

Sampling of probability measures in the convex order by Wasserstein projection

Aurélien Alfonsi, Jacopo Corbetta and Benjamin Jourdain

Université Paris-Est, Cermics (ENPC), INRIA, F-77455 Marne-la-Vallée, France.
E-mail: aurelien.alfonsi@enpc.fr; j.corbetta@zeliade.com; benjamin.jourdain@enpc.fr

Received 24 September 2018; revised 28 June 2019; accepted 30 June 2019

Abstract. In this paper, for μ and ν two probability measures on \mathbb{R}^d with finite moments of order $\varrho \geq 1$, we define the respective projections for the W_ϱ -Wasserstein distance of μ and ν on the sets of probability measures dominated by ν and of probability measures larger than μ in the convex order. The W_2 -projection of μ can be easily computed when μ and ν have finite support by solving a quadratic optimization problem with linear constraints. In dimension $d = 1$, Gozlan et al. (*Ann. Inst. Henri Poincaré Probab. Stat.* **54** (3) (2018) 1667–1693) have shown that the projection of μ does not depend on ϱ . We explicit their quantile functions in terms of those of μ and ν . The motivation is the design of sampling techniques preserving the convex order in order to approximate Martingale Optimal Transport problems by using linear programming solvers. We prove convergence of the Wasserstein projection based sampling methods as the sample sizes tend to infinity and illustrate them by numerical experiments.

Résumé. Soient μ et ν deux mesures de probabilité sur \mathbb{R}^d ayant un moment d'ordre $\varrho \geq 1$ fini. Dans ce papier, nous définissons respectivement les projections de μ et ν pour la distance de Wasserstein W_ϱ sur l'ensemble des probabilités dominées par ν et sur l'ensemble des probabilités dominant μ pour l'ordre convexe. Pour $\varrho = 2$, la projection de μ peut facilement être calculée lorsque μ et ν ont un support fini en résolvant un problème de minimisation quadratique avec des contraintes linéaires. En dimension $d = 1$, Gozlan et al. (*Ann. Inst. Henri Poincaré Probab. Stat.* **54** (3) (2018) 1667–1693) ont montré que la projection de μ ne dépend pas de ϱ . Nous donnons ici l'expression de la fonction quantile de cette projection à l'aide des fonctions quantiles de μ et ν . La motivation de cette étude est de fournir une méthode d'échantillonnage permettant de préserver l'ordre convexe. Cela permet ensuite d'approcher les problèmes de transport optimal martingale en utilisant un solveur de programmation linéaire. Nous prouvons la convergence des méthodes d'échantillonnage basées sur la projection Wasserstein lorsque la taille des échantillons tend vers l'infini, et illustrons cette convergence par des exemples numériques.

MSC: 91G60; 90C08; 60G42; 60E15

Keywords: Convex order; Martingale optimal transport; Wasserstein distance; Sampling techniques; Linear programming

1. Introduction

For μ, ν in the set $\mathcal{P}(\mathbb{R}^d)$ of probability measures on \mathbb{R}^d , we say that μ is smaller than ν for the convex order and denote $\mu \leq_{cx} \nu$ if $\int_{\mathbb{R}^d} \phi(x) \mu(dx) \leq \int_{\mathbb{R}^d} \phi(y) \nu(dy)$ for each convex function $\phi: \mathbb{R}^d \rightarrow \mathbb{R}$ non-negative or integrable with respect to $\mu + \nu$. Let us note that when ν has a finite first order moment, this implies the same for μ and that $\int_{\mathbb{R}^d} x \mu(dx) = \int_{\mathbb{R}^d} y \nu(dy)$ by taking $\phi(x) = |x|$ for the first claim and $\phi(x_1, \dots, x_d) = \pm x_i$ for the second. Up to our knowledge, few studies consider the problem of preserving the convex order while approximating two such probability measures. We can mention the one-dimensional method based on the quantile functions proposed by David Baker in his PhD thesis [6] (see the beginning of Section 2.2 for more details). The dual quantization introduced by Pagès and Wilbertz [23] gives another way to preserve the convex order in dimension one (see the remark after Proposition 10 in [23]). This is unfortunately no longer true for higher dimensions. Take for example the case of the probability laws $\mu = \delta_{(0,0)}$ and

ν the distribution on $(U, 0)$ with U uniform on $[-1, 1]$. We have $\mu \leq_{\text{cx}} \nu$. We calculate their dual quantizers $\bar{\mu}$ and $\bar{\nu}$ on the two triangles \mathcal{T}_1 and \mathcal{T}_2 with vertices $\{(-1, 0), (0, -1), (0, 1)\}$ and $\{(0, -1), (1, 0), (0, 1)\}$. We easily obtain $\bar{\mu} = \frac{1}{2}(\delta_{(0,-1)} + \delta_{(0,1)})$, $\bar{\nu} = \frac{1}{4}(\delta_{(0,-1)} + \delta_{(0,1)} + \delta_{(-1,0)} + \delta_{(1,0)})$. Thus, we have $\int x^2 \bar{\mu}(dx, dy) = 1$, $\int x^2 \bar{\nu}(dx, dy) = \frac{1}{2}$, which proves that the convex order is not preserved. However, the quantization and the dual quantization give a possible way to approximate μ and ν in the convex order. Precisely, the quantization of μ gives a probability measure $\underline{\mu}$ with finite support such that $\underline{\mu} \leq_{\text{cx}} \mu$ while the dual quantization of ν gives a probability measure $\bar{\nu}$ with finite support such that $\nu \leq_{\text{cx}} \bar{\nu}$. We therefore have $\underline{\mu} \leq_{\text{cx}} \bar{\nu}$. Though being general, this construction has several drawbacks. First, to define the dual quantization, ν and therefore μ must have a compact support. This is a very restrictive assumption. Second, the calculation of the quantization of μ and of the dual quantization of ν is in general not obvious in dimension $d \geq 2$ and may require an important computation time. This is why one usually pre-calculates the quantization for standard distributions, see [22] for the Gaussian case. Third, this method only works for two measures and does not generalize to design approximations of $\mu, \nu, \eta \in \mathcal{P}(\mathbb{R}^d)$ preserving the convex order when $\mu \leq_{\text{cx}} \nu \leq_{\text{cx}} \eta$.

To avoid the curse of dimension, it is natural to look at the Monte-Carlo method and to consider the empirical measures $\mu_I = \frac{1}{I} \sum_{i=1}^I \delta_{X_i}$ and $\nu_J = \frac{1}{J} \sum_{j=1}^J \delta_{Y_j}$, where X_1, \dots, X_I (resp. Y_1, \dots, Y_J) are i.i.d. random variables with distribution μ (resp. ν). Clearly, there is no reason to have $\frac{1}{I} \sum_{i=1}^I X_i = \frac{1}{J} \sum_{j=1}^J Y_j$ (a necessary condition for the convex order from the choices $\phi(x_1, \dots, x_d) = \pm x_k$ with $k \in \{1, \dots, d\}$) and even more to have $\mu_I \leq_{\text{cx}} \nu_J$. In dimension $d = 1$, according to Kertz and Rösler [19,20], the set of probability measures with a finite first moment is a complete lattice for the increasing and decreasing convex orders. The present paper stems from our preprint [1] (Sections 3 and 4), where, in Section 2 devoted to the one-dimensional case, we also investigate the approximation of μ_I by $\mu_I \wedge \nu_J$ (resp. ν_J by $\mu_I \vee \nu_J$) defined as the infimum of μ_I and ν_J for the decreasing convex order when $\frac{1}{I} \sum_{i=1}^I X_i \leq \frac{1}{J} \sum_{j=1}^J Y_j$ and for the increasing convex order otherwise so that $\mu_I \wedge \nu_J \leq_{\text{cx}} \nu_J$ (resp. $\mu_I \leq_{\text{cx}} \mu_I \vee \nu_J$). Unfortunately, this approach does not generalize to dimension $d \geq 2$, where, according to Proposition 4.5 [21], even the set of probability measures with a constant expectation is no longer a lattice for the convex order. In the present paper, still looking for modifications of μ_I smaller than ν_J in the convex order, we introduce the following minimization problem where $\varrho \geq 1$

$$\begin{cases} \text{minimize } \frac{1}{I} \sum_{i=1}^I |X_i - \sum_{j=1}^J r_{ij} Y_j|^\varrho \\ \text{under the constraints } \forall i, j, r_{ij} \geq 0, \forall i, \sum_{j=1}^J r_{ij} = 1 \text{ and } \forall j, \sum_{i=1}^I r_{ij} = \frac{1}{J}. \end{cases} \tag{1.1}$$

For $\varrho = 2$, this is a quadratic optimization problem with linear constraints which can be solved efficiently numerically (see Section 5). In general, this is the minimization of a continuous function on a compact set and there exists a minimizer r^* . We then define

$$\mu_{I,J}^{\varrho,*} = \frac{1}{I} \sum_{i=1}^I \delta_{X_i^*}, \quad \text{with } X_i^* = \sum_{j=1}^J r_{ij}^* Y_j.$$

By construction, we have $\mu_{I,J}^{\varrho,*} \leq_{\text{cx}} \nu_J$. In the next section, we generalize this problem by considering, in place of the point measures μ_I and ν_J , general elements of $\mathcal{P}_\varrho(\mathbb{R}^d) = \{\eta \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^\varrho \eta(dx) < \infty\}$ with $\varrho \geq 1$ denoted (with a slight abuse of notation) by μ and ν . This leads us to define the projection $\mu_{\mathcal{P}(v)}^\varrho$ of μ on the set $\mathcal{P}(v) = \{\eta \in \mathcal{P}(\mathbb{R}^d) : \eta \leq_{\text{cx}} \nu\}$ of probability measures dominated by ν in the convex order for the Wasserstein distance with index ϱ :

$$W_\varrho(\mu, \eta) = \min_{\pi \in \Pi(\mu, \eta)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^\varrho \pi(dx, dy) \right)^{1/\varrho},$$

where $\Pi(\mu, \nu)$ the set of probability measures π on $\mathbb{R}^d \times \mathbb{R}^d$ with marginal laws μ and ν , i.e. $\pi(A \times \mathbb{R}^d) = \mu(A)$ and $\pi(\mathbb{R}^d \times A) = \nu(A)$ for any Borel set $A \subset \mathbb{R}^d$. We show that this projection is well defined for $\varrho > 1$ and study some of its properties. Notice that after our preprint [1], Gozlan and Juillet [13] and Backhoff-Veraguas et al. [5] have recently considered the projection for $\varrho = 2$. In dimension $d = 1$, according to Gozlan et al. [14] Theorem 1.5, the projection does not depend on ϱ . We explicit its quantile function in terms of the quantile functions of μ and ν so that it can be computed by efficient algorithms when μ and ν have finite supports. In Section 3, we prove that, when $\mu \leq_{\text{cx}} \nu$, then $W_\varrho(\mu, (\mu_I)_{\mathcal{P}(v_J)}^\varrho) \leq 2W_\varrho(\mu, \mu_I) + W_\varrho(\nu, \nu_J)$ and deduce that $(\mu_I)_{\mathcal{P}(v_J)}^\varrho$ converges weakly to μ as $I, J \rightarrow +\infty$. Moreover, we extend the construction to the sampling of several probability measures ranked in the convex order. Section 4 is devoted to the projection $\nu_{\mathcal{P}(\mu)}^\varrho$ of ν on the set $\mathcal{P}(\mu) = \{\eta \in \mathcal{P}(\mathbb{R}^d) : \mu \leq_{\text{cx}} \eta\}$ of probability measures larger than μ in the convex order for the Wasserstein distance with index ϱ . Last, in Section 5, we illustrate by numerical experiments the Wasserstein projection based sampling methods and their application to approximate Martingale Optimal Transport problems. One important motivation of this paper is indeed to tackle numerically the Martingale Optimal Transport (MOT)

problem introduced in [7], which has received a recent and great attention in finance to get model-free bounds on option prices. A family of probability measures on \mathbb{R}^d ($Q(x, dy)_{x \in \mathbb{R}^d}$) is called a Markov kernel on \mathbb{R}^d if for any Borel set $A \subset \mathbb{R}^d$, $\mathbb{R}^d \ni x \mapsto Q(x, A)$ is measurable. We define $\Pi^M(\mu, \nu) = \{\pi \in \Pi(\mu, \nu) : \forall x \in \mathbb{R}^d, \int_{\mathbb{R}^d} |y| \pi_{Y|X}(x, dy) < \infty \text{ and } \int_{\mathbb{R}^d} y \pi_{Y|X}(x, dy) = x\}$ where $\pi_{Y|X}$ denotes a Markov kernel such that $\pi(dx, dy) = \mu(dx) \pi_{Y|X}(x, dy)$, the set of martingale couplings. Theorem 8 in Strassen [27] ensures that, when $\nu \in \mathcal{P}_1(\mathbb{R}^d)$, $\mu \leq_{cx} \nu \iff \Pi^M(\mu, \nu) \neq \emptyset$. For a measurable payoff function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, the MOT problem consists in finding an optimal coupling $\pi^* \in \Pi^M(\mu, \nu)$ that minimizes (or maximizes)

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \pi(dx, dy) \tag{1.2}$$

among all couplings $\pi \in \Pi^M(\mu, \nu)$. In finance, this problem arises naturally if one considers the prices of d assets S_{T_1}, S_{T_2} at dates $T_1 < T_2$. We assume zero interest rates and suppose that we can observe the marginal laws μ (resp. ν) of S_{T_1} (resp. S_{T_2}) from option prices on the market and that we want to price an option that pays $c(S_{T_1}, S_{T_2})$ at date T_2 . Any martingale coupling $\pi \in \Pi^M(\mu, \nu)$ is an arbitrage free pricing model: the supremum and the infimum of $\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \pi(dx, dy)$ over all these couplings give model free bounds on the option price. From the dual formulation of the problem, Beiglböck, Penkner and Henry-Labordère [7] have proved that the upper (resp. lower) bound is the cheapest (resp. most expensive) initial value among superhedging (resp. subhedging) strategies. To compute the model free bounds on the option price, one may consider approximating the probability measures μ and ν by probability measures with finite supports (typically the empirical measures of i.i.d. samples) $\mu_I = \sum_{i=1}^I p_i \delta_{x_i}$ and $\nu_J = \sum_{j=1}^J q_j \delta_{y_j}$, with $I, J \in \mathbb{N}^*$, $x_i, y_j \in \mathbb{R}^d$, $p_i, q_j > 0$ for any i, j and $\sum_{i=1}^I p_i = \sum_{j=1}^J q_j = 1$ and solve the approximate MOT problem: to minimize (or maximize)

$$\sum_{i=1}^I \sum_{j=1}^J r_{ij} c(x_i, y_j) \tag{1.3}$$

over $(r_{ij})_{1 \leq i \leq I, 1 \leq j \leq J}$ under the constraints

$$r_{ij} \geq 0, \quad \sum_{i=1}^I p_i r_{ij} = q_j, \quad \sum_{j=1}^J r_{ij} = 1 \quad \text{and} \quad \sum_{j=1}^J r_{ij} y_j = x_i.$$

This problem falls into the realm of linear programming: powerful algorithms have been developed to solve it numerically. The key issue to run these algorithms is the existence of such matrices $(r_{ij})_{1 \leq i \leq I, 1 \leq j \leq J}$, that amounts to the existence of a martingale coupling between μ_I and ν_J . By Strassen's theorem, this is equivalent to have $\mu_I \leq_{cx} \nu_J$, which motivates the interest of preserving the convex order when sampling both the probability measures μ and ν . It is very natural in the financial application to consider empirical measures with $I = J$: once a stochastic model is calibrated to European option market prices, one basically samples it at different times to price exotic options, which gives the empirical measures at those times.

2. Wasserstein projection of μ on the set of probability measures dominated by ν in the convex order

2.1. Definition, existence and uniqueness

Let us first recall that a family of probability measures on \mathbb{R}^d ($R(x, dy)_{x \in \mathbb{R}^d}$) is called a Markov kernel on \mathbb{R}^d if for any Borel set $A \subset \mathbb{R}^d$, $\mathbb{R}^d \ni x \mapsto R(x, A)$ is measurable. Besides, we say that the Markov kernel R is a martingale Markov kernel if for all $x \in \mathbb{R}^d$, $\int_{\mathbb{R}^d} |y| R(x, dy) < \infty$ and $\int_{\mathbb{R}^d} y R(x, dy) = x$.

For a Markov kernel $R(x, dy)$ on \mathbb{R}^d , we set

$$m_R(x) = \int_{\mathbb{R}^d} y R(x, dy) \quad \text{for } x \in \mathbb{R}^d \text{ s.t. } \int_{\mathbb{R}^d} |y| R(x, dy) < \infty.$$

It is well known (see [11] pages 78–80 or [24] page 117) that if $\pi \in \Pi(\mu, \nu)$ (resp. $\pi \in \Pi^M(\mu, \nu)$), there exists a $\mu(dx)$ -a.e. unique Markov kernel (resp. martingale Markov kernel) R such that $\mu(dx) R(x, dy) = \pi(dx, dy)$. This kernel satisfies obviously $\int_{x \in \mathbb{R}^d} \mu(dx) R(x, dy) = \nu(dy)$, which we note $\mu R = \nu$ later on. Conversely, if R is a kernel satisfying

$\mu R = \nu$ then $\mu(dx)R(x, dy)$ defines a probability measure in $\Pi(\mu, \nu)$. We define $\mathcal{P}(\mathbb{R}^d)$ the set of probability measures on \mathbb{R}^d and, for $\varrho \geq 1$,

$$\mathcal{P}_\varrho(\mathbb{R}^d) = \left\{ \mu \in \mathcal{P}(\mathbb{R}^d), \int_{\mathbb{R}^d} |x|^\varrho \mu(dx) < \infty \right\},$$

the set of probability measures with finite moment of order ϱ .

Suppose that $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ and R is a Markov kernel such that $\mu R = \nu$. Then

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |y| R(x, dy) \mu(dx) = \int_{\mathbb{R}^d} |y| \nu(dy) < \infty$$

so that $m_R(x)$ is defined $\mu(dx)$ -a.e. Moreover for each convex function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $\sup_{x \in \mathbb{R}^d} \frac{|\phi(x)|}{1+|x|} < \infty$, by Jensen's inequality,

$$\begin{aligned} \int_{\mathbb{R}^d} \phi(y) \nu(dy) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \phi(y) \mu(dx) R(x, dy) \\ &\geq \int_{\mathbb{R}^d} \phi \left(\int_{\mathbb{R}^d} y R(x, dy) \right) \mu(dx) = \int_{\mathbb{R}^d} \phi(m_R(x)) \mu(dx). \end{aligned}$$

Despite the restriction on the growth of the convex function ϕ , by Lemma A.1 below, this ensures that $m_R \# \mu \leq_{\text{cx}} \nu$.

For $\varrho \geq 1$ and $\mu, \nu \in \mathcal{P}_\varrho(\mathbb{R}^d)$, we consider the following generalization of the minimization problem (1.1):

$$\begin{cases} \text{Minimize } \mathcal{J}_\varrho(R) := \int_{\mathbb{R}^d} |x - m_R(x)|^\varrho \mu(dx) \\ \text{under the constraint that } R \text{ is a Markov kernel such that } \mu R = \nu. \end{cases}$$

Note that this problem is a particular case of the general transport costs considered by Gozlan et al. [15] and Alibert et al. [3], who are interested in duality results and by Backhoff-Veraguas et al. [5] who deal with existence of optimal transport plans and necessary and sufficient optimality conditions in the spirit of cyclical monotonicity. Gozlan and Juillet [13] characterize optimal transport plans between μ and ν for the cost J_2 . When the X_i are distinct, (1.1) is recovered by setting

$$R(x, dy) = \begin{cases} \sum_{j=1}^J r_{ij} \delta_{Y_j}(dy) & \text{if } x = X_i \text{ for some } i \in \{1, \dots, I\} \\ \delta_x(dy) & \text{if } x \notin \{X_1, \dots, X_I\}. \end{cases}$$

At optimality in (1.1), by Jensen's inequality $\sum_{j=1}^J r_{ij} Y_j = \sum_{j=1}^J r_{kj} Y_j$ when $X_i = X_k$ for $1 \leq k \neq i \leq I$ and the problem (1.1) modified with the additional constraint $\sum_{j=1}^J r_{ij} Y_j = \sum_{j=1}^J r_{kj} Y_j$ when $X_i = X_k$ is recovered by setting

$$R(x, dy) = \begin{cases} \frac{1}{\sum_{i=1}^I 1_{\{X_i=x\}}} \sum_{i: X_i=x} \sum_{j=1}^J r_{ij} \delta_{Y_j}(dy) & \text{if } x \in \{X_1, \dots, X_I\} \\ \delta_x(dy) & \text{if } x \notin \{X_1, \dots, X_I\}. \end{cases}$$

According to the next theorem the generalized problem is equivalent to the computation of the projection of μ on the set of probability measures dominated by ν in the convex order for the ϱ -Wasserstein distance.

Theorem 2.1. *Let $\varrho \geq 1$, $\mu, \nu \in \mathcal{P}_\varrho(\mathbb{R}^d)$. One has $\inf_{R: \mu R = \nu} \mathcal{J}_\varrho(R) = \inf_{\eta \in \mathcal{P}(\nu)} W_\varrho^\varrho(\mu, \eta)$ where both infima are attained. If $\varrho > 1$, then the functions $\{m_{R_\star} : \mu R_\star = \nu \text{ and } \mathcal{J}_\varrho(R_\star) = \inf_{R: \mu R = \nu} \mathcal{J}_\varrho(R)\}$ are $\mu(dx)$ a.e. equal, $\mu_{\mathcal{P}(\nu)}^\varrho := m_{R_\star} \# \mu$ is the unique $\eta \leq_{\text{cx}} \nu$ minimizing $W_\varrho^\varrho(\mu, \eta)$ and $\mu(dx) \delta_{m_{R_\star}(x)}(dy)$ the unique optimal transport plan $\pi \in \Pi(\mu, \mu_{\mathcal{P}(\nu)}^\varrho)$ such that $W_\varrho^\varrho(\mu, \mu_{\mathcal{P}(\nu)}^\varrho) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^\varrho \pi(dx, dy)$.*

When $\varrho > 1$, $\mu_{\mathcal{P}(\nu)}^\varrho$ is the projection of μ on the set of probability measures dominated by ν in the convex order and $\mu_{\mathcal{P}(\nu)}^\varrho \leq_{\text{cx}} \nu$. The proof relies on the next lemma, and is postponed after its proof.

Lemma 2.2. *Let $\nu \in \mathcal{P}_1(\mathbb{R}^d)$. Then, the set $\mathcal{P}(\nu)$ is compact for the weak convergence topology.*

Proof. We have $\sup_{\eta \in \underline{\mathcal{P}}(\nu)} \int_{\mathbb{R}^d} |x|^\varrho \eta(dx) = \int_{\mathbb{R}^d} |x|^\varrho \nu(dx)$. By the Markov inequality, we deduce that for $K \in (0, +\infty)$, $\sup_{\eta \in \underline{\mathcal{P}}(\nu)} \eta(\{x : |x| > K\}) \leq \frac{1}{K^\varrho} \int_{\mathbb{R}^d} |x|^\varrho \nu(dx)$. By the Prokhorov theorem, this implies that $\underline{\mathcal{P}}(\nu)$ is relatively compact for the weak convergence topology. For $K \in (0, \infty)$ and $\eta \leq_{\text{cx}} \nu$, denoting by R a martingale Markov kernel such that $\eta R = \nu$, we have

$$\begin{aligned} \int_{\mathbb{R}^d} |x| 1_{\{|x| \geq K\}} \eta(dx) &= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} y R(x, dy) \right| 1_{\{|x| \geq K\}} \eta(dx) \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |y| 1_{\{|x| \geq K\}} R(x, dy) \eta(dx) \\ &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |y| 1_{\{|y| \geq \sqrt{K}\}} R(x, dy) \eta(dx) + \sqrt{K} \int_{\mathbb{R}^d} 1_{\{|x| \geq K\}} \eta(dx) \\ &\leq \int_{\mathbb{R}^d} |y| 1_{\{|y| \geq \sqrt{K}\}} \nu(dy) + \frac{\int_{\mathbb{R}^d} |x| \eta(dx)}{\sqrt{K}} \\ &\leq \int_{\mathbb{R}^d} |y| 1_{\{|y| \geq \sqrt{K}\}} \nu(dy) + \frac{\int_{\mathbb{R}^d} |y| \nu(dy)}{\sqrt{K}}. \end{aligned}$$

For $(\eta_n)_n$ a sequence in $\{\eta \in \underline{\mathcal{P}}(\mathbb{R}^d) : \eta \leq_{\text{cx}} \nu\}$ weakly converging to η_∞ , this implies uniform integrability ensuring that for $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ continuous and such that $\sup_{x \in \mathbb{R}^d} \frac{|\phi(x)|}{1+|x|} < \infty$, $\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} \phi(x) \eta_n(dx) = \int_{\mathbb{R}^d} \phi(x) \eta_\infty(dx)$. With Lemma A.1 below and the continuity of real valued convex functions on \mathbb{R}^d , we deduce that $\eta_\infty \in \underline{\mathcal{P}}(\nu)$. Hence $\underline{\mathcal{P}}(\nu)$ is compact for the weak convergence topology. \square

Proof of Theorem 2.1. For $\eta \in \underline{\mathcal{P}}(\mathbb{R}^d)$,

$$W_\varrho^\varrho(\mu, \eta) \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^\varrho \mu(dx) \eta(dy) \leq 2^{\varrho-1} \left(\int_{\mathbb{R}^d} |x|^\varrho \mu(dx) + \int_{\mathbb{R}^d} |y|^\varrho \eta(dy) \right),$$

where the right-hand side is finite if $\eta \in \underline{\mathcal{P}}(\nu)$ since $\sup_{\eta \in \underline{\mathcal{P}}(\nu)} \int_{\mathbb{R}^d} |x|^\varrho \eta(dx) = \int_{\mathbb{R}^d} |x|^\varrho \nu(dx)$. Since $\eta \mapsto W_\varrho^\varrho(\mu, \eta)$ is lower-semicontinuous for the weak convergence topology, by Lemma 2.2, there exists $\eta_\star \in \underline{\mathcal{P}}(\nu)$ such that $W_\varrho^\varrho(\mu, \eta_\star) = \inf_{\eta \in \underline{\mathcal{P}}(\nu)} W_\varrho^\varrho(\mu, \eta)$. Let P be a martingale Markov kernel such that $\eta_\star P = \nu$ and Q a Markov kernel such that $\mu Q = \eta_\star$ and $W_\varrho^\varrho(\mu, \eta_\star) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^\varrho Q(x, dy) \mu(dx)$. One has $\mu Q P = \eta_\star P = \nu$ and, by martingality of P ,

$$m_{QP}(x) = \int_{\mathbb{R}^d \times \mathbb{R}^d} z P(y, dz) Q(x, dy) = \int_{\mathbb{R}^d} y Q(x, dy).$$

With Jensen's inequality, we deduce that

$$W_\varrho^\varrho(\mu, \eta_\star) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^\varrho Q(x, dy) \mu(dx) \geq \int_{\mathbb{R}^d} \left| x - \int_{\mathbb{R}^d} y Q(x, dy) \right|^\varrho \mu(dx) = \mathcal{J}_\varrho(QP). \quad (2.1)$$

On the other hand, for any Markov kernel R such that $\mu R = \nu$, $m_R \# \mu \leq_{\text{cx}} \nu$ and $\mathcal{J}_\varrho(R) = \int_{\mathbb{R}^d} |x - m_R(x)|^\varrho \mu(dx) \geq W_\varrho^\varrho(\mu, m_R \# \mu)$. Hence

$$\inf_{R: \mu R = \nu} \mathcal{J}_\varrho(R) \geq \inf_{\eta \in \underline{\mathcal{P}}(\nu)} W_\varrho^\varrho(\mu, \eta) = W_\varrho^\varrho(\mu, \eta_\star) \geq \mathcal{J}_\varrho(QP) \geq \inf_{R: \mu R = \nu} \mathcal{J}_\varrho(R)$$

so that both infima are equal and $\mathcal{J}_\varrho(QP) = \inf_{R: \mu R = \nu} \mathcal{J}_\varrho(R)$. Moreover, the inequality in (2.1) is an equality. If $\varrho > 1$, by strict convexity of $x \mapsto |x|^\varrho$, this implies that $\mu(dx)$ a.e. $R(x, dy) = \delta_{m_{QP}(x)}(dy)$ so that $\eta_\star = \mu Q = m_{QP} \# \mu$.

For $\varrho > 1$, the uniqueness of m_{R_\star} is also obtained from the strict convexity of $x \mapsto |x|^\varrho$. Namely, for any optimal kernel R_\star we have

$$\begin{aligned} \mathcal{J}_\varrho((R_\star + QP)/2) &= \int_{\mathbb{R}^d} \left| x - \frac{m_{R_\star}(x) + m_{QP}(x)}{2} \right|^\varrho \mu(dx) \\ &\leq \int_{\mathbb{R}^d} \frac{1}{2} |x - m_{R_\star}(x)|^\varrho + \frac{1}{2} |x - m_{QP}(x)|^\varrho \mu(dx) \\ &= \frac{1}{2} (\mathcal{J}_\varrho(R_\star) + \mathcal{J}_\varrho(QP)) = \inf_{R: \mu R = \nu} \mathcal{J}_\varrho(R). \end{aligned}$$

Since $\mu \frac{R_\star + QP}{2} = \nu$, we necessarily have $\mathcal{J}_\varrho((R_\star + QP)/2) = \inf_{R:\mu R = \nu} \mathcal{J}_\varrho(R)$ and then $m_{R_\star}(x) = m_{QP}(x)$, $\mu(dx)$ -a.e. □

Remark 2.3. When $\varrho = 1$, let us give an example of non-uniqueness for the optimal functions m_R and the probability measures $\eta_\star \in \mathcal{P}(\nu)$ such that $W_1(\mu, \eta_\star) = \inf_{\eta \in \mathcal{P}(\nu)} W_1(\mu, \eta)$. Let $\mu(dx) = 1_{[0,1]}(dx)$ (resp. $\nu(dy) = 1_{[1,2]}(dy)$) be the uniform law on $[0, 1]$ (resp. $[1, 2]$). We have

$$\inf_{R:\mu R = \nu} \mathcal{J}_1(R) \geq \inf_{R:\mu R = \nu} \left| \int_{\mathbb{R}} x\mu(dx) - \int_{\mathbb{R}} m_R(x)\mu(dx) \right| = \left| \int_{\mathbb{R}} x\mu(dx) - \int_{\mathbb{R}} y\nu(dy) \right| = 1.$$

For $\lambda \in [0, 1]$, $R_\lambda(x, dy) = (1 - \lambda)\delta_{1+x}(dy) + \lambda\delta_{2-x}(dy)$ is such that $\mu R_\lambda = \nu$, $m_{R_\lambda}(x) = (1 + \lambda) + (1 - 2\lambda)x$ and $m_{R_\lambda} \# \mu$ is the uniform law on $[(1 + \lambda) \wedge (2 - \lambda), (1 + \lambda) \vee (2 - \lambda)]$. Using that $m_{R_\lambda}(x) \geq 1$ for $x \in (0, 1)$ for the first equality, we have

$$W_1(\mu, m_{R_\lambda} \# \mu) \leq \mathcal{J}_1(R_\lambda) = \int_0^1 1 + \lambda - 2\lambda x \, dx = 1 = \inf_{\eta \in \mathcal{P}(\nu)} W_1(\mu, \eta).$$

Thus all the kernels R_λ and pushforward measures $m_{R_\lambda} \# \mu$ are optimal.

Example 2.4. Let $\mu \leq_{cx} \nu$ and $\varrho \geq 1$. We assume that $\nu \in \mathcal{P}_\varrho$, which implies that $\mu \in \mathcal{P}_\varrho$. For $\alpha \in \mathbb{R}^d$, let μ^α be the image of μ by $x \mapsto x + \alpha$. Then, for any kernel R such that $\mu^\alpha R = \nu$,

$$\begin{aligned} \int_{\mathbb{R}^d} |x - m_R(x)|^\varrho \mu^\alpha(dx) &\geq \left| \int_{\mathbb{R}^d} x - m_R(x) \mu^\alpha(dx) \right|^\varrho \\ &= \left| \int_{\mathbb{R}^d} x \mu^\alpha(dx) - \int_{\mathbb{R}^d} y \nu(dy) \right|^\varrho = \left| \int_{\mathbb{R}^d} x \mu^\alpha(dx) - \int_{\mathbb{R}^d} x \mu(dx) \right|^\varrho = |\alpha|^\varrho. \end{aligned}$$

This lower bound is attained for $R(x, dy) = Q(x - \alpha, dy)$, where Q is any martingale Markov kernel such that $\mu Q = \nu$, since $m_R(x) = x - \alpha$ for this choice. Therefore, for $\varrho > 1$, $(\mu^\alpha)_{\mathcal{P}(\nu)}^\varrho = \mu$.

Let us observe that if $\mu, \nu \in \mathcal{P}_\varrho(\mathbb{R}^d)$ with $\varrho > 1$, then we have $\mu, \nu \in \mathcal{P}_{\varrho'}(\mathbb{R}^d)$ for any $\varrho' \in (1, \varrho)$. In general, as in the next example, $\mu_{\mathcal{P}(\nu)}^{\varrho'}$ is different from $\mu_{\mathcal{P}(\nu)}^\varrho$.

Example 2.5. Let $d = 2$, $\mu = \frac{1}{2}(\delta_{(1,0)} + \delta_{(0,a)})$ with $a \in \mathbb{R}$ and $\nu = \frac{1}{2}(\delta_{(1,0)} + \delta_{(-1,0)})$. For $\varrho > 1$, since, by Theorem 2.1, $\mu_{\mathcal{P}(\nu)}^\varrho$ is the image of μ by some transport map and $\mu_{\mathcal{P}(\nu)}^\varrho \leq_{cx} \nu$, one has $\mu_{\mathcal{P}(\nu)}^\varrho = \frac{1}{2}(\delta_{(x_\varrho,0)} + \delta_{(-x_\varrho,0)})$ for some $x_\varrho \in [0, 1]$. For $x \in [0, 1]$, since the distances between $(x, 0)$ and $(0, a)$ and between $(-x, 0)$ and $(0, a)$ are equal whereas $(x, 0)$ is closer to $(1, 0)$ than $(-x, 0)$, one has $2W_\varrho^\varrho(\mu, \frac{1}{2}(\delta_{(x,0)} + \delta_{(-x,0)})) = (1-x)^\varrho + (a^2 + x^2)^{\varrho/2}$. Since the unique minimizer of $x \mapsto (1-x)^2 + (a^2 + x^2)$ on $[0, 1]$ is $\frac{1}{2}$, one has $x_2 = \frac{1}{2}$ and $\mu_{\mathcal{P}(\nu)}^2 = \frac{1}{2}(\delta_{(1/2,0)} + \delta_{(-1/2,0)})$. Since the unique minimizer of $x \mapsto (1-x)^3 + (5 + x^2)^{3/2}$ on $[0, 1]$ is $\frac{1}{4}$, for $a \in \{-\sqrt{5}, \sqrt{5}\}$, one has $x_3 = \frac{1}{4}$ and $\mu_{\mathcal{P}(\nu)}^3 = \frac{1}{2}(\delta_{(1/4,0)} + \delta_{(-1/4,0)})$.

Nevertheless, the situation is strikingly different in dimension $d = 1$ where, according to Gozlan et al. [14] Theorem 1.5, the projection does not depend on ϱ . We are going to explicit this projection by characterizing its quantile function in terms of the quantile functions of μ and ν .

2.2. Dimension $d = 1$

Let $F_\mu(x) = \mu((-\infty, x])$ and $F_\nu(x) = \nu((-\infty, x])$ be the cumulative distribution functions and for $p \in (0, 1)$, $F_\mu^{-1}(p) = \inf\{x \in \mathbb{R} : F_\mu(x) \geq p\}$ and $F_\nu^{-1}(p) = \inf\{x \in \mathbb{R} : F_\nu(x) \geq p\}$ their left-continuous and non-decreasing generalised inverses also called quantile functions. The convex order is characterized as follows in terms of the quantile functions (see Theorem 3.A.5 [26]): for $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$,

$$\begin{aligned} \mu \leq_{cx} \nu \quad \text{iff} \quad & \int_0^1 F_\mu^{-1}(p) \, dp = \int_0^1 F_\nu^{-1}(p) \, dp \\ & \text{and } \forall q \in (0, 1), \quad \int_q^1 F_\mu^{-1}(p) \, dp \leq \int_q^1 F_\nu^{-1}(p) \, dp. \end{aligned} \tag{2.2}$$

Notice that, as a consequence of this characterization, if $\mu \leq_{\text{cx}} \nu$, then for $I, k \geq 1$, $\frac{1}{I} \sum_{i=1}^I \delta_{\int_{\frac{i-1}{I}}^{\frac{i}{I}} F_{\mu}^{-1}(u) du} \leq_{\text{cx}} \frac{1}{kI} \sum_{j=1}^{kI} \delta_{\int_{\frac{j-1}{kI}}^{\frac{j}{kI}} F_{\nu}^{-1}(u) du}$, as stated by Baker in Theorem 2.4.11 [6].

Theorem 2.6. For $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$, let ψ denote the convex hull (largest convex function bounded from above by) of the function $[0, 1] \ni q \mapsto \int_0^q F_{\mu}^{-1}(p) - F_{\nu}^{-1}(p) dp$. There exists a probability measure $\mu_{\underline{\mathcal{P}}(\nu)}$ such that $\forall q \in [0, 1]$, $\int_0^q F_{\mu_{\underline{\mathcal{P}}(\nu)}}^{-1}(p) dp = \int_0^q F_{\mu}^{-1}(p) dp - \psi(q)$. Moreover, $\mu_{\underline{\mathcal{P}}(\nu)} \in \underline{\mathcal{P}}(\nu)$ and for each $\varrho > 1$ such that $\mu, \nu \in \mathcal{P}_{\varrho}(\mathbb{R})$, $\mu_{\underline{\mathcal{P}}(\nu)}^{\varrho} = \mu_{\underline{\mathcal{P}}(\nu)}$. Last, $T(x) = F_{\mu}^{-1}(F_{\mu}(x)) - \psi'(F_{\mu}(x)-)$ is non-decreasing and is an optimal transport map: $T\# \mu = \mu_{\underline{\mathcal{P}}(\nu)}$ and for all $\varrho \geq 1$, $W_{\varrho}^{\varrho}(\mu, \mu_{\underline{\mathcal{P}}(\nu)}) = \int_{\mathbb{R}} |T(x) - x|^{\varrho} \mu(dx)$.

Remark 2.7. The fact that the Wasserstein projection $\mu_{\underline{\mathcal{P}}(\nu)}^{\varrho}$ does not depend on ϱ is in line with a remarkable feature of the classical optimal transport in dimension 1. Indeed, whatever $\varrho > 1$, the optimal W_{ϱ} coupling between μ and ν is the image π^{OT} of the uniform law on $[0, 1]$ by $p \mapsto (F_{\mu}^{-1}(p), F_{\nu}^{-1}(p))$ (see for instance Proposition 2.17 [25]). This is deeply related to the complete lattice structure of $\mathcal{P}_1(\mathbb{R})$ for the convex order (see top of p. 162 [20]). In fact, the set of images of $\pi \in \Pi(\mu, \nu)$ by $(x, y) \mapsto y - x$ has an infimum for the convex order. Strikingly, it turns out that this infimum is a minimum (given by π^{OT}), which ensures that the optimal coupling does not depend on ϱ . Below, to prove that $\mu_{\underline{\mathcal{P}}(\nu)}^{\varrho}$ does not depend on ϱ , we show that the set of images of $\pi \in \Pi(\mu, \eta)$ with $\eta \leq_{\text{cx}} \nu$ by $(x, y) \mapsto y - x$ also has a minimum for the convex order.

For probability measures $\mu_I = \sum_{i=1}^I p_i \delta_{x_i}$ (resp. $\nu_J = \sum_{j=1}^J q_j \delta_{y_j}$) on the real line with $(p_1, \dots, p_I) \in (0, 1]^I$ and $x_1 < x_2 < \dots < x_I$ (resp. $(q_1, \dots, q_J) \in (0, 1]^J$ and $y_1 < y_2 < \dots < y_J$), the continuous and piecewise affine function $q \mapsto \int_0^q F_{\mu_I}^{-1}(p) - F_{\nu_J}^{-1}(p) dp$ changes slope at $q \in \{\sum_{k=1}^i p_k : 1 \leq i \leq I - 1\} \cup \{\sum_{k=1}^j q_k : 1 \leq j \leq J - 1\}$ with a change equal to $\sum_{i=1}^{I-1} 1_{\{q = \sum_{k=1}^i p_k\}}(x_{i+1} - x_i) - \sum_{j=1}^{J-1} 1_{\{q = \sum_{k=1}^j q_k\}}(y_{j+1} - y_j)$ (which can be equal to zero if $q = \sum_{k=1}^i p_k = \sum_{k=1}^j q_k$ and $x_{i+1} - x_i = y_{j+1} - y_j$). Clearly, ψ is piecewise affine and changes slope at most at points $q \in \{\sum_{k=1}^i p_k : 1 \leq i \leq I - 1\}$ with changes not greater than $x_{i+1} - x_i$ so that $(\mu_I)_{\underline{\mathcal{P}}(\nu_J)} = \sum_{i=1}^I p_i \delta_{z_i}$ with $z_1 \leq z_2 \leq \dots \leq z_I$. The convex hull ψ can be computed by Andrew’s monotone chain algorithm and the points $(z_i)_{i \in I}$ are easily deduced.

The proof of Theorem 2.6 relies on the following lemma and is postponed after its proof.

Lemma 2.8. Let $\varrho > 1$ and $\mu, \nu \in \mathcal{P}_{\varrho}(\mathbb{R})$. Then $(0, 1) \ni p \mapsto F_{\mu_{\underline{\mathcal{P}}(\nu)}^{\varrho}}^{-1}(p) - F_{\mu}^{-1}(p)$ is non-increasing.

Proof. It is enough to check that if $p \mapsto F_{\eta}^{-1}(p) - F_{\mu}^{-1}(p)$ is not non-increasing for some $\eta \in \underline{\mathcal{P}}(\nu)$, one can find $\tilde{\eta} \in \underline{\mathcal{P}}(\nu)$ such that $W_{\varrho}^{\varrho}(\mu, \tilde{\eta}) < W_{\varrho}^{\varrho}(\mu, \eta)$ where, according to Proposition 2.17 [25], $W_{\varrho}^{\varrho}(\mu, \eta) = \int_0^1 |F_{\eta}^{-1}(p) - F_{\mu}^{-1}(p)|^{\varrho} dp$. With the left-continuity of $p \mapsto F_{\eta}^{-1}(p) - F_{\mu}^{-1}(p)$, the lack of monotonicity of this function is equivalent to

$$0 < \int_{(0,1)^2} 1_{I_{\eta}}(p, q) dp dq \quad \text{where } I_{\eta} = \{(p, q) : (p - q)(F_{\eta}^{-1}(p) - F_{\mu}^{-1}(p) - F_{\eta}^{-1}(q) + F_{\mu}^{-1}(q)) > 0\}.$$

Let $\alpha(p, q) = 1_{I_{\eta}}(p, q) \frac{F_{\eta}^{-1}(p) - F_{\mu}^{-1}(p) - F_{\eta}^{-1}(q) + F_{\mu}^{-1}(q)}{2(F_{\eta}^{-1}(p) - F_{\eta}^{-1}(q))}$, where one easily checks that the denominator does not vanish on I_{η} and that $0 \leq \alpha(p, q) = \alpha(q, p) < 1$. For $(p, q) \in I_{\eta}$,

$$\alpha(p, q)F_{\eta}^{-1}(q) + (1 - \alpha(q, p))F_{\eta}^{-1}(p) - F_{\mu}^{-1}(p) = \frac{F_{\eta}^{-1}(p) - F_{\mu}^{-1}(p) + F_{\eta}^{-1}(q) - F_{\mu}^{-1}(q)}{2},$$

so that by strict convexity,

$$\begin{aligned} & \frac{1}{2} (|F_{\eta}^{-1}(p) - F_{\mu}^{-1}(p)|^{\varrho} + |F_{\eta}^{-1}(q) - F_{\mu}^{-1}(q)|^{\varrho}) \\ & > |\alpha(p, q)F_{\eta}^{-1}(q) + (1 - \alpha(q, p))F_{\eta}^{-1}(p) - F_{\mu}^{-1}(p)|^{\varrho}. \end{aligned}$$

With Jensen’s inequality, we deduce that

$$\begin{aligned} W_\varrho^\eta(\mu, \eta) &= \frac{1}{2} \int_{(0,1)^2} |F_\eta^{-1}(p) - F_\mu^{-1}(p)|^\varrho + |F_\eta^{-1}(q) - F_\mu^{-1}(q)|^\varrho dp dq \\ &> \int_{(0,1)^2} |\alpha(p, q)F_\eta^{-1}(q) + (1 - \alpha(q, p))F_\eta^{-1}(p) - F_\mu^{-1}(p)|^\varrho dp dq \\ &\geq \int_0^1 \left| \int_0^1 \alpha(p, q)F_\eta^{-1}(q) + (1 - \alpha(q, p))F_\eta^{-1}(p) dq - F_\mu^{-1}(p) \right|^\varrho dp. \end{aligned}$$

The right-hand side is not smaller than $W_\varrho^\eta(\mu, \tilde{\eta})$ where $\tilde{\eta}$ denotes the image of the Lebesgue measure on $(0, 1)$ by $p \mapsto \int_0^1 \alpha(p, q)F_\eta^{-1}(q) + (1 - \alpha(q, p))F_\eta^{-1}(p) dq$. For $\phi : \mathbb{R} \rightarrow \mathbb{R}$ convex and such that $\sup_{x \in \mathbb{R}} \frac{|\phi(x)|}{1+|x|^\varrho} < \infty$, by Jensen’s inequality,

$$\int_{\mathbb{R}} \phi(x)\tilde{\eta}(dx) \leq \int_{(0,1)^2} \alpha(p, q)\phi(F_\eta^{-1}(q)) + (1 - \alpha(q, p))\phi(F_\eta^{-1}(p))dqdp = \int_0^1 \phi(F_\eta^{-1}(q)) dq.$$

Since the right-hand side is equal to $\int_{\mathbb{R}} \phi(x)\eta(dx)$, by Lemma A.1 below, one has $\tilde{\eta} \in \mathcal{P}(v)$. □

Proof of Theorem 2.6. Let U be uniformly distributed on $(0, 1)$.

Since for all $q \in [0, 1]$, $\int_0^q F_\mu^{-1}(p) - F_\nu^{-1}(p) dp \geq \int_0^q F_\mu^{-1}(p) dp - q \int_0^1 F_\nu^{-1}(p) dp$ where the right-hand side is a convex function of q , one has $\psi(0) = 0$ and $\psi(1) = \int_0^1 F_\mu^{-1}(p) - F_\nu^{-1}(p) dp$. By Lemma A.2 below, the convexity of both $q \mapsto \int_0^q F_\mu^{-1}(p) dp$ and $q \mapsto \int_0^q F_\nu^{-1}(p) dp$ implies that $q \mapsto \int_0^q F_\mu^{-1}(p) dp - \psi(q)$ is convex. Let f denote the left-hand derivative of this function and $\mu_{\mathcal{P}(v)}$ the probability distribution of $f(U)$. By Lemma A.3 below, f is equal to $F_{\mu_{\mathcal{P}(v)}}^{-1}$ so that $\forall q \in [0, 1]$, $\int_0^q F_{\mu_{\mathcal{P}(v)}}^{-1}(p) dp = \int_0^q F_\mu^{-1}(p) dp - \psi(q)$.

Let $q \in [0, 1]$. Since $\psi(q) \leq \int_0^q F_\mu^{-1}(p) - F_\nu^{-1}(p) dp$ with equality when $q = 1$, one has $\int_0^q F_{\mu_{\mathcal{P}(v)}}^{-1}(p) dp = \int_0^q F_\mu^{-1}(p) dp - \psi(q) \geq \int_0^q F_\nu^{-1}(p) dp$ with equality when $q = 1$ so that by (2.2), $\mu_{\mathcal{P}(v)} \leq_{cx} \nu$. By concavity of $[0, 1] \ni q \mapsto -\psi(q) = \int_0^q F_{\mu_{\mathcal{P}(v)}}^{-1}(p) - F_\mu^{-1}(p) dp$, the left-continuous function $(0, 1) \ni p \mapsto F_{\mu_{\mathcal{P}(v)}}^{-1}(p) - F_\mu^{-1}(p)$ is non-increasing.

The set

$$\tilde{\mathcal{P}}(v) := \{ \eta \in \mathcal{P}(v) : (0, 1) \ni p \mapsto F_\eta^{-1}(p) - F_\mu^{-1}(p) \text{ is non-increasing} \}$$

is not empty since $\mu_{\mathcal{P}(v)}, \delta_{\int_{\mathbb{R}} x\nu(dx)} \in \tilde{\mathcal{P}}(v)$. Let $D(\eta)$ denote the distribution of $F_\eta^{-1}(1 - U) - F_\mu^{-1}(1 - U)$ for $\eta \in \tilde{\mathcal{P}}(v)$. For all $\eta \in \tilde{\mathcal{P}}(v)$, $\int_{\mathbb{R}} |x|D(\eta)(dx) < \infty$ and $\int_{\mathbb{R}} xD(\eta)(dx) = \mathbb{E}[F_\eta^{-1}(1 - U) - F_\mu^{-1}(1 - U)] = \int_{\mathbb{R}} x(v - \mu)(dx)$. By Lemma A.4 below, the set $\{D(\eta) : \eta \in \tilde{\mathcal{P}}(v)\}$ admits an infimum π for the convex order and for all $q \in [0, 1]$, $\int_q^1 F_\pi^{-1}(p) dp = \inf_{\eta \in \tilde{\mathcal{P}}(v)} \int_q^1 F_{D(\eta)}^{-1}(p) dp$. For $\eta \in \tilde{\mathcal{P}}(v)$, since $(0, 1) \ni p \mapsto F_\eta^{-1}(1 - p) - F_\mu^{-1}(1 - p)$ is non-decreasing, by Lemma A.3, $p \mapsto F_{D(\eta)}^{-1}(p)$ and $p \mapsto F_\eta^{-1}(1 - p) - F_\mu^{-1}(1 - p)$ coincide away from the at most countable set of their common discontinuities, with the former left-continuous and the latter right-continuous. Hence for $q \in [0, 1]$,

$$\int_q^1 F_\pi^{-1}(p) dp = \inf_{\eta \in \tilde{\mathcal{P}}(v)} \int_0^{1-q} F_\eta^{-1}(p) - F_\mu^{-1}(p) dp = - \sup_{\eta \in \tilde{\mathcal{P}}(v)} \int_0^{1-q} F_\mu^{-1}(p) - F_\eta^{-1}(p) dp$$

where the right-hand side is not greater than $\int_0^{1-q} F_{\mu_{\mathcal{P}(v)}}^{-1}(p) - F_\mu^{-1}(p) dp$ since $\mu_{\mathcal{P}(v)} \in \tilde{\mathcal{P}}(v)$. Since $\eta \in \tilde{\mathcal{P}}(v)$ iff $F_\mu^{-1} - F_\eta^{-1}$ is non-decreasing, $\int_0^1 |F_\eta^{-1}(p)| dp < \infty$, $\int_0^1 F_\eta^{-1}(p) dp = \int_0^1 F_\nu^{-1}(p) dp$ and for all $q \in [0, 1]$, $\int_0^{1-q} F_\eta^{-1}(p) dp \geq \int_0^{1-q} F_\nu^{-1}(p) dp$ (see (2.2)), the definition of ψ implies that for all $q \in [0, 1]$, $\sup_{\eta \in \tilde{\mathcal{P}}(v)} \int_0^{1-q} F_\mu^{-1}(p) - F_\eta^{-1}(p) dp \leq \psi(1 - q) = \int_0^{1-q} F_\mu^{-1}(p) - F_{\mu_{\mathcal{P}(v)}}^{-1}(p) dp$. Hence $\int_q^1 F_\pi^{-1}(p) dp = \int_0^{1-q} F_{\mu_{\mathcal{P}(v)}}^{-1}(p) - F_\mu^{-1}(p) dp = \int_q^1 F_{D(\mu_{\mathcal{P}(v)})}^{-1}(p) dp$ for all $q \in [0, 1]$ which ensures that π is the distribution $D(\mu_{\mathcal{P}(v)})$ of $F_{\mu_{\mathcal{P}(v)}}^{-1}(1 - U) - F_\mu^{-1}(1 - U)$. Therefore, if $\varrho > 1$

is such that $\mu, \nu \in \mathcal{P}_\varrho(\mathbb{R})$,

$$\begin{aligned} W_\varrho^e(\mu, \mu_{\underline{\mathcal{P}}(\nu)}) &= \mathbb{E}[|F_{\mu_{\underline{\mathcal{P}}(\nu)}}^{-1}(1-U) - F_\mu^{-1}(1-U)|^\varrho] = \int_{\mathbb{R}} |x|^\varrho \pi(dx) \\ &\leq \inf_{\eta \in \underline{\mathcal{P}}(\nu)} \mathbb{E}[|F_\mu^{-1}(1-U) - F_\eta^{-1}(1-U)|^\varrho] = \inf_{\eta \in \underline{\mathcal{P}}(\nu)} W_\varrho^e(\mu, \eta) = \inf_{\eta \in \underline{\mathcal{P}}(\nu)} W_\varrho^e(\mu, \eta), \end{aligned}$$

where we used the definition of π and the convexity of $\mathbb{R} \ni x \mapsto |x|^\varrho$ for the inequality and Lemma 2.8 for the final equality. Since, by Theorem 2.1, $\mu_{\underline{\mathcal{P}}(\nu)}^e$ is the unique minimizer of $W_\varrho^e(\mu, \eta)$ on $\underline{\mathcal{P}}(\nu)$, we conclude that $\mu_{\underline{\mathcal{P}}(\nu)} = \mu_{\underline{\mathcal{P}}(\nu)}^e$.

From the left-continuity of the quantile functions, we get $F_{\mu_{\underline{\mathcal{P}}(\nu)}}^{-1}(p) = F_\mu^{-1}(p) - \psi'(p-)$ for $p \in (0, 1)$, and this function is non-decreasing. Thus, T is nondecreasing. To conclude the proof, it is now sufficient to check that $T(F_\mu^{-1}(p)) = F_\mu^{-1}(p) - \psi'(p-)$ for a.e. $p \in (0, 1)$. Indeed, combined with the inverse transform sampling and Proposition 2.17 [25], this ensures that $T\#\mu = \mu_{\underline{\mathcal{P}}(\nu)}$ and

$$W_\varrho^e(\mu, \mu_{\underline{\mathcal{P}}(\nu)}) = \int_0^1 |F_\mu^{-1}(p) - T(F_\mu^{-1}(p))|^\varrho dp = \int_{\mathbb{R}} |x - T(x)|^\varrho \mu(dx).$$

By definition of the quantile function F_μ^{-1} , for all $x \in \mathbb{R}$, $F_\mu^{-1}(F_\mu(x)) \leq x$ and by right-continuity of F_μ , for all $p \in (0, 1)$, $F_\mu(F_\mu^{-1}(p)) \geq p$. With the monotonicity of F_μ we deduce that for all $x \in \mathbb{R}$ such that $F_\mu(x) \in (0, 1)$, $F_\mu(F_\mu^{-1}(F_\mu(x))) = F_\mu(x)$. Therefore, if $p \in (0, 1)$ is such that $p = F_\mu(x)$ for some $x \in \mathbb{R}$, then $T(F_\mu^{-1}(p)) = F_\mu^{-1}(p) - \psi'(p-)$. Otherwise, $p \in [F_\mu(x-), F_\mu(x)]$ for some $x \in \mathbb{R}$ such that $\mu(\{x\}) > 0$. We observe that $F_\mu^{-1}(q)$ and $\psi'(q-)$ are constant on $(F_\mu(x-), F_\mu(x)]$ since $q \mapsto \int_0^q F_\mu^{-1}(u) - F_\nu^{-1}(u) du$ is concave on this interval. For $p \in (F_\mu(x-), F_\mu(x)]$, we have $F_\mu^{-1}(p) = x$ and we get $T(F_\mu^{-1}(p)) = F_\mu^{-1}(F_\mu(x)) - \psi'(F_\mu(x)-) = F_\mu^{-1}(p) - \psi'(p-)$. Therefore the equality holds for p outside the countable set $\{F_\mu(x-) : x \in \mathbb{R} \text{ s.t. } \mu(\{x\}) > 0\}$. \square

3. Approximations in the convex order

The next proposition is the key result to construct approximations of probability measures that preserve the convex order.

Proposition 3.1. *Let $\varrho \geq 1$, $\mu, \nu, \mu_I, \nu_J \in \mathcal{P}_\varrho(\mathbb{R}^d)$ such that $\mu \leq_{\text{cx}} \nu$. Then, we have*

$$W_\varrho(\mu, (\mu_I)_{\underline{\mathcal{P}}(\nu_J)}^\varrho) \leq 2W_\varrho(\mu, \mu_I) + W_\varrho(\nu, \nu_J),$$

where, for $\varrho = 1$, by a slight abuse of notation, $(\mu_I)_{\underline{\mathcal{P}}(\nu_J)}^1$ denotes any $\eta_\star \in \underline{\mathcal{P}}(\nu_J)$ such that $W_1(\mu_I, \eta_\star) = \inf_{\eta \in \underline{\mathcal{P}}(\nu_J)} W_1(\mu_I, \eta)$.

Let $\mu, \nu \in \mathcal{P}_\varrho(\mathbb{R}^d)$ be such that $\mu \leq_{\text{cx}} \nu$. From Proposition 3.1, if we have approximations μ_I and ν_J that satisfy $W_\varrho(\mu, \mu_I) \xrightarrow{I \rightarrow +\infty} 0$ and $W_\varrho(\nu, \nu_J) \xrightarrow{J \rightarrow +\infty} 0$, then $(\mu_I)_{\underline{\mathcal{P}}(\nu_J)}^\varrho$ also approximates μ since we have $W_\varrho(\mu, (\mu_I)_{\underline{\mathcal{P}}(\nu_J)}^\varrho) \xrightarrow{I, J \rightarrow +\infty} 0$. In particular, if we take i.i.d. samples $(X_i)_{i \geq 1}$ (resp. $(Y_j)_{j \geq 1}$) distributed according to μ (resp. ν), the empirical measure $\mu_I = \frac{1}{I} \sum_{i=1}^I \delta_{X_i}$ (resp. $\nu_J = \frac{1}{J} \sum_{j=1}^J \delta_{Y_j}$) satisfy $W_\varrho(\mu, \mu_I) \xrightarrow{I \rightarrow +\infty} 0$ (resp. $W_\varrho(\nu, \nu_J) \xrightarrow{J \rightarrow +\infty} 0$) almost surely. Indeed, the law of large numbers gives the almost sure weak convergence of μ_I towards μ as well as the almost sure convergence of $\frac{1}{I} \sum_{i=1}^I |X_i|^\varrho$ to $\int_{\mathbb{R}^d} |x|^\varrho \mu(dx)$. By Proposition 7.1.5 of [4], we get $W_\varrho(\mu, \mu_I) \xrightarrow{I \rightarrow +\infty} 0$ almost surely. Under more restrictive assumptions on the measures μ and ν , we can have almost sure estimates on the rate of convergence. Let us assume that μ is such that $\mathcal{E}_{\alpha, \gamma} = \int_{\mathbb{R}^d} e^{\gamma|x|^\alpha} \mu(dx) < \infty$ for some $\alpha > \varrho$ and $\gamma > 0$. Then, by Theorem 2 of Fournier and Guillin [12], there are constants $c, C > 0$ depending on $\varrho, d, \alpha, \gamma, \mathcal{E}_{\alpha, \gamma}$ such that

$$\forall x \in (0, 1), \mathbb{P}(W_\varrho(\mu, \mu_I) > x) = \mathbb{P}(W_\varrho^e(\mu, \mu_I) > x^\varrho) \leq C \exp(-cIx^{d\nu(2\varrho)}).$$

Therefore we have $\sum_{I=2}^\infty \mathbb{P}(W_\varrho(\mu, \mu_I) > (\frac{2 \log(I)}{cI})^{\frac{1}{d\nu(2\varrho)}}) \leq C \sum_{I=2}^\infty I^{-2} < \infty$, which gives that almost surely, there exists I_0 such that $\forall I \geq I_0, W_\varrho(\mu, \mu_I) \leq (\frac{2 \log(I)}{cI})^{\frac{1}{d\nu(2\varrho)}}$. Since $x \mapsto e^{\gamma|x|^\alpha}$ is convex, $\int_{\mathbb{R}^d} e^{\gamma|x|^\alpha} \nu(dx) < \infty \implies$

$\int_{\mathbb{R}^d} e^{\gamma|x|^\alpha} \mu(dx) < \infty$, in which case we have both $W_\varrho(\mu, \mu_I) = \mathcal{O}\left(\left(\frac{\log(I)}{I}\right)^{\frac{1}{d\nu(2\varrho)}}\right)$ and $W_\varrho(v, v_J) = \mathcal{O}\left(\left(\frac{\log(J)}{J}\right)^{\frac{1}{d\nu(2\varrho)}}\right)$ and thus

$$W_\varrho(\mu, (\mu_I)_{\underline{\mathcal{P}}(v_J)}^{\varrho})_{I, J \rightarrow +\infty} = \mathcal{O}\left(\left(\frac{\log(I \wedge J)}{I \wedge J}\right)^{\frac{1}{d\nu(2\varrho)}}\right), \quad \text{a.s.}$$

Theorem 2 of [12] also gives upper bounds of $\mathbb{P}(W_\varrho(\mu, \mu_I) > x)$ under different weaker assumptions on μ . We can repeat the same argument in those cases and get a weaker rate of convergence of $W_\varrho(\mu, \mu_I)$ towards 0.

We now briefly consider the multi-marginal case. Let $\varrho \geq 1$, $\ell \geq 2$, I_1, \dots, I_ℓ be positive integers and μ^1, \dots, μ^ℓ be probability measures on \mathbb{R}^d such that $\mu^1 \leq_{\text{cx}} \dots \leq_{\text{cx}} \mu^\ell$ and $\int_{\mathbb{R}^d} |x|^\varrho \mu^\ell(dx) < \infty$. We consider for $1 \leq k \leq \ell$, $\mu_{I_k}^k = \frac{1}{I_k} \sum_{i=1}^{I_k} \delta_{X_i^k}$ the empirical measure of an i.i.d. sample $X_1^k, \dots, X_{I_k}^k$ distributed according to μ^k . Let us set $\mu_{I_\ell}^{\ell, \varrho} = \mu_{I_\ell}^\ell$ and define (using for $\varrho = 1$ the abuse of notation made in Proposition 3.1) by backward induction for $k \in \{1, \dots, \ell - 1\}$, the projection $\mu_{I_k, \dots, I_\ell}^{k, \varrho}$ of $\mu_{I_k}^k$ on the set $\underline{\mathcal{P}}(\mu_{I_{k+1}, \dots, I_\ell}^{k+1, \varrho})$ for the W_ϱ -Wasserstein distance. Then, by Proposition 3.1, we have for $1 \leq k \leq \ell - 1$,

$$W_\varrho(\mu^k, \mu_{I_k, \dots, I_\ell}^{k, \varrho}) \leq 2W_\varrho(\mu^k, \mu_{I_k}^k) + W_\varrho(\mu^{k+1}, \mu_{I_{k+1}, \dots, I_\ell}^{k+1, \varrho}).$$

Therefore, we deduce by induction that

$$W_\varrho(\mu^k, \mu_{I_k, \dots, I_\ell}^{k, \varrho}) \leq 2 \sum_{k'=k}^{\ell-1} W_\varrho(\mu^{k'}, \mu_{I_{k'}}^{k'}) + W_\varrho(\mu^\ell, \mu_{I_\ell}^\ell).$$

We eventually get the following result.

Corollary 3.2. *Let $\varrho \geq 1$, μ^1, \dots, μ^ℓ be probability measures on \mathbb{R}^d such that $\mu^1 \leq_{\text{cx}} \dots \leq_{\text{cx}} \mu^\ell$ and $\int_{\mathbb{R}^d} |x|^\varrho \mu^\ell(dx) < \infty$. For $1 \leq k \leq \ell$, let $(X_i^k)_{i \geq 1}$ be i.i.d. according to μ^k and $\mu_{I_k}^k = \frac{1}{I_k} \sum_{i=1}^{I_k} \delta_{X_i^k}$ for $I_k \in \mathbb{N}^*$.*

Then, as $I_1, \dots, I_\ell \rightarrow +\infty$, $\sum_{k=1}^\ell W_\varrho(\mu^k, \mu_{I_k, \dots, I_\ell}^{k, \varrho})$ converges almost surely to 0. Besides, if $\int_{\mathbb{R}^d} e^{\gamma|x|^\alpha} \mu^\ell(dx)$ for some $\alpha > \varrho$ and $\gamma > 0$, we have a.s.

$$\sum_{k=1}^\ell W_\varrho(\mu^k, \mu_{I_k, \dots, I_\ell}^{k, \varrho})_{\min_{k=1, \dots, \ell} I_k \rightarrow +\infty} = \mathcal{O}\left(\left(\frac{\log(\min_{k=1, \dots, \ell} I_k)}{\min_{k=1, \dots, \ell} I_k}\right)^{\frac{1}{d\nu(2\varrho)}}\right).$$

Proof of Proposition 3.1. We consider $\varrho > 1$. Let $Q_{\mu_I}^\varrho$ (resp. Q_v^ϱ) be a Markov kernel such that $\mu_I(dx) Q_{\mu_I}^\varrho(x, dy)$ (resp. $\nu(dx) Q_v^\varrho(x, dy)$) is an optimal transport plan for $W_\varrho(\mu_I, \mu)$ (resp. $W_\varrho(v, v_J)$). Let $R(x, dy)$ be a martingale Markov kernel such that $\nu = \mu R$. We observe that $Q_{\mu_I}^\varrho R Q_v^\varrho$ is a Markov kernel such that $\mu_I Q_{\mu_I}^\varrho R Q_v^\varrho = \mu R Q_v^\varrho = \nu Q_v^\varrho = \nu_J$. By Theorem 2.1, then using the martingale property of R , the Jensen and Minkowski inequalities, we get

$$\begin{aligned} W_\varrho(\mu_I, (\mu_I)_{\underline{\mathcal{P}}(v_J)}^{\varrho}) &\leq \mathcal{J}_\varrho^{1/\varrho}(Q_{\mu_I}^\varrho R Q_v^\varrho) \\ &= \left(\int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} (x - w + w - y) Q_{\mu_I}^\varrho(x, dw) R(w, dz) Q_v^\varrho(z, dy) \right|^\varrho \mu_I(dx) \right)^{1/\varrho} \\ &= \left(\int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} (x - w + z - y) Q_{\mu_I}^\varrho(x, dw) R(w, dz) Q_v^\varrho(z, dy) \right|^\varrho \mu_I(dx) \right)^{1/\varrho} \\ &\leq \left(\int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} |x - w + z - y|^\varrho Q_{\mu_I}^\varrho(x, dw) R(w, dz) Q_v^\varrho(z, dy) \mu_I(dx) \right)^{1/\varrho} \\ &\leq \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - w|^\varrho Q_{\mu_I}^\varrho(x, dw) \mu_I(dx) \right)^{1/\varrho} + \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |z - y|^\varrho \nu(dz) Q_v^\varrho(z, dy) \right)^{1/\varrho} \\ &= W_\varrho(\mu_I, \mu) + W_\varrho(v_J, \nu). \end{aligned}$$

The claim follows since $W_\varrho(\mu, (\mu_I)_{\underline{\mathcal{P}}(v_J)}^{\varrho}) \leq W_\varrho(\mu, \mu_I) + W_\varrho(\mu_I, (\mu_I)_{\underline{\mathcal{P}}(v_J)}^{\varrho})$. □

4. Wasserstein projection of ν on the set of probability measures larger than μ in the convex order

Let $\mu, \nu \in \mathcal{P}_\varrho(\mathbb{R}^d)$. We have just presented a construction of a measure $\mu_{\underline{\mathcal{P}}(\nu)}^\varrho$ such that $\mu_{\underline{\mathcal{P}}(\nu)}^\varrho \leq_{\text{cx}} \nu$. Then, a natural question is: can we construct similarly a measure $\nu_{\bar{\mathcal{P}}(\mu)}^\varrho$ such that $\mu \leq_{\text{cx}} \nu_{\bar{\mathcal{P}}(\mu)}^\varrho$? Let us start again with two empirical measures $\mu^I = \frac{1}{I} \sum_{i=1}^I \delta_{X_i}$ and $\nu^J = \frac{1}{J} \sum_{j=1}^J \delta_{Y_j}$. A natural construction would be to take $(\nu^J)_{\bar{\mathcal{P}}(\mu^I)}^\varrho = \frac{1}{J} \sum_{j=1}^J \delta_{\tilde{Y}_j}$, where $(\tilde{Y}_j, j = 1, \dots, J) \in (\mathbb{R}^d)^J$ minimizes $\sum_{j=1}^J |\tilde{Y}_j - Y_j|^\varrho$ under the constraint $\mu^I \leq_{\text{cx}} \frac{1}{J} \sum_{j=1}^J \delta_{\tilde{Y}_j}$ (this constraint can always be satisfied when $J = I$ by taking $\tilde{Y}_j = X_j$ for $j = 1, \dots, J$ or when $J \geq d + 1$ by taking $\tilde{Y}_j, j = 1, \dots, d + 1$ as the images of the vertices of the canonical simplex by some similarity transformation). The analogous construction for $\nu_{\bar{\mathcal{P}}(\mu)}^\varrho$ would be then to take $\nu_{\bar{\mathcal{P}}(\mu)}^\varrho = T\#\nu$, where $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a measurable map that minimizes $\int_{\mathbb{R}^d} |y - T(y)|^\varrho \nu(dy)$, under the constraint $\mu \leq_{\text{cx}} T\#\nu$. More generally, we define

$$\nu_{\bar{\mathcal{P}}(\mu)}^\varrho := \arg \min_{\eta \in \bar{\mathcal{P}}(\mu)} W_\varrho(\nu, \eta) \quad \text{where } \bar{\mathcal{P}}(\mu) = \{\eta \in \mathcal{P}(\mathbb{R}^d) : \mu \leq_{\text{cx}} \eta\}.$$

Let us now assume that $\varrho > 1$. The latter problem coincides with the former one when ν is absolutely continuous with respect to the Lebesgue measure (i.e. $\nu(A) = 0$ for any Borel set A with zero Lebesgue measure), since we know in this case that the optimal coupling for the Wasserstein distance W_ϱ is given by a transport map, see e.g. Theorem 6.2.4 in [4]. We now check that it is well defined. Let $(\eta_n)_{n \geq 1} \in (\mathcal{P}_\varrho(\mathbb{R}^d))^{\mathbb{N}}$ be such that $\eta_n \in \bar{\mathcal{P}}(\mu)$ and $W_\varrho(\nu, \eta_n) \xrightarrow{n \rightarrow +\infty} \inf_{\eta \in \bar{\mathcal{P}}(\mu)} W_\varrho(\nu, \eta)$. Let $\pi_n \in \Pi(\nu, \eta_n)$ denote an optimal transport plan between ν and η_n for W_ϱ . We have $(\int |x|^\varrho \eta_n(x))^{1/\varrho} = W_\varrho(\eta_n, \delta_0) \leq W_\varrho(\eta_n, \nu) + W_\varrho(\nu, \delta_0)$: the boundedness of the moments ensures that there is a subsequence such that $\pi_{\varphi(n)}$ and $\eta_{\varphi(n)}$ weakly converges to π_∞ and η_∞ . This gives $\inf_{\eta \in \bar{\mathcal{P}}(\mu)} W_\varrho^\eta(\nu, \eta) \geq \lim_{n \rightarrow +\infty} \int (|x - y|^\varrho \wedge K) \pi_{\varphi(n)}(dx, dy) = \int (|x - y|^\varrho \wedge K) \pi_\infty(dx, dy)$ for any $K > 0$. By monotone convergence, we deduce that $\inf_{\eta \in \bar{\mathcal{P}}(\mu)} W_\varrho(\nu, \eta) \geq \int |x - y|^\varrho \pi_\infty(dx, dy)$. Clearly, π_∞ is a coupling between ν and η_∞ . Besides, from the uniform integrability given by the bounds on the ϱ -th moment, we get that for any convex function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $\sup_{x \in \mathbb{R}^d} \frac{|\phi(x)|}{1+|x|} < \infty$, $\int \phi(x) \mu(dx) \leq \int \phi(x) \eta_{\varphi(n)}(dx) \xrightarrow{n \rightarrow +\infty} \int \phi(x) \eta_\infty(dx)$. Therefore, by Lemma A.1 below, $\eta_\infty \in \bar{\mathcal{P}}(\mu)$, which shows the existence of a minimum. When ν is absolutely continuous with respect to the Lebesgue measure, we can show that this minimum is unique. Let us consider $\eta_1, \eta_2 \in \bar{\mathcal{P}}(\mu)$ such that $W_\varrho(\nu, \eta_1) = W_\varrho(\nu, \eta_2) = \inf_{\eta \in \bar{\mathcal{P}}(\mu)} W_\varrho(\nu, \eta)$. One has $\frac{1}{2}(\eta_1 + \eta_2) \in \bar{\mathcal{P}}(\mu)$, and, by Lemma A.5 below, we get $W_\varrho(\nu, \frac{1}{2}(\eta_1 + \eta_2)) \leq \inf_{\eta \in \bar{\mathcal{P}}(\mu)} W_\varrho(\nu, \eta)$ and $\eta_1 = \eta_2$ since the inequality is necessarily an equality. In dimension 1, uniqueness still holds without any assumption on ν . Indeed, by (2.2), the probability measure $\bar{\eta}_{12}$ defined by $F_{\bar{\eta}_{12}}^{-1} = \frac{1}{2}(F_{\eta_1}^{-1} + F_{\eta_2}^{-1})$ is such that $\mu \leq_{\text{cx}} \bar{\eta}_{12}$. Again by Lemma A.5, $W_\varrho(\nu, \bar{\eta}_{12}) \leq \inf_{\eta \in \bar{\mathcal{P}}(\mu)} W_\varrho(\nu, \eta)$ and $\eta_1 = \eta_2$ since the inequality is necessarily an equality. We have just proved the following result.

Theorem 4.1. *For $\varrho > 1$, if $\mu, \nu \in \mathcal{P}_\varrho(\mathbb{R}^d)$, then $\inf_{\eta \in \bar{\mathcal{P}}(\mu)} W_\varrho^\eta(\nu, \eta)$ is attained by some probability measure $\nu_{\bar{\mathcal{P}}(\mu)}^\varrho$ which is unique when ν is absolutely continuous with respect to the Lebesgue measure or $d = 1$.*

Let us check now that in dimension $d = 1$, like $\mu_{\underline{\mathcal{P}}(\nu)}^\varrho$, the projection $\nu_{\bar{\mathcal{P}}(\mu)}^\varrho$ does not depend on $\varrho > 1$.

Proposition 4.2. *If $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$, then there is a probability $\nu_{\bar{\mathcal{P}}(\mu)}$ such that for all $q \in [0, 1]$, $\int_q^1 F_{\nu_{\bar{\mathcal{P}}(\mu)}}^{-1}(p) dp = \tilde{\psi}(q) + \int_q^1 F_\nu^{-1}(p) dp$ where $\tilde{\psi}$ denotes the concave hull of the function $q \mapsto \int_q^1 F_\mu^{-1}(p) - F_\nu^{-1}(p) dp$. Moreover, $\nu_{\bar{\mathcal{P}}(\mu)}^\varrho = \nu_{\bar{\mathcal{P}}(\mu)}$ for each $\varrho > 1$ such that $\mu, \nu \in \mathcal{P}_\varrho(\mathbb{R})$.*

Proof. Let

$$\tilde{\mathcal{P}}(\mu) := \{\eta \in \bar{\mathcal{P}}(\mu) \cap \mathcal{P}_1(\mathbb{R}) : (0, 1) \ni p \mapsto F_\eta^{-1}(p) - F_\nu^{-1}(p) \text{ is non-decreasing}\}.$$

Let $\tilde{\psi}$ denote the concave hull (smallest concave function larger than) of the function $q \mapsto \int_q^1 F_\mu^{-1}(p) - F_\nu^{-1}(p) dp$. There is a probability measure $\nu_{\tilde{\mathcal{P}}(\mu)}$ such that $\int_q^1 F_{\nu_{\tilde{\mathcal{P}}(\mu)}}^{-1}(p) dp = \tilde{\psi}(q) + \int_q^1 F_\nu^{-1}(p) dp$. Moreover, $\nu_{\tilde{\mathcal{P}}(\mu)} \in \tilde{\mathcal{P}}(\mu)$. For $\eta \in \tilde{\mathcal{P}}(\mu)$, let $D(\eta)$ denote the distribution of $F_\eta^{-1}(U) - F_\nu^{-1}(U)$ for U uniformly distributed on $(0, 1)$. By Lemma A.4 below, the set $\{D(\eta) : \eta \in \tilde{\mathcal{P}}(\mu)\}$ admits an infimum π for the convex order and for all $q \in [0, 1]$, $\int_q^1 F_\pi^{-1}(p) dp =$

$\inf_{\eta \in \tilde{\mathcal{P}}(\mu)} \int_q^1 F_{D(\eta)}^{-1}(p) dp$. For $\eta \in \tilde{\mathcal{P}}(\mu)$, one has $F_{D(\eta)}^{-1} = F_\eta^{-1} - F_v^{-1}$ by Lemma A.3 below. With the fact that $\eta \in \tilde{\mathcal{P}}(\mu)$ if and only if $\int_q^1 F_\eta^{-1}(p) dp \geq \int_q^1 F_\mu^{-1}(p) dp$ for all $q \in (0, 1)$ with equality for $q = 0$ and $[0, 1] \ni q \mapsto \int_q^1 F_\eta^{-1}(p) - F_v^{-1}(p) dp$ is concave, one deduces that for $q \in (0, 1)$,

$$\int_q^1 F_\pi^{-1}(p) dp = \inf_{\eta \in \tilde{\mathcal{P}}(\mu)} \int_q^1 F_\eta^{-1}(p) - F_v^{-1}(p) dp = \tilde{\psi}(q) = \int_q^1 F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(p) - F_v^{-1}(p) dp.$$

Hence $\pi = D(v_{\tilde{\mathcal{P}}(\mu)})$. If $\mu, \nu \in \mathcal{P}_\varrho(\mathbb{R})$ for some $\varrho > 1$, then

$$\begin{aligned} W_\varrho^e(\nu, v_{\tilde{\mathcal{P}}(\mu)}) &= \mathbb{E}[|F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(U) - F_v^{-1}(U)|^\varrho] = \int_{\mathbb{R}} |x|^\varrho \pi(dx) \\ &\leq \inf_{\eta \in \tilde{\mathcal{P}}(\mu)} \mathbb{E}[|F_\eta^{-1}(U) - F_v^{-1}(U)|^\varrho] = \inf_{\eta \in \tilde{\mathcal{P}}(\mu)} W_\varrho^e(\nu, \eta). \end{aligned}$$

By Lemma A.6 below, $\inf_{\eta \in \tilde{\mathcal{P}}(\mu)} W_\varrho(\nu, \eta) = \inf_{\eta \in \tilde{\mathcal{P}}(\mu)} W_\varrho(\nu, \eta)$. Therefore $W_\varrho(\nu, v_{\tilde{\mathcal{P}}(\mu)}) = \inf_{\eta \in \tilde{\mathcal{P}}(\mu)} W_\varrho(\nu, \eta)$ and $v_{\tilde{\mathcal{P}}(\mu)}^\varrho = v_{\tilde{\mathcal{P}}(\mu)}$. □

For probability measures $\mu_I = \sum_{i=1}^I p_i \delta_{x_i}$ (resp. $\nu_J = \sum_{j=1}^J q_j \delta_{y_j}$) on the real line with $(p_1, \dots, p_I) \in (0, 1]^I$ and $x_1 < x_2 < \dots < x_I$ (resp. $(q_1, \dots, q_J) \in (0, 1]^J$ and $y_1 < y_2 < \dots < y_J$), $\tilde{\psi}$ is equal to $\int_0^1 F_{\mu_I}^{-1}(p) - F_{\nu_J}^{-1}(p) dp$ minus the convex hull ψ of $q \mapsto \int_0^q F_{\mu_I}^{-1}(p) - F_{\nu_J}^{-1}(p) dp$ which has already been discussed after Theorem 2.6 and can be computed by Andrew’s monotone chain algorithm. One then may compute the probability measure $(\nu_J)_{\tilde{\mathcal{P}}(\mu_I)}$ which writes $\sum_{k=1}^K r_k \delta_{z_k}$ with $K \leq I + J$, $z_1 \leq z_2 \leq \dots \leq z_K$ and $(r_k)_{1 \leq k \leq K}$ denoting the differences between the successive elements of the increasing reordering of $\{0\} \cup \{\sum_{k=1}^i p_k : 1 \leq i \leq I\} \cup \{\sum_{k=1}^j q_k : 1 \leq j \leq J\}$.

In contrast, when $d \geq 2$, the measure(s) $v_{\tilde{\mathcal{P}}(\mu)}^\varrho$ do(es) not seem easy to be calculated numerically, even for $\varrho = 2$. In fact, the constraint of the convex order is not simple to handle in a minimization program. More precisely, in the case of empirical measures, one would have to minimize $\sum_{j=1}^J |\tilde{Y}_j - Y_j|^2$ under the constraint $\frac{1}{J} \sum_{i=1}^I \delta_{X_i} \leq_{\text{cx}} \frac{1}{J} \sum_{j=1}^J \delta_{\tilde{Y}_j}$. Even in dimension 1, this constraint is not linear since it is equivalent to $\max_i X_i \leq \max_j \tilde{Y}_j$, $\min_i X_i \leq \min_j \tilde{Y}_j$, $\frac{1}{I} \sum_{i=1}^I X_i = \frac{1}{J} \sum_{j=1}^J \tilde{Y}_j$, and $\frac{1}{I} \sum_{i=1}^I (X_i - \tilde{Y}_{j'})^+ \leq \frac{1}{J} \sum_{j=1}^J (\tilde{Y}_j - \tilde{Y}_{j'})^+$ for any $1 \leq j' \leq J$, see e.g. Corollary 2.2 in [1]. Nethertheless, we can state error estimations similar to the ones obtained for $\mu_{\tilde{\mathcal{P}}(\nu)}^\varrho$ in Section 3.

Proposition 4.3. *Let $\varrho \geq 1$, $\mu, \nu, \mu_I, \nu_J \in \mathcal{P}_\varrho(\mathbb{R}^d)$ such that $\mu \leq_{\text{cx}} \nu$. Then, we have*

$$W_\varrho((\nu_J)_{\tilde{\mathcal{P}}(\mu_I)}^\varrho, \nu) \leq W_\varrho(\mu, \mu_I) + 2W_\varrho(\nu, \nu_J), \tag{4.1}$$

where by a slight abuse of notation, $(\nu_J)_{\tilde{\mathcal{P}}(\mu_I)}^\varrho$ denotes any minimizing probability measure when uniqueness is not shown.

