

Research Article

Johannes Wiesel*, and Erica Zhang

An optimal transport-based characterization of convex order

<https://doi.org/10.1515/demo-2023-0102>

received March 8, 2023; accepted August 30, 2023

Abstract: For probability measures μ , ν , and ρ , define the cost functionals

$$C(\mu, \rho) = \sup_{\pi \in \Pi(\mu, \rho)} \int \langle x, y \rangle \pi(dx, dy) \quad \text{and} \quad C(\nu, \rho) = \sup_{\pi \in \Pi(\nu, \rho)} \int \langle x, y \rangle \pi(dx, dy),$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product and $\Pi(\cdot, \cdot)$ is the set of couplings. We show that two probability measures μ and ν on \mathbb{R}^d with finite first moments are in convex order (i.e., $\mu \preceq_c \nu$) iff $C(\mu, \rho) \leq C(\nu, \rho)$ holds for all probability measures ρ on \mathbb{R}^d with bounded support. This generalizes a result by Carlier. Our proof relies on a quantitative bound for the infimum of $\int f d\nu - \int f d\mu$ over all 1-Lipschitz functions f , which is obtained through optimal transport (OT) duality and the characterization result of OT (couplings) by Rüschendorf, by Rachev, and by Brenier. Building on this result, we derive new proofs of well known one-dimensional characterizations of convex order. We also describe new computational methods for investigating convex order and applications to model-independent arbitrage strategies in mathematical finance.

Keywords: convex order, optimal transport, Wasserstein distance, model-independent finance**MSC 2020:** 60G46, 60G42, 91G80

1 Introduction and main result

Fix two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ with

$$\int |x| \mu(dx) < \infty \quad \text{and} \quad \int |y| \nu(dy) < \infty.$$

Recall that μ and ν are in convex order (denoted by $\mu \preceq_c \nu$) iff

$$\int f d\mu \leq \int f d\nu \quad \text{for all convex functions } f: \mathbb{R}^d \rightarrow \mathbb{R}.$$

As any convex function is bounded from below by an affine function, the aforementioned integrals take values in $(-\infty, \infty]$. The notion of convex order is very well studied (see, e.g., [5, 26, 32, 35] and the references therein for an overview). It plays a pivotal role in mathematical finance since [36] established that $\mu \preceq_c \nu$ if and only if $\mathcal{M}(\mu, \nu)$ – the set of martingale laws on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν – is non-empty. This result is also

A first version of this article was submitted to the ArXiv on July 04, 2022, under the title “A characterisation of convex order using the 2-Wasserstein distance.”

* **Corresponding author: Johannes Wiesel**, Department of Mathematics, Carnegie Mellon University, Wean Hall, 5000 Forbes Ave, Pittsburgh, PA 15213, USA, e-mail: wiesel@cmu.edu

Erica Zhang: Department of Statistics, Columbia University, 1255 Amsterdam Avenue, New York, NY 10027, USA, e-mail: y4232@columbia.edu

the reason why convex order has taken the center stage in the field of martingale optimal transport (OT) (see, e.g., [2–4,6,7,12,18,19,21,23,28] and references therein). Furthermore, convex order plays a pivotal role in dependence modeling and risk aggregation (see, e.g., [8,17,34,37,40]). Over the past 50 years or so, several other properties of convex order and extensions of [36] have been found. To mention just a few, [15,16] give an equivalent condition of convex order of μ and ν . This condition is based on the so-called fusions of probability measures and is most instructive in the case of finitely supported measures μ and ν . Furthermore, [25, Section 4] give a constructive proof of Strassen's theorem in the univariate case.

While there is an abundance of explicit characterizations of convex order available in one dimension (i.e., $d = 1$) – (see, e.g., [35, Chapter 3]), the case $d > 1$ seems to be less studied to the best of our knowledge. The main goal of this article is to fill this gap: we discuss a characterization of convex order that holds in general dimensions, and is based on the theory of OT. OT goes back to the seminal works of [24] and [22]. It is concerned with the problem of transporting probability distributions in a cost-optimal way. We refer to [31] and [38,39] for an overview. For this study, we only need a few basic concepts from OT. Most importantly, we will need the cost functionals

$$C(\mu, \rho) := \sup_{\pi \in \Pi(\mu, \rho)} \int \langle x, y \rangle \pi(dx, dy) \quad \text{and} \quad C(\nu, \rho) := \sup_{\pi \in \Pi(\nu, \rho)} \int \langle x, y \rangle \pi(dx, dy).$$

Here, $\Pi(\mu, \nu)$ denotes the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν . Our main result is the following:

Theorem 1.1. *Assume that $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ have finite first moments. Then,*

$$\inf_{f \in C^1(\mathbb{R}^d)} \left(\int f d\nu - \int f d\mu \right) = \inf_{\rho \in \mathcal{P}^1(\mathbb{R}^d)} (C(\nu, \rho) - C(\mu, \rho)), \quad (1)$$

where

$$\mathcal{P}^1(\mathbb{R}^d) := \{\rho \in \mathcal{P}(\mathbb{R}^d) : \text{supp}(\rho) \subseteq B_1(0)\}$$

and $B_1(0)$ denotes the unit ball in \mathbb{R}^d , as well as

$$C^1(\mathbb{R}^d) := \{f : \mathbb{R}^d \rightarrow \mathbb{R} \text{ convex, 1-Lipschitz}\}.$$

Theorem 1.1 states that convex order of μ and ν is equivalent to an order relation $C(\cdot, \cdot)$ on the space of probability measures. Contrary to standard characterizations of convex order using potential functions or cumulative distribution functions, it holds in any dimension and can be seen as a natural generalization of the following result:

Corollary 1.2. *Denote the 2-Wasserstein metric by:*

$$\mathcal{W}_2(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \sqrt{\int |x - y|^2 \pi(dx, dy)}.$$

If μ and ν have finite second moment, then they are in convex order if and only if

$$\mathcal{W}_2(\nu, \rho)^2 - \mathcal{W}_2(\mu, \rho)^2 \leq \int |y|^2 \nu(dy) - \int |x|^2 \mu(dx) \quad (2)$$

holds for all probability measures ρ on \mathbb{R}^d with bounded support.

Corollary 1.2 has an interesting history. To the best of our knowledge, it was first stated in [11] for compactly supported measures μ and ν . The proof in [11] relies on a well known connection between convex functions and OT for the squared Euclidean distance due to [33] and [10] together with a certain probabilistic first-order condition (see [11, Proposition 1]).

