# 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

In mathematical finance, Knightian uncertainty [30] refers to financial risk resulting from mis-specification or uncertainty about the true model of the physical world. People became more concerned about such risks following the 2007-2008 financial crisis, according to [35]. One approach to addressing this issue is model independent finance, which was initially discussed in [27], which offered a novel way of pricing financial derivatives based on the concept of model-independent arbitrage.

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

Problems are well researched when the payoff function is dependent on the values of two (or more) assets at a single future time, with known individual distributions but uncertain dependence structure. In this scenario, the model independent pricing problem is equivalent to the classical problem known as optimal transport in the mathematics literature. The variant occurring

*Date*: July 17, 2023.

where the payout is dependent on the value of a single asset at two (or more) future times has received a lot of attention in recent years. In this situation, the absence of arbitrage compels the unknown coupling between the known distributions to be a martingale, and the ensuing optimization problem with 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, individual asset price distributions can typically be estimated at many different times, and this information is not incorporated in the standard optimal transport problem.[1] Our attention is drawn here to a situation in which the distributions of multiple individual asset prices are known at two future dates, but nothing about the dependence structure is known (either between distinct assets or between different assets at different times). As options on individual stocks with a specific maturity and a wide range of strike prices are often frequently 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 (JH: refence needed, to be added). The optimization problem we investigate here, known in the literature as vectorial martingale optimal transport (or multi-marginal martingale optimal transport), yields upper and lower bounds on the arbitrage free prices of contracts depending on several assets at two future times.

A conjecture formulated in [15, 33] is that, for a certain class of payoff functions, the maximium arbitrage free price arises when the coupling of the assets at the first maturity time is perfectly co-monotone. If true, not only does the conjecture provide insight into the extremal dependence structure of asset prices (specifically, the conjecture asserts the existence of a single factor market model leading to the maximum price), but it also significantly reduces the complexity of computing the overall solution; because the dependence

[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.

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

In this paper, we address this conjecture by demonstrating that it is true for derivatives depending on two assets and providing a counterexample for derivatives depending on three or more assets. We then exploit the monotone structure in the case of two assets to refine a numerical method developed in [16], resulting in faster and more accurate computations.

The payoff functions studied in this paper cover a wide range of contracts that naturally arise in applications. These include the model independent pricing of European calls and puts on a basket, as well as the maximums and minimums of numerous assets, when the distributions of the individual asset prices are known both at maturity and earlier.

This work is organized as follows: in the following section, we present and review the vectorial martingale optimal transport problem. Section 3 examines the formulation and resolution of the monotonicity conjecture, while Section 4 presents our numerical method and results.

## 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 $\Omega$. 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 $\mu_i, \nu_i$, $i \in [d]$. We consider the following space of *Vectorial Martingale Transportations* (VMT) from $\vec{\mu}$ to $\vec{\nu}$ (see [33]):

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

where $X = (X_1, ..., X_d)$, $Y = (Y_1, ..., Y_d) \in \mathbb{R}^d$ are random vectors. For a measure $\pi \in \mathcal{P}(\mathbb{R}^{2d})$, we 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 $\pi = \text{Law}(X, Y)$, then $\pi^X = \text{Law}(X)$ and $\pi^Y = \text{Law}(Y)$. We will denote by $\Pi(\vec{\mu})$ the set of couplings of the $\mu_i$, that is,

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

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 $\mu_i, \nu_i$ 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 $\pi$ to (2.3) will be called a vectorial martingale optimal transport, or VMOT.

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 by assuming zero interest rate, 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 an $\mathbb{R}^d$-valued (one-period) martingale. We call $\pi$ a vectorial martingale transport, or VMT, if its one-dimensional marginals $\vec{\mu}, \vec{\nu}$ are given, which condition is inspired by [2], [14], [17], [25], [27], [12]. [6] 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 $\pi$ 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 arbitrage-free price bound for the option $c$ 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 $|c(x,y)| \le \sum_{i=1}^d \big(v_i(x_i) + w_i(y_i)\big)$. Note that this ensures, for any $\pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$,

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

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

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

$$c(x,y) = \sum_{1 \le i,j \le d} (a_{ij} x_i x_j + b_{ij} x_i y_j + c_{ij} y_i y_j). \tag{2.4}$$

Note that for any $\pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$, we have $\mathbb{E}_\pi[X_i Y_j] = \mathbb{E}_\pi[\mathbb{E}_\pi[X_i Y_j | X]] = \mathbb{E}_\pi[X_i \mathbb{E}_\pi[Y_j|X]] = \mathbb{E}_\pi[X_i X_j]$ 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

$$c(x,y) = \sum_{1 \le i < j \le d} (a_{ij} x_i x_j + b_{ij} y_i y_j). \tag{2.5}$$

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

$$c(x,y) = a x_1 x_2 + b y_1 y_2 \tag{2.6}$$

so that $\mathbb{E}_\pi[c] = a\mathbb{E}_\pi[X_1 X_2] + b\mathbb{E}_\pi[Y_1 Y_2]$ represents a weighted sum of mutual covariances between $X_1, X_2$ and between $Y_1, Y_2$ under the market model $\pi$. We note that this cost function is part of a larger class of financial derivative known as a cap, the holder of which receives a payoff at each of the predetermined periods $t = 1, ..., N$:

$$c(X_1, ..., X_N) = \sum_{t=1}^N c_t(X_{t,1}, ..., X_{t,d}), \tag{2.7}$$

where $X_t = (X_{t,1}, ..., X_{t,d})$ represents the vector of asset prices at time $t$. In this case, our prior notation $(X,Y)$ corresponds to $X = X_1$ and $Y = X_2$. We will further discuss more exotic options in Section 3.

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

$$\inf_{(\phi_i,\psi_i,h_i)\in\Xi} \sum_{i=1}^{d} \Big( \int \phi_i d\mu_i + \int \psi_i d\nu_i \Big) \tag{2.8}$$

where $\Xi$ 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

$$\sum_{i=1}^{d} \big(\phi_i(x_i) + \psi_i(y_i) + h_i(x)(y_i - x_i)\big) \geq c(x,y) \quad \forall (x,y) \in \mathbb{R}^{2d} \tag{2.9}$$

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 also has an interpretation in financial terms. Assume that a financial institution is bound to pay $c(X,Y)$ at the terminal maturity $t_2$ to an option holder who holds an option $c$. To hedge its risk, the institution may consider purchasing European options $\phi_i, \psi_j$, 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 $i^{\text{th}}$ 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 (2.9) imposes that the position must superhedge the liability for all possible realizations of the assets price $(X,Y)$, and the left hand side of (2.9) yields the overall payoff of the hedging portfolio $(\phi_i, \psi_i, h_i)_{i=1}^d$. On the other hand, if the problem (2.3) is a minimization problem, then the dual problem is an optimal subhedging problem for which the inequality in (2.9) is reversed.

The value (2.8) 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 (2.8) cannot be solved within the class $\Xi$ in general even when $d = 1$, i.e., the option $c$ depends on a single asset (see [3], [5]), unless some suitable regularity assumption is made on the payoff function $c$ [4]. As a result, a generalized notion of *dual attainment*, i.e., solvability of the dual problem, was introduced in [5] for the case $d = 1$ and then in [33] for $d \geq 2$. More specifically, [33] presents the following dual attainment result.

**Theorem 2.1** ([33])**.** *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* (2.9) *tightly in the following pathwise manner:*

$$\sum_{i=1}^{d} \big(\phi_i(x_i) + \psi_i(y_i) + h_i(x)(y_i - x_i)\big) = c(x, y) \quad \pi - a.s. \tag{2.10}$$

*for every* VMOT $\pi$ *which solves the primal problem* (2.3)*.*

The term *pathwise* denotes that (2.9) and (2.10) hold in a pathwise manner, that is, the equality (2.10) is satisfied for $\pi$ - almost every "price path" $(x, y)$, and that we do not impose an integrability condition on the dual minimizer. $(\phi_i, \psi_i, h_i)$ are only measurable functions that are almost surely real-valued with regard to given marginals.

We note that 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 it happens that the pair is not irreducible, one can perturb it in an arbitrarily small way to make the perturbed pair irreducible. We refer to [5] for more details about irreducibility.[2]

While readers can move on to the next section without interruption of understanding, we shall conclude this section with a brief discussion of two other scenarios in finance where the VMOT approach is potentially useful.

### 2.1. Detecting Mispricing in Foreign Exchange Cross-Rates and Risk Management in Global Supply Chains.

Not all currency pairs in the foreign exchange market are actively traded, resulting in cross-rates that may not accurately reflect market information and could be mispriced. For instance, consider the US dollar price of one Japanese Yen $(X_1)$ and the

[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 $\mu$, i.e., $\mu(I) = \mu(\mathbb{R})$.

US dollar price of one New Zealand Dollar ($X_2$). While $X_1$ and $X_2$ are actively traded, the volume of exchange between the Japanese Yen and the New Zealand Dollar is relatively low [10]. Assuming no transaction costs, we can calculate the Yen price of the New Zealand Dollar ($Y_{t,12}$) as $Y_{t,12} = X_{t,2}/X_{t,1}$.

The call option payoff for the cross-rate between the Japanese Yen and the New Zealand Dollar, maturing at time $t$ with a strike of $K$, is defined as

$$\mathrm{CR}(t, X_{t,1}, X_{t,2}, K) := (X_{t,2} - KX_{t,1})^+.$$

The price bounds of this function, obtained through the VMOT approach, can indicate mispricing if the option price of the illiquid cross-rate exceeds the price bounds implied by the liquid cross-rate options. (TL: I don't get this.)

Foreign exchange risk has a considerable impact on supply chain performance in global supply chains that span multiple nations [38], and decision makers in a multi-period supply chain would therefore include foreign exchange as a critical consideration [20]. To mitigate this risk, a foreign exchange $N$-period cap can be used as a risk management instrument for locking in the exchange rate. This cap's payoff is given by

$$\mathrm{CCR}(X_1, ..., X_N) := \sum_{i=1}^{N} \lambda_i \mathrm{CR}(t_i, X_{t_i,1}, X_{t_i,2}, K_i) = \sum_{i=1}^{N} \lambda_i (X_{t_i,2} - K_i X_{t_i,1})^+$$

where a predetermined sequence of times $\{t_i\}_{i=1}^N$, weights $\{\lambda_i\}_{i=1}^N$, and strikes $\{K_i\}_{i=1}^N$ are used. However, if the instrument itself is mispriced, its effectiveness in risk mitigation is reduced. The VMOT approach for detecting mispricing can be useful in this regard.

2.2. **Enhancing Risk Analysis in Mean-Variance Protfolio with VMOT's Implied Distribution Covariance Matrix.** Given the option-implied marginals of $X_i$ and $Y_i$ for $i = 1, 2$, the VMOT provides a method for calculating option implied ex-ante covariance bounds among different stocks via the mutual covariances cost function: TL: Isn't this just the same as (2.5) or (2.6) if $i = 1, 2$?

$$c(x, y) = \sum_{i<j} a_{ij} x_i x_j + \sum_{i<j} b_{ij} y_i y_j \tag{2.11}$$

which providing valuable insights into the implied variance of the associated portfolio. Unlike the covariance matrix used in the mean-variance portfolio, which is typically based on historical stock movements (under the physical measure), VMOT constructs the covariance matrix (TL: How? And what is the "covariance matrix"?) using the implied distribution derived from option prices (under the risk-neutral measure). These bounds provide valuable insights into the ex-ante joint distribution of prices for multiple assets at distinct future times.

It is important to note that the implied variance derived from option prices differs from physical variance. However, the risk-neutral variance, which is widely accepted as a reference for true variance, can be derived from option prices. Empirical evidence suggests that option implied ex-ante higher moments hold predictive power for future stock returns [11]. This indicates that option prices encompass valuable market information and reflect investors' expectations of future stock moments [28].

One significant advantage of VMOT is its ability to tighten price bounds by setting the $a_{ij}$ values to zero while still considering the information from the first period. Even the cost function depends solely on the terminal period, VMOT incorporates valuable insights from both periods. This innovative approach surpasses the limitations of classical optimal transport methodologies, providing a more accurate assessment of risk.

## 3. Monotone geometry of VMOT

In this section, we present our theoretical findings. To begin, we will discuss the concept 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 $\beta$ 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. In addition, 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 $\beta$ is twice differentiable, then $\beta$ 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.

**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* $\mu$ *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* $\mu_i$ *(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.

**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* $\pi$ *for the problem* (2.3) *whose first marginal* $\pi^X$ *is the monotone coupling of* $\vec{\mu} = (\mu_1, ..., \mu_d)$*. Moreover, if* $c_1$ *is strictly supermodular, then every VMOT* $\pi$ *has monotone first marginal* $\pi^X = \chi_{\vec{\mu}}$.

Monotonicity of optimizers in the classical optimal transport problem for supermodular costs is well known [9][34]. Results asserting higher dimensional determininstic solutions, such as those of Brenier [7] (for two marginals) and Gangbo-Świech [19] (for three or more marginals) regarding the cost function $c(x) = \sum_{1 \leq i < j \leq d} x_i \cdot x_j$, and generalizations to other cost functions [32], [18] [8] (for two marginals) [24][36][29][37] (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 $\mu_i$ is a dirac mass (corresponding to the case when the first time is the present), then the VMOT problem reduces to the classical (multi-marginal) optimal transport problem on the $\nu_i$'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 $\pi^X, \pi^Y$) are equal to the monotone coupling of $\vec{\mu}$ and $\vec{\nu}$ respectively, i.e., $\pi^X = \chi_{\vec{\mu}}$ and $\pi^Y = \chi_{\vec{\nu}}$, 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 $\pi$ implies the convex order condition $\pi^X \preceq_c \pi^Y$. 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 $[-1, 1]$, $\nu_1$ be uniform on $[-3, 3]$, and $\nu_2$ 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 $l_1 \not\subseteq l_2$. This shows that the ideal situation, i.e., $\chi_{\vec{\mu}} \preceq_c \chi_{\vec{\nu}}$, is infeasible in general.

Nevertheless, it is plausible that a VMOT $\pi$ may still couple the marginals $\vec{\mu}$ monotonically[3], i.e., a VMOT $\pi$ sets $\pi^X = \chi_{\vec{\mu}}$ thereby maximizing $\mathbb{E}_{\pi^X}[c_1(X)]$, then seek $\pi_Y$ which satisfies $\chi_{\vec{\mu}} \preceq_c \pi_Y$ while $\pi_Y$ is as close as the ideal $\chi_{\vec{\nu}}$, so that $\pi_Y$ 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.

[15] showed that the conjecture is correct when the marginals $\mu_i, \nu_i$ satisfy a special criterion known as linear increment of marginals, which is the case if e.g. $\mu_i, \nu_i$ are gaussians with increasing variance; see also [33, Example 4.5] for a related discussion. In this paper, we prove that the conjecture is indeed

[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 $\pi^X = \gamma$.

correct if $d = 2$ (without any particular condition imposed on the marginals), 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 every $d \geq 2$. The distinction is due to the convex ordering constraint $\pi^X \preceq_c \pi^Y$, which every martingale transport $\pi$ 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 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 semi-continuous 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 semi-continuous function satisfying $f^{**} \leq f$. We call $f^{**}$ the *convex envelope* of $f$.

**Proposition 3.4.** *i) If $\beta$ on $\mathbb{R}^d$ is submodular, then $\beta^*$ is supermodular.*
*ii) If $d = 2$ and $\beta$ on $\mathbb{R}^2$ is supermodular, then $\beta^*$ is submodular.*
*iii) If $\beta$ on $\mathbb{R}^2$ is sub/supermodular, then $\beta^{**}$ 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 but for the sake of readers it may be ok)

**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. Assume the same condition as in Theorem 2.1, and that the second moments of $\mu_1, \mu_2$ are finite. Then:*
*i) There exists a VMOT $\pi$ such that its first time marginal $\pi^X$ is the monotone coupling of $\mu_1, \mu_2$.*
*ii) If $c_1$ is strictly supermodular, then every VMOT $\pi$ satisfies that its first time marginal $\pi^X$ is the monotone coupling of $\mu_1, \mu_2$.*

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

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

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

for every VMOT $\pi$ which solves the problem (2.3). We define

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

and rewrite the above as

$$c_1(x) - \sum_{i=1}^{d} \big(\phi_i(x_i) + h_i(x)(y_i - x_i)\big) \leq \beta(y) \quad \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d \tag{3.4}$$

$$c_1(x) - \sum_{i=1}^{d} \big(\phi_i(x_i) + h_i(x)(y_i - x_i)\big) = \beta(y) \quad \pi - a.s.. \tag{3.5}$$

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

$$c_1(x) - \sum_{i=1}^{d} \big(\phi_i(x_i) + h_i(x)(y_i - x_i)\big) \leq \beta^{**}(y) \quad \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d \tag{3.6}$$

$$c_1(x) - \sum_{i=1}^{d} \big(\phi_i(x_i) + h_i(x)(y_i - x_i)\big) = \beta^{**}(y) \quad \pi - a.s.. \tag{3.7}$$

Then by equating $y_i = x_i$, (3.6) yields

$$c_1(x) - \sum_{i=1}^{d} \phi_i(x_i) \leq \beta^{**}(x) \quad \forall x \in \mathbb{R}^d. \tag{3.8}$$

On the other hand, for any VMOT $\pi = \pi_x \otimes \pi^X$ (TL: did we introduce kernel and disintegration?), by integrating (3.7) with respect to the martingale kernel $\pi_x(dy)$, we obtain

$$c_1(x) - \sum_{i=1}^{d} \phi_i(x_i) = \int \beta^{**}(y) d\pi_x(y) \quad \pi^X - a.e.\, x, \tag{3.9}$$

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 $\int \beta^{**}(y) d\pi_x(y) \geq \beta^{**}(x)$, since $\beta^{**}$ is convex. In view of this, (3.8) yields

$$c_1(x) - \sum_{i=1}^d \phi_i(x_i) = \beta^{**}(x) \quad \pi^X - a.s.. \tag{3.10}$$

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

$$\sum_{i=1}^d \phi_i(x_i) \geq \tilde{c}(x) \quad \forall x \in \mathbb{R}^d, \tag{3.11}$$

$$\sum_{i=1}^d \phi_i(x_i) = \tilde{c}(x) \quad \pi^X - a.s.. \tag{3.12}$$

(3.11) and (3.12) implies that for any VMOT $\pi$, its first time marginal $\pi^X$ solves the optimal transport problem with the cost $\tilde{c}$ and marginals $\mu_1, ..., \mu_d$, that is, $\pi^X$ 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 the fact that $\beta$ is submodular and Proposition 3.4 iii), $\tilde{c}$ is also strictly supermodular. Then Fact 1 implies that $\pi^X$ must be the monotone coupling of $\mu_1, \mu_2$. This proves part ii) of the theorem.

To prove part i), fix $\delta > 0$, and choose a VMOT $\pi$ 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., $\pi^X$ is the monotone coupling of $\vec{\mu} = (\mu_1, \mu_2)$. Moreover, since $\pi$ is a VMOT, its second time marginal $\pi^Y$ 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 in turn implies that $\pi$ is a VMOT for the cost $c_\delta(x, y)$ for every $\delta > 0$. Letting $\delta \searrow 0$, we deduce that $\pi$ is still a VMOT for the cost $c(x, y) = c_1(x) + c_2(y)$. This proves part i). $\square$

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 $\mu$ 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. Assume the same condition as in Theorem 2.1, and that the second moments of $\mu_1, \mu_2$ are finite. Then:*
*i) There exists a VMOT $\pi$ such that its first time marginal $\pi^X$ is the anti-monotone coupling of $\mu_1, \mu_2$.*

*ii) If $c_1$ is strictly submodular, then every VMOT $\pi$ satisfies that its first time marginal $\pi^X$ is the anti-monotone coupling of $\mu_1, \mu_2$.*

**Remark 3.7.** *A financial interpretation of these results may be given as follows. For a two-period cap whose payoff depends on two underlying assets, Theorem 3.5 together with corollary 3.6 describes the extreme market model which attains the extremal of the price bounds: the (anti-)monotonicity implies that there exists a market model under which the two assets are controlled by a single factor in the first period. Furthremore, if $c_1$ is strictly sub or supermodular, then every extremal market model exhibits this property.*

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

**Example 3.8.** *The strict supermodularity of $c_1$ is necessary for part ii) of Theorem 3.5. To construct a counterexample VMOT $\pi$ 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*

$$\begin{aligned} \Gamma_{\{\psi_1,\psi_2\}} &:= \{(y_1, y_2) \in \mathbb{R}^2 \mid \psi_1(y_1) + \psi_2(y_2) = y_1 y_2\} \\ &= \{(y_1, y_2) \mid y_2 = \psi_1'(y_1)\} \\ &= \{(y_1, y_2) \mid y_2 = |y_1|^2 \textit{ if } y_1 \geq 0,\ y_2 = -|y_1|^2 \textit{ if } y_1 \leq 0\}. \end{aligned} \tag{3.13}$$

*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*

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

*Such a choice is possible because $\mathrm{conv}(\Gamma_{\{g_1,g_2\}}) = \mathbb{R}^2$. We then define a martingale measure $\pi$ via $\pi = \pi_x \otimes \pi^X$, i.e., its first marginal is $\pi^X$ and its kernel is $\{\pi_z, \pi_w\}$. Now take $c(x, y) = y_1 y_2$ (so that $c_1(x) = 0$), $\phi_1 = \phi_2 = h_1 = h_2 = 0$, and notice that $\{\phi_i, \psi_i, h_i\}_{i=1,2}$ and $\pi$ then jointly satisfy the optimality condition (2.9), (2.10). This implies that $\pi$ 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 $\pi$. 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 3.9.** *In $\mathbb{R}^3$, there is a supermodular function $\beta : \mathbb{R}^3 \to \mathbb{R}$ for which $\beta^*$ is not submodular. As an example, for $x \in \mathbb{R}^3$, let $\beta$ 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, hence $\beta$ is a strongly convex function, in which case $\beta^*$ is given by*

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

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

$$\frac{1}{206}\begin{bmatrix} 23 & 4 & -9 \\ 4 & 144 & -118 \\ -9 & -118 & 111 \end{bmatrix}. \tag{3.14}$$

