SAMPLING OF ONE-DIMENSIONAL PROBABILITY MEASURES IN THE CONVEX ORDER AND COMPUTATION OF ROBUST OPTION PRICE BOUNDS

AURÉLIEN ALFONSI, JACOPO CORBETTA AND BENJAMIN JOURDAIN

ABSTRACT. For μ and ν two probability measures on the real line such that μ is smaller than ν in the convex order, this property is in general not preserved at the level of 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 $(X_i)_{1\le i\le I}$ (resp. $(Y_j)_{1\le j\le J}$) are independent and identically distributed according to μ (resp. ν). We investigate modifications of μ_I (resp. ν_J) smaller than ν_J (resp. greater than μ_I) in the convex order and weakly converging to μ (resp. ν) as $I, J \to \infty$. According to Kertz and Rösler (1992), the set of probability measures on the real line with a finite first order moment is a complete lattice for the increasing and the decreasing convex orders. For μ and ν in this set, this enables us to define a probability measure $\mu\vee\nu$ (resp. $\mu\wedge\nu$) greater than μ (resp. smaller than ν) in the convex order. We give efficient algorithms permitting to compute $\mu\vee\nu$ and $\mu\wedge\nu$ (and therefore $\mu_I\vee\nu_J$ and $\mu_I\wedge\nu_J$) when μ and ν have finite supports. Last, we illustrate by numerical experiments the resulting sampling methods that preserve the convex order and their application to approximate martingale optimal transport problems and in particular to calculate robust option price bounds.

Keywords: *convex order, martingale optimal transport, robust option price bounds, sampling techniques, linear programming*
AMS Subject Classification (2010): *91G60, 90C08, 60G42.*

1. Introduction and motivations

We start with a practical motivation. Let us consider a trader who sells and hedges exotic options on an asset $(S_t)_{t\ge0}$. To calculate option prices, the trader typically picks her favourite model, then calibrates this model to vanilla options prices and uses this calibrated model $(\widetilde{S}_t)_{t\ge0}$ to calculate exotic option prices. Then, a natural question is to know what is the range of prices that can be attained by this procedure. To make things simple, we assume zero interest rates and consider a trader who wants to deal an exotic option that pays $c(S_{T_1}, S_{T_2})$ at time T_2 (with $T_1 < T_2$ and $c : \mathbb{R}\times\mathbb{R}\to\mathbb{R}$) and observes on the market European call option prices at times T_1 and T_2 for all strikes. This amounts to know exactly μ and ν, the respective probability laws of S_{T_1} and S_{T_2} under all risk-neutral measures, and we suppose that $\widetilde{S}_{T_1}$ and $\widetilde{S}_{T_2}$ are respectively distributed according to μ and ν (this is achieved by the local volatility model and extensions). Then, all the models that are

Date: December 21, 2018.
Université Paris-Est, Cermics (ENPC), INRIA, F-77455 Marne-la-Vallée, France,
e-mails : aurelien.alfonsi@enpc.fr, j.corbetta@zeliade.com, benjamin.jourdain@enpc.fr.
This research benefited from the support of the "Chaire Risques Financiers", Fondation du Risque, the French National Research Agency under the program ANR-12-BS01-0019 (STAB) and was completed after the hiring of Jacopo Corbetta by Zeliade Systems.

compatible with the absence of arbitrage are described by the set of martingale couplings

$$\Pi^M(\mu,\nu)=\left\{\pi\in\Pi(\mu,\nu):\forall x\in\mathbb{R},\ \int_{\mathbb{R}}|y|\pi_{Y|X}(x,dy)<\infty \text{ and } \int_{\mathbb{R}}y\pi_{Y|X}(x,dy)=x\right\},$$

where $\pi_{Y|X}$ denotes a Markov kernel such that $\pi(dx,dy)=\mu(dx)\pi_{Y|X}(x,dy)$. Notice that the joint law of $(\widetilde{S}_{T_1},\widetilde{S}_{T_2})$ belongs to $\Pi^M(\mu,\nu)$. The range of prices of the exotic option is given by the set

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

and we are interested in calculating the upper and lower bounds of this set, which are usually called the robust option price bounds. This problem is precisely the martingale optimal transport (MOT) problem studied by Beiglböck et al. (2013). They have shown that the upper (resp. lower) bound is also the cheapest (resp. most expensive) price of a superhedging (resp. subhedging) strategy from the dual formulation of the problem. In practice, the trader is interested in calculating these bounds and comparing them to her model price, in order to evaluate the model risk. Unless for particular payoff functions c, she has to use a Monte-Carlo method and approximate the exotic price in her model by $\frac{1}{I}\sum_{i=1}^I c(\widetilde{S}^i_{T_1},\widetilde{S}^i_{T_2})$, where $(\widetilde{S}^i)_{1\le i\le I}$ are independent copies of $\widetilde{S}$. Let us note that by construction, the empirical measures $\mu_I=\frac{1}{I}\sum_{i=1}^I\delta_{\widetilde{S}^i_{T_1}}$ and $\nu_I=\frac{1}{I}\sum_{i=1}^I\delta_{\widetilde{S}^i_{T_2}}$ approximate μ and ν. The goal of the paper is to give a way to calculate from μ_I and ν_I the robust price option bounds. Our approach is to consider the corresponding discrete martingale optimal transport problem and to rely on the powerful linear programming library that have been developed. To do so, we need to modify slightly the empirical measures μ_I and/or ν_I in order to recover the convex order, as we explain now.

We now present the mathematical framework and consider X and Y, two random variables on $\mathbb{R}$, with respective probability distributions μ and ν. The random variable X is said to be smaller than Y in the convex order if $\mathbb{E}[\phi(X)]\le\mathbb{E}[\phi(Y)]$ for every convex function $\phi:\mathbb{R}\to\mathbb{R}$, provided that both expectations exist. In this case, we use the notation $X\le_{\text{cx}}Y$ or $\mu\le_{\text{cx}}\nu$, since the convex order only involves the probability distributions. Theorem 8 in Strassen (1965) ensures that, when $\int_{\mathbb{R}}|y|\nu(dy)<\infty$, $\Pi^M(\mu,\nu)\neq\emptyset\iff\mu\le_{\text{cx}}\nu$. In this work, we consider the approximation of the probability measures μ and ν by probability measures with finite supports $\mu_I=\sum_{i=1}^I p_i\delta_{x_i}$ and $\nu_J=\sum_{j=1}^J q_j\delta_{y_j}$, with $I,J\in\mathbb{N}^*$, $x_i,y_j\in\mathbb{R}$ $p_i,q_j>0$ for any i,j and $\sum_{i=1}^I p_i=\sum_{j=1}^J q_j=1$. Note that we simply have $p_i=1/I$ and $q_j=1/J$ for empirical measures. We consider the approximation of the MOT problem: to minimize (or maximize)

$$\sum_{i=1}^I\sum_{j=1}^J r_{ij}c(x_i,y_j) \tag{1.1}$$

under the constraints

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

Thus, we can use a linear programming library to solve this problem numerically. The key issue to run these algorithms is the existence of such matrices $(r_{ij})_{1\le i\le I,1\le j\le J}$, that amounts to the existence of a martingale coupling between μ_I and ν_J. Otherwise, the set of matrices satisfying the constraints is void. By Strassen's theorem, this is equivalent to have

$\mu_I \leq_{\rm cx} \nu_J$, which motivates the interest of preserving the convex order when approximating μ and ν.

Up to now, we have dealt with only two measures $\mu \leq_{\rm cx} \nu$. More generally, if $\mu^1 \leq_{\rm cx} \ldots \leq_{\rm cx} \mu^\ell$ are ℓ probability measures, one may be interested in constructing approximations that preserve the convex order. In the same manner, such approximations would enable to tackle numerically multi-marginal martingale optimal transport problems, see again Beiglböck et al. (2013).

Up to our knowledge, few studies consider the problem of preserving the convex order while approximating two probability measures. We mention the thesis of Baker (2012) who proposes the following construction for $\mu \leq_{\rm cx} \nu$ that are integrable. 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 generalized inverses also called quantile functions. For $I \geq 1$, setting

$$\hat{\mu}_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{I\int_{\frac{i-1}{I}}^{\frac{i}{I}} F_\mu^{-1}(u)du} \text{ and } \hat{\nu}_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{I\int_{\frac{i-1}{I}}^{\frac{i}{I}} F_\nu^{-1}(u)du}, \tag{1.2}$$

one has by Theorem 2.4.11 in Baker (2012) that $\hat{\mu}_I \leq_{\rm cx} \hat{\nu}_I$ (this result is consequence of the characterization of the convex order in terms of the quantile functions that we recall below in Theorem 2.1 (v)). This construction is easy to implement and can be obviously generalized to the multi-marginal case, but it requires an explicit calculation of the integrals of the quantile functions. Notice that nevertheless, these integrals may be deduced by calibration if one observes for $i \in \{1, \ldots, I-1\}$ the price C_i^X of the European call option with strike $K_i^X = F_\mu^{-1}(i/I)$ written on the asset X as well as the current price $s_0 = \mathbb{E}[X]$ of this asset. Since $C_i^X = \mathbb{E}[(X - F_\mu^{-1}(i/I))^+] = \int_{\frac{i}{I}}^1 F_\mu^{-1}(p)dp - \left(1 - \frac{i}{I}\right) F_\mu^{-1}(i/I)$, we get $I\int_{\frac{i-1}{I}}^{\frac{i}{I}} F_\mu^{-1}(p)dp = I(C_{i-1}^X - C_i^X) + (I-i+1)K_{i-1}^X - (I-i)K_i^X$ (with the convention $C_I^X = 0$) for $i = 2, \ldots, I$ and $I\int_0^{\frac{1}{I}} F_\mu^{-1}(p)dp = I(s_0 - C_1^X) - (I-1)K_1^X$.

The dual quantization introduced by Pagès and Wilbertz (2012) gives another way to preserve the convex order in dimension one (see the remark after Proposition 10 in Pagès and Wilbertz (2012)). Suppose that μ and ν have a bounded support included in $[x_1, x_I]$ and that $x_1 < \cdots < x_I$. The dual quantization of μ is defined by

$$\hat{\mu}_I = \mu(\{x_1\})\delta_{x_1} + \sum_{i=1}^{I-1} \int_{(x_i, x_{i+1}]} \frac{x - x_i}{x_{i+1} - x_i}\mu(dx)\delta_{x_{i+1}} + \int_{(x_i, x_{i+1}]} \frac{x_{i+1} - x}{x_{i+1} - x_i}\mu(dx)\delta_{x_i}.$$

This is the law $\hat{X} = X1_{X=x_1} + \sum_{i=1}^{I-1} 1_{X\in(x_i,x_{i+1}]}\left(x_{i+1}1_{U\leq\frac{X-x_i}{x_{i+1}-x_i}} + x_i 1_{U>\frac{X-x_i}{x_{i+1}-x_i}}\right)$, where U is sampled independently from X according to the uniform distribution on $[0,1]$. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a convex function. We define $\hat{\phi} : [x_1, x_I] \to \mathbb{R}$ by $\hat{\phi}(x) = \frac{x-x_i}{x_{i+1}-x_i}\phi(x_{i+1}) + \frac{x_{i+1}-x}{x_{i+1}-x_i}\phi(x_i)$ for $x \in [x_i, x_{i+1}]$, $i = 1, \ldots, I-1$. This is a convex function, and we have by using $\mu \leq_{\rm cx} \nu$ that

$$\mathbb{E}[\phi(\hat{X})] = \mathbb{E}[\hat{\phi}(X)] \leq \mathbb{E}[\hat{\phi}(Y)] = \mathbb{E}[\phi(\hat{Y})].$$

Note that both methods have serious drawbacks to be used in practice. Baker's construction requires to calculate explicitly the integrals of the quantile functions, which is often not

possible. Dual quantization assumes probability distributions with bounded support, which is not satisfied by most of the asset models. On the other side, there are many recent works that deals with robust option price bounds for specific options. Without being exhaustive, Hobson and Neuberger (2012), Hobson and Klimmek (2015) characterize the bounds for forward start options. Henry-Labordère and Touzi (2016) describe explicitly the bounds for an abstract class of payoff functions c satisfying the so-called Spence-Mirrlees condition, for which the left and right curtain couplings introduced by Beiglböck and Juillet (2016) are optimal. These works crucially exploits particular properties of the payoff functions. In contrast, the numerical method that we propose in this paper to approximate robust option price bounds works for general payoff functions.

The paper is structured as follows. In Section 2, we recall the characterization of the convex, increasing convex and decreasing convex orders in dimension one, and give a way to test if two probability measures with finite support are in the convex order. Then, in Section 3, we shift the approximating measures μ_I and ν_J so that they have the same mean and show that they are then in the convex order for I, J large enough, under suitable assumptions. However, this method is not reliable in practice since the dimension $I \times J$ of the linear programming problem (1.1) can then be too large for this problem to be solved quickly. It is thus crucial to use non asymptotic methods that gives approximating measures in the convex order for any $I, J \in \mathbb{N}^*$. In Section 4, we propose such a method based on the following result stated in Kertz and Rösler (1992): the set of probability measures with a finite first order moment is a lattice for the increasing and the decreasing convex orders. Given μ and ν in this set, we define $\mu \vee \nu$ (resp. $\mu \wedge \nu$) as the supremum (resp. infimum) of μ and ν for the decreasing convex order when $\int_{\mathbb{R}} x\mu(dx) \leq \int_{\mathbb{R}} x\nu(dx)$ and for the increasing convex order otherwise. This way, $\mu \vee \nu$ (resp. $\mu \wedge \nu$) is greater then μ (resp. smaller than ν) in the convex order. In Section 5, we give efficient algorithms permitting to compute $\mu \vee \nu$ and $\mu \wedge \nu$ (and therefore $\mu_I \vee \nu_J$ and $\mu_I \wedge \nu_J$) when μ and ν are convex combinations of Dirac masses. We prove that, when $\mu \leq_{\text{cx}} \nu$, $\mu_I \vee \nu_J$ and $\mu_I \wedge \nu_J$ respectively converge weakly to ν and μ as $I, J \to \infty$, which shows that either μ_I and $\mu_I \vee \nu_J$ or $\mu_I \wedge \nu_J$ and ν_J respectively approximate μ and ν and are in the convex order. Section 6 presents the numerical approximation of different MOT problems. We first consider academic examples where the MOT is known explicitly. Then, we address some practical examples with two or three marginals, when the marginal laws are the ones given by the Black and Scholes model.

This paper only deals with the one-dimensional problem. Of course, the same problem occurs in higher dimension when one consider $d \geq 2$ assets instead of one. Thus, for two probability measures μ and ν on $\mathbb{R}^d$ such that $\mu \leq_{\text{cx}} \nu$, and for discrete probability measures μ_I and ν_J that respectively approximate μ and ν, we want to slightly modify either μ_I or ν_J to recover the convex order between the approximating measures. Unfortunately, we can no longer rely on the lattice structure to construct such a modification since it has been shown by Müller and Scarsini (2006) that the set of probability measures with a given mean is not a lattice for the convex order when $d \geq 2$. In a companion paper Alfonsi et al. (2018), we propose a suitable modification of μ_I that is obtained by the mean of a projection for the Wasserstein distance and that can be efficiently computed.

2. Characterization of the convex, decreasing convex and increasing orders

Let $\mathcal{P}(\mathbb{R})$ denote the set of probability measures on the real line. We define

$$\mathcal{P}_-(\mathbb{R}) = \{\mu \in \mathcal{P}(\mathbb{R}) : \int_{\mathbb{R}} x^- \mu(dx) < \infty\},$$
$$\mathcal{P}_+(\mathbb{R}) = \{\mu \in \mathcal{P}(\mathbb{R}) : \int_{\mathbb{R}} x^+ \mu(dx) < \infty\},$$
$$\mathcal{P}_1(\mathbb{R}) = \{\mu \in \mathcal{P}(\mathbb{R}) : \int_{\mathbb{R}} |x| \mu(dx) < \infty\} = \mathcal{P}_-(\mathbb{R}) \cap \mathcal{P}_+(\mathbb{R}).$$

For $\mu \in \mathcal{P}(\mathbb{R})$, we set $F_\mu(x) = \mu((-\infty, x])$, $F_\mu(x-) = \lim_{y \to x-} F_\mu(y) = \mu((-\infty, x))$. For $t \in \mathbb{R}$, we define $\varphi_\mu(t) = \int_{-\infty}^t F_\mu(x)dx$. Fubini's theorem ensures that for $t \in \mathbb{R}$,

$$\varphi_\mu(t) = \int_{\mathbb{R}} 1_{\{x \le t\}} \int_{\mathbb{R}} 1_{\{y \le x\}} \mu(dy) = \int_{\mathbb{R}} (t-y)^+ \mu(dy) \tag{2.1}$$

where the right-hand side is finite iff $\mu \in \mathcal{P}_-(\mathbb{R})$. In a symmetric way, we define $\bar{F}_\mu(x) = \mu([x, +\infty))$, $\bar{\varphi}_\mu(t) = \int_t^{+\infty} \bar{F}_\mu(x)dx$ and notice that

$$\bar{\varphi}_\mu(t) = \int_t^{+\infty} \mu((x, +\infty))dx = \int_{\mathbb{R}} (y-t)^+ \mu(dy) \tag{2.2}$$

where the right-hand side is finite iff $\mu \in \mathcal{P}_+(\mathbb{R})$. Let us note that $\mu((x, +\infty)) = \mu([x, +\infty))$ dx-a.e., and equation (2.2) is written with the open interval to have the same convention as Kertz and Rösler (2000). Denoting by $\bar{\mu}$ the image of μ by $x \mapsto -x$, one has

$$\forall x \in \mathbb{R},\ F_{\bar{\mu}}(x) = \bar{\mu}((-\infty, x]] = \mu([-x, +\infty)) = \bar{F}_\mu(-x) \text{ and thus } \forall t \in \mathbb{R},\ \bar{\varphi}_\mu(t) = \varphi_{\bar{\mu}}(-t). \tag{2.3}$$

Last, we define the function

$$\pi_\mu(t) = \int_{\mathbb{R}} |t-x| \mu(dx) = \varphi_\mu(t) + \bar{\varphi}_\mu(t),$$

that is usually called the potential of μ. It is finite when $\mu \in \mathcal{P}_1(\mathbb{R})$. Since $t - \int_{\mathbb{R}} x\mu(dx) = \varphi_\mu(t) - \bar{\varphi}_\mu(t) = 2\varphi_\mu(t) - \pi_\mu(t) = \pi_\mu(t) - 2\bar{\varphi}_\mu(t)$, we have

$$\pi_\mu(t) = 2\varphi_\mu(t) - t + \int_{\mathbb{R}} x\mu(dx) = 2\bar{\varphi}_\mu(t) + t - \int_{\mathbb{R}} x\mu(dx).$$

One important particularity of the dimension 1 is the following result (see e.g. Theorems 3.A.1., 3.A.2. and 3.A.5. in Shaked and Shanthikumar (2007)).

Theorem 2.1. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$. The following conditions are equivalent:*

(i) $\mu \le_{\text{cx}} \nu$,
(ii) $\int_{\mathbb{R}} x\mu(dx) = \int_{\mathbb{R}} x\nu(dx)$ *and* $\forall t \in \mathbb{R}, \pi_\mu(t) \le \pi_\nu(t)$,
(iii) $\int_{\mathbb{R}} x\mu(dx) = \int_{\mathbb{R}} x\nu(dx)$ *and* $\forall t \in \mathbb{R}, \varphi_\mu(t) \le \varphi_\nu(t)$,
(iv) $\int_{\mathbb{R}} x\mu(dx) = \int_{\mathbb{R}} x\nu(dx)$ *and* $\forall t \in \mathbb{R}, \bar{\varphi}_\mu(t) \le \bar{\varphi}_\nu(t)$,
(v) $\int_0^1 F_\mu^{-1}(p)dp = \int_0^1 F_\nu^{-1}(p)dp$ *and* $\forall q \in (0,1), \int_q^1 F_\mu^{-1}(p)dp \le \int_q^1 F_\nu^{-1}(p)dp$.

Thus, to check if two probability measures are in the convex order, it is sufficient to focus on the the family of convex functions $\phi(x) = |t-x|$, for $t \in \mathbb{R}$. Thanks to this result,

we obtain an interesting corollary that gives a necessary and sufficient condition for two probability measures with finite support to be in the convex order.

Corollary 2.2. *Let $\mu = \sum_{i=1}^I p_i \delta_{x_i}$ and $\nu = \sum_{j=1}^J q_j \delta_{y_j}$ be two probability measures on $\mathbb{R}$. Without loss of generality, we assume that $x_1 < \cdots < x_I$, $y_1 < \cdots < y_J$ and $p_1 p_I q_1 q_J > 0$. Then we have $\mu \leq_{\rm cx} \nu$ if, and only if*

(i) $y_1 \leq x_1$ *and* $y_J \geq x_I$,

(ii) for all j *such that* $x_1 \leq y_j \leq x_I$, $\pi_\mu(y_j) \leq \pi_\nu(y_j)$,

(iii) $\sum_{i=1}^I p_i x_i = \sum_{j=1}^J q_j y_j$.

By using (iii) and the link between π_μ and φ_μ, this corollary is still true if we replace (ii) by one of the following conditions:

(ii') for all j such that $x_1 \leq y_j \leq x_I$, $\varphi_\mu(y_j) \leq \varphi_\nu(y_j)$,

(ii'') for all j such that $x_1 \leq y_j \leq x_I$, $\bar\varphi_\mu(y_j) \leq \bar\varphi_\nu(y_j)$.

Proof. The necessary condition is obvious. Let us check the sufficient condition. Since $y_1 = \min_{1\leq j\leq J} y_j$ and $y_J = \max_{1\leq j\leq J} y_j$, we obtain

$$\forall t \leq y_1,\ \pi_\nu(t) = \sum_{j=1}^J q_j y_j - t \text{ and } \forall t \geq y_M,\ \pi_\nu(t) = t - \sum_{j=1}^J q_j y_j.$$

Similarly, we have $\pi_\mu(t) = \sum_{i=1}^I p_i x_i - t$ for $t \leq x_1$ and $\pi_\mu(t) = t - \sum_{i=1}^I p_i x_i$ for $t \geq x_I$. By (i) and (iii), we get $\pi_\mu(t) = \pi_\nu(t)$ for $t \in (-\infty, y_1] \cup [y_M, +\infty)$ and

$$\pi_\mu(t) = \left| t - \sum_{i=1}^I p_i x_i \right| = \left| \sum_{j=1}^J q_j (t - y_j) \right| \leq \sum_{j=1}^J q_j |t - y_j| = \pi_\nu(t)$$

for $t \in (-\infty, x_1] \cup [x_M, +\infty)$ by Jensen's inequality. Moreover, since π_μ is convex and π_ν is affine on $[y_j; y_{j+1}]$ for every $j = 1, \ldots, M-1$, we get from (ii) that $\pi_\mu(t) \leq \pi_\nu(t)$ for every $t \in [y_j; y_{j+1}]$ for every $j = 1, \ldots, M-1$. Thus, we have $\pi_\mu(t) \leq \pi_\nu(t)$ for all $t \in \mathbb{R}$. □

In dimension one, the increasing and decreasing convex orders may be defined as follows.

