left-hand side is uniformly bounded in $n \in \mathbb{N}$. Recall that $\frac{r}{\theta(r)}$ vanishes for $r \to \infty$ and so does $\sup_{s \geq r} \frac{s}{\theta(s)}$. Since

$$\int \mathbb{1}_{[r,\infty)} \left(\frac{|x|^p}{4^p} \right) \frac{|x|^p}{4^p} \, \mathcal{J}(\mu_n, \nu_n)(dx) \leq \sup_{s \geq r} \frac{s}{\theta(s)} \int \theta \left(\frac{|x|^p}{4^p} \right) \mathcal{J}(\mu_n, \nu_n)(dx),$$

with the integrals on the right-hand side uniformly bounded in $n \in \mathbb{N}$, we deduce that

$$\lim_{r \to \infty} \sup_{n \in \mathbb{N}} \int \mathbb{1}_{[r,\infty)}(|x|^p)|x|^p \, \mathcal{J}(\mu_n, \nu_n)(dx) = 0.$$

In particular, by Markov's inequality, we get that the sequence $(\mathcal{J}(\mu_n, \nu_n))_{n \in \mathbb{N}}$ is tight and by [23, Definition 6.8] precompact in $\mathcal{P}_p(\mathbb{R})$.

Step 2. Precompactness allows us to pass to a subsequence convergent in $\mathcal{P}_p(\mathbb{R})$ with limit γ. Consider the continuous, increasing, and strictly convex function $\hat{\theta}(x) := \sqrt{x^{2p} + 1}$ on $\mathbb{R}_+$ with $\hat{\theta}(x) \leq x^p + 1$. By stability, that is [5, Theorem 1.5], we obtain

$$\mathcal{V}_{\hat{\theta}}(\mu, \nu) = \lim_{n \to \infty} \mathcal{V}_{\hat{\theta}}(\mu_n, \nu_n) = \lim_{n \to \infty} \mathcal{W}_{\hat{\theta}}(\mathcal{J}(\mu_n, \nu_n), \nu_n) = \mathcal{W}_{\hat{\theta}}(\gamma, \nu).$$

Since $\mathcal{W}_p$-convergence preserves convex ordering, we get that $\mu \leq_c \gamma$ and by uniqueness of the optimizer of $\mathcal{V}_{\hat{\theta}}(\mu, \nu)$, $\gamma = \mathcal{J}(\mu, \nu)$. Hence, $(\mathcal{J}(\mu_n, \nu_n))_{n \in \mathbb{N}}$ converges to $\mathcal{J}(\mu, \nu)$ in $\mathcal{P}_p(\mathbb{R})$. □

The proof of Proposition 2.2 relies on the next three lemmas. The proof of the first one is postponed after the one of the proposition.

Lemma 2.3. *Let $0 \leq a < b$ and $F, G : [0, b] \to \mathbb{R}$ be continuous on $[0, b]$, convex on $[0, a)$ and affine on $[a, b]$. Then for any non-decreasing convex function $\theta : \mathbb{R} \to \mathbb{R}$ we have*

$$\int_0^b \theta\left(|\partial_+(\mathrm{co}(F) - \mathrm{co}(G))|\right)(u) \, du \leq \int_0^b \theta\left(|\partial_+(F - G)|\right)(u) \, du \tag{2.4}$$

where co denotes the convex hull and ∂_+ the right-hand derivative.

Lemma 2.4. *Let $0 \leq a < b < \infty$, $F : [0, b] \to \mathbb{R}$, and*

$$H(x) := \begin{cases} \mathrm{co}(F|_{[0,a)})(x) & \text{if } x \in [0, a) \\ F(x) & \text{if } x \in [a, b]. \end{cases} \tag{2.5}$$

Then $\mathrm{co}(H) = \mathrm{co}(F)$.