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

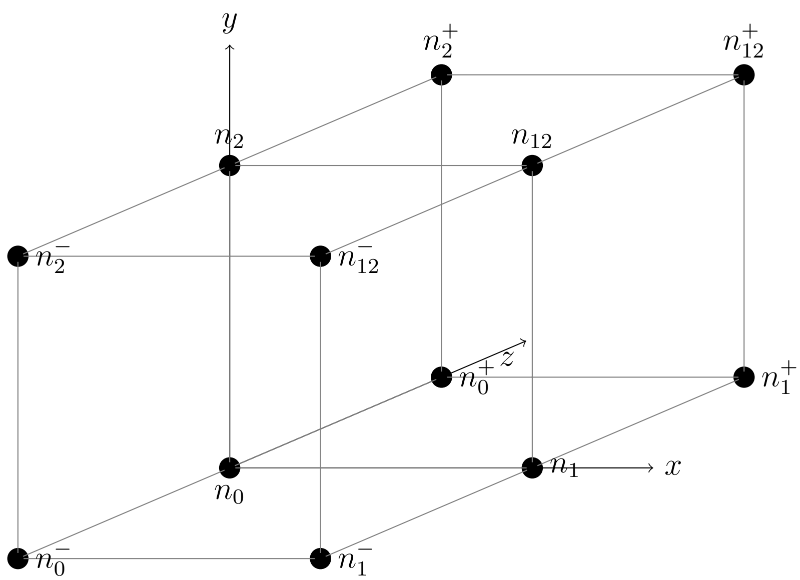
$$\frac{\partial^2 \beta^*}{\partial x_i \partial x_j} \leq 0 \quad \textit{for every } i \neq j.$$

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

Despite Example 3.9, 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 3.10** (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:*

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




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

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

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

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

*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* $\beta_0^{**}$ *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*

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

*yielding that* $\beta_0^{**}$ *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 3.11.** *Conjecture 1 is false if* $d \geq 3$*. Specifically, there exist vectorial marginals* $\vec{\mu} = (\mu_1,\mu_2,\mu_3)$*,* $\vec{\nu} = (\nu_1,\nu_2,\nu_3)$ *satisfying* $\mu_i \preceq_c \nu_i$*,* $i = 1,2,3$*, such that for every VMOT* $\pi$ *to the problem* (2.3) *with the cost* $c = c(y) = y_1y_2 + y_2y_3 + y_3y_1$ *and the marginals* $\vec{\mu},\vec{\nu}$*, its first time marginal* $\pi^X$ *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 $c$? (TL: revised, is it ok now?)

*Proof.* Let $\mathcal{Y}$ be the set of twelve points in $\mathbb{R}^3$ and $\beta_0$ be the submodular function as in Example 3.10. 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 3.10, 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 3.10, 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 $\pi$. For this, take $\pi^X := \frac{1}{2}(\delta_u + \delta_{u'})$ as the first marginal of $\pi$. Then we take the martingale kernel $\pi_x$ as the unique probability measure supported on $\mathcal{Z}$ with its barycenter $x$ in $\operatorname{conv}(\mathcal{Z})$. Now define $\pi = \pi_x \otimes \pi^X$ (note that we only need $\pi_x$ for $x = u, u'$), and let $\vec{\mu} := (\mu_1, \mu_2, \mu_3)$ be the 1D marginals of $\pi^X$, and let $\vec{\nu} := (\nu_1, \nu_2, \nu_3)$ be the 1D marginals of $\pi^Y = \frac{1}{2}(\pi_u + \pi_{u'})$. We now claim that $\pi$ is a VMOT solving the problem (2.3) with the cost $c$ and the marginals $\vec{\mu}, \vec{\nu}$.

We will prove the optimality of $\pi$ by locating an associated dual optimizer $(\phi_i, \psi_i, h_i)_{i=1,2,3}$, where $\psi_i$ has already been defined above. To define $\phi_i$, recall the affine functions $L_1, L_2, L_3$ in Example 3.10. 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

$$h(x) = \begin{cases} \nabla L_1 = (0,0,0) \ \text{ if } \ L_1(x) > L_2(x), \\ \nabla L_2 = (1,1,-1) \ \text{ if } \ L_1(x) < L_2(x), \\ \nabla \frac{L_1+L_2}{2} = (\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}) \ \text{ if } \ L_1(x) = L_2(x). \end{cases} \tag{3.15}$$

In order to prove the optimality of $\pi$ and $(\phi, \psi, h)$ simultaneously, we need to confirm the optimality conditions (2.9) and (2.10). To see (2.9), 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 (2.10) follows by the fact that on $\mathcal{Z}$, $L_3 \le L_1 = L_2 = \beta$, such that the above inequalities become equality for $x = u, u'$ and $y \in \mathcal{Z}$. This simultaneously proves the optimality of $\pi$ and $(\phi_i, \psi_i, h_i)_i$ for the primal and dual problem respectively with the cost $c$ and the marginals $\vec{\mu}, \vec{\nu}$.

Finally, take any $\gamma \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$, such that its first time marginal $\gamma^X$ is monotone, i.e., $\gamma^X = \frac{1}{2}(\delta_{\bar{u}} + \delta_{\bar{u}'})$. We claim that $\gamma$ cannot be optimal. If $\gamma$ were optimal, it must satisfy the optimality condition (2.10) with the optimal dual $(\phi, \psi, h)$ 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

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

This indicates that $\gamma$ cannot satisfy (2.10), thereby completing the proof. $\square$

**Remark 3.12.** *We showed that the VMOT $\pi$ constructed in Proposition 3.11 cannot have a monotone first marginal $\pi^X$. In other words, we showed*

$$\mathbb{E}_\pi[c(Y)] > \max_{\gamma \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})} \left\{ \mathbb{E}_\gamma[c(Y)] \;\middle|\; \gamma^1 \textit{ is monotone; } \gamma^1 = \frac{\delta_{\bar{u}} + \delta_{\bar{u}'}}{2} \right\} \tag{3.16}$$

*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* $\lambda \geq 0$*. Because the inequality* (3.16) *is strict, it remains strict for the cost* $c_\lambda$ *with sufficiently small positive* $\lambda$*. In other words, even if the cost function involves* $\lambda c(x)$ *which is strictly supermodular,* $\pi^X$ *is still not monotone for every VMOT* $\pi$*, as long as* $\lambda$ *is not too large. On the other hand, since*

$$c(u) + c(u') < c(\bar{u}) + c(\bar{u}'), \tag{3.17}$$

*inequality* (3.16) *is reversed for* $c = c_\lambda$ *with sufficiently large* $\lambda$*, in which case every VMOT* $\pi$ *now has 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.*

*An intuition can be given as follows: Because it is more important to maximize* $\mathbb{E}[c(Y)]$ *for small* $\lambda$*, a VMOT* $\pi$ *promotes its second time marginal* $\pi^Y$ *to be supported on the monotone set* $\mathcal{Z}$*, even if this necessitates supporting its first marginal* $\pi^X$ *on a non-monotone set* $\{u, u'\}$*. However, as* $\lambda$ *grows larger, maximizing* $\mathbb{E}[c(X)]$ *becomes more important, so a VMOT* $\pi$ *promotes its first marginal* $\pi^X$ *to be supported on the monotone set* $\{\bar{u}, \bar{u}'\}$ *even if this requires* $\pi^Y$ *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 3.13.** *The functions* $\psi_i$ *appearing as part of dual optimizers in Example 3.11 appear quite singular. However, they can be made continuous and convex by applying the martingale Legendre transform [22]. Recall* (2.9)*, which we rewrite in this case as*

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

*which holds for all* $x, y \in \mathbb{R}^3$*. In view of this, the martingale Legendre transform of* $\psi_1$ *can be naturally defined by*

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

*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* $\psi_1$ *is finite on* $\mathcal{Y}$, *convexity of* $\tilde{\psi}_1$ *implies that it is finite in* $\operatorname{conv}(\mathcal{Y})$. *Similarly, we can replace* $\psi_2$ *and* $\psi_3$ *with their martingale Legendre transforms. Then* $(\phi_i, \tilde{\psi}_i, h_i)_i$ *continues to be a dual optimizer.*

**Remark 3.14.** *[15] showed that if each pair of marginals* $(\mu_i, \nu_i)$ *is Gaussian with equal mean and increasing variance* $\operatorname{Var}(\mu_i) < \operatorname{Var}(\nu_i)$ *(or more generally if each pair satisfies the linear increment of marginals condition), then the first marginal* $\pi^X$ *of any VMOT* $\pi$ *with respect to the cost* $c = c(y) = \sum_{1 \leq i < j \leq d} y_i y_j$ *is the monotone coupling of* $\vec{\mu} = (\mu_1, ..., \mu_d)$. *But the linear increment is a very restrictive assumption on the marginals. In view of this, we believe finding other sufficient conditions on the cost and marginals for which the conjecture holds true 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? (TL: revised. Is it ok now?)

Table 1 concludes this section by presenting several submodular payoff functions for exotic options, some of which are also seen in [1]. Theorem 3.5 shows that in the first period, the extremal model for the two-period cap over the two-asset option with a sub- or super-modular payout reduces to a single factor model. In addition, we document the payoff of a cross rate option between two illiquid currencies as well as the sum of covariances functional.

## 4. Numerics (Improved section / subsection titles ?)

TL: We may want to provide some intro / summary of this section here, explaining the structure of the section concisely.

### 4.1. Reformulation exploiting monotonicity.

An immediate consequence of Theorem 3.5 is that when $d = 2$ and $c(x, y) = c_1(x_1, x_2) + c_2(y_1, y_2)$ with supermodular $c_1, c_2$, the VMOT problem (2.3) is equivalent to

| 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 among $d$ stocks | $\sum_{i,j} a_{ij} X_i X_j + b_{ij} Y_i Y_j$ |
| Call on cross rate among illiquid two currencies | $(X_{t,2} - K X_{t,1})^+$ |

TABLE 1. A list of submodular derivative payouts

$$\text{maximize } \ \mathbb{E}_\pi[c(X,Y)] \ \text{ over } \ \pi \in \mathrm{VMT}(\mu, \vec{\nu}) \tag{4.1}$$

where (notice there is no arrow on $\mu$)

$$\mathrm{VMT}(\mu,\vec{\nu}) := \{\pi \in \mathcal{P}(\mathbb{R}^{2d}) \mid \pi = \mathrm{Law}(X,Y),\, \mathbb{E}_\pi[Y|X] = X, \\ \mathrm{Law}(X) = \mu,\, \mathrm{Law}(Y_i) = \nu_i \text{ for all } i \in [d]\}, \tag{4.2}$$

and $\mu \in \mathcal{P}(\mathbb{R}^d)$ is the monotone coupling of the $\{\mu_i\}_{i\in[d]}$. We note that $\mu$ 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 $\mu_i$ and $\mathcal{L}_{[0,1]}$ denotes the Lebesgue, or uniform probability, measure on the unit interval $[0,1]$. For simplicity, we also replace each $\nu_i$ with ${G_i^{-1}}_\# \mathcal{L}_{[0,1]}$ where $G_i^{-1}$ is the inverse cumulative of $\nu_i$. With this, observe that we can rewrite (4.1) 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 $\pi$ in the original VMOT space.)

$$\text{maximize } \ \mathbb{E}_\pi[\tilde{c}(U,V)] \ \text{ over } \ \tilde{\pi} \in \mathrm{CVMT}(\Phi,\Psi) \tag{4.3}$$

where $U \in (0,1)$, $V = (V_1, ..., V_d) \in (0,1)^d$ are random variables, and

$$\mathrm{CVMT}(\Phi,\Psi) := \{\tilde{\pi} \in \mathcal{P}([0,1] \times [0,1]^d) \mid \tilde{\pi} = \mathrm{Law}(U,V), \\ \mathbb{E}_\pi[\Psi(V) \,|\, U] = \Phi(U),\ \mathrm{Law}(U) = \mathrm{Law}(V_i) = \mathcal{L}_{[0,1]} \text{ for all } i \in [d]\} \tag{4.4}$$

where $\Phi(u) := (F_1^{-1}(u), F_2^{-1}(u), ..., F_d^{-1}(u))$, $v = (v_1, v_2, ..., v_d)$, $\Psi(v) := (G_1^{-1}(v_1), G_2^{-1}(v_2), ..., G_d^{-1}(v_d))$ and $\tilde{c}(u,v) := c(\Phi(u), \Psi(v))$.[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...(TL: I removed $\mathcal{L}_{[0,1]}$ notations)

Now the dual problem to (4.3) is formulated as

$$\inf_{(\tilde{\phi}, \tilde{\psi}_i, \tilde{h}_i) \in \tilde{\Xi}} \int \tilde{\phi}(u)du + \sum_{i=1}^{d} \int \tilde{\psi}_i dv_i \tag{4.5}$$

where $\tilde{\Xi}$ 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 the following inequality holds for every $(u,v) \in (0,1) \times (0,1)^d$:

$$\tilde{\phi}(u) + \sum_{i=1}^{d} \left( \tilde{\psi}_i(v_i) + \tilde{h}_i(u)(G_i^{-1}(v_i) - F_i^{-1}(u)) \right) \geq \tilde{c}(u,v). \tag{4.6}$$

As with the primal problem, this dual is much simpler than the original (2.8), as we replace the $d$ functions $\phi_i$ with the single function $\tilde{\phi}$, while the functions $h_i$ now depend on the single variable $u \in (0,1)$ rather than $x \in \mathbb{R}^d$. In the following, we exploit this simplified structure to develop a numerical method to compute solutions to (4.3) (and, consequently, to 2.3 for $d = 2$ and appropriate cost $c$.)

We note that several numerical methods have been proposed in the literature to solve the martingale optimal transport problem. One approach involves transforming the problem into a relaxed linear programming (LP) formulation through discretization [23]. Another strategy incorporates entropic regularization [13] which still necessitates discretization. A major potential disadvantage of these approaches is the curse of dimensionality when dealing with an increasing number of marginals or dimensions.

[4]We note carefully that although this problem can be formulated for any cost $c$ and dimension $d$, 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.

In recent years, the application of neural networks to solve optimal transport problems has gained attention [21, 31, 39]. This neural network-based approach has also been extended to the context of martingale transport [15, 16, 26]. One notable advantage of employing neural networks is their ability to effectively handle the curse of dimensionality through powerful approximation techniques. It is worth noting that [15], [26] have also applied neural network-based numerical martingale transport methods in finance.

4.2. **Numeric.** We propose a numerical method that based on neural networks with penalization for calculating the optimal value, optimal distribution and the corresponding dual functions. The trained neural networks can be viewed as the portfolio of optimal semi-static trading strategy which super/sub-replicate the given payoff function. It will provide a ready-to-use instruction for hedgers to replicate the corresponding derivatives. We apply the framework of [16], further developed in [15] and analyze potential performance improvements allowed by our result about the geometry of the optimal couplings. Our tests run across various dimensions, cost functions and marginal types, including marginals derived from actual stock market option prices.

We develop several examples of robust pricing problems designed to show the improvement allowed by our result. We consider cost functions based on the the European call option applied to a basket of assets and a theoretical covariance (or cross-product) contract. The first two examples use hypothetical uniform and normal marginal distributions, while the third uses real-world marginal distributions implied by option prices of individual assets at two distinct future points in time.