Definition 2.3. *For $\mu, \nu \in \mathcal{P}(\mathbb{R})$, we say that μ is smaller than ν in the increasing (resp. decreasing) convex order and denote $\mu \leq_{\rm icx} \nu$ (resp. $\mu \leq_{\rm dcx} \nu$) if $\int_{\mathbb{R}} \phi(x)\mu(dx) \leq \int_{\mathbb{R}} \phi(x)\nu(dx)$ for each increasing (resp. decreasing) convex function $\phi : \mathbb{R} \to \mathbb{R}$ such that the integrals make sense.*

Lemmas 2.2 and 2.5 of Kertz and Rösler (2000) give the following characterization of the increasing and decreasing convex orders.

Theorem 2.4. *Let $\mu, \nu \in \mathcal{P}_+(\mathbb{R})$ (resp. $\mathcal{P}_-(\mathbb{R})$). The following statements are equivalent :*

(i) $\mu \leq_{\rm icx} \nu$ *(resp.* $\mu \leq_{\rm dcx} \nu$*),*

(ii) $\forall t \in \mathbb{R}$, $\bar\varphi_\mu(t) \leq \bar\varphi_\nu(t)$ *(resp.* $\varphi_\mu(t) \leq \varphi_\nu(t)$*),*

(iii) $\forall q \in [0,1]$, $\int_q^1 F_\mu^{-1}(p)dp \leq \int_q^1 F_\nu^{-1}(p)dp$ *(resp.* $\int_0^q F_\mu^{-1}(p)dp \geq \int_0^q F_\nu^{-1}(p)dp$*),*

(iv) $\bar\mu \leq_{\rm dcx} \bar\nu$ *(resp.* $\bar\mu \leq_{\rm icx} \bar\nu$*).*

Note that the equivalence between (ii) and (iii) is a direct consequence that φ_μ is the Legendre transform of the convex function $q \mapsto \int_0^q F_\mu^{-1}(p)dp$, and conversely, see e.g. Lemma A.22 of Föllmer and Schied (2011).

3. An asymptotic approach

Now, we turn to our practical problem. We assume that $\mu \leq_{\mathrm{cx}} \nu$ with $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$. We consider two i.i.d. samples $X_1, \dots, X_I$ with distribution μ and $Y_1, \dots, Y_J$ with distribution ν, and we set

$$\hat{\mu}_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{X_i} \text{ and } \hat{\nu}_J = \frac{1}{J}\sum_{j=1}^{J} \delta_{X_j}.$$

At least when μ and ν have densities, the empirical means $\bar{X}_I = \frac{1}{I}\sum_{i=1}^{I} X_i$ and $\bar{Y}_J = \frac{1}{J}\sum_{j=1}^{J} Y_j$ are almost surely distinct, and we cannot have $\hat{\mu}_I \leq_{\mathrm{cx}} \hat{\nu}_J$. However, we know that when $I, J \to +\infty$, both empirical means converge almost surely to $\int_{\mathbb{R}} x\mu(dx)$. A natural idea is then to slightly modify $\hat{\mu}_I$, or $\hat{\nu}_J$, or both empirical measures so that they both have the same mean. The hope is then that for I and J large enough, the modified empirical measures would be naturally in the convex order. In this paragraph, we will take

$$\widetilde{\mu}_{I,J} = \frac{1}{I}\sum_{i=1}^{I} \delta_{X_i + \bar{Y}_J - \bar{X}_I}, \tag{3.1}$$

and keep $\hat{\nu}_J$. Other choices are of course possible and some of them are presented in Subsection 4.2.

For a probability measure μ on the real line, let $F_\mu^{-1}(0+)$ and $F_\mu^{-1}(1-)$ respectively denote the left-hand and the right-hand limits of the function F_μ^{-1} as $p \to 0$ and $p \to 1$. We have the following result.

Proposition 3.1. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \leq_{\mathrm{cx}} \nu$ and $F_\nu^{-1}(0+) < F_\mu^{-1}(0+) \leq F_\mu^{-1}(1-) < F_\nu^{-1}(1-)$. The probability measure $\widetilde{\mu}_{I,J}$ defined by (3.1) (resp. $\hat{\nu}_J$) converges weakly to μ (resp. ν), almost surely when $I, J \to +\infty$. Suppose moreover that there exists $\varepsilon > 0$ such that for every $t \in [F_\mu^{-1}(0+), F_\mu^{-1}(1-)]$*

$$\int_{-\infty}^{t} (F_\nu(x) - F_\mu(x))dx \geq \varepsilon. \tag{3.2}$$

Then, almost surely, there exists M such that for all $I, J \geq M$,

$$\widetilde{\mu}_{I,J} \leq_{\mathrm{cx}} \hat{\nu}_J. \tag{3.3}$$

Proof. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous bounded function. The strong law of large number gives the (almost sure) weak convergence of $\hat{\mu}_I$ (resp. $\hat{\nu}_J$) towards μ (resp. ν), as well as the almost sure convergence of $\bar{Y}_J - \bar{X}_I$ towards 0 as $I, J \to \infty$. Now, we use that f is uniformly continuous on $[F_\mu^{-1}(0+) - 1, F_\mu^{-1}(1-) + 1]$ to obtain that

$$\frac{1}{I}\sum_{i=1}^{I} f(X_i + \bar{Y}_J - \bar{X}_I) \underset{I,J\to+\infty}{\to} \int f(x)\mu(dx)$$

almost surely. Thus, $\widetilde{\mu}_{I,J}$ converges weakly to μ, almost surely.

We now turn to the convex order. Let $\delta > 0$ be such that

$$\forall t \in [F_\mu^{-1}(0+) - \delta, F_\mu^{-1}(1-) + \delta], \ \int_{-\infty}^{t} (F_\nu(x) - F_\mu(x))dx \geq \varepsilon/2.$$

We take $\underline{t} \in (-\infty, F_\mu^{-1}(0+) - \delta)$ such that $\int_{-\infty}^{\underline{t}} F_\nu(x)dx < \frac{\varepsilon}{6}$, which implies that

$$\forall t \in [F_\mu^{-1}(0+) - \delta, F_\mu^{-1}(1-) + \delta], \ \int_{\underline{t}}^{t} (F_\nu(x) - F_\mu(x))dx \geq \frac{\varepsilon}{3}.$$

From the weak convergence of $\hat{\nu}_J$ (resp. $\widetilde{\mu}_{I,J}$) to ν (resp. μ), we obtain that dx-a.e., $F_{\hat{\nu}_J}(x) \to F_\nu(x)$ (resp. $F_{\widetilde{\mu}_{I,J}}(x) \to F_\mu(x)$). Then, the dominated convergence theorem gives the pointwise convergence of $\int_{\underline{t}}^t F_{\hat{\nu}_J}(x)dx$ (resp. $\int_{\underline{t}}^t F_{\widetilde{\mu}_{I,J}}(x)dx$) towards $\int_{\underline{t}}^t F_\nu(x)dx$ (resp. $\int_{\underline{t}}^t F_\mu(x)dx$). Since these functions are nondecreasing and the limit is continuous with respect to t, we obtain the uniform convergence on each compact set, and deduce that, almost surely, there exists M such that

$$\forall I, J \geq M,\ \forall t \in [F_\mu^{-1}(0+) - \delta, F_\mu^{-1}(1-) + \delta],\ \int_{\underline{t}}^t (F_{\hat{\nu}_J}(x) - F_{\widetilde{\mu}_{I,J}}(x))dx \geq \frac{\varepsilon}{4}.$$

Taking also M large enough so that for $I, J \geq M$, $|\bar{Y}_J - \bar{X}_I| < \delta$, we have $F_{\widetilde{\mu}_{I,J}}(x) = 0$ and $F_{\hat{\nu}_J}(x) - F_{\widetilde{\mu}_{I,J}}(x) \geq 0$ for $x \leq F_\mu^{-1}(0+) - \delta$ and deduce that

$$\forall t \in (-\infty, F_\mu^{-1}(1-) + \delta],\ \int_{-\infty}^t (F_{\hat{\nu}_J}(x) - F_{\widetilde{\mu}_{I,J}}(x))dx \geq 0.$$

Last, for $I, J \geq M$, $F_{\widetilde{\mu}_{I,J}}(x) = 1$ for $x \geq F_\mu^{-1}(1-) + \delta$ and since $\lim_{t\to+\infty} \varphi_{\hat{\nu}_J}(t) - \varphi_{\widetilde{\mu}_{I,J}}(t) = 0$, we get that for $t \geq F_\mu^{-1}(1-) + \delta$, $\int_{-\infty}^t (F_{\hat{\nu}_J}(x) - F_{\widetilde{\mu}_{I,J}}(x))dx = \int_t^{+\infty} (1 - F_{\hat{\nu}_J}(x))dx \geq 0$. We conclude by using Theorem 2.1 (iii). □

We now give a sufficient condition for (3.2).

Lemma 3.2. *Let μ, ν be two probability measures on the real line with $\mu \leq_{\mathrm{cx}} \nu$ and $F_\nu^{-1}(0+) < F_\mu^{-1}(0+) \leq F_\mu^{-1}(1-) < F_\nu^{-1}(1-)$. Suppose that*

$$\exists x_0 \in \mathbb{R}, \forall x \in (-\infty, x_0], F_\nu(x) - F_\mu(x) \geq 0 \text{ and } \forall x \in [x_0, +\infty), F_\nu(x) - F_\mu(x) \leq 0.$$

Then, (3.2) *holds.*

Proof. We know that $F_\nu(x) - F_\mu(x) = 0$ when $x \notin [F_\nu^{-1}(0+), F_\nu^{-1}(1-))$, $F_\nu(x) - F_\mu(x) = F_\nu(x) > 0$ when $x \in (F_\nu^{-1}(0+), F_\mu^{-1}(0+))$, $F_\nu(x) - F_\mu(x) = F_\nu(x) - 1 < 0$ when $x \in [F_\mu^{-1}(1-), F_\nu^{-1}(1-))$, and by the equality of the means we know $\int_{-\infty}^{F_\nu^{-1}(1-)} (F_\nu(x) - F_\mu(x))dx = 0$. In particular, we necessarily have $x_0 \in [F_\mu^{-1}(0+), F_\mu^{-1}(1-))$ and we get that $t \mapsto \int_{-\infty}^t (F_\nu(x) - F_\mu(x))dx$ is nondecreasing on $(\infty, x_0]$, nonincreasing on $[x_0, \infty)$, increasing on $[F_\nu^{-1}(0+), F_\mu^{-1}(0+)]$, decreasing on $[F_\mu^{-1}(1-), F_\nu^{-1}(1-)]$. Thus $\int_{-\infty}^{F_\mu^{-1}(0+)} (F_\nu(x) - F_\mu(x))dx > 0$, $\int_{-\infty}^{F_\mu^{-1}(1-)} (F_\nu(x) - F_\mu(x))dx > 0$ and (3.2) holds for

$$\varepsilon = \min\left(\int_{-\infty}^{F_\mu^{-1}(0+)} (F_\nu(x) - F_\mu(x))dx > 0, \int_{-\infty}^{F_\mu^{-1}(1-)} (F_\nu(x) - F_\mu(x))dx \right).$$

□

In particular, we see that under the assumptions of Lemma 3.2, $\varphi_\nu(t) > \varphi_\mu(t)$ for any $t \in (F_\nu^{-1}(0+), F_\nu^{-1}(1-))$, which means that (μ, ν) is irreducible, see Definition A.3 of Beiglböck and Juillet (2016).

Proposition 3.1 gives a framework under which the empirical measures, up to a modification that equalizes their means, are asymptotically in the convex order. The assumptions of Proposition 3.1 are quite restrictive and could be refined. Nonetheless, it is not very convenient in practice to reach the convex order only asymptotically. If we have in mind to solve the discrete MOT problem (1.1), which is a linear programming problem in dimension $I \times J$. To fix the ideas, if we need $I, J \gtrsim 10^3$ for the modified empirical measures to be in

the convex order, the resolution of the discrete MOT problem (1.1) is already too greedy in time and memory. Thus, it would be much more convenient if we could guarantee that the approximating measure are in the convex order. This is why we prefer to focus on non asymptotic methods that ensures the convex order for any I, J.

4. A non asymptotic approach based on the lattice structure

4.1. **The lattice structure for the increasing and decreasing convex orders.** It has been observed by Kertz and Rösler that for any $a \in \mathbb{R}$,

$$\left\{\eta \in \mathcal{P}_1(\mathbb{R}) \text{ such that } \int_{\mathbb{R}} x\eta(dx) = a\right\}$$

is a lattice for the convex order (Proposition 1.6 of Kertz and Rösler (1992)) and even a complete lattice (see top of p162 of Kertz and Rösler (2000)). By Proposition 4.5 of Müller and Scarsini (2006), this property is no longer true in higher dimension. Thus, if $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ share the same expectation, there is a unique probability measure $\mu \wedge_{\rm cx} \nu \in \mathcal{P}_1(\mathbb{R})$ (resp. $\mu \vee_{\rm cx} \nu \in \mathcal{P}_1(\mathbb{R})$) such that $\mu \wedge_{\rm cx} \nu \leq_{\rm cx} \mu$, $\mu \wedge_{\rm cx} \nu \leq_{\rm cx} \nu$ (resp. $\mu \leq_{\rm cx} \mu \vee_{\rm cx} \nu$, $\mu \leq_{\rm cx} \mu \vee_{\rm cx} \nu$) and $\eta \leq_{\rm cx} \mu \wedge_{\rm cx} \nu$ for all $\eta \in \mathcal{P}_1(\mathbb{R})$ such that $\eta \leq_{\rm cx} \mu$ and $\eta \leq_{\rm cx} \nu$ (resp. $\mu \vee_{\rm cx} \nu \leq_{\rm cx} \eta$ for all $\eta \in \mathcal{P}_1(\mathbb{R})$ such that $\mu \leq_{\rm cx} \eta$ and $\nu \leq_{\rm cx} \eta$). By Proposition 1.6 of Kertz and Rösler (1992), $\bar\varphi_{\mu\wedge_{\rm cx}\nu}$ is the greatest convex function below $\bar\varphi_\mu \wedge \bar\varphi_\nu$ and $\bar\varphi_{\mu\vee_{\rm cx}\nu} = \bar\varphi_\mu \vee \bar\varphi_\nu$.

To deal with the case when $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ are such that $\int_{\mathbb{R}} x\mu(dx) \neq \int_{\mathbb{R}} x\nu(dx)$, we are going to use the complete lattice property of $\mathcal{P}_+(\mathbb{R})$ (resp. $\mathcal{P}_-(\mathbb{R})$) endowed with the increasing (resp. decreasing) convex order stated in Theorem 3.4 of Kertz and Rösler (2000) (resp. Theorem 3.7 (a) of Kertz and Rösler (2000)). For $\mu, \nu \in \mathcal{P}_+(\mathbb{R})$ (resp. $\mathcal{P}_-(\mathbb{R})$), let $\mu \wedge_{\rm icx} \nu, \mu \vee_{\rm icx} \nu \in \mathcal{P}_+(\mathbb{R})$ (resp. $\mu \wedge_{\rm dcx} \nu, \mu \vee_{\rm dcx} \nu \in \mathcal{P}_-(\mathbb{R})$) satisfy the above properties of $\mu \wedge_{\rm cx} \nu, \mu \vee_{\rm cx} \nu$ but with $\leq_{\rm cx}$ replaced by $\leq_{\rm icx}$ (resp. $\leq_{\rm dcx}$). By Proposition 1.4 of Kertz and Rösler (1992), for $\mu, \nu \in \mathcal{P}_+(\mathbb{R})$, $\bar\varphi_{\mu\wedge_{\rm icx}\nu}$ is the greatest convex function below $\bar\varphi_\mu \wedge \bar\varphi_\nu$ and $\bar\varphi_{\mu\vee_{\rm icx}\nu} = \bar\varphi_\mu \vee \bar\varphi_\nu$. By Theorem 3.7 (a) of Kertz and Rösler (2000), for $\mu, \nu \in \mathcal{P}_-(\mathbb{R})$, $\mu \wedge_{\rm dcx} \nu = \overline{\bar\mu \wedge_{\rm icx} \bar\nu}$ and $\mu \vee_{\rm dcx} \nu = \overline{\bar\mu \vee_{\rm icx} \bar\nu}$, so that by (2.3), $\varphi_{\mu\wedge_{\rm dcx}\nu}$ is the greatest convex function below $\varphi_\mu \wedge \varphi_\nu$ and $\varphi_{\mu\vee_{\rm dcx}\nu} = \varphi_\mu \vee \varphi_\nu$. Let us now compute the expectations of $\mu \wedge_{\rm icx} \nu, \mu \vee_{\rm icx} \nu, \mu \wedge_{\rm dcx} \nu$ and $\mu \vee_{\rm dcx} \nu$ when $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$.

Lemma 4.1. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$. Then $\mu \wedge_{\rm icx} \nu, \mu \vee_{\rm icx} \nu, \mu \wedge_{\rm dcx} \nu, \mu \vee_{\rm dcx} \nu \in \mathcal{P}_1(\mathbb{R})$ and*

$$\int_{\mathbb{R}} x\mu \vee_{\rm icx} \nu(dx) = \int_{\mathbb{R}} x\mu(dx) \vee \int_{\mathbb{R}} x\nu(dx), \quad \int_{\mathbb{R}} x\mu \vee_{\rm dcx} \nu(dx) = \int_{\mathbb{R}} x\mu(dx) \wedge \int_{\mathbb{R}} x\nu(dx),$$
$$\int_{\mathbb{R}} x\mu \wedge_{\rm icx} \nu(dx) = \int_{\mathbb{R}} x\mu(dx) \wedge \int_{\mathbb{R}} x\nu(dx), \quad \int_{\mathbb{R}} x\mu \wedge_{\rm dcx} \nu(dx) = \int_{\mathbb{R}} x\mu(dx) \vee \int_{\mathbb{R}} x\nu(dx).$$

Proof. By (2.1), we have $t - \varphi_\mu(t) = \int_{\mathbb{R}} x\mu(dx) - \int_{\mathbb{R}} (x-t)^+\mu(dx)$, where the last term tends to 0 as $t \to +\infty$ by Lebesgue theorem. Since, in the same way, $t - \varphi_\nu(t)$ tends to $\int_{\mathbb{R}} x\nu(dx)$, we deduce that $t - \varphi_\mu \vee \varphi_\nu(t)$ tends to $\int_{\mathbb{R}} x\mu(dx) \wedge \int_{\mathbb{R}} x\nu(dx)$. On the other hand, by (2.1) with μ replaced by $\mu \vee_{\rm dcx} \nu$,

$$t - \varphi_\mu \vee \varphi_\nu(t) = t - \int_{\mathbb{R}} (t-x)^+\mu \vee_{\rm dcx} \nu(dx) = \int_{\mathbb{R}} (t \wedge x)\mu \vee_{\rm dcx} \nu(dx)$$

For $t \geq 0$, the right-hand side is equal to $-\int_{\mathbb{R}} x^-\mu \vee_{\rm dcx} \nu(dx) + \int_{\mathbb{R}} (x^+ \wedge t)\mu \vee_{\rm dcx} \nu(dx)$, where the second term converges to $\int_{\mathbb{R}} x^+\mu \vee_{\rm dcx} \nu(dx)$ as $t \to \infty$ by monotone convergence. Therefore $\int_{\mathbb{R}} |x|\mu \vee_{\rm dcx} \nu(dx) < \infty$ and $\int_{\mathbb{R}} x\mu \vee_{\rm dcx} \nu(dx) = \int_{\mathbb{R}} x\mu(dx) \wedge \int_{\mathbb{R}} x\nu(dx)$. The fact that $\mu \vee_{\rm icx} \nu \in \mathcal{P}_1(\mathbb{R})$ and the formula giving its expectation are deduced from the equality

$\mu \vee_{\mathrm{icx}} \nu = \overline{\bar{\mu} \vee_{\mathrm{dcx}} \bar{\nu}}$. For $\mu \wedge_{\mathrm{icx}} \nu$ and $\mu \wedge_{\mathrm{dcx}} \nu$, this is an easy consequence of the inverse transform sampling method combined with the more convenient characterization of these probability measures based on the quantile functions stated in Lemma 4.2 just below. □

Lemma 4.2.

$$\textit{For } \mu,\nu \in \mathcal{P}_+(\mathbb{R}),\ \forall q \in [0,1],\ \int_q^1 F^{-1}_{\mu\wedge_{\mathrm{icx}}\nu}(p)dp = \int_q^1 F^{-1}_\mu(p)dp \wedge \int_q^1 F^{-1}_\nu(p)dp,$$

$$\textit{and for } \mu,\nu \in \mathcal{P}_-(\mathbb{R}),\ \forall q \in [0,1],\ \int_0^q F^{-1}_{\mu\wedge_{\mathrm{dcx}}\nu}(p)dp = \int_0^q F^{-1}_\mu(p)dp \vee \int_0^q F^{-1}_\nu(p)dp.$$

Moreover, for all $p \in (0,1)$, $F^{-1}_{\mu\wedge_{\mathrm{icx}}\nu}(p), F^{-1}_{\mu\wedge_{\mathrm{dcx}}\nu}(p) \in \{F^{-1}_\mu(p), F^{-1}_\nu(p)\}$.

Proof. The first statement is deduced from the second one and the equality $\mu \wedge_{\mathrm{dcx}} \nu = \overline{\bar{\mu} \wedge_{\mathrm{icx}} \bar{\nu}}$. For $f : \mathbb{R} \to (-\infty, +\infty]$, we define $f^*(y) = \sup_{x\in\mathbb{R}} xy - f(x)$ the Fenchel-Legendre transform of f and recall that $f^{**} := (f^*)^*$ is the greatest convex function below f. Thus, we have $\varphi_{\mu\wedge_{\mathrm{dcx}}\nu} = (\varphi_\mu \wedge \varphi_\nu)^{**}$. This gives $\varphi^*_{\mu\wedge_{\mathrm{dcx}}\nu} = (\varphi_\mu \wedge \varphi_\nu)^*$. By using a standard property of the Fenchel-Legendre transform, we have $(\varphi_\mu \wedge \varphi_\nu)^* = \varphi^*_\mu \vee \varphi^*_\nu$, and we deduce that $\int_0^q F^{-1}_{\mu\wedge_{\mathrm{dcx}}\nu}(p)dp = \int_0^q F^{-1}_\mu(p)dp \vee \int_0^q F^{-1}_\nu(p)dp$ for all $q \in [0,1]$ by Lemma A.23 of Föllmer and Schied (2011).

Let us check now that $F^{-1}_{\mu\wedge_{\mathrm{icx}}\nu}(p) \in \{F^{-1}_\mu(p), F^{-1}_\nu(p)\}$ for $p \in (0,1)$. The function

$$q \mapsto g(q) := \int_q^1 F^{-1}_\mu(p)dp - \int_q^1 F^{-1}_\nu(p)dp$$

is locally bounded with a locally bounded derivative on $(0,1)$ so that the distribution derivative of $g^+(q)$ is equal to $1_{\{g(q)>0\}}(F^{-1}_\nu(q) - F^{-1}_\mu(q))$ and the one of

$$q \mapsto \int_q^1 F^{-1}_{\mu\wedge_{\mathrm{icx}}\nu}(p)dp = \int_q^1 F^{-1}_\mu(p)dp - g^+(q)$$

is equal to $-1_{\{g(q)\le 0\}}F^{-1}_\mu(q) - 1_{\{g(q)>0\}}F^{-1}_\nu(q)$. Therefore dq a.e. on $(0,1)$, $F^{-1}_{\mu\wedge_{\mathrm{icx}}\nu}(q) = 1_{\{g(q)\le 0\}}F^{-1}_\mu(q) + 1_{\{g(q)>0\}}F^{-1}_\nu(q)$. Let now $p \in (0,1)$. There is a sequence $(q_n)_{n\in\mathbb{N}}$ of elements in $(0,p)$ such that $\lim_{n\to\infty} q_n = p$ and $F^{-1}_{\mu\wedge_{\mathrm{icx}}\nu}(q_n) \in \{F^{-1}_\mu(q_n), F^{-1}_\nu(q_n)\}$ for all $n \in \mathbb{N}$. Either $F^{-1}_{\mu\wedge_{\mathrm{icx}}\nu}(q_n) = F^{-1}_\mu(q_n)$ for infinitely many n and, by left-continuity of the quantile functions, $F^{-1}_{\mu\wedge_{\mathrm{icx}}\nu}(p) = F^{-1}_\mu(p)$ or $F^{-1}_{\mu\wedge_{\mathrm{icx}}\nu}(q_n) = F^{-1}_\nu(q_n)$ for infinitely many n and $F^{-1}_{\mu\wedge_{\mathrm{icx}}\nu}(p) = F^{-1}_\nu(p)$.

Similarly, we show that $F^{-1}_{\mu\wedge_{\mathrm{dcx}}\nu}(p) \in \{F^{-1}_\mu(p), F^{-1}_\nu(p)\}$ by considering the derivative of $q \mapsto \left(\int_0^q F^{-1}_\mu(p)dp - \int_0^q F^{-1}_\nu(p)dp\right)^+ + \int_0^q F^{-1}_\nu(p)dp$. □

Definition 4.3. *For* $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$, *we define* $\mu \wedge \nu, \mu \vee \nu \in \mathcal{P}_1(\mathbb{R})$ *by*

$$\mu \wedge \nu = 1_{\{\int_{\mathbb{R}} x\mu(dx) \le \int_{\mathbb{R}} x\nu(dx)\}}\mu \wedge_{\mathrm{dcx}} \nu + 1_{\{\int_{\mathbb{R}} x\mu(dx) > \int_{\mathbb{R}} x\nu(dx)\}}\mu \wedge_{\mathrm{icx}} \nu, \tag{4.1}$$

$$\mu \vee \nu = 1_{\{\int_{\mathbb{R}} x\mu(dx) \le \int_{\mathbb{R}} x\nu(dx)\}}\mu \vee_{\mathrm{dcx}} \nu + 1_{\{\int_{\mathbb{R}} x\mu(dx) > \int_{\mathbb{R}} x\nu(dx)\}}\mu \vee_{\mathrm{icx}} \nu, \tag{4.2}$$

By Lemma 4.1, $\int_{\mathbb{R}} x\mu\wedge\nu(dx) = \int_{\mathbb{R}} x\nu(dx)$ and $\int_{\mathbb{R}} x\mu\vee\nu(dx) = \int_{\mathbb{R}} x\mu(dx)$. By Lemma 1.5 of Kertz and Rösler (1992), for fixed $a \in \mathbb{R}$, the convex, increasing convex and decreasing convex orders coincide on $\{\eta \in \mathcal{P}_1(\mathbb{R}) \text{ such that } \int_{\mathbb{R}} x\eta(dx) = a\}$. Therefore

$$\mu \wedge \nu \le_{\mathrm{cx}} \nu \text{ and } \mu \le_{\mathrm{cx}} \mu \vee \nu, \tag{4.3}$$

and, when $\int_{\mathbb{R}} x\mu(dx) = \int_{\mathbb{R}} x\nu(dx)$, $\mu \wedge \nu = \mu \wedge_{\mathrm{cx}} \nu$ and $\mu \vee \nu = \mu \vee_{\mathrm{cx}} \nu$.

Remark 4.4. *Let $\mu,\nu\in\mathcal{P}_1(\mathbb{R})$. Then $\nu\leq_{\rm cx}\mu\vee\nu \iff \int_{\mathbb{R}}x\nu(dx)=\int_{\mathbb{R}}x\mu(dx)$. Let $\widetilde{\nu}$ be the image of ν by $t\mapsto t+\int_{\mathbb{R}}x\mu(dx)-\int_{\mathbb{R}}x\nu(dx)$. One has $\varphi_{\widetilde{\nu}}(t)=\varphi_\nu\left(t+\int_{\mathbb{R}}x\nu(dx)-\int_{\mathbb{R}}x\mu(dx)\right)$ and $\bar{\varphi}_{\widetilde{\nu}}(t)=\bar{\varphi}_\nu\left(t+\int_{\mathbb{R}}x\nu(dx)-\int_{\mathbb{R}}x\mu(dx)\right)$ for all $t\in\mathbb{R}$. When $\int_{\mathbb{R}}x\mu(dx)\leq\int_{\mathbb{R}}x\nu(dx)$, then $\varphi_\nu\leq\varphi_{\widetilde{\nu}}$ and $\mu\vee\nu=\mu\vee_{\rm dcx}\nu$ so that $\varphi_{\mu\vee\nu}=\varphi_\mu\vee\varphi_\nu\leq\varphi_\mu\vee\varphi_{\widetilde{\nu}}=\varphi_{\mu\vee\widetilde{\nu}}$. When $\int_{\mathbb{R}}x\mu(dx)>\int_{\mathbb{R}}x\nu(dx)$, then $\bar{\varphi}_\nu\leq\bar{\varphi}_{\widetilde{\nu}}$ and $\mu\vee\nu=\mu\vee_{\rm icx}\nu$ so that $\bar{\varphi}_{\mu\vee\nu}=\bar{\varphi}_\mu\vee\bar{\varphi}_\nu\leq\bar{\varphi}_\mu\vee\bar{\varphi}_{\widetilde{\nu}}=\bar{\varphi}_{\mu\vee\widetilde{\nu}}$. Therefore, in both cases, $\mu\leq_{\rm cx}\mu\vee\nu\leq_{\rm cx}\mu\vee\widetilde{\nu}$.*

In the same way, $\mu\wedge\nu\leq_{\rm cx}\mu$ if and only if $\int_{\mathbb{R}}x\mu(dx)=\int_{\mathbb{R}}x\nu(dx)$. Let $\widetilde{\mu}$ denote the image of μ by $t\mapsto t+\int_{\mathbb{R}}x\nu(dx)-\int_{\mathbb{R}}x\mu(dx)$. One has $F^{-1}_{\widetilde{\mu}}(p)=F^{-1}_\mu(p)+\int_{\mathbb{R}}x\nu(dx)-\int_{\mathbb{R}}x\mu(dx)$ for $p\in(0,1)$. When $\int_{\mathbb{R}}x\mu(dx)\leq\int_{\mathbb{R}}x\nu(dx)$, then $F^{-1}_\mu\leq F^{-1}_{\widetilde{\mu}}$ and $\mu\wedge\nu=\mu\wedge_{\rm dcx}\nu$, so that, by Lemma 4.2, $\int_0^q F^{-1}_{\mu\wedge\nu}(p)dp=\int_0^q F^{-1}_\mu(p)dp\vee\int_0^q F^{-1}_\nu(p)dp\leq\int_0^q F^{-1}_{\widetilde{\mu}}(p)dp\vee\int_0^q F^{-1}_\nu(p)dp=\int_0^q F^{-1}_{\widetilde{\mu}\wedge\nu}(p)dp$ for $q\in[0,1]$. When $\int_{\mathbb{R}}x\mu(dx)>\int_{\mathbb{R}}x\nu(dx)$, then $F^{-1}_\mu>F^{-1}_{\widetilde{\mu}}$ and $\mu\wedge\nu=\mu\wedge_{\rm icx}\nu$, so that, again by Lemma 4.2, $\int_q^1 F^{-1}_{\mu\wedge\nu}(p)dp=\int_q^1 F^{-1}_\mu(p)dp\wedge\int_q^1 F^{-1}_\nu(p)dp\geq\int_q^1 F^{-1}_{\widetilde{\mu}}(p)dp\wedge\int_q^1 F^{-1}_\nu(p)dp=\int_q^1 F^{-1}_{\widetilde{\mu}\wedge\nu}(p)dp$ for $q\in[0,1]$. With Theorem 2.4 (iii), we deduce that in both cases, $\widetilde{\mu}\wedge\nu\leq_{\rm cx}\mu\wedge\nu\leq_{\rm cx}\nu$.

Let us now check that the diameter of the set $\{\mu,\nu,\mu\vee_{\rm dcx}\nu,\mu\vee_{\rm icx}\nu,\mu\wedge_{\rm dcx}\nu,\mu\wedge_{\rm icx}\nu\}$ (resp. $\{\mu,\nu,\mu\vee_{\rm dcx}\nu,\mu\wedge_{\rm dcx}\nu\},\{\mu,\nu,\mu\vee_{\rm icx}\nu,\mu\wedge_{\rm icx}\nu\}$) when $\mu,\nu\in\mathcal{P}_1(\mathbb{R})$ (resp. $\mathcal{P}_-(\mathbb{R}),\mathcal{P}_+(\mathbb{R})$)) in Wasserstein distance is equal to the Wasserstein distance between μ and ν.

Lemma 4.5. *For $\eta\in\{\mu,\nu,\mu\vee_{\rm dcx}\nu,\mu\vee_{\rm icx}\nu,\mu\wedge_{\rm dcx}\nu,\mu\wedge_{\rm icx}\nu\}$, $\forall p\in(0,1)$, $F^{-1}_\mu(p)\wedge F^{-1}_\nu(p)\leq F^{-1}_\eta(p)\leq F^{-1}_\mu(p)\vee F^{-1}_\nu(p)$.*

Let $\varrho\geq1$. Since, by Proposition 2.17 of Santambrogio (2015), the ϱ-Wasserstein distance between two probability measures $\eta_1,\eta_2\in\mathcal{P}(\mathbb{R})$ is given by

$$W_\varrho(\eta_1,\eta_2)=\left(\int_0^1|F^{-1}_{\eta_1}(p)-F^{-1}_{\eta_2}(p)|^\varrho dp\right)^{1/\varrho},$$

one easily deduces that for all $\eta_1,\eta_2\in\{\mu,\nu,\mu\vee_{\rm dcx}\nu,\mu\vee_{\rm icx}\nu,\mu\wedge_{\rm dcx}\nu,\mu\wedge_{\rm icx}\nu\}$,

$$W_\varrho(\eta_1,\eta_2)\leq\left(\int_0^1(F^{-1}_\mu(p)\vee F^{-1}_\nu(p)-F^{-1}_\mu(p)\wedge F^{-1}_\nu(p))^\varrho dp\right)^{1/\varrho}=W_\varrho(\mu,\nu).$$

Proof. For $\eta\in\{\mu\wedge_{\rm icx}\nu,\mu\wedge_{\rm dcx}\nu\}$ this is a consequence of the last statement in Lemma 4.2 and the left-continuity of the quantile functions. Let $\eta\in\{\mu\vee_{\rm icx}\nu,\mu\vee_{\rm dcx}\nu\}$. Since for all $t\in\mathbb{R}$, $\int_{-\infty}^t F_{\mu\wedge_{\rm dcx}\nu}(x)dx=\int_{-\infty}^t F_\mu(x)dx\vee\int_{-\infty}^t F_\nu(x)dx$ and $\int_t^{+\infty}\bar{F}_{\mu\wedge_{\rm icx}\nu}(x)dx=\int_t^{+\infty}\bar{F}_\mu(x)dx\vee\int_t^{+\infty}\bar{F}_\nu(x)dx$ a reasoning analogous to the proof of this last statement ensures that $\forall x\in\mathbb{R}$, $F_\eta(x)\in\{F_\mu(x),F_\nu(x)\}$. For $p\in(0,1)$, since

$$\begin{aligned}&\{x\in\mathbb{R}:F_\mu(x)\geq p\}\cap\{x\in\mathbb{R}:F_\nu(x)\geq p\}\\&\qquad=\{x\in\mathbb{R}:F_\mu(x)\wedge F_\nu(x)\geq p\}\subset\{x\in\mathbb{R}:F_\eta(x)\geq p\},\\&\{x\in\mathbb{R}:F_\mu(x)\geq p\}\cup\{x\in\mathbb{R}:F_\nu(x)\geq p\}\\&\qquad=\{x\in\mathbb{R}:F_\mu(x)\vee F_\nu(x)\geq p\}\supset\{x\in\mathbb{R}:F_\eta(x)\geq p\},\end{aligned}$$

we get $F^{-1}_\mu(p)\wedge F^{-1}_\nu(p)\leq F^{-1}_\eta(p)\leq F^{-1}_\mu(p)\vee F^{-1}_\nu(p)$. □

4.2. **Approximations in convex order.** Let $\mu,\nu\in\mathcal{P}_1(\mathbb{R})$ be two probability measures such that $\mu\leq_{\rm cx}\nu$. We want to construct a couple of measures with finite supports and in convex order which approximate (μ,ν). We first generate $(\widetilde{x}^I_i)_{1\leq i\leq I}$ and $(\widetilde{y}^J_j)_{1\leq j\leq J}$ such that $(\frac{1}{I}\sum_{i=1}^I\delta_{\widetilde{x}^I_i},\frac{1}{J}\sum_{j=1}^J\delta_{\widetilde{y}^J_j})$ approximates (μ,ν) in one of the following ways :

1. $(\widetilde{x}_1^I,\dots,\widetilde{x}_I^I)=(X_1,\dots,X_I)$ and $(\widetilde{y}_J^J,\dots,\widetilde{y}_J^J)=(Y_1,\dots,Y_J)$ where $((X_i,Y_i))_{i\geq 1}$ are random vectors i.i.d. according to $\mu\otimes\nu$,
2. $\widetilde{x}_i^I=F_\mu^{-1}\left(\frac{2i-1}{2I}\right)$ for $i\in\{1,\dots,I\}$ and $\widetilde{y}_j^J=F_\nu^{-1}\left(\frac{2j-1}{2J}\right)$ for $j\in\{1,\dots,J\}$.

We may want either to simply keep these samples with, in general, distinct empirical means by setting

a. $(x_1^I,\dots,x_I^I)=(\widetilde{x}_1^I,\dots,\widetilde{x}_I^I)$ and $(y_1^J,\dots,y_J^J)=(\widetilde{y}_1^J,\dots,\widetilde{y}_J^J)$,

or to modify the approximate measures to ensure that they have the same expectation. We then propose one of the following modifications

b. set $(x_1^I,\dots,x_I^I)=(\widetilde{x}_1^I+\int_{\mathbb{R}}x\mu(dx)-\frac{1}{I}\sum_{i=1}^I\widetilde{x}_i^I,\dots,\widetilde{x}_I^I+\int_{\mathbb{R}}x\mu(dx)-\frac{1}{I}\sum_{i=1}^I\widetilde{x}_i^I)$ and $(y_1^J,\dots,y_J^J)=(\widetilde{y}_1^J+\int_{\mathbb{R}}y\nu(dy)-\frac{1}{J}\sum_{j=1}^J\widetilde{y}_j^J,\dots,\widetilde{y}_J^J+\int_{\mathbb{R}}y\nu(dy)-\frac{1}{J}\sum_{j=1}^J\widetilde{y}_j^J)$.

c. set $(y_1^J,\dots,y_J^J)=(\widetilde{y}_1^J+\frac{1}{I}\sum_{i=1}^I\widetilde{x}_i^I-\frac{1}{J}\sum_{k=1}^J\widetilde{y}_k^J,\dots,\widetilde{y}_J^J+\frac{1}{I}\sum_{i=1}^I\widetilde{x}_i^I-\frac{1}{J}\sum_{k=1}^J\widetilde{y}_k^J)$ and $(x_1^I,\dots,x_I^I)=(\widetilde{x}_1^I,\dots,\widetilde{x}_I^I)$,

d. compute the estimators of the variances $\widehat{\sigma_x^2}=\frac{1}{I-1}\sum_{i=1}^I(\widetilde{x}_i^I)^2-\frac{1}{I(I-1)}(\sum_{i=1}^I\widetilde{x}_i^I)^2$ and $\widehat{\sigma_y^2}=\frac{1}{J-1}\sum_{j=1}^J(\widetilde{y}_j^J)^2-\frac{1}{J(J-1)}(\sum_{j=1}^J\widetilde{y}_j^J)^2$ and add $\frac{J\widehat{\sigma_x^2}}{I\widehat{\sigma_y^2}+J\widehat{\sigma_x^2}}\left(\frac{1}{J}\sum_{j=1}^J\widetilde{y}_j^J-\frac{1}{I}\sum_{i=1}^I\widetilde{x}_i^I\right)$ to each $\widetilde{x}_i^I$ to obtain x_i^I and $\frac{I\widehat{\sigma_y^2}}{I\widehat{\sigma_y^2}+J\widehat{\sigma_x^2}}\left(\frac{1}{I}\sum_{i=1}^I\widetilde{x}_i^I-\frac{1}{J}\sum_{j=1}^J\widetilde{y}_j^J\right)$ to each $\widetilde{y}_j^J$ to obtain y_j^J.

The first modification necessitates the knowledge of $\int_{\mathbb{R}}x\mu(dx)=\int_{\mathbb{R}}y\nu(dy)$ which is generally the case in financial applications, since this is the current price of the underlying asset. The second modification seems suitable for the deterministic initialization of points with $I=J$, since formally, because of the inequality $\mu\leq_{\mathrm{cx}}\nu$, the tails of μ should be lighter than the tails of ν enabling a quicker convergence of $\frac{1}{I}\sum_{i=1}^I F_\mu^{-1}\left(\frac{2i-1}{2I}\right)$ than of $\frac{1}{I}\sum_{i=1}^I F_\nu^{-1}\left(\frac{2i-1}{2I}\right)$ to the common expectation of μ and ν as $I\to\infty$. When $\int_{\mathbb{R}}y^2\nu(dy)<\infty$, the convex combination of $\frac{1}{I}\sum_{i=1}^I X_i$ and $\frac{1}{J}\sum_{j=1}^J Y_j$ with minimal variance is

$$\frac{\mathrm{Var}(Y_1)}{I\mathrm{Var}(Y_1)+J\mathrm{Var}(X_1)}\sum_{i=1}^I X_i+\frac{\mathrm{Var}(X_1)}{I\mathrm{Var}(Y_1)+J\mathrm{Var}(X_1)}\sum_{j=1}^J Y_j$$

and is approximated by the common mean $\frac{\widehat{\sigma_y^2}}{I\widehat{\sigma_y^2}+J\widehat{\sigma_x^2}}\sum_{i=1}^I\widetilde{x}_i^I+\frac{\widehat{\sigma_x^2}}{I\widehat{\sigma_y^2}+J\widehat{\sigma_x^2}}\sum_{j=1}^J\widetilde{y}_j^J$ of the two samples $(x_1^I,\dots,x_I^I)$ and $(y_1^J,\dots,y_J^J)$ for the random initialization of points.

We then set

$$\mu_I=\frac{1}{I}\sum_{i=1}^I\delta_{x_i^I}\text{ and }\nu_J=\frac{1}{J}\sum_{j=1}^J\delta_{y_j^J},$$

and consider either $(\mu_I,\mu_I\vee\nu_J)$ or $(\mu_I\wedge\nu_J,\nu_J)$ to approximate (μ,ν) in the convex order. The next proposition shows without giving any speed of convergence that the measures μ and ν are then well approximated asymptotically. Notice that, according to Remark 4.4, the measure $\mu_I\vee\nu_I$ obtained with the second modification dominates in the convex order the one obtained without modification.

Proposition 4.6. *Let μ and ν be two probability measures on $\mathbb{R}$ such that $\mu\leq_{\mathrm{cx}}\nu$ and $\int_{\mathbb{R}}|x|\nu(dx)<\infty$. As $I,J\to\infty$, μ_I and $\mu_I\vee\nu_J$ (resp. $\mu_I\wedge\nu_J$ and ν_J) converges a.s. weakly to μ and ν.*

From this proposition, it seems then natural to approximate the continuous MOT

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

by the minimum of the discrete MOT (1.1) for the measures $\mu_I \wedge \nu_J$ and ν_J when I, J are large. We will do this with $I = J$ in Section 6 on numerical examples. However, up to our knowledge, no convergence result is known in the literature for the MOT cost between probability measures $\hat{\mu}$ and $\hat{\nu}$ such that $\hat{\mu} \leq_{\mathrm{cx}} \hat{\nu}$, in the limit $\hat{\mu} \to \mu$ and $\hat{\nu} \to \nu$ for the weak sense. We intend to address this important question in future research.

To prove Proposition 4.6, we need the following lemma to deal with the deterministic choice of points.

Lemma 4.7. *Let μ be such that $\int_{\mathbb{R}} |x|\mu(dx) < \infty$. As $I \to \infty$, $\frac{1}{I}\sum_{i=1}^{I} \delta_{F_\mu^{-1}\left(\frac{2i-1}{2I}\right)}$ converges weakly to μ, $\frac{1}{I}\sum_{i=1}^{I} F_\mu^{-1}\left(\frac{2i-1}{2I}\right)$ converges to $\int_{\mathbb{R}} x\mu(dx)$ and for all $t \in \mathbb{R}$, $\frac{1}{I}\sum_{i=1}^{I}(t - F_\mu^{-1}\left(\frac{2i-1}{2I}\right))^+$ converges to $\int_{\mathbb{R}}(t-x)^+\mu(dx)$.*

Proof. The weak convergence follows from the continuity of $(0,1) \ni p \mapsto F_\mu^{-1}(p)$ outside of a set with zero Lebesgue measure and the weak convergence of $\frac{1}{I}\sum_{i=1}^{I}\delta_{\frac{2i-1}{2I}}$ to the Lebesgue measure on $(0,1)$. We get by the weak convergence

$$\liminf_{I\to\infty} \frac{1}{I}\sum_{i=1}^{I}(F_\mu^{-1}\left((2i-1)/2I\right))^- \geq \int_0^1 (F_\mu^{-1}(p))^- dp.$$

From the inequality

$$\frac{1}{I}\sum_{i=1}^{I}(F_\mu^{-1}\left((2i-1)/2I\right))^- \leq \int_0^{\frac{1}{2I}} (F_\mu^{-1}(p))^- + \int_0^{1-\frac{1}{2I}} (F_\mu^{-1}(p))^- dp,$$

and since $\int_0^1 |F_\mu^{-1}(p))|dp < \infty$, we deduce that

$$\lim_{I\to\infty} \frac{1}{I}\sum_{i=1}^{I}(F_\mu^{-1}\left((2i-1)/2I\right))^- = \int_0^1 (F_\mu^{-1}(p))^- dp. \tag{4.4}$$

Dealing in a symmetric way with the positive part ensures the convergence of the expectations. The inequality $(t-x)^+ \leq t^+ + x^-$ for $t, x \in \mathbb{R}$ combined with (4.4) yields some uniform integrability property ensuring the last convergence. □

Proof of Proposition 4.6. Let us deal with the second modification, the proof being similar for the first or third modifications or with no modification. We also deal with the random choice of points, the previous lemma replacing the strong law of large numbers for the deterministic choice. The a.s. weak convergence of the empirical measure μ_I to μ as $I \to \infty$ is a consequence of the strong law of large numbers. Still by the strong law of large numbers, $\frac{1}{I}\sum_{i=1}^{I} X_i - \frac{1}{J}\sum_{j=1}^{J} Y_j$ converges a.s. to $\int_{\mathbb{R}} x\mu(dx) - \int_{\mathbb{R}} x\nu(dx) = 0$ as $I, J \to \infty$. For $t \in \mathbb{R}$,

since $x \mapsto (t-x)^+$ is Lipschitz continuous with constant 1,

$$\left|\int_{\mathbb{R}}(t-x)^+\nu_J(dx) - \int_{\mathbb{R}}(t-x)^+\nu(dx)\right| \le \left|\frac{1}{I}\sum_{i=1}^{I} X_i - \frac{1}{J}\sum_{j=1}^{J} Y_j\right| + \left|\frac{1}{J}\sum_{j=1}^{J}(t-Y_j)^+ - \int_{\mathbb{R}}(t-x)^+\nu(dx)\right|.$$

Dealing with the second term in the right-hand side by the strong law of large numbers, one deduces that $\varphi_{\nu_J}(t)$ converges a.s. to $\varphi_\nu(t)$ as $I, J \to \infty$. The monotonicity in t and the continuity of the limit in the same variable ensures that a.s. the convergence holds for all $t \in \mathbb{R}$. In the same way a.s. for all $t \in \mathbb{R}$, $\varphi_{\mu_I}(t)$ converges to $\varphi_\mu(t)$. Since $\varphi_{\mu_I \vee \nu_J}(t) = \max(\varphi_{\mu_I}(t), \varphi_{\nu_J}(t))$, we deduce that as $I, J \to \infty$, a.s., for all $t \in \mathbb{R}$, $\varphi_{\mu_I\vee\nu_J}(t)$ converges to $\max(\varphi_\mu(t), \varphi_\nu(t)) = \varphi_\nu(t)$ since $\mu \le_{\mathrm{cx}} \nu$. For $t \in \mathbb{R}$ and $h > 0$,

$$\frac{1}{h}(\varphi_{\mu_I\vee\nu_J}(t) - \varphi_{\mu_I\vee\nu_J}(t-h)) = \frac{1}{h}\int_{t-h}^{t} F_{\mu_I\vee\nu_J}(x)dx$$
$$\le F_{\mu_I\vee\nu_J}(t) \le \frac{1}{h}\int_{t}^{t+h} F_{\mu_I\vee\nu_J}(x)dx = \frac{1}{h}(\varphi_{\mu_I\vee\nu_J}(t+h) - \varphi_{\mu_I\vee\nu_J}(t))$$

where the left-hand and the right-hand sides of the inequality respectively converge to $\frac{1}{h}\int_{\mathbb{R}}(t-x)^+ - (t-h-x)^+\nu(dx) = \frac{1}{h}\int_{t-h}^{t} F_\nu(x)dx$ and $\frac{1}{h}\int_{\mathbb{R}}(t+h-x)^+ - (t-x)^+\nu(dx) = \frac{1}{h}\int_{t}^{t+h} F_\nu(x)dx$. Letting $h \to 0$, we conclude that a.s., $F_{\mu_I\vee\nu_J}(t)$ converges to $F_\nu(t)$ for all $t \in \mathbb{R}$ such that $\nu(\{t\}) = 0$ which is equivalent to the weak convergence of $\mu_I \vee \nu_J$ to ν.

We now prove the a.s. weak convergence of $\mu_I \wedge \nu_J$ to μ. We have by construction $\varphi_{\mu_I\wedge\nu_J}(t) \le \min(\varphi_{\mu_I}(t), \varphi_{\nu_J}(t))$. This gives a.s, that for all $t \in \mathbb{R}$, $\limsup_{I,J\to+\infty} \varphi_{\mu_I\wedge\nu_J}(t) \le \varphi_\mu(t)$ since $\mu \le_{\mathrm{cx}} \nu$. Now, we observe that the convergence of φ_{μ_I} (resp. φ_{ν_J}) to φ_μ (resp. φ_ν) is uniform on $\mathbb{R}$. In fact, φ_{μ_I} and φ_μ are nondecreasing functions such that $\lim_{t\to-\infty}\varphi_{\mu_I}(t) = \lim_{t\to-\infty}\varphi_\mu(t) = 0$ and φ_μ is continuous. From the a.s. simple convergence, we deduce that a.s., φ_{μ_I} converges uniformly to φ_μ on $(-\infty, T]$, for any $T > 0$. The same reasoning gives that a.s., $\bar\varphi_{\mu_I}$ converges uniformly to $\bar\varphi_\mu$ on $[T, +\infty)$. Since, by (2.1) and (2.2),

$$t - \int_{\mathbb{R}} x\mu(dx) = \int_{\mathbb{R}}(t-x)^+\mu(dx) - \int_{\mathbb{R}}(x-t)^+\mu(dx) = \varphi_\mu(t) - \bar\varphi_\mu(t),$$

$t - \frac{1}{I}\sum_{i=1}^{I} X_i = \varphi_{\mu_I}(t) - \bar\varphi_{\mu_I}(t)$ and $\frac{1}{I}\sum_{i=1}^{I} X_i \underset{I\to+\infty}{\to} \int_{\mathbb{R}} x\mu(dx)$, we get that a.s., φ_{μ_I} converges uniformly to φ_μ on $\mathbb{R}$. The same argument works for the uniform convergence of φ_{ν_J} to φ_ν. Thus, for $\varepsilon > 0$ there exists N such that for $I, J \ge N$,

$$\forall t \in \mathbb{R}, \max(|\varphi_{\mu_I} - \varphi_\mu(t)|, |\varphi_{\nu_J} - \varphi_\nu(t)|) \le \varepsilon.$$

We deduce

$$\forall t \in \mathbb{R}, \varphi_{\mu_I}(t) = \varphi_{\mu_I}(t) \wedge \varphi_{\nu_J}(t) + \varphi_{\mu_I}(t) \vee \varphi_{\nu_J}(t) - \varphi_{\nu_J}(t)$$
$$\le \varphi_{\mu_I}(t) \wedge \varphi_{\nu_J}(t) + \varphi_\mu(t) \vee \varphi_\nu(t) - \varphi_\nu(t) + 2\varepsilon = \varphi_{\mu_I}(t) \wedge \varphi_{\nu_J}(t) + 2\varepsilon,$$

since $\mu \le_{\mathrm{cx}} \nu$. Since $\varphi_{\mu_I} - 2\varepsilon$ is a convex function below $\varphi_{\mu_I} \wedge \varphi_{\nu_J}$, we get that

$$\forall I, J \ge N, t \in \mathbb{R}, \varphi_{\mu_I}(t) - 2\varepsilon \le \varphi_{\mu_I\wedge\nu_J}(t).$$

Since $\varepsilon > 0$ is arbitrary, this gives a.s, that for all $t \in \mathbb{R}$, $\liminf_{I,J\to+\infty} \varphi_{\mu_I\wedge\nu_J}(t) \geq \varphi_\mu(t)$. From the a.s. convergence for all $t \in \mathbb{R}$ of $\varphi_{\mu_I\wedge\nu_J}(t)$ to $\varphi_\mu(t)$, we deduce as before the convergence for all $t \in \mathbb{R}$ such that $\mu(\{t\}) = 0$ of $F_{\mu_I\wedge\nu_J}(t)$ to $F_\mu(t)$. □

Remark 4.8. *The proof of Proposition 4.6 does not use the independence between $(X_i)_{i\geq1}$ and $(Y_i)_{i\geq1}$ when they are sampled. In fact, among the three proposed modifications, only the third one exploits the independence in order to get the best estimation of the common expectation $\int_{\mathbb{R}} x\mu(dx) = \int_{\mathbb{R}} y\mu(dy)$ from the samples $(X_i)_{1\leq i\leq I}$ and $(Y_j)_{1\leq j\leq J}$.*

An interesting thing to notice is that the construction of approximate probability measures that preserve the convex order can be easily extended to the multi-marginal case. More precisely, let $\ell \geq 2$, $I_1, \ldots, I_\ell$ be positive integers and $\mu^1, \ldots, \mu^\ell$ be probability measures on $\mathbb{R}$ such that $\mu^1 \leq_{\rm cx} \ldots \leq_{\rm cx} \mu^\ell$ and $\int_{\mathbb{R}} |x|\mu^\ell(dx) < \infty$. We either set $\widetilde{x}_i^k = F^{-1}_{\mu^k}\left(\frac{2i-1}{I_k}\right)$ for $i \in \{1, \ldots, I_k\}$ and $k \in \{1, \ldots, \ell\}$ or $\widetilde{x}_i^k = X_i^k$ where the random vectors $((X_i^1, \ldots, X_i^\ell))_{i\geq1}$ are i.i.d. according to $\mu^1 \otimes \ldots \otimes \mu^\ell$. Next we choose $(x_i^k)_{1\leq i\leq I_k, 1\leq k\leq \ell}$ as one of the following vectors :

- $(\widetilde{x}_i^k)_{1\leq i\leq I_k, 1\leq k\leq \ell}$,
- $(\widetilde{x}_i^k - \frac{1}{I_k}\sum_{i=1}^{I_k} \widetilde{x}_i^k + \int_{\mathbb{R}} x\mu_1(dx))_{1\leq i\leq I_k, 1\leq k\leq \ell}$,
- $(\widetilde{x}_i^k - \frac{1}{I_k}\sum_{i=1}^{I_k} \widetilde{x}_i^k + \frac{1}{I_1}\sum_{i=1}^{I_1} \widetilde{x}_i^1)_{1\leq i\leq I_k, 1\leq k\leq \ell}$,
- $\left(\widetilde{x}_i^k + \sum_{j=1}^{\ell} \frac{I_j/\widehat{\sigma_j^2}}{\sum_{l=1}^{\ell} I_l/\widehat{\sigma_l^2}}(\frac{1}{I_j}\sum_{i=1}^{I_j} \widetilde{x}_i^j - \frac{1}{I_k}\sum_{i=1}^{I_k} \widetilde{x}_i^k)\right)_{1\leq i\leq I_k, 1\leq k\leq \ell}$,

where $\widehat{\sigma_j^2} = \frac{1}{I_j-1}\sum_{i=1}^{I_j}(\widetilde{x}_i^j)^2 - \frac{1}{I_j(I_j-1)}(\sum_{i=1}^{I_j} \widetilde{x}_i^j)^2$ and set $\mu_{I_k}^k = \frac{1}{I_k}\sum_{i=1}^{I_k} \delta_{x_i^k}$ for $k \in \{1, \ldots, \ell\}$. For the first choice, the expectations of the measures $\mu_{I_k}^k$ are in general distinct and, to obtain an increasing sequence of approximate probability measures for the convex order, we compute either

$$\mu_{I_1}^1, \mu_{I_1}^1 \vee \mu_{I_2}^2, (\mu_{I_1}^1 \vee \mu_{I_2}^2) \vee \mu_{I_3}^3, \ldots, ((\mu_{I_1}^1 \vee \mu_{I_2}^2) \vee \ldots) \vee \mu_{I_\ell}^\ell$$

or $\mu_{I_1}^1 \wedge (\ldots \wedge (\mu_{I_{\ell-1}}^{\ell-1} \wedge \mu_{I_\ell}^\ell)), \ldots, \mu_{I_{\ell-2}}^{\ell-2} \wedge (\mu_{I_{\ell-1}}^{\ell-1} \wedge \mu_{I_\ell}^\ell), \mu_{I_{\ell-1}}^{\ell-1} \wedge \mu_{I_\ell}^\ell, \mu_{I_\ell}^\ell$. For the other choices, the expectations are equal and the order in which the suprema $\vee$ (resp. the infima $\wedge$) are computed no longer matters. Indeed, for $f, g, h : \mathbb{R} \to \mathbb{R}$, we have $\max(f, g, h) = \max(\max(f, g), h)$ and $\min(f, g, h) = \min((\min(f, g)), h))$. One easily generalizes the proof of Proposition 4.6 to obtain the following result.

Proposition 4.9. *Let $\mu^1, \ldots, \mu^\ell$ be probability measures on $\mathbb{R}$ such that $\mu^1 \leq_{\rm cx} \ldots \leq_{\rm cx} \mu^\ell$ and $\int_{\mathbb{R}} |x|\mu^\ell(dx) < \infty$. Then, as $I_1, \ldots, I_\ell \to +\infty$, $\mu_{I_1}^1$, $\mu_{I_1}^1, \mu_{I_1}^1 \vee \mu_{I_2}^2, (\mu_{I_1}^1 \vee \mu_{I_2}^2) \vee \mu_{I_3}^3, \ldots, ((\mu_{I_1}^1 \vee \mu_{I_2}^2) \vee \ldots) \vee \mu_{I_\ell}^\ell$ (resp. $\mu_{I_1}^1 \wedge (\ldots \wedge (\mu_{I_{\ell-1}}^{\ell-1} \wedge \mu_{I_\ell}^\ell)), \ldots, \mu_{I_{\ell-2}}^{\ell-2} \wedge (\mu_{I_{\ell-1}}^{\ell-1} \wedge \mu_{I_\ell}^\ell), \mu_{I_{\ell-1}}^{\ell-1} \wedge \mu_{I_\ell}^\ell, \mu_{I_\ell}^\ell$) converge a.s. weakly to $\mu^1, \ldots, \mu^\ell$.*

5. Algorithms and complexity

In this section, we present the algorithms to calculate $\mu_I \wedge \nu_J$ and $\mu_I \vee \nu_J$ defined in (4.1) and (4.2) when $\mu_I = \sum_{i=1}^I p_i\delta_{x_i}$ and $\nu_J = \sum_{j=1}^J q_j\delta_{y_j}$ are probability measures with finite support. In this case, the probability measures $\mu_I \wedge \nu_J$ and $\mu_I \vee \nu_J$ also have a finite support, and the number of points in their support is discussed in Subsection 5.3. When $x_1 < \cdots < x_I$ and $y_1 < \cdots < y_J$, the number of operations required to calculate $\mu_I \wedge \nu_J$ or $\mu_I \vee \nu_J$ is in $O(I + J)$. Otherwise, the computation cost is in $O(I \log(I) + J \log(J))$ since one has first to sort the points.

5.1. **Algorithm for the calculation of the supremum of two probability measures with finite support.** Let $\mu_I = \sum_{i=1}^I p_i \delta_{x_i}$ with $(p_1, \dots, p_I) \in (0,1]^I$ and $\nu_J = \sum_{j=1}^J q_j \delta_{y_j}$ with $(q_1, \dots, q_J) \in (0,1]^J$ be two probability measures on the real line. Up to reducing the number of points and sorting them, we may suppose that $x_1 < x_2 < \ldots < x_I$ (resp $y_1 < y_2 < \ldots < y_J$). The equality $\mu_I \vee_{\rm icx} \nu_J = \overline{\overline{\mu}_I \vee_{\rm dcx} \overline{\nu}_J}$ implies that it is enough to focus on the computation of $\mu_I \vee_{\rm dcx} \nu_J$.

The functions φ_{μ_I} and φ_{ν_J} are piecewise affine with slopes taking respectively their values in $\{0, p_1, p_1+p_2, \dots, \sum_{i=1}^{I-1} p_i, 1\}$ and $\{0, q_1, q_1+q_2, \dots, \sum_{j=0}^{J-1} q_j, 1\}$. As a consequence $\varphi_{\mu_I} \vee \varphi_{\nu_J}$ is piecewise affine with slopes taking values in the union of these two sets so that $\mu_I \vee \nu_J$ is a probability measure supported on $K \le I+J-1$ points $z_1 < z_2 < \ldots < z_K$. Our aim is now to provide an efficient algorithm aimed at expliciting $\mu_I \vee_{\rm dcx} \nu_J = \sum_{k=1}^K \varrho_k \delta_{z_k}$. Let $w_1 < w_2 < \ldots < w_L$ with $L \ge I \vee J$ be such that

$$\{w_1, \dots, w_L\} = \{x_1, \dots, x_I\} \cup \{y_1, \dots, y_J\}.$$

For convenience, we set $w_0 = w_1 - 1$. Of course, $\varphi_{\mu_I}(w_0) = F_{\mu_I}(w_0) = \varphi_{\nu_J}(w_0) = F_{\nu_J}(w_0) = 0$. We now proceed inductively on $\ell \in \{1, \dots, L\}$ remarking that the functions φ_{μ_I} and φ_{ν_J} are affine with respective slopes $F_{\mu_I}(w_{\ell-1})$ and $F_{\nu_J}(w_{\ell-1})$ on $[w_{\ell-1}, w_\ell]$:

1. We first compute
$$\varphi_{\mu_I}(w_\ell) = \varphi_{\mu_I}(w_{\ell-1}) + F_{\mu_I}(w_{\ell-1})(w_\ell - w_{\ell-1}), \quad F_{\mu_I}(w_\ell) = F_{\mu_I}(w_{\ell-1}) + \mu_I(\{w_\ell\})$$
$$\varphi_{\nu_J}(w_\ell) = \varphi_{\nu_J}(w_{\ell-1}) + F_{\nu_J}(w_{\ell-1})(w_\ell - w_{\ell-1}), \quad F_{\nu_J}(w_\ell) = F_{\nu_J}(w_{\ell-1}) + \nu_J(\{w_\ell\}).$$
2. If $(\varphi_{\nu_J}(w_\ell) - \varphi_{\mu_I}(w_\ell))(\varphi_{\nu_J}(w_{\ell-1}) - \varphi_{\mu_I}(w_{\ell-1})) < 0$, then, computing the intersection point between φ_{μ_I} and φ_{ν_J} on $[w_{\ell-1}, w_\ell]$ and the increase of the slope of $\varphi_{\mu_I} \vee \varphi_{\nu_J}$ at this point, we obtain that
$$\mu_I \vee_{\rm dcx} \nu_J \left(\left\{ \frac{|\varphi_{\nu_J}(w_\ell) - \varphi_{\mu_I}(w_\ell)| w_{\ell-1} + |\varphi_{\nu_J}(w_{\ell-1}) - \varphi_{\mu_I}(w_{\ell-1})| w_\ell}{|\varphi_{\nu_J}(w_\ell) - \varphi_{\mu_I}(w_\ell)| + |\varphi_{\nu_J}(w_{\ell-1}) - \varphi_{\mu_I}(w_{\ell-1})|} \right\} \right)$$
$$= \mu_I \vee_{\rm dcx} \nu_J((w_{\ell-1}, w_\ell)) = \frac{|\varphi_{\mu_I}(w_\ell) - \varphi_{\mu_I}(w_{\ell-1}) - \varphi_{\nu_J}(w_\ell) + \varphi_{\nu_J}(w_{\ell-1})|}{w_\ell - w_{\ell-1}}$$
$$= \frac{|\varphi_{\nu_J}(w_\ell) - \varphi_{\mu_I}(w_\ell)| + |\varphi_{\nu_J}(w_{\ell-1}) - \varphi_{\mu_I}(w_{\ell-1})|}{w_\ell - w_{\ell-1}} = |F_\nu(w_{\ell-1}) - F_\mu(w_{\ell-1})|.$$
Otherwise, we either have $\forall t \in (w_{\ell-1}, w_\ell)$ $\varphi_{\mu_I \vee_{\rm dcx} \nu_J}(t) = \varphi_{\mu_I}(t)$ or $\forall t \in (w_{\ell-1}, w_\ell)$ $\varphi_{\mu_I \vee_{\rm dcx} \nu_J}(t) = \varphi_{\nu_J}(t)$, and the slope of $\varphi_{\mu_I} \vee \varphi_{\nu_J}$ is constant on $(w_{\ell-1}, w_\ell)$ so that $\mu_I \vee_{\rm dcx} \nu_J((w_{\ell-1}, w_\ell)) = 0$.
3. The mass $\mu_I \vee_{\rm dcx} \nu_J(\{w_\ell\})$ is obtained in the following way.
   (i) If $\varphi_{\mu_I}(w_\ell) = \varphi_{\nu_J}(w_\ell)$, then the slopes of $\varphi_{\mu_I} \vee \varphi_{\nu_J}$ on $(w_{\ell-1}, w_\ell)$ and on $(w_\ell, w_{\ell+1})$ (convention $w_{L+1} = w_L + 1$) are respectively equal to $F_{\mu_I}(w_{\ell-1}) \wedge F_{\nu_J}(w_{\ell-1})$ and $F_{\mu_I}(w_\ell) \vee F_{\nu_J}(w_\ell)$ so that
   $$\mu_I \vee_{\rm dcx} \nu_J(\{w_\ell\}) = F_{\mu_I}(w_\ell) \vee F_{\nu_J}(w_\ell) - F_{\mu_I}(w_{\ell-1}) \wedge F_{\nu_J}(w_{\ell-1}).$$
   (ii) If $\varphi_{\mu_I}(w_\ell) > \varphi_{\nu_J}(w_\ell)$ then $\varphi_{\mu_I} \vee \varphi_{\nu_J}$ is equal to φ_{μ_I} on a neighbourhood of w_ℓ so that $\mu_I \vee_{\rm dcx} \nu_J(\{w_\ell\}) = \mu_I(\{w_\ell\})$.
   (iii) If $\varphi_{\mu_I}(w_\ell) < \varphi_{\nu_J}(w_\ell)$, then, in a symmetric way, $\mu_I \vee_{\rm dcx} \nu_J(\{w_\ell\}) = \nu_J(\{w_\ell\})$.