Proof. The function $\mathrm{co}(H)$ is convex and satisfies $\mathrm{co}(H) \leq H \leq F$. By definition of the convex hull, we deduce that $\mathrm{co}(H) \leq \mathrm{co}(F)$. Conversely, $\mathrm{co}(F)$ is convex and bounded from above by F, so that the restriction $\mathrm{co}(F)|_{[0,a)}$ is convex and bounded from above by $F|_{[0,a)}$. Hence $\mathrm{co}(F)|_{[0,a)} \leq \mathrm{co}(F|_{[0,a)})$ and, since $\mathrm{co}(F) \leq F$, we have $\mathrm{co}(F) \leq H$ by (2.5). By definition of the convex hull, $\mathrm{co}(F) \leq \mathrm{co}(H)$, which concludes the proof. □

Lemma 2.5. *Let $f : [0, 1] \to \mathbb{R}$ be a càdlàg function. Then there exists for every $\varepsilon > 0$ a piecewise constant, càdlàg function $g : [0, 1] \to \mathbb{R}$ with at most finitely many jumps such that $\sup_{x \in [0,1]} |f(x) - g(x)| < \varepsilon$.*

Proof. This follows from [11, Section 12, Lemma 1] and the discussion below. □

Proof of Proposition 2.2. For the moment we assume that the assertion of the proposition holds true for antiderivatives of piecewise constant, càdlàg functions. Since f and g are càdlàg, there exist by Lemma 2.5 for every $n \in \mathbb{N}^* := \mathbb{N} \setminus \{0\}$ piecewise constant, càdlàg functions $f_n, g_n : [0, 1] \to \mathbb{R}$ with finitely many discontinuities such that

$$\sup_{x \in [0,1]} \{|f(x) - f_n(x)| + |g(x) - g_n(x)|\} < 1/n.$$

Therefore, $(f_n)_{n \in \mathbb{N}^*}$ and $(g_n)_{n \in \mathbb{N}^*}$ converge in $L^p(\lambda)$ to f and g, respectively. Let, for $n \in \mathbb{N}^*, u \in [0, 1]$,

$$F_n(u) := F(0) + \int_0^u f_n(v)\, dv \text{ and } G_n(u) := G(0) + \int_0^u g_n(v)\, dv.$$

We have $\|F - F_n\|_\infty < \frac{1}{n}$, $\|G - G_n\|_\infty < \frac{1}{n}$. Since $\mathrm{co}(F) - \|F - F_n\|_\infty$ (resp. $\mathrm{co}(F_n) - \|F - F_n\|_\infty$) is a convex function bounded from above by $F - \|F - F_n\|_\infty \le F_n$ (resp. F), $\|\mathrm{co}(F) - \mathrm{co}(F_n)\|_\infty \le \|F - F_n\|_\infty < \frac{1}{n}$ and, in the same way, $\|\mathrm{co}(G) - \mathrm{co}(G_n)\|_\infty < \frac{1}{n}$. By [18, Theorem 6.2.7], we have λ-almost sure convergence of $(\partial_+ \mathrm{co}(F_n))_{n \in \mathbb{N}^*}$ and $(\partial_+ \mathrm{co}(G_n))_{n \in \mathbb{N}^*}$ to $\partial_+ \mathrm{co}(F)$ and $\partial_+ \mathrm{co}(G)$, respectively. Again, as f and g are càdlàg, we have $\max(\|f\|_\infty, \|g\|_\infty) + 1 =: K < \infty$, and $\max(\|f_n\|_\infty, \|g_n\|_\infty) \le K$ for all $n \in \mathbb{N}^*$, which yields by definition of the convex hull that $\mathrm{co}(F_n)(u) \ge F_n(0) - Ku$ and $\mathrm{co}(F_n)(u) \ge F_n(1) - K(1 - u)$. Hence,

$$\frac{\mathrm{co}(F_n)(u) - \mathrm{co}(F_n)(0)}{u} \ge -K, \quad \frac{\mathrm{co}(F_n)(1) - \mathrm{co}(F_n)(u)}{1 - u} \le K,$$