The model introduced in [16] uses penalization and neural network optimization to solve the dual problem of Optimal Transport. More precisely, the model approximates the potential functions $(\phi_i, \psi_i, h_i)_i$ of equation 2.6 using neural networks and use penalization to impose condition 2.7. [15] introduces the martingale constraint to the optimal transport problem, providing a dual solution to $\mathsf{VMOT}$.

The novelty of our method is in the improvement of the sampling measure $\theta$, used for penalization: since we know that a solution (unique in the case of strict submodularity) is attained with monotone coupling of the $X$ marginals $(\vec{\mu}_i)$, we determine $\theta^1$, the $X$-side of $\theta$, by those marginals.

**Numerical implementation.** Below we numerically solve the VMOT problem for several choices of marginals and costs. We apply the numerical framework introduced in [16], both to the original VMOT problem 2.3 and the reduced-dimension version 4.3 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 better on the lower dimensional problem 4.3. The method in [16] 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:

$$\max_{\mu \in \mathcal{Q}} \int \tilde{c} d\mu - \frac{1}{\gamma} \int \beta^*(\frac{d\mu}{d\theta}) d\theta \tag{4.7}$$

where $\beta^*$ is the convex conjugate of the function $\beta$ which assumed to be a differentiable non-decreasing convex function such that $\lim_{x\to\infty} \beta(x)/x = \infty$. $\theta$ 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 $\theta$ to be the independent coupling of the marginals $\mu_1, ..., \mu_d$.

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

$$\inf_{\varphi \in \mathcal{H}} \int \varphi d\mu_0 + \int b_\gamma (\tilde{c} - \varphi) d\theta \tag{4.8}$$

where $b_\gamma := \frac{1}{\gamma}\beta(\gamma x)$ is a penalty function parameterized by $\gamma > 0$. [16, 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 (4.8), then the measure $\hat{\mu} \in \mathbb{R}^d$ given by

$$\frac{d\hat{\mu}}{d\theta} := b'_\gamma(\tilde{c} - \hat{\varphi}) \tag{4.9}$$

is an optimal measure to the problem (4.7). 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:

$$\varphi(u,v) = \tilde{\phi}(u) + \sum_{i=1}^{d} \left( \tilde{\psi}_i(v_i) + \tilde{h}_i(u)(G_i(v_i) - F_i^{-1}(u)) \right) \tag{4.10}$$

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:

$$\begin{aligned} \text{Loss} &= \int \varphi d\mu_0 + \int b_\gamma \left( c - \varphi \right) d\theta \\ &= \int \tilde{\phi}(u) du + \sum_{i=1}^{d} \int \tilde{\psi}_i d\nu_i + \int b_\gamma \left( c - \varphi \right) d\theta \end{aligned} \tag{4.11}$$

We fix the penalization function $b$ as

$$b\left(t\right) = \frac{1}{2\gamma} \left( (\gamma t)^+ \right)^2 \qquad \gamma = 1000 \tag{4.12}$$

Computationally, each of $\theta$, $\psi_i$, $h_i$ is replaced by some approximation $\theta^m$, $\psi_i^m$, $h_i^m$ 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 $\mu_0$ and $\theta$ – 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 [16].

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

**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 [15]. 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 simplification to the problem. In fact, using the dimension of the sample domain as a measure of computational complexity, this one is proportional to $2d$ in the full dimension case and to $d + 1$ in the reduced dimension.

To introduce some variability, we use randomization to define the cost functions and marginal dispersion parameters. We set $c(x, y) = \sum_{i<j} b_{ij} y_i y_j$ where 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 $\sigma_i$ is a random number between 1 and 2 and each $\rho_i$ 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 the resulting coefficients and parameters are available in the referenced code, to allow the examples to be reproduced.

Figure 1 shows the convergence of the dual value over the training epochs in the two formulations, namely, the full-dimension version 2.3 and our simplified, reduced-dimension version 4.3, 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.

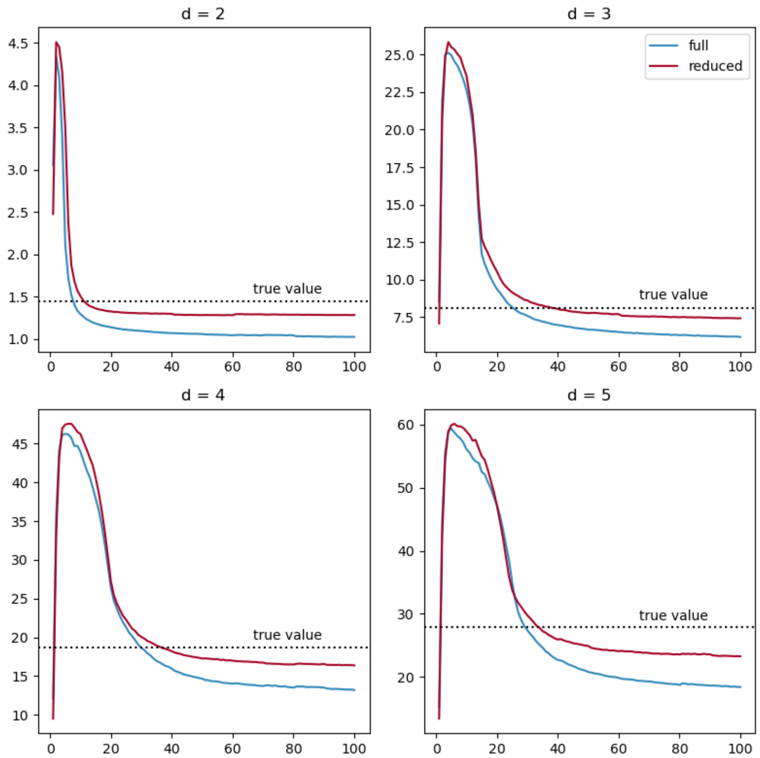



FIGURE 1. Convergence of the dual value – normal marginals, $d = 2$ to 5.

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.

| 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 $\varphi^m$ in both versions for $d = 2$ according to Theorem 2.2 of [16]. The left side shows the full dimension case, where we observe a convergence of the mass of $\pi^X$ 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.

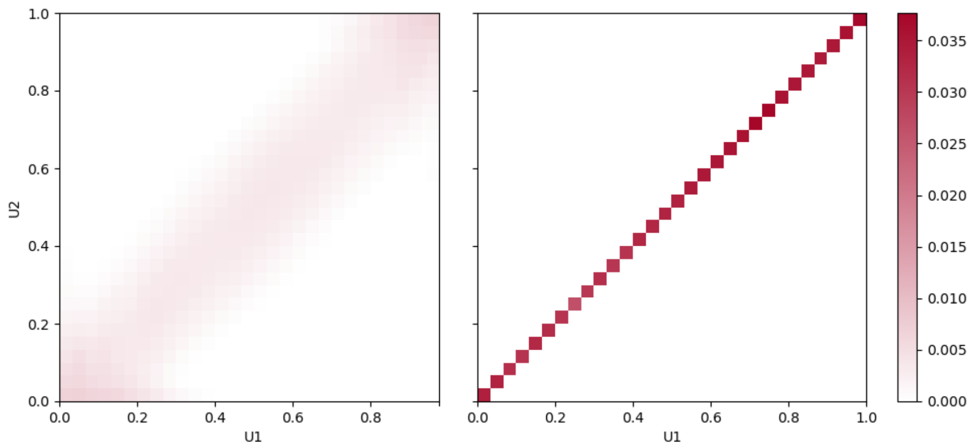



FIGURE 2. Heat map of $(\pi^X)^\star$, full (left) and reduced dimension (right).

