

# DIMENSION REDUCTION IN MARTINGALE OPTIMAL TRANSPORT: GEