GEOMETRY OF VECTORIAL MARTINGALE OPTIMAL TRANSPORT AND ROBUST OPTION PRICING

JOSHUA ZOEN-GIT HIEW, TONGSEOK LIM, BRENDAN PASS, AND MARCELO CRUZ DE SOUZA

Abstract.

Keywords:
MSC2020 Classification: Primary XXXXX. Secondary XXXXX

1. Introduction

Model independent finance (TL: remove?)

In mathematical finance, the Knightian uncertainty [43] is the financial risk coming from mis-specification or uncertainty about the real model of the physical world. People put more attention on such risk after the financial crisis in 2007-2008 [6]. One way to tackle this problem is the so-called model independent finance which first introduced in [37] that proposed a new way of pricing financial derivatives based on the idea of model-independent arbitrage.

In many situations, available data allows one to reconstruct distributions of individual assets at particular times; uncertainly arises concerning the joint distirbutions between different assets, or the same asset at different times. This dependence structure may then be modeled in different ways; model independent pricing problems determine the largest or smallest possible price of a particular derivative (depending on the values of multiple assets and/or multiple mulitple times) which is consistent with the available data.

Problems when the payoff function depends on the values of two (or more) assets at a single future time, with known individual distributions but unknown dependence structure, are well studied. In this case, the model independent pricing problem is equivalent a classical problem known in the

Date: June 22, 2023.

mathematics literature as optimal transport. The variant arising when the payoff depends on the value of a single asset at two (or more) future times has received a great deal of attention in the past several years. In this case, the absence of arbitrage forces the unknown coupling between the known distributions to be a martingale; the resulting optimization problem, incorporating this additional constraint, is known as martingale optimal transport.

In real markets, there are many important pricing and risk management problems that fall outside the scope of these situations. For instance, distributions of individual assets prices can typically be estimated at many different times, and this information is not incorporated in the standard optimal transport problem [1] Our focus here is on a situation where the distributions of several individual asset prices are known at two future times, but nothing about the dependence structure is known (either between different assets, or between the different assets at different times). Such as situation is reasonably realistic, as options on individual stocks with a given maturitty and a wide variety of strike prices are typically widely traded; the prices of these can be used to infer the distribution of the stock price at that maturity, known as the implied risk netural measure. On the other hand, traded contracts depending on the values of two assets at a future time, or a single asset at multiple future times, are much scarcer. The optimization problem we study here, known in the literature as vectorial martingale optimal transport, yields upper and lower bounds on the arbitrage prices free of contracts depending on several assets at two future times.

A conjecture formulated in [24] is that, for a certain class of payoff functions, the maximium arises when the coupling of the assets at the first maturity time is perfectly co-monotone. If true, aside from providing insight into the extremal dependence structure of asset prices (in particular, note that the conjecture asserts the esistence of a single factor market model leading to the maximum price) this result would significantly reduce the complexity of computing the overall solution; since the dependence structure of assets

[1]Even if the payoff depends only on the values of two assets at a single time, information about the distributions at earlier times affects the allowable dependence structures at that time; these constraints are not reflected in the formulation of the standard optimal transport problem, but are captured by the model we study here.

at the first time is known explicitly, only the dependence structure at the second time as well as the coupling between the times needs to computed.

We resolve this conjecture here by proving that it is true in the case of only two assets, and providing a counterexample when there are three or more assets. In the two asset case, we exploit the monotone structure to refine a numerical method developed in [], resulting in faster and more accurate computations

The form of payoff function considered here emcompass a wide range of contracts arising naturally in applications. These include the model independent pricing of European calls and puts on a baskets, maximums and minimums of several assets, when the distributions of the individual asset prices are known at the maturity time as well as at an earlier time.

BP: I've commented out the numerical discussion here. I'd suggest incorporating a substantially shorter version into the numerical section of the paper. Something missing from the current discussion is the advantage of our method over others (why did we choose to use neural networks rather than, say, entropic regularization?

This paper is organized as follows: in the next section, we introduce the vectorial martingale optimal transport problem and recall some basic facts about it. Section 3 is reserved for the fomulation and resolution of the monotoncity conjecture, while our numerical method and results are presented in Section 4.

BP: I woul[d] incorporate streamlined dis[-]cussions of ou[r] applications (cur[-]rently section 4[)] into other sec[-]tion. Should w[e] also add a conclu[-]sion/discussion section at th[e] end? (TL: Hav[-]ing a conclusion section will hel[p] readers I believe.

2. Model

We denote $[n] := \{1, 2, ..., n\}$ for $n \in \mathbb{N}$, and let $\mathcal{P}(\Omega)$ denote the set of all probability measures (distributions) over a set Ω. Let $\vec{\mu} = (\mu_1, ..., \mu_d)$, $\vec{\nu} = (\nu_1, ..., \nu_d)$ denote vectors of probability measures (called marginals) on $\mathbb{R}$. Throughout the paper, we assume that all distributions have a finite first moment, including the marginals μ_i, ν_i, $i \in [d]$. We consider the following space of *Vectorial Martingale Transportations* (VMT) from $\vec{\mu}$ to $\vec{\nu}$ (see [48]):

$$\begin{aligned} \mathrm{VMT}(\vec{\mu}, \vec{\nu}) := \{\pi \in \mathcal{P}(\mathbb{R}^{2d}) \mid \pi = \mathrm{Law}(X, Y),\, \mathbb{E}_\pi[Y|X] = X, \\ X_i \sim \mu_i,\, Y_i \sim \nu_i \quad \forall\, i \in [d]\}. \end{aligned} \tag{2.1}$$

For a measure $\pi \in \mathcal{P}(\mathbb{R}^{2d})$, we will denote by $\pi_X \in \mathcal{P}(\mathbb{R}^d)$ and $\pi_Y \in \mathcal{P}(\mathbb{R}^d)$, respectively, its first and second time marginals; that is, if $\pi = \mathrm{Law}(X,Y)$, $\pi_X = \mathrm{Law}(X)$ and $\pi_Y = \mathrm{Law}(Y)$. We will denote by $\Pi(\vec{\mu})$ the set of couplings of the μ_i; that is,

$$\Pi(\vec{\mu}) := \{\sigma \in \mathcal{P}(\mathbb{R}^d) \mid \sigma = \mathrm{Law}(X),\ X_i \sim \mu_i \quad \forall\, i \in [d]\} \tag{2.2}$$

Clearly, if $\pi \in \mathrm{VMT}(\vec{\mu},\vec{\nu})$, then $\pi_X \in \Pi(\vec{\mu})$ and $\pi_Y \in \Pi(\vec{\nu})$ It is known that the set $\mathrm{VMT}(\vec{\mu},\vec{\nu})$ is nonempty if only if every pair of marginals μ_i, ν_i is in *convex order*, defined by

$$\mu_i \preceq_c \nu_i \text{ if and only if } \int f d\mu_i \le \int f d\nu_i \text{ for every convex function } f.$$

Thus, we will always assume $\mu_i \preceq_c \nu_i$ for all $i \in [d]$ in the VMOT problem.

Let $c : \mathbb{R}^{2d} \to \mathbb{R}$ be a (cost) function. We define the VMOT problem as

$$\text{maximize } \ \mathbb{E}_\pi[c(X,Y)] \ \text{ over } \ \pi \in \mathrm{VMT}(\vec{\mu},\vec{\nu}). \tag{2.3}$$

A solution π to (2.3) will be called a vectorial martingale optimal transport, or VMOT.

As explained in the introduction (true?), each pair of random variables (X_i, Y_i) represents an asset price process at two future maturity times $0 < t_1 < t_2$, and each martingale measure $\pi \in \mathrm{VMT}(\vec{\mu},\vec{\nu})$ represents the risk neutral probability under which $(X,Y) \in \mathbb{R}^{2d}$ becomes a vector-valued martingale. We call π a vectorial martingale transport, or VMT, if its one-dimensional marginals $\vec{\mu},\vec{\nu}$ are given, which condition is inspired by [8], [23], [26], [34], [37], [38]. [11] demonstrated that such marginal distribution information can be obtained from market data, providing theoretical support for the model-free martingale optimal transportation approach we consider in this paper. Finally, in financial terms, the cost function $c = c(x_1, ..., x_d, y_1, ..., y_d$ can represent an exotic option whose payoff is fully determined at the terminal maturity time t_2 by prices (X,Y) of the d assets at the two times, t_1, t_2. Because π cannot be observed in the financial market, we are led to consider the set of all possible laws $\mathrm{VMT}(\vec{\mu},\vec{\nu})$ given the marginal information $\vec{\mu},\vec{\nu}$. With this knowledge, the max / min value in (2.3) can be interpreted as the upper / lower price bound for c derived from the market data. We defined (2.3) as a maximization problem, but note that it can also describe a minimization problem by simply changing c to $-c$.

To ensure that the problem (2.3) is well-defined, we will make the following assumptions throughout the paper: When considering a VMOT problem given a cost function c, we assume that the marginals satisfy the following condition: there exist continuous functions $v_i \in L^1(\mu_i)$, $w_i \in L^1(\nu_i)$, $i \in [d]$, such that $|c(x,y)| \leq \sum_{i=1}^d \big(v_i(x_i) + w_i(y_i)\big)$. Note that this ensures, for any $\pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$,

$$\big|\mathbb{E}_\pi[c(X,Y)]\big| \leq \sum_i \big(\mathbb{E}_{\mu_i}[v_i(X_i)] + \mathbb{E}_{\nu_i}[w_i(Y_i)]\big) < \infty.$$

This implies that the problem (2.3) is attained (i.e., admits an optimizer) whenever c is upper-semicontinuous.

The following simple example cost function will be useful to illustrate many of the results in this paper. It relfects the mutual covariances of the assets at two future times:

$$c(x,y) = \sum_{1\leq i,j\leq d} (a_{ij}x_ix_j + b_{ij}x_iy_j + c_{ij}y_iy_j). \tag{2.4}$$

Note that for any $\pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$, we have $\mathbb{E}_\pi[X_iY_j] = \mathbb{E}_\pi[\mathbb{E}_\pi[X_iY_j|X]] = \mathbb{E}_\pi[X_i\mathbb{E}_\pi[Y_j|X]] = \mathbb{E}_\pi[X_iX_j]$ by the martingale constraint $\mathbb{E}_\pi[Y|X] = X$, and $\mathbb{E}_\pi[X_i^2] = \int_\mathbb{R} x^2 d\mu_i(x)$, $\mathbb{E}_\pi[Y_j^2] = \int_\mathbb{R} y^2 d\nu_j(y)$ are fixed by marginal constraint. Hence, we can reduce the cost (2.4) as the following form

$$c(x,y) = \sum_{1\leq i<j\leq d} (a_{ij}x_ix_j + b_{ij}y_iy_j). \tag{2.5}$$

We shall assume $a_{ij} \geq 0$, $b_{ij} \geq 0$. In particular, if $d = 2$, this becomes

$$c(x,y) = ax_1x_2 + by_1y_2. \tag{2.6}$$

The VMOT problem belongs to the class of infinite-dimensional linear programming, thus the problem admits a *dual programming* problem. For the maximization problem in (2.3), its dual problem is given by the following minimization problem

$$\inf_{(\phi_i,\psi_i,h_i)\in\Psi} \sum_{i=1}^d \Big(\int \phi_i d\mu_i + \int \psi_i d\nu_i\Big) \tag{2.7}$$

where Ψ consists of triplets $\phi_i, \psi_i : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ and $h_i : \mathbb{R}^d \to \mathbb{R}$ such that $\phi_i \in L^1(\mu_i)$, $\psi_i \in L^1(\nu_i)$, h_i is bounded for every $i \in [d]$, and

$$\sum_{i=1}^{d} \big(\phi_i(x_i) + \psi_i(y_i) + h_i(x)(y_i - x_i)\big) \geq c(x,y) \quad \forall (x,y) \in \mathbb{R}^{2d} \tag{2.8}$$

where $x = (x_1, ., , , x_d)$, $y = (y_1, ..., y_d)$. In this regard, (2.3) may be referred to as the *primal problem*.

The dual problem can also be interpreted in financial terms. Assume a financial firm is obligated to pay $c(X,Y)$ at the terminal maturity t_2 for an option c. To hedge its risk, the company may consider purchasing European options ϕ_i, ψ_j, where each payoff is based solely on a single asset at a specific time. Furthermore, the firm may consider holding h_i shares of the i^{th} asset between the t_1 and t_2 maturities, so that the payoff at t_2 is $h_i(X) \cdot (Y_i - X_i)$. Take note that h_i is a function of the past prices of all assets $\{X_i\}_{i=1}^d$. Now the inequality (2.8) imposes that the position must superhedge the liability for all possible realizations of the assets price (X,Y), and the left hand side of (2.8) yields the overall payoff of the hedging portfolio $(\phi_i, \psi_i, h_i)_{i=1}^d$. Finally, if the problem (2.3) is a minimization problem, then the dual problem is an optimal subhedging problem for which the inequality in (2.8) is reversed.

The value (2.7) represents the lowest possible cost to construct a superhedging portfolio, thus we are naturally interested in finding such an optimal (cheapest) superhedging portfolio. However, it has already been shown that the dual problem (2.7) cannot be solved within the class Ψ in general even when $d = 1$, i.e., the option c depends on a single asset (see [9], [12]), unless some suitable regularity assumption is made on the payoff function c [10]. As a result, a generalized notion of *dual attainment*, i.e., solvability of the dual problem, was introduced in [12] for the case $d = 1$ and then in [48] for $d \geq 2$. More specifically, [48] presents the following dual attainment result.