Of course, the slope of $\varphi_{\mu_I} \vee \varphi_{\nu_J}$ is equal to 0 on $(-\infty, w_1)$ and to 1 on $(w_L, +\infty)$ so that $\mu_I \vee_{\rm dcx} \nu_J((-\infty, w_1) \cup (w_L, +\infty)) = 0$. Moreover, $\mu_I \vee_{\rm dcx} \nu_J(\{w_1\}) = \mu_I(\{w_1\}) \vee \nu_J(\{w_1\})$. On the other hand, by (2.1), $\varphi_{\mu_I}(w_L) = w_L - \sum_{i=1}^I p_i x_i$, $\varphi_{\nu_J}(w_L) = w_L - \sum_{j=1}^J q_j y_j$ and

when either $\sum_{i=1}^I p_i x_i < \sum_{j=1}^J q_j y_j$ and $\mu_I(\{w_L\}) = 0$ or $\sum_{i=1}^I p_i x_i > \sum_{j=1}^J q_j y_j$ and $\nu_J(\{w_L\}) = 0$, then $\mu_I \vee_{\rm dcx} \nu_J(\{w_L\}) = 0$. Otherwise, and in particular when $\sum_{i=1}^I p_i x_i = \sum_{j=1}^J q_j y_j$, $\mu_I \vee_{\rm dcx} \nu_J(\{w_L\}) > 0$. Algorithm 1 below implements these ideas to compute $\mu_I \vee_{\rm dcx} \nu_J$.

5.2. **Algorithm for the calculation of the infimum of two probability measures with finite support.** We consider the same setting as in Subsection 5.1. The equality $\mu_I \wedge_{\rm icx} \nu_J = \overline{\bar{\mu}_I \wedge_{\rm dcx} \bar{\nu}_J}$ implies that it is enough to focus on the computation of $\mu_I \wedge_{\rm dcx} \nu_J$. Recall that $w_1 < w_2 < \ldots < w_L$ with $L \geq I \vee J$ are such that $\{w_1, \ldots, w_L\} = \{x_1, \ldots, x_I\} \cup \{y_1, \ldots, y_J\}$. Since convex hulls will appear in the characterization, when $d = 1$, of the Wasserstein projections considered in Section 3, rather than using the characterization of $\mu_I \wedge_{\rm dcx} \nu_J$ deduced from Lemma 4.2 which would lead to some algorithm similar to the previous one, we compute the convex hull $\varphi_{\mu_I \wedge_{\rm dcx} \nu_J}$ of $\varphi_{\mu_I} \wedge \varphi_{\nu_J}$. To do so, we observe that its epigraph is the convex hull of the union of the epigraphs of φ_{μ_I} and φ_{ν_J}. Thus, we can use Andrew's monotone chain convex hull algorithm. We describe here briefly how it works in our case. We exclude the trivial case $L = 1$ and assume that $L \geq 2$. The principle is to calculate for $\ell \in \{2, \ldots L\}$ the convex hull of $\{(y, w) : w \in [w_1, w_\ell], y \geq \varphi_{\mu_I} \wedge \varphi_{\nu_J}(w)\}$. It is fully described by indices $i_\ell(1) < \cdots < i_\ell(n_\ell)$ such that this convex hull is equal to $\cup_{n=1}^{n_\ell - 1} \{(y, w) : w \in [w_{i_\ell(n)}, w_{i_\ell(n+1)}], y \geq \frac{w_{i_\ell(n+1)} - w}{w_{i_\ell(n+1)} - w_{i_\ell(n)}} \varphi_{\mu_I} \wedge \varphi_{\nu_J}(w_{i_\ell(n)}) + \frac{w - w_{i_\ell(n)}}{w_{i_\ell(n+1)} - w_{i_\ell(n)}} \varphi_{\mu_I} \wedge \varphi_{\nu_J}(w_{i_\ell(n+1)})\}$. Obviously $n_2 = 2$, $i_2(1) = 1$, $i_2(2) = 2$ and from this initialization, we proceed by induction. To calculate the indices $i_{\ell+1}$ from the indices i_ℓ, Andrew's monotone chain convex hull algorithm works as follows: $\{i_{\ell+1}(1), \ldots, i_\ell(n_{\ell+1})\}$ is equal to $\{1\} \cup \{i_\ell(n) : 2 \leq n \leq n_\ell, \frac{\varphi_{\mu_I} \wedge \varphi_{\nu_J}(w_{\ell+1}) - \varphi_{\mu_I} \wedge \varphi_{\nu_J}(w_{i_\ell(n)})}{w_{\ell+1} - w_{i_\ell(n)}} > \frac{\varphi_{\mu_I} \wedge \varphi_{\nu_J}(w_{i_\ell(n)}) - \varphi_{\mu_I} \wedge \varphi_{\nu_J}(w_{i_\ell(n-1)})}{w_{i_\ell(n)} - w_{i_\ell(n-1)}}\} \cup \{\ell + 1\}$.

This means that we remove all the previous points that would give a nonincreasing slope. Once we have calculated the indices, the weights of the measure $\mu_I \wedge_{\rm dcx} \nu_J$ are obtained as the difference of two successive slopes. This is presented in Algorithm 2.

5.3. **Examples.** In this subsection, we discuss on different examples the number K of points in the support of $\mu_I \vee \nu_J$ (resp. $\mu_I \wedge \nu_J$), in the case $\sum_{i=1}^I p_i x_i = \sum_{j=1}^J q_j y_j$. We show that the maximal value of K is $I + J - 1$ (resp. $I + J - 2$).

We start by studying examples for $\mu_I \vee \nu_J$ and we suppose for convenience that $I \leq J$.

1. If $I = 1$, then necessarily $\mu_I \leq_{\rm cx} \nu_J$ so that $\mu_I \vee \nu_J = \nu_J$ and $K = J$,
2. If $I \geq 2$, then the support of $\mu_I \vee \nu_J$ contains at least the two distinct points $x_1 \wedge y_1$ and $x_I \vee y_J$. Now either $\mu_I \vee \nu_J \in \{\mu_I, \nu_J\}$ or neither $\mu_I \geq_{\rm cx} \nu_J$ nor $\nu_J \geq_{\rm cx} \mu_I$ which implies that $K \geq 3$: the sign of $\varphi_{\nu_J} - \varphi_{\mu_I}$ changes on the interval $(x_1 \wedge y_1, x_I \vee y_J)$, which implies that the cumulative distribution function F equal to the slope of $\varphi_{\mu_I} \vee \varphi_{\nu_J}$ is not constant on this interval. In particular, when $I \geq 3$ then $K \geq 3$. Notice that K may be equal to 3 even if I is arbitrarily large. For instance, for the centered probability measures

$$\mu_I = \frac{1}{3}\delta_{-1} + \frac{2}{3(I-1)} \sum_{i=2}^{I} \delta_{\frac{i-2}{I-2}} \text{ and } \nu_I = \frac{2}{3(I-1)} \sum_{i=1}^{I-1} \delta_{\frac{1-I+i}{I-2}} + \frac{1}{3}\delta_1$$

with $I \geq 3$, $\mu_I \vee \nu_I = \frac{1}{3}(\delta_{-1} + \delta_0 + \delta_1)$. Indeed $\varphi_{\mu_I}(t) = \frac{1+t}{3}$ for $t \in [-1, 0]$ whereas $\varphi_{\nu_I}(0) = \frac{2}{3(I-1)} \sum_{i=1}^{I-1} \frac{i-1}{I-2} = \frac{1}{3}$ so that $\varphi_{\nu_I}(t) = \frac{1+2t}{3}$ for $t \in [0, 1]$. With the convexity of φ_{μ_I} and φ_{ν_I} and the equality of these functions at the extreme points -1 and

Algorithm 1 Calculate $z_1 < z_2 < \ldots < z_K$ and $(\varrho_1, \ldots, \varrho_K) \in (0,1]^K$ s.t. $\mu_I \vee_{\mathrm{dcx}} \nu_J = \sum_{k=1}^K \varrho_k \delta_{z_k}$

$max \leftarrow x_I \vee y_J$
$x_{I+1} \leftarrow max$
$p_{I+1} \leftarrow 0$
$y_{J+1} \leftarrow max$
$q_{J+1} \leftarrow 0$
$i \leftarrow 1$
$j \leftarrow 1$
$k \leftarrow 1$
$w \leftarrow x_1 \wedge y_1 - 1$
$F_\mu \leftarrow 0$
$\varphi_\mu \leftarrow 0$
$F_\nu \leftarrow 0$
$\varphi_\nu \leftarrow 0$
while $w < max$ **do**
  $\widetilde{w} \leftarrow x_i \wedge y_j$
  $\widetilde{F}_\mu \leftarrow F_\mu + 1_{\{\widetilde{w}=x_i\}} p_i$
  $\widetilde{\varphi}_\mu \leftarrow \varphi_\mu + F_\mu(\widetilde{w} - w)$
  $\widetilde{F}_\nu \leftarrow F_\nu + 1_{\{\widetilde{w}=y_j\}} q_j$
  $\widetilde{\varphi}_\nu \leftarrow \varphi_\nu + F_\nu(\widetilde{w} - w)$
  if $(\widetilde{\varphi}_\nu - \widetilde{\varphi}_\mu)(\varphi_\nu - \varphi_\mu) < 0$ **then**
    $z_k \leftarrow \frac{|\widetilde{\varphi}_\nu - \widetilde{\varphi}_\mu| w + |\varphi_\nu - \varphi_\mu| \widetilde{w}}{|\widetilde{\varphi}_\nu - \widetilde{\varphi}_\mu| + |\varphi_\nu - \varphi_\mu|}$
    $\varrho_k \leftarrow |F_\nu - F_\mu|$
    $k \leftarrow k + 1$
  else if $\widetilde{\varphi}_\nu - \widetilde{\varphi}_\mu = 0$ **then**
    $z_k \leftarrow \widetilde{w}$
    $\varrho_k \leftarrow (\widetilde{F}_\nu \vee \widetilde{F}_\mu) - (F_\nu \wedge F_\mu)$
    $k \leftarrow k + 1$
  end if
  if $(\widetilde{\varphi}_\mu - \widetilde{\varphi}_\nu) p_i 1_{\{\widetilde{w}=x_i\}} > 0$ **then**
    $z_k \leftarrow \widetilde{w}$
    $\varrho_k \leftarrow p_i$
    $k \leftarrow k + 1$
  end if
  if $(\widetilde{\varphi}_\nu - \widetilde{\varphi}_\mu) q_j 1_{\{\widetilde{w}=y_j\}} > 0$ **then**
    $z_k \leftarrow \widetilde{w}$
    $\varrho_k \leftarrow q_j$
    $k \leftarrow k + 1$
  end if
  $i \leftarrow i + 1_{\{\widetilde{w}=x_i\}}$
  $j \leftarrow j + 1_{\{\widetilde{w}=y_j\}}$
  $w \leftarrow \widetilde{w}$
  $F_\mu \leftarrow \widetilde{F}_\mu$
  $\varphi_\mu \leftarrow \widetilde{\varphi}_\mu$
  $F_\nu \leftarrow \widetilde{F}_\nu$
  $\varphi_\nu \leftarrow \widetilde{\varphi}_\nu$
end while

Algorithm 2 Calculate $z_1 < z_2 < \ldots < z_K$ and $(\varrho_1, \ldots, \varrho_K) \in (0,1]^K$ s.t. $\mu_I \wedge_{\mathrm{dcx}} \nu_J = \sum_{k=1}^K \varrho_k \delta_{z_k}$

$max \leftarrow x_I \vee y_J$
$x_{I+1} \leftarrow max$
$p_{I+1} \leftarrow 0$
$y_{J+1} \leftarrow max$
$q_{J+1} \leftarrow 0$
$i \leftarrow 1$
$j \leftarrow 1$
$L \leftarrow \{\}$
$w \leftarrow x_1 \wedge y_1 - 1$
$F_\mu \leftarrow 0$
$\varphi_\mu \leftarrow 0$
$F_\nu \leftarrow 0$
$\varphi_\nu \leftarrow 0$
while $w < max$ **do**
  $\widetilde{w} \leftarrow x_i \wedge y_j$
  $\varphi_\mu \leftarrow \varphi_\mu + F_\mu(\widetilde{w} - w)$
  $\varphi_\nu \leftarrow \varphi_\nu + F_\nu(\widetilde{w} - w)$
  while $s(L) > 1$ **and** $\frac{\psi_{s(L)} - \psi_{s(L)-1}}{z_{s(L)} - z_{s(L)-1}} \geq \frac{\varphi_\mu \wedge \varphi_\nu - \psi_{s(L)}}{\widetilde{w} - z_{s(L)}}$ **do**
    Remove the last element of L.
  end while
  $w \leftarrow \widetilde{w}$
  $F_\mu \leftarrow F_\mu + 1_{\{\widetilde{w} = x_i\}} p_i$
  $F_\nu \leftarrow F_\nu + 1_{\{\widetilde{w} = y_j\}} q_j$
  Add $(\widetilde{w}, \varphi_\mu \wedge \varphi_\nu, 0, 0)$ to the list L.
  $i \leftarrow i + 1_{\{\widetilde{w} = x_i\}}$
  $j \leftarrow j + 1_{\{\widetilde{w} = y_j\}}$
end while
if $s(L) = 1$ **then**
  $\varrho_1 = 1$
else
  $F^0 \leftarrow 0$
  for $1 \leq k \leq s(L)$ **do**
    $F^k \leftarrow 1_{\{k < s(L)\}} \frac{\psi_{k+1} - \psi_k}{z_{k+1} - z_k} + 1_{\{k = s(L)\}}$
    $\varrho_k \leftarrow F^k - F^{k-1}$
  end for
end if
if $\varrho_{s(L)} = 0$ **then**
  Remove the last element of L.
end if
if $\varrho_1 = 0$ **then**
  Remove the first element of L.
end if

1 (deduced at point 1 from the equality of the expectations), one concludes that $\varphi_{\mu_I} \vee \varphi_{\nu_I}(t) = \frac{(1+t)^+ + t^+ + (t-1)^+}{3}$.

3. As explained in Subsection 5.1, we necessarily have $K \leq I + J - 1$. For $\alpha \in (\frac{2}{I}, \frac{2}{I-1})$,

$$\mu_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{x_i} \text{ with } x_i = \frac{(1+\alpha)i}{I+1} - \frac{\alpha}{2} \text{ and } \nu_{I+1} = \frac{1}{I+1}\sum_{i=0}^{I} \delta_{\frac{i}{I}},$$

the cardinality of the support of $\mu_I \vee \nu_{I+1}$ is $2I$. Indeed, it is easy to check that the condition $\alpha < \frac{2}{I-1}$ is necessary and sufficient to ensure $\frac{i-1}{I} < x_i < \frac{i}{I}$ for all $i \in \{1, \dots, I\}$. Under this condition, it is enough to check that for all $i \in \{1, \dots, I\}$, $\varphi_{\nu_{I+1}}(x_i) > \varphi_{\mu_I}(x_i)$ and for all $i \in \{1, \dots I-1\}$, $\varphi_{\mu_I}(\frac{i}{I}) > \varphi_{\nu_{I+1}}(\frac{i}{I})$ to ensure that F indeed takes the $2I+1$ possible different values. Notice that on $[\frac{i-1}{I}, \frac{i}{I}]$, $\varphi_{\nu_{I+1}}(t) = \frac{i}{I+1}\left(t - \frac{i-1}{2I}\right)$ and that on $[x_i, x_{i+1}]$, $\varphi_{\mu_I}(t) = \frac{i}{I}\left(t - \frac{(1+\alpha)(i+1)}{2(I+1)} + \frac{\alpha}{2}\right)$. The inequality $\varphi_{\nu_{I+1}}(x_i) > \varphi_{\mu_I}(x_i)$ multiplied by $\frac{2I(I+1)^2}{i}$ therefore writes $(2-\alpha(I-1))(I+1-i) > 0$. On the other hand, the inequality $\varphi_{\mu_I}(\frac{i}{I}) > \varphi_{\nu_{I+1}}(\frac{i}{I})$ multiplied by $\frac{2I^2(I+1)}{i}$ rewrites $(\alpha I - 2)(I - i) > 0$.

4. When the probability measures μ_I and ν_J are uniform : $p_i = \frac{1}{I}$ for all $i \in \{1, \dots, I\}$ and $q_j = \frac{1}{J}$ for all $j \in \{1, \dots, J\}$. Then $K \leq I + J - \gcd(I, J)$ (in particular $K \leq I$ when $J = I$) since the cardinality of $\{\frac{i}{I} : i \in \{0, \dots, I\}\} \cup \{\frac{j}{J} : j \in \{0, \dots, J\}\}$ is $I + J + 1 - \gcd(I, J)$. Let $I' = I/\gcd(I, J)$ and $J' = J/\gcd(I, J)$. These integers are prime together and we have

$$\frac{i}{I} = \frac{j}{J} \iff iJ' = jI' \iff i = \kappa I' \text{ and } j = \kappa J',$$

with $\kappa \in \{0, \dots, \gcd(I, J)\}$ when $0 \leq i \leq I$ and $0 \leq j \leq J$. Therefore, the cardinal of $\{\frac{i}{I} : i \in \{0, \dots, I\}\} \cap \{\frac{j}{J} : j \in \{0, \dots, J\}\}$ is $1 + \gcd(I, J)$, which gives the claim.

We now discuss the number K of points in the support of $\mu_I \wedge \nu_J$ and suppose again $I \leq J$. We assume $I \geq 2$, otherwise $\mu_I \wedge \nu_J$ has clearly one element ($K = 1$) and is the Dirac mass at x_1.

1′. If $I = J$, $x_i = y_i$ and $p_i = q_i$ for $1 \leq i \leq I$, we have $\mu_I \wedge \nu_J = \mu_I$ and thus $K = I$.

2′. If $I = 2$, $x_1 \leq \min_{j=1,\dots,J} y_j$ and $x_2 \geq \max_{j=1,\dots,J} y_j$, then for any $j \in \{1, \dots, J\}$, we have $y_j = \frac{x_2 - y_j}{x_2 - x_1} x_1 + \frac{y_j - x_1}{x_2 - x_1} x_2$. Thus, if $p_1 = \sum_{j=1}^{J} q_j \frac{x_2 - y_j}{x_2 - x_1}$ and $p_2 = \sum_{j=1}^{J} q_j \frac{y_j - x_1}{x_2 - x_1}$, we have $\nu_J \leq_{\text{cx}} \mu_I$ and thus $K = J$.

3′. We have $K \leq I + J - 2$. It is clear that $K \leq I + J$ since the support of $\mu_I \wedge \nu_J$ is included in $\{x_1, \dots, x_I\} \cup \{y_1, \dots, y_J\}$. If $x_1 < y_1$, then x_1 is not in the support of $\mu_I \wedge \nu_J$ since $\varphi_{\mu_I \wedge \nu_J}(y_1) = 0$. Thus, at most one element of $\{x_1, y_1\}$ is in the support of $\mu_I \wedge \nu_J$. If $x_I > y_J$, then x_I is not in the support of $\mu_I \wedge \nu_J$, otherwise we would have $\varphi'_{\mu_I \wedge \nu_J}(x_I-) < 1$ and thus

$$\varphi_{\mu_I \wedge \nu_J}(y_J) > \varphi_{\mu_I \wedge \nu_J}(x_I) + y_J - x_I = x_I - \sum_{i=1}^{I} p_i x_i + y_J - x_I = y_J - \sum_{j=1}^{J} q_j y_j = \varphi_{\nu_J}(y_J).$$

Thus, at most one element of $\{x_I, y_J\}$ is in the support of $\mu_I \wedge \nu_J$. We eventually get $K \leq I + J - 2$. This bound is attained by the previous example: for $\alpha \in (\frac{2}{I}, \frac{2}{I-1})$,

$\mu_I = \frac{1}{I}\sum_{i=1}^I \delta_{x_i}$ with $x_i = \frac{(1+\alpha)i}{I+1} - \frac{\alpha}{2}$ and $\nu_{I+1} = \frac{1}{I+1}\sum_{i=0}^I \delta_{\frac{i}{I}}$, the cardinality of the support of $\mu_I \wedge \nu_{I+1}$ is $2I-1$. Let us recall that we have already shown that for all $i \in \{1,\dots I\}$, $x_i \in (\frac{i-1}{I}, \frac{i}{I})$ and $\varphi_{\nu_{I+1}}(x_i) > \varphi_{\mu_I}(x_i)$ and, for all $i \in \{1,\dots I-1\}$, $\varphi_{\mu_I}(\frac{i}{I}) > \varphi_{\nu_{I+1}}(\frac{i}{I})$. Thus, to prove that the support of $\mu_I \wedge \nu_J$ is $\{x_i, 1 \le i \le I\} \cup \{\frac{i}{I}, 1 \le i \le I-1\}$, it is necessary and sufficient to prove that

$$\frac{\varphi_{\mu_I}(x_i) - \varphi_{\nu_{I+1}}\left(\frac{i-1}{I}\right)}{x_i - \frac{i-1}{I}} < \frac{\varphi_{\nu_{I+1}}\left(\frac{i}{I}\right) - \varphi_{\mu_I}(x_i)}{\frac{i}{I} - x_i},\ i \in \{1,\dots,I\},$$
$$\frac{\varphi_{\nu_{I+1}}\left(\frac{i}{I}\right) - \varphi_{\mu_I}(x_i)}{\frac{i}{I} - x_i} < \frac{\varphi_{\mu_I}(x_{i+1}) - \varphi_{\nu_{I+1}}\left(\frac{i}{I}\right)}{x_{i+1} - \frac{i}{I}},\ i \in \{1,\dots,I-1\}.$$

We observe that $\varphi_{\mu_I}(x_i) = \frac{1+\alpha}{2I(I+1)} i(i-1)$ and $\varphi_{\nu_{I+1}}\left(\frac{i}{I}\right) = \frac{1}{2I(I+1)} i(i+1)$. After simple but tedious calculations, we can check that the first condition is equivalent to $0 < (2-\alpha(I-1))(I+1-i)$, which is true since $\alpha < \frac{2}{I-1}$. Similarly, the second condition is equivalent to $(I-i)(2-\alpha I) < 0$, which is true since $\alpha > \frac{2}{I}$.