and by monotonicity of the one-sided derivatives (and the same reasoning for G_n) we obtain that $\max(\|\partial_+ \mathrm{co}(F_n)\|_\infty, \|\partial_+ \mathrm{co}(G_n)\|_\infty) \le K$. Then dominated convergence yields that $(\partial_+ \mathrm{co}(F_n))_{n \in \mathbb{N}^*}$ and $(\partial_+ \mathrm{co}(G_n))_{n \in \mathbb{N}^*}$ converge in $L^p(\lambda)$ to $\partial_+ \mathrm{co}(F)$ and $\partial_+ \mathrm{co}(G)$, respectively. Finally, by applying (2.1) and the triangle inequality we get the desired inequality:

$$\|\partial_+ (\mathrm{co}(F) - \mathrm{co}(G))\|_p = \lim_{n \to \infty} \|\partial_+ (\mathrm{co}(F_n) - \mathrm{co}(G_n))\|_p$$
$$\le \lim_{n \to +\infty} \|f_n - g_n\|_p = \|f - g\|_p.$$

It remains to show (2.1) for piecewise constant, càdlàg functions f and g with finitely many jumps. To this end, let $(a_k)_{0 \le k \le n}$ be a partition of $[0, 1]$ adapted to f and g, i.e., $0 = a_0 < \ldots < a_n = 1$ and for all $k \in \{0, \cdots, n - 1\}$, $f|_{[a_k, a_{k+1})}$ and $g|_{[a_k, a_{k+1})}$ are constant. For $k \in \{0, \cdots, n\}$, we consider

$$F^k : x \mapsto \begin{cases} \mathrm{co}(F|_{[0, a_k)})(x) & \text{if } x \in [0, a_k), \\ F(x) & \text{else;} \end{cases}$$
$$G^k : x \mapsto \begin{cases} \mathrm{co}(G|_{[0, a_k)})(x) & \text{if } x \in [0, a_k), \\ G(x) & \text{else,} \end{cases}$$

and write $f^k = \partial_+ F^k$ and $g^k = \partial_+ G^k$.

Note that $F^0 = F^1 = F$, $G^0 = G^1 = G$ and $F^n = \mathrm{co}(F)$, $G^n = \mathrm{co}(G)$. By induction we will show that, for $k \in \{0, \ldots n - 1\}$.

$$\|f^{k+1} - g^{k+1}\|_p \le \|f^k - g^k\|_p. \tag{2.6}$$

As the initial case is trivial, we assume that (2.6) holds for $0 \le k \le n - 2$. First, observe

$$\|f^{k+1} - g^{k+1}\|_p^p = \|(f^{k+1} - g^{k+1})|_{[0, a_{k+1})}\|_p^p + \|(f^{k+1} - g^{k+1})|_{[a_{k+1}, 1]}\|_p^p$$
$$= \|\partial_+ (\mathrm{co}(F|_{[0, a_{k+1})}) - \mathrm{co}(G|_{[0, a_{k+1})}))\|_p^p + \|(f^k - g^k)|_{[a_{k+1}, 1]}\|_p^p.$$

Applying Lemma 2.4 with $a = a_k$, $b = a_{k+1}$ yields $\mathrm{co}(F|_{[0, a_{k+1})}) = \mathrm{co}(F^k|_{[0, a_{k+1})})$ and $\mathrm{co}(G|_{[0, a_{k+1})}) = \mathrm{co}(G^k|_{[0, a_{k+1})})$, so that,

$$\|f^{k+1} - g^{k+1}\|_p^p = \|\partial_+ (\mathrm{co}(F^k|_{[0, a_{k+1})}) - \mathrm{co}(G^k|_{[0, a_{k+1})}))\|_p^p + \|(f^k - g^k)|_{[a_{k+1}, 1]}\|_p^p.$$

