GEOMETRY OF VECTORIAL MARTINGALE OPTIMAL TRANSPORT AND ROBUST OPTION PRICING

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

Abstract. This paper addresses robust finance, which is concerned with the development of models and approaches that account for market uncertainties. Specifically, we investigate the Vectorial Martingale Optimal Transport (VMOT) problem, the geometry of its solutions, and its application with robust option pricing problems in finance. To this end, we consider two-period market models and show that when the spatial dimension d (the number of underlying assets) is 2, the extremal model for the cap option with a sub- or super-modular payout reduces to a single factor model in the first period, but not in general when $d > 2$. The result demonstrates a subtle relationship between spatial dimension, cost function supermodularity, and their effect on the geometry of solutions to the VMOT problem. We investigate applications of the model to financial problems and demonstrate how the dimensional reduction caused by monotonicity can be used to improve existing computational methods.

Keywords: Robust finance, Hedging, Martingale, Optimal transport, Duality, Monotonicity, Neural Network, Infinite-dimensional linear programming
MSC2020 Classification: 90Bxx, 90Cxx, 60Dxx, 60Gxx, 49Mxx, 65Kxx

---

Date: September 18, 2023.

All authors have contributed equally and are listed alphabetically. T.L. wishes to express gratitude to the Korea Institute of Advanced Study (KIAS) AI research group and the director Hyeon, Changbong for their hospitality and support during his stay at KIAS in 2023, where parts of this work were performed. B.P. is pleased to acknowledge the support of Natural Sciences and Engineering Research Council of Canada Discovery Grant number 04658-2018. The work of J.H. and M.S. is in partial fulfillment of their doctoral degrees.

1. Introduction

In mathematical finance, Knightian uncertainty [28] 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 [33]. One approach to addressing this issue is model independent finance, which was initially discussed in [25], 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 [6]. Nevertheless, uncertainty arises concerning the joint distributions 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 [23], 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 literature. The variant occurring 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 resulting optimization problem with this additional constraint is known as martingale optimal transport problem.

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. The optimization problem we investigate here, known in the literature as Vectorial Martingale Optimal Transport (also known as multi-marginal martingale optimal transport) problem, yields upper and lower bounds on the fair prices of contracts depending on several assets at two future times.

Regarding the geometry of solutions to the VMOT problems, examples worked out in [14, 31] suggest the following conjecture: for a certain class of payoff functions, the maximium arbitrage free price arises when the joint distribution (or coupling) of the underlying assets at the first maturity time is perfectly co-monotone. If this is 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 computational complexity; because the dependence structure of assets at the first time is known explicitly, only the dependence structure at the second time, as well as the martingale coupling structure between the times, needs to be 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 [15], 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

[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.

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 paper is organized as follows. In Section 2, we explain the vectorial martingale optimal transport problem. Section 3 examines the formulation and resolution of the monotonicity conjecture as well as counterexamples. Section 4 introduces a hybrid version of the VMOT problem and presents our numerical method and results. Section 6 provides miscellaneous proofs.

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 Transports* (VMT) from $\vec{\mu}$ to $\vec{\nu}$ (see [31]):

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

where $X = (X_1, ..., X_d)$, $Y = (Y_1, ..., Y_d) \in \mathbb{R}^d$ are random vectors. For a distribution $\pi \in \mathcal{P}(\mathbb{R}^{2d})$, 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 μ_i, that is,

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

Clearly, if $\pi \in \text{VMT}(\vec{\mu}, \vec{\nu})$, then $\pi^X \in \Pi(\vec{\mu})$ and $\pi^Y \in \Pi(\vec{\nu})$.

It is known that the set $\text{VMT}(\vec{\mu}, \vec{\nu})$ is nonempty if and 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 this paper.

Let $c : \mathbb{R}^{2d} \to \mathbb{R}$ be a (cost, or option payoff) function. We define the vectorial martingale optimal transport (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.

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 π a vectorial martingale transport, or VMT, if its one-dimensional marginals $\vec{\mu},\vec{\nu}$ are given, which condition is inspired by [2, 11, 13, 16, 23, 25]. [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 option whose payoff is fully determined at the terminal maturity t_2 by prices (X,Y) of the d assets at the two future 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 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

$$\big|\mathbb{E}_\pi[c(X,Y)]\big| \le \sum_i \big(\mathbb{E}_{\mu_i}[v_i(X_i)] + \mathbb{E}_{\nu_i}[w_i(Y_i)]\big) < \infty \;\text{ for any } \pi \in \mathrm{VMT}(\vec{\mu},\vec{\nu}).$$

This in turn 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:

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

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

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

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

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

so that $\mathbb{E}_\pi[c] = a\mathbb{E}_\pi[X_1X_2] + b\mathbb{E}_\pi[Y_1Y_2]$ represents a weighted sum of mutual covariances between X_1, X_2 and between Y_1, Y_2 under the market model π.

To provide a motivation for studying the cost function of the form (2.5), consider a portfolio consisting of the assets with prices $Y_1, ..., Y_d$ at the terminal maturity t_2, with weights $w_1, ..., w_d > 0$, so that the price is $\sum_{i=1}^d w_iY_i$. The variance is a commonly used measure of the risk of the portfolio:

$$\mathrm{Var}_\pi\left(\sum_{i=1}^d w_iY_i\right) = \mathbb{E}_\pi\left[\left(\sum_{i=1}^d w_iY_i\right)^2\right] - \left(\sum_{i=1}^d w_i\mathbb{E}_\pi[Y_i]\right)^2 \text{ given } \pi \in \mathrm{VMT}(\vec{\mu},\vec{\nu}).$$

In our VMOT framework, the term $(\sum_{i=1}^d w_i\mathbb{E}[Y_i])^2$ does not depend on π, since the individual distributions are assumed to be known. Finding the VMOT π yielding the maximal variance, or risk, is thus equivalent to maximizing $\mathbb{E}_{\pi^Y}[(\sum_i w_iY_i)^2] = \int_{\mathbb{R}^d}(\sum_i w_iy_i)^2 d\pi^Y(y)$. This represents a classical, (multi-marginal) optimal transport (OT) problem with cost function

$$c(x,y) = c(y) = \left(\sum_{i=1}^d w_iy_i\right)^2. \tag{2.7}$$

Notice this cost is equivalent to the cost (2.5) with each $a_{ij} = 0$ and $b_{ij} = w_iw_j$ in the VMOT problem, again due to the given marginal information.

Similar to OT, the VMOT problem reflects a worst case risk scenario in a situation when, in addition to the distributions ν_i at some future time t_2, the distributions μ_i of the assets at an earlier time $t_1 < t_2$ are known. In this situation, one may still try to evaluate the risk, or variance of the portfolio's value $\sum_{i=1}^d w_i Y_i$ at time t_2, but now incorporating the extra information coming from knowledge of distributions μ_i at t_1. This leads to the VMOT problem with cost (2.7). As a result, the VMOT approach can yield tighter bounds on the variance of the portfolio than the classical OT approach even though the cost functions here depend only on the terminal prices Y (that is, the intermediate prices X do not directly affect the value of the portfolio at the terminal time). The reason is that only couplings π^Y of the ν_i which dominate some coupling π^X of the μ_i in convex order can arise as feasible couplings in VMOT, whereas all coupings in $\Pi(\vec{\nu})$ are allowed as candidate optimizers in the classical OT. In Section 4, this phenomenon will be demonstrated numerically using empirical financial data.

It is worth noting that the maximal implied variance, which is the solution of the VMOT problem when the marginals are the risk-neutral distributions of the Y_i, differs from the physical variance, which is derived by a physical joint distributions of the Y_i. Importantly, option prices can be used to calculate the risk-neutral variance, which is widely used as a proxy for true variance. Empirical evidence suggests that option implied ex-ante higher moments hold predictive power for future stock returns [10]. This indicates that option prices encompass market information and reflect investors' expectations of future stock moments [26]. The VMOT approach, which derives marginal distributions from option prices, resonates with these findings.

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 Ξ 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 a concrete financial meaning. Specifically, (2.9) describes a semi-static super-replicate of the path-dependent payoff c, where $\phi_i(X_i), \psi_j(Y_j)$ are the European options written on the individual underlying at a specific time, and h_i is the amount to hold i^{th} asset for the delta hedging. It is worth noting that h_i is a function of the past prices of all assets $\{X_i\}_{i=1}^d$, and the left hand side of (2.9) yields the overall payoff of the superhedging portfolio $(\phi_i, \psi_i, h_i)_{i=1}^d$ given the price path (x, y). If the problem (2.3), on the other hand, is a minimization problem, then the dual problem represents 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 Ξ 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 [31] for $d \geq 2$. More specifically, [31] presents the following dual attainment result.

Theorem 2.1 ([31]). *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 *π 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 π - almost every price path (x, y), and that we do not impose an integrability condition on the dual minimizer.

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]

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 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. 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 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.

Definition 3.2. *i) $\{a \vee b, a \wedge b\}$ is called monotone rearrangement of $\{a, b\}$.*
ii) A set $A \subseteq \mathbb{R}^d$ is called monotone if for any $a, b \in A$, $\{a, b\} = \{a \vee b, a \wedge b\}$.
iii) A measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ is called 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 $\mu_1, ..., \mu_d$

[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 $I := \{x \in \mathbb{R} \mid u_\mu(x) < u_\nu(x)\}$ is a connected interval containing the full mass of μ, i.e., $\mu(I) = \mu(\mathbb{R})$.

as its marginals (i.e., $\chi_{\vec{\mu}} \in \Pi(\vec{\mu})$), is called the monotone coupling of $\vec{\mu}$. Note that $\chi_{\vec{\mu}} = (F^{-1}_{\mu_1}, F^{-1}_{\mu_2}, , ..., F^{-1}_{\mu_d})_{\#}\mathcal{L}_{[0,1]}$,[3] *where each $F^{-1}_{\mu_i}$ denotes the inverse of the cumulative distribution function of the μ_i (i.e., the quantile function), and $\mathcal{L}_{[0,1]}$ denotes the uniform probability measure on the interval $[0, 1]$.*

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.*

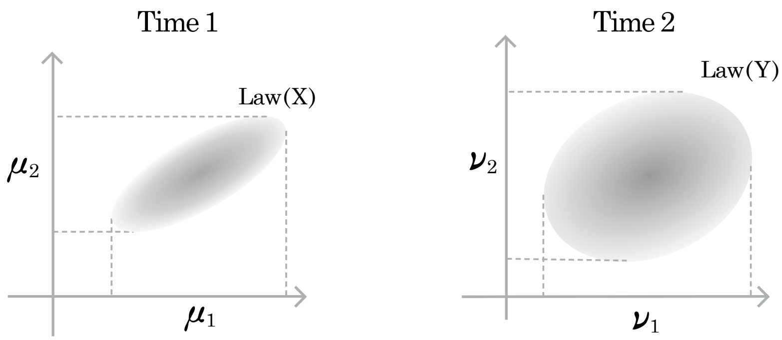
The following is the main question we will investigate in this section.

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 π for the problem (2.3) whose first time marginal π^X 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 $\pi^X = \chi_{\vec{\mu}}$.*

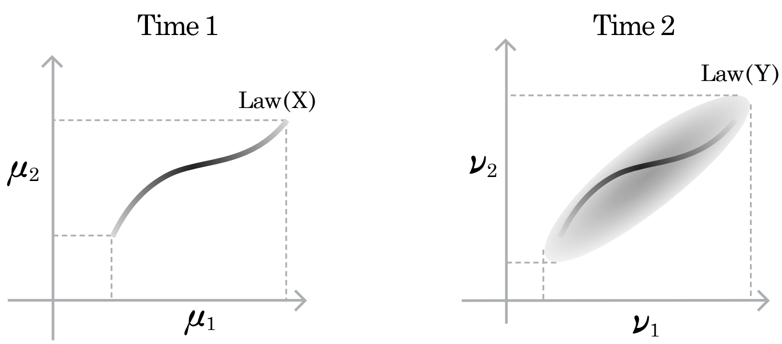
Monotonicity of optimizers in the classical optimal transport problem for supermodular costs is well known [9, 32]. Results asserting higher dimensional determininstic solutions, such as those of Brenier [7] (for two marginals) and Gangbo-Święch [18] (for three or more marginals) regarding the cost function $c(x) = \sum_{1 \leq i < j \leq d} x_i \cdot x_j$, and generalizations to other cost functions [8, 17, 30] (for two marginals) [22, 27, 34, 35] (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), then 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 $\vec{\mu}$

[3] Given a measurable map $F : \mathcal{X} \to \mathcal{Y}$ and a measure μ on $\mathcal{X}$, the push-forward of μ by F, denoted by $F_{\#}\mu$, is a measure on $\mathcal{Y}$ satisfying $F_{\#}\mu(A) = \mu(F^{-1}(A))$ for every $A \subseteq \mathcal{Y}$.



(A) Generic marginals of a vectorial martingale transport.



(B) A martingale transport with monotone first time marginal.

FIGURE 1. Conjecture 1 asserts that the martingale transport in (1b) can be superior to the one in (1a) for the problem (2.3) due to the supermodularity of c_1, c_2. We overlap Law(X) with Law(Y) in (1b) to emphasize they must be in convex order.

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 π 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 $[-1,1]$, ν_1 be uniform on $[-3,3]$, and ν_2 be uniform on $[-2,2]$. Then $\chi_{\vec{\mu}}$ is the uniform probability 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 situation, i.e., $\chi_{\vec{\mu}} \preceq_c \chi_{\vec{\nu}}$, is infeasible in general.

Nevertheless, it is plausible that a VMOT π may still couple the marginals $\vec{\mu}$ monotonically,[4] i.e., a VMOT π sets $\pi^X = \chi_{\vec{\mu}}$ thereby maximizing $\mathbb{E}_{\pi^X}[c_1(X)]$, then seek π^Y which satisfies $\chi_{\vec{\mu}} \preceq_c \pi^Y$ while π^Y is as close as the ideal $\chi_{\vec{\nu}}$, so that π^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 1.

[14] showed that the conjecture is correct when the cost function is of the quadratic form (2.7) and the marginals μ_i, ν_i satisfy a special condition known as linear increment of marginals, which is the case if e.g. μ_i, ν_i are gaussians with increasing variance. In fact, [14] showed that under these conditions, π^X of a VMOT π is distributed on a straight line in $\mathbb{R}^d$. This implies that the assumption imposed on the cost and marginal is very restrictive and the conclusion cannot be extended to general marginals; see [31, Example 4.5] for a related discussion. In this paper, we will prove that the conjecture is indeed correct when $d = 2$ (without any particular condition imposed on the marginals), but incorrect in general when $d \geq 3$. 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 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 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 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.

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

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

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 the second moments of μ_1,μ_2 are finite. Then:*
i) There exists a VMOT π such that its first time 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 time marginal π^X is the monotone coupling of μ_1,μ_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 π 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

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

On the other hand, for any VMOT $\pi = \pi_x \otimes \pi^X$,[5] 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 $\int \beta^{**}(y)d\pi_x(y) \ge \beta^{**}(x)$, since β^{**} 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

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

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

(3.11) and (3.12) implies that for any VMOT π, its first time 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 the fact that β is submodular and 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.

[5] Any $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ can be written as $\pi = \pi_x \otimes \pi^X$, where $\pi_x \in \mathcal{P}(\mathbb{R}^d)$ is called a kernel (or disintegration) of π. When π represents the joint distribution of the $\mathbb{R}^d$-valued random variables X and Y, denoted as $\pi = \mathrm{Law}(X, Y)$, then π_x represents the conditional distribution of Y given $X = x$, that is, $\pi_x(B) = \mathcal{P}(Y \in B \,|\, X = x)$. The disintegration allows us to iteratively integrate as $\int f(x, y)d\pi(x, y) = \iint f(x, y)d\pi_x(y)d\pi^X(x)$.

To prove part i), fix $\delta > 0$, and choose a VMOT π for the cost $c_\delta(x,y) = c_1(x) + \delta x_1 x_2 + c_2(y)$.[6] 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 time marginal π^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 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). $\square$

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 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. Assume the same condition as in Theorem 2.1, and that the second moments of μ_1, μ_2 are finite. Then:*
i) There exists a VMOT π such that its first time 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 time marginal π^X is the anti-monotone coupling of μ_1, μ_2.

Remark 3.7. *A financial interpretation of these results may be given as follows. For a two-period cap[7] 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. Furthermore, 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 marginal π^X of a VMOT π is not necessarily monotone, as explained in the following.

Example 3.8. *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*

[6]Only here is the finite second moment assumption used to ensure that $x_1 x_2$ is integrable.
[7]Option payoffs of the form $c(x,y) = c_1(x) + c_2(y)$ are sometimes referred to as a two-period cap, where $x = (x_1, ..., x_d)$ denote d-assets price at the first maturity t_1 and $y = (y_1, ..., y_d)$ are the corresponding assets price at the second maturity t_2.

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*

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

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*

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

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 time marginal is* π^X *and its kernel is* $\{\pi_z, \pi_w\}$*. Now take* $c(x, y) = y_1 y_2$ *(so that* $c_1(x) = 0$*),* $\phi_1 = \phi_2 = h_1 = h_2 = 0$*, and notice that* $\{\phi_i, \psi_i, h_i\}_{i=1,2}$ *and* π *then jointly satisfy the optimality condition* (2.9), (2.10)*. 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.*

As previously shown, 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 if part ii) can be extended for $d \geq 3$ as part i). But unfortunately, this is not the case.

Remark 3.9. *There exists a supermodular function* $\beta : \mathbb{R}^3 \to \mathbb{R}$ *for which* β^* *is not submodular. To see this, let* $\beta(x) = \frac{1}{2}x \cdot Ax$ *with a symmetric positive-definite* $d \times d$ *matrix* A*. Then* $\beta^*(z) = \frac{1}{2}z \cdot A^{-1}z$*. Now if* $d \geq 3$*, it is easy to find such an* A *with all positive entries and that* A^{-1} *also has some positive off-diagonal entries. This implies that* β *is supermodular and* $\frac{\partial^2 \beta^*}{\partial x_i \partial x_j} > 0$ *for some* $i \neq j$*. But the latter implies that* β^* *is not submodular.*

Despite Remark 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. In the following, $\operatorname{conv}(A)$ denotes the convex hull of a set A in a vector space.

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:*

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

It is clear 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*

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

One can further check that $\beta_0 = \beta_0^{**}$ *on* $\mathcal{Y}$*, and*

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

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* 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*

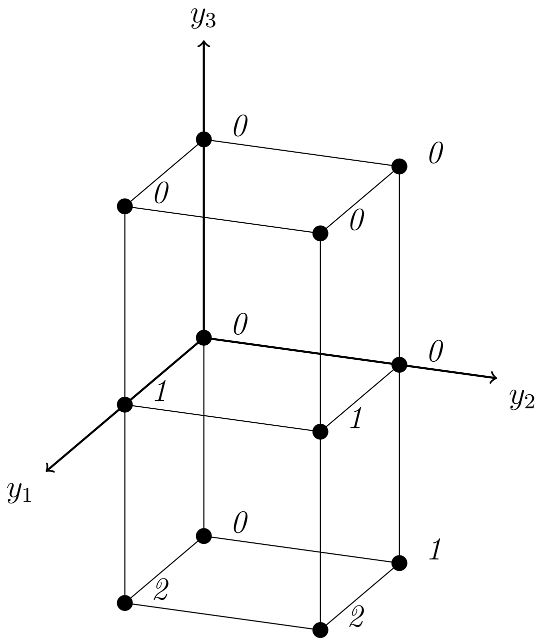


FIGURE 2. β_0 values on the vertices.

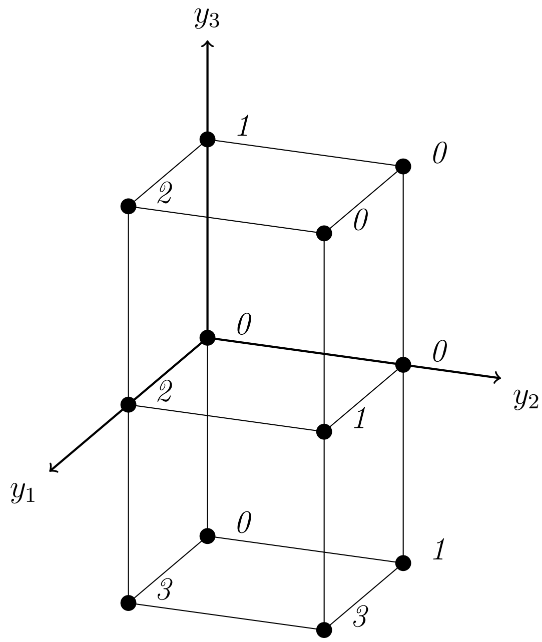


FIGURE 3. β values on the vertices.

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

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

where the second equality is because $\frac{L_1+L_2}{2}$ is affine and $\{\bar{u}, \bar{u}'\}$ is the monotone rearrangement of $\{u, u'\}$, and the strict inequality is because $\frac{L_1+L_2}{2}(\bar{u}) < L_1(\bar{u})$ and $\frac{L_1+L_2}{2}(\bar{u}') < L_2(\bar{u}')$. This yields that β_0^{**} is not submodular.

The failure of Proposition 3.4 iii) for $d \geq 3$ reduces the plausibility of Conjecture 1. Nonetheless, we continue to suspect that the conjecture may still be true 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 Example 3.10 eventually allow us to construct a 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* π *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* π^X *fails to be the monotone coupling of* $\vec{\mu} = (\mu_1, \mu_2, \mu_3)$.

Proof. Let $\mathcal{Y}$ be the set of twelve points in $\mathbb{R}^3$ and β_0 be the submodular function as in Example 3.10. Define $\psi_1, \psi_2, \psi_3 : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ by

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

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

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

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 $\{u, u'\} \subseteq \mathrm{conv}(\mathcal{Z})$, while none of $\bar{u}, \bar{u}'$ lies on the plane containing $\mathcal{Z}$.

We now construct a vectorial martingale transport π. For this, take $\pi^X := \frac{1}{2}(\delta_u + \delta_{u'})$ as the first time 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 $\mathrm{conv}(\mathcal{Z})$. Now define $\pi = \pi_x \otimes \pi^X$ (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^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 π 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 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 3.10. 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

$$h(x) = \begin{cases} \nabla L_1 = (0,0,0) \ \text{ if } \ L_1(x) > L_2(x), \\ \nabla L_2 = (1,1,-1) \ \text{ if } \ L_1(x) < L_2(x), \\ \nabla \frac{L_1+L_2}{2} = (\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}) \ \text{ if } \ L_1(x) = L_2(x). \end{cases} \tag{3.14}$$

In order to prove the optimality of π and (ϕ, ψ, 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 π 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 γ^X is monotone, that is, $\gamma^X = \frac{1}{2}(\delta_{\bar{u}} + \delta_{\bar{u}'})$. We claim that γ cannot be optimal. If γ were optimal, it must satisfy the optimality condition (2.10) 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

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

This shows γ cannot satisfy (2.10), hence cannot be a VMOT for (2.3). $\square$

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

$$\mathbb{E}_\pi[c(Y)] > \max_{\gamma \in \mathrm{VMT}(\vec{\mu},\vec{\nu})} \left\{ \mathbb{E}_\gamma[c(Y)] \;\middle|\; \gamma^X \text{ is monotone; } \gamma^X = \frac{\delta_{\bar{u}} + \delta_{\bar{u}'}}{2} \right\} \tag{3.15}$$

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 (3.15) *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^X is still not monotone for every VMOT π, as long as λ is not too large. On the other hand, since*

$$c(u) + c(u') < c(\bar{u}) + c(\bar{u}'), \tag{3.16}$$

inequality (3.15) *is reversed for $c = c_\lambda$ with sufficiently large λ, in which case every VMOT π 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 λ, a VMOT π promotes its second time marginal π^Y to be supported on the monotone set $\mathcal{Z}$, even if this necessitates supporting its first marginal π^X 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^X to be supported on the monotone set $\{\bar{u}, \bar{u}'\}$ even if this requires π^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 ψ_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 [20]. Recall* (2.9)*, which we rewrite in this case as*

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

Name	Payoff function
European basket call option	$(\sum_{i=1}^{d} a_i X_i - K)^+$
European basket put option	$(K - \sum_{i=1}^{d} a_i X_i)^+$
Put on the maximum among d stocks	$(K - \max_{1\le i\le d}\{X_i\})^+$
Call on the minimum among d stocks	$(\min_{1\le i\le d}\{X_i\} - K)^+$
Covariance among d stocks	$\sum_{i,j} a_{ij}X_iX_j + b_{ij}Y_iY_j$

TABLE 1. A list of supermodular derivative payouts ($a, b \ge 0$).

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

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

By definition, we have $\psi_1 \ge \tilde{\psi}_1$. 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 $(\phi_i, \tilde{\psi}_i, h_i)_i$ continues to be a dual optimizer.*

Remark 3.14. *[14] 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 $c = c(y) = \sum_{1\le i<j\le d} y_iy_j$ 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.*

Table 1 concludes this section by presenting several supermodular payoff functions for exotic options, some of which can also be found 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.

4. Numerical implementation via neural network

4.1. **A hybrid version of the VMOT Problem.** Theorem 3.5 motivates us to consider the following variant problem:

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

where

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

This formulation assumes that we know the joint distribution of $X = (X_1, ..., X_d)$ which represents asset prices at the first maturity, but only the individual distributions of Y_i which represents asset prices at the second maturity. This version of the VMOT problem appears relevant in practice, because in some situations we may have more information about the joint distribution, or fairly confidently model it, at a time in the near future (as being monotone, for instance), whereas the dependence structure in the far future is much less certain. One might therefore consider hybrid pricing problems, between the model dependent and model independent settings, where the dependence structure at the first time is assumed to be known but other dependency (between different asset prices at the second time, or the same asset price at the first and second time) is not. Problem (4.1) captures exactly this scenario.

Now a 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 the above problem (4.1) in which $\chi \in \mathcal{P}(\mathbb{R}^d)$ is the monotone coupling of the $\{\mu_i\}_{i\in[d]}$, i.e., $\chi = \chi_{\vec{\mu}}$. We recall that χ can 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. Likewise, we can also replace each ν_i with $G_{i\,\#}^{-1}\mathcal{L}_{[0,1]}$ where G_i^{-1} is the inverse cumulative of ν_i. With this, observe that when $\chi = \chi_{\vec{\mu}}$, which we will assume from now on, we can rewrite (4.1) as

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

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

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

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 $\tilde{c}(u,v) := c(\Phi(u), \Psi(v))$.[8] We note that $\text{VMT}(\chi, \vec{\nu}) = (\Phi, \Psi)_{\#}\text{CVMT}(\Phi, \Psi)$; elements in $\text{VMT}(\chi, \vec{\nu})$ are the push-forward of elements in $\text{CVMT}(\Phi, \Psi)$ by the map $(\Phi, \Psi) : (0,1) \times (0,1)^d \to \mathbb{R}^d \times \mathbb{R}^d$. 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 $U \sim \mathcal{L}_{[0,1]}$.

Now the dual problem to (4.3) is formulated as

$$\inf_{(\tilde{\phi}, \tilde{\psi}_i, \tilde{h}_i) \in \tilde{\Xi}} \int_0^1 \tilde{\phi}(u) du + \sum_{i=1}^d \int_0^1 \tilde{\psi}_i(v_i) dv_i \tag{4.5}$$

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 $\tilde{\phi}, \tilde{\psi}_i \in L^1(\mathcal{L}_{[0,1]})$, $\tilde{h}_i$ is bounded for every $i \in [d]$, and the following inequality holds for every $(u,v) \in (0,1) \times (0,1)^d$:

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

As with the primal problem, this dual is much simpler than the original (2.8), 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 $u \in (0,1)$ rather than $x \in \mathbb{R}^d$. 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.)

4.2. **Numerical method.** We offer a numerical method which is based on neural networks with penalization for calculating the optimal value, optimal joint distribution, and the corresponding dual optimizers. The trained neural networks can be viewed as approximate functions describing an optimal semi-static trading strategy that super/sub-replicates the given payout function.

[8]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.

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 [21]. Another strategy incorporates entropic regularization [12] 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.

In recent years, the application of neural networks to solve optimal transport problems has gained attention [19, 29, 36]. This neural network-based approach has also been extended to the context of martingale transport [14, 15, 24]. One notable advantage of employing neural networks is their ability to effectively handle the curse of dimensionality through powerful approximation techniques. In particular, [14, 24] has applied neural network-based numerical martingale transport methods in financial framework.

We apply the framework of [15], which was further developed in [14], and examine potential performance enhancements enabled by our solution, which takes advantage of the monotone geometry of the optimal couplings through the formulation (4.5). More specifically, we optimize the following *regularized super-hedging functional*:

$$\inf_{\varphi\in\mathcal{H}} \int \varphi \, d\pi_0 + \int b_\gamma(c-\varphi)\, d\theta \tag{4.7}$$

where π_0 is any fixed element in $\mathcal{Q} := \mathrm{VMT}(\vec{\mu},\vec{\nu})$, $\mathcal{X} := \mathbb{R}^d \times \mathbb{R}^d$, and

$$\mathcal{H} = \{\varphi \in C_b(\mathcal{X}) \,|\, \varphi(x,y) = \sum_{i=1}^d \big(\phi_i(x_i) + \psi_i(y_i) + h_i(x)(y_i - x_i)\big)\}$$

where $\phi_i, \psi_i \in C_b(\mathbb{R})$ and $h_i \in C_b(\mathbb{R}^d)$ are continuous and bounded functions. Note that $\int \varphi d\pi_0 = \sum_i \int \phi_i d\mu_i + \int \psi_i d\nu_i$ as π_0 is a martingale measure. The second term in (4.7) is designed to penalize the violation of the inequality $c \le \varphi$ via the penalty function $b_\gamma(t) := \frac{1}{\gamma} b(\gamma t)$, where $\gamma > 0$ is a parameter and $b : \mathbb{R} \to \mathbb{R}_+$ is a differentiable non-decreasing convex function satisfying $\lim_{t\to\infty} b(t)/t = \infty$. Finally, $\theta \in \mathcal{P}(\mathcal{X})$ is referred to as a *reference/sampling measure*, which is used to sample points at which the inequality constraint $c \le \varphi$ can be tested. We note that the problem (4.7) is dual to the following *regularized optimal transport problem*:

$$\sup_{\mu\in\mathcal{Q}} \int c\, d\mu - \frac{1}{\gamma}\int \beta^*(\frac{d\mu}{d\theta})\, d\theta \tag{4.8}$$

where β^* is the convex conjugate of β; see [15] for more details.

The novelty of our strategy lies in applying the dimensional reduction described earlier to the regularized implementation. More specifically, in view of our reformulation (4.5), we let $\tilde{\mathcal{Q}} := \mathrm{CVMT}(\Phi, \Psi)$, and $\tilde{\mathcal{H}}$ be the class of functions of the form:

$$\tilde{\varphi}(u,v) = \tilde{\phi}(u) + \sum_{i=1}^{d} \big(\tilde{\psi}_i(v_i) + \tilde{h}_i(u)(G_i^{-1}(v_i) - F_i^{-1}(u))\big) \tag{4.9}$$

where $(\tilde{\phi}, \tilde{\psi}_i, \tilde{h}_i) \in \tilde{\Xi}$. By the martingale condition, for any $\tilde{\pi}_0 \in \tilde{\mathcal{Q}}$, we have $\int (\sum_{i=1}^d \tilde{h}_i(u)(G_i^{-1}(v_i) - F_i^{-1}(u))d\tilde{\pi}_0 = 0$, thus $\int \tilde{\varphi} d\tilde{\pi}_0 = \int_0^1 \tilde{\phi} du + \sum_i \int_0^1 \tilde{\psi}_i dv_i$. This yields that the objective/loss function in (4.7) can be written as

$$\begin{aligned} \mathrm{Loss} &= \int \varphi\, d\pi_0 + \int b_\gamma(c-\varphi)\, d\theta \\ &= \int_0^1 \tilde{\phi}(u)du + \sum_{i=1}^d \int_0^1 \tilde{\psi}_i dv_i + \int b_\gamma(\tilde{c} - \tilde{\varphi})d\tilde{\theta} \end{aligned}$$

where $\tilde{\theta}$ is the independent coupling of the $(d+1)$-copy of $\mathcal{L}_{[0,1]}$ (rather than $2d$-copy of $\mathcal{L}_{[0,1]}$). This results in the desired dimension reduction. With this, the regularized dual problem (4.7) is reformulated as

$$\inf_{\tilde{\varphi}\in\tilde{\mathcal{H}}} \int_0^1 \tilde{\phi}(u)du + \sum_{i=1}^d \int_0^1 \tilde{\psi}_i dv_i + \int b_\gamma(\tilde{c} - \tilde{\varphi})d\tilde{\theta}. \tag{4.10}$$

An appendix details our implementation, including penalization and neural network parameters and the computational technology employed.

4.3. **Examples.** We develop several examples to demonstrate the improvement enabled by our numerical method. We solve the VMOT problem for the cost function 2.7, which is related to finding model-free bounds to the variance of a portfolio given the marginal distributions of underlying assets. The first example employs normal marginal distributions, while the second employs real-world marginal distributions implied by option prices for individual assets at two distinct future points in time.

4.3.1. *Normal marginals.* We compute (4.10) (which is equivalent to (4.7)) with the cost (2.7) assuming that the marginals are normal distributions centered at zero. It is known that our result on the monotone support of π^X of a VMOT π extends to general dimension d in specific cases such as when the marginals are normally distributed; [14, Theorem 5.3] shows that π^X is in fact normally distributed on a straight line in $\mathbb{R}^d$ when the marginals are normal. Motivated by this, and to further illustrate the positive effects of dimension reduction, besides the case $d = 2$, we add examples with $d = 3$, 4 and 5. We note that higher-dimensional cases give opportunity to a greater relative simplification to the problem; as remarked earlier, using the dimension of the sample domain as a measure of computational complexity, it is proportional to $2d$ in the full dimension case and to $d+1$ in the reduced dimension.

For each value of d, we randomly generate the vector of portfolio weights $w = (w_1, ..., w_d)$. The marginal distributions are defined as

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

where σ_i and ρ_i are randomly generated in the intervals $[1, 2]$ and $[2, 3]$ respectively. The fact $\sigma_i < \rho_i$ guarantees that the marginals are in convex order for each i. The resulting coefficients and parameters are available in the code (see appendix for details), to allow the examples to be reproduced. The exact optimal value of (2.3) is as follows; see an appendix for a proof.

Proposition 4.1. *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*

$$\max_{\pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})} \mathbb{E}_\pi[c(X, Y)] = \sum_{1 \leq i < j \leq d} \big((a_{ij} + b_{ij}) \sigma_i \sigma_j + b_{ij} \lambda_i \lambda_j \big). \tag{4.11}$$

In view of the cost (2.7), we set $a_{ij} = 0$ and $b_{ij} = w_i w_j$. Figure 4 shows the convergence of the dual value (4.10) over the training epochs in the two formulations, namely, the full-dimension version 2.3 and our reduced dimension version 4.3, for each value of d. The true optimal value (4.11) 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 accuracy. However, we observe that this degradation is much less pronounced in the reduced dimension case. The latter also demands less computational resources, notably memory. Table 2 compares the true and the mean numeric values for all cases. For each d and each version of the problem, we report the mean and the standard deviation of the numerical output of the 100 last epochs.

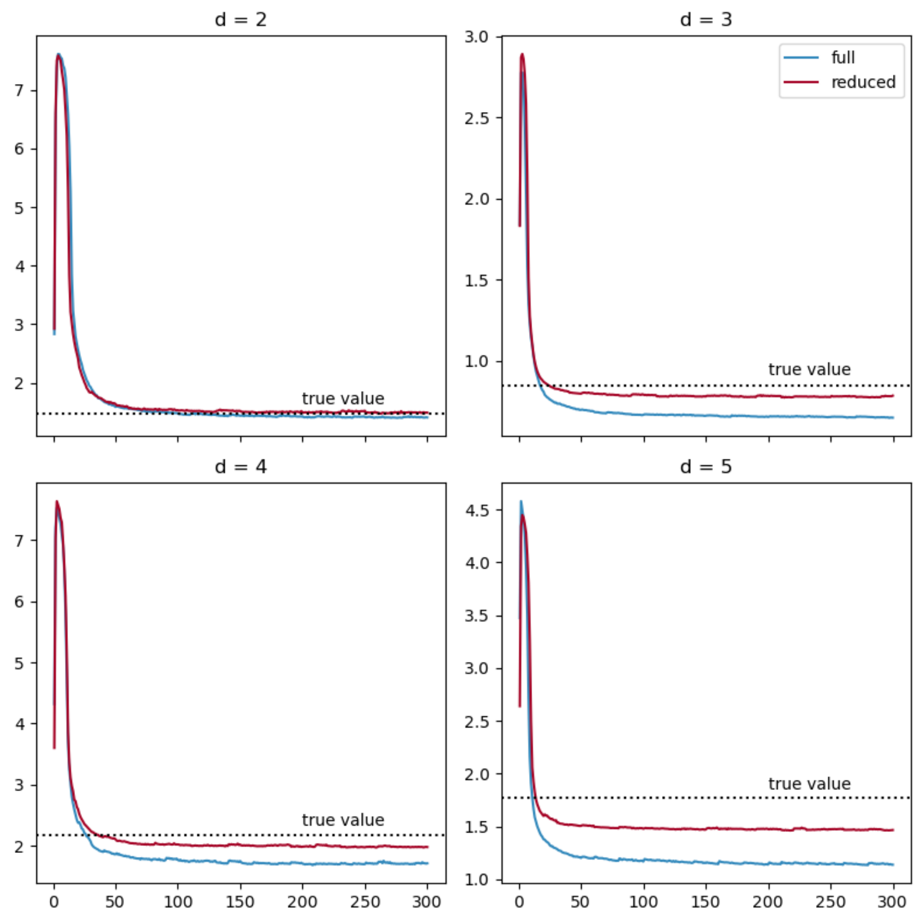


FIGURE 4. Dual value convergence, full vs. reduced dimension – normal marginals, $d = 2$ to 5.

Figure 5 shows the distribution of π^X of a VMOT π in the quantile domain inferred by the numerical implementation of the potential functions in both versions for $d = 2$ based on [15, Theorem 2.2]. The left side shows the full dimension case, where we observe an accumulation of the distribution of π^X

Method	d	Exact solution	Numerical value	(Std.)
Full	2	1.4729	1.4135	(0.0075)
	3	0.8486	0.6536	(0.0025)
	4	2.1673	1.7105	(0.0111)
	5	1.7716	1.1472	(0.0057)
Reduced	2	1.4729	1.4958	(0.0115)
	3	0.8486	0.7784	(0.0028)
	4	2.1673	1.9891	(0.0109)
	5	1.7716	1.4709	(0.0054)

TABLE 2. Comparison between exact and numerical values.

around the main diagonal, consistent with our theoretical result. The right side shows the reduced dimension case, where π^X is restricted to the diagonal.