Theorem 2.1 ([48])**.** *Let $(\mu_i, \nu_i)_{i \in [d]}$ be irreducible pairs of marginals on $\mathbb{R}$. Let $c(x_1, ..., x_d, y_1, ..., y_d)$ be an upper-semicontinuous cost such that $|c(x,y)| \leq \sum_{i=1}^d \big(v_i(x_i) + w_i(y_i)\big)$ for some continuous $v_i \in L^1(\mu_i)$, $w_i \in L^1(\nu_i)$. Then there exists a* dual minimizer*, which is a triplet of functions $(\phi_i, \psi_i, h_i)_{i=1}^d$*

that satisfies (2.8) *tightly in the following pointwise manner:*

$$\sum_{i=1}^{d} \big(\phi_i(x_i) + \psi_i(y_i) + h_i(x)(y_i - x_i)\big) = c(x, y) \quad \pi - a.s. \tag{2.9}$$

for every VMOT π *which solves the primal problem* (2.3).

Here, the term *pointwise* denotes that (2.8) and (2.9) hold in a pointwise manner (that is, the equality (2.9) is satisfied for π - almost every "point" or "path" (x, y) of the assets price), and that we do not impose an integrability condition on the dual minimizer. (ϕ_i, ψ_i, h_i) are only measurable functions.

The irreducibility condition imposed on each pair of marginals generically holds for any pair of probability distributions $\mu \preceq_c \nu$ on the line in convex order. Furthermore, if the pair is not irreducible, one can perturb it in an arbitrarily small way to make the perturbed pair irreducible. We refer to [12] for more details about irreducibility. [2]

3. Monotone geometry of VMOT

Let us start with the notion of sub/supermodularity of functions on $\mathbb{R}^d$.

Definition 3.1. *For $a, b \in \mathbb{R}^d$, set $a \vee b$ to be the componentwise maximum of a, b and $a \wedge b$ to be the componentwise minimum, so that $(a \vee b)_i = \max\{a_i, b_i\}$ and $(a \wedge b)_i = \min\{a_i, b_i\}$. Let $d \geq 2$, and $\beta : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a function. Then submodularity and supermodularity of β reads, for all $a, b \in \mathbb{R}^d$,*

$$\beta(a) + \beta(b) \geq \beta(a \vee b) + \beta(a \wedge b), \tag{3.1}$$

$$\beta(a) + \beta(b) \leq \beta(a \vee b) + \beta(a \wedge b), \tag{3.2}$$

respectively. Finally, a function is called strictly sub / supermodular if the above inequality is strict for all $a, b \in \mathbb{R}^d$ with $\{a, b\} \neq \{a \vee b, a \wedge b\}$.

If β is twice differentiable, then β is supermodular if $\frac{\partial^2 \beta}{\partial x_i \partial x_j} \geq 0$ for all $i \neq j$, and is strictly supermodular if $\frac{\partial^2 \beta}{\partial x_i \partial x_j} > 0$ for all $i \neq j$. Hence, for example, the function $x \mapsto \sum_{1 \leq i < j \leq d} x_i x_j$ is strictly supermodular.

[2]For $\xi \in \mathcal{P}(\mathbb{R})$, its potential function is given by $u_\xi(x) := \int |x - y| d\xi(y)$. Then we say that a pair of probabilities $\mu \preceq_c \nu$ in convex order is *irreducible* if the set $I := \{x \in \mathbb{R} \,|\, u_\mu(x) < u_\nu(x)\}$ is a connected interval containing the full mass of μ, i.e., $\mu(I) = \mu(\mathbb{R})$.

Definition 3.2. *i)* $\{a \vee b, a \wedge b\}$ *is called the monotone rearrangement of* $\{a, b\}$.
ii) A set $A \subseteq \mathbb{R}^d$ *is monotone if for any* $a, b \in A$, $\{a, b\} = \{a \vee b, a \wedge b\}$.
iii) A probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ *is monotone (or monotonically supported) if there is a monotone set* A *such that* μ *is supported on* A, *i.e.,* $\mu(A) = 1$.
iv) Given a vector of probabilities $\vec{\mu} = (\mu_1, ..., \mu_d)$ *where each* $\mu_i \in \mathcal{P}(\mathbb{R})$, *the unique probability measure* $\chi_{\vec{\mu}} \in \mathcal{P}(\mathbb{R}^d)$, *which is monotone and has* $\mu_1, ..., \mu_d$ *as its marginals (i.e.,* $\chi_{\vec{\mu}} \in \Pi(\vec{\mu})$*), is called the monotone coupling of* $\vec{\mu}$. *Note that* $\chi_{\vec{\mu}} = (F_{\mu_1}^{-1}, F_{\mu_2}^{-1}, , ..., F_{\mu_d}^{-1})_{\#}\mathcal{L}_{[0,1]}$, *where each* $F_{\mu_i}^{-1}$ *denotes the inverse of the cumulative distribution function of the* μ_i *(that is, the quantile function).*

Fact 1. *It is known that if* c *is supermodular, the monotone coupling* $\chi_{\vec{\mu}}$ *arises as a maximizer of* $\mathbb{E}_\gamma[c(X)] = \int_{\mathbb{R}^d} c(x)d\gamma(x)$ *among all* $\gamma \in \Pi(\vec{\mu})$, *and that* $\chi_{\vec{\mu}}$ *is the unique maximizer if* c *is strictly supermodular.*

The following is the main question we will investigate in this section.BP: Did [24] formulate this conjecture? If not, we should at least say that it is suggested by their work.

Conjecture 1. *Let* $d \geq 2$, $\mu_i \preceq_c \nu_i$ *for* $i = 1, ..., d$, *and the cost function be given by* $c(x, y) = c_1(x) + c_2(y)$ *where* $x, y \in \mathbb{R}^d$ *and* c_1, c_2 *are supermodular. Then there exists a VMOT* π *for the problem* (2.3) *whose first marginal* π^1 *is the monotone coupling of* $\vec{\mu} = (\mu_1, ..., \mu_d)$. *Moreover, if* c_1 *is strictly supermodular, then every VMOT* π *has monotone first marginal* $\pi^1 = \chi_{\vec{\mu}}$.

Monotonicity of optimizers in the classical optimal transport problem for supermodular costs is well known [17][49]. Results asserting higher dimensional determininstic solutions, such as those of Brenier [13] (for two marginals) and Gangbo-Świech [28] (for three or more marginals) the cost function $c(x) = \sum_{1 \leq i < j \leq d} x_i \cdot x_j$, and generalizations to other cost functions [47], [27] [15] (for two marginals) [33][54][40][53] (for several marginals) can be thought of as higher dimensional analogues of this monotonicity. Our conjecture can be thought of as a vectorial martingale transport version of such a stream of results; indeed, note that if each μ_i is a dirac mass (corresponding to the case when the first time is the present), the the VMOT problem reduces to the classical (multi-marginal) optimal transport problem on the ν_i's. The following is a heuristic for the conjecture:

Heuristic. Given marginals $\vec{\mu} = (\mu_1, ..., \mu_d)$ and $\vec{\nu} = (\nu_1, ..., \nu_d)$ and cost function $c(x, y) = c_1(x) + c_2(y)$ for the vectorial martingale optimal transport problem (2.3), where c_1, c_2 are both supermodular, in view of Fact 1, the ideal situation would be that the first and second time marginals of $\pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$ (denoted as π^X, π^Y) are equal to the monotone coupling of $\vec{\mu}$ and $\vec{\nu}$ respectively, i.e., $\pi^X = \chi_{\vec{\mu}}$ and $\pi^2 = \chi_{\vec{\nu}}$, such that $\mathbb{E}_\pi[c(X, Y)] = \mathbb{E}_{\chi_{\vec{\mu}}}[c_1(X)] + \mathbb{E}_{\chi_{\vec{\nu}}}[c_2(Y)]$. However, the martingale constraint imposed on π implies the convex order condition $\pi^X \preceq_c \pi^2$. Now even if $\mu_i \preceq_c \nu_i$ for all i, the monotone couplings $\chi_{\vec{\mu}}$ and $\chi_{\vec{\nu}}$ may not satisfy the convex order in general, in which case the ideal case is not feasible. As an example, consider the following: Let $d = 2$ and $\mu_1 = \mu_2$ be the uniform probability measure on the interval $[-1, 1]$, ν_1 be uniform on $[-3, 3]$, and ν_2 be uniform on $[-2, 2]$. Then $\chi_{\vec{\mu}}$ is uniform on $l_1 = \{(x_1, x_2) \in \mathbb{R}^2 \,|\, x_1 = x_2,\, x_1 \in [-1, 1]\}$, and $\chi_{\vec{\nu}}$ is uniform on $l_2 = \{(y_1, y_2) \in \mathbb{R}^2 \,|\, y_2 = \frac{2}{3} y_1,\, y_1 \in [-3, 3]\}$. Then $\chi_{\vec{\mu}}$, $\chi_{\vec{\nu}}$ cannot be in convex order because $l_1 \nsubseteq l_2$. This shows that the ideal case is infeasible in general.

Nevertheless, it is plausible that a VMOT π may couple the marginals $\vec{\mu}$ monotonically[3], i.e., a VMOT π sets $\pi^X = \chi_{\vec{\mu}}$ thereby maximizing $\mathbb{E}_{\pi^X}[c_1(X)]$, then seek π^2 which satisfies $\chi_{\vec{\mu}} \preceq_c \pi^Y$ while π^Y is as close as the ideal $\chi_{\vec{\nu}}$, so that πY maximizes $\mathbb{E}_{\pi^Y}[c_2(Y)]$ under the constraint $\pi^Y \in \Pi(\vec{\nu})$ and $\chi_{\vec{\mu}} \preceq_c \pi^Y$. This is our heuristic behind the conjecture; see Figure xxx.

It turns out that the conjecture is indeed correct if $d = 2$, but incorrect in general if $d \geq 3$. This dimensional bifurcation stands in stark contrast to the standard optimal transport problem, in which Fact 1 holds for all $d \geq 2$. The distinction is due to the convex ordering constraint $\pi^X \preceq_c \pi^Y$, which every martingale transport π must satisfy. The rest of this section will go over our findings in greater detail. To begin, we recognize that the following relationship between modularity and convex conjugate is closely

[3]It can be shown that for any $\gamma \in \Pi(\vec{\mu})$, there exists $\pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$ such that $\pi^X = \gamma$.

related to our conjecture, which also contrasts intriguingly with standard optimal transport problems.

Definition 3.3. *For a proper function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, its convex conjugate f^* is the following convex lower-semicontinuous function*

$$f^*(y) = \sup_{x \in \mathbb{R}^d} x \cdot y - f(x), \quad y \in \mathbb{R}^d. \tag{3.3}$$

It is well known that $f^{**} = (f^*)^*$ is the largest convex lower-semicontinuous function satisfying $f^{**} \leq f$. We call f^{**} the *convex envelope* of f.

Proposition 3.4. *i) If β on $\mathbb{R}^d$ is submodular, then β^* is supermodular.*
ii) If $d = 2$ and β on $\mathbb{R}^2$ is supermodular, then β^* is submodular.
iii) If β on $\mathbb{R}^2$ is sub/supermodular, then β^{**} is also sub/supermodular.

An appendix contains a proof of the proposition. Now we present our first main result, which provides an affirmative case for the conjecture.

BP: Do we really need the proof (since we have a reference)? (TL: technically it is not necessary...we can remove and give a reference)

Theorem 3.5. *Conjecture 1 is true if $d = 2$. More specifically, let $c(x, y) = c_1(x_1, x_2) + c_2(y_1, y_2)$ where c_1, c_2 are supermodular, and assume the same condition as in Theorem 2.1. Assume that the second moments of μ_1, μ_2 are finite. Then:*
i) There exists a VMOT π such that its first marginal π^X is the monotone coupling of μ_1, μ_2.
ii) If c_1 is strictly supermodular, then every VMOT π satisfies that its first marginal π^X is the monotone coupling of μ_1, μ_2.