Interestingly, Carlier's result does not seem to be very well known in the literature on stochastic order. We conjecture that this is mainly due to his use of the French word “balayée” instead of convex order, so that the connection is not immediately apparent. For this reason, one aim of this note is to popularize Carlier's result, making it accessible to a wider audience, while simultaneously showcasing potential applications. As it turns

out, Corollary 1.2 is at least partially known to the mathematical finance community: indeed, the “only if” direction of Corollary 1.2 was rediscovered in [4, Equation (2.2)] for (not necessarily compactly supported) probability measures μ and ν with finite second moments. Recently, Carlier’s result has also been used in [14] for statistical estimation of convex order for compactly supported probability measures with the so-called input convex maxout neural networks. This note differs from Carlier’s work in three aspects: first, as convex order is classically embedded in $\mathcal{P}_1(\mathbb{R}^d)$ – the space of probability measures on \mathbb{R}^d with finite first moments – and does not require moments of higher order or compact support assumptions (see, e.g., [27]), Theorem 1.1 is simultaneously more concise and arguably more natural than Corollary 1.2. Second, our proof of Theorem 1.1 (and thus also Corollary 1.2) follows a different route than Carlier’s original proof, who argues purely on the space probability measures (i.e., the “primal side” in OT). Instead, we directly work with the result of [33] and [10] – in particular, the classical OT duality. Finally, we discuss three implications of Theorem 1.1: we first give a proof of a characterization of convex order in one dimension through quantile functions. Then, we use Theorem 1.1 to derive new computational methods for testing convex order between μ and ν . For the computation, we exploit state-of-the-art computational OT methods, which are efficient for potentially high-dimensional problems. These have recently seen a spike in research activity. We refer to [29] for an overview. Finally, we discuss the applications of Theorem 1.1 to the theory of the so-called model-independent arbitrages (see [1, Definition 1.2]).

This article is structured as follows: in Section 2, we state examples and consequences of Theorem 1.1. In particular, we connect it to some well known results in the theory of convex order. The proof of the main results is given in Section 3. Sections 4 and 5 discuss numerical and mathematical finance applications of Theorem 1.1, respectively. Remaining proofs are collected in Section 6.

2 Discussion and consequences of main results

To sharpen intuition, let us first discuss the case $d = 1$. By Theorem 1.1, we can obtain a new proof of a well known representation of convex order on the real line (see, e.g., [35, Theorem 3.A.5]). Here, we denote the quantile function of a probability measure μ by:

$$F_\mu^{-1}(x) := \inf\{y \in \mathbb{R} : \mu((-\infty, y]) \geq x\}.$$

Corollary 2.1. *For $d = 1$, we have*

$$\mu \preceq_c \nu \Leftrightarrow \int_0^x [F_\mu^{-1}(y) - F_\nu^{-1}(y)] dy \geq 0$$

for all $x \in [0, 1]$, with equality for $x = 1$.

The proofs of all results of this section are collected in Section 6. We continue with general $d \in \mathbb{N}$ and give a geometric interpretation of Corollary 1.2 by restating it as follows: $\mu \preceq_c \nu$ holds iff

$$\mathcal{W}_2(\nu, \rho)^2 - \mathcal{W}_2(\mu, \rho)^2 \leq \mathcal{W}_2(\nu, \delta_z)^2 - \mathcal{W}_2(\mu, \delta_z)^2 \quad (3)$$

for all $\rho \in \mathcal{P}(\mathbb{R}^d)$ with bounded support, where δ_z , $z \in \mathbb{R}^d$ is a Dirac measure. Indeed, varying ρ over Dirac measures in (2) implies that the means of μ and ν have to be equal; equation (3) then follows from simple algebra. This implies, in particular, that the difference between squared Wasserstein cost from ν and μ to ρ is maximized at the point masses. More generally, it is natural to ask if one can restrict the class of $\rho \in \mathcal{P}^1(\mathbb{R}^d)$ to a subclass with a more tractable representation. This is true in $d = 1$, where it can be easily checked from Corollary 2.1 that it is enough to consider $\rho = x\delta_0 + (1 - x)\delta_1$ for $x \in [0, 1]$. For $d > 1$, this question is beyond the scope of this article and we leave it for future research. Theorem 1.1 can also be reformulated as: $\mu \preceq_c \nu$ iff

$$\sup_{\pi \in \Pi(\mu, \rho)} \int \langle x, z \rangle \pi(dx, dz) \leq \sup_{\pi \in \Pi(v, \rho)} \int \langle y, z \rangle \pi(dy, dz),$$

i.e., for any $\rho \in \mathcal{P}(\mathbb{R}^d)$ with bounded support, the maximal covariance between μ and ρ is less than the one between v and ρ . This provides a natural intuition for a classical pedestrian description of convex order, namely, that “ v being more spread out than μ .”

We next give a simple example for Corollary 1.2.

Example 2.2. Let us take $\mu = \delta_0$ and v with mean zero. Now, recalling (4) and bounding $\mathcal{W}_2(v, \rho)$ from above by choosing the product coupling, we obtain that for any ρ with finite second moment,

$$\begin{aligned} \mathcal{W}_2(v, \rho)^2 - \mathcal{W}_2(\mu, \rho)^2 &= \mathcal{W}_2(v, \rho)^2 - \int |x|^2 \rho(dx) \\ &\leq \int |y|^2 v(dy) - \int 2\langle x, y \rangle v(dx) \rho(dy) \\ &= \int |y|^2 v(dy) \\ &= \int |y|^2 v(dy) - \int |x|^2 \mu(dx). \end{aligned}$$

In conclusion, we recover the well known fact $\delta_0 \preceq_c v$.

We now state two direct corollaries of Corollary 1.2. We consider the cost $c(x, y) = |x - y|^2/2$ and recall that a function f is c -concave, if

$$f(x) = \inf_{y \in \mathbb{R}^d} (g(y) - c(x, y))$$

for some function $g : \mathbb{R}^d \rightarrow \mathbb{R}$. We then have the following:

Corollary 2.3. *We have*

$$\int g dv \leq \int g d\mu \quad \text{for all } c\text{-concave functions } g : \mathbb{R}^d \rightarrow \mathbb{R}$$

if and only if

$$\mathcal{W}_2(v, \rho)^2 \leq \mathcal{W}_2(\mu, \rho)^2 \quad \text{for all } \rho \in \mathcal{P}(\mathbb{R}^d) \text{ with compact support.}$$

Corollary 1.2 also directly implies the following well known result:

Corollary 2.4. *If $\mu \preceq_c v$, then*

$$\mathcal{W}_2(\mu, v)^2 \leq \int |y|^2 v(dy) - \int |x|^2 \mu(dx).$$

In particular, $\mu \preceq_c v$ implies

$$\sup_{\pi \in \Pi(\mu, v)} \int \langle x, y \rangle \pi(dx, dy) \geq \int |x|^2 \mu(dx).$$

3 Proof of Theorem 1.1 and Corollary 1.2

Let us start by setting up some notation. We denote the scalar product on \mathbb{R}^d by $\langle \cdot, \cdot \rangle$. We write $|\cdot|$ for the Euclidean norm on \mathbb{R}^d . The ball in \mathbb{R}^d around x of radius $r > 0$ will be denoted by $B_r(x)$.