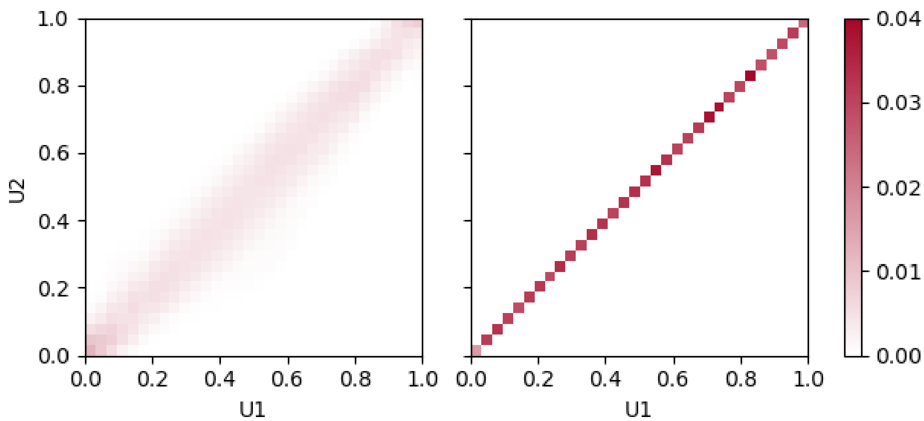


FIGURE 5. Heat map of π^X for a VMOT π on the $U = (U_1, U_2)$ space, full (left) and reduced dimension (right).

4.3.2. *Empirical marginals.* We move on to a real world problem where we calculate upper and lower bounds for $\mathbb{E}_\pi[c]$ with c in 2.7 considering a portfolio composed half of Apple and half of Amazon shares. As of Dec. 16$^{\text{th}}$, 2022, we use the call and put option closing prices to compute the market-implied, risk-neutral marginal distributions of prices on Jan. 20$^{\text{th}}$ and Feb. 17$^{\text{th}}$, 2023. Details about the calculation of the marginals are given in an appendix. We consider the returns on both future dates with respect to the closing prices of Dec. 16^{th}, 2022, that is,

$$\begin{aligned} X_i &= \left(\text{price}_i(\text{Jan } 20^{\text{th}}) - \text{price}_i(\text{Dec } 16^{\text{th}})\right)/\text{price}_i(\text{Dec } 16^{\text{th}}) \\ Y_i &= \left(\text{price}_i(\text{Feb } 17^{\text{th}}) - \text{price}_i(\text{Dec } 16^{\text{th}})\right)/\text{price}_i(\text{Dec } 16^{\text{th}}) \\ i &= 1\,(\text{Amazon}),\ 2\,(\text{Apple}). \end{aligned}$$

The resulting marginal distributions are shown in Figure 6.

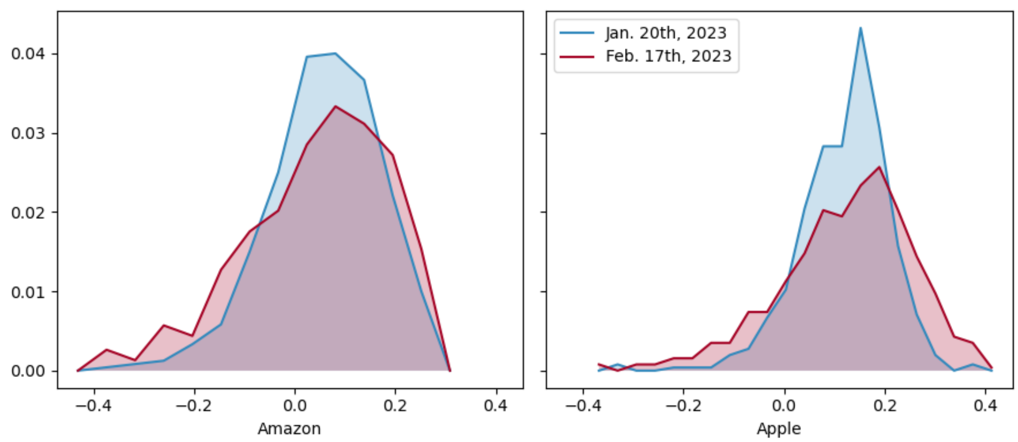


FIGURE 6. Marginal distributions of future returns implied by the call and put option prices as of Dec. 16th, 2022 (US$).

Finally, we compare the upper and lower bounds of VMOT values to OT. First, we can easily compute upper and lower OT value bounds using the positive and negative monotone couplings of the marginal distributions of Y_1, Y_2. Second, we apply the method described in this section to calculate the VMOT bounds. Figure 7 shows the convergence of the dual value for both in full and reduced dimension cases, with the OT bounds shown as grey lines. We also plot the (independent coupling) sample mean cost $\mathbb{E}_\theta[(\frac{Y_1+Y_2}{2})^2] = \frac{1}{2}\mathbb{E}_{\nu_1}[Y_1]\mathbb{E}_{\nu_2}[Y_2] + \frac{1}{4}\mathbb{E}_{\nu_1}[Y_1^2] + \frac{1}{4}\mathbb{E}_{\nu_2}[Y_2^2]$ as a reference. The following table summarizes the outputs. In both cases, the standard deviation of the sample means is negligible.

We note that the reduced dimension method generates slightly wider bounds. This is consistent with the intuition that the method that (correctly) considers monotone first time marginals is closer to the extremes in comparison

Method	Upper bound	(Std.)	Lower bound	(Std.)
Full	0.0288	(0.0009)	0.0131	(0.0003)
Reduced	0.0301	(0.0008)	0.0125	(0.0004)
MOT	0.0308		0.0114	
Sample mean	0.0207			

TABLE 3. Comparison of numeric values between VMOT and OT.

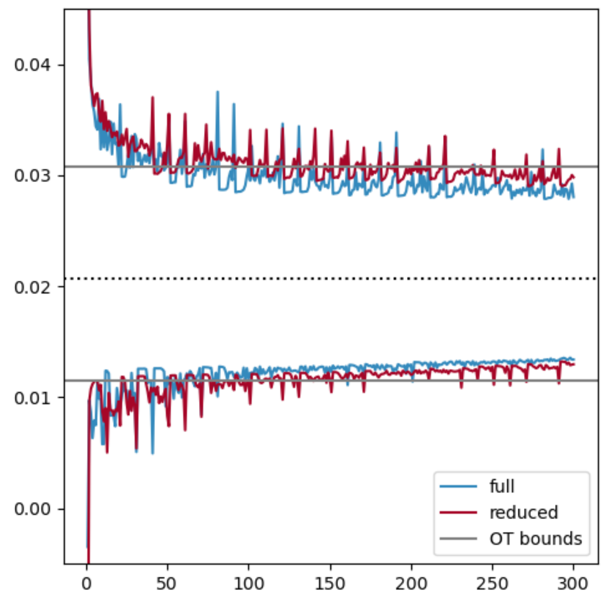


FIGURE 7. Dual value convergence, full vs. reduced dimension – empirical marginals, $d = 2$.

with the one whose first time marginals only approximate the monotone disposition. Although accuracy cannot be compared without knowing the true solution, from a risk management perspective the reduced dimension bounds can be considered more conservative. We also observe that that the bounds from both methods are located within the OT bounds, as expected.