Proof. Theorem 2.1 implies there exists an optimal dual $(\varphi_i, \psi_i, h_i)_i$ such that

$$\sum_{i=1}^{d} \big(\varphi_i(x_i) + \psi_i(y_i) + h_i(x)(y_i - x_i)\big) \geq c(x, y) \quad \forall x = (x_1, ..., x_d),\ y = (y_1, ..., y_d),$$

$$\sum_{i=1}^{d} \big(\varphi_i(x_i) + \psi_i(y_i) + h_i(x)(y_i - x_i)\big) = c(x, y) \quad \pi - a.s.,$$

for every VMOT π which solves the problem (2.3). We define

$$\beta(y) = \sum_{i=1}^{d} \psi_i(y_i) - c_2(y), \tag{3.4}$$

and rewrite the above as

$$c_1(x) - \sum_{i=1}^{d} \big(\varphi_i(x_i) + h_i(x)(y_i - x_i)\big) \le \beta(y) \quad \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d \tag{3.5}$$

$$c_1(x) - \sum_{i=1}^{d} \big(\varphi_i(x_i) + h_i(x)(y_i - x_i)\big) = \beta(y) \quad \pi - a.s.. \tag{3.6}$$

As a result of the left hand side being linear in y, we have

$$c_1(x) - \sum_{i=1}^{d} \big(\varphi_i(x_i) + h_i(x)(y_i - x_i)\big) \le \beta^{**}(y) \quad \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d \tag{3.7}$$

$$c_1(x) - \sum_{i=1}^{d} \big(\varphi_i(x_i) + h_i(x)(y_i - x_i)\big) = \beta^{**}(y) \quad \pi - a.s.. \tag{3.8}$$

Then by equating $y_i = x_i$, (3.7) yields

$$c_1(x) - \sum_{i=1}^{d} \varphi_i(x_i) \le \beta^{**}(x) \quad \forall x \in \mathbb{R}^d. \tag{3.9}$$

On the other hand, for any VMOT $\pi = \pi_x \otimes \pi^X$, by integrating (3.8) with respect to the martingale kernel $\pi_x(dy)$, we obtain

$$c_1(x) - \sum_{i=1}^{d} \varphi_i(x_i) = \int \beta^{**}(y) d\pi_x(y) \quad \pi^X - a.e.\, x, \tag{3.10}$$

since $\int h(x) \cdot (y - x) d\pi_x(y) = 0$ due to the martingale property $\int y\, d\pi_x(y) = x$. Now we have $\int \beta^{**}(y) d\pi_x(y) \ge \beta^{**}(x)$, since β^{**} is convex. Then (3.9) yields

$$c_1(x) - \sum_{i=1}^{d} \varphi_i(x_i) = \beta^{**}(x) \quad \pi^X - a.s.. \tag{3.11}$$

Set $\tilde{c}(x) := c_1(x) - \beta^{**}(x)$. We arrive at

$$\sum_{i=1}^{d} \varphi_i(x_i) \ge \tilde{c}(x) \quad \forall x \in \mathbb{R}^d, \tag{3.12}$$

$$\sum_{i=1}^{d} \varphi_i(x_i) = \tilde{c}(x) \quad \pi^X - a.s.. \tag{3.13}$$

(3.12) and (3.13) implies that for any VMOT π, its first marginal π^X solves the optimal transport problem with the cost $\tilde{c}$ and marginals $\mu_1, ..., \mu_d$, that is, π^X maximizes $\mathbb{E}[\tilde{c}(X)]$ among all couplings of $\mu_1, ..., \mu_d$.

Now assume $d = 2$ and c_1 is strictly supermodular. Then by Proposition 3.4 iii), $\tilde{c}$ is also strictly supermodular. Then Fact 1 implies that π^X must be the monotone coupling of μ_1, μ_2. This proves part ii) of the theorem.

To prove part i), fix $\delta > 0$, and choose a VMOT π for the cost $c_\delta(x, y) = c_1(x) + \delta x_1 x_2 + c_2(y)$. Then by part ii) we have $\pi^X = \chi_{\vec{\mu}}$, i.e., π^X is the monotone coupling of $\vec{\mu} = (\mu_1, \mu_2)$. Moreover, since π is a VMOT, its second marginal π^2 must maximize $\mathbb{E}_\gamma[c_2(Y)]$ among all couplings $\gamma \in \Pi(\nu_1, \nu_2)$ satisfying the convex order $\chi_{\vec{\mu}} \preceq_c \gamma$. This then implies that π is a VMOT for the cost $c_\delta(x, y)$ for every $\delta > 0$. Letting $\delta \searrow 0$, we deduce that π is still a VMOT for the cost $c(x, y) = c_1(x) + c_2(y)$. This proves part i). □

The preceding proof, combined with Proposition 3.4 iii), also yields the following mirror statement. We say that a set A in $\mathbb{R}^2$ is called *anti-monotone* if the set $\{x = (x_1, x_2) \in \mathbb{R}^2 \,|\, (-x_1, x_2) \in A\}$ is monotone. Then a measure $\mu \in \mathcal{P}(\mathbb{R}^2)$ is called anti-monotone if μ is supported on an anti-monotone set.

Corollary 3.6. *Let $d = 2$, $c(x, y) = c_1(x_1, x_2) + c_2(y_1, y_2)$ where c_1, c_2 are submodular, and assume the same condition as in Theorem 2.1. Assume that the second moments of μ_1, μ_2 are finite. Then:*
i) There exists a VMOT π such that its first marginal π^X is the anti-monotone coupling of μ_1, μ_2.
ii) If c_1 is strictly submodular, then every VMOT π satisfies that its first marginal π^X is the anti-monotone coupling of μ_1, μ_2.

In Theorem 3.5, if c_1 is not strictly supermodular, then the first marginal π^X of a VMOT π is not necessarily monotone. The following example illustrates this point.

Example 3.7. *The strict supermodularity of c_1 is necessary for part ii) of Theorem 3.5. To construct a counterexample VMOT π via duality, we take a convex function $\psi_1(y_1) = \frac{1}{3}|y_1|^3$ and its convex conjugate $\psi_2(y_2) = \psi_1^*(y_2) =$*

$\frac{2}{3}|y_2|^{\frac{3}{2}}$. *Then* $\psi_1(y_1) + \psi_2(y_2) \geq y_1 y_2$ *for all* $y_1, y_2 \in \mathbb{R}$, *and*

$$\begin{aligned} \Gamma_{\{\psi_1,\psi_2\}} &:= \{(y_1, y_2) \in \mathbb{R}^2 \mid \psi_1(y_1) + \psi_2(y_2) = y_1 y_2\} \\ &= \{(y_1, y_2) \mid y_2 = \psi_1'(y_1)\} \\ &= \{(y_1, y_2) \mid y_2 = |y_1|^2 \text{ if } y_1 \geq 0,\ y_2 = -|y_1|^2 \text{ if } y_1 \leq 0\}. \end{aligned} \tag{3.14}$$

Let $z = (-1, 1)$, $w = (1, -1)$, *and take* $\pi^X = \frac{1}{2}\delta_z + \frac{1}{2}\delta_w \in \mathcal{P}(\mathbb{R}^2)$. *Then choose a martingale kernel* $\pi_z, \pi_w \in \mathcal{P}(\mathbb{R}^2)$ *that satisfies*

$$\int_{\mathbb{R}^2} x\, \pi_z(dx) = z, \int_{\mathbb{R}^2} x\, \pi_w(dx) = w, \text{ and } \pi_z(\Gamma_{\{\psi_1,\psi_2\}}) = \pi_w(\Gamma_{\{\psi_1,\psi_2\}}) = 1.$$

Such a choice is possible because $\operatorname{conv}(\Gamma_{\{g_1,g_2\}}) = \mathbb{R}^2$. *We then define a martingale measure* π *via* $\pi = \pi_x \otimes \pi^X$, *i.e., its first marginal is* π^X *and its kernel is* $\{\pi_z, \pi_w\}$. *Now take* $c(x, y) = y_1 y_2$ *(so that* $c_1(x) = 0$*),* $\varphi_1 = \varphi_2 = h_1 = h_2 = 0$, *and notice that* $\{\phi_i, \psi_i, h_i\}_{i=1,2}$ *and* π *then jointly satisfy the optimality condition* (2.8), (2.9). *This implies that* π *is a VMOT in the class* $\mathrm{VMT}(\mu_1, \mu_2, \nu_1, \nu_2)$, *where* $\mu_1, \mu_2, \nu_1, \nu_2$ *are the one-dimensional marginals of* π. *However, by construction,* $\pi^X = \frac{1}{2}\delta_z + \frac{1}{2}\delta_w$ *is not monotone. (Figure?)*

As previously demonstrated, part iii) of Proposition 3.4 was used as a key to the proof of Theorem 3.5. Because iii) is a direct consequence of i) and ii), where ii) is restricted to the two-dimensional domain, we are led to ask whether part ii) can be extended for $d \geq 3$ as part i). Unfortunately, this is not the case, as demonstrated by the following example.

Example 3.8. *In* $\mathbb{R}^3$, *there is a supermodular function* $\beta : \mathbb{R}^3 \to \mathbb{R}$ *for which* β^* *is not submodular. As an example, for* $x \in \mathbb{R}^3$, *let* β *be the quadratic function* $\beta(x) = \frac{1}{2}x \cdot Ax$ *with a symmetric matrix* A *given by*

$$A = \begin{bmatrix} 10 & 3 & 4 \\ 3 & 12 & 13 \\ 4 & 13 & 16 \end{bmatrix}.$$

It can be checked that all the eigenvalues of A *are positive, so* β *is a strongly convex function, in which case* β^* *is given by*

$$\beta^*(z) = \frac{1}{2} z \cdot A^{-1} z.$$

In this case, A^{-1} is given by

$$\frac{1}{206}\begin{bmatrix} 23 & 4 & -9 \\ 4 & 144 & -118 \\ -9 & -118 & 111 \end{bmatrix}. \tag{3.15}$$

If β^* is submodular, then its all mixed partials should be nonpositive, i.e.,

$$\frac{\partial^2 \beta^*}{\partial x_i \partial x_j} \leq 0 \quad \text{for every } i \neq j.$$

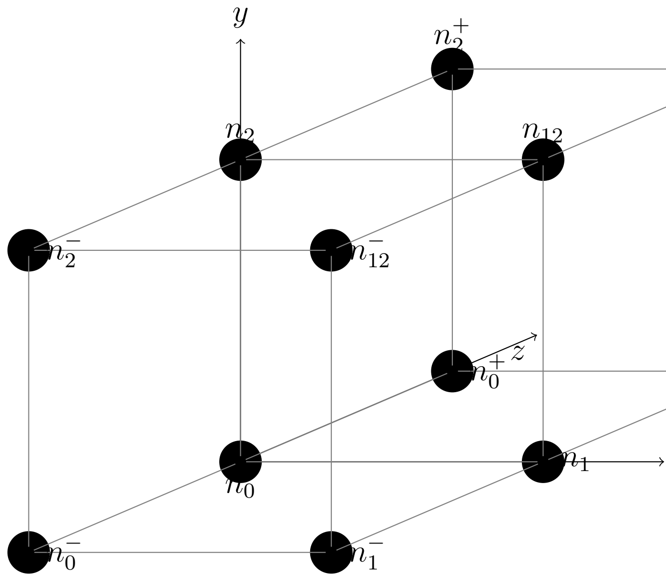
However, we see that A^{-1} has some positive off-diagonal entries, which implies that β^* is not submodular.

Despite Example 3.8, we continue to ask whether part iii) of Proposition 3.4 can hold true for $d \geq 3$, as it is the only part required to prove Theorem 3.5. It turns out this is also not the case, as the following example shows.

Example 3.9 (Existence of a submodular function on $\mathbb{R}^3$ whose convex envelope is not submodular)**.** *Let $n_0^+ = (0,0,1)$, $n_1^+ = (1,0,1)$, $n_2^+ = (0,1,1)$, $n_{12}^+ = (1,1,1)$, $n_0 = (0,0,0)$, $n_1 = (1,0,0)$, $n_2 = (0,1,0)$, $n_{12} = (1,1,0)$, $n_0^- = (0,0,-1)$, $n_1^- = (1,0,-1)$, $n_2^- = (0,1,-1)$, $n_{12}^- = (1,1,-1)$ be the vertices of two vertically stacked cubes in $\mathbb{R}^3$, and let $\mathcal{Y} \subseteq \mathbb{R}^3$ be the set of these twelve vertices. We then define $\beta_0 : \mathbb{R}^3 \to \mathbb{R} \cup \{+\infty\}$ as follows:*

$$\begin{aligned}
&\beta_0(n_0^+) = 0,\ \beta_0(n_1^+) = 0,\ \beta_0(n_2^+) = 0,\ \beta_0(n_{12}^+) = 0,\\
&\beta_0(n_0) = 0,\ \beta_0(n_1) = 1,\ \beta_0(n_2) = 0,\ \beta_0(n_{12}) = 1,\\
&\beta_0(n_0^-) = 0,\ \beta_0(n_1^-) = 2,\ \beta_0(n_2^-) = 1,\ \beta_0(n_{12}^-) = 2,\\
&\beta_0 = +\infty \quad \text{on} \quad \mathbb{R}^3 \setminus \mathcal{Y}.
\end{aligned}$$

(TL: need to modify the figure...little smaller and the dot size reduced, centered, etc)



It is easy to check that β_0 is submodular, and moreover, β_0^{**} is given by the supremum of three affine functions; $\beta_0^{**} = \max(L_1, L_2, L_3)$ in $\mathrm{conv}(\mathcal{Y})$, where

$$L_1(y) = 0 \quad \textit{for} \quad y = (y_1, y_2, y_3) \in \mathbb{R}^3,$$
$$L_2(y) = y_1 + y_2 - y_3 - 1,$$
$$L_3(y) = 2y_1 - y_3 - 1.$$

One can further check that $\beta_0 = \beta_0^{**}$ on $\mathcal{Y}$, and (check all claims carefully)

$$H_{12} := \{y \in \mathrm{conv}(\mathcal{Y}) \mid L_1(y) = L_2(y) \geq L_3(y)\} = \mathrm{conv}(\{n_0^-, n_2, n_{12}^+\}),$$
$$H_{13} := \{y \in \mathrm{conv}(\mathcal{Y}) \mid L_1(y) = L_3(y) \geq L_2(y)\} = \mathrm{conv}(\{n_0^-, n_1^+, n_{12}^+\}),$$
$$H_{23} := \{y \in \mathrm{conv}(\mathcal{Y}) \mid L_2(y) = L_3(y) \geq L_1(y)\} = \mathrm{conv}(\{n_0^-, n_{12}^-, n_{12}^+\}).$$

We can see that H_{12} has the normal direction $(1,1,-1)$, which is not of the form $(a,-b,0)$, $(a,0,-b)$, or $(0,a,-b)$ for any $a,b \geq 0$. This implies that β_0^{**} is not a submodular function. To provide details, we can find two distinct points u, u' in H_{12} such that its monotone rearrangement $\bar{u}, \bar{u}'$ is not in the plane containing H_{12}. For example, one may take $u = \frac{1}{4}n_0^- + \frac{1}{4}n_2 + \frac{1}{2}n_{12}^+ = (\frac{1}{2}, \frac{3}{4}, \frac{1}{4})$, $u' = \frac{1}{5}n_0^- + \frac{2}{5}n_2 + \frac{2}{5}n_{12}^+ = (\frac{2}{5}, \frac{4}{5}, \frac{1}{5})$, so that $\bar{u} = (\frac{2}{5}, \frac{3}{4}, \frac{1}{5})$, $\bar{u}' = (\frac{1}{2}, \frac{4}{5}, \frac{1}{4})$. We see that none of $\bar{u}, \bar{u}'$ lies on the plane containing H_{12}, and we have

$L_1(\bar u) > L_2(\bar u)$ and $L_1(\bar u') < L_2(\bar u')$. Then we have

$$\begin{aligned}\beta_0^{**}(u) + \beta_0^{**}(u') &= \left(\frac{L_1+L_2}{2}\right)(u+u') \\ &= \left(\frac{L_1+L_2}{2}\right)(\bar u+\bar u') \\ &< L_1(\bar u) + L_2(\bar u') \\ &\le \beta_0^{**}(\bar u) + \beta_0^{**}(\bar u'),\end{aligned}$$

yielding that β_0^{**} is not submodular.