Since f and g are absolutely bounded by some constant $C > 0$, we have $\mathrm{co}(F|_{[0,a_k)})(u) \geq F(a_k) - C(a_k - u)$, $\mathrm{co}(G|_{[0,a_k)})(u) \geq G(a_k) - C(a_k - u)$, thus, $\lim_{u \nearrow a_k} \mathrm{co}(F|_{[0,a_k)})(u) = F(a_k)$ and $\lim_{u \nearrow a_k} \mathrm{co}(G|_{[0,a_k)})(u) = G(a_k)$. In particular, F^k and G^k are continuous. We can apply Lemma 2.3 with $a = a_k$, $b = a_{k+1}$ to obtain

$$\left\| \partial_+ (\mathrm{co}(F^k|_{[0,a_{k+1})}) - \mathrm{co}(G^k|_{[0,a_{k+1})})) \right\|_p \leq \left\| (f^k - g^k)|_{[0,a_{k+1})} \right\|_p,$$

from which we deduce (2.6). In particular, we have shown the assertion:

$$\left\| \partial_+ (\mathrm{co}(F) - \mathrm{co}(G)) \right\|_p = \|f^n - g^n\|_p \leq \left\| f^0 - g^0 \right\|_p = \|f - g\|_p. \qquad \square$$

The proof of Lemma 2.3 relies on the next two lemmas.

Lemma 2.6. *Let $\theta \colon \mathbb{R} \to \mathbb{R}$ be convex, $x, z \in \mathbb{R}$ with $x < z$, and $y, \hat{y} \in [x, z]$. Then*

$$\theta(y) - \theta(x) \leq \tfrac{y-x}{z-\hat{y}} \left(\theta(z) - \theta(\hat{y}) \right). \tag{2.7}$$

Proof. Since θ is convex, we have

$$\theta(y) - \theta(x) \leq \tfrac{y-x}{z-x} \left(\theta(z) - \theta(x) \right) \text{ and } \tfrac{z-\hat{y}}{z-x} \left(\theta(z) - \theta(x) \right) \leq \theta(z) - \theta(\hat{y}).$$

Combining these two inequalities yields (2.7). $\qquad \square$

Lemma 2.7. *Let $\mu \in \mathcal{P}_1(\mathbb{R})$ and $f : \mathbb{R} \to \mathbb{R}$ be a measurable map such that μ is equal to the image $f_\# \lambda$ of the Lebesgue measure λ on $(0,1)$ by f. Then*

$$\forall v \in [0,1], \quad \int_v^1 f(u)\, du \leq \int_v^1 F_\mu^{-1}(u)\, du.$$

Proof. The conclusion being obvious when $v = 1$, we suppose $v \in [0,1)$. The image η of $\mathbb{1}_{(v,1)}(u)\lambda(du)$ by f is such that $\eta \leq \mu$. We have $\frac{\mu}{1-v} \geq \frac{\eta}{1-v} =: \hat{\eta} \in \mathcal{P}_1(\mathbb{R})$ and thus

$$1 - F_{\hat{\eta}}(F_\mu^{-1}(1-(1-v)u)) \leq \tfrac{\mu((F_\mu^{-1}(1-(1-v)u)),+\infty))}{1-v} = \tfrac{1 - F_\mu(F_\mu^{-1}(1-(1-v)u))}{1-v} \leq u,$$

where $u \in (0,1)$, so that $F_{\hat{\eta}}^{-1}(1-u) \leq F_\mu^{-1}(1-(1-v)u)$. With the inverse transform sampling, this implies

$$\int_v^1 f(u)\, du = (1-v) \int_0^1 F_{\hat{\eta}}^{-1}(1-u)\, du \leq (1-v) \int_0^1 F_\mu^{-1}(1-(1-v)u)\, du = \int_v^1 F_\mu^{-1}(u)\, du. \ \square$$