In order to keep this article self-contained, we summarize some properties of OT at the beginning of this section (see [38, Chapter 2.1] for a more detailed treatment). First we set

$$\mathcal{P}_p(\mathbb{R}^d) := \left\{ \rho \in \mathcal{P}(\mathbb{R}^d) : \int |x|^p \mu(dx) < \infty \right\}$$

for all $p \geq 1$. By definition, we have for any $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ that

$$\mathcal{W}_2(\mu, \rho)^2 = \int |x|^2 \mu(dx) + \int |y|^2 \rho(dy) - 2 \sup_{\pi \in \Pi(\mu, \rho)} \int \langle x, y \rangle \pi(dx, dy). \quad (4)$$

In this section, we thus (re-)define the cost function $c(x, y) := \langle x, y \rangle$ and recall that the convex conjugate $f^* : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ of a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is given by:

$$f^*(y) := \sup_{x \in \mathbb{R}^d} (\langle y, x \rangle - f(x)).$$

The subdifferential of a proper convex function $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ is defined as:

$$\partial f(x) := \{y \in \mathbb{R}^d : f(x') - f(x) \geq \langle y, x' - x \rangle \text{ for all } x' \in \mathbb{R}^d\}.$$

It is non-empty if x belongs to the interior of the domain of f . We have

$$f(x) + f^*(y) - \langle x, y \rangle = 0 \Leftrightarrow y \in \partial f(x). \quad (5)$$

Finally, we recall the OT duality (see [33, (12)])

$$\begin{aligned} C(\mu, \rho) &= \sup_{\pi \in \Pi(\mu, \rho)} \int \langle x, y \rangle \pi(dx, dy) \\ &= \inf_{f \oplus g \geq c} \int f d\mu + \int g d\rho \\ &= \inf_{f \oplus g \geq c, f, g \text{ proper, convex}} \int f d\mu + \int g d\rho \end{aligned} \quad (6)$$

and the existence of an optimal pair (f, f^*) of (lower semicontinuous, proper) convex conjugate functions (see, e.g., [33, proof of Theorem 1]). Replacing μ by ν in the aforementioned display, we obtain a similar duality for $C(\nu, \rho)$.

3.1 Proof of Theorem 1.1

We now prove Theorem 1.1. As $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ and $\rho \in \mathcal{P}^1(\mathbb{R}^d)$, the domain of the optimizing potential f for $C(\mu, \rho)$ (resp. $C(\nu, \rho)$) is \mathbb{R}^d in this case. We write

$$\|\partial f\|_\infty := \sup_{x \in \mathbb{R}^d} \sup_{y \in \partial f(x)} |y|.$$

Proof of Theorem 1.1. As μ and ν have finite first moment and ρ is compactly supported, $|C(\mu, \rho)|, |C(\nu, \rho)| < \infty$ follows from Hölder's inequality. We now fix $\rho \in \mathcal{P}^1(\mathbb{R}^d)$ and take an optimal convex pair (\hat{f}, \hat{g}) in (6) for $C(\nu, \rho)$. Next, we apply [33, Theorem 1], which states that $\rho = \partial \hat{f}(x)_* \nu$.¹ Furthermore, as $\text{supp}(\rho) \subseteq B_1(0)$, we conclude $\|\partial \hat{f}\|_\infty \leq 1$ and

¹ We note that the result is stated under the additional requirement that $\nu, \rho \in \mathcal{P}_2(\mathbb{R}^d)$. However, as ρ is supported on the unit ball and thus $|C(\mu, \rho)|, |C(\nu, \rho)| < \infty$, it can be checked that the arguments of [33, Proof of Theorem 1] carry over (see also [33, Example 1(a)]).

$$\begin{aligned}
C(\nu, \rho) - C(\mu, \rho) &\geq \int \hat{f} d\nu + \int \hat{g} d\rho - \left(\int \hat{f} d\mu + \int \hat{g} d\rho \right) \\
&= \int \hat{f} d\nu - \int \hat{f} d\mu \\
&\geq \inf_{f \in C^1(\mathbb{R}^d)} \left(\int f d\nu - \int f d\mu \right).
\end{aligned}$$

Taking the infimum over $\rho \in \mathcal{P}^1(\mathbb{R}^d)$ shows that

$$\inf_{\rho \in \mathcal{P}^1(\mathbb{R}^d)} (C(\nu, \rho) - C(\mu, \rho)) \geq \inf_{f \in C^1(\mathbb{R}^d)} \left(\int f d\nu - \int f d\mu \right).$$

On the other hand, fix $f \in C^1(\mathbb{R}^d)$ and set $g := f^*$. Define $\hat{\rho} := \partial f_* \mu$ and note that $\hat{\rho} \in \mathcal{P}^1(\mathbb{R}^d)$. Then, again by [33, Theorem 1], we obtain optimality of the pair (f, g) for $C(\mu, \hat{\rho})$, and thus,

$$\begin{aligned}
\int f d\nu - \int f d\mu &= \left(\int f d\nu + \int g d\hat{\rho} \right) - \left(\int g d\hat{\rho} + \int f d\mu \right) \\
&\geq C(\nu, \hat{\rho}) - C(\mu, \hat{\rho}) \\
&\geq \inf_{\rho \in \mathcal{P}^1(\mathbb{R}^d)} (C(\nu, \rho) - C(\mu, \rho)).
\end{aligned}$$

Taking the infimum over $f \in C^1(\mathbb{R}^d)$ shows

$$\inf_{f \in C^1(\mathbb{R}^d)} \left(\int f d\nu - \int f d\mu \right) \geq \inf_{\rho \in \mathcal{P}^1(\mathbb{R}^d)} (C(\nu, \rho) - C(\mu, \rho)).$$

This concludes the proof. \square

3.2 Proof of Corollary 1.2

We now detail the proof of Corollary 1.2. We start with a preliminary result, which is an immediate corollary of Theorem 1.1.

Corollary 3.1. Assume $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$. Then, we have

$$\inf_{f \text{ convex}} \left(\int f d\nu - \int f d\mu \right) = \inf_{\rho \in \mathcal{P}^\infty(\mathbb{R}^d)} (C(\nu, \rho) - C(\mu, \rho)), \quad (7)$$

where $\mathcal{P}^\infty(\mathbb{R}^d)$ denotes the set of probability measures with bounded support. In particular,

$$\int f d\mu \leq \int f d\nu \quad \text{for all convex functions } f: \mathbb{R}^d \rightarrow \mathbb{R}$$

if and only if

$$C(\mu, \rho) \leq C(\nu, \rho) \quad \text{for all } \rho \in \mathcal{P}^\infty(\mathbb{R}^d).$$

Proof. Multiplying both sides of (1) by $k > 0$ yields

$$\inf_{f \in C^k(\mathbb{R}^d)} \left(\int f d\nu - \int f d\mu \right) = \inf_{\rho \in \mathcal{P}^k(\mathbb{R}^d)} (C(\nu, \rho) - C(\mu, \rho)),$$

with the definitions

$$\mathcal{P}^k(\mathbb{R}^d) = \{\rho \in \mathcal{P}(\mathbb{R}^d) : \text{supp}(\rho) \subseteq B_k(0)\}$$

and

$$C^k(\mathbb{R}^d) = \{f: \mathbb{R}^d \rightarrow \mathbb{R} \text{ convex}, \|\partial f\|_\infty \leq k\}.$$