BP: Are all of these examples really needed?

The failure of Proposition 3.4 iii) for $d \geq 3$ reduces the plausibility of Conjecture 1. Nevertheless, we continue to suspect that the conjecture may still be true for $d \geq 3$ because, while Proposition 3.4 iii) is sufficient to yield the conjecture, it may not be strictly necessary. Moreover, the heuristic is still appealing.

However, a closer examination of the submodular function and its convex envelop in the preceding example eventually lead us to construct the following counterexample.

Proposition 3.10. *Conjecture 1 is false if $d \geq 3$. More specifically, there exist vectorial marginals $\vec\mu = (\mu_1, \mu_2, \mu_3)$, $\vec\nu = (\nu_1, \nu_2, \nu_3)$ satisfying $\mu_i \preceq_c \nu_i$, $i = 1, 2, 3$, and a supermodular cost function $c = c(y) : \mathbb{R}^3 \to \mathbb{R}$ (e.g., $c(y) = y_1y_2 + y_2y_3 + y_3y_1$), such that for every VMOT π to the problem (2.3), its first marginal π^1 fails to be the monotone coupling of $\vec\mu = (\mu_1, \mu_2, \mu_3)$.*

BP: I don't completely understand the statement of the proposition. Is it true for every supermodular c?

Proof. Let $\mathcal{Y}$ be the set of twelve points in $\mathbb{R}^3$ and β_0 be the submodular function as in Example 3.9. Define $\psi_1, \psi_2, \psi_3 : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ by

$$\begin{aligned}&\psi_1(0) = 0,\ \psi_1(1) = 2,\ \psi_1 = +\infty \ \text{ else},\\ &\psi_2(0) = 0,\ \psi_2(1) = 0,\ \psi_2 = +\infty \ \text{ else},\\ &\psi_3(-1) = 0,\ \psi_3(0) = 0,\ \psi_3(1) = 1,\ \psi_3 = +\infty \ \text{ else}.\end{aligned}$$

Set $\beta(y) = \sum_{i=1}^{3} \psi_i(y_i) - c(y)$, where $c(y) = y_1y_2 + y_2y_3 + y_3y_1$. We have

$$\begin{aligned}
&\beta(n_0^+) = 0,\ \beta(n_1^+) = 2,\ \beta(n_2^+) = 0,\ \beta(n_{12}^+) = 0,\\
&\beta(n_0) = 0,\ \beta(n_1) = 2,\ \beta(n_2) = 0,\ \beta(n_{12}) = 1,\\
&\beta(n_0^-) = 0,\ \beta(n_1^-) = 3,\ \beta(n_2^-) = 1,\ \beta(n_{12}^-) = 3,\\
&\beta = +\infty \ \text{ on } \ \mathbb{R}^3 \setminus \mathcal{Y}.
\end{aligned}$$

Notice $\beta \geq \beta_0$, and $\beta = \beta_0$ on $\mathcal{Z} := \{n_0^-, n_2, n_{12}^+\}$. In Example 3.9, we observed $\beta_0 = \beta_0^{**}$ on $\mathcal{Z}$ (in fact also on $\mathcal{Y}$), hence $\beta = \beta^{**}$ on $\mathcal{Z}$ as well. Then as in Example 3.9, we may take $u = (\frac{1}{2}, \frac{3}{4}, \frac{1}{4})$, $u' = (\frac{2}{5}, \frac{4}{5}, \frac{1}{5})$ and their monotone rearrangement $\bar{u} = (\frac{2}{5}, \frac{3}{4}, \frac{1}{5})$, $\bar{u}' = (\frac{1}{2}, \frac{4}{5}, \frac{1}{4})$, such that $\{u, u'\} \subseteq \operatorname{conv}(\mathcal{Z})$, while none of $\bar{u}, \bar{u}'$ lies on the plane containing $\mathcal{Z}$.

We now construct a vectorial martingale transport π. For this, we take $\pi^1 := \frac{1}{2}(\delta_u + \delta_{u'})$ as the first marginal of π. Then we take the martingale kernel π_x as the unique probability measure supported on $\mathcal{Z}$ with its barycenter x in $\operatorname{conv}(\mathcal{Z})$. Now define $\pi = \pi_x \otimes \pi^1$ (note that we only need π_x for $x = u, u'$), and let $\vec{\mu} := (\mu_1, \mu_2, \mu_3)$ be the 1D marginals of π^1, and let $\vec{\nu} := (\nu_1, \nu_2, \nu_3)$ be the 1D marginals of $\pi^2 = \frac{1}{2}(\pi_u + \pi_{u'})$. We now claim that π is a VMOT solving the problem (2.3) with the cost c and the marginals $\vec{\mu}, \vec{\nu}$.

We will demonstrate the optimality of π by locating an associated dual optimizer $(\phi_i, \psi_i, h_i)_{i=1,2,3}$, where ψ_i has already been defined above. To define ϕ_i, recall the affine functions L_1, L_2, L_3 in Example 3.9. We have $\frac{L_1+L_2}{2}(x) = \frac{1}{2}x_1 + \frac{1}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}$. We take $\phi_1(x_1) = \frac{1}{2}x_1$, $\phi_2(x_2) = \frac{1}{2}x_2$, $\phi_3(x_3) = -\frac{1}{2}x_3 - \frac{1}{2}$, so that $\sum_{i=1}^{3} \phi_i = \frac{L_1+L_2}{2}$. Then we take $h(x) = (h_1(x), h_2(x), h_3(x))$ as

$$h(x) = \begin{cases} \nabla L_1 = 0 \ \text{ if } \ L_1(x) > L_2(x), \\ \nabla L_2 = (1, 1, -1) \ \text{ if } \ L_1(x) < L_2(x), \\ \nabla \frac{L_1+L_2}{2} = (\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}) \ \text{ if } \ L_1(x) = L_2(x). \end{cases} \tag{3.16}$$

In order to prove the optimality of π and (ϕ, ψ, h) simultaneously, we need to confirm the optimality conditions (2.8) and (2.9). To see (2.8), observe

$$\begin{aligned}\sum_{i=1}^{3} \phi_i(x_i) + h(x)\cdot(y-x) &= \left(\frac{L_1+L_2}{2}\right)(x) + h(x)\cdot(y-x)\\ &\le \max(L_1, L_2)(y)\\ &\le \beta_0^{**}(y) = \max(L_1, L_2, L_3)(y)\\ &\le \beta^{**}(y)\\ &\le \beta(y) = \sum_{i=1}^{3} \psi_i(y_i) - c(y).\end{aligned}$$

Observe further that (2.9) follows by the fact that on $\mathcal{Z}$, $L_3 \le L_1 = L_2 = \beta$, such that the above equality become equality for $x = u, u'$ and $y \in \mathcal{Z}$.

Finally, take any $\gamma \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$, such that its first marginal γ^1 is monotone, i.e., $\gamma^1 = \frac{1}{2}(\delta_{\bar{u}} + \delta_{\bar{u}'})$. We claim that γ cannot be optimal. If γ were optimal, it must satisfy the optimality condition (2.9) with the optimal dual (ϕ, ψ, h) constructed above. However, we have $L_1(\bar{u}) > L_2(\bar{u})$ and $L_1(\bar{u}') < L_2(\bar{u}')$, and this clearly implies the strict inequality

$$\left(\frac{L_1+L_2}{2}\right)(x) + h(x)\cdot(y-x) < \max(L_1, L_2)(y) \quad \text{for } x = \bar{u}, \bar{u}' \text{ and } y \in \mathcal{Y}.$$

This indicates that γ cannot satisfy (2.9), thereby completing the proof. $\square$

Remark 3.11. *We showed that the VMOT π constructed in Proposition 3.10 cannot have a monotone first marginal π^1. In other words, we showed*

$$\mathbb{E}_\pi[c(Y)] > \max_{\gamma \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})} \left\{ \mathbb{E}_\gamma[c(Y)] \mid \gamma^1 \textit{ is monotone; } \gamma^1 = \frac{\delta_{\bar{u}} + \delta_{\bar{u}'}}{2} \right\} \tag{3.17}$$

where $c(y) = y_1y_2 + y_2y_3 + y_3y_1$. Now let us consider the cost function $c_\lambda(x, y) = \lambda c(x) + c(y) = \lambda(x_1x_2 + x_2x_3 + x_3x_1) + y_1y_2 + y_2y_3 + y_3y_1$ for $\lambda \ge 0$. Because the inequality (3.17) is strict, it remains strict for the cost c_λ with sufficiently small positive λ. In other words, even if the cost function involves $\lambda c(x)$ which is strictly supermodular, π^1 is still not monotone for every VMOT π, as long as λ is not too large. However, since

$$c(u) + c(u') < c(\bar{u}) + c(\bar{u}'), \tag{3.18}$$

inequality (3.17) *is reversed for* $c = c_\lambda$ *with sufficiently large* λ*, in which case every VMOT* π *now have the monotone first marginal* $\frac{1}{2}(\delta_{\bar{u}} + \delta_{\bar{u}'})$*. We thus observe a tension between* $\mathbb{E}[c(x)]$ *and* $\mathbb{E}[c(y)]$ *for the geometry of VMOT.*

The intuition is as follows: Because it is more important to maximize $\mathbb{E}[c(Y)]$ *for small* λ*, a VMOT* π *promotes its second marginal* π^2 *to be supported on the monotone set* $\mathcal{Z}$*, even if this necessitates supporting its first marginal* π^1 *on a non-monotone set* $\{u, u'\}$*. However, as* λ *grows larger, maximizing* $\mathbb{E}[c(X)]$ *becomes more important, so a VMOT* π *promotes its first marginal* π^1 *to be supported on the monotone set* $\{\bar{u}, \bar{u}'\}$ *even if this requires* π^2 *to be supported on a non-monotone set (while each of the kernels* $\pi_{\bar{u}}, \pi_{\bar{u}'}$ *being kept monotone supported). The tension is caused by the martingale constraint of the problem* (2.3)*, which distinguishes the vectorial martingale optimal transport problem from the standard multi-marginal optimal transport problem in an interesting way.*

Remark 3.12. *The functions* ψ_i *appearing as part of dual optimizers in Example 3.10 appear quite singular. However, they can be made continuous and convex by applying the martingale Legendre transform [30]. Recall* (2.8)*, which we rewrite in this case as*

$$\psi_1(y_1) \geq c(y) - \sum_{i=2}^{3} \psi_i(y_i) - \sum_{i=1}^{3} \big(\varphi_i(x_i) + h_i(x)(y_i - x_i)\big),$$

which holds for all $x, y \in \mathbb{R}^3$*. In view of this, the martingale Legendre transform of* ψ_1 *can be naturally defined by*

$$\tilde{\psi}_1(y_1) := \sup_{x_1,x_2,x_3,y_2,y_3} \left\{ c(y) - \sum_{i=2}^{3} \psi_i(y_i) - \sum_{i=1}^{3} \big(\varphi_i(x_i) + h_i(x)(y_i - x_i)\big) \right\}.$$

By definition, we have $\psi_1 \geq \tilde{\psi}_1$*. Furthermore, if* $c(y) = y_1y_2 + y_2y_3 + y_3y_1$*, we see that* $\tilde{\psi}_1$ *is convex (in this case, it is the supremum of finitely many affine functions of* y_1*). Because* ψ_1 *is finite on* $\mathcal{Y}$*, convexity of* $\tilde{\psi}_1$ *implies that it is finite in* $\operatorname{conv}(\mathcal{Y})$*. Similarly, we can replace* ψ_2 *and* ψ_3 *with their martingale Legendre transforms. Then* $(\varphi_i, \tilde{\psi}_i, h_i)_i$ *continues to be a dual optimizer.*

Remark 3.13. *[24] showed that if each pair of marginals (μ_i, ν_i) is Gaussian with equal mean and increasing variance $\mathrm{Var}(\mu_i) < \mathrm{Var}(\nu_i)$ (or more generally if each pair satisfies the linear increment of marginals condition), then the first marginal π^X of any VMOT π with respect to the cost $c = c(y) = \sum_{1\le i<j\le d} y_i y_j$ is the monotone coupling of $\vec{\mu} = (\mu_1, ..., \mu_d)$. This result, along with the heuristic for the conjecture, prompts us to suspect that, even though Conjecture 1 is false in general, as demonstrated by Proposition 3.10, the first marginal π^X of VMOT π can be monotone at least approximately for various generic marginals $(\mu_i, \nu_i)_i$ in convex order and supermodular cost c, in particular, for $c(x, y) = \lambda \sum_{1\le i<j\le d} x_i x_j + \sum_{1\le i<j\le d} y_i y_j$ with $\lambda \ge 0$.*

We believe that quantifying the approximate monotonicity of π^X of VMOT *π in terms of marginal distributions and supermodularity of the cost is an intriguing question for future research.*

BP: I'm not sure why we discuss, or exactly what we mean by, approximate optimaility here. I think it's reasonable to expect (and indeed [**?**] gives evidence for this) that the conjecture is true for some choices of marginals and supermodular costs, and determining exactly which ones is an interesting research direction, but what does bringing up potential approximate optimality really add?