Proof of Lemma 2.3. Let $f, g \colon [0,b) \to \mathbb{R}$ be given by the right-hand derivative of F and G resp., that is $f := \partial_+ F$, $g := \partial_+ G$. Since F and G are convex on $[0,a)$ and affine on $[a,b]$, f and g are non-decreasing on $[0,a)$ and constant on $[a,b)$. Our first goal is to find an explicit representation of the convex hulls of F and G. To this end, consider the infimum

$$c := \inf \left\{ x \in [0,b) \mid f(x) \geq \tfrac{F(b)-F(x)}{b-x} \right\}, \tag{2.8}$$

which is well-defined and not greater than a as $f(a) = \frac{F(b)-F(a)}{b-a}$. Moreover, the infimum in (2.8) is attained by continuity of F and right-continuity of f. When $c > 0$, we get by (2.8) and monotonicity of f on $[0,a)$ that $\sup_{x \in [0,c)} f(x) = \lim_{x \nearrow c} f(x) \leq \lim_{x \nearrow c} \frac{F(b)-F(x)}{b-x} = \frac{F(b)-F(c)}{b-c}$. Under the convention $\sup_{x \in [0,c)} f(x) = -\infty$ when $c = 0$, we therefore find

$$\sup_{x \in [0,c)} f(x) \leq \tfrac{F(b)-F(c)}{b-c} =: \phi \leq f(c). \tag{2.9}$$

For $x \in [c,a]$, using (2.9) and the fact that f is non-decreasing on $[0,a)$, we obtain

$$\begin{aligned} F(b) - F(x) &= F(b) - F(c) + F(c) - F(x) = (b-c)\phi - \int_c^x f(u)\, du \\ &\leq (b-c)\phi + (c-x)\phi = (b-x)\phi \leq (b-x)f(x). \end{aligned} \tag{2.10}$$

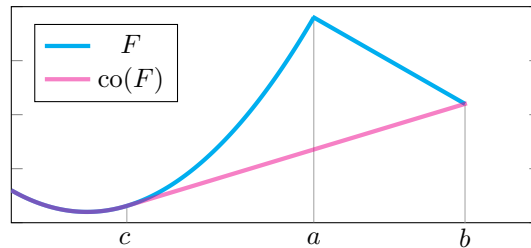Figure 1: Illustration of F and its convex hull $\mathrm{co}(F)$.

We claim that

$$\mathrm{co}(F)(x) = F(\min(x,c)) + \max(x-c,0)\phi, \quad x \in [0,b]. \tag{2.11}$$

Denote the right-hand side of (2.11) by $\tilde{F}$. Note that $\tilde{F}$ is convex on $[0,b]$ since the right-hand derivative $\partial_+ \tilde{F}(x) = \mathbb{1}_{[0,c)} f(x) + \mathbb{1}_{[c,b)}\phi$ is non-decreasing by (2.9). We calculate

$$F(x) = F(c) + F(x) - F(b) + F(b) - F(c) \geq F(c) - (b-x)\phi + (b-c)\phi = \tilde{F}(x),$$

for $x \in [c,a]$, which yields that $\tilde{F} \leq F$ on $[0,a] \cup \{b\}$. Since both functions are affine on $[a,b]$ we conclude that $\tilde{F} \leq F$ and therefore, by definition of the convex hull, $\tilde{F} \leq \mathrm{co}(F)$.

In order to show (2.11), it remains to verify that $\mathrm{co}(F) \leq \tilde{F}$. By convexity of $\mathrm{co}(F)$ and the inequality $\mathrm{co}(F) \leq F$, we have, for $x \in [c,b]$,

$$\mathrm{co}(F)(x) \leq \tfrac{x-c}{b-c} F(b) + \tfrac{b-x}{b-c} F(c) = F(c) + (x-c)\phi = \tilde{F}(x),$$

and $\mathrm{co}(F)(x) \leq F(x) = \tilde{F}(x)$ for $x \in [0,c]$, thus, $\mathrm{co}(F) \leq \tilde{F}$.