4′. If $I \ge 2$, the support of $\mu_I \wedge \nu_J$ has at least two elements, i.e. $K \ge 2$. Since $m := \sum_{i=1}^I p_i x_i = \sum_{j=1}^J q_j y_j$ and the elements x_i (resp. y_j) are distinct, we have $\varphi_{\mu_I}(m) > 0$ (resp. $\varphi_{\nu_J}(m) > 0$). If $K = 1$, we would have then necessary $\mu_I \wedge \nu_J = \delta_m$, which would be in contradiction with the previous statement. Notice that K may be equal to 2 even if I is arbitrarily large. For $I \ge 2$, we take when $I-1 = 2k$ is even

$$\mu_I = \frac{1}{I-1}\sum_{i=1}^k \frac{1}{2}\left(\delta_{-1-\frac{i}{k}} + \delta_{-1+\frac{i}{k}}\right) + \frac{1}{2}\delta_1,\ \nu_I = \frac{1}{2}\delta_{-1} + \frac{1}{I-1}\sum_{i=1}^k \frac{1}{2}\left(\delta_{1-\frac{i}{k}} + \delta_{1+\frac{i}{k}}\right),$$

and when $I-1 = 2k+1$ is odd,

$$\mu_I = \frac{1}{I}\sum_{i=0}^k \frac{1}{2}\left(\delta_{-1-\frac{i}{k}} + \delta_{-1+\frac{i}{k}}\right) + \frac{1}{2}\delta_1,\ \nu_I = \frac{1}{2}\delta_{-1} + \frac{1}{I}\sum_{i=0}^k \frac{1}{2}\left(\delta_{1-\frac{i}{k}} + \delta_{1+\frac{i}{k}}\right).$$

By construction, the support of μ_I (resp. ν_I) has exactly I points. Moreover, one easily checks that $\mu_I \wedge \nu_I = \frac{1}{2}(\delta_{-1} + \delta_1)$.

5.4. **Irreducible components.** Specializing Theorem A.4 of Beiglböck and Juillet (2016) and its proof to the case of probability measures, one has the following decomposition.

Proposition 5.1. *Let $\mu \le_{\rm cx} \nu$ be two distinct probability measures such that $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$. There exists a countable family of disjoint intervals $(\underline{t}_n, \bar{t}_n)$, $1 \le n \le N$, with $N \in \mathbb{N}^* \cup \{\infty\}$ such that*

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

We have $\mu((\underline{t}_n, \bar{t}_n)) > 0$ for any $1 \le n \le N$.

Let $E = \cap_{1\le n\le N}\{(-\infty, \underline{t}_n] \cup [\bar{t}_n, +\infty)\}$, $\mu^0(dx) = 1_E(x)\mu(dx)$, $\mu^n(dx) = 1_{(\underline{t}_n, \bar{t}_n)}(x)\mu(dx)$ so that $\mu = \sum_{n=0}^N \mu^n$. Then, there exists a unique decomposition $\nu = \sum_{n=0}^N \nu^n$ such that $\nu^n(\mathbb{R}) = \mu^n(\mathbb{R})$ and $\frac{\mu^n}{\mu^n(\mathbb{R})} \le_{\rm cx} \frac{\nu^n}{\mu^n(\mathbb{R})}$, for $1 \le n \le N$ and, when $\mu^0(\mathbb{R}) > 0$, $\frac{\mu^0}{\mu^0(\mathbb{R})} \le_{\rm cx} \frac{\nu^0}{\mu^0(\mathbb{R})}$. It is given by $\nu^0 = \mu^0$ and, for $1 \le n \le N$, by

$$\nu^n(dx) = 1_{(\underline{t}_n, \bar{t}_n)}(x)\nu(dx) + (F_\nu(\underline{t}_n) - F_\mu(\underline{t}_n))\delta_{\underline{t}_n}(dx) + (F_\mu(\bar{t}_n-) - F_\nu(\bar{t}_n-))\delta_{\bar{t}_n}(dx).$$

Last, if P is a martingale coupling between μ and ν,

$$P(dx,dy) = \mu^0(dx)\delta_x(dy) + \sum_{n=1}^{N} \mu((\underline{t}_n, \bar{t}_n))P_n(dx,dy),$$

where P_n is a martingale coupling between $\frac{\mu^n}{\mu^n(\mathbb{R})}$ and $\frac{\nu^n}{\mu^n(\mathbb{R})}$.

The intervals $((\underline{t}_n, \bar{t}_n))_{1\le n\le N}$ are called the irreducible components of the couple (μ,ν). According to Corollary A.7 of Beiglböck and Juillet (2016), solving a martingale optimal transport problem between μ and ν is equivalent to solving the martingale optimal transport problem with the same payoff function between $\frac{\mu^n}{\mu^n(\mathbb{R})}$ and $\frac{\nu^n}{\mu^n(\mathbb{R})}$ for each $1 \le n \le N$. This motivates our interest in studying the irreducible components of $(\mu_I \wedge \nu_J, \nu_J)$ and $(\mu_I, \mu_I \vee \nu_J)$: to solve numerically the corresponding discrete MOT problem, this replace a linear programming problem by N linear programming problems of smaller dimension, where N is the corresponding number of irreducible components.

We now consider two probability measures with finite supports $\mu_I = \sum_{i=1}^I p_i\delta_{x_i}$ and $\nu_J = \sum_{j=1}^J q_j\delta_{y_j}$ with $x_1 < \cdots < x_I$, $y_1 < \cdots < y_J$, $p_i > 0$ for $i \in \{1,\dots,I\}$ and $q_j > 0$ for $j \in \{1,\dots,J\}$ and $\sum_{i=1}^I p_i x_i \le \sum_{j=1}^J q_j y_j$. Since, by Lemma 4.1, $\int_{\mathbb{R}} x\mu_I \vee_{\mathrm{dcx}} \nu_J(dx) = \int_{\mathbb{R}} x\mu_I(dx)$ and $\int_{\mathbb{R}} x\mu_I \wedge_{\mathrm{dcx}} \nu_J(dx) = \int_{\mathbb{R}} x\nu_J(dx)$, $\mu_I \le_{\mathrm{cx}} \mu_I \vee_{\mathrm{dcx}} \nu_J$ and $\mu_I \wedge_{\mathrm{dcx}} \nu_J \le_{\mathrm{cx}} \nu_J$. We are interested in describing precisely and computing the irreducible components of $(\mu_I, \mu_I \vee_{\mathrm{dcx}} \nu_J)$ (resp. $(\mu_I \wedge_{\mathrm{dcx}} \nu_J, \nu_J)$). Of course, there are finitely many components. The computation of the irreducible components of $(\mu_I, \mu_I \vee_{\mathrm{icx}} \nu_J)$ (resp. $(\mu_I \wedge_{\mathrm{icx}} \nu_J, \nu_J)$) when $\sum_{i=1}^I p_i x_i > \sum_{j=1}^J q_j y_j$ is easily deduced from the equality $\mu_I \vee_{\mathrm{icx}} \nu_J = \overline{\bar{\mu}_I \vee_{\mathrm{dcx}} \bar{\nu}_J}$ (resp. $\mu_I \wedge_{\mathrm{icx}} \nu_J = \overline{\bar{\mu}_I \wedge_{\mathrm{dcx}} \bar{\nu}_J}$).

Irreducible components of $(\mu_I, \mu_I \vee_{\mathrm{dcx}} \nu_J)$ when $\sum_{i=1}^I p_i x_i \le \sum_{j=1}^J q_j y_j$. The irreducible components are defined as the largest intervals on which $\varphi_{\mu_I} < \varphi_{\mu_I \vee_{\mathrm{dcx}} \nu_J}$. We have $\mu_I \vee_{\mathrm{dcx}} \nu_J = \sum_{k=1}^K \varrho_k \delta_{z_k}$ with $z_1 < \cdots < z_K$, $\varrho_k > 0$, and Algorithm 1 calculates the points z_k and the weights ϱ_k. Among these points, we focus on the points such that $\varphi_{\mu_I}(z_k) = \varphi_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(z_k)$. Thus, we define $v_1 < v_2 < \ldots < v_M$ such that

$$\{v_1,\dots,v_M\} = \{z_k : 1 \le k \le K \text{ and } \varphi_{\mu_I}(z_k) = \varphi_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(z_k)\}. \tag{5.1}$$

For $t \ge z_K$, $\varphi_{\mu_I}(t) \le \varphi_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(t) = t - \sum_{i=1}^I p_i x_i$. Since the function φ_{μ_I} is convex and equal to $t - \sum_{i=1}^I p_i x_i$ for $t \ge x_I$, one deduces that $\varphi_{\mu_I}(t) = t - \sum_{i=1}^I p_i x_i$ for $t \ge z_K$. Since $\varphi_{\mu_I}(t) = \varphi_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(t) = 0$ for $t \le z_1$ and we have $M \ge 1$ and all the irreducible components lie in (z_1, z_K). If $K = 1$, $\mu_I = \mu_I \vee_{\mathrm{dcx}} \nu_J = \delta_{x_1}$ and there is no irreducible component. Otherwise, we have $M \ge 2$, which we assume now.

Remark 5.2. *When $M \ge 3$, looking at the irreducible components reduces the dimension of the discrete MOT problem. We notice that if $I \ge 2$ and $\sum_{i=1}^I p_i x_i = \sum_{j=1}^J q_j y_j$, then $M = 2 \implies \mu_I \le_{\mathrm{cx}} \nu_J$. In fact, $M = 2$ implies that $\varphi_{\mu_I}(z_k) < \varphi_{\nu_J}(z_k)$ for $2 \le k \le K-1$. Since $0 = \varphi_{\mu_I}(z_1) \ge \varphi_{\nu_J}(z_1)$ and $z_K - \sum_{i=1}^I p_i x_i = \varphi_{\mu_I}(z_K) \ge \varphi_{\nu_J}(z_K) \ge \left(z_K - \sum_{i=1}^I p_i x_i\right)^+$ by Jensen's inequality, we get $\varphi_{\mu_I}(z_1) = \varphi_{\nu_J}(z_1)$ and $\varphi_{\mu_I}(z_K) = \varphi_{\nu_J}(z_K)$ and thus $\mu_I \le_{\mathrm{cx}} \nu_J$.*

We have $\varphi_{\mu_I} \le \varphi_{\mu_I} \vee \varphi_{\nu_J} = \varphi_{\mu_I \vee_{\mathrm{dcx}} \nu_J}$ and, for any $m \in \{1,\dots,M\}$, $\varphi_{\mu_I}(v_m) = \varphi_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(v_m)$. We can therefore compare the derivatives and get

$$F_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(v_m-) \le F_{\mu_I}(v_m-),\ F_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(v_m) \ge F_{\mu_I}(v_m).$$

Let $m \leq M-1$ and k_m be such that $v_m = z_{k_m}$. If $F_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(v_m) = F_{\mu_I}(v_m)$ then, since $\varphi_{\mu_I \vee_{\mathrm{dcx}} \nu_J} \geq \varphi_{\mu_I}$ and the derivative $F_{\mu_I \vee_{\mathrm{dcx}} \nu_J}$ of $\varphi_{\mu_I} \vee \varphi_{\nu_J}$ is constant and equal to $F_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(v_m)$ on (z_{k_m}, z_{k_m+1}) whereas the one F_{μ_I} of φ_{μ_I} is nondecreasing, one has $\forall t \in [z_{k_m}, z_{k_m+1}]$, $\varphi_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(t) = \varphi_{\mu_I}(t)$ and, in particular, $k_{m+1} = k_m + 1$ and $\mu_I((v_m, v_{m+1})) = 0$. In a symmetric way, if $F_{\mu_I}(v_{m+1}-) = F_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(v_{m+1}-)$, then

$$k_m = k_{m+1} - 1,\ \mu_I((v_m, v_{m+1})) = 0 \text{ and } \varphi_{\mu_I}(t) = \varphi_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(t) \text{ for } t \in [v_m, v_{m+1}].$$

Thus, $\mu_I((v_m, v_{m+1})) > 0$ implies that $F_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(v_m) > F_{\mu_I}(v_m)$ and $F_{\mu_I}(v_{m+1}-) > F_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(v_{m+1}-)$, which gives that $\forall t \in (v_m, v_{m+1})$, $\varphi_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(t) > \varphi_{\mu_I}(t)$.

Proposition 5.3. *We assume $\sum_{i=1}^I p_i x_i \leq \sum_{j=1}^J q_j y_j$ so that $\mu_I \leq_{\mathrm{cx}} \mu_I \vee_{\mathrm{dcx}} \nu_J$. Let $v_1 < v_2 < \ldots < v_M$ be defined by (5.1). We have*

$$\{t \in \mathbb{R}, \varphi_{\mu_I}(t) < \varphi_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(t)\} = \bigcup_{\substack{1 \leq m \leq M-1, \\ \mu_I((v_m, v_{m+1}))>0}} (v_m, v_{m+1}),$$

i.e. the intervals (v_m, v_{m+1}) such that $\mu_I((v_m, v_{m+1})) > 0$ are the irreducible components of $(\mu_I, \mu_I \vee_{\mathrm{dcx}} \nu_J)$. Moreover $\mu_I^0 = \sum_{m=1}^M \mu_I(\{v_m\})\delta_{v_m}$ and any martingale coupling P between μ_I and $\mu_I \vee_{\mathrm{dcx}} \nu_J$ is such that for all $m \in \{1, \ldots, M\}$, $P(\{(v_m, v_m)\}) = \mu_I(\{v_m\})$ and for all $m \in \{1, \ldots, M-1\}$, $P((v_m, v_{m+1}) \times [v_m, v_{m+1}]) = \mu_I((v_m, v_{m+1}))$.

Any martingale optimal transport problem between the marginals μ_I and $\mu_I \vee_{\mathrm{dcx}} \nu_J$ may be decomposed into the martingale optimal transport problems with the same payoff function but between the marginals $\frac{1}{\mu_I((v_m, v_{m+1}))} \sum_{i=1}^I p_i 1_{\{v_m < x_i < v_{m+1}\}} \delta_{x_i}$ and

$$\frac{1}{\mu_I((v_m, v_{m+1}))} \Bigg((F_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(v_m) - F_{\mu_I}(v_m))\delta_{v_m} + \sum_{k=1}^K \varrho_k 1_{\{v_m < z_k < v_{m+1}\}} \delta_{z_k} + (F_{\mu_I}(v_{m+1}-) - F_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(v_{m+1}-))\delta_{v_{m+1}} \Bigg)$$

for m such that $\mu_I((v_m, v_{m+1})) > 0$. Doing so, we replace one linear programming problem by N linear programming problems of smaller size, where $N = \sum_{m=1}^{M-1} 1_{\{\mu_I((v_m, v_{m+1}))>0\}}$ is the number of irreducible components.

Setting $\eta_{m-} = F_{\mu_I}(v_m-) - F_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(v_m-)$, $\eta_m = \mu_I(\{v_m\})$ and $\eta_{m+} = F_{\mu_I \vee_{\mathrm{dcx}} \nu_J}(v_m) - F_{\mu_I}(v_m)$, Algorithm 1 can be modified in the following way to compute $(v_m, \eta_{m-}, \eta_m, \eta_{m+})_{1 \leq m \leq M}$. First initialize $m \leftarrow 1$ before the while loop. In this while loop,

1. if $(\widetilde{\varphi}_\nu - \widetilde{\varphi}_\mu)(\varphi_\nu - \varphi_\mu) < 0$, add $v_m \leftarrow z_k$, $\eta_{m-} \leftarrow 1_{\{\varphi_\nu > \varphi_\mu\}} \varrho_k$, $\eta_m \leftarrow 0$, $\eta_{m+} \leftarrow 1_{\{\varphi_\nu < \varphi_\mu\}} \varrho_k$, $m \leftarrow m+1$ just before $k \leftarrow k+1$,
2. if $\widetilde{\varphi}_\nu - \widetilde{\varphi}_\mu = 0$, add $v_m \leftarrow \widetilde{w}$, $\eta_{m-} \leftarrow (F_\mu - F_\nu)^+$, $\eta_m \leftarrow \widetilde{F}_\mu - F_\mu$, $\eta_{m+} \leftarrow (\widetilde{F}_\nu - \widetilde{F}_\mu)^+$, $m \leftarrow m+1$ just before $k \leftarrow k+1$.
3. if $(\widetilde{\varphi}_\mu - \widetilde{\varphi}_\nu) p_i 1_{\{\widetilde{z} = x_i\}} > 0$, add $v_m \leftarrow \widetilde{w}$, $\eta_{m-} \leftarrow 0$, $\eta_m \leftarrow p_i$, $\eta_{m+} \leftarrow 0$, $m \leftarrow m+1$ just before $k \leftarrow k+1$.

Irreducible components of $(\mu_I \wedge_{\mathrm{dcx}} \nu_J, \nu_J)$ when $\sum_{i=1}^I p_i x_i \leq \sum_{j=1}^J q_j y_j$. We start with a simple observation.

Lemma 5.4. *For $j \in \{1, \ldots, J-1\}$, we have*

$$\exists t_0 \in (y_j, y_{j+1}),\ \varphi_{\nu_J}(t_0) = \varphi_{\mu_I \wedge_{\mathrm{dcx}} \nu_J}(t_0) \iff \forall t \in [y_j, y_{j+1}], \varphi_{\nu_J}(t) = \varphi_{\mu_I \wedge_{\mathrm{dcx}} \nu_J}(t).$$

Proof. The function φ_{ν_J} is affine on $[y_j, y_{j+1}]$. Since the function $\varphi_{\mu_I \wedge_{\mathrm{dcx}} \nu_J}$ is below φ_{ν_J} and convex, the equality of both functions at t_0 implies that they coincide on $[y_j, y_{j+1}]$. □

Now, we assume that $\sum_{i=1}^I p_i x_i \leq \sum_{j=1}^J q_j y_j$ so that $\mu_I \wedge_{\mathrm{dcx}} \nu_J \leq_{\mathrm{cx}} \nu_J$ and we consider $v_1 < \cdots < v_M$ such that

$$\{v_1, \ldots, v_M\} = \{y_j : 1 \leq j \leq J \text{ and } \varphi_{\nu_J}(y_j) = \varphi_{\mu_I \wedge_{\mathrm{dcx}} \nu_J}(y_j)\}, \tag{5.2}$$

and we denote by j_m the index such that $v_m = y_{j_m}$. We have $\varphi_{\nu_J}(y_1) = \varphi_{\mu_I \wedge_{\mathrm{dcx}} \nu_J}(y_1) = 0$ and thus $M \geq 1$. We have $\varphi_{\nu_J}(y_J) = y_J - \sum_{j=1}^J q_j y_j$ and $\varphi_{\mu_I \wedge_{\mathrm{dcx}} \nu_J}(y_J) \geq y_J - \sum_{j=1}^J q_j y_j$ since $\mu_I \wedge_{\mathrm{dcx}} \nu_J$ and μ_I have the same means. This gives $\varphi_{\nu_J}(y_J) = \varphi_{\mu_I \wedge_{\mathrm{dcx}} \nu_J}(y_J)$ and $v_M = y_J$. We consider $M \geq 2$ and $1 \leq m \leq M-1$. By Lemma 5.4, we either have $j_{m+1} = j_m + 1$ or $j_{m+1} > j_m + 1$ and $\varphi_{\nu_J}(t) > \varphi_{\mu_I \wedge_{\mathrm{dcx}} \nu_J}(t)$ for $t \in (v_m, v_{m+1})$. Since $j_{m+1} = j_m + 1 \iff \nu_J(v_m, v_{m+1}) = 0$, we get the following result.

Proposition 5.5. *We assume $\sum_{i=1}^I p_i x_i \leq \sum_{j=1}^J q_j y_j$ so that $\mu_I \wedge_{\mathrm{dcx}} \nu_J \leq_{\mathrm{cx}} \nu_J$. Let $v_1 < v_2 < \ldots < v_M$ be defined by (5.2). We have*

$$\{t \in \mathbb{R}, \varphi_{\mu_I \wedge_{\mathrm{dcx}} \nu_J}(t) < \varphi_{\mu_I}(t)\} = \bigcup_{\substack{1 \leq m \leq M-1, \\ \nu_J((v_m, v_{m+1}))>0}} (v_m, v_{m+1}).$$

i.e. the intervals (v_m, v_{m+1}) such that $\nu_J((v_m, v_{m+1})) > 0$ are the irreducible components of $(\mu_I \wedge_{\mathrm{dcx}} \nu_J, \nu_J)$. Moreover $\nu_J^0 = \sum_{m=1}^M \mu_I \wedge_{\mathrm{dcx}} \nu_J(\{v_m\})\delta_{v_m}$ and any martingale coupling P between $\mu_I \wedge_{\mathrm{dcx}} \nu_J$ and ν_J is such that for all $m \in \{1, \ldots, M\}$, $P(\{(v_m, v_m)\}) = \mu_I \wedge_{\mathrm{dcx}} \nu_J(\{v_m\})$ and for all $m \in \{1, \ldots, M-1\}$, $P((v_m, v_{m+1}) \times [v_m, v_{m+1}]) = \mu_I \wedge_{\mathrm{dcx}} \nu_J((v_m, v_{m+1}))$.

Any MOT problem between the marginals $\mu_I \wedge_{\mathrm{dcx}} \nu_J = \sum_{k=1}^K \varrho_k \delta_{z_k}$ and $\nu_J = \sum_{j=1}^J q_j \delta_{y_j}$ may be decomposed into the MOT problems with the same payoff function but between the marginals $\frac{1}{\sum_{k=1}^K \varrho_k 1_{\{v_m < z_k < v_{m+1}\}}} \sum_{k=1}^K \varrho_k 1_{\{v_m < z_k < v_{m+1}\}} \delta_{z_k}$ and

$$\frac{1}{\sum_{k=1}^K \varrho_k 1_{\{v_m < z_k < v_{m+1}\}}} \Big((F_{\nu_J}(v_m) - F_{\mu_I \wedge_{\mathrm{dcx}} \nu_J}(v_m))\delta_{v_m} + \sum_{j=1}^J q_j 1_{\{v_m < y_j < v_{m+1}\}} \delta_{y_j} + (F_{\mu_I \wedge_{\mathrm{dcx}} \nu_J}(v_{m+1}-) - F_{\nu_J}(v_{m+1}-))\delta_{v_{m+1}} \Big)$$

for m such that $\nu_J((v_m, v_{m+1})) > 0$. Doing so, we replace one linear programming problem by N linear programming problems of smaller size, where $N = \sum_{m=1}^{M-1} 1_{\{\nu_J((v_m, v_{m+1}))>0\}}$ is the number of irreducible components.

To compute the irreducible components, it is convenient to add two coordinates η_k and F_ν^k to each element in the list L in Algorithm 2. Instead of adding $(\widetilde{w}, \varphi_\mu \wedge \varphi_\nu, 0, 0)$ to the list, one should add $(\widetilde{w}, \varphi_\mu \wedge \varphi_\nu, 0, 0, 1_{\{\widetilde{w} = y_j, \varphi_\nu \leq \varphi_\mu\}} q_j, F_\nu)$. At the end of the algorithm, one computes $\{v_1, \ldots, v_m\}$ by $m = 1$, for $k = 1$ to $s(L)$ if $\eta_k > 0$, then $v_m = z_k$, $\mu_I \wedge_{\mathrm{dcx}} \nu_J(\{v_m\}) = \varrho_k$, $F_{\mu_I \wedge_{\mathrm{dcx}} \nu_J}(v_m-) - F_{\nu_J}(v_m-) = F^k - \varrho_k - F_\nu^k + \eta_k$, $F_{\nu_J}(v_m) - F_{\mu_I \wedge_{\mathrm{dcx}} \nu_J}(v_m) = F_\nu^k - F^k$, $m = m + 1$.