4. Application in finance

4.1. Pricing multi-asset multi-period derivative.

BP: I would strongly recommend removing this section. A more concise version could be incorporated into Section 3, say, as examples of supermodular cost functions with financial relevance. A cap is a financial derivative that receives a payoff at each of the predetermined period $t = 1, ..., N$:

$$c(X_1, ..., X_N) = \sum_{t=1}^{N} c_t(X_{t,1}, ..., X_{t,d}). \tag{4.1}$$

If each of the c_t is sub/super-modular, then the corresponding 2-period cap $c(X, Y) = c_1(X_1, ..., X_d) + c_2(Y_1, ..., Y_d)$ is the financial derivative which is of our main interest.

For such a cap contract, a consequence of Theorem 2.1 implies that the robust super/sub-replication hedging do give the model independent price

Name	Payoff function
European basket call option	$(\sum_{i=1}^{d} \alpha_{t,i} X_{t,i} - K)^{+}$
European basket put option	$(K - \sum_{i=1}^{d} \alpha_i X_{t,i})^{+}$
Put on the minimum among d stocks	$(K - \min_{1\le i\le d}\{X_{t,i}\})^{+}$
Call on the maximum among d stocks	$(\max_{1\le i\le d}\{X_{t,i}\} - K)^{+}$
Covariance contract among d stocks	$\sum_i \sum_{j>i} a_{ij} X_i X_j + b_{ij} Y_i Y_j$
Call on illiquid cross rate among two currency	$(X_{t,2} - K X_{t,1})^{+}$

TABLE 1. A table of sub-modular payoff

bounds. Therefore, any arbitrage-free price $AF[c(X,Y)]$ of such a derivative should be bounded by the solutions of the VMOT problems:

$$\min_{\pi\in VMT(\vec{\mu},\vec{\nu})} \mathbb{E}_\pi[c(X,Y)] \le AF[c(X,Y)] \le \max_{\pi\in VMT(\vec{\mu},\vec{\nu})} \mathbb{E}_\pi[c(X,Y)]. \tag{4.2}$$

Moreover, if we only focus on the 2-period cap which has only two underlying assets, Theorem 3.5 can further describe the extreme market model which attains the extremal of the price bounds: there exists a market model that the two assets are indeed controlled by a single factor in the first period. If c_1 is strictly monotone, then every extremal market model enjoys this property.

To close this section, we present the table 1 which records some sub-modular payoff function which we are interested. (TL: "which we are interested" may need to rewrite)

4.2. **Pricing multi-period Foreign Exchange Derivatives with Illiquid Cross-Rates.** In foreign exchange market, not the cross-rate of every pair of currencies are actively traded. In consequence, those cross-rates may not fully reflect the information in market and miss priced. For example let X_1 be the dollar price of one Japanese Yen, and X_2 be the dollar price of one New Zealand Dollar. Let Y_{12} be the Yen price of New Zealand Dollar. We can check from data [20] that X_1 and X_2 are both actively traded pair while the volume of the exchange between Japanese Yen and New Zealand Dollar only count for a low proportion.

If we assume that there is no transaction cost for foreign exchange, then we can recover the $Y_{t,12}$ by $Y_{t,12} = X_{t,2}/X_{t,1}$. The payoff of the call option of cross rate between Japanese Yen and New Zealand Dollar maturing at t with strike K is defined as

$$CR(t, X_{t,1}, X_{t,2}) = (X_{t,2} - KX_{t,1})^+ \tag{4.3}$$

Note that this is a sub-modular payoff. Let us assuming it is zero-interest rate. With a pre-determined sequence of time $\{t_i\}_{i=1}^N$ then a foreign exchange N-period cap has the payoff:

$$CCR(X_1, ..., X_N) = \sum_{i=1}^{N} CR(t_i, X_{t_i,1}, X_{t_i,2}) = \sum_{i=1}^{N} (X_{t_i,2} - KX_{t_i,1})^+ \tag{4.4}$$

which is a contract that the cost function is of our interest. If we denote $AF[CCR(X_1, ..., X_N)]$ as the arbitrage-free price of the N-period cross-rate cap, then the VMOT of the associated payoff function will be the price bounds of it.

$$\begin{aligned}\min_{\pi \in VMT(\overrightarrow{\mu_1}, ..., \overrightarrow{\mu_N})} & \mathbb{E}_\pi[CCR(X_1, ..., X_N)] \\ & \leq AF[CCR(X_1, ..., X_N)] \\ & \quad \leq \max_{\pi \in VMT(\overrightarrow{\mu_1}, ..., \overrightarrow{\mu_N})} \mathbb{E}_\pi[CCR(X_1, ..., X_N)]\end{aligned}$$

Specifically, we provide a theoretical structural result about the optimal measures and a numerical method about the execution of the sub/super-replication of the price bounds for the 2-period cross-rate call option cap on two illiquid currency.

5. Numerics

Reformulation exploiting monotonicity An immediate consequence of Theorem 3.5 is that when $d = 2$ and $c(x, y) = c_1(x_1, x_2) + c_2(y_1, y_2)$, (2.3) is equivalent to

$$\text{maximize} \;\; \mathbb{E}_\pi[c(X, Y)] \;\; \text{over} \;\; \pi \in \text{VMT}(\mu, \vec{\nu}). \tag{5.1}$$

where

$$\begin{aligned}\mathrm{VMT}(\mu, \vec{\nu}) := \{\pi \in \mathcal{P}(\mathbb{R}^{2d}) \mid \pi = \mathrm{Law}(X, Y),\, \mathbb{E}_\pi[Y|X] = X,\\ X \sim \mu,\, Y_i \sim \nu_i \quad \forall\, i \in [d]\}.\end{aligned} \tag{5.2}$$

and $\mu \in \mathcal{P}(\mathbb{R}^d)$ is the monotone coupling of the μ_i. We note that μ may be written as $(F_1^{-1}, ..., F_d^{-1})_\# \mathcal{L}_{[0,1]}$, where each F_i^{-1} is the inverse cumulative distribution function of the corresponding μ_i. For simplicity, we also replace each ν_i for ${G_i^{-1}}_\# \mathcal{L}_{[0,1]}$ where G_i^{-1} is the inverse cumulative of ν_i. We may therefore rewrite (5.1) as (JH: Should we explicit include F, G a variable in the name, like $\mathrm{CVMT}(F, G, \mathcal{L}_{[0,1]}, \mathcal{L}_{[0,1]^d})$, in order to specific the original marginals? Then the reader may easier to agree that each $\tilde{\pi}$ in this set is associate to some martingale π in the original VMOT space.)

$$\text{maximize } \ \mathbb{E}_\pi[\tilde{c}(U, V)] \ \text{ over } \ \tilde{\pi} \in \mathrm{CVMT}(\mathcal{L}_{[0,1]}, \mathcal{L}_{[0,1]^d}), \tag{5.3}$$

where

(5.4)

$$\begin{aligned}\mathrm{CVMT}(\mathcal{L}_{[0,1]}, \mathcal{L}_{[0,1]^d}) := \{\tilde{\pi} \in \mathcal{P}([0,1] \times [0,1]^d) \mid \tilde{\pi} = \mathrm{Law}(U, V),\\ \mathbb{E}_\pi[G(V)|U] = F(U), U \sim \mathcal{L}_{[0,1]},\, V_i \sim \mathcal{L}_{[0,1]} \quad \forall\, i \in [d]\}.\end{aligned}$$

for $F(u) := (F_1^{-1}(u), F_2^{-1}(u), ..., F_d^{-1}(u))$, $v := v_1, v_2, \ldots, v_d$, $G(v) := (G_1^{-1}(v_1), G_2^{-1}(v_2), ..., G_d^{-1}(v$ and $\tilde{c}(u, v) = c(F(u), G(v))$ (shouldn't we call them F^{-1}, G^{-1}?). [4] This problem is significantly simpler than the original (2.3), due to the dimensional reduction in going from $X \in \mathbb{R}^d$ to $U \in \mathbb{R}$ and the reduction from the d constraints $X_i \sim \mu_i$ to the single constraint $U \sim \mathcal{L}_{[0,1]}$.

BP: The notation $\mathcal{L}_{[0,1]^d}$ suggests that Lebesgue measure on $[0,1]^d$ is a marginal in some sense...

The dual to (5.3) is

$$\inf_{(\tilde{\phi}, \tilde{\psi}_i, \tilde{h}_i) \in \tilde{\Psi}} \int \tilde{\phi}(u) du + \sum_{i=1}^d \int \tilde{\psi}_i dv_i \tag{5.5}$$

[4]We note carefully that although this problem can be formulated for any c and d, it is equivalent to 2.3 in general only for $d = 2$ and $c(x, y) = c_1(x_1, x_2) + c_2(y_1, y_2)$ with c_1 and c_2 supermodular.

where $\tilde{\Psi}$ consists of triplets $\tilde{\phi}, \tilde{\psi}_i : [0,1] \to \mathbb{R} \cup \{+\infty\}$ and $\tilde{h}_i : [0,1] \to \mathbb{R}$ such that $\tilde{\phi} \in L^1(\mathcal{L}_{[0,1]})$, $\tilde{\psi}_i \in L^1(\nu_i)$, $\tilde{h}_i$ is bounded for every $i \in [d]$, and

$$\tilde{\phi}(u) + \sum_{i=1}^{d} \big(\tilde{\psi}_i(v_i) + \tilde{h}_i(u)(G_i^{-1}(v_i) - F_i^{-1}(u))\big) \geq \tilde{c}(u,v) \quad \forall (u,v) \in [0,1] \times [0,1]^d. \tag{5.6}$$

As with the primal problem, this dual is much simpler than the original (2.7), as we replace the d functions ϕ_i with the single function $\tilde{\phi}$, while the functions h_i now depend on the single variable u rather than $x \in \mathbb{R}^d$. In the following, we exploit this simplified structure to develop a numerical method to compute solutions to (5.3) (and, consequently, to 2.3 for $d = 2$ and appropriate c.)

———————— BP: Somewhere around here I suggest incorporating a shorter version of the numerical literature review I removed from the introduction. It is important to justify why we choose to use the neural network approach rather than other existing methods.

Numerical implementation. Below we numerically solve the VMOT problem for several choices of marginals and costs. We apply the numerical framework introduced in [25], both to the original VMOT problem 2.3 and the reduced-dimension version 5.3 above and compare the results. Note that the solutions to the two versions are equivalent, under the conditions in Theorem 3.5, but we expect the numerical scheme to perform bettwe on the lower dimensional problem 5.3. The method in [25] uses neural network optimization with penalization to solve the dual problem of Optimal Transport. Let $\mathcal{H} \subseteq C_b(\mathbb{R}^d)$ be a linear space of bounded continuous functions which contains the constant functions. Assume $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$ is some probability measure which agree with all the given marginals $\mu_1, ..., \mu_d$, we let $\mathcal{Q} = \{\mu \in \mathcal{P}(\mathbb{R}^d) | \int \varphi d\mu = \int \varphi d\mu_0, \forall \varphi \in \mathcal{H}\}$. A regularized optimal transport problem is:

$$\max_{\mu \in \mathcal{Q}} \int \tilde{c} d\mu - \frac{1}{\gamma} \int \beta^*(\frac{d\mu}{d\theta}) d\theta \tag{5.7}$$

where β^* is the convex conjugate of the function β which assumed to be a differentiable non-decreasing convex function such that $\lim_{x\to\infty} \beta(x)/x = \infty$.

θ can be any measure in $\mathcal{P}(\mathbb{R}^d)$ as long as there exist a $\mu \in \mathcal{Q}$ such that $\mu \ll \theta$. We can choose θ to be the independent coupling of the marginals $\mu_1, \ldots, \mu_d$.