Proof. We proceed like in the proof of Proposition 3.1. Let $Q_{\mu_I}^\varrho$ (resp. Q_ν^ϱ) be a Markov kernel such that $\mu_I(dx)Q_{\mu_I}^\varrho(x, dy)$ (resp. $\nu(dx)Q_\nu^\varrho(x, dy)$) is an optimal transport plan for $W_\varrho(\mu_I, \mu)$ (resp. $W_\varrho(\nu, \nu_J)$) and R be a martingale Markov kernel such that $\mu R = \nu$. We obviously have $\nu_J = \mu_I Q_{\mu_I}^\varrho R Q_\nu^\varrho$. By Jensen inequality and using the martingale property of R , we have $\mu_I \leq_{\text{cx}} ((x, w, z) \mapsto x + z - w) \# \mu_I(dx) Q_{\mu_I}^\varrho(x, dw) R(w, dz) Q_\nu^\varrho(z, dy)$, so that

$$\inf_{\eta \in \tilde{\mathcal{P}}(\mu_I)} W_\varrho(\nu_J, \eta) \leq \left(\int_{(\mathbb{R}^d)^4} |x + z - w - y|^\varrho \mu_I(dx) Q_{\mu_I}^\varrho(x, dw) R(w, dz) Q_\nu^\varrho(z, dy) \right)^{1/\varrho}.$$

We get (4.1) using Minkowski’s inequality and the triangle inequality $W_\varrho((\nu_J)_{\tilde{\mathcal{P}}(\mu_I)}^\varrho, \nu) \leq W_\varrho((\nu_J)_{\tilde{\mathcal{P}}(\mu_I)}^\varrho, \nu_J) + W_\varrho(\nu, \nu_J)$. □

In the multi-marginal case, defining inductively $\mu_{I_1}^{1,\varrho} = \mu_{I_1}^1$ and for $k \in \{2, \dots, \ell\}$, $\mu_{I_1, \dots, I_k}^{k,\varrho}$ as the W_ϱ projection of $\mu_{I_k}^k$ on $\tilde{\mathcal{P}}(\mu_{I_1, \dots, I_{k-1}}^{k-1,\varrho})$, we deduce that for $k \in \{2, \dots, \ell\}$,

$$W_\varrho(\mu^k, \mu_{I_1, \dots, I_k}^{k,\varrho}) \leq W_\varrho(\mu^1, \mu_{I_1}^1) + 2 \sum_{k'=2}^k W_\varrho(\mu^{k'}, \mu_{I_{k'}}^{k'}).$$

Comparing $W_\varrho(v_{\bar{\mathcal{P}}(\mu)}^\varrho, \nu)$ and $W_\varrho(\mu, \mu_{\underline{\mathcal{P}}(\nu)}^\varrho)$ leads to interesting properties.

Corollary 4.4. *For $\varrho > 1$, $\mu, \nu \in \mathcal{P}_\varrho(\mathbb{R}^d)$, we have*

$$W_\varrho(v_{\bar{\mathcal{P}}(\mu)}^\varrho, \nu) = W_\varrho(\mu, \mu_{\underline{\mathcal{P}}(\nu)}^\varrho)$$

and there is a measurable transport map $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that the only optimal transport plan between $v_{\bar{\mathcal{P}}(\mu)}^\varrho$ and ν is $v_{\bar{\mathcal{P}}(\mu)}^\varrho(dz)\delta_{T(z)}(dy)$. Moreover, for any martingale Markov kernel R such that $\mu R = v_{\bar{\mathcal{P}}(\mu)}^\varrho$, $\mu(dx)R(x, dz)$ a.e., $T(z) - z = \int_{\mathbb{R}^d} T(z)R(x, dz) - x$. Last, in dimension $d = 1$, when $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$, we also have for all $\varrho \geq 1$, $W_\varrho(v_{\bar{\mathcal{P}}(\mu)}^\varrho, \nu) = W_\varrho(\mu, \mu_{\underline{\mathcal{P}}(\nu)}^\varrho) = (\int_0^1 |\psi'(u-)|^\varrho du)^{1/\varrho}$ where $\psi'(u-)$ is the left-hand derivative of the convex hull ψ of the function $[0, 1] \ni q \mapsto \int_0^q F_\mu^{-1}(p) - F_\nu^{-1}(p) dp$.

Proof. Since $\mu \leq_{\text{cx}} v_{\bar{\mathcal{P}}(\mu)}^\varrho$, we may replace (μ, μ_I, ν, ν_J) by $(\mu, \mu, v_{\bar{\mathcal{P}}(\mu)}^\varrho, \nu)$ in Proposition 3.1 to get $W_\varrho(\mu, \mu_{\underline{\mathcal{P}}(\nu)}^\varrho) \leq W_\varrho(v_{\bar{\mathcal{P}}(\mu)}^\varrho, \nu)$. Using that $\mu_{\underline{\mathcal{P}}(\nu)}^\varrho \leq_{\text{cx}} \nu$ to replace (μ, μ_I, ν, ν_J) by $(\mu_{\underline{\mathcal{P}}(\nu)}^\varrho, \mu, \nu, \nu)$ in Proposition 4.3, we obtain the converse inequality.

Now, let $\varrho > 1$, R denote a martingale Markov kernel such that $\mu R = v_{\bar{\mathcal{P}}(\mu)}^\varrho$ and Q a Markov kernel such that $v_{\bar{\mathcal{P}}(\mu)}^\varrho(dz)Q(z, dy)$ is an optimal transport plan for $W_\varrho(v_{\bar{\mathcal{P}}(\mu)}^\varrho, \nu)$. Repeating the arguments of Proposition 3.1 (replacing again (μ, μ_I, ν, ν_J) by $(\mu, \mu, v_{\bar{\mathcal{P}}(\mu)}^\varrho, \nu)$), we get

$$\begin{aligned} W_\varrho^\varrho(\mu, \mu_{\underline{\mathcal{P}}(\nu)}^\varrho) &\leq \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d \times \mathbb{R}^d} (x - y)R(x, dz)Q(z, dy) \right|^\varrho \mu(dx) \\ &= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d \times \mathbb{R}^d} (z - y)R(x, dz)Q(z, dy) \right|^\varrho \mu(dx) \leq \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (z - y)Q(z, dy) \right|^\varrho v_{\bar{\mathcal{P}}(\mu)}^\varrho(dz) \\ &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |z - y|^\varrho v_{\bar{\mathcal{P}}(\mu)}^\varrho(dz)Q(z, dy) = W_\varrho^\varrho(v_{\bar{\mathcal{P}}(\mu)}^\varrho, \nu) = W_\varrho^\varrho(\mu, \mu_{\underline{\mathcal{P}}(\nu)}^\varrho). \end{aligned}$$

The equality in the last inequality ensures that $v_{\bar{\mathcal{P}}(\mu)}^\varrho(dz)$ a.e., $Q(z, dy) = \delta_{T(z)}(dy)$ where $T(z) = \int_{\mathbb{R}^d} yQ(z, dy)$. Moreover, the equality in the second inequality implies that $\mu(dx)R(x, dz)$ a.e., $T(z) - z = \int_{\mathbb{R}^d} T(z)R(x, dz) - x$.

If \tilde{Q} is another Markov kernel such that $v_{\bar{\mathcal{P}}(\mu)}^\varrho(dz)\tilde{Q}(z, dy)$ is an optimal transport plan for $W_\varrho(v_{\bar{\mathcal{P}}(\mu)}^\varrho, \nu)$, then $v_{\bar{\mathcal{P}}(\mu)}^\varrho(dz)\frac{\tilde{Q}+Q}{2}(z, dy)$ is also an optimal transport plan and $v_{\bar{\mathcal{P}}(\mu)}^\varrho(dz)$ a.e., $\frac{\tilde{Q}+Q}{2}(z, dy)$ is a Dirac mass so that $\tilde{Q}(z, dy) = Q(z, dy)$.

In dimension 1, we observe that

$$\begin{aligned} \forall q \in [0, 1], \quad \int_0^q F_{\mu_{\underline{\mathcal{P}}(\nu)}^\varrho}^{-1}(p) dp &= \int_0^q F_\mu^{-1}(p) dp - \psi(q) \\ \text{and} \quad \int_0^q F_{v_{\bar{\mathcal{P}}(\mu)}^\varrho}^{-1}(p) dp &= \int_0^q F_\nu^{-1}(p) dp + \psi(q). \end{aligned} \tag{4.2}$$

Thus, we have $F_{\mu_{\underline{\mathcal{P}}(\nu)}^\varrho}^{-1}(p) - F_\mu^{-1}(p) = -\psi'(p-)$ and $F_{v_{\bar{\mathcal{P}}(\mu)}^\varrho}^{-1}(p) - F_\nu^{-1}(p) = \psi'(p-)$ for $p \in (0, 1)$, which gives the claim. \square

The property $T(z) - z = \int_{\mathbb{R}^d} T(z)R(x, dz) - x$, $\mu(dx)R(x, dz)$ a.e., in Corollary 4.4 indicates that in dimension 1, an optimal transport map T between $v_{\bar{\mathcal{P}}(\mu)}^\varrho$ and ν should be piecewise affine with slope 1 on the irreducible components of $(\mu, v_{\bar{\mathcal{P}}(\mu)}^\varrho)$ introduced in Theorem A.4 [8], provided that we can find a martingale Markov kernel R that spans the whole components. The irreducible components are the countable disjoint open intervals, the union of which is equal to the open set $\{t \in \mathbb{R}, \int_{\mathbb{R}} (t - x)^+ \mu(dx) < \int_{\mathbb{R}} (t - x)^+ v_{\bar{\mathcal{P}}(\mu)}^\varrho(dx)\}$. This is indeed the case according to the following proposition which moreover exhibits a common optimal transport map for $W_\varrho(v_{\bar{\mathcal{P}}(\mu)}^\varrho, \nu)$ and $W_\varrho(\mu, \mu_{\underline{\mathcal{P}}(\nu)}^\varrho)$.

Proposition 4.5. *Let $\varrho > 1$, $\mu, \nu \in \mathcal{P}_\varrho(\mathbb{R})$. Let $(\underline{t}_n, \bar{t}_n)$, $1 \leq n \leq N$, (resp. $(\underline{t}'_n, \bar{t}'_n)$, $1 \leq n \leq N'$) be the irreducible components of $(\mu, v_{\bar{\mathcal{P}}(\mu)}^\varrho)$ (resp. $(\mu_{\underline{\mathcal{P}}(\nu)}^\varrho, \nu)$). Then, we have $N = N'$ and $F_\mu(\underline{t}_n) = F_{\mu_{\underline{\mathcal{P}}(\nu)}^\varrho}(\underline{t}'_n)$, $F_\mu(\bar{t}_n-) = F_{\mu_{\underline{\mathcal{P}}(\nu)}^\varrho}(\bar{t}'_n-)$ up to a renumbering of $(\underline{t}'_n)_{1 \leq n \leq N}$.*

Let ψ be the convex hull of the function $[0, 1] \ni q \mapsto \int_0^q F_\mu^{-1}(p) - F_\nu^{-1}(p) dp$. Then, the function $T : \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$\forall x \notin \bigcup_{1 \leq n \leq N} (\underline{t}_n, \bar{t}_n), \quad T(x) = F_\nu^{-1}(F_{\nu_{\bar{\mathcal{P}}(\mu)}}(x)) \quad \text{and}$$

$$\forall 1 \leq n \leq N, \forall x \in (\underline{t}_n, \bar{t}_n), \quad T(x) = x - \frac{\psi(F_\mu(\bar{t}_n-)) - \psi(F_\mu(\underline{t}_n))}{F_\mu(\bar{t}_n-) - F_\mu(\underline{t}_n)}$$

is an optimal transport map for $W_\varrho(\nu_{\bar{\mathcal{P}}(\mu)}, \nu)$ and $W_\varrho(\mu, \mu_{\underline{\mathcal{P}}(\nu)})$.

Proof. We set $\underline{q}_n = F_\mu(\underline{t}_n)$ and $\bar{q}_n = F_\mu(\bar{t}_n-)$. From (4.2) and Lemma A.8 below which characterizes the irreducible components in terms of the quantile functions, we get

$$\begin{aligned} \bigcup_{1 \leq n \leq N} (\underline{q}_n, \bar{q}_n) &= \left\{ q \in [0, 1], \int_0^q F_\mu^{-1}(p) dp > \int_0^q F_{\nu_{\bar{\mathcal{P}}(\mu)}}^{-1}(p) dp \right\} \\ &= \left\{ q \in [0, 1], \psi(q) < \int_0^q F_\mu^{-1}(p) dp - \int_0^q F_\nu^{-1}(p) dp \right\} \\ &= \left\{ q \in [0, 1], \int_0^q F_{\mu_{\underline{\mathcal{P}}(\nu)}}^{-1}(p) dp > \int_0^q F_\nu^{-1}(p) dp \right\} \\ &= \bigcup_{1 \leq n \leq N'} (F_{\mu_{\underline{\mathcal{P}}(\nu)}}(\underline{t}'_n), F_{\mu_{\underline{\mathcal{P}}(\nu)}}(\bar{t}'_n-)), \end{aligned}$$

which gives the first claim. From the second equality and since ψ is the convex hull of $[0, 1] \ni q \mapsto \int_0^q F_\mu^{-1}(p) - F_\nu^{-1}(p) dp$, we get

$$\forall q \notin \bigcup_{1 \leq n \leq N} (\underline{q}_n, \bar{q}_n), \quad \psi(q) = \int_0^q F_\mu^{-1}(p) dp - \int_0^q F_\nu^{-1}(p) dp \tag{4.3}$$

and $\psi(q) = \psi(\underline{q}_n) + \frac{\psi(\bar{q}_n) - \psi(\underline{q}_n)}{\bar{q}_n - \underline{q}_n}(q - \underline{q}_n)$ for $q \in [\underline{q}_n, \bar{q}_n]$. From (4.2), this gives

$$\forall q \in (\underline{q}_n, \bar{q}_n], \quad F_{\nu_{\bar{\mathcal{P}}(\mu)}}^{-1}(q) = F_\nu^{-1}(q) + \frac{\psi(\bar{q}_n) - \psi(\underline{q}_n)}{\bar{q}_n - \underline{q}_n}$$

$$\text{and} \quad F_{\mu_{\underline{\mathcal{P}}(\nu)}}^{-1}(q) = F_\mu^{-1}(q) - \frac{\psi(\bar{q}_n) - \psi(\underline{q}_n)}{\bar{q}_n - \underline{q}_n}. \tag{4.4}$$

Any point q in $(0, 1) \setminus \bigcup_{1 \leq n \leq N} (\underline{q}_n, \bar{q}_n]$ is the limit of an increasing sequence $(q_k)_{k \geq 1}$ of points in $(0, 1) \setminus \bigcup_{1 \leq n \leq N} (\underline{q}_n, \bar{q}_n)$. Since, by (4.2) and (4.3), $\frac{1}{q-q_k} \int_{q_k}^q F_{\nu_{\bar{\mathcal{P}}(\mu)}}^{-1}(p) dp = \frac{1}{q-q_k} \int_{q_k}^q F_\mu^{-1}(p) dp$ and $\frac{1}{q-q_k} \int_{q_k}^q F_{\mu_{\underline{\mathcal{P}}(\nu)}}^{-1}(p) dp = \frac{1}{q-q_k} \int_{q_k}^q F_\nu^{-1}(p) dp$, the left-continuity of the quantile functions implies that $F_{\nu_{\bar{\mathcal{P}}(\mu)}}^{-1}(q) = F_\mu^{-1}(q)$ and $F_{\mu_{\underline{\mathcal{P}}(\nu)}}^{-1}(q) = F_\nu^{-1}(q)$. We deduce that

$$\forall q \in (0, 1) \setminus \bigcup_{1 \leq n \leq N} (\underline{q}_n, \bar{q}_n], \quad F_{\nu_{\bar{\mathcal{P}}(\mu)}}^{-1}(q) = F_\mu^{-1}(q) \quad \text{and} \quad F_{\mu_{\underline{\mathcal{P}}(\nu)}}^{-1}(q) = F_\nu^{-1}(q). \tag{4.5}$$

By Corollary 4.4, there exists an optimal transport map \tilde{T} between $\nu_{\bar{\mathcal{P}}(\mu)}$ and ν . By Proposition 2.17 in [25], we have dq -a.e. $\tilde{T}(F_{\nu_{\bar{\mathcal{P}}(\mu)}}^{-1}(q)) = F_\nu^{-1}(q)$. For $x \in \mathbb{R}$ such that $F_{\nu_{\bar{\mathcal{P}}(\mu)}}(x-) < F_{\nu_{\bar{\mathcal{P}}(\mu)}}(x)$, since $F_{\nu_{\bar{\mathcal{P}}(\mu)}}^{-1}$ is constant (equal to x) on $(F_{\nu_{\bar{\mathcal{P}}(\mu)}}(x-), F_{\nu_{\bar{\mathcal{P}}(\mu)}}(x)]$, we deduce that the left-continuous function F_ν^{-1} is also constant on this interval. Let now $x \in \mathbb{R} \cap \bigcup_{1 \leq n \leq N} \{\underline{t}_n, \bar{t}_n\}$. By definition of the irreducible components, we have

$$F_{\nu_{\bar{\mathcal{P}}(\mu)}}(x-) \leq F_\mu(x-) \leq F_\mu(x) \leq F_{\nu_{\bar{\mathcal{P}}(\mu)}}(x). \tag{4.6}$$

If $F_{v_{\tilde{\mathcal{P}}(\mu)}}(x-) < F_{v_{\tilde{\mathcal{P}}(\mu)}}(x)$, $(F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}, F_v^{-1})$ is constant and equal to $(x, F_v^{-1}(F_{v_{\tilde{\mathcal{P}}(\mu)}}(x)))$ on the interval $(F_{v_{\tilde{\mathcal{P}}(\mu)}}(x-), F_{v_{\tilde{\mathcal{P}}(\mu)}}(x))$ and, by definition of T , $T(F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1})$ and F_v^{-1} are equal on this interval.

We are now going to prove that dq a.e., $T(F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q)) = F_v^{-1}(q)$, which, by Proposition 2.17 in [25], ensures that T is an optimal transport map between $v_{\tilde{\mathcal{P}}(\mu)}$ and v .

If $q \in (F_{v_{\tilde{\mathcal{P}}(\mu)}}(\underline{t}_n), F_{v_{\tilde{\mathcal{P}}(\mu)}}(\bar{t}_n-))$ then $F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q) \in (\underline{t}_n, \bar{t}_n)$ and $T(F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q)) = F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q) - \frac{\psi(\bar{q}_n) - \psi(\underline{q}_n)}{\bar{q}_n - \underline{q}_n}$ with the right-hand side equal to $F_v^{-1}(q)$ by (4.4) since, by (4.6),

$$(F_{v_{\tilde{\mathcal{P}}(\mu)}}(\underline{t}_n), F_{v_{\tilde{\mathcal{P}}(\mu)}}(\bar{t}_n-)) \subset (\underline{q}_n, \bar{q}_n).$$

By the above reasoning for $x \in \mathbb{R} \cap \bigcup_{1 \leq n \leq N} \{\underline{t}_n, \bar{t}_n\}$, the equality between $T(F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1})$ and F_v^{-1} still holds on $(F_{v_{\tilde{\mathcal{P}}(\mu)}}(\underline{t}_n-), F_{v_{\tilde{\mathcal{P}}(\mu)}}(\bar{t}_n-)) \cup (F_{v_{\tilde{\mathcal{P}}(\mu)}}(\bar{t}_n-), F_{v_{\tilde{\mathcal{P}}(\mu)}}(\bar{t}_n))$.

If $q \notin (F_{v_{\tilde{\mathcal{P}}(\mu)}}(\underline{t}_n-), F_{v_{\tilde{\mathcal{P}}(\mu)}}(\bar{t}_n))$, then $F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q) \leq \underline{t}_n$ or $F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q) > \bar{t}_n$. We deduce that for $q \notin \bigcup_{1 \leq n \leq N} (F_{v_{\tilde{\mathcal{P}}(\mu)}}(\underline{t}_n-), F_{v_{\tilde{\mathcal{P}}(\mu)}}(\bar{t}_n))$, $F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q) \notin \bigcup_{1 \leq n \leq N} (\underline{t}_n, \bar{t}_n)$ and $T(F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q)) = F_v^{-1}(F_{v_{\tilde{\mathcal{P}}(\mu)}}(F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q)))$. The right-hand side is equal to $F_v^{-1}(q)$ when $F_{v_{\tilde{\mathcal{P}}(\mu)}}(F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q)-) = F_{v_{\tilde{\mathcal{P}}(\mu)}}(F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q))$ since then $F_{v_{\tilde{\mathcal{P}}(\mu)}}(F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q)) = q$ and otherwise when $q > F_{v_{\tilde{\mathcal{P}}(\mu)}}(F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q)-)$ since, then, the interval $(F_{v_{\tilde{\mathcal{P}}(\mu)}}(F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q)-), F_{v_{\tilde{\mathcal{P}}(\mu)}}(F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q)))$ on which F_v^{-1} is constant contains q .

In conclusion $T(F_{v_{\tilde{\mathcal{P}}(\mu)}}^{-1}(q)) = F_v^{-1}(q)$ for q outside the at most countable set $\{F_{v_{\tilde{\mathcal{P}}(\mu)}}(\bar{t}_n-): 1 \leq n \leq N\} \cup \{F_{v_{\tilde{\mathcal{P}}(\mu)}}(x-): x \in \mathbb{R} \text{ s.t. } F_{v_{\tilde{\mathcal{P}}(\mu)}}(x-) < F_{v_{\tilde{\mathcal{P}}(\mu)}}(x)\}$ and therefore dq a.e.