6. Numerical experiments

In this section, we show how the different constructions of probability measures presented in the paper can be used to approximate the solution of a martingale optimal transport problem. In each case, we have used as approximating measures the empirical measures

obtained of i.i.d. samples. This corresponds to the most generic practical case for the financial application: martingale asset prices are simulated to compute option prices by the Monte-Carlo method, and we want to reuse these simulations to compute bounds on these option prices. To solve numerically (1.1),we have used the linear programming solver GLPK[†]. With this solver, we are limited to use samples with $I = J \approx 100$: the dimension of the matrix r_{ij} is then about 10000. Other solvers may be more efficient, but since the dimension of the linear problem is the square of the number of points, the size of the samples which can be treated by exact solvers is limited. Approximated solvers such as the one proposed by Benamou et al. (2015) with entropic regularization seem to be a promising way to overcome this issue. The discussion on the maximal size of the samples that is possible to handle is of course interesting, but outside the scope of the paper. Here, we rather want to compare the different constructions and illustrate their relevance.

Last, we only consider in our numerical experiments approximations constructed from i.i.d. samples of the measures μ and ν. When the quantile functions F_μ^{-1} and F_ν^{-1} are known explicitly, it is possible to consider approximations with more evenly distributed points such as (1.2) or $\hat{\eta}_I = \frac{1}{I}\sum_{i=0}^{I-1} \delta_{F_\eta^{-1}((i+1/2)/I)}$, $\eta \in \{\mu, \nu\}$. The former approximation, which requires in addition to know explicitly the antiderivative of the quantile functions, preserves the convex order. The latter is likely to do the same, but one has to check this property by using for instance Corollary 2.2 before using the linear programming solver. This check has the same computational complexity as calculating the probability measure $\hat{\mu}_I \wedge \hat{\nu}_I$ equal to $\hat{\mu}_I$ when $\hat{\mu}_I \leq_{\mathrm{cx}} \hat{\nu}_I$, which ensures in any case the convex ordering. Besides, for the financial application, the quantile functions coming from the marginal distributions of a local volatility model or a stochastic volatility model are not known explictly in general. Moreover, when considering large i.i.d. samples with $I = J$ it is possible to use Baker's construction (1.2) on the re-ordered modifications of empirical measures. This approach is made precise in Example6.3 and leads to very evenly spaced points in Figure 7. We have noticed in our numerical experiments almost no difference for the same final sample size between this procedure and the direct use of (1.2) when the (integrated) quantile functions can be calculated explicitly. This is why we only present here approximations obtained from i.i.d. samples.

6.1. **An example with an explicit MOT.** We start with an example where the martingale optimal transport is known explicitly. Let $\varrho > 2$,

$$\mu(dx) = \frac{1}{2}1_{[-1,1]}(x)dx \text{ and } \nu(dx) = \frac{1}{4}1_{[-2,2]}(x)dx.$$

We consider the following martingale optimal transport problem:

$$\min_{\pi \in \Pi^M(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} |y-x|^\varrho \pi(dx,dy).$$

For any $\pi \in \Pi^M(\mu,\nu)$, we have $\int_{\mathbb{R}\times\mathbb{R}} |y-x|^2\pi(dx,dy) = \int_{\mathbb{R}} y^2\nu(dy) - \int_{\mathbb{R}} x^2\mu(dx) = 1$. For $\varrho > 2$, Jensen's inequality gives

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

[†]`https://www.gnu.org/software/glpk/`

We observe that

$$\pi^\star(dx,dy) = \frac{1}{2}1_{[-1,1]}(x)\frac{\delta_{x+1}(dy)+\delta_{x-1}(dy)}{2}dx \tag{6.1}$$

achieves this lower bound. Let $\pi \in \Pi^M(\mu,\nu)$ be an optimal coupling (i.e. $\int_{\mathbb{R}\times\mathbb{R}} |y-x|^\varrho \pi(dx,dy) = 1$). The equality condition in Jensen's inequality gives that $|y-x|$ is constant $\pi(dx,dy)$-a.s. and thus equal to 1. The martingale condition then gives $\pi = \pi^\star$. Let us note that for $\varrho < 2$, the same arguments with $\int_{\mathbb{R}\times\mathbb{R}} |y-x|^\varrho \pi(dx,dy) \le \left(\int_{\mathbb{R}\times\mathbb{R}} |y-x|^2 \pi(dx,dy)\right)^{\frac{\varrho}{2}} = 1$ gives that $\pi^\star$ is the unique martingale coupling that maximizes $\int_{\mathbb{R}\times\mathbb{R}} |y-x|^\varrho \pi(dx,dy)$. The optimality of $\pi^\star$ for $\varrho = 1$ was obtained by Hobson and Neuberger (2012) in Example 6.1. Note that the optimal coupling is neither the left-curtain coupling $\pi^{\mathrm{lc}}(dx,dy) = \frac{1}{2}1_{[-1,1]}(x)dx\left(\frac{1}{4}\delta_{-\frac{x}{2}-\frac{3}{2}}(dy) + \frac{3}{4}\delta_{\frac{3}{2}x+\frac{1}{2}}(dy)\right)$ nor the right-curtain coupling $\pi^{\mathrm{rc}}(dx,dy) = \frac{1}{2}1_{[-1,1]}(x)dx\left(\frac{3}{4}\delta_{\frac{3}{2}x-\frac{1}{2}}(dy) + \frac{1}{4}\delta_{\frac{3}{2}-\frac{x}{2}}(dy)\right)$ introduced by Beiglböck and Juillet (2016), Theorem 1.5, see also Henry-Labordère and Touzi (2016) for an explicit calculation of those couplings.

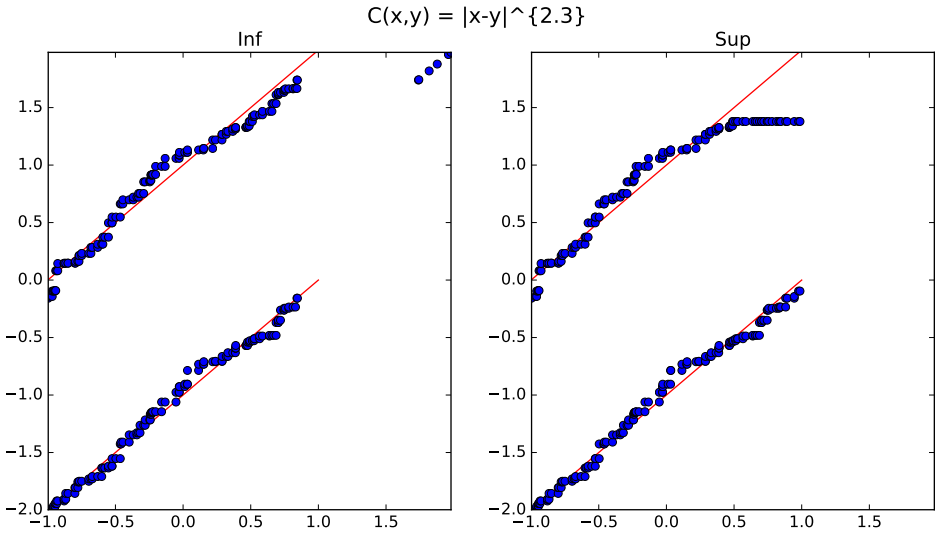


FIGURE 1. Points (x_i, y_j) such that $r^\star_{ij} > 0$, where $r^\star$ is the optimal solution of the MOT minimization (1.1) with $c(x,y) = |y-x|^{2.3}$ for $(\mu_I \wedge \nu_I, \nu_I)$ at left and $(\mu_I, \mu_I \vee \nu_I)$ at right, with $I = 100$. The segments $y = x \pm 1$, $x \in [-1,1]$ indicate the support of $\pi^\star$ given by (6.1).

For the implementation, we consider independent random samples $X_1,\dots,X_I$ distributed according to μ and $Y_1,\dots,Y_I$ distributed according to ν, with $I = 100$. We set

$$\mu_I = \frac{1}{I}\sum_{i=1}^I \delta_{X_i},\ \nu_I = \frac{1}{I}\sum_{i=1}^I \delta_{Y_i},$$

$$\widetilde{\mu}_I = \frac{1}{I}\sum_{i=1}^I \delta_{X_i-\bar{X}_I},\ \widetilde{\nu}_I = \frac{1}{I}\sum_{i=1}^I \delta_{Y_i-\bar{Y}_I},$$

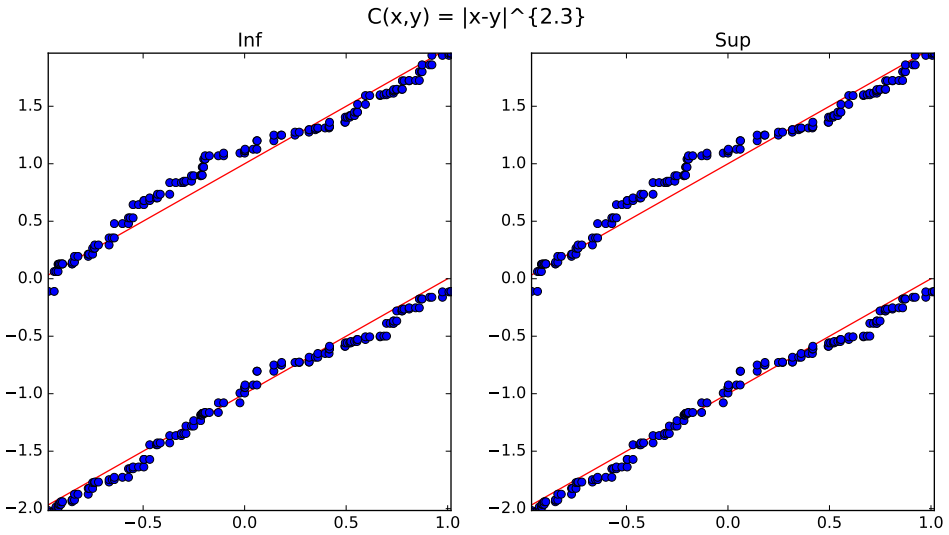


FIGURE 2. Points (x_i, y_j) such that $r^\star_{ij} > 0$, where $r^\star$ is the optimal solution of the MOT minimization (1.1) with $c(x,y) = |y-x|^{2.3}$ for $(\widetilde{\mu}_I \wedge \widetilde{\nu}_I, \widetilde{\nu}_I)$ at left and $(\widetilde{\mu}_I, \widetilde{\mu}_I \vee \widetilde{\nu}_I)$ at right, with $I = 100$. The segments $y = x \pm 1$, $x \in [-1,1]$ indicate the support of $\pi^\star$ given by (6.1).

with $\bar{X}_I = \frac{1}{I}\sum_{i=1}^I X_i$ and $\bar{Y}_I = \frac{1}{I}\sum_{i=1}^I Y_i$. The measures $\widetilde{\mu}_I$ and $\widetilde{\nu}_I$ are the ones suggested by the first modification presented in Subsection 4.2. This modification requires to know explicitly the common mean of μ and ν. Note that this is usually the case for the financial application: discounted assets prices are martingales, and their mean is given by the present values. We consider $\varrho = 2.3$. We have run the linear programming solver to solve (1.1) with $(\mu_I \wedge \nu_I, \nu_I)$ and $(\mu_I, \mu_I \vee \nu_I)$ in Figure 1 and with $(\widetilde{\mu}_I \wedge \widetilde{\nu}_I, \widetilde{\nu}_I)$ and $(\widetilde{\mu}_I, \widetilde{\mu}_I \vee \widetilde{\nu}_I)$ in Figure 2. In those figures, we have drawn the points with positive probability under the optimal transport, i.e. with the notation of (1.1), the points (x_i, y_j) such that $r^\star_{ij} > 0$, where $r^\star$ is the optimal solution of (1.1). We expect to recover the line segments $y = x+1$ and $y = x-1$ for $x \in [-1,1]$. From those figures, we observe that the martingale optimal transport map is much better approximated when we equalize the means. This can be heuristically explained as follows. Since $\widetilde{\mu}_I$ and $\widetilde{\nu}_I$ have the same mean, we have $\widetilde{\mu}_I \wedge \widetilde{\nu}_I \leq_{\mathrm{cx}} \widetilde{\nu}_I$ and $\widetilde{\mu}_I \wedge \widetilde{\nu}_I \leq_{\mathrm{cx}} \widetilde{\mu}_I$. The latter inequality ensures from Corollary 2.2 that the lowest (resp. highest) point weighted by $\widetilde{\mu}_I \wedge \widetilde{\nu}_I$ is greater (resp. smaller) than the lowest (resp. highest) point weighted by $\widetilde{\mu}_I$. Thus, $\widetilde{\mu}_I \wedge \widetilde{\nu}_I$ mainly weights points in $[-1,1]$. Instead, when we do not equalize the means, we only have $\mu_I \wedge \nu_I \leq_{\mathrm{cx}} \nu_I$, but there is absolutely no reason to have $\mu_I \wedge \nu_I \leq_{\mathrm{cx}} \mu_I$. Therefore, $\mu_I \wedge \nu_I$ may weight points that are significantly outside $[-1,1]$: we observe in Figure 1 (left) many points close to 2. A similar phenomenon happens for the supremum. When we equalize the means, we have both $\widetilde{\mu}_I \leq_{\mathrm{cx}} \widetilde{\mu}_I \vee \widetilde{\nu}_I$ and $\widetilde{\nu}_I \leq_{\mathrm{cx}} \widetilde{\mu}_I \vee \widetilde{\nu}_I$, and the latter condition gives that $\widetilde{\mu}_I \vee \widetilde{\nu}_I$ mainly weights points across $[-2,2]$. In contrast, we do not have $\nu_I \leq_{\mathrm{cx}} \mu_I \vee \nu_I$ and the points weighted by $\mu_I \vee \nu_I$ may not span $[-2,2]$. In Figure 1 (right), the highest point weighted by $\mu_I \vee \nu_I$ is 1.215876, which is far from 2.

To complement this discussion, we now compare the value of the discrete optimal cost to the continuous one, i.e. to $\int_{\mathbb{R}\times\mathbb{R}} |y-x|^\varrho \pi^\star(dx,dy) = 1$. On the run of Figures 1 and 2, we have obtained respectively 0.9294, 0.9101, 1.0773 and 1.0773 for $(\mu_I \wedge \nu_I, \nu_I)$, $(\mu_I, \mu_I \vee \nu_I)$,

$(\widetilde{\mu}_I \wedge \widetilde{\nu}_I, \widetilde{\nu}_I)$ and $(\widetilde{\mu}_I, \widetilde{\mu}_I \vee \widetilde{\nu}_I)$. On 100 independent samples, we have obtained respectively the mean costs 0.7506, 0.7319, 1.0020, 1.0020 with the standard deviations 0.2148, 0.2148, 0.1400 and 0.1400. These results confirm that the cost is much better approximated in this example when we equalize the mean. Let us note here that the costs (and the optimal probability measure) obtained with $(\widetilde{\mu}_I \wedge \widetilde{\nu}_I, \widetilde{\nu}_I)$ and $(\widetilde{\mu}_I, \widetilde{\mu}_I \vee \widetilde{\nu}_I)$ are the same because we already have on our samples $\widetilde{\mu}_I \leq_{\rm cx} \widetilde{\nu}_I$, which is due to the fact that ν largely dominates μ for the convex order. From now on, we will only present numerical results for which we equalize the means.

6.2. **The at-the-money discrete Asian call option.** We consider the payoff function $c(x,y) = (x+y)^+ = 2((x+y)/2)^+$, which corresponds to the case of a discret Asian call option where the average is calculated on two dates. We consider probability distributions $\mu \leq_{\rm cx} \nu$ with zero mean, so that the call is at-the-money.

Before to consider shifted lognormal laws, we first study the case

$$\mu(dx) = \frac{1}{2}1_{[-1,1]}dx,\ \nu(dx) = \left(\frac{1}{12}1_{[-3,-1]}(x) + \frac{1}{3}1_{[-1,1]}(x) + \frac{1}{12}1_{[1,3]}(x)\right)dx,$$

because we know then an explicit optimal martingale transport. If X is a uniform random variable on $[-1,1]$ and ξ an independent random variable such that $\mathbb{P}(\xi = -2) = \mathbb{P}(\xi = 2) = 1/6$ and $\mathbb{P}(\xi = 0) = 2/3$, $X + \xi \sim \nu$. Since $\mathbb{E}[\xi] = 0$, this shows that $\mu \leq_{\rm cx} \nu$. We are interested in the optimal coupling $\pi \in \Pi^M(\mu,\nu)$ that minimizes $\int_{\mathbb{R}\times\mathbb{R}}(x+y)^+\pi(dx,dy)$. By Jensen's inequality, we have

$$\int_{\mathbb{R}\times\mathbb{R}}(x+y)^+\mu(dx)\pi_{Y|X}(x,dy) \geq \int_{\mathbb{R}}\left(x + \int_{\mathbb{R}} y\pi_{Y|X}(x,dy)\right)^+\mu(dx) = \int_{\mathbb{R}}(2x)^+\mu(dx) = \frac{1}{2}.$$

The equality in Jensen's inequality is equivalent to have either $\pi_{Y|X}(x,[-x,+\infty)) = 1$ or $\pi_{Y|X}(x,(-\infty,-x]) = 1$, $\mu(dx)$-a.e.. The martingale kernel

$$\pi^\star_{Y|X}(x,dy) = \frac{1}{2}(\delta_{-x}(dy) + \delta_{3x}(dy)) \tag{6.2}$$

satisfies this condition and is therefore optimal. We observe that $\pi^\star(dx,dy)$ has the right-monotone property for $x \in [-1,0]$ (see Definition 1.4 of Beiglböck and Juillet (2016)) and the left-monotone property for $x \in [0,1]$. Let us note that we have not shown that $\pi^\star$ is the unique optimal martingale coupling. We also mention that, by similar arguments, $\pi^\star$ is also an optimal martingale coupling that minimizes $\int_{\mathbb{R}\times\mathbb{R}}(3x-y)^+\pi(dx,dy)$. We have plotted in Figure 3 the points weighted by the discrete MOT problem with the measures $\widetilde{\mu}_I = \frac{1}{I}\sum_{i=1}^I \delta_{X_i - \bar{X}_I}$ and $\widetilde{\nu}_I = \frac{1}{I}\sum_{i=1}^I \delta_{Y_i - \bar{Y}_I}$, where $X_1,\dots,X_I$ and $Y_1,\dots,Y_I$ are sampled independently according to μ and ν. Note that for $I = 100$, like the one plotted in Figure 3, most of the samples already satisfy $\widetilde{\mu}_I \leq_{\rm cx} \widetilde{\nu}_I$. As expected, the plotted points are close to the segment lines $y = -x$ and $y = 3x$ for $x \in [-1,1]$. The corresponding cost is 0.5085, which is not so far from the theoretical cost $1/2$ obtained for the continuous MOT. On 100 independent runs, we have obtained a mean cost of 0.5106 with a standard deviation equal to 0.0326.

Now, we carry on our numerical study and consider lognormal distributions (Black and Scholes model). Namely, we assume that $X_i \overset{(d)}{=} \exp\left(\sigma_X G - \frac{1}{2}\sigma_X^2\right) - 1$ and $Y_i \overset{(d)}{=} \exp\left(\sigma_Y G - \frac{1}{2}\sigma_Y^2\right) - 1$, with $G \sim \mathcal{N}(0,1)$, $\sigma_X = 0.24$ and $\sigma_Y = 0.28$. We have estimated for $I = 100$ the probability that $\widetilde{\mu}_I \leq_{\rm cx} \widetilde{\nu}_I$ on 10^5 runs: we have obtained $\mathbb{P}(\widetilde{\mu}_I \leq_{\rm cx} \widetilde{\nu}_I) \approx 0.4601$, with a 95% confidence interval $[0.4570, 0.4632]$. Therefore, it is crucial to use either $\widetilde{\mu}_I \wedge \widetilde{\nu}_I$ or $\widetilde{\mu}_I \vee \widetilde{\nu}_I$ on this example to recover discrete probability measures in the convex

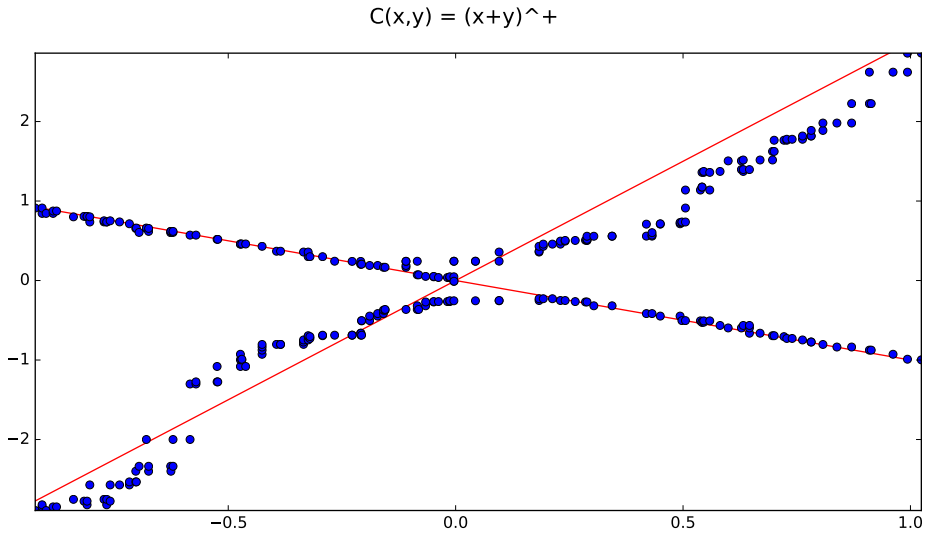


FIGURE 3. Points (x_i, y_j) such that $r^\star_{ij} > 0$, where $r^\star$ is the optimal solution of the MOT minimization (1.1) with $c(x,y) = (y+x)^+$ for $(\widetilde{\mu}_I \wedge \widetilde{\nu}_I, \widetilde{\nu}_I)$, with $I = 100$. The segments $y = 3x$ and $y = -x$, $x \in [-1,1]$ indicate the support of $\pi^\star$ given by (6.2).

order. In Figure 4, we have plotted the points weighted by the solution to the discrete MOT minimization of $c(x,y) = (x+y)^+$ for $(\widetilde{\mu}_I \wedge \widetilde{\nu}_I, \widetilde{\nu}_I)$ and $(\widetilde{\mu}_I, \widetilde{\mu}_I \vee \widetilde{\nu}_I)$ in a case where $\widetilde{\mu}_I \not\leq_{\text{cx}} \widetilde{\nu}_I$. Though being different, the two graphs are very similar. The costs are the same up to 9 digits (compared to 12 digits when $\widetilde{\mu}_I \leq_{\text{cx}} \widetilde{\nu}_I$) and are equal to 0.1809. Thus, in our experiments, we have not observed any important differences between the MOT problems on $(\widetilde{\mu}_I \wedge \widetilde{\nu}_I, \widetilde{\nu}_I)$ and $(\widetilde{\mu}_I, \widetilde{\mu}_I \vee \widetilde{\nu}_I)$, and we will work later on with $(\widetilde{\mu}_I \wedge \widetilde{\nu}_I, \widetilde{\nu}_I)$. Now, let us comment qualitatively Figure 4. We have plotted, as in Figure 3 the segment lines $y = -x$ and $y = 3x$ for $x \in [-1,1]$. As one may expect, there are still many points on the first segment line, which is the boundary of the set of points where the cost function vanishes. Instead, the other segment line was more related to the specific laws that we have considered in Figure 3. In Figure 4, there are points close to $y = 3x$ at the origin, but then they move away from this line. Points at the top right or at bottom left are mostly gathered around the line $y = x$, which is due to the martingale constraint. This is particularly noticeable on our example for the points at the bottom left.