The goal for the numerical calculation is to optimize the regularized dual (sub)super-hedging functional:

$$\inf_{\varphi \in \mathcal{H}} \int \varphi d\mu_0 + \int b_\gamma \left(\tilde{c} - \varphi\right) d\theta \tag{5.8}$$

where $b_\gamma := \frac{1}{\gamma}\beta(\gamma x)$ is a penalty function parameterized by $\gamma > 0$. [25, Theorem 2.2] shows that this penalized version converges to the optimal value of the original, constrained problem. Moreover, it allows the construction of an optimal joint distribution for the primal problem. In fact, assume $\hat{\varphi} \in \mathcal{H}$ is such an optimal function to (5.8), then the measure $\hat{\mu} \in \mathbb{R}^d$ given by

$$\frac{d\hat{\mu}}{d\theta} := b'_\gamma(\tilde{c} - \hat{\varphi}) \tag{5.9}$$

is an optimal measure to the problem (5.7). This construction helps us understanding and visualizing the geometry of the primal problem.

In our setting, we let $\mathcal{Q} = \mathrm{CVMT}(\mathcal{L}_{[0,1]}, \mathcal{L}_{[0,1]^d})$ and $\mathcal{H}$ be the class of functions of the form:

$$\varphi(u, v) = \tilde{\phi}(u) + \sum_{i=1}^{d} \left(\tilde{\psi}_i(v_i) + \tilde{h}_i(u)(G_i(v_i) - F_i^{-1}(u))\right) \tag{5.10}$$

where $(\tilde{\phi}, \tilde{\psi}_i, \tilde{h}_i) \in \tilde{\Psi}$. Notice that, by the martingale condition, for any $\mu_0 \in \mathcal{Q}$, we have $\int \sum_{i=1}^{d} \tilde{h}_i(u)(G_i(v_i) - F_i^{-1}(u)) d\mu_0 = 0$. Therefore, the loss function can be written as:

$$\begin{aligned} \text{Loss} &= \int \varphi d\mu_0 + \int b_\gamma \left(c - \varphi\right) d\theta \\ &= \int \tilde{\phi}(u) du + \sum_{i=1}^{d} \int \tilde{\psi}_i d\nu_i + \int b_\gamma \left(c - \varphi\right) d\theta \end{aligned} \tag{5.11}$$

We fix the penalization function b as

$$b\left(t\right) = \frac{1}{2\gamma}\left(\left(\gamma t\right)^+\right)^2 \qquad \gamma = 1000 \tag{5.12}$$

Computationally, each of θ, ψ_i, h_i is replaced by some approximation θ^m, ψ_i^m, h_i^m implemented as a neural network with an internal size parameterized by m. We chose to use a fixed number of 2 ReLU-network layers with 64 Neurons each. Variations of this arrangement did not bring significant change, and we did not perform a hyper parameter search. The integrals are approximated by the mean over samples drown from the distributions μ_0 and θ – in both cases, we adopt the independent coupling of the marginals $\vec{\mu}$ and $\vec{\nu}$. As a standard procedure, we run the neural network calibration, or "training", for a certain number of epochs until an acceptable level of convergence is reached. We used random samples of 1 million points renewed at every ten epochs, and found a number of 100 epochs to be sufficient for all examples. At each epoch, we store the sample mean of the dual value and the penalty term. We employ Python and the Pytorch neural network package with the standard Adam gradient descent optimizer. [5] The convergence to the true optimal value is guaranteed by Proposition 2.4 and Remark 3.5 in [25].

Examples.BP: I think this example costs and its financial interpretation should be introduced earlier in the paper, and then only the numerical solution presented here. Also, I think the interpretation in terms of maximizing the variance of the sum, as a measure of risk of a portfolio,, for instance, is more natural (I not sure what "pricing the covariance" means). We are especially interested in the cross-product function $c\,(x, y) = \sum_{i \neq j} b_{ij} y_i y_j$, which relates to the problem of robustly pricing the covariance of financial assets. In fact, the covariance of two assets Y_i, Y_j under the measure π is $\mathrm{COV}_\pi(Y_i, Y_j) = \mathbb{E}_\pi[Y_i, Y_j] - \mathbb{E}_\pi[Y_i]\mathbb{E}_\pi[Y_j]$.

It has been shown that the option implied ex-ante higher moment have a predictive power for the future stock returns [19]. Some argue that [39] this observation is due to the fact that option prices are aggregation of market information and investors' expectation to future stock moment.

From an asset pricing point of view, knowing the option-implied marginals of X_i and Y_i, $i = 1, 2$, means knowing the ex-ante distribution of prices of two assets at two distinct future times $t = 1, 2$. When their joint distribution is unknown, VMOT provides an upper bound to the expected covariance at $t = 2$. In fact, we can derive the marginal distributions from market data, as

[5]Source code available at https://github.com/souza-m/vmot.

we do in our second example below. Notice that, even though the prices at $t = 1$ are not included in the cost function, knowledge about the distribution at $t = 1$ provide more information by considering the history of the assets. By considering the filtration, this potentially tightening the bounds for the expected value of the cost function, which is also the subject in [50].

1. Normal marginals. We start with a set of theoretical examples with normal marginals, whose exact solutions are given by proposition 6.1 in the appendix. It is known that our result on the monotonicity support extends to general dimension in specific cases such as when the marginals are normally distributed – see for instance Example 5.2 and Theorem 5.3 of [24]. Motivated by that, and to further illustrate the positive effects of dimension reduction, besides the case $d = 2$ we add examples with $d = 3$, 4 and 5. Notice that higher dimension cases give opportunity to a greater relative dimensional reduction to the problem. In fact, using the dimension of the sample domain as a measure of computational complexity, this one is proportional to $d + 1$ in the reduced dimension and to $2d$ in the full dimension case.

To introduce some variability, we use randomization to define the cost functions and marginal dispersion parameters. Each coefficient b_{ij} is a randomly generated number in $[0, 1]$ rounded to the second decimal. The marginals are defined as

$$X_i \sim N\left(0, \sigma_i^2\right)$$
$$Y_i \sim N\left(0, \rho_i^2\right)$$

where each σ_i is a random number between 1 and 2 and each ρ_i is a random number between 2 and 3, all of them rounded to the second decimal. This guarantees that the marginals are in convex order for each i, as required. The resulting coefficients and parameters are available in the referenced code.

Figure 1 shows the convergence of the dual value over the training epochs in the two formulations, namely, with our simplified, reduced-dimension version 5.3 shown in blue and the full-dimension version 2.3 shown in orange, for each value of d. The true value is shown as a dotted line for reference. It is noteworthy that accuracy is significantly sensitive to the dimension of the sample domain. As higher values of d are used, we see a degradation in the convergence pattern, especially in the full dimension version of the problem. The latter also uses more memory and demands slightly more time to process the same number of epochs. A greater accuracy is observed in the reduced dimension version in all cases. (Is there a way to make Overleaf locate the figures where we want them to be?)

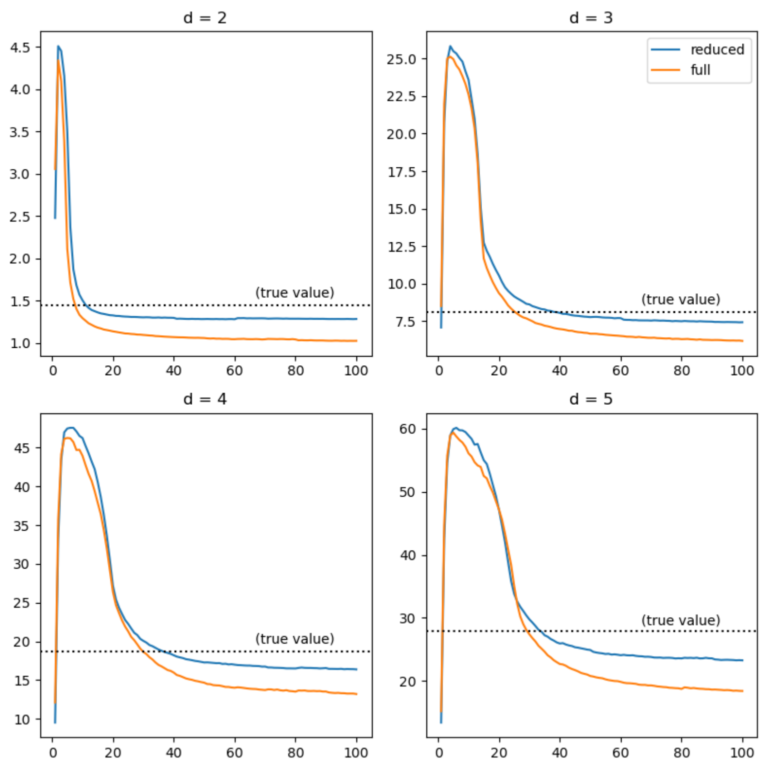


FIGURE 1. Convergence of the dual value – normal marginals, $d = 2$ to 5.

The following table compares the true and the mean numeric values for all cases. For each version of the problem, we ran the method over 10 samples of 1 million points each and registered the global mean, the standard deviation across the 10 samples and the mean penalty term. (Less important columns may be excluded.)

Method	d	True value	Numerical value	Numerical STD	Penalty
reduced	2	1.45	1.26	0.0027	0.1007
	3	8.13	7.41	0.0077	0.5082
	4	18.72	16.54	0.0160	1.0160
	5	27.92	23.61	0.0112	1.5909
full	2	1.45	1.02	0.0020	0.1632
	3	8.13	6.00	0.0086	0.9898
	4	18.72	13.20	0.0071	1.7106
	5	27.92	18.53	0.0162	2.1185

Figure 2 shows the mass of the underlying coupling $(\pi^X)^\star$ implied by φ^n in both versions for $d = 2$ according to Theorem 2.2 of [25]. The left side shows the full dimension case, where we observe a convergence of the mass of π^1 towards the main diagonal, consistent with our result. The right side shows the reduced dimension case, where the support of $(\pi^X)^\star$ is restricted to the diagonal.

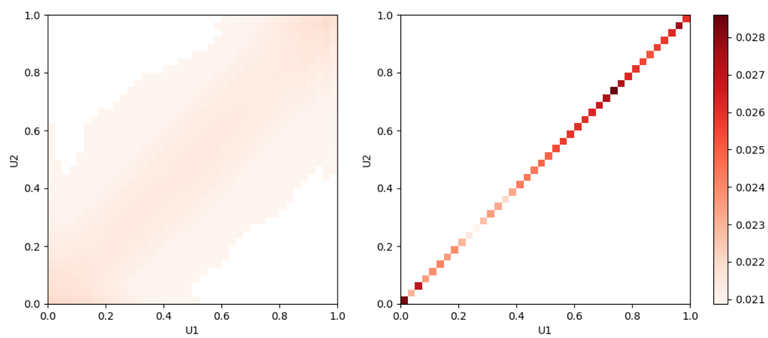


FIGURE 2. Heat map of $(\pi^X)^\star$, full (left) and reduced dimension (right).

2. Empirical marginals. We move on to a real world problem where we calculate upper and lower bounds p^+ and p^- (discuss) an upper bound for the price of a hypothetical contract that pays the product of the prices of two assets – Apple and Amazon – at some future date. Our model is consistent with the fact that the distribution of each price at each date is given and implied by the market, while the joint distribution is unknown. We construct implicit distributions for the prices at times Jan. 20th and Feb. 17th, 2023 based on call and put option prices as of Dec. 16th, 2022, shown in Figure 3. It

BP: Again, think consid ering varianc of a portfoli is more natura than the price o a hypothetica contract.

is well known that the risk-neutral density function is the second derivative of the option price curve with respect to strike price. Since we can only observe call prices at a finite number of strike prices, we will use second-order finite differences method to approximate the second derivative. Details about the calculation are provided in the Appendix.

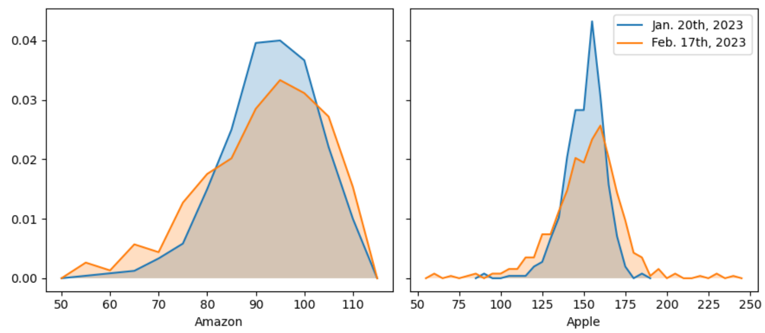


FIGURE 3. Marginal distributions of future prices as of Dec. 16th, 2022.

We maximize the cost function $c(x, y) = y_1 y_2$. We use the inverse cumulative of the empirical distributions as an input to our method. Figure 4 shows the convergence of the dual value in both cases as in previous examples. Since the optimal cost is unknown, we plot the (independent coupling) sample mean cost as a reference, observing that this value must be a lower bound reference to the true value rather than a central target. We observe a slightly better convergence speed for the reduced-dimension version and virtually identical accuracy results, as measured by the standard deviation and penalty over sets of 10 samples, as in the previous example.

Alternative: upper and lower bounds – Figure 5.