Taking $k \rightarrow \infty$, we obtain

$$\inf_{f \text{ convex, Lipschitz}} \left(\int f d\nu - \int f d\mu \right) = \inf_{\rho \in \mathcal{P}^\infty(\mathbb{R}^d)} (C(\nu, \rho) - C(\mu, \rho)).$$

Finally, any convex function $f: \mathbb{R}^d \rightarrow \mathbb{R}$ can be approximated in a pointwise sense by convex Lipschitz functions. Thus,

$$\inf_{f \text{ convex}} \left(\int f d\nu - \int f d\mu \right) = \inf_{\rho \in \mathcal{P}^\infty(\mathbb{R}^d)} (C(\nu, \rho) - C(\mu, \rho)).$$

The claim thus follows. \square

Remark 3.2. If $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ for some $p \geq 1$, then by Hölder's inequality and density of finitely supported measures in the q -Wasserstein space, we also obtain

$$\inf_{f \text{ convex}} \left(\int f d\nu - \int f d\mu \right) = \inf_{\rho \in \mathcal{P}_q(\mathbb{R}^d)} (C(\nu, \rho) - C(\mu, \rho)),$$

where $1/p + 1/q = 1$.

Proof of Corollary 1.2. Recall from (4) that

$$\begin{aligned} C(\mu, \rho) &= \frac{1}{2} \left(\int |x|^2 \mu(dx) + \int |z|^2 \rho(dz) - \mathcal{W}_2(\mu, \rho)^2 \right), \\ C(\nu, \rho) &= \frac{1}{2} \left(\int |y|^2 \nu(dy) + \int |z|^2 \rho(dz) - \mathcal{W}_2(\nu, \rho)^2 \right). \end{aligned}$$

Combining this with (7) from Corollary 3.1 yields

$$\begin{aligned} & \inf_{f \text{ convex}} \left(\int f d\nu - \int f d\mu \right) \\ &= \inf_{\rho \in \mathcal{P}^\infty(\mathbb{R}^d)} (C(\nu, \rho) - C(\mu, \rho)) \\ &= \frac{1}{2} \inf_{\rho \in \mathcal{P}^\infty(\mathbb{R}^d)} \left(\int |y|^2 \nu(dy) + \int |z|^2 \rho(dz) - \mathcal{W}_2(\nu, \rho)^2 - \int |x|^2 \mu(dx) - \int |z|^2 \rho(dz) + \mathcal{W}_2(\mu, \rho)^2 \right) \\ &= \frac{1}{2} \inf_{\rho \in \mathcal{P}^\infty(\mathbb{R}^d)} \left(\mathcal{W}_2(\mu, \rho)^2 - \mathcal{W}_2(\nu, \rho)^2 + \int |y|^2 \nu(dy) - \int |x|^2 \mu(dx) \right). \end{aligned}$$

Thus,

$$\begin{aligned} & \int f d\mu \leq \int f d\nu \quad \text{for all convex functions } f: \mathbb{R}^d \rightarrow \mathbb{R} \\ & \Leftrightarrow \inf_{f \text{ convex}} \left(\int f d\nu - \int f d\mu \right) \geq 0 \\ & \Leftrightarrow \inf_{\rho \in \mathcal{P}^\infty(\mathbb{R}^d)} (\mathcal{W}_2(\mu, \rho)^2 - \mathcal{W}_2(\nu, \rho)^2) \geq \int |x|^2 \mu(dx) - \int |y|^2 \nu(dy) \\ & \Leftrightarrow \sup_{\rho \in \mathcal{P}^\infty(\mathbb{R}^d)} (\mathcal{W}_2(\nu, \rho)^2 - \mathcal{W}_2(\mu, \rho)^2) \leq \int |y|^2 \nu(dy) - \int |x|^2 \mu(dx). \end{aligned}$$

The claim follows. \square

4 Numerical examples

In this section, we illustrate Theorem 1.1 numerically. We focus on the following toy examples, where convex order or its absence is easy to establish.

Example 4.1. $\mu = \mathcal{N}(0, \sigma^2 I)$ and $\nu = \mathcal{N}(0, I)$ for $\sigma^2 \in [0, 2]$ for $d = 1, 2$.

Example 4.2. $d = 2$ and $\mu = \mathcal{N}(0, I)$ and $\nu = \mathcal{N}(0, \Sigma)$ for

$$\Sigma = \begin{bmatrix} 2 & 1+s \\ 1+s & 2 \end{bmatrix}$$

for $s \in [-1, 1]$.

Example 4.3. $\mu = \frac{1}{2}(\delta_{-1-s} + \delta_{1+s})$ and $\nu = \frac{1}{2}(\delta_{-1} + \delta_1)$ for $s \in [-1, 1]$.

Example 4.4.

$$\mu = \frac{1}{4}(\delta_{(-1-s,0)} + \delta_{(1+s,0)} + \delta_{(0,1+s)} + \delta_{(0,-1-s)})$$

and

$$\nu = \frac{1}{4}(\delta_{(-1,0)} + \delta_{(1,0)} + \delta_{(0,1)} + \delta_{(0,-1)})$$

for $s \in [-1, 1]$.

A general numerical implementation for testing convex order of the two measures $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ in general dimensions and the examples discussed here can be found in the Github repository <https://github.com/johanneswiesel/Convex-Order>. In the implementation, we use the POT package (<https://pythonot.github.io>) to compute OT distances.

Let us set

$$V(\mu, \nu) := \inf_{\rho \in \mathcal{P}^1(\mathbb{R}^d)} (C(\nu, \rho) - C(\mu, \rho))$$

and note that by Theorem 1.1, we have the relationship

$$\mu \preceq_c \nu \Leftrightarrow V(\mu, \nu) \geq 0.$$

Clearly, the computation of $V(\mu, \nu)$ hinges on the numerical exploration of the convex set of probability measures $\mathcal{P}^1(\mathbb{R}^d)$. We propose two methods for this: our first method only considers finitely supported measures ρ , which are dense in $\mathcal{P}^1(\mathbb{R}^d)$ in the Wasserstein topology. It relies on the *Dirichlet distribution* on the space \mathbb{R}^{g-1} , $g \in \mathbb{N}$, with density

$$f(x_1, \dots, x_g; \alpha_1, \dots, \alpha_g) = \frac{1}{B(\alpha)} \prod_{i=1}^g x_i^{\alpha_i-1}$$

for $x_1, \dots, x_g \in [0, 1]$ satisfying $\sum_{i=1}^g x_i = 1$. Here, $\alpha_1, \dots, \alpha_g > 0$, $\alpha = (\alpha_1, \dots, \alpha_g)$ and $B(\alpha)$ denotes the Beta function. Fixing g grid points $\{k_1, \dots, k_g\}$ in $B_1(0)$, we consider any realization of a Dirichlet random variable (X_1, \dots, X_g) as a probability distribution assigning probability mass X_i to the grid point k_i , $i \in \{1, \dots, g\}$. This leads to the following algorithm:

Algorithm 1. Basic algorithm for the indirect Dirichlet method**Input:** probability measures μ, ν , maximal number of evaluations N , number of grid points g **Output:** $V(\mu, \nu)$

Generate a grid G of $B_1(0)$ of g equidistant points, and consider a Dirichlet random variable modeling ρ supported on G . Use Bayesian optimization to solve

$$\inf[C(\rho, \nu) - C(\rho, \mu)]$$

over the set of Dirichlet distributions on \mathbb{R}^{g-1} . Terminate after N steps.

return $\inf[C(\rho, \nu) - C(\rho, \mu)]$.