6.3. **Lognormal distributions with $c(x,y) = |x-y|^\varrho$.** We consider the cases $\varrho = 2.1$ and $\varrho = 1.9$, and take again μ and ν as the lognormal distributions of $\exp\left(\sigma_X G - \frac{1}{2}\sigma_X^2\right) - 1$ and $\exp\left(\sigma_Y G - \frac{1}{2}\sigma_Y^2\right) - 1$, with $G \sim \mathcal{N}(0,1)$, $\sigma_X = 0.24$ and $\sigma_Y = 0.28$. In Figure 5, we have plotted the weighted points of the discrete MOT for the minimization problem. We still observe for $\varrho = 2.1$ two curves that do not cross like in Figure 2, but the curves obtained for the lognormal distributions are quite far from the lines $y = x+1$ and $y = x-1$ obtained for the uniform distribution. In Figure 6 we have plotted the weighted points for the maximization program. Interestingly, the optimal minimizing coupling for $\varrho = 2.1$ (resp. $\varrho = 1.9$) is very close to the optimal maximizing coupling for $\varrho = 1.9$ (resp. $\varrho = 2.1$).

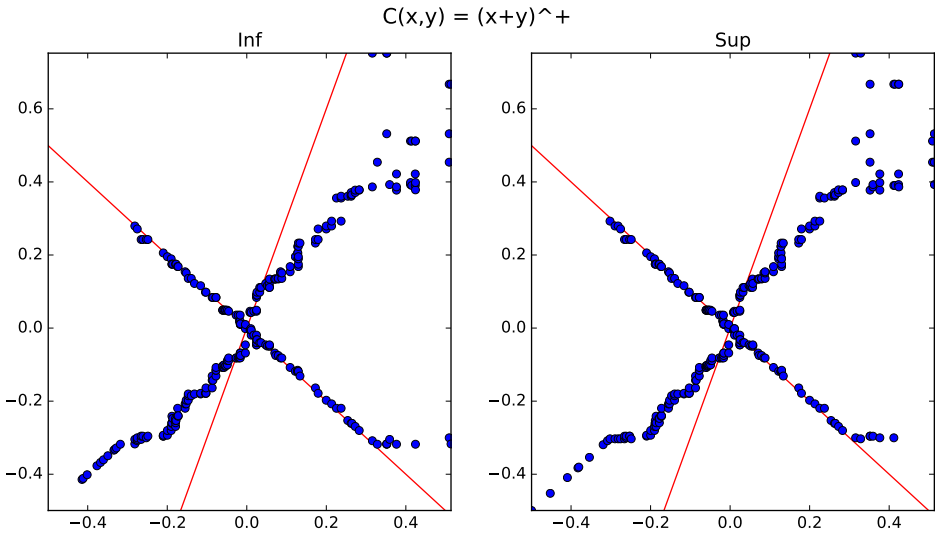


FIGURE 4. Points (x_i, y_j) such that $r^\star_{ij} > 0$, where $r^\star$ is the optimal solution of the MOT minimization (1.1) with $c(x,y) = (y+x)^+$ for $(\widetilde{\mu}_I \wedge \widetilde{\nu}_I, \widetilde{\nu}_I)$ at left and $(\widetilde{\mu}_I, \widetilde{\mu}_I \vee \widetilde{\nu}_I)$ at right, with $I = 100$. The segments $y = 3x$ and $y = -x$, $x \in [-1, 1]$ indicate the support of $\pi^\star$ given by (6.2).

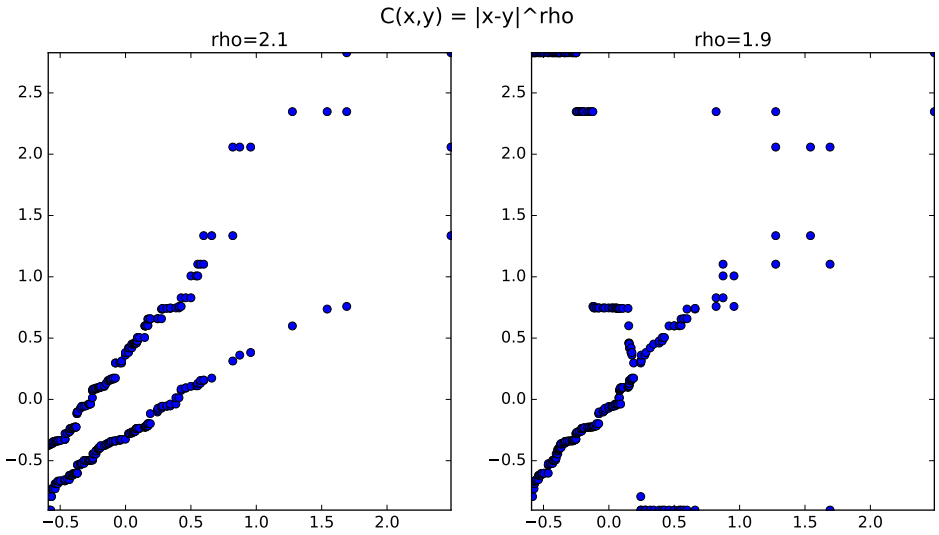


FIGURE 5. Points (x_i, y_j) such that $r^\star_{ij} > 0$, where $r^\star$ is the optimal solution of the MOT minimization (1.1) for $(\widetilde{\mu}_I \wedge \widetilde{\nu}_I, \widetilde{\nu}_I)$ with $c(x,y) = |y-x|^{2.1}$ at left and $c(x,y) = |y-x|^{1.9}$ at right, with $I = 100$.

The graphs that we have produced up to now depend on the samples $X_1, \dots, X_I$ and $Y_1, \dots, Y_I$. From a run to another, we observe the same patterns but with different points. One may like to have a discrete MOT with smaller statistical error. It is possible to do this by using Baker's construction on probability measures with finite supports. Namely,

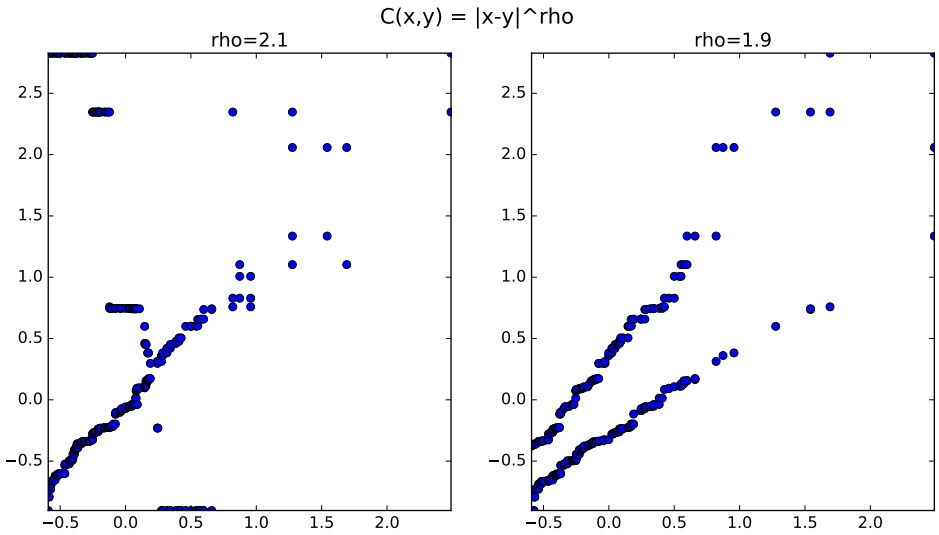


FIGURE 6. Points (x_i, y_j) such that $r^\star_{ij} > 0$, where $r^\star$ is the optimal solution of the MOT maximization (1.1) for $(\widetilde{\mu}_I \wedge \widetilde{\nu}_I, \widetilde{\nu}_I)$ with $c(x,y) = |y-x|^{2.1}$ at left and $c(x,y) = |y-x|^{1.9}$ at right, with $I = 100$.

we consider a not too small integer $I' < I$ and set

$$\hat{\mu}_{I'} = \frac{1}{I'}\sum_{i=1}^{I'} \delta_{I' \int_{\frac{i-1}{I'}}^{\frac{i}{I'}} F^{-1}_{\widetilde{\mu}_I \wedge \widetilde{\nu}_I}(u)du}, \ \hat{\nu}_{I'} = \frac{1}{I'}\sum_{i=1}^{I'} \delta_{I' \int_{\frac{i-1}{I'}}^{\frac{i}{I'}} F^{-1}_{\widetilde{\nu}_I}(u)du}. \tag{6.3}$$

The calculation of the integrals is easy and explicit since the quantile function of a discrete probability measure is piecewise constant. Even better, this construction applied to $\frac{1}{I}\sum_{i=1}^{I} \delta_{x_i}$ with $x_1 \le x_2 \le \ldots \le x_I$ and I' a divisor of I simply leads to

$$\frac{1}{I'}\sum_{i=1}^{I'} \delta_{\frac{I'}{I}\sum_{j=1}^{I/I'} x_{\frac{(i-1)I}{I'}+j}}.$$

By Theorem 2.4.11 of Baker (2012), we have $\hat{\mu}_{I'} \le_{\mathrm{cx}} \hat{\nu}_{I'}$ since $\widetilde{\mu}_I \wedge \widetilde{\nu}_I \le_{\mathrm{cx}} \widetilde{\nu}_I$. We have plotted in Figure 7 the same discrete MOT as in Figure 5, but using $(\hat{\mu}_{I'}, \hat{\nu}_{I'})$ with $I' = 100$ and $I = 10000$. The plot is clearly less noisy. This method can also be used to reduce variance on the calculation of the optimal cost. On the example of Figure 2, with 100 samples, $I' = 100$ and $I = 10000$, we have a mean cost equal to 0.9981 with a standard deviation equal to 0.0148. This is to compare with the mean cost of 1.0020 and the standard deviation of 0.1400 obtained with $(\widetilde{\mu}_I \wedge \widetilde{\nu}_I, \widetilde{\nu}_I)$ and $I = 100$.

6.4. **An example with three marginal laws.** We first write the discrete MOT problem in this case. We have to minimize (or maximize)

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} r_{ijk} c(x_i, y_j, z_k) \tag{6.4}$$

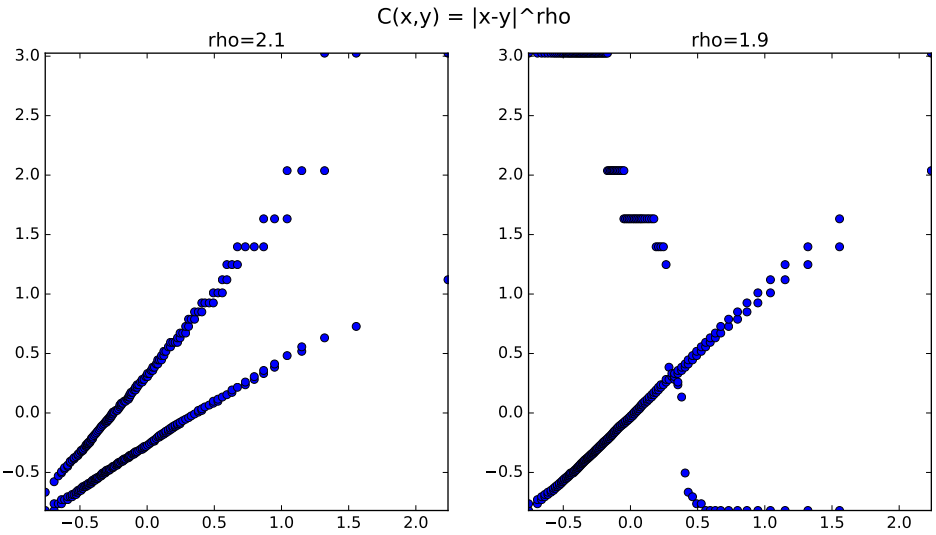


FIGURE 7. Points (x_i, y_j) such that $r^\star_{ij} > 0$, where $r^\star$ is the optimal solution of the MOT minimization (1.1) with $c(x,y) = |y-x|^{2.1}$ at left and $c(x,y) = |y-x|^{1.9}$ at right for $(\hat{\mu}_{I'}, \hat{\nu}_{I'})$ defined by (6.3) with $I' = 100$ and $I = 10000$.

under the constraints

$$\forall i,j,k,\ r_{ijk} \geq 0,\ \forall i, \sum_{j=1}^{J}\sum_{k=1}^{K} r_{ijk} = p_i,\ \forall j, \sum_{i=1}^{I}\sum_{k=1}^{K} r_{ijk} = q_j,\ \forall k, \sum_{i=1}^{I}\sum_{j=1}^{J} r_{ijk} = s_k,$$

$$\forall i, \sum_{j=1}^{J}\sum_{k=1}^{K} r_{ijk}(y_j - x_i) = 0,\ \forall i,j, \sum_{k=1}^{K} r_{ijk}(z_k - y_j) = 0.$$

For a solution to exist, the measures $\mu = \sum_{i=1}^{I} p_i \delta_{x_i}$, $\nu = \sum_{j=1}^{J} q_j \delta_{y_j}$ and $\eta = \sum_{k=1}^{K} s_k \delta_{z_k}$ have to satisfy $\mu \leq_{\text{cx}} \nu \leq_{\text{cx}} \eta$.

For $i = 1, \dots, I$, we consider independent samples $X_i \overset{(d)}{=} \exp\left(\sigma_X G - \frac{1}{2}\sigma_X^2\right) - 1$, $Y_i \overset{(d)}{=} \exp\left(\sigma_Y G - \frac{1}{2}\sigma_Y^2\right) - 1$ and $Z_i \overset{(d)}{=} \exp\left(\sigma_Y G - \frac{1}{2}\sigma_Y^2\right) - 1$, with $G \sim \mathcal{N}(0,1)$, $\sigma_X = 0.24$, $\sigma_Y = 0.28$, $\sigma_Z = 0.32$. In the financial application, the variables $X+1$, $Y+1$ and $Z+1$ represent the values of an asset at three different dates $t_1 < t_2 < t_3$ in a Black-Scholes model, and we are interested in calculating price bounds for the option that pays $\left(Z - \frac{X+Y}{2}\right)^+$, i.e. $c(x,y,z) = \left(z - \frac{x+y}{2}\right)^+$. The price of this option in the Black-Scholes model can easily be calculated with a Monte-Carlo method and is approximately equal to 0.0681.

We define as before $\bar{X}_I = \frac{1}{I}\sum_{i=1}^{I} X_i$, $\bar{Y}_I = \frac{1}{I}\sum_{i=1}^{I} Y_i$, $\bar{Z}_I = \frac{1}{I}\sum_{i=1}^{I} Z_i$, $\widetilde{\mu}_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{X_i - \bar{X}_I}$, $\widetilde{\nu}_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{Y_i - \bar{Y}_I}$ and $\widetilde{\eta} = \frac{1}{I}\sum_{i=1}^{I} \delta_{Z_i - \bar{Z}_I}$. Following Proposition 4.9, we then consider the measures $\widetilde{\mu}_I \wedge (\widetilde{\nu}_I \wedge \widetilde{\eta}_I), \widetilde{\nu}_I \wedge \widetilde{\eta}_I, \widetilde{\eta}_I$ to solve (6.4). In fact, instead of working with $(\widetilde{\mu}_I \wedge (\widetilde{\nu}_I \wedge \widetilde{\eta}_I), \widetilde{\nu}_I \wedge \widetilde{\eta}_I, \widetilde{\eta}_I)$, we will work with

$$\hat{\mu}_{I'} = \frac{1}{I'}\sum_{i=1}^{I'} \delta_{I' \int_{\frac{i-1}{I'}}^{\frac{i}{I'}} F^{-1}_{\widetilde{\mu}_I \wedge (\widetilde{\nu}_I \wedge \widetilde{\eta}_I)}(u)du},\ \hat{\nu}_{I'} = \frac{1}{I'}\sum_{i=1}^{I'} \delta_{I' \int_{\frac{i-1}{I'}}^{\frac{i}{I'}} F^{-1}_{\widetilde{\nu}_I \wedge \widetilde{\eta}_I}(u)du},\ \hat{\eta}_{I'} = \frac{1}{I'}\sum_{i=1}^{I'} \delta_{I' \int_{\frac{i-1}{I'}}^{\frac{i}{I'}} F^{-1}_{\widetilde{\eta}_I}(u)du}. \tag{6.5}$$

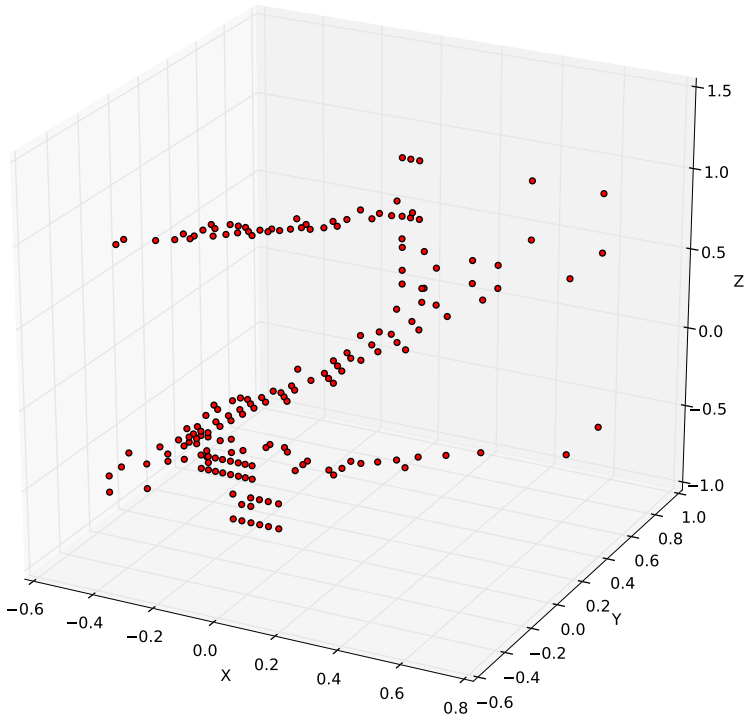


FIGURE 8. Points (x_i, y_j, z_k) such that $r^\star_{ijk} > 0$, where $r^\star$ is the optimal solution of the MOT minimization (6.4) with $c(x, y, z) = \left(z - \frac{x+y}{2}\right)^+$ for $(\hat{\mu}_{I'}, \hat{\nu}_{I'})$ defined by (6.3) with $I' = 25$ and $I = 2500$.

This has two advantages. First, as discussed in the example of Figure 7, the result is less noisy. Second, the dimension of r in the problem (6.4) is fixed and equal to $(I')^3$, while we do not know a priori the number of points weighted by $\widetilde{\mu}_I \wedge (\widetilde{\nu}_I \wedge \widetilde{\eta}_I)$ and $\widetilde{\nu}_I \wedge \widetilde{\eta}_I$, see Subsection 5.3. Thus, we keep the control on the time needed by the linear programming solver.

We have plotted in Figure 8 the points weighted by the discrete MOT that minimizes the cost. As one may expect, many points are on the hyperplane $z = \frac{x+y}{2}$ which is the boundary of the set of points where the cost function vanishes. With this example, we have obtained a cost of 0.0303, which is a lower bound for the price. Running the maximization program on the same sample, we have obtained 0.0856 for the price upper bound. As expected, the Black-Scholes price 0.0681 is between these bounds: the position of the Black-Scholes price with respect to the bounds is an indication on how conservative the model is to price the

option.

References

Alfonsi, A., Corbetta, J., and Jourdain, B. (2018). Sampling of probability measures in the convex order by Wasserstein projection and approximation of martingale optimal transport problems. *Submitted*.

Baker, D. (2012). *Martingales with specified marginals*. Theses, Université Pierre et Marie Curie - Paris VI.

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

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

Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., and Peyré, G. (2015). Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):A1111–A1138.

Föllmer, H. and Schied, A. (2011). *Stochastic finance, An introduction in discrete time*. Walter de Gruyter & Co., Berlin, third edition.

Henry-Labordère, P. and Touzi, N. (2016). An explicit martingale version of the one-dimensional Brenier theorem. *Finance Stoch.*, 20(3):635–668.

Hobson, D. and Klimmek, M. (2015). Robust price bounds for the forward starting straddle. *Finance Stoch.*, 19(1):189–214.

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

Kertz, R. P. and Rösler, U. (1992). Stochastic and convex orders and lattices of probability measures, with a martingale interpretation. *Israel J. Math.*, 77(1-2):129–164.

Kertz, R. P. and Rösler, U. (2000). Complete lattices of probability measures with applications to martingale theory. In *Game theory, optimal stopping, probability and statistics*, volume 35 of *IMS Lecture Notes Monogr. Ser.*, pages 153–177. Inst. Math. Statist., Beachwood, OH.

Müller, A. and Scarsini, M. (2006). Stochastic order relations and lattices of probability measures. *SIAM J. Optim.*, 16(4):1024–1043.

Pagès, G. and Wilbertz, B. (2012). Intrinsic stationarity for vector quantization: foundation of dual quantization. *SIAM J. Numer. Anal.*, 50(2):747–780.

Santambrogio, F. (2015). *Optimal transport for applied mathematicians*. Progress in Nonlinear Differential Equations and their Applications, 87. Birkhäuser/Springer.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic orders*. Springer Series in Statistics. Springer, New York.

Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439.