6. APPENDIX

Proof of Proposition 3.4. Let us prove part i) first. If $\beta \equiv \infty$, then $\beta^* \equiv -\infty$ and there is nothing to prove. And if $\beta \not\equiv \infty$ but β^* is not proper, i.e., $\beta^* \equiv \infty$, again there is nothing to prove. So we assume β and β^* are proper.

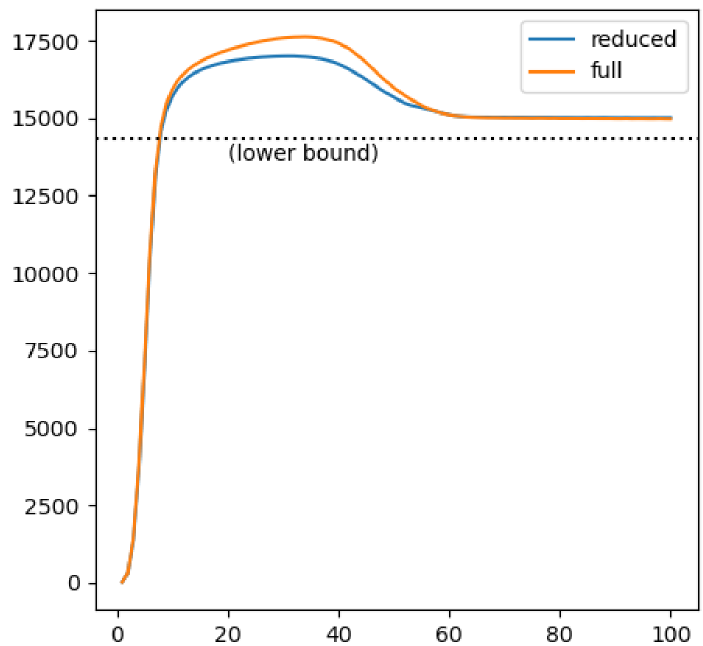


FIGURE 4. Convergence of the dual value – empirical marginals, $d = 2$.

Suppose β is submodular. For $R \geq 0$, define

$$\beta_R(x) = \begin{cases} \beta(x) & \text{if } x \in [-R, R]^d \\ +\infty & \text{otherwise.} \end{cases} \tag{6.1}$$

Notice β_R is submodular. And for large enough R, β_R is proper. Now β_R being compactly supported implies that β_R^* will be Lipschitz unless $\beta_R^* \equiv +\infty$, but the latter is excluded since $\beta^* \geq \beta_R^*$. In particular, β_R^* is real-valued everywhere. 3.4 We will first show the supermodularity of β_R^*. Assume $y_i, \bar{y}_i \in \mathbb{R}$ and $y_i \leq \bar{y}_i$ for all $i = 1, ..., d$. Let $\hat{y}_i$ be any number between $y_i, \bar{y}_i$ and $\hat{y}_i^c = \{y_i, \bar{y}_i\} \setminus \{\hat{y}_i\}$ be the other number. We denote $y = (y_1, ..., y_d)$, $\hat{y} = (\hat{y}_1, ..., \hat{y}_d)$, $\hat{y} = (\hat{y}_1, ..., \hat{y}_d)$, $\hat{y}^c = (\hat{y}_1^c, ..., \hat{y}_d^c)$ by the elements in $\mathbb{R}^d$. We need to prove:

$$\beta_R^*(y) + \beta_R^*(\bar{y}) \geq \beta_R^*(\hat{y}) + \beta_R^*(\hat{y}^c). \tag{6.2}$$

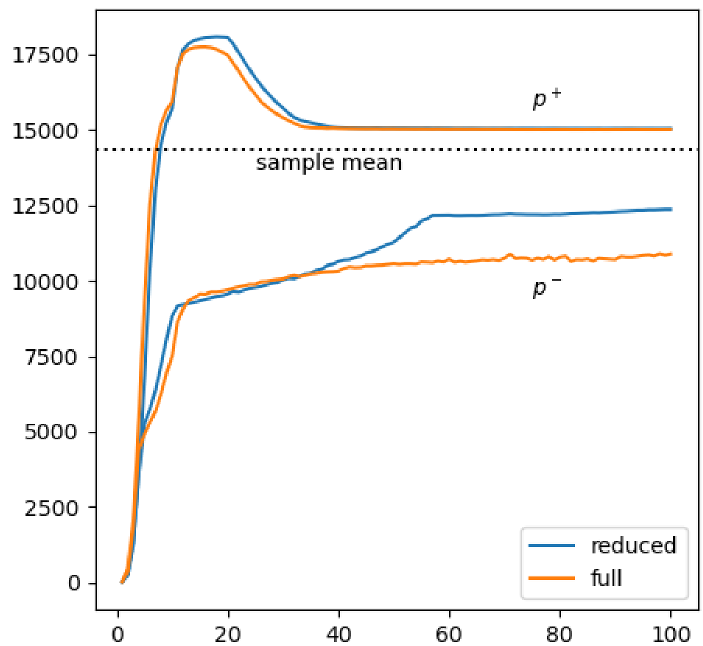


FIGURE 5. Convergence of the dual value – empirical marginals, $d = 2$.

By definition of Legendre transform, given $\epsilon > 0$, there exists $x, \bar{x} \in \mathbb{R}^d$ such that

$$\beta_R^*(\hat{y}) < x \cdot \hat{y} - \beta_R(x) + \epsilon, \quad \beta_R^*(\hat{y}^c) < \bar{x} \cdot \hat{y}^c - \beta_R(\bar{x}) + \epsilon. \tag{6.3}$$

Then we deduce the following, where the first inequality is by definition of Legendre transform, and the second is by submodularity of β_R:

$$\begin{aligned}
&\beta_R^*(y) + \beta_R^*(\bar{y}) \\
&\geq (x \wedge \bar{x}) \cdot y - \beta_R(x \wedge \bar{x}) + (x \vee \bar{x}) \cdot \bar{y} - \beta_R(x \vee \bar{x}) \\
&\geq (x \wedge \bar{x}) \cdot y + (x \vee \bar{x}) \cdot \bar{y} - \beta_R(x) - \beta_R(\bar{x}) \\
&= (x \wedge \bar{x}) \cdot (y - \hat{y}) + \hat{y} \cdot (x \wedge \bar{x} - x) + x \cdot \hat{y} \qquad (6.4) \\
&\quad + (x \vee \bar{x}) \cdot (\bar{y} - \hat{y}^c) + \hat{y}^c \cdot (x \vee \bar{x} - \bar{x}) + \bar{x} \cdot \hat{y}^c - \beta_R(x) - \beta_R(\bar{x}).
\end{aligned}$$

We claim that for each i, we have

$$\begin{aligned}
&\min(x_i,\bar{x}_i)(y_i-\hat{y}_i)+\hat{y}_i(\min(x_i,\bar{x}_i)-x_i)\\
&+\max(x_i,\bar{x}_i)(\bar{y}_i-\hat{y}_i^c)+\hat{y}_i^c(\max(x_i,\bar{x}_i)-\bar{x}_i)\geq 0.
\end{aligned}\tag{6.5}$$

To see this, we investigate the following four possible cases: i) $x_i=\min(x_i,\bar{x}_i)$, $y_i=\hat{y}_i$, ii) $\bar{x}_i=\min(x_i,\bar{x}_i)$, $y_i=\hat{y}_i$, iii) $x_i=\min(x_i,\bar{x}_i)$, $\bar{y}_i=\hat{y}_i$, iv) $\bar{x}_i=\min(x_i,\bar{x}_i)$, $\bar{y}_i=\hat{y}_i$.

i) $x_i=\min(x_i,\bar{x}_i)$, $y_i=\hat{y}_i$:

$$\begin{aligned}
&\min(x_i,\bar{x}_i)(y_i-\hat{y}_i)+\hat{y}_i(\min(x_i,\bar{x}_i)-x_i)\\
&+\max(x_i,\bar{x}_i)(\bar{y}_i-\hat{y}_i^c)+\hat{y}_i^c(\max(x_i,\bar{x}_i))-\bar{x}_i)\\
=\;&x_i(y_i-y_i)+y_i(x_i-x_i)+\bar{x}_i(\bar{y}_i-\bar{y}_i)+\bar{y}_i(\bar{x}_i-\bar{x}_i)=0,
\end{aligned}$$

ii) $\bar{x}_i=\min(x_i,\bar{x}_i)$, $y_i=\hat{y}_i$:

$$\begin{aligned}
&\min(x_i,\bar{x}_i)(y_i-\hat{y}_i)+\hat{y}_i(\min(x_i,\bar{x}_i)-x_i)\\
&+\max(x_i,\bar{x}_i)(\bar{y}_i-\hat{y}_i^c)+\hat{y}_i^c(\max(x_i,\bar{x}_i))-\bar{x}_i)\\
=\;&\bar{x}_i(y_i-y_i)+y_i(\bar{x}_i-x_i)+x_i(\bar{y}_i-\bar{y}_i)+\bar{y}_i(x_i-\bar{x}_i)\\
=\;&\underbrace{(y_i-\bar{y}_i)}_{\leq 0}\underbrace{(\bar{x}_i-x_i)}_{\leq 0}\geq 0,
\end{aligned}$$

iii) $x_i=\min(x_i,\bar{x}_i)$, $\bar{y}_i=\hat{y}_i$:

$$\begin{aligned}
&\min(x_i,\bar{x}_i)(y_i-\hat{y}_i)+\hat{y}_i(\min(x_i,\bar{x}_i)-x_i)\\
&+\max(x_i,\bar{x}_i)(\bar{y}_i-\hat{y}_i^c)+\hat{y}_i^c(\max(x_i,\bar{x}_i))-\bar{x}_i)\\
=\;&x_i(y_i-\bar{y}_i)+\bar{y}_i(x_i-x_i)+\bar{x}_i(\bar{y}_i-y_i)+y_i(\bar{x}_i-\bar{x}_i)\\
=\;&\underbrace{(x_i-\bar{x}_i)}_{\leq 0}\underbrace{(y_i-\bar{y}_i)}_{\leq 0}\geq 0,
\end{aligned}$$

iv) $\bar{x}_i=\min(x_i,\bar{x}_i)$, $\bar{y}_i=\hat{y}_i$:

$$\begin{aligned}
&\min(x_i,\bar{x}_i)(y_i-\hat{y}_i)+\hat{y}_i(\min(x_i,\bar{x}_i)-x_i)\\
&+\max(x_i,\bar{x}_i)(\bar{y}_i-\hat{y}_i^c)+\hat{y}_i^c(\max(x_i,\bar{x}_i))-\bar{x}_i)\\
=\;&\bar{x}_i(y_i-\bar{y}_i)+\bar{y}_i(\bar{x}_i-x_i)+x_i(\bar{y}_i-y_i)+y_i(x_i-\bar{x}_i)=0.
\end{aligned}$$

We conclude that (6.5) holds. Combining (6.5) with (6.4) and (6.3), we have

$$\begin{aligned}
&\beta_R^*(y) + \beta_R^*(\bar{y}) \\
&\geq (x \wedge \bar{x}) \cdot (y - \hat{y}) + \hat{y} \cdot (x \wedge \bar{x} - x) + x \cdot \hat{y} \\
&\quad + (x \vee \bar{x}) \cdot (\bar{y} - \hat{y}^c) + \hat{y}^c \cdot (x \vee \bar{x} - \bar{x}) + \bar{x} \cdot \hat{y}^c - \beta_R(x) - \beta_R(\bar{x}) \\
&\geq x \cdot \hat{y} - \beta_R(x) + \bar{x} \cdot \hat{y}^c - \beta_R(\bar{x}) \\
&\geq \beta_R^*(\hat{y}) + \beta_R^*(\hat{y}^c) - 2\epsilon.
\end{aligned}$$

Taking $\epsilon \to 0$ yields the desired supermodularity of β_R^*. Now as $R \to \infty$, we have $\beta_R \searrow \beta$ pointwise on $\mathbb{R}^d$ thus $\beta_R^* \nearrow \beta^*$, hence obtaining supermodularity of β^*.

Now we prove part ii). For $d = 2$, if $\beta(x_1, x_2)$ is supermodular, then $\tilde{\beta}(x_1, x_2) \coloneqq \beta(x_1, -x_2)$ is submodular. Hence $\tilde{\beta}^*$ is supermodular by part i), yielding $(y_1, y_2) \mapsto \tilde{\beta}^*(y_1, -y_2)$ is submodular. We then compute

$$\begin{aligned}
\tilde{\beta}^*(y_1, -y_2) &= \sup_{x_1, x_2} x_1 y_1 + x_2(-y_2) - \tilde{\beta}(x_1, x_2) \\
&= \sup_{x_1, x_2} x_1 y_1 + (-x_2)(-y_2) - \tilde{\beta}(x_1, -x_2) \\
&= \sup_{x_1, x_2} x_1 y_1 + x_2 y_2 - \beta(x_1, x_2) \\
&= \beta^*(y_1, y_2)
\end{aligned}$$

which shows that $\beta^*(y_1, y_2) = \tilde{\beta}^*(y_1, -y_2)$ and the result follows. □

Proposition 6.1. *Let $\mu_i \sim N(0, \sigma_i^2)$, $\nu_i \sim N(0, \rho_i^2)$, $0 < \sigma_i < \rho_i$, $i = 1, ..., d$. Let $\lambda_i = \sqrt{\rho_i^2 - \sigma_i^2}$. Let $c(x, y) = \sum_{i<j} a_{ij} x_i x_j + b_{ij} y_i y_j$, $a_{ij} \geq 0$, $b_{ij} \geq 0$. Then*

$$\max_{\pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})} \mathbb{E}_\pi[c(X, Y)] = \sum_{1 \leq i < j \leq d} \big((a_{ij} + b_{ij})\sigma_i \sigma_j + b_{ij} \lambda_i \lambda_j\big). \tag{6.6}$$

Construction of the empirical measure. It is well known that if we had an infinite number of call option prices $C(K, t)$ or put option prices $P(K, t)$ with t time to maturity, across all possible strike prices K, we could determine the risk-neutral density function. As noted by Breeden and Litzenberger

[11], the risk-neutral density $f(K,t)$, for period t is essentially the second derivative of the curve of call or put prices with respect to strike price.

$$f(K,t) = \frac{\partial^2 C(X,t)}{\partial X^2}\Big|_{X=K} = \lim_{h\to 0} \frac{[C(X+h,t) - C(X,t)] - [C(X,t) - C(X-h,t)]}{h^2} \tag{6.7}$$

This formular is also true when we replace the curve of call price C by curve of put price P. Importantly, because of the no arbitrage condition, both call and put price curves are convex as a function of K, meaning the second derivative almost surely exists.