The main computational challenge in Algorithm 1 is the efficient evaluation of $C(\rho, \nu)$ and $C(\rho, \mu)$. For this, we aim to write $C(\rho, \nu)$ and $C(\rho, \mu)$ as linear programs. We offer two different variants of Algorithm 1:

- *Indirect Dirichlet method with histograms:* if we have access to finitely supported approximations \mathbf{a} and \mathbf{b} of μ and ν , respectively, and the measure ρ is supported on G as mentioned earlier, then we solve the linear programs $C(\mathbf{a}, \rho)$ and $C(\mathbf{b}, \rho)$ as is standard in OT theory.
- *Indirect Dirichlet method with samples:* here, we draw a number of samples from μ and ν , respectively, and denote the respective empirical distributions of these samples by \mathbf{a} and \mathbf{b} . As before, we assume that we have access to a probability measure ρ supported on G . We then solve the linear programs $C(\mathbf{a}, \rho)$ and $C(\mathbf{b}, \rho)$.

An alternative to Algorithm 1 is to directly draw samples from a distribution $\rho \in \mathcal{P}^1(\mathbb{R}^d)$. We call this the *direct randomized Dirichlet method* (see Algorithm 2).

Algorithm 2. Direct randomized Dirichlet method**Input:** probability measures μ, ν , maximal number of evaluations N **Output:** $V(\mu, \nu)$

Draw samples from μ and ν and denote the empirical distributions of these samples by \mathbf{a} and \mathbf{b} , respectively. Draw samples from a Dirichlet distribution and randomize their signs, under the constraint that the empirical distribution ρ of these samples is an element of $\mathcal{P}^1(\mathbb{R}^d)$. Use Bayesian optimization to solve

$$\inf[C(\rho, \nu) - C(\rho, \mu)]$$

over the set of these distributions. Terminate after N steps.

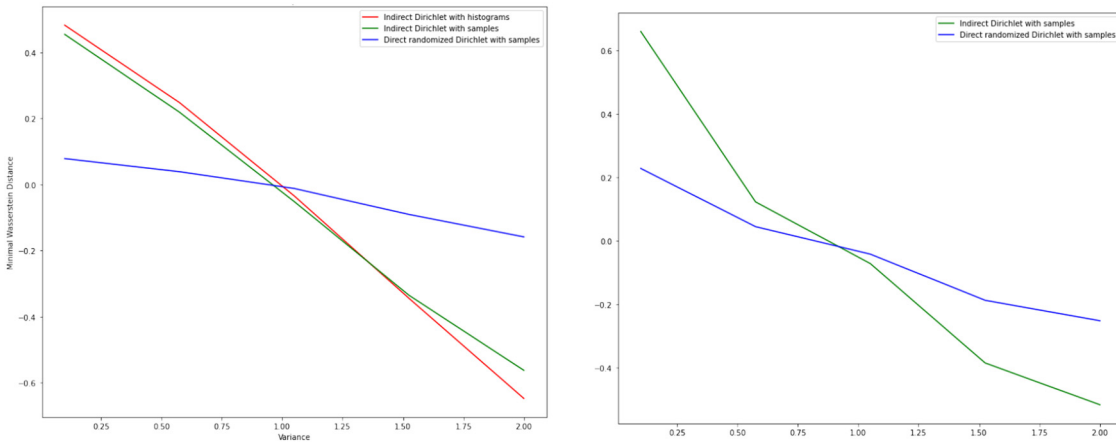
return minimal value of $\inf [C(\rho, \mu) - C(\rho, \nu)]$.

Figure 1: Values of different estimators of $V(\mu, \nu)$ plotted against σ for Example 4.1. Both plots use $N = 100$ samples.

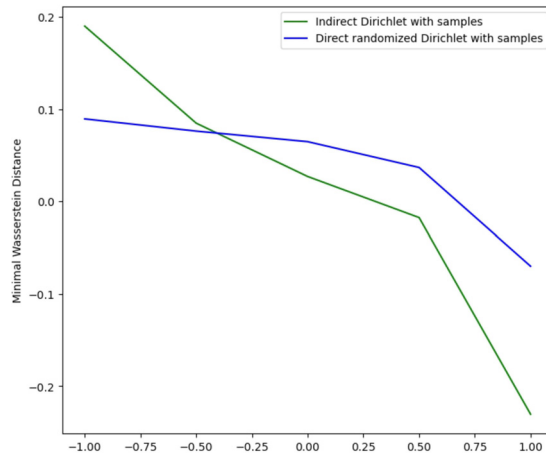


Figure 2: Values of different estimators of $V(\mu, \nu)$ plotted against s for Example 4.2. The plot uses $N = 100$ samples.

We refer to the Github repository for a more detailed discussion, in particular for the implementation and further comments. For each example stated at the beginning of this section and each pair (μ, ν) , we plot $V(\mu, \nu)$ for the three methods discussed earlier (see Figures 1–3).

Discounting numerical errors, all estimators seem to detect convex order. The direct randomized Dirichlet method is less complex; however, it does not seem to explore the $\mathcal{P}^1(\mathbb{R}^d)$ space as well as the two indirect Dirichlet methods. This can be seen in particular in Figure 2, where the direct randomized Dirichlet method does not identify convex order around $s \approx 0$ correctly. On the other hand, both of the indirect Dirichlet methods yield very similar results for the examples considered. As the name suggests, the “indirect Dirichlet method with samples” works on samples directly, which might be more convenient for practical applications on real data.

As can be expected from the numerical implementation, the histogram method consistently yields the lowest runtimes, while runtimes of the other methods are much higher. Indeed, when working with samples, the weights of the empirical distributions are constant, while the OT cost matrices \mathbf{M}_a and \mathbf{M}_b in the implementation have to be re-computed in each iteration and this is very costly; for the histogram method, the weights ρ change, while the grid stays constant – and thus also \mathbf{M}_a and \mathbf{M}_b . As an alternative, one can also use the swapping algorithm of [30]. In the aforementioned examples, this does not lead to a significant improvement of accuracy, however.