Reasoning the same way for G, we deduce that d defined analogously to (2.8)

$$d := \inf \left\{ x \in [0,b) : g(x) \geq \tfrac{G(b)-G(x)}{b-x} \right\}$$

is not greater than a and has the properties

$$\sup_{x \in [0,d)} g(x) \leq \tfrac{G(b)-G(d)}{b-d} =: \gamma \leq g(d), \tag{2.12}$$

$$\mathrm{co}(G)(x) = G(\min(x,d)) + \max(x-d,0)\gamma. \tag{2.13}$$

After this preparatory work we proceed to show the assertion, that is (2.4). Without loss of generality, we assume that $c \leq d$. Note that, by (2.11) and (2.9), $\partial_+ \mathrm{co}(F)(x) = \min(f(\min(x,c)), \phi)$ and, by (2.13) and (2.12), $\partial_+ \mathrm{co}(G)(x) = \min(g(\min(x,d)), \gamma)$. Therefore, the left-hand side of (2.4) coincides with

$$\int_0^c \theta\left(|f(u) - g(u)|\right) du + \int_c^d \theta\left(|\phi - g(u)|\right) du + \int_d^b \theta\left(|\phi - \gamma|\right) du.$$

To conclude, we thus have to show

$$\int_c^d \theta\left(|\phi - g(u)|\right) du + \int_d^b \theta\left(|\phi - \gamma|\right) du \leq \int_c^b \theta\left(|f(u) - g(u)|\right) du. \tag{2.14}$$

Case 1: Suppose that $\phi \geq \gamma$. By (2.9), the monotonicity of f on $[0,a)$ and (2.12),

$$g(u) \leq \gamma \leq \phi \leq f(u) \quad u \in [c,d).$$

Then, by applying Lemma 2.6 (with, in the notation of this lemma, $x = \phi - \gamma$, $y = f(u) - \gamma$, $\hat{y} = \phi - g(u)$, $z = f(u) - g(u)$, which satisfy $\frac{y-x}{z-\hat{y}} = 1$), we find

$$\theta\left(f(u) - g(u)\right) - \theta\left(\phi - g(u)\right) \geq \theta\left(f(u) - \gamma\right) - \theta\left(\phi - \gamma\right), \quad u \in [c,d)$$

so that $\int_c^d \theta\left(|f(u) - g(u)|\right) - \theta\left(\phi - g(u)\right) du \geq \int_c^d \theta\left(f(u) - \gamma\right) - \theta(\phi - \gamma) du.$ \hfill (2.15)

Denoting by $\tilde{\phi}$ the fraction $\frac{F(b)-F(d)}{b-d}$, we have, since θ is non-decreasing and convex,

$$\int_d^b \theta\left(|f(u)-g(u)|\right) \frac{du}{b-d} \geq \theta\left(\left|\int_d^b \frac{f(u)-g(u)}{b-d}\,du\right|\right) = \theta\left(|\tilde{\phi}-\gamma|\right). \tag{2.16}$$

We bring all terms of (2.14) to one side and find

$$\int_c^b \theta\left(|f(u)-g(u)|\right)\,du - \int_c^d \theta(\phi-g(u))\,du - \int_b^d \theta(\phi-\gamma)\,du$$
$$\geq \int_c^d \theta(f(u)-\gamma) - \theta(\phi-\gamma)\,du + \int_d^b \theta(|\tilde{\phi}-\gamma|) - \theta(\phi-\gamma)\,du$$
$$\geq (b-c)\left(\theta\left(\frac{1}{b-c}\left(\int_c^d f(u) - \gamma\,du + \int_d^b |\tilde{\phi}-\gamma|du\right)\right) - \theta(\phi-\gamma)\right), \tag{2.17}$$

where we use (2.15) and (2.16) for the first inequality and then convexity of θ. Since θ is non-decreasing and $\frac{1}{b-c}\left(\int_c^d f(u)-\gamma\,du + \int_d^b \tilde{\phi}-\gamma\,du\right) = \phi-\gamma$, we find that the right-hand side of (2.17) is non-negative, from which we derive (2.14).