5. Conclusion

In this paper we presented a geometrical result that can be viewed as the martingale version of classical results in optimal transport, namely that a solution to the VMOT problem with a sub or supermodular cost function has the monotone first time marginal if $d = 2$. We then presented examples that negate the result for higher dimension, at least when no additional condition is imposed on the cost function and marginals. We provided examples of robust pricing problems that can be described as the VMOT problem with sub/supermodular option payouts. We demonstrated that our result on the geometry of VMOT allows for dimensional reduction in the dual formulation of the problem, and used both synthetic and real data to calculate numerical solutions to the portfolio variance cost function on the dual side with and without the benefit of dimensional reduction. A comparison of the outputs in the synthetic case demonstrated how dimensional reduction improves numerical precision. Finally, our main result and the counterexample prompt us to consider whether further structure on the cost function and marginals can be imposed to allow the result to be extended to higher dimensions.

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.

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 that 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. 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 $y = (y_1, ..., y_d)$, $\hat{y} = (\hat{y}_1, ..., \hat{y}_d)$, $\bar{y} = (\bar{y}_1, ..., \bar{y}_d)$, $\hat{y}^c = (\hat{y}_1^c, ..., \hat{y}_d^c)$ be 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 β_R:

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

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

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

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 $\beta_R \searrow \beta$ 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. $\square$

Proof of Proposition 4.1. [14, Theorem 5.3] showed by duality argument that under the assumption of Proposition 4.1, there is a VMOT $\pi = \mathrm{Law}(X, Y)$ under which the martingale (X, Y) is given by

$$X = (X_1, ..., X_d) \text{ with } X_i = \sigma_i U \text{ where } U \sim N(0, 1), \tag{6.6}$$

and the conditional law of $Y = (Y_1, ..., Y_d)$ given X is as follows:

$$\begin{aligned} &\text{Given } X, \quad Y_i \sim N(X_i, \lambda_i^2), \text{ and} \\ &Y_j - X_j = \frac{\lambda_j}{\lambda_i}(Y_i - X_i), \quad 1 \le i < j \le d. \end{aligned} \tag{6.7}$$

Notice the marginal condition is satisfied with this martingale. (6.6) yields

$$\mathbb{E}_\pi[X_i X_j] = \sigma_i \sigma_j \mathbb{E}_\pi[U^2] = \sigma_i \sigma_j, \quad i < j. \tag{6.8}$$

On the other hand, for each $i < j$, (6.7) yields

$$\begin{aligned} \mathbb{E}_\pi[Y_i Y_j | X] &= \mathbb{E}_\pi[Y_i\big(\frac{\lambda_j}{\lambda_i}(Y_i - X_i) + X_j\big)|X] \\ &= \frac{\lambda_j}{\lambda_i}\mathbb{E}_\pi[Y_i^2|X] - \frac{\lambda_j}{\lambda_i}X_i\mathbb{E}_\pi[Y_i|X] + X_j\mathbb{E}_\pi[Y_i|X] \\ &= \frac{\lambda_j}{\lambda_i}(\lambda_i^2 + X_i^2) - \frac{\lambda_j}{\lambda_i}X_i^2 + X_i X_j \\ &= \lambda_i \lambda_j + X_i X_j. \end{aligned}$$

Hence, $\mathbb{E}_\pi[Y_i Y_j] = \mathbb{E}[\mathbb{E}_\pi[Y_i Y_j | X]] = \lambda_i \lambda_j + \sigma_i \sigma_j$, completing the proof. □

Numerical methods – computational details.

In our implementation, 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 drawn from the reference measure θ. As a standard procedure, we run the neural network optimization, or "training," for a certain number of epochs until an acceptable level of convergence is reached. We used 30 random samples of 1 million points each for each example and version, and ran 10 training epochs with each sample, for a total of 300 training epochs, which proved sufficient for all examples.

We employ Python and the Pytorch neural network package with the standard Adam gradient descent optimizer.[9] The convergence to the true optimal value is guaranteed by Proposition 2.4 and Remark 3.5 in [15].

A typical (and our) choice of θ is $\theta := \mu_1 \otimes \cdots \otimes \mu_d \otimes \nu_1 \otimes \cdots \otimes \nu_d$, i.e., the independent coupling of the marginals. Our choice of b is $b(t) = \frac{1}{2}(t^+)^2$, so that $b(t) = \frac{1}{2\gamma}((\gamma t)^+)^2$. In our tests, the choice of γ affected the convergence pattern significantly; we finally fixed $\gamma := 1000$ in the examples with normal marginals and $\gamma := 100000$ in the examples with empirical marginals.

Construction of the empirical marginal distributions. 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 of the underlying asset. [6] noted that 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:

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

This is also valid if we replace the curve of call price C by that of put price P. Due to the no arbitrage condition, both call and put price curves are convex as a function of K, meaning the second derivative exists almost everywhere.

In practice, however, we only see call prices at a limited range of strike prices, restricting our ability to calculate the risk-neutral density directly. Instead, we use second-order finite differences at observable strike prices to approximate it and linearly interpolate between different strike prices. Because out-of-the-money options have lesser activity and may be mispriced, we calculate the density function using put options with strike prices lower than the spot price and call options with strike prices higher than the spot price. In addition, near the extremes of the call and put price curves where prices approach zero, we exclude some option prices that clearly violate no arbitrage due to illiquidity. We may also normalize the resulting function to ensure that it is a probability density, i.e., its integral equals one, if needed. Through these steps, we obtain an empirical risk-neutral density function.

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

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

[11] 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.

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] Wilfrid Gangbo and Andrzej Święch. 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.

[19] 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.

[20] 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.

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

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

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

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

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

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

[27] 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.

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

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

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

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

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

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

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

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

[36] 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.

Joshua Zoen-Git Hiew: Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton AB Canada
Email address: joshuazo@ualberta.ca

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

Brendan Pass: Department of Mathematical and Statistical Sciences
University of Alberta, Edmonton AB Canada
Email address: pass@ualberta.ca

Marcelo Cruz de Souza: Mitchell E. Daniels, Jr. School of Business
Purdue University, West Lafayette, Indiana 47907, USA
Email address: souzam@purdue.edu