BP: Again, I think considering variance of a portfolio is more natural than the price of a hypothetical contract.

**2. Empirical marginals.** We move on to a real world problem where we calculate upper and lower bounds for the covariance of the stock prices of Apple and Amazon at future date Feb. $17^{\text{th}}$, 2023 given their individual distributions on Jan. $20^{\text{th}}$ and Feb. $17^{\text{th}}$ implied by the call and put option prices as of Dec. $16^{\text{th}}$, 2022. Our model is consistent with the fact that the distribution of each price at each date is given and implied by the market, while their joint distribution is unknown. It 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. The resulting marginal distributions are shown in Figure 3.

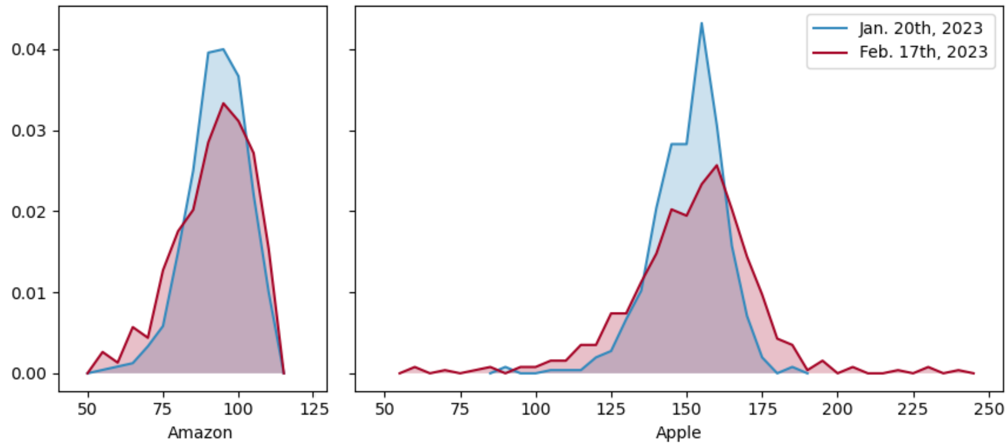



FIGURE 3. Marginal distributions of future prices as of Dec. 16th, 2022 (US$).

We calculate $p^+$ and $p^-$, the maximum and minimum for the cost function $c(x, y) = y_1 y_2$ given the marginals. Figure 5 shows the convergence of the dual value for both in simplified and full dimension cases. Since the optimal cost is unknown, we plot the (independent coupling) sample mean cost as a reference. We observe a significantly better convergence speed and accuracy for $p^-$, and slightly better convergence speed for $p^+$ in the reduced-dimension version.

(Alternative: graph showing the bounds with lower information. The problem is that we expected wider bounds with lower information, but we got narrower ones, probably indicating low accuracy in the numerics. To discuss.) We calculate $p^+$ and $p^-$, the maximum and minimum for the cost function $c(x, y) = y_1 y_2$ given the marginals. Figure 5 shows the convergence of the dual value for both in simplified and full dimension cases. For the sake of comparison, we include the upper and lower bounds that would be calculated if we only knew the distribution at $t = 2$ as a grey solid lines. It is known that, in the single time case, the positive and negative monotone couplings of the marginals are solutions to the optimal maximum and minimum costs. Thus, we approximate each of these values by taking the average cost over ten million points from its respective coupling. We also show the

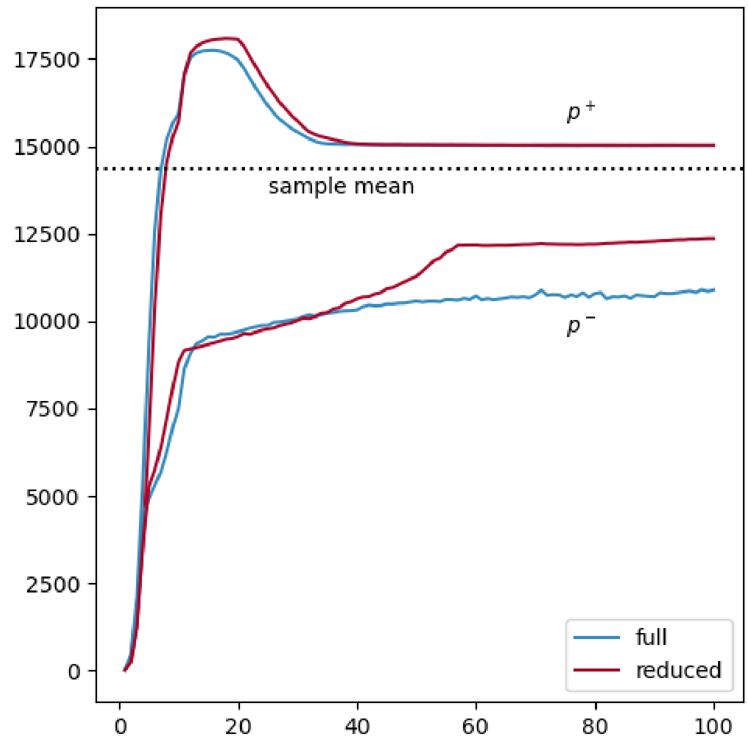



FIGURE 4. Convergence of the dual value – empirical marginals, $d = 2$.

(independent coupling) sample mean cost as a dotted line. We observe a significantly better convergence speed and accuracy for $p^-$, and slightly better convergence speed for $p^+$ in the reduced-dimension version.

## 5. CONCLUSION

In this paper we presented a geometrical result that can be seen as the vectorial martingale version of classical results in optimal transport, namely, that a solution to the VMOT problem with a sub or supermodular cost function has its first time-marginal monotonically supported when $d = 2$. We laid out a counterexample that negates the result for general dimension, at least when no additional requirement is imposed on the cost function. We provided examples of robust pricing problems that can be written as the VMOT of submodular functions(*). We noted that the main result gives

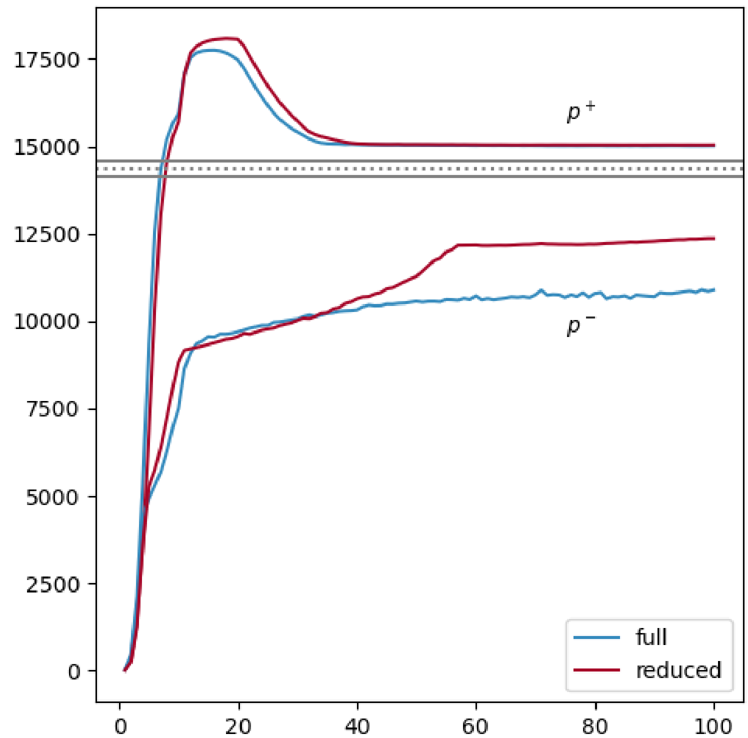



FIGURE 5. Convergence of the dual value – empirical marginals, $d = 2$.

opportunity to a dimensional reduction in the dual formulation of the problem, and calculated numerical solutions to the cross-product cost function on the dual side with and without the benefit of dimensional reduction, using both simulated and real data. A comparison between the outputs illustrated how the dimensional reduction helps improve accuracy. Our main result and the counterexample rise the question of whether additional structure on the cost function could be imposed to allow the result to be extended to higher dimensions with arbitrary marginal distributions.