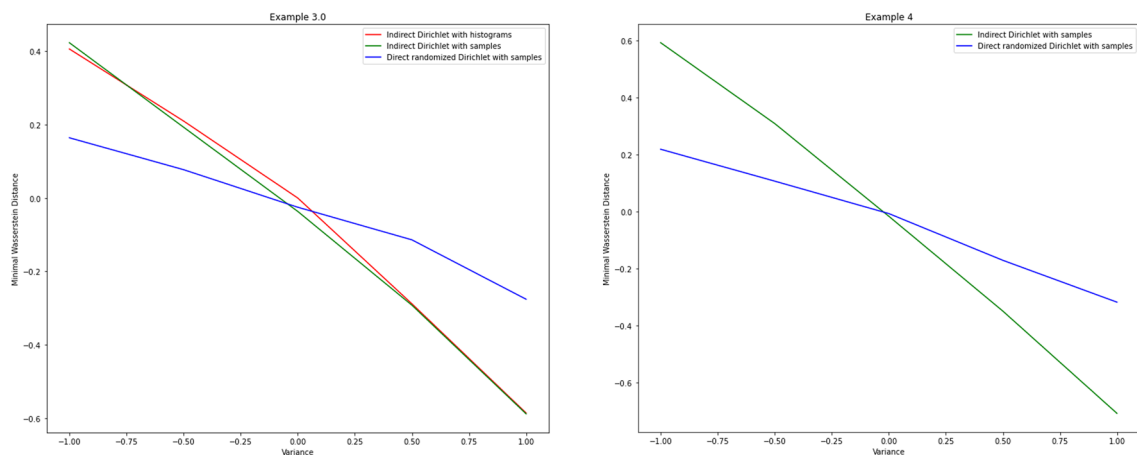


Figure 3: Values of different estimators of $V(\mu, \nu)$ plotted against s for Example 4.3 (left) and 4.4 (right). Both plots use $N = 100$ samples.

5 Model-independent arbitrage strategies

Let us consider a financial market with d financial assets and denote its price process by $(S_t)_{t \geq 0}$. Let us assume $S_0 = s_0 \in \mathbb{R}$ and fix two maturities $T_1 < T_2$. If call options with these maturities are traded at all strikes, then the prices of the call options determine the distribution of S_{T_1} and S_{T_2} under any martingale measure; this fact was first established by [9]. Let us denote the laws of S_{T_1} and S_{T_2} by μ and ν , respectively. If trading is only allowed at 0, T_1 , and T_2 , the following definition is natural and will be crucial for our analysis.

Definition 5.1. The triple of measurable functions (u_1, u_2, Δ) is a model-independent arbitrage if $u_1 \in L^1(\mu)$, $u_2 \in L^1(\nu)$, and

$$u_1(x) - \int u_1 d\mu + u_2(y) - \int u_2 d\nu + \Delta(x)(y - x) > 0, \quad \text{for all } (x, y) \in \mathbb{R}^d \times \mathbb{R}^d.$$

If no such strategies exist, then we call the market *free of model-independent arbitrage*.

In the aforementioned equation, u_1 and u_2 can be interpreted as payoffs of Vanilla options with market prices $\int u_1 d\mu$ and $\int u_2 d\nu$, respectively, while the term $\Delta(x)(y - x)$ denotes the gains or losses from buying $\Delta(x)$ assets at time T_1 and holding them until T_2 .

The following theorem makes the connection between model-independent arbitrages and convex order of μ and ν apparent. It can essentially be found in [20, Theorem 3.4].

Theorem 5.2. *The following are equivalent:*

- (i) *The market is free of model-independent arbitrage.*
- (ii) $\mathcal{M}(\mu, \nu) \neq \emptyset$.
- (iii) $\mu \preceq_c \nu$.

In particular, if $\mu \not\preceq_c \nu$, then there exists a convex function f , such that the triple $(-f(x), f(y), -g(x))$ is a model-independent arbitrage. Here, g is a measurable selector of the subdifferential of f .

The strategy $(-f(x), f(y), -g(x))$ is often called a *calendar spread*. As our setting is not quite exactly covered by [20, Theorem 3.4] and the proof is not hard, we include it here.

Proof of Theorem 5.2. (ii) \Leftrightarrow (iii) is Strassen's theorem (see [36]). If $\mu \not\preceq_c \nu$, then by definition, there exists a convex function f such that

$$\int f d\mu > \int f d\nu.$$

On the other hand, f is convex and thus satisfies

$$f(y) - f(x) \geq g(x)(y - x) \quad \text{for all } (x, y) \in \mathbb{R}^d \times \mathbb{R}^d.$$

Combining the two aforementioned equations shows that $(-f(x), f(y), -g(x))$ is a model-independent arbitrage, and thus, (i) \Rightarrow (iii). It remains to show (ii) \Rightarrow (i), which is well known. Indeed, taking expectations in the inequality

$$u_1(x) - \int u_1 d\mu + u_2(y) - \int u_2 d\nu + \Delta(x)(y - x) > 0, \quad \text{for all } (x, y) \in \mathbb{R}^d \times \mathbb{R}^d$$

under any martingale measure with marginals μ and ν leads to a contradiction. This concludes the proof. \square

As a direct consequence of Theorem 5.2, we can use Theorem 1.1 to detect model-independent arbitrages in the market under consideration: indeed, Theorem 1.1 states that $\mu \not\preceq_c \nu$ implies the existence of a probability measure $\rho \in \mathcal{P}^1(\mathbb{R}^d)$ satisfying

$$C(\rho, \nu) - C(\rho, \mu) < 0.$$

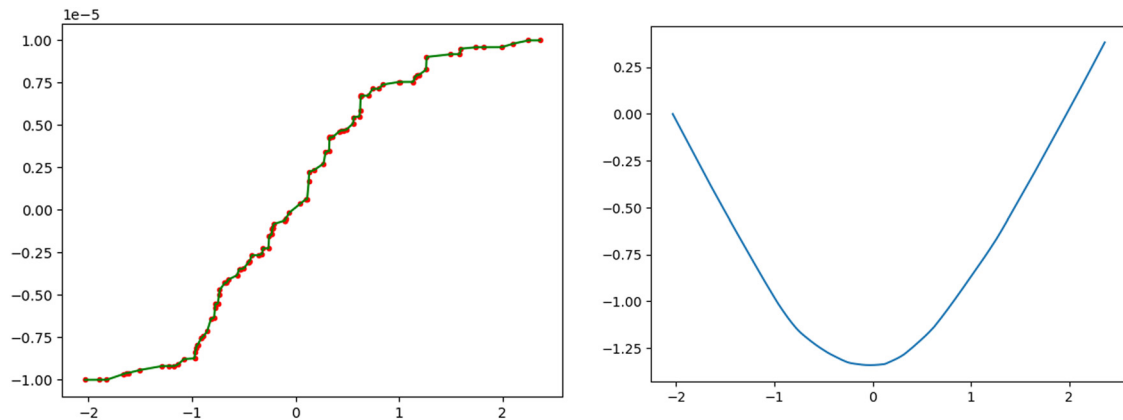


Figure 4: Plot of estimates for $\partial \hat{f}$ and \hat{f} for $\mu = \mathcal{N}(0, 2)$, $\nu = \mathcal{N}(0, 1)$, and $d = 1$. Both plots use $N = 100$ samples.