With (4.5), we deduce that dq a.e. on $(0, 1) \setminus \bigcup_{1 \leq n \leq N} (\underline{q}_n, \bar{q}_n]$, $T(F_{\mu}^{-1}(q)) = F_{\mu_{\mathcal{P}(v)}}^{-1}(q)$. If $q \in (\underline{q}_n, \bar{q}_n)$ for some $1 \leq n \leq N$, then $F_{\mu}^{-1}(q) \in (\underline{t}_n, \bar{t}_n)$ and, by definition of T and (4.4), $T(F_{\mu}^{-1}(q)) = F_{\mu}^{-1}(q) - \frac{\psi(\bar{q}_n) - \psi(\underline{q}_n)}{\bar{q}_n - \underline{q}_n} = F_{\mu_{\mathcal{P}(v)}}^{-1}(q)$. Hence dq a.e. $T(F_{\mu}^{-1}(q)) = F_{\mu_{\mathcal{P}(v)}}^{-1}(q)$ and T is an optimal transport map between μ and $\mu_{\mathcal{P}(v)}$. \square

5. Numerical experiments

5.1. Wasserstein distance

We start by illustrating numerically the convergences obtained in Corollary 3.2, and deduced from Proposition 4.3. We present on an example the convergence of the Wasserstein projection $(\mu_I)_{\mathcal{P}(v_I)}^e$ (resp. $(v_I)_{\tilde{\mathcal{P}}(\mu_I)}^e$) toward μ (resp. v) for the Wasserstein distance when μ_I and v_I are the respective empirical measures of μ and v with $\mu \leq_{\text{cx}} v$. To do so we consider an example in dimension one, so that the projections can be calculated explicitly according to Theorem 2.6 and Proposition 4.2. We take $\mu = \mathcal{N}(0, 1)$ and $v = \mathcal{N}(0, 1.1)$. For $I \geq 1$, we consider independent samples X_1, \dots, X_I and Y_1, \dots, Y_I distributed respectively according to μ and v . Then, we set $\mu_I = \frac{1}{I} \sum_{i=1}^I \delta_{X_i}$, $v_I = \frac{1}{I} \sum_{i=1}^I \delta_{Y_i}$, $\bar{X}_I = \frac{1}{I} \sum_{i=1}^I X_i$, $\bar{Y}_I = \frac{1}{I} \sum_{i=1}^I Y_i$, $\tilde{\mu}_I = \frac{1}{I} \sum_{i=1}^I \delta_{X_i - \bar{X}_I}$ and $\tilde{v}_I = \frac{1}{I} \sum_{i=1}^I \delta_{Y_i - \bar{Y}_I}$. Notice that, to define $\tilde{\mu}_I$ and \tilde{v}_I , we took advantage of the knowledge of the common mean of μ and v . This situation is usual in financial applications: discounted asset prices are martingales and their means are given by the present values. Equalizing the means can be seen as a first step towards the restoration of the convex order and we have noticed in our numerical experiments that it improves the approximation of the limiting problem (see for instance Figure 1). We calculate the Wasserstein projections $(\mu_I)_{\mathcal{P}(v_I)}$ and $(\tilde{\mu}_I)_{\tilde{\mathcal{P}}(\tilde{v}_I)}$ (resp. $(v_I)_{\tilde{\mathcal{P}}(\mu_I)}$ and $(\tilde{v}_I)_{\tilde{\mathcal{P}}(\tilde{\mu}_I)}$) and the 2-Wasserstein distance between each of these measures and μ (resp. v), as explained below.

As a comparison to these projections, we consider the respective approximations of μ and v by $\mu_I \wedge v_I$ and $\mu_I \vee v_I$, where $\mu_I \wedge v_I$ and $\mu_I \vee v_I$ are respectively defined as the infimum and the supremum of μ_I and v_I for the decreasing convex order when $\frac{1}{I} \sum_{i=1}^I X_i \leq \frac{1}{I} \sum_{i=1}^I Y_i$ and for the increasing convex order otherwise so that $\mu_I \wedge v_I \in \mathcal{P}(v_I)$ and $\mu_I \vee v_I \in \tilde{\mathcal{P}}(\mu_I)$. We also consider the approximations by $\tilde{\mu}_I \wedge \tilde{v}_I$ and $\tilde{\mu}_I \vee \tilde{v}_I$. These approximations can be calculated explicitly for probability measures with finite support (see [2] or [1]) and are natural alternatives to the Wasserstein projections in dimension 1.

The graph at left (resp. right) of Figure 1 illustrates the convergence of $W_2(\mu, \mu_I \wedge v_I)$, $W_2(\mu, (\mu_I)_{\mathcal{P}(v_I)})$, $W_2(\mu_I \vee v_I, v)$ and $W_2((v_I)_{\tilde{\mathcal{P}}(\mu_I)}, v)$ (resp. $W_2(\mu, \tilde{\mu}_I \wedge \tilde{v}_I)$, $W_2(\mu, (\tilde{\mu}_I)_{\tilde{\mathcal{P}}(\tilde{v}_I)})$, $W_2(\tilde{\mu}_I \vee \tilde{v}_I, v)$ and $W_2((\tilde{v}_I)_{\tilde{\mathcal{P}}(\tilde{\mu}_I)}, v)$) toward zero as $I \rightarrow \infty$. The corresponding curves are respectively in red, blue, green and magenta. The star (resp. cross) points indicate the upper bound for $W_2(\mu, (\mu_I)_{\mathcal{P}(v_I)})$ (left) and $W_2(\mu, (\tilde{\mu}_I)_{\tilde{\mathcal{P}}(\tilde{v}_I)})$ (right) (resp. $W_2((v_I)_{\tilde{\mathcal{P}}(\mu_I)}, v)$ (left) and $W_2((\tilde{v}_I)_{\tilde{\mathcal{P}}(\tilde{\mu}_I)}, v)$ (right)) given by Proposition 3.1 (resp. Proposition 4.3). As expected, the curves in blue and magenta

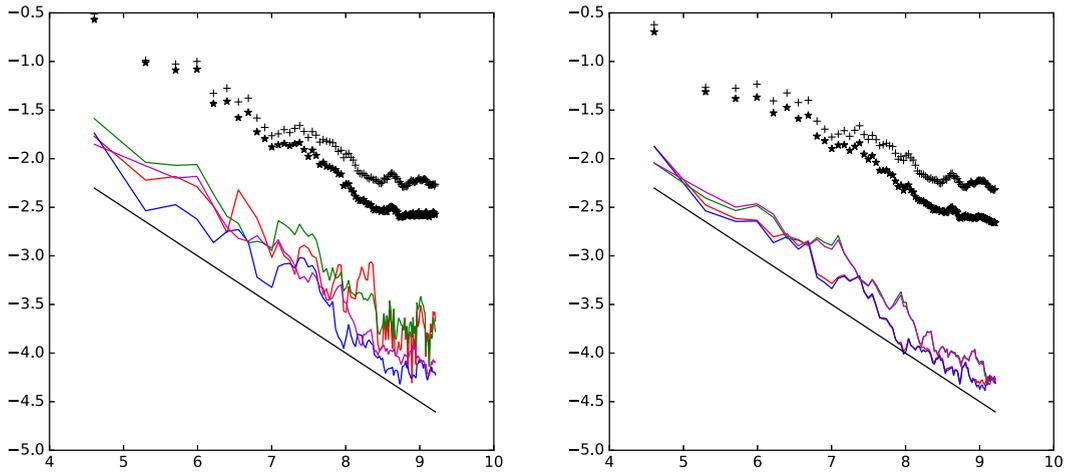


Fig. 1. Plot of the logarithms of the Wasserstein distances in function of $\log(I)$.

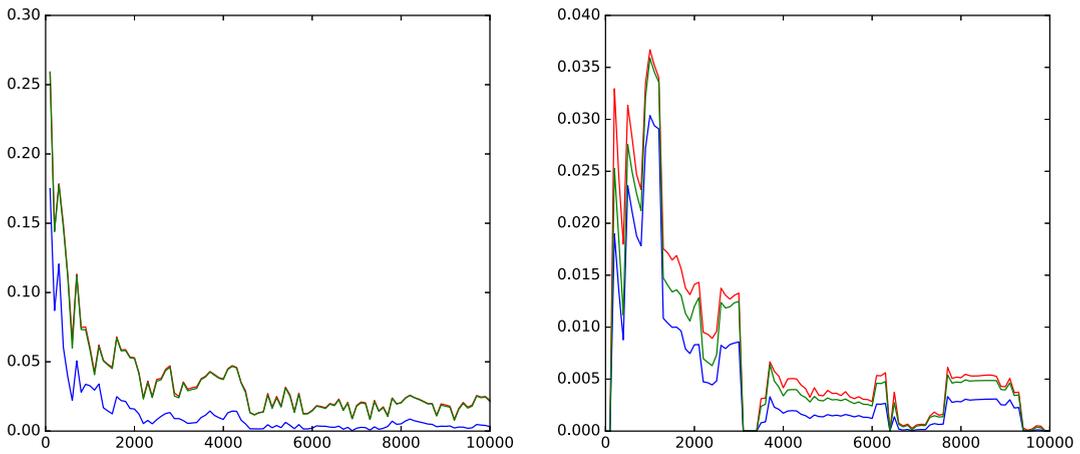


Fig. 2. Plot of the Wasserstein distances $W_2(\mu_I, \mu_I \wedge \nu_I)$, $W_2(\mu_I, (\mu_I)_{\mathcal{P}(\nu_I)})$, $W_2(\nu_I, \mu_I \vee \nu_I)$ (left) and $W_2(\tilde{\mu}_I, \tilde{\mu}_I \wedge \tilde{\nu}_I)$, $W_2(\tilde{\mu}_I, (\tilde{\mu}_I)_{\mathcal{P}(\tilde{\nu}_I)})$, $W_2(\tilde{\nu}_I, \tilde{\mu}_I \vee \tilde{\nu}_I)$ (right) in function of I .

are below these points. Let us mention that all these Wasserstein distances are calculated exactly by using the quantile function \mathcal{N}^{-1} of the standard normal variable. For instance, if $\eta = \sum_{i=1}^I p_i \delta_{Z_i}$ with $Z_1 \leq Z_2 \leq \dots \leq Z_I$, $P_0 = 0$ and $P_i = P_{i-1} + p_i$ for $1 \leq i \leq I$,

$$\begin{aligned} W_2^2(\mu, \eta) &= \int_{\mathbb{R}} x^2(\mu(dx) + \eta(dx)) - 2 \sum_{i=1}^I Z_i \int_{P_{i-1}}^{P_i} \mathcal{N}^{-1}(p) dp \\ &= 1 + \sum_{i=1}^I p_i Z_i^2 + \frac{\sqrt{2}}{\sqrt{\pi}} \sum_{i=1}^I Z_i (e^{-(\mathcal{N}^{-1}(P_i))^2/2} - e^{-(\mathcal{N}^{-1}(P_{i-1}))^2/2}). \end{aligned}$$

Asymptotically, the measure $(\mu_I)_{\mathcal{P}(\nu_I)}$ (resp. $(\nu_I)_{\mathcal{P}(\mu_I)}$) seems to slightly better approximate μ (resp. ν) than $\mu_I \wedge \nu_I$ (resp. $\mu_I \vee \nu_I$). Nonetheless, all these measures seem to converge for the Wasserstein distance at a rate close to $O(I^{-1/2})$ as indicated by the line in black with equation $y = -x/2$. This rate is better than the theoretical one stated in Corollary 3.2. In the right figure, we first observe that equalizing the means improves the approximations and reduces the Wasserstein distances (see the distances to the black lines). However, the rate of convergence is still roughly in $O(I^{-1/2})$. We also observe that there are only very small differences between using $\tilde{\mu}_I \wedge \tilde{\nu}_I$ or $(\tilde{\mu}_I)_{\mathcal{P}(\tilde{\nu}_I)}$ (resp. $\tilde{\mu}_I \vee \tilde{\nu}_I$ or $(\tilde{\nu}_I)_{\mathcal{P}(\tilde{\mu}_I)}$).

In Figure 2 are plotted at left (resp. right) the values of $W_2(\mu_I, \mu_I \wedge \nu_I)$, $W_2(\mu_I, (\mu_I)_{\mathcal{P}(\nu_I)}) = W_2(\nu_I, (\nu_I)_{\mathcal{P}(\mu_I)})$, $W_2(\nu_I, \mu_I \vee \nu_I)$ (resp. $W_2(\tilde{\mu}_I, \tilde{\mu}_I \wedge \tilde{\nu}_I)$, $W_2(\tilde{\mu}_I, (\tilde{\mu}_I)_{\mathcal{P}(\tilde{\nu}_I)}) = W_2(\tilde{\nu}_I, (\tilde{\nu}_I)_{\mathcal{P}(\tilde{\mu}_I)})$, $W_2(\tilde{\nu}_I, \tilde{\mu}_I \vee \tilde{\nu}_I)$) in function of I . The corresponding curves are in red, blue and green. We observe that the values of $W_2(\mu_I, \mu_I \wedge \nu_I)$ and $W_2(\nu_I, \mu_I \vee \nu_I)$ are very close. As expected, the blue curve is below the two other ones. At right, we observe that all the Wasserstein distances

Table 1

Comparison of the numerical minimizer of (1.1) for $\varrho = 2$ with the explicit solution $(\mu_I)_{\mathcal{P}(\tilde{\nu}_I)}$

	I				
	10	50	100	200	300
W ₂ -Wasserstein distance	4.4×10^{-5}	1.4×10^{-6}	4.5×10^{-6}	4.1×10^{-7}	4.2×10^{-7}

are equal to 0 on our sample for $I \approx 3200$, but take again positive values for larger values of I . This shows that the value of I from which we have $\tilde{\mu}_I \leq_{cx} \tilde{\nu}_I$, if it exists, depends on the sample and may be large.

Now, we conclude this section by checking the accuracy of the solver COIN-OR¹ for the quadratic optimization problem (1.1) with $\varrho = 2$. In fact, in dimension 1, we know that $(\mu_I)_{\mathcal{P}(\tilde{\nu}_I)}$ can be calculated explicitly as described below Theorem 2.6. In Table 1, we calculate the Wasserstein distance between $(\mu_I)_{\mathcal{P}(\tilde{\nu}_I)}$ and the measure obtained by solving numerically (1.1) with COIN-OR for different sample sizes I . As expected, the difference is very small. This validates numerically our theoretical results. More importantly, this indicates that the solver is reliable for finding the optimal solution with the values of I that we have considered in this paper.

5.2. MOT problems in dimension 2 with two marginal laws

5.2.1. An explicit example

Let μ and ν be respectively the uniform distributions on $[-1, 1]^2$ and $[-2, 2]^2$. For $x = (x^1, x^2) \in \mathbb{R}^2$ and $y = (y^1, y^2) \in \mathbb{R}^2$, we consider the minimization of the cost function $c(x, y) = |x^1 - y^1|^\varrho + |x^2 - y^2|^\varrho$, with $\varrho > 2$. For any $\pi \in \Pi^M(\mu, \nu)$, we have $\int_{\mathbb{R}^2 \times \mathbb{R}^2} \|y - x\|_2^2 \pi(dx, dy) = \int_{\mathbb{R}^2} \|y\|_2^2 \nu(dy) - \int_{\mathbb{R}^2} \|x\|_2^2 \mu(dx) = 2$. Jensen’s inequality gives

$$\begin{aligned} & \int_{\mathbb{R}^2 \times \mathbb{R}^2} |x^1 - y^1|^\varrho + |x^2 - y^2|^\varrho \pi(dx, dy) \\ & \geq \left(\int_{\mathbb{R}^2 \times \mathbb{R}^2} |x^1 - y^1|^2 \pi(dx, dy) \right)^{\frac{\varrho}{2}} + \left(\int_{\mathbb{R}^2 \times \mathbb{R}^2} |x^2 - y^2|^2 \pi(dx, dy) \right)^{\frac{\varrho}{2}} = 2. \end{aligned}$$

The equality condition in Jensen’s equality gives that $|x^1 - y^1| = |x^2 - y^2| = 1$, $\pi(dx, dy)$ -almost surely. Now, let us consider $X = (X^1, X^2)$ be distributed according to μ and $Z = (Z^1, Z^2)$ a couple of independent Rademacher random variables which is independent of X . Then $Y = X + Z$ is distributed according to ν and satisfies $|Y^1 - X^1| = |Y^2 - X^2| = 1$. The probability distribution π^* of (X, Y) is the unique martingale optimal coupling that minimizes $\int_{\mathbb{R}^2 \times \mathbb{R}^2} c(x, y) \pi(dx, dy)$. Indeed, if (\tilde{X}, \tilde{Y}) is distributed according to an optimal coupling, then $\tilde{Y}^1 - \tilde{X}^1$ and $\tilde{Y}^2 - \tilde{X}^2$ follow the Rademacher distribution, and both these random variables are necessarily independent of \tilde{X} in order to satisfy the martingale property. Last, $\tilde{Y}^1 - \tilde{X}^1$ and $\tilde{Y}^2 - \tilde{X}^2$ are necessarily independent, otherwise \tilde{Y} would not follow ν .

We now illustrate the MOT and consider independent samples $(X_1^1, X_1^2), \dots, (X_I^1, X_I^2)$ and $(Y_1^1, Y_1^2), \dots, (Y_I^1, Y_I^2)$ respectively distributed according to μ and ν . We set $\tilde{\mu}_I = \frac{1}{I} \sum_{i=1}^I \delta_{(X_i^1 - \bar{X}_I^1, X_i^2 - \bar{X}_I^2)}$ and $\tilde{\nu}_I = \frac{1}{I} \sum_{i=1}^I \delta_{(Y_i^1 - \bar{Y}_I^1, Y_i^2 - \bar{Y}_I^2)}$, with $\bar{X}_I^\ell = \frac{1}{I} \sum_{i=1}^I X_i^\ell$ and $\bar{Y}_I^\ell = \frac{1}{I} \sum_{i=1}^I Y_i^\ell$. We work with $\tilde{\mu}_I$ and $\tilde{\nu}_I$ rather than with the empirical measures μ_I and ν_I since we have noticed on our experiments that they better approximate μ and ν (see Figure 1) and give better results for the approximation of MOT problems (see [2]). Let us mention here that in financial applications, it is generally possible to calculate $\tilde{\mu}_I$ and $\tilde{\nu}_I$ from the empirical measures μ_I and ν_I since the mean of μ and ν is given by the current price of the underlying assets. To calculate $(\tilde{\mu}_I)_{\mathcal{P}(\tilde{\nu}_I)}$, we have to solve the quadratic optimization problem with linear constraints described in equation (1.1) for $\varrho = 2$. The dimension of the problem is thus equal to I^2 . We have used the COIN-OR solver in our numerical experiments, which enables us to solve (1.1) for I up to 500. Once $(\tilde{\mu}_I)_{\mathcal{P}(\tilde{\nu}_I)} = \frac{1}{I} \sum_{i=1}^I \delta_{(\tilde{x}_i^1, \tilde{x}_i^2)*}$ is calculated, we can then solve the discrete MOT problem between $(\tilde{\mu}_I)_{\mathcal{P}(\tilde{\nu}_I)}$ and $\tilde{\nu}_I$.

In Figure 3, for $\varrho = 2.5$, we have plotted $y_i^2 - x_i^2$ in function of $y_i^1 - x_i^1$ for the points (x_i, y_i) with positive probability in the MOT for $((\tilde{\mu}_I)_{\mathcal{P}(\tilde{\nu}_I)}, \nu_I)$. We recall that the optimal coupling for the continuous MOT is given by (X, Y) with $X \sim \mu$ and $Y = X + Z$, Z being a couple of independent Rademacher random variables. Since $Y_2 - Y_1 = X_2 - X_1 + Z_2 - Z_1$ and $Z_2 - Z_1$ takes values in $\{-2, 0, 2\}$, we expect to observe that the points are gathered around the lines $y = x - 2$, $y = x$ and $y = x + 2$, which is the case on Figure 3. This checks our implementation of the algorithm. Besides, we have

¹<https://www.coin-or.org/>

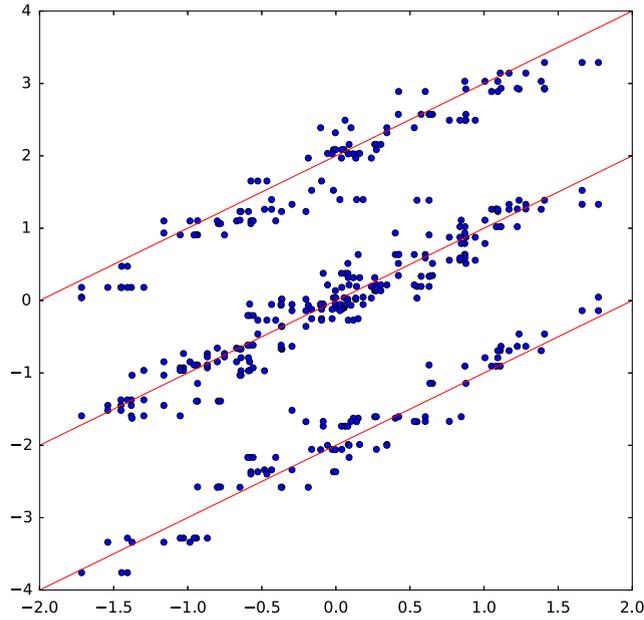


Fig. 3. Plot of $y_i^2 - y_i^1$ in function of $x_i^2 - x_i^1$ for the points (x_i, y_i) with positive probability in the MOT for $((\tilde{\mu}_I)^2_{\mathcal{P}(\tilde{\nu}_I)}, \nu_I)$, with $I = 100$. In red are drawn the lines $y = x - 2$, $y = x$ and $y = x + 2$.

calculated on 100 independent runs the value of the discrete MOT for $((\tilde{\mu}_I)^2_{\mathcal{P}(\tilde{\nu}_I)}, \nu_I)$ with $I = 100$: the average is equal to 2.0064 and the standard deviation is equal to 0.2213, which gives $[1.9631, 2.0498]$ as 95% confidence interval, which approximates well the value of the continuous MOT.