## 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 $\beta^*$ is not proper, i.e., $\beta^* \equiv \infty$, again there is nothing to prove. So we assume $\beta$ and $\beta^*$ are proper.

Suppose $\beta$ 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 $\beta_R$ is submodular. And for large enough $R$, $\beta_R$ is proper. Now $\beta_R$ being compactly supported implies that $\beta_R^*$ will be Lipschitz unless $\beta_R^* \equiv +\infty$, but the latter is excluded since $\beta^* \geq \beta_R^*$. In particular, $\beta_R^*$ is real-valued everywhere. 3.4 We will first show the supermodularity of $\beta_R^*$. 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:

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

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 $\beta_R$:

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

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

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

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

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

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 $\beta_R^*$. 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 $\beta^*$.

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 [6], 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.

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

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] Jonathan Ansari, Eva Lütkebohmert, Ariel Neufeld, and Julian Sester. Improved robust price bounds for multi-asset derivatives under market-implied dependence information. *arXiv preprint arXiv:2204.01071*, 2022.

[2] Mathias Beiglböck, Pierre Henry-Labordère, and Friedrich Penkner. Model-independent bounds for option prices—a mass transport approach. *Finance and Stochastics*, 17(3):477–501, Jul 2013.

[3] Mathias Beiglböck and Nicolas Juillet. On a problem of optimal transport under marginal martingale constraints. *The Annals of Probability*, 44(1):42 – 106, 2016.

[4] Mathias Beiglböck, Tongseok Lim, and Jan Obłój. Dual attainment for the martingale transport problem. *Bernoulli*, 25(3):1640 – 1658, 2019.

[5] Mathias Beiglböck, Marcel Nutz, and Nizar Touzi. Complete duality for martingale optimal transport on the line. *The Annals of Probability*, 45(5):3038 – 3074, 2017.

[6] Douglas T Breeden and Robert H Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of business*, pages 621–651, 1978.

[7] Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *CR Acad. Sci. Paris Sér. I Math.*, 305:805–808, 1987.

[8] Luis Caffarelli. Allocation maps with general cost functions. *Partial differential equations and applications*, 177:29–35, 1996.

[9] G. Carlier. On a class of multidimensional optimal transportation problems. *J. Convex Anal.*, 10(2):517–529, 2003.

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

[11] Jennifer Conrad, Robert F. Dittmar, and Eric Ghysels. Ex ante skewness and expected stock returns. *The Journal of Finance*, 68(1):85–124, 2013.

[12] Areski Cousin, Stéphane Crépey, Olivier Guéant, David Hobson, Monique Jeanblanc, Jean-Michel Lasry, Jean-Paul Laurent, Pierre-Louis Lions, Peter Tankov, and David Hobson. The skorokhod embedding problem and model-independent bounds for option prices. *Paris-Princeton lectures on mathematical finance 2010*, pages 267–318, 2011.

[13] Hadrien De March. Entropic approximation for multi-dimensional martingale optimal transport. *arXiv preprint arXiv:1812.11104*, 2018.

[14] Yan Dolinsky and H Mete Soner. Martingale optimal transport and robust hedging in continuous time. *Probability Theory and Related Fields*, 160(1-2):391–427, 2014.

[15] Stephan Eckstein, Gaoyue Guo, Tongseok Lim, and Jan Obłój. Robust pricing and hedging of options on multiple assets and its numerics. *SIAM Journal on Financial Mathematics*, 12(1):158–188, 2021.

[16] Stephan Eckstein and Michael Kupper. Computation of optimal transport and related hedging problems via penalization and neural networks. *Applied Mathematics & Optimization*, 83:639–667, 2021.

[17] A. Galichon, P. Henry-Labordère, and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *The Annals of Applied Probability*, 24(1):312 – 336, 2014.

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

[19] Wilfrid Gangbo and Andrzej Świech. Optimal maps for the multidimensional monge-kantorovich problem. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 51(1):23–45, 1998.

[20] Srinagesh Gavirneni. Periodic review inventory control with fluctuating purchasing costs. *Operations Research Letters*, 32(4):374–379, 2004.

[21] Aude Genevay, Gabriel Peyré, and Marco Cuturi. Learning generative models with sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*, pages 1608–1617. PMLR, 2018.

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

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

[24] Henri Heinich. Problème de monge pour n probabilités. *Comptes Rendus Mathematique*, 334(9):793–795, 2002.

[25] Pierre Henry-Labordère. *Model-free hedging: A martingale optimal transport viewpoint*. CRC Press, 2017.

[26] Pierre Henry-Labordere. (martingale) optimal transport and anomaly detection with neural networks: A primal-dual algorithm. *arXiv preprint arXiv:1904.04546*, 2019.

[27] David G Hobson. Robust hedging of the lookback option. *Finance and Stochastics*, 2:329–347, 1998.

[28] Jens Carsten Jackwerth and Mark Rubinstein. Recovering probability distributions from option prices. *The Journal of Finance*, 51(5):1611–1631, 1996.

[29] Young-Heon Kim and Brendan Pass. A general condition for monge solutions in the multi-marginal optimal transport problem. *SIAM Journal on Mathematical Analysis*, 46(2):1538–1550, 2014.

[30] Frank Hyneman Knight. *Risk, uncertainty and profit*, volume 31. Houghton Mifflin, 1921.

[31] Alexander Korotin, Daniil Selikhanovych, and Evgeny Burnaev. Neural optimal transport. *arXiv preprint arXiv:2201.12220*, 2022.

[32] Vladimir Levin. Abstract cyclical monotonicity and monge solutions for the general monge–kantorovich problem. *Set-Valued Analysis*, 7(1):7–32, 1999.

[33] Tongseok Lim. Geometry of vectorial martingale optimal transportations and duality. *Mathematical Programming*, pages 1–35, 2023.

[34] Georg Gunther Lorentz. An inequality for rearrangements. *The American Mathematical Monthly*, 60(3):176–179, 1953.

[35] Basle Committee on Banking Supervision. *Revisions to the Basel II market risk framework*. Bank for International Settlements, 2009.

[36] Brendan Pass. Uniqueness and monge solutions in the multimarginal optimal transportation problem. *SIAM Journal on Mathematical Analysis*, 43(6):2758–2775, 2011.

[37] Brendan Pass and Adolfo Vargas-Jiménez. Monge solutions and uniqueness in multi-marginal optimal transport via graph theory. *Advances in Mathematics*, 428, 2023.

[38] Roberta Pellegrino, Barbara Gaudenzi, and George A. Zsidisin. Mitigating foreign exchange risk exposure with supply chain flexibility: A real option analysis. *Journal of Business Logistics*, 2023.

[39] Vivien Seguy, Bharath Bhushan Damodaran, Rémi Flamary, Nicolas Courty, Antoine Rolet, and Mathieu Blondel. Large-scale optimal transport and mapping estimation. *arXiv preprint arXiv:1711.02283*, 2017.

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