Next, the proof of Theorem 1.1 shows that $\rho = \partial \hat{f}(x)_* \nu$ for some convex function $\hat{f} : \mathbb{R}^d \rightarrow \mathbb{R}$ and

$$\int \hat{f} d\nu - \int \hat{f} d\mu \leq C(\rho, \nu) - C(\rho, \mu) < 0, \quad \text{i.e. } \int \hat{f} d\nu < \int \hat{f} d\mu.$$

In particular, a model-independent arbitrage strategy is given by calendar spread $(-\hat{f}(x), \hat{f}(x), -\partial \hat{f}(x))$. In particular, we can use the same methods as in Section 4 to find ρ . We then estimate $\partial \hat{f}(x)$ from the optimizing transport plan $\pi \in \Pi(\rho, \nu)$ of $C(\rho, \nu)$ by taking the conditional expectation $\int x \pi_y(dx)$, where $(\pi_y)_{y \in \mathbb{R}^d}$ denotes the conditional probability distribution of π with respect to its second marginal ν . This is a standard technique (see, e.g., [13] for details). In conclusion, we can obtain an explicit arbitrage strategy.

To illustrate the ideas outlined earlier, we return to Example 4.1, i.e., $\mu = \mathcal{N}(0, \sigma^2 I)$ and $\nu = \mathcal{N}(0, I)$ for $\sigma^2 > 0$ and $d = 1, 2$. Having determined ρ such that $C(\rho, \nu) - C(\rho, \mu) < 0$, we estimate $\partial \hat{f}$ numerically. We show estimates for $\partial \hat{f}$ and \hat{f} in the plots in Figures 4 and 5.

6 Remaining proofs

Proof of Corollary 2.3. Recall that a function $g : \mathbb{R}^d \rightarrow \mathbb{R}$ is c -concave, iff $f(x) := |x|^2/2 - g(x)$ is convex. In particular,

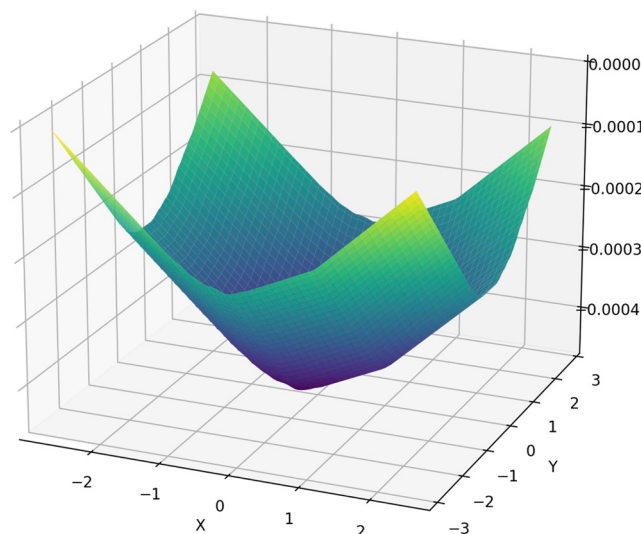


Figure 5: Plot of estimate for \hat{f} for $\mu = \mathcal{N}(0, 4I)$, $\nu = \mathcal{N}(0, I)$, $d = 2$. Both plots use $N = 100$ samples.

$$\begin{aligned}\int g d\mu - \int g d\nu &= \left(- \int \left[\frac{|x|^2}{2} - g(x) \right] \mu(dx) + \int \left[\frac{|y|^2}{2} - g(y) \right] \nu(dy) \right) + \int \frac{|x|^2}{2} \mu(dx) - \int \frac{|y|^2}{2} \nu(dy) \\ &= \int f d\nu - \int f d\mu + \int \frac{|x|^2}{2} \mu(dx) - \int \frac{|y|^2}{2} \nu(dy).\end{aligned}$$

By (7), we obtain

$$\begin{aligned}& \inf_{g \text{ c-concave}} \left(\int g d\mu - \int g d\nu \right) \\ &= \inf_{f \text{ convex}} \left(\int f d\nu - \int f d\mu \right) + \int \frac{|x|^2}{2} \mu(dx) - \int \frac{|y|^2}{2} \nu(dy) \\ &= \frac{1}{2} \inf_{\rho \in \mathcal{P}^\infty(\mathbb{R}^d)} \left(\mathcal{W}_2^2(\mu, \rho) - \mathcal{W}_2^2(\nu, \rho) + \int |y|^2 \nu(dy) - \int |x|^2 \mu(dx) + \int |x|^2 \mu(dx) - \int |y|^2 \nu(dy) \right) \\ &= \frac{1}{2} \inf_{\rho \in \mathcal{P}^\infty(\mathbb{R}^d)} (\mathcal{W}_2(\mu, \rho)^2 - \mathcal{W}_2(\nu, \rho)^2).\end{aligned}$$

This concludes the proof. \square

Proof of Corollary 2.4. The first claim follows from Corollary 1.2 by setting $\rho = \mu$. By (4), the above implies

$$2 \int |x|^2 \mu(dx) \leq 2 \sup_{\pi \in \Pi(\mu, \nu)} \int \langle x, y \rangle \pi(dx, dy),$$

so the second claim follows. \square

Proof of Corollary 2.1. First, [41, Theorem 2 and Lemma 1] show that $\mu \preccurlyeq_c \nu$ iff

$$\int_0^1 [F_\nu^{-1}(1-u) - F_\mu^{-1}(1-u)] dh(u) \geq 0 \quad (8)$$

for all concave functions h such that the aforementioned integral is finite. As any concave function is Lebesgue-almost surely differentiable, standard approximation arguments imply that (8) holds iff

$$\int_0^1 g(u) [F_\nu^{-1}(u) - F_\mu^{-1}(u)] du \geq 0$$

for all bounded increasing left-continuous functions $g : (0, 1) \rightarrow \mathbb{R}$. But

$$\{F_\rho^{-1} : \rho \in \mathcal{P}(\mathbb{R}) \text{ with bounded support}\}$$

is exactly the set of all bounded increasing left-continuous functions on $(0, 1)$. Note that by [38, Equation (2.47)]

$$\begin{aligned}\mathcal{W}_2(\nu, \rho)^2 &= \int_0^1 (F_\nu^{-1}(x) - F_\rho^{-1}(x))^2 dx \\ &= \int y^2 \nu(dy) - 2 \int_0^1 F_\nu^{-1}(x) F_\rho^{-1}(x) dx + \int z^2 \rho(dz),\end{aligned}$$

we calculate

$$\begin{aligned}\mathcal{W}_2(\nu, \rho)^2 - \mathcal{W}_2(\mu, \rho)^2 &= \int y^2 \nu(dy) - 2 \int_0^1 F_\rho^{-1}(u) F_\nu^{-1}(u) du + \int z^2 \rho(dz) \\ &\quad - \int x^2 \mu(dy) + 2 \int_0^1 F_\rho^{-1}(u) F_\mu^{-1}(u) du - \int z^2 \rho(dz) \\ &= 2 \int_0^1 F_\rho^{-1}(u) [F_\mu^{-1}(u) - F_\nu^{-1}(u)] du + \int y^2 \nu(dy) - \int x^2 \mu(dy).\end{aligned}$$

This concludes the proof. \square

Acknowledgments: JW thanks Beatrice Acciaio, Guillaume Carlier, Max Nendel, Gudmund Pammer, and Ruodu Wang for helpful discussions.