5.2.2. Model-free bounds on a best-of option

Let (G^1, G^2) be a centered Gaussian vector with covariance matrix Σ . We denote by μ the law of (X^1, X^2) with $X^\ell = \exp(G^\ell - \Sigma_{\ell\ell}/2)$ for $\ell \in \{1, 2\}$, and by ν the law of (Y^1, Y^2) with $Y^\ell = \exp(\sqrt{2}G^\ell - \Sigma_{\ell\ell})$. In the financial context, this choice of marginal laws is usual and corresponds to a two-dimensional Black–Scholes model: (X^1, X^2) is the price of two assets at time $t > 0$ and (Y^1, Y^2) is the price of these assets at time $2t$. We are interested in an option that pays $\max(Y^1 - X^1, Y^2 - X^2, 0)$, i.e. the best arithmetic performance of the two assets, if positive. The price of this option in the Black–Scholes model can easily be calculated by using a Monte-Carlo algorithm.

Let $(X_1^1, X_1^2), \dots, (X_I^1, X_I^2)$ and $(Y_1^1, Y_1^2), \dots, (Y_I^1, Y_I^2)$ denote independent samples respectively distributed according to μ and ν . We set $\tilde{\mu}_I = \frac{1}{I} \sum_{i=1}^I \delta_{(\tilde{X}_i^1, \tilde{X}_i^2)}$ and $\tilde{\nu}_I = \frac{1}{I} \sum_{i=1}^I \delta_{(\tilde{Y}_i^1, \tilde{Y}_i^2)}$, with $(\tilde{X}_i^1, \tilde{X}_i^2) = (X_i^1 + 1 - \bar{X}_I^1, X_i^2 + 1 - \bar{X}_I^2)$, $(\tilde{Y}_i^1, \tilde{Y}_i^2) = (Y_i^1 + 1 - \bar{Y}_I^1, Y_i^2 + 1 - \bar{Y}_I^2)$, $\bar{X}_I^\ell = \frac{1}{I} \sum_{i=1}^I X_i^\ell$ and $\bar{Y}_I^\ell = \frac{1}{I} \sum_{i=1}^I Y_i^\ell$. We calculate $(\tilde{\mu}_I)^2_{\mathcal{P}(\tilde{\nu}_I)}$ numerically by using again the quadratic optimization solver COIN-OR, and then solve the discrete MOT problem between $(\tilde{\mu}_I)^2_{\mathcal{P}(\tilde{\nu}_I)}$ and $\tilde{\nu}_I$.

We now turn to our example illustrated in Figure 4. We have considered the following covariance matrix $\Sigma = \begin{bmatrix} 0.5 & 0.1 \\ 0.1 & 0.1 \end{bmatrix}$. With this choice, the Black–Scholes price of the option is approximately equal to 0.345. With $I = 100$, we have calculated on 100 independent runs the value of the minimization and the maximization programs, and then computed the mean values. We have thus obtained 0.2293 for the lower bound price and 0.4111 for the upper bound price. The corresponding standard variations are respectively 0.0848 and 0.1422, which makes 95% confidence intervals with half lengths 0.017 and 0.028. In Figure 4, we have plotted the discrete MOT on the same sample for the minimization and the maximization problem. Precisely, we have plotted the points $(\tilde{X}_i^1, \tilde{X}_i^2)^*$, $i \in \{1, \dots, I\}$ in the hyperplane $z = 0$ and the points $(\tilde{Y}_i^1, \tilde{Y}_i^2)$ in the hyperplane $z = 1$. The edges between the points $(\tilde{X}_i^1, \tilde{X}_i^2)^*$ and $(\tilde{Y}_i^1, \tilde{Y}_i^2)$ indicate that the optimal coupling gives a positive weight to the corresponding transitions. The difference between the two optimal couplings is clear. We can heuristically explain the graphs as follows. The cost function $c(x, y) = \max(y^1 - x^1, y^2 - x^2, 0)$ will anyway be positive for a large increase of one of the two assets. Therefore, to minimize the cost, one has to gather the large increases of Asset 1 and Asset 2. Instead, to maximize the cost, it is better to gather an increase of one asset with a decrease of the other one.

The CPU time needed for the computation of the Wasserstein projection and for the linear programming problem is reported in Table 2. The dimension $d = 1$ rows of the table correspond to the MOT problem between the laws of X_1 and Y_1 for the cost function $\max(y - x, 0)$. What mainly influences the computation time is the dimension I^2 in which

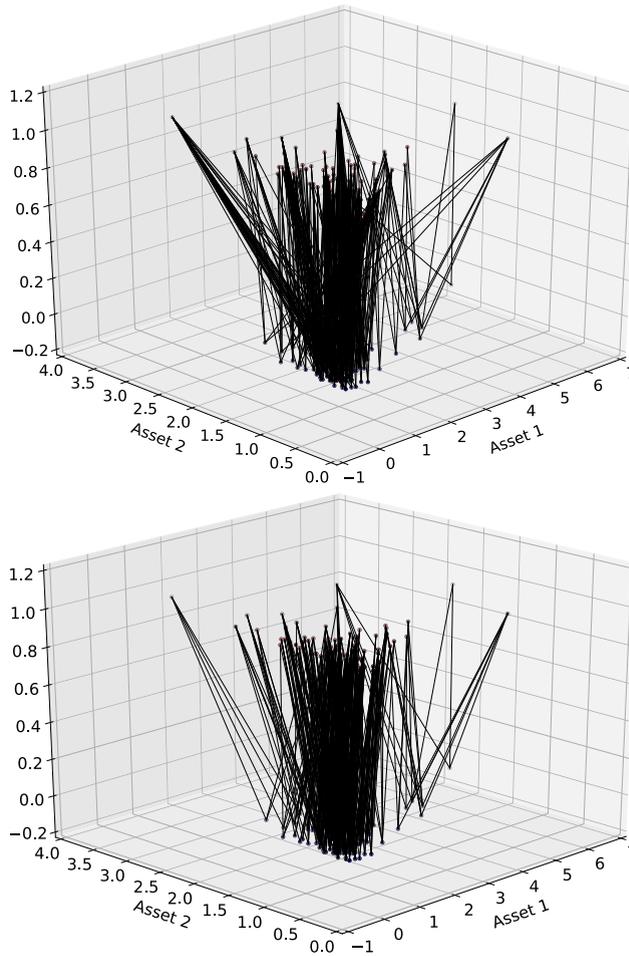


Fig. 4. Discrete MOT ($I = 100$) in dimension 2 for the minimization problem (top) and the maximization problem (bottom).

Table 2

Computation time on a CPU Intel Core i7 at 2.6 GHz with COIN-OR of the quadratic and linear problems in dimensions $d = 1$ and $d = 2$

	I				
	100	150	200	300	500
Quadratic problem (1.1), $d = 1$	1.5 s	4.8 s	18 s	88 s	673 s
Quadratic problem (1.1), $d = 2$	1.3 s	10 s	22 s	105 s	807 s
Linear problem (1.3), $d = 1$	0.3 s	0.78 s	2 s	6.6 s	41 s
Linear problem (1.3), $d = 2$	0.43 s	2 s	4.5 s	19.5 s	120 s

the optimal matrix (r_{ij}) has to be found. The dimension d of the underlying space of the probability measures has a low impact on the computation time for the quadratic problem (1.1), since the number of equality constraints $2I$ does not change with d . Instead, it has some impact on the linear programming problem (1.3), since the number of equality constraints $(2+d)I$ increases with d . Nonetheless, since the resolution of the linear problem is much less time consuming than the resolution of the quadratic problem, the impact of the dimension d on the overall computation time is rather mild.

5.3. Further directions

In view of Proposition 3.1 and Corollary 3.2, it would be nice to prove the stability of

$$\inf_{\pi \in \Pi^M(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \pi(dx, dy)$$

with respect to μ and ν in $\mathcal{P}(\mathbb{R}^d)$ for the weak convergence topology or the Wasserstein distance. On our numerical example of Figure 3 where the continuous MOT is explicit, the convergence of the discrete optimal cost towards the continuous one seems to hold. We plan to investigate this property in a future work. Note that for cost functions satisfying the so-called Spence–Mirrlees condition (see [17]), the stability of left-curtain couplings obtained by Juillet [18] is an important step in that direction.

To overcome the sample size limitation for the linear programming solvers to compute the solution of problem (1.3), one can contemplate introducing an entropic regularization of this problem similar to the one proposed by Benamou et al. [9] for discrete optimal transport. For $\mu_I = \sum_{i=1}^I p_i \delta_{x_i} \leq_{\text{cx}} \nu_J = \sum_{j=1}^J q_j \delta_{y_j}$ and $\varepsilon > 0$, the regularized problem is the minimization of

$$\sum_{i=1}^I \sum_{j=1}^J r_{ij}^\varepsilon (c(x_i, y_j) + \varepsilon (\ln r_{ij}^\varepsilon - 1))$$

under the constraints $r_{ij}^\varepsilon \geq 0$, $\sum_{i=1}^I r_{ij}^\varepsilon = q_j$ for $j \in \{1, \dots, J\}$, $\sum_{j=1}^J r_{ij}^\varepsilon = p_i$ and $\sum_{j=1}^J r_{ij}^\varepsilon y_j = p_i x_i$ for $i \in \{1, \dots, I\}$. Since the constraints are affine, this problem can be solved by the iterative Bregman projections presented in [9]. In particular the solution is obtained by iterating successive entropic projections on the first marginal law constraints, on the second marginal law constraints and on the martingale constraints. The two first projections are explicit (see for instance Proposition 1 [9]). The entropic projection on the martingale constraints can be computed using the generalized iterative scaling algorithm introduced by Darroch and Ratcliff [10]. Such an approach combined with a relaxation of the martingale constraint has been recently investigated by Guo and Oblòj [16].

Appendix: Technical lemmas

Lemma A.1. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$. Then, we have $\mu \leq_{\text{cx}} \nu$ if, and only if,*

$$\forall \phi : \mathbb{R}^d \rightarrow \mathbb{R} \text{ convex and such that } \sup_{x \in \mathbb{R}^d} \frac{|\phi(x)|}{1 + |x|} < \infty, \int_{\mathbb{R}^d} \phi(x) \mu(dx) \leq \int_{\mathbb{R}^d} \phi(x) \nu(dx).$$

Proof. Let $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ be a convex function. We define $\phi^*(y) = \sup_{x \in \mathbb{R}^d} x \cdot y - \phi(x)$ the Legendre–Fenchel transform of ϕ and have

$$\phi(x) = \phi^{**}(x) = \sup_{y \in \mathbb{R}^d} x \cdot y - \phi^*(y).$$

The function $\phi^* : \mathbb{R}^d \rightarrow [-\phi(0), +\infty]$ is a convex lower semicontinuous function. Therefore, for any $n \geq 1$, there exists y_n with Euclidean norm $|y_n| \leq n$ and $\inf_{|y| \leq n} \phi^*(y) = \phi^*(y_n)$. There exists $n_0 \in \mathbb{N}^*$ such that $\phi^*(y_n) < \infty$ for $n \geq n_0$, otherwise we would have $\phi^* = +\infty$ and then $\phi = -\infty$. We set $\phi_n(x) = \sup_{|y| \leq n} x \cdot y - \phi^*(y)$ and have for $n \geq n_0$

$$x \cdot y_n - \phi^*(y_n) \leq \phi_n(x) \leq n|x| + \phi(0).$$

Thus, ϕ_n is with affine growth and therefore $\int_{\mathbb{R}^d} \phi_n(x) \mu(dx) \leq \int_{\mathbb{R}^d} \phi_n(x) \nu(dx)$. By the monotone convergence theorem the integrals $\int_{\mathbb{R}^d} (\phi_n - \phi_{n_0})(x) \mu(dx)$ (resp. $\int_{\mathbb{R}^d} (\phi_n - \phi_{n_0})(x) \nu(dx)$) converge to $\int_{\mathbb{R}^d} (\phi - \phi_{n_0})(x) \mu(dx)$ (resp. $\int_{\mathbb{R}^d} (\phi - \phi_{n_0})(x) \nu(dx)$) as $n \rightarrow \infty$. We conclude that $\int_{\mathbb{R}^d} \phi(x) \mu(dx) \leq \int_{\mathbb{R}^d} \phi(x) \nu(dx)$. \square

Lemma A.2. *Let $f, g : [0, 1] \rightarrow \mathbb{R}$ be two convex functions and h denote the convex hull of $f - g$. Then $f - h$ is convex.*

Proof. Let $0 \leq p < q \leq 1$ and $\alpha \in [0, 1]$. If $h(\alpha p + (1 - \alpha)q) = (f - g)(\alpha p + (1 - \alpha)q)$, then, using the convexity of g , then the fact that h is bounded from above by $f - g$ for the two inequalities, we obtain that

$$\begin{aligned} (f - h)(\alpha p + (1 - \alpha)q) &= g(\alpha p + (1 - \alpha)q) \leq \alpha g(p) + (1 - \alpha)g(q) \\ &= \alpha(f(p) - (f - g)(p)) + (1 - \alpha)(f(q) - (f - g)(q)) \\ &\leq \alpha(f - h)(p) + (1 - \alpha)(f - h)(q). \end{aligned} \tag{A.1}$$

Otherwise, h is affine on some interval $[r, s]$ with $0 \leq r < \alpha p + (1 - \alpha)q < s \leq 1$, $h(r) = (f - g)(r)$ and $h(s) = (f - g)(s)$. If $r \in (p, \alpha p + (1 - \alpha)q)$, then replacing α by $\frac{q-r}{q-p}$ in (A.1), we get $(f - h)(r) \leq \frac{q-r}{q-p}(f - h)(p) + \frac{r-p}{q-p}(f - h)(q)$

so that $(f - h)(r \vee p) \leq \frac{q-r \vee p}{q-p}(f - h)(p) + \frac{r \vee p - p}{q-p}(f - h)(q)$. In a symmetric way, $(f - h)(s \wedge q) \leq \frac{q-s \wedge q}{q-p}(f - h)(p) + \frac{s \wedge q - p}{q-p}(f - h)(q)$. Hence,

$$\begin{aligned} & \frac{s \wedge q - (\alpha p + (1 - \alpha)q)}{s \wedge q - r \vee p}(f - h)(r \vee p) + \frac{(\alpha p + (1 - \alpha)q) - r \vee p}{s \wedge q - r \vee p}(f - h)(s \wedge q) \\ & \leq \alpha(f - h)(p) + (1 - \alpha)(f - h)(q). \end{aligned}$$

By convexity of f and the affine property of h on the interval $[r \vee p, s \wedge q]$ containing $\alpha p + (1 - \alpha)q$, the left-hand side is not smaller than $(f - h)(\alpha p + (1 - \alpha)q)$. □

Lemma A.3. *Let $f : (0, 1) \rightarrow \mathbb{R}$ be a non-decreasing function and η denote the probability distribution of $f(U)$ for U uniformly distributed on $(0, 1)$. Then f and the quantile function F_η^{-1} coincide away from the at most countable set of their common discontinuities and even everywhere on $(0, 1)$ if f is moreover left-continuous.*

Proof. The random variables $f(U)$ and $F_\eta^{-1}(U)$ are both distributed according to η . Hence for $p \in (0, 1)$, $\mathbb{P}(f(U) \leq F_\eta^{-1}(p)) = \mathbb{P}(F_\eta^{-1}(U) \leq F_\eta^{-1}(p)) \geq p$ so that $F_\eta^{-1}(p) \geq \sup_{q \in (0, p)} f(q)$. By symmetry, $f(p) \geq \sup_{q \in (0, p)} F_\eta^{-1}(q)$ with the supremum equal to $F_\eta^{-1}(p)$ by left-continuity and monotonicity of F_η^{-1} . Hence $f(p) \geq F_\eta^{-1}(p) \geq \sup_{q \in (0, p)} f(q)$ with the supremum equal to $f(p)$ when f is left-continuous. □

Lemma A.4. *For $x \in \mathbb{R}$, any non empty subset \mathcal{P}^x of $\{\eta \in \mathcal{P}_1(\mathbb{R}) : \int_{\mathbb{R}} y \eta(dy) = x\}$ has an infimum π for the convex order. Moreover for all $q \in [0, 1]$, $\int_q^1 F_\pi^{-1}(p) dp = \inf_{\eta \in \mathcal{P}^x} \int_q^1 F_\eta^{-1}(p) dp$.*

Proof. The existence of the infimum is given by Kertz and Rösler [20] p. 162. These authors work with the characterization of the convex order in terms of the cumulative distribution functions. By the more convenient characterization in terms of the quantile functions recalled in (2.2), it is enough to check that for all $q \in [0, 1]$, $\tilde{\psi}(q) := \inf_{\eta \in \mathcal{P}^x} \int_q^1 F_\eta^{-1}(p) dp = \int_q^1 F_\pi^{-1}(p) dp$ for some probability measure $\pi \in \mathcal{P}_1(\mathbb{R})$ such that $\int_{\mathbb{R}} y \pi(dy) = x$. For $\eta \in \mathcal{P}^x$, $\int_0^1 F_\eta^{-1}(p) dp = x$ and for all $q \in [0, 1]$, $\int_q^1 F_\eta^{-1}(p) dp \geq (1 - q)x$. Therefore for all $q \in [0, 1]$, $\tilde{\psi}(q) \geq (1 - q)x$, $\tilde{\psi}(0) = x$ and $\tilde{\psi}(1) = 0$. The function $\tilde{\psi}$ being concave on $[0, 1]$ as the infimum of concave functions it is continuous on $(0, 1)$. Since for $\eta \in \mathcal{P}^x$, $\tilde{\psi}(q) \leq \int_q^1 F_\eta^{-1}(p) dp$, $\tilde{\psi}$ is continuous at 0 and 1 and therefore on $[0, 1]$. Denoting its left-hand derivative by f , one has $\int_0^1 |f(p)| dp < \infty$ and for all $q \in [0, 1]$, $\tilde{\psi}(q) = \int_q^1 f(p) dp$ with f non-decreasing. One concludes by defining π as the image of the Lebesgue measure on $(0, 1)$ by f . □

Lemma A.5. *Let $\varrho > 1$ and $\eta, \eta_1, \eta_2 \in \mathcal{P}_\varrho(\mathbb{R}^d)$. Then*

$$W_\varrho^\varrho\left(\eta, \frac{\eta_1 + \eta_2}{2}\right) \leq \frac{1}{2}(W_\varrho^\varrho(\eta, \eta_1) + W_\varrho^\varrho(\eta, \eta_2)), \tag{A.2}$$

Besides, when η is absolutely continuous with respect to the Lebesgue measure or $d = 1$ and η has no atom, equality holds if and only if $\eta_1 = \eta_2$. Last, when $d = 1$, the statements remain valid with $\frac{\eta_1 + \eta_2}{2}$ replaced by the distribution $\bar{\eta}_{12}$ of $\frac{F_{\eta_1}^{-1} + F_{\eta_2}^{-1}}{2}(U)$ with U uniformly distributed on $[0, 1]$.

Proof. Let $\eta_3 = \frac{\eta_1 + \eta_2}{2}$. For $i \in \{1, 2, 3\}$, there exists an optimal probability measure $\pi_i \in \Pi(\eta, \eta_i)$ that satisfies $W_\varrho^\varrho(\eta, \eta_i) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - x|^\varrho \pi_i(dx, dy)$. Since $\frac{\pi_1 + \pi_2}{2} \in \Pi(\eta, \eta_3)$, we have

$$W_\varrho^\varrho\left(\eta, \frac{\eta_1 + \eta_2}{2}\right) \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - x|^\varrho \frac{\pi_1 + \pi_2}{2}(dx, dy) = \frac{1}{2}(W_\varrho^\varrho(\eta, \eta_1) + W_\varrho^\varrho(\eta, \eta_2)). \tag{A.3}$$

We now suppose that η is absolutely continuous with respect to the Lebesgue measure. We know by Theorem 6.2.4 in [4] that the probability measure $\pi_i \in \Pi(\eta, \eta_i)$ satisfying $W_\varrho^\varrho(\eta, \eta_i) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - x|^\varrho \pi_i(dx, dy)$ is unique, and writes $\pi_i(dx, dy) = \eta(dx) \delta_{T_i(x)}(dy)$ for some Borel map $T_i : \mathbb{R}^d \rightarrow \mathbb{R}^d$. If (A.2) is an equality, then the inequality in (A.3) is also an equality and, by uniqueness, $\frac{\pi_1 + \pi_2}{2} = \pi_3$. Hence we get $\eta(dx) \delta_{T_3(x)}(dy) = \eta(dx) \frac{1}{2}(\delta_{T_1(x)}(dy) + \delta_{T_2(x)}(dy))$, which gives $T_1(x) = T_2(x) = T_3(x)$, $\eta(dx)$ -a.e., and implies $\eta_1 = \eta_2$.