However, in practice we only observe call prices at a finite number of strike prices, limiting our ability to directly calculate the risk-neutral density. Instead, we approximate it using second-order finite differences at observable strike prices and linearly interpolate between different strike prices. We note that out-of-the-money options tend to have lower volume and may be mispriced, so we calculate the density function using put options for strike prices below the spot price and call options for strike prices above the spot price.

Additionally, near the extremes of the call and put price curves where prices approach zero, we exclude some option prices that obviously violate no arbitrage due to illiquidity. We may also normalize the resulting function to ensure it is a valid probability density function (i.e. its integral equals one), if needed. Through these steps, we obtain an empirical risk-neutral density function.

References

[1] Aurélien Alfonsi, Jacopo Corbetta, Benjamin Jourdain, *Sampling of probability measures in the convex order by Wasserstein projection*, volume 56 (3) of *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*. 1706-1729, 2020.

[2] Jonathan Ansari, Eva Lütkebohmert, Ariel Neufeld, Julian Sester, *Improved Robust Price Bounds for Multi-asset Derivatives under Market-implied Dependence Information*, Improved Robust Price Bounds for Multi-asset Derivatives under Market-implied Dependence Information *arXiv:2204.01071*. 2022.

[3] Gurdip Bakshi, Nikunj Kapadia, Dilip Madan, *Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options*, volume 16 (1) of *The Review of Financial Studies.* 101-143, 2003.

[4] Francis Bach, *Submodular functions: from discrete to continuous domains*, volume 175 of *Mathematical Programming.* 419-459, 2019.

[5] Nicole Bäuerle, Daniel Schmithals, *Martingale optimal transport in the discrete case via simple linear programming techniques*, volume 90 of *Mathematical Methods of Operations Research.* 453–476, 2019.

[6] Bank for International Settlements, Basel *Revisions to the Basel II market risk framework*, *Technical report.* 2009.

[7] J. Backhoff-Veraguas, G. Pammer, *Stability of martingale optimal transport and weak optimal transport*, volume 32(1) of *The Annals of Applied Probability.* 721-752, 2022.

[8] M. Beiglböck, Pierre Henry-Labordère and Friedrich Penkner Model-independent bounds for option prices—a mass transport approach *Finance and Stochastics*, 17, 477–501, 2013

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

[10] M. Beiglböck, T. Lim and J. Obłój. Dual attainment for the martingale transport problem. *Bernoulli.* 25 (2019), no. 3, 1640–1658.

[11] Douglas T. Breeden and Robert H. Litzenberger. Prices of State-Contingent Claims Implicit in Option Prices *The Journal of Business*, 51(4) (1978) 621-651.

[12] M. Beiglböck, M. Nutz and N. Touzi. Complete duality for martingale optimal transport on the line. *Ann. Probab.*, 45(5) (2017) 3038–3074.

[13] Yann Brenier, *Décomposition polaire et réarrangement monotone des champs de vecteurs*, volume 305 of *C. R. Acad. Sci. Paris Sér. I Math..* 805–808, 1987.

[14] Martin Bruckerhoff, Nicolas Juillet, *Instability of martingale optimal transport in dimension* $d \leq 2$, volume 27 of *Electronic Communications in Probability.* 1-10, 2022.

[15] Luis Caffarelli, *Allocation maps with general cost functions*, in *Partial differential equations and applications* 29–35, 1996.

[16] Guglielmo Maria Caporale, Luis Gil-Alana, Alex Plastun, *Is market fear persistent? A long-memory analysis*, volume 27 of *Finance Research Letters*. 140-147, 2018.

[17] Guillaume Carlier *On a class of multidimensional optimal transportation problems* volume 10 of *Journal of Convex Analysis* 517–529, 2003.

[18] Guillaume Carlier, Alfred Galichon, Filippo Santambrogio, *From Knothe's Transport to Brenier's Map and a Continuation Method for Optimal Transport*, volume 41 (6) of *The Journal of Finance*. 2554-2576, 2010.

[19] Jennifer Conard, Robert F. Dittmar, Eric Ghysels, *Ex Ante Skewness and Expected Stock Returns*, volume 68 (1) of *The Journal of Finance*. 85–124, 2013.

[20] Giovanni Cespa, Antonio Gargano, Steven J Riddiough, Lucio Sarno, *Foreign Exchange Volume*, volume 35 (5) of *The Review of Financial Studies*. 2386–2427, 2022.

[21] Marco Cuturi *'Sinkhorn distances: Lightspeed computation of optimal transport*, In: *In: Advances in Neural Information Processing Systems*. 26, 2292–2300, 2013.

[22] Hadrien De March, *Entropic approximation for multi-dimensional martingale optimal transpor*, 2018 arXiv preprint, `https://arxiv.org/pdf/1812.11104.pdf`

[23] Yan Dolinsky, H. Mete Soner, *Martingale optimal transport and robust hedging in continuous time*, volume 160 of *Probability Theory and Related Fields*. 391–427, 2014.

[24] Stephan Eckstein, Gaoyue Guo, Tongseok Lim, Jan Obłój, *Robust Pricing and Hedging of Options on Multiple Assets and Its Numerics*, volume 12 (1) of *SIAM Journal on Financial Mathematics*. 158-188, 2021.

[25] Stephan Eckstein, Michael Kupper, Computation of Optimal Transport and Related Hedging Problems via Penalization and Neural Networks *Applied Mathematics and Optimization*, 2021

[26] A. Galichon, P. Henry-Labordère, and N. Touzi. A Stochastic Control Approach to No-Arbitrage Bounds Given Marginals, with an Application to Lookback Options. *Annals of Applied Probability.*, Volume 24, Number 1 (2014) 312–336.

[27] Wilfrid Gangbo, Robert J. McCann, The geometry of optimal transportation, *Acta Mathematica*, 177(2): 113–161, 1996

[28] Wilfrid Gangbo, Andrzej Świech,Optimal maps for the multidimensional Monge-Kantorovich problem, *Communications on Pure and Applied Mathematics*, 1998

[29] Aude Genevay, Gabriel Peyré, Marco Cuturi, Learning Generative Models with Sinkhorn Divergences, 2017 arXiv preprint, `https://arxiv.org/pdf/1706.00292.pdf`

[30] Nassif Ghoussoub, Young-Heon Kim, Tongseok Lim, Structure of optimal martingale transport plans in general dimensions *The Annals of Probability*, 47(1),109-164, 2019

[31] Gaoyue Guo, Jan Obłój *Computational methods for martingale optimal transport problems.*, volume 29(6) of *The Annals of Applied Probability.* 3311-3347, 2019.

[32] Christopher Harshaw, Ehsan Kazemi, Moran Feldman, Amin Karbasi The Power of Subsampling in Submodular Maximization *Mathematics of Operations Research*, 47(2) (2021) 1365-1393.

[33] Henri Heinich, *Problème de Monge pour n probabilités* volume 334 of *C. R. Math. Acad. Sci. Paris.* 793–795, 2002.

[34] Pierre Henry-Labordère, *Model-free Hedging: A Martingale Optimal Transport Viewpoint*, Chapman and Hall/CRC. 2017.

[35] Pierre Henry-Labordère, Xiaolu Tan, Nizar Touzi, *An explicit martingale version of the one-dimensional Brenier's Theorem with full marginals constraint*, volume 126 (9) of *Stochastic Processes and their Applications.* 2800-2834, 2016.

[36] Pierre Henry-Labordère, *(Martingale) Optimal Transport And Anomaly Detection With Neural Networks: A Primal-dual Algorithm*, 2019 arXiv preprint, `https://arxiv.org/abs/1904.04546`

[37] David G. Hobson Robust hedging of the lookback option *Finance and Stochastics*, 2(4), 329–347, 1998

[38] David G. Hobson The Skorokhod embedding problem and model-independent bounds for option prices. In: *Carmona, R.A., Çinlar, E., Ekeland, I., Jouini, E., Scheinkman, J.A., Touzi, N. (eds.) Paris–Princeton Lectures on Mathematical Finance 2010.*, Lecture Notes

in Mathematics, vol. 2003, 267–318, 2011

[39] Jens Carsten Jackwerth, Mark Rubinstein, *Recovering probability distributions from option*, volume 51 of *The Journal of Finance*. 1611-1631, 1996.

[40] Young-Heon Kim, Brendan Pass *A general condition for Monge solutions in the multi-marginal optimal transport problem* volume 46 of *SIAM J. Math. Anal.* 1538–1550, 2014.

[41] Jun Kitagawa, Quentin Merigot, Boris Thibert, *Convergence of a Newton algorithm for semi-discrete optimal transport*, volume 21 (9) of *Journal of the European Mathematical Society*. 2603–2651, 2019.

[42] Alexander Korotin, Daniil Selikhanovych, Evgeny Burnaev, *Neural Optimal Transport*, 2022 arXiv preprint, `https://arxiv.org/abs/2201.12220`

[43] Frank H. Knight, *Risk, Uncertainty and Profit*, Houghton Mifflin, Boston

[44] Thomas Maurer, Thuy-Duong Tô, Ngoc-Khanh Tran, *Pricing Implications of Covariances and Spreads in Currency Markets*, volume 12 (1) of *The Review of Asset Pricing Studies*. 336–388, 2022.

[45] Quentin Merigot, *A multiscale approach to optimal transport*, volume 30(5) of *Computer Graphics Forum*. .1584-1592, 2011.

[46] Christian Leonard, From the Schrödinger problem to the Monge–Kantorovich problem *The Journal of Functional Analysis*, 262(4), 1879-1920, 2012

[47] Vladimir Levin, *Abstract cyclical monotonicity and Monge solutions for the general Monge-Kantorovich problem*, volume 7 of *Set-Valued Anal.* 7–32, 1999.

[48] Tongseok Lim, *Geometry of vectorial martingale optimal transportations and duality*, *Mathematical Programming series A*, to appear. arXiv preprint, `https://arxiv.org/pdf/1611.01496.pdf`

[49] G.G Lorentz, *An inequality for rearrangements.*, volume 60 of *Amer. Math. Monthly*. 176–179, 1953.

[50] Eva Lütkebohmert and Julian Sester, *Tightening robust price bounds for exotic derivatives*, volume 19(11) of *Quantitative Finance*. 1797-1815, 2019.

[51] Fatemeh Navidi, Prabhanjan Kambadur, Viswanath Nagarajan, *Adaptive Submodular Ranking and Routing*, volume 68(3) of *Operations Research*. 856-877, 2020.

[52] Jan Obłój, Siorpaes, Pietro *Structure of martingale transports in finite dimensions*, 2017 arXiv preprint, https://arxiv.org/pdf/1702.08433.pdf

[53] Brendan Pass, Adolfo Vargas-Jimenez *Monge solutions and uniqueness in multi-marginal optimal transport via graph theory*, to appear in *Adv. Math.*

[54] Brendan Pass *Uniqueness and Monge solutions in the multimarginal optimal transportation problem*, volume 43 of *SIAM J. Math. Anal.* 2758–2775, 2011.

[55] Filippo Santambrogio, *Optimal Transport for Applied Mathematicians. Progress in Nonlinear Differential Equations and Their Applications*, vol 87. Birkhäuser, Cham. 2015

[56] Ghulam Sarwar *Is VIX an investor fear gauge in BRIC equity markets?*, volume 22 (3) of *Journal of Multinational Financial Management*. 55-65, 2012.

[57] Vivien Seguy, Bharath Bhushan Damodaran, Rémi Flamary, Nicolas Courty, Antoine Rolet, Mathieu Blondel, *Large-Scale Optimal Transport and Mapping Estimation*, 2017 arXiv preprint, https://arxiv.org/abs/1711.02283

[58] Jarno Talponen, Lauri Viitasaari, Note on multidimensional Breeden–Litzenberger representation for state price densities *The Mathematics and Financial Economics*, volume 8,153–157, 2014

[59] Cédric Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 2003.

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

Tongseok Lim: Mitchell E. Daniels, Jr. School of Business
Purdue University, West Lafayette, Indiana 47907, USA
Email address: lim336@purdue.edu