Case 2: Suppose that $\phi < \gamma$ and let $e := \inf\{u \in [c,d] \mid g(u) \geq \phi\}$ where, by convention, the infimum over the empty set is defined as d. By (2.9), the monotonicity of f on $[0,a)$ and $d \leq a$, we have $g(u) \leq \phi \leq f(u)$ for $u \in [c,e)$, thus, by monotonicity of θ,

$$\int_c^e \theta(|\phi-g(u)|)\,du = \int_c^e \theta(\phi-g(u))\,du \leq \int_c^e \theta(f(u)-g(u))\,du = \int_c^e \theta(|f(u)-g(u)|)\,du. \tag{2.18}$$

On the other hand, by (2.11),

$$\forall x \in [c,b], \ F(x) \geq \mathrm{co}(F)(x) = F(b) + (x-b)\phi \tag{2.19}$$

so that $\tilde{\phi} = \frac{F(b)-F(d)}{b-d} \leq \phi$. As θ is non-decreasing and convex, we obtain

$$\int_d^b \theta\left(|f(u)-g(u)|\right) \frac{du}{b-d} \geq \theta\left(\gamma-\tilde{\phi}\right). \tag{2.20}$$

As consequence of (2.18) and (2.20), the following inequality suffices to get (2.14):

$$\int_e^d \theta(g(u)-\phi)\,du + (b-d)\theta(\gamma-\phi) \leq \int_e^d \theta\left(|f(u)-g(u)|\right)\,du + (b-d)\theta(\gamma-\tilde{\phi}). \tag{2.21}$$

Showing (2.21) is equivalent to proving that the respective images μ and ν of the Lebesgue measure λ on $(0,1)$ by the maps

$$T^1(u) := \begin{cases} g(e+(b-e)u) - \phi & u < \frac{d-e}{b-e}, \\ \gamma - \phi & \text{else}, \end{cases}$$

$$T^2(u) := \begin{cases} |f(e+(b-e)u) - g(e+(b-e)u)| & u < \frac{d-e}{b-e}, \\ \gamma - \tilde{\phi} & \text{else} \end{cases}$$

are in the increasing convex order. By [21, Theorem 4.A.3], this is equivalent to

$$\int_v^1 F_\mu^{-1}(u)\,du \leq \int_v^1 F_\nu^{-1}(u)\,du, \quad v \in [0,1]. \tag{2.22}$$

Since T^1 is non-decreasing, we have by [2, Lemma A.3] that $T^1(u) = F_\mu^{-1}(u)$ for λ-almost every $u \in (0,1)$. This observation combined with Lemma 2.7 leads to

$$\int_v^1 F_\mu^{-1}(u)\,du = \int_v^1 T^1(u)\,du \text{ and } \int_v^1 F_\nu^{-1}(u)\,du \geq \int_v^1 T^2(u)\,du, \ v \in [0,1]$$

and (2.22) is implied by

$$\int_v^1 T^1(u)\,du \leq \int_v^1 T^2(u)\,du, \ v \in [0,1]. \tag{2.23}$$

Recall that $\tilde{\phi} \le \phi$, so that this inequality holds for $v \in [\frac{d-e}{b-e}, 1]$. Next, abbreviate $\frac{d-e}{b-e} =: w$, let $v \in [0, w)$, and write $\hat{e} := e + (b - e)v$. Using $g \le f + |f - g|$, we have that

$$\int_v^1 T^1(u)\, du \le \int_v^w f(e + (b-e)u) + T^2(u) - \phi\, du + \int_w^1 \gamma - \phi\, du$$
$$= \frac{F(d) - F(\hat{e}) - (d - \hat{e})\phi + (b-d)(\gamma - \phi)}{b - e} + \int_v^w T^2(u)\, du.$$

Remember that $(b - d)\tilde{\phi} = F(b) - F(d)$ and (2.19) applies to $x = \hat{e}$ since $\hat{e} \ge e \ge c$. Thus,

$$F(d) - F(\hat{e}) = F(b) - F(\hat{e}) - (b - d)\tilde{\phi} \le (b - \hat{e})\phi - (b - d)\tilde{\phi}.$$