Funding information: JW acknowledges the support by NSF Grant DMS-2205534. Part of this research was performed while JW was visiting the Institute for Mathematical and Statistical Innovation (IMSI), which is supported by the National Science Foundation (Grant No. DMS-1929348). EZ acknowledges the support through the summer internship program of the Columbia University Statistics Department.

Conflict of interest: There are no conflicts of interest.

Data availability statement: Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

References

- [1] Acciaio, B., Beiglböck, M., Penkner, F., & Schachermayer, W. (2013). A model-free version of the fundamental theorem of asset pricing and the super-replication theorem. *Mathematical Finance*, 26, 233–251. doi: 10.1111/mafi.12060.
- [2] Alfonsi, A., Corbetta, J., & Jourdain, B. (2019). Sampling of one-dimensional probability measures in the convex order and computation of robust option price bounds. *International Journal of Theoretical and Applied Finance*, 22(3), 1950002.
- [3] Alfonsi, A., Corbetta, J., & Jourdain, B. (2020). Sampling of probability measures in the convex order by Wasserstein projection. *Annales Henri Poincaré*, 56(3), 1706–1729.
- [4] Alfonsi, A., & Jourdain, B. (2020). Squared quadratic Wasserstein distance: optimal Couplings and Lions differentiability. *ESAIM: Probability and Statistics*, 24, 703–717.
- [5] Arnold, B. (2012). *Majorization and the Lorenz order: A brief introduction*, (Vol. 43). Berlin: Springer Science & Business Media.
- [6] Beiglböck, M., Henry-Labordère, P., & Penkner, F. (2013). Model-independent bounds for option prices—a mass transport approach. *Finance and Stochastics*, 17(3), 477–501.
- [7] Beiglböck, M., Nutz, M., & Touzi, N. (2015). Complete duality for martingale optimal transport on the line. *Annals of Probability*, 45(5), 3038–3074.
- [8] Bernard, C., Rüschendorf, L., & Vanduffel, S. (2017). Value-at-risk bounds with variance constraints. *Journal of Risk and Insurance*, 84(3), 923–959.
- [9] Breeden, D., & Litzenberger, R. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51, 621–651.
- [10] Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4), 375–417.
- [11] Carlier, G. (2008). Remarks on Toland's duality, convexity constraint and optimal transport. *Pacific Journal of Optimization*, 4(3), 423–432.
- [12] De March, H., & Touzi, N. (2019). Irreducible convex paving for decomposition of multidimensional martingale transport plans. *Annals of Probability*, 47(3), 1726–1774.
- [13] Deb, N., Ghosal, P., & Sen, B. (2021). Rates of estimation of optimal transport maps using plug-in estimators via barycentric projections. *Advances in Neural Information Processing Systems*, 34, 29736–29753.
- [14] Domingo-Enrich, C., Schiff, Y., & Mroueh, Y. (2022). *Learning with Stochastic Orders*. arXiv: <http://arXiv.org/abs/arXiv:2205.13684>.
- [15] Elton, J., & Hill, T.P. (1992). Fusions of a probability distribution. *The Annals of Probability*, 20, 421–454.
- [16] Elton, J., & Hill, T.P. (1998). On the basic representation theorem for convex domination of measures. *Journal of Mathematical Analysis and Applications*, 228(2), 449–466.
- [17] Embrechts, P., Puccetti, G., & Rüschendorf, L. (2013). Model uncertainty and var aggregation. *Journal of Banking & Finance*, 37(8), 2750–2764.
- [18] Galichon, A., Henry-Labordère, P., & Touzi, N. (2014). A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Annals of Applied Probability*, 24(1), 312–336.
- [19] Guo, G., & Obłój, J. (2019). Computational methods for martingale optimal transport problems. *Annals of Applied Probability*, 29(6), 3311–3347.
- [20] Guyon, J., Menegaux, R., & Nutz, M. (2017). Bounds for VIX futures given S&P 500 smiles. *Finance and Stochastics*, 21, 593–630.
- [21] Jourdain, B., & Margheriti, W. (2022). Martingale Wasserstein inequality for probability measures in the convex order. *Bernoulli*, 28(2), 830–858.
- [22] Kantorovich, L. (1958). On the translocation of masses. *Management Science*, 5, 1–4.
- [23] Massa, M., & Siorpaes, P. (2022). *How to quantise probabilities while preserving their convex order*. arXiv: <http://arXiv.org/abs/arXiv:2206.10514>.

- [24] Monge, G. (1781). *Mémoire sur la théorie des déblais et des remblais*. Paris: De l'Imprimerie Royale.
- [25] Müller, A., & Rüschendorf, L. (2001). On the optimal stopping values induced by general dependence structures. *Journal of Applied Probability*, 38(3), 672–684.
- [26] Müller, A., & Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*, (Vol. 389), New York: Wiley.
- [27] Nendel, M. (2020). A note on stochastic dominance, uniform integrability and lattice properties. *Bulletin of the London Mathematical Society*, 52(5), 907–923.
- [28] Obłój, J., & Siorpaes, P. (2017). *Structure of martingale transports in finite dimensions*. arXiv: <http://arXiv.org/abs/arXiv:1702.08433>.
- [29] Peyré, G., & Cuturi, M. (2019). Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5–6), 355–607.
- [30] Puccetti, G. (2017). An algorithm to approximate the optimal expected inner product of two vectors with given marginals. *Journal of Mathematical Analysis and Applications*, 451(1), 132–145.
- [31] Rachev, S., & Rüschendorf, L. (1998). *Mass Transportation Problems: Volume I: Theory*, (Vol. 1). New York: Springer Science & Business Media.
- [32] Ross, S. (1996). *Stochastic Processes*, (Vol. 2). New York: Wiley.
- [33] Rüschendorf, L., & Rachev, S. (1990). A characterization of random variables with minimum L2-distance. *Journal of Multivariate Analysis*, 32(1), 48–54.
- [34] Rüschendorf, L., & Uckelmann, L. (2002). Variance minimization and random variables with constant sum. In: *Distributions with given marginals and statistical modeling*, (pp. 211–222) Dordrecht: Springer.
- [35] Shaked, M., & Shanthikumar, J. (2007). *Stochastic Orders*. New York: Springer.
- [36] Strassen, V. (1965). The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36, 423–439.
- [37] Tchen, A. (1980). Inequalities for distributions with given marginals. *Annals of Probability*, 8, 814–827.
- [38] Villani, C. (2003). *Topics in optimal transportation*. (Vol. 58). Providence, Rhode Island: American Mathematical Society.
- [39] Villani, C. (2008). *Optimal Transport: Old and New*, (Vol. 338). Berlin: Springer.
- [40] Wang, B., & Wang, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, 102(10), 1344–1360.
- [41] Wang, Q., Wang, R., & Wei, Y. (2020). Distortion riskmetrics on general spaces. *ASTIN Bulletin*, 50(3), 827–851.