When $d = 1$, if η has no atom, according to Theorem 2.9 in [25], π_i is still unique and given by $\eta(dx)\delta_{F_{\eta_1}^{-1}(F_\eta(x))}(dy)$, so that the same conclusion holds. Still when $d = 1$, since $F_{\tilde{\eta}_{12}}^{-1} = \frac{F_{\eta_1}^{-1} + F_{\eta_2}^{-1}}{2}$, by Proposition 2.17 [25] and strict convexity of $x \mapsto |x|^e$,

$$\begin{aligned} W_\varrho^e(\tilde{\eta}_{12}, \eta) &= \int_0^1 \left| \frac{1}{2}(F_{\eta_1}^{-1}(p) + F_{\eta_2}^{-1}(p)) - F_\eta^{-1}(p) \right|^e dp \\ &\leq \frac{1}{2} \left(\int_0^1 |F_{\eta_1}^{-1}(p) - F_\eta^{-1}(p)|^e dp + \int_0^1 |F_{\eta_2}^{-1}(p) - F_\eta^{-1}(p)|^e dp \right) = \frac{1}{2} (W_\varrho^e(\eta_1, \eta) + W_\varrho^e(\eta_2, \eta)) \end{aligned}$$

with equality iff dp a.e. $F_{\eta_1}^{-1}(p) = F_{\eta_2}^{-1}(p)$ i.e. $\eta_1 = \eta_2$. □

Lemma A.6. Let $\varrho > 1$ and $\mu, \nu \in \mathcal{P}_\varrho(\mathbb{R})$. The function $(0, 1) \ni p \mapsto F_{\nu_{\tilde{\mathcal{P}}(\mu)}}^{-1}(p) - F_\nu^{-1}(p)$ is non-decreasing.

Proof. It is enough to check that for $\eta \in \mathcal{P}_\varrho(\mathbb{R}) \cap \tilde{\mathcal{P}}(\mu)$ such that $p \mapsto F_\eta^{-1}(p) - F_\nu^{-1}(p)$ is not non-decreasing then $W_\varrho^e(\nu, \nu_{\tilde{\mathcal{P}}(\eta)}) < W_\varrho^e(\nu, \eta)$ (indeed $F_{\nu_{\tilde{\mathcal{P}}(\eta)}}^{-1}(p) - F_\nu^{-1}(p)$ is non-decreasing and $\nu_{\tilde{\mathcal{P}}(\eta)} \in \tilde{\mathcal{P}}(\eta) \subset \tilde{\mathcal{P}}(\mu)$). By Proposition 2.17 [25] and the definition of $\nu_{\tilde{\mathcal{P}}(\eta)}$,

$$W_\varrho^e(\nu, \nu_{\tilde{\mathcal{P}}(\eta)}) = \int_0^1 |F_{\nu_{\tilde{\mathcal{P}}(\eta)}}^{-1}(p) - F_\nu^{-1}(p)|^e dp = \int_0^1 |f(p)|^e dp,$$

where $f(p)$ denotes the left-hand derivative of the concave hull $\tilde{\psi}(q)$ of $[0, 1] \ni q \mapsto \phi(q) := \int_q^1 F_\eta^{-1}(p) - F_\nu^{-1}(p) dp$. Since $\forall q \in [0, 1], \int_q^1 F_\eta^{-1}(p) - F_\nu^{-1}(p) dp \leq \int_q^1 F_\eta^{-1}(p) dp - q \int_0^1 F_\nu^{-1}(p) dp$ where the right-hand side is a concave function of q , $\tilde{\psi}(1) = \phi(1) = 0$ and $\tilde{\psi}(0) = \phi(0) = \int_0^1 F_\eta^{-1}(p) - F_\nu^{-1}(p) dp$. Now either $\tilde{\psi}$ and ϕ coincide on $[0, 1]$ and $F_\eta^{-1} - F_\nu^{-1}$ is non-decreasing or the open set $\{q \in [0, 1] : \tilde{\psi}(q) > \phi(q)\}$ is non empty and writes as the at most countable union $\bigcup_{i \in I} (p_i, q_i)$ of disjoint intervals with $0 \leq p_i < q_i \leq 1$, $\tilde{\psi}(p_i) = \phi(p_i)$, $\tilde{\psi}(q_i) = \phi(q_i)$ and $\tilde{\psi}$ affine on $[p_i, q_i]$. For each i in the non empty set I , for all $p \in (p_i, q_i)$, $f(p) = \frac{\tilde{\psi}(q_i) - \tilde{\psi}(p_i)}{q_i - p_i} = \frac{\phi(q_i) - \phi(p_i)}{q_i - p_i} = \frac{\int_{p_i}^{q_i} F_\nu^{-1}(q) - F_\eta^{-1}(q) dq}{q_i - p_i}$ so that, by Jensen's inequality,

$$\forall i \in I, \int_{p_i}^{q_i} |f(p)|^e dp < \int_{p_i}^{q_i} |F_\nu^{-1}(p) - F_\eta^{-1}(p)|^e dp. \tag{A.4}$$

For $p \in (0, 1] \setminus \bigcup_{i \in I} (p_i, q_i]$, either $\tilde{\psi}$ is equal to ϕ on a left-hand neighbourhood of p or there is an accumulation of intervals $((p_{i_n}, q_{i_n}))_{n \in \mathbb{N}}$ at the left of p with $(i_n)_{n \in \mathbb{N}}$ a sequence of distinct elements of I , $q_{i_n} < p$ for all $n \in \mathbb{N}$ and $\lim_{n \rightarrow \infty} q_{i_n} = p$. For q in the left-hand neighbourhood of p in the first case and in $\{q_{i_n} : n \in \mathbb{N}\}$ in the second one, $\tilde{\psi}(p) - \tilde{\psi}(q) = \phi(p) - \phi(q) = \int_q^p F_\nu^{-1}(r) - F_\eta^{-1}(r) dr$. By the left-continuity of $q \mapsto F_\nu^{-1}(q) - F_\eta^{-1}(q)$ and the definition of f , one concludes that $f(p) = F_\nu^{-1}(p) - F_\eta^{-1}(p)$. Therefore $\int_0^1 \mathbb{1}_{\{p \notin \bigcup_{i \in I} (p_i, q_i]\}} |f(p)|^e dp = \int_0^1 \mathbb{1}_{\{p \notin \bigcup_{i \in I} (p_i, q_i]\}} |F_\nu^{-1}(p) - F_\eta^{-1}(p)|^e dp$ which combined with (A.4) and Proposition 2.17 [25] leads to $\int_0^1 |f(p)|^e dp < \int_0^1 |F_\nu^{-1}(p) - F_\eta^{-1}(p)|^e dp = W_\varrho^e(\nu, \eta)$ when $\tilde{\psi}$ and ϕ do not coincide on $[0, 1]$. □

Remark A.7. Lemma 2.8 can be proved by similar arguments. But to exhibit $\tilde{\eta} \in \underline{\mathcal{P}}(\nu)$ with $W_\varrho^e(\mu, \tilde{\eta}) \leq W_\varrho^e(\mu, \eta)$ and $F_{\tilde{\eta}}^{-1} - F_\mu^{-1}$ non-increasing when $\eta \in \underline{\mathcal{P}}(\nu)$ is such that $F_\eta^{-1} - F_\mu^{-1}$ is not non-increasing, we chose a more elementary transformation exploiting directly the lack of monotonicity in place of $\mu_{\underline{\mathcal{P}}(\eta)}$.

Lemma A.8. Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be two distinct probability measures such that $\mu \leq_{\text{cx}} \nu$ and $(\underline{t}_n, \bar{t}_n)$, $1 \leq n \leq N \in \mathbb{N}^* \cup \{\infty\}$ be the irreducible components of (μ, ν) . Then, we have

$$\left\{ q \in [0, 1], \int_0^q F_\mu^{-1}(p) dp > \int_0^q F_\nu^{-1}(p) dp \right\} = \bigcup_{n=1}^N (F_\mu(\underline{t}_n), F_\mu(\bar{t}_n -)).$$

Proof. For $\eta \in \mathcal{P}_1(\mathbb{R})$, let $\varphi_\eta(t) = \int_{-\infty}^t F_\eta(x) dx$ for $t \in \mathbb{R}$, $\psi_\eta(q) = \int_0^q F_\eta^{-1}(p) dp$ for $q \in [0, 1]$ and $\psi_\eta(q) = +\infty$ for $q \notin [0, 1]$. One has $\varphi_\mu(t) = \int_{\mathbb{R}} (t-x)^+ \mu(dx) \leq \int_{\mathbb{R}} (t-y)^+ \nu(dy) = \varphi_\nu(t)$ for all $t \in \mathbb{R}$ and $(\underline{t}_n, \bar{t}_n)$, $1 \leq n \leq N \in \mathbb{N}^* \cup \{\infty\}$ is the countable family of disjoint intervals such that

$$\{t \in \mathbb{R} : \varphi_\mu(t) < \varphi_\nu(t)\} = \bigcup_{n=1}^N (\underline{t}_n, \bar{t}_n). \tag{A.5}$$

Since φ_η and ψ_η are the antiderivatives of two reciprocal non-decreasing functions, it is well known they are the Legendre–Fenchel transforms of each other i.e. $\varphi_\eta(t) = \sup_{q \in \mathbb{R}} \{qt - \psi_\eta(q)\}$. In fact, for $t \in \mathbb{R}$, if $F_\eta(t-) > 0$ then $F_\eta^{-1}(q) < t$ for $q \in (0, F_\eta(t-))$, if $F_\eta(t) < 1$ then $F_\eta^{-1}(q) > t$ for $q \in (F_\eta(t), 1)$ and if $F_\eta(t-) < F_\eta(t)$ then $F_\eta^{-1}(q) = t$ for $q \in (F_\eta(t-), F_\eta(t))$. We deduce that $\sup_{q \in \mathbb{R}} \{qt - \psi_\eta(q)\} = F_\eta(t) t - \psi_\eta(F_\eta(t)) = \int_0^{F_\eta(t)} (t - F_\eta^{-1}(p)) dp = \int_0^1 (t - F_\eta^{-1}(p))^+ dp = \varphi_\eta(t)$ and

$$\forall t \in \mathbb{R}, \quad \{q \in \mathbb{R}, qt - \psi_\eta(q) = \varphi_\eta(t)\} = [F_\eta(t-), F_\eta(t)]. \tag{A.6}$$

Therefore, we have

$$\begin{aligned} \{t \in \mathbb{R}, \varphi_\mu(t) < \varphi_\nu(t)\} &\subset \{t \in \mathbb{R}, \forall q \in [F_\nu(t-), F_\nu(t)], qt - \psi_\mu(q) < qt - \psi_\nu(q)\} \\ &= \{t \in \mathbb{R}, \forall q \in [F_\nu(t-), F_\nu(t)], \psi_\mu(q) > \psi_\nu(q)\}. \end{aligned}$$

Hence

$$\bigcup_{1 \leq n \leq N} (F_\nu(\underline{t}_n), F_\nu(\bar{t}_n-)) \subset \bigcup_{1 \leq n \leq N} \bigcup_{t \in (\underline{t}_n, \bar{t}_n)} [F_\nu(t-), F_\nu(t)] \subset \{q \in [0, 1], \psi_\mu(q) > \psi_\nu(q)\}. \tag{A.7}$$

Now, we observe that $(0, 1) \subset \bigcup_{t \in \mathbb{R}} [F_\mu(t-), F_\mu(t)]$ and, for $t \in \mathbb{R}$ such that $F_\mu(t-) < F_\mu(t)$, $\psi_\mu(q)$ is affine for $q \in [F_\mu(t-), F_\mu(t)]$. Using the convexity of ψ_ν , we get

$$\{q \in [0, 1], \psi_\mu(q) > \psi_\nu(q)\} \subset \bigcup_{t \in \mathbb{R}: \psi_\mu(F_\mu(t)) > \psi_\nu(F_\mu(t)) \text{ or } \psi_\mu(F_\mu(t-)) > \psi_\nu(F_\mu(t-))} [F_\mu(t-), F_\mu(t)].$$

If $\psi_\mu(F_\mu(t)) > \psi_\nu(F_\mu(t))$, we have $\varphi_\mu(t) = F_\mu(t)t - \psi_\mu(F_\mu(t)) < F_\mu(t)t - \psi_\nu(F_\mu(t)) \leq \varphi_\nu(t)$ by using that φ_ν is the Legendre transform of ψ_ν . Similarly, $\psi_\mu(F_\mu(t-)) > \psi_\nu(F_\mu(t-)) \implies \varphi_\mu(t) < \varphi_\nu(t)$, and we get

$$\{q \in [0, 1], \psi_\mu(q) > \psi_\nu(q)\} \subset \bigcup_{t \in \mathbb{R}: \varphi_\mu(t) < \varphi_\nu(t)} [F_\mu(t-), F_\mu(t)] \subset \bigcup_{1 \leq n \leq N} [F_\mu(\underline{t}_n), F_\mu(\bar{t}_n-)].$$

By (A.6), if $\underline{t}_n > -\infty$, $\psi_\mu(F_\mu(\underline{t}_n)) = \underline{t}_n F_\mu(\underline{t}_n) - \varphi_\mu(\underline{t}_n)$. Since $\varphi_\mu(\underline{t}_n) = \varphi_\nu(\underline{t}_n)$ and the Legendre transform ψ_ν of φ_ν is not greater than ψ_μ , we deduce that $\psi_\mu(F_\mu(\underline{t}_n)) = \psi_\nu(F_\mu(\underline{t}_n))$. In the same way, if $\bar{t}_n < +\infty$, then $\psi_\mu(F_\mu(\bar{t}_n-)) = \psi_\nu(F_\mu(\bar{t}_n-))$ so that

$$\{q \in [0, 1], \psi_\mu(q) > \psi_\nu(q)\} \subset \bigcup_{1 \leq n \leq N} (F_\mu(\underline{t}_n), F_\mu(\bar{t}_n-)). \tag{A.8}$$

Now, (A.5) implies that $F_\mu(\underline{t}_n) \leq F_\nu(\underline{t}_n)$. If $F_\mu(\underline{t}_n) < F_\nu(\underline{t}_n)$, we necessarily have $\underline{t}_n > -\infty$, and for $q \in (F_\mu(\underline{t}_n), F_\nu(\underline{t}_n))$, we have $F_\nu^{-1}(q) \leq \underline{t}_n$ and $F_\mu^{-1}(q) > \underline{t}_n$ since F_μ is right-continuous. Therefore, we have $\int_{F_\mu(\underline{t}_n)}^p F_\mu^{-1}(q) dq > \int_{F_\mu(\underline{t}_n)}^p F_\nu^{-1}(q) dq$ and thus $\psi_\mu(p) > \psi_\nu(p)$ for $p \in (F_\mu(\underline{t}_n), F_\nu(\underline{t}_n))$. Similarly, we show that $\psi_\mu(q) > \psi_\nu(q)$ for $q \in [F_\nu(\bar{t}_n-), F_\mu(\bar{t}_n-))$, which, with (A.7) and (A.8), gives the claim. \square

References

[1] A. Alfonsi, J. Corbetta and B. Jourdain. Sampling of probability measures in the convex order and approximation of martingale optimal transport problems. arXiv eprint, 2017. Available at [arXiv:1709.05287](https://arxiv.org/abs/1709.05287). MR3951948 <https://doi.org/10.1142/S021902491950002X>

[2] A. Alfonsi, J. Corbetta and B. Jourdain. Sampling of one-dimensional probability measures in the convex order and computation of robust option price bounds probability measures in the convex order and computation of robust option prices bounds. *Int. J. Theor. Appl. Finance* **22** (3) (2019). MR3951948 <https://doi.org/10.1142/S021902491950002X>

- [3] J.-J. Alibert, G. Bouchitté and T. Champion. A new class of cost for optimal transport planning. Preprint hal-01741688, 2018.
- [4] L. Ambrosio, N. Gigli and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures. Lectures in Mathematics ETH Zürich*. Birkhäuser Verlag, Basel, 2005. MR2129498
- [5] J. Backhoff Veraguas, M. Beiglböck and G. Pammer. Existence and cyclical monotonicity for weak transport costs. arXiv e-print, 2018. Available at arXiv:1809.05893.
- [6] D. Baker. Martingales with specified marginals. Theses Université Pierre et Marie Curie – Paris VI, 2012.
- [7] M. Beiglböck, P. Henry-Labordère and F. Penkner. Model-independent bounds for option prices – a mass transport approach. *Finance Stoch.* **17** (3) (2013) 477–501. MR3066985 <https://doi.org/10.1007/s00780-013-0205-8>
- [8] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.* **44** (1) (2016) 42–106. MR3456332 <https://doi.org/10.1214/14-AOP966>
- [9] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna and G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.* **37** (2) A1111–A1138. MR3340204 <https://doi.org/10.1137/141000439>
- [10] J. N. Darroch and D. Ratcliff. Generalized iterative scaling for log-linear models. *Ann. Math. Stat.* **43** (1972) 1470–1480. MR0345337 <https://doi.org/10.1214/aoms/1177692379>
- [11] C. Dellacherie and P.-A. Meyer. *Probabilities and Potential. North-Holland Mathematics Studies*. **29**. North-Holland Publishing Co., Amsterdam–New York, 1978. MR0521810
- [12] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** (3–4) (2015) 707–738. MR3383341 <https://doi.org/10.1007/s00440-014-0583-7>
- [13] N. Gozlan and N. Juillet. On a mixture of Brenier and Strassen Theorems. *Proceedings of the London Mathematical Society* **120** (3) (2020) 434–463. <https://doi.org/10.1112/plms.12302>
- [14] 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) (2018) 1667–1693. MR3825894 <https://doi.org/10.1214/17-AIHP851>
- [15] N. Gozlan, C. Roberto, P.-M. Samson and P. Tetali. Kantorovich duality for general transport costs and applications. arXiv eprint arXiv:1412.7480v4, 2015. MR3706606 <https://doi.org/10.1016/j.jfa.2017.08.015>
- [16] G. Guo and J. Oblöj. Computational methods for martingale optimal transport problems. arXiv e-prints arXiv:1710.07911, 2017. MR3432145 <https://doi.org/10.1137/140960724>
- [17] P. Henry-Labordère and N. Touzi. An explicit martingale version of the one-dimensional Brenier theorem. *Finance Stoch.* **20** (3) (2016) 635–668. MR3519164 <https://doi.org/10.1007/s00780-016-0299-x>
- [18] N. Juillet. Stability of the shadow projection and the left-curtain coupling. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** (4) (2016) 1823–1843. MR3573297 <https://doi.org/10.1214/15-AIHP700>
- [19] R. P. Kertz and U. Rösler. Stochastic and convex orders and lattices of probability measures, with a martingale interpretation. *Israel J. Math.* **77** (1–2) (1992) 129–164. MR1194790 <https://doi.org/10.1007/BF02808015>
- [20] R. P. Kertz and U. Rösler. Complete lattices of probability measures with applications to martingale theory. In *Game Theory, Optimal Stopping, Probability and Statistics* 153–177. *IMS Lecture Notes Monogr. Ser.* **35**. Inst. Math. Statist, Beachwood, OH, 2000. MR1833858 <https://doi.org/10.1214/Inms/1215089751>
- [21] A. Müller and M. Scarsini. Stochastic order relations and lattices of probability measures. *SIAM J. Optim.* **16** (4) (2006) 1024–1043. MR2219130 <https://doi.org/10.1137/040611021>
- [22] G. Pagès and J. Printems. Optimal quadratic quantization for numerics: The Gaussian case. *Monte Carlo Methods Appl.* **9** (2) (2003) 135–165. MR2006483 <https://doi.org/10.1163/156939603322663321>
- [23] G. Pagès and B. Wilbertz. Intrinsic stationarity for vector quantization: Foundation of dual quantization. *SIAM J. Numer. Anal.* **50** (2) (2012) 747–780. MR2914285 <https://doi.org/10.1137/110827041>
- [24] D. Pollard. *A User’s Guide to Measure Theoretic Probability. Cambridge Series in Statistical and Probabilistic Mathematics* **8**. Cambridge University Press, Cambridge, 2002. MR1873379
- [25] F. Santambrogio. *Optimal Transport for Applied Mathematicians. Progress in Nonlinear Differential Equations and Their Applications* **87**. Birkhäuser/Springer, 2015. MR3409718 <https://doi.org/10.1007/978-3-319-20828-2>
- [26] M. Shaked and J. G. Shanthikumar. *Stochastic Orders. Springer Series in Statistics*. Springer, New York, 2007. MR2265633 <https://doi.org/10.1007/978-0-387-34675-5>
- [27] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Stat.* **36** (1965) 423–439. MR0177430 <https://doi.org/10.1214/aoms/1177700153>