As $1 - w = \frac{b-d}{b-e}$, we obtain

$$\int_v^1 T^1(u)\, du \le \frac{(b-d)(\phi - \tilde{\phi} + \gamma - \phi)}{b - e} + \int_v^w T^2(u)\, du$$
$$= \frac{b-d}{b-e}(\gamma - \tilde{\phi}) + \int_v^w T^2(u)\, du = \int_v^1 T^2(u)\, du. \qquad \square$$

References

[1] A. Alfonsi, J. Corbetta, and B. Jourdain. 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(03):1950002, 2019. MR3951948

[2] A. Alfonsi, J. Corbetta, and B. Jourdain. Sampling of probability measures in the convex order by Wasserstein projection. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 56, pages 1706–1729. Institut Henri Poincaré, 2020. MR4116705

[3] J.-J. Alibert, G. Bouchitté, and T. Champion. A new class of costs for optimal transport planning. *European Journal of Applied Mathematics*, 30(6):1229–1263, 2019. MR4028478

[4] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Existence, duality, and cyclical monotonicity for weak transport costs. *Calculus of Variations and Partial Differential Equations*, 58(6):1–28, 2019. MR4029731

[5] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Weak monotone rearrangement on the line. *Electronic Communications in Probability*, 25, 2020. MR4069738

[6] J. Backhoff-Veraguas and G. Pammer. Applications of weak transport theory. *Bernoulli*, 28(1):370–394, 2022. MR4337709

[7] J. Backhoff-Veraguas and G. Pammer. Stability of martingale optimal transport and weak optimal transport. *Ann. Appl. Probab.*, 32(1):721–752, 2022. MR4386541

[8] 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. MR3066985

[9] M. Beiglböck, B. Jourdain, W. Margheriti, and G. Pammer. Approximation of martingale couplings on the line in the adapted weak topology. *Probab. Theory Related Fields*, 183(1-2):359–413, 2022. MR4421177

[10] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.*, 44(1):42–106, 2016. MR3456332

[11] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication. MR1700749

[12] V. I. Bogachev. *Measure theory*, volume 1. Springer, 2007. MR2267655

[13] M. Fathi, N. Gozlan, and M. Prod'homme. A proof of the Caffarelli contraction theorem via entropic regularization. *Calculus of Variations and Partial Differential Equations*, 59:1–18, 2020. MR4098037

[14] 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. MR3161649

[15] N. Gozlan and N. Juillet. On a mixture of Brenier and Strassen theorems. *Proceedings of the London Mathematical Society*, 120(3):434–463, 2020. MR4008375

[16] N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, and P. Tetali. Characterization of a class of weak transport-entropy inequalities on the line. *Ann. Inst. Henri Poincaré Probab. Stat.*, 54(3):1667–1693, 2018. MR3825894

[17] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *J. Funct. Anal.*, 273(11):3327–3405, 2017. MR3706606

[18] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of convex analysis*. Springer Science & Business Media, 2004. MR1865628

[19] R. P. Kertz and U. Rösler. Complete lattices of probability measures with applications to martingale theory. *Lecture Notes-Monograph Series*, pages 153–177, 2000. MR1833858

[20] Y.-H. Kim and Y. L. Ruan. Backward and forward Wasserstein projections in stochastic order. *arXiv preprint arXiv:2110.04822*, 2021.

[21] M. Shaked and J. G. Shanthikumar. *Stochastic orders*. Springer, 2007. MR2265633

[22] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965. MR0177430

[23] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009. MR2459454

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4], ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1] LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2] EJMS: Electronic Journal Management System: https://vtex.lt/services/ejms-peer-review/
[3] IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4] BS: Bernoulli Society http://www.bernoulli-society.org/
[5] Project Euclid: https://projecteuclid.org/
[6] IMS Open Access Fund: https://imstat.org/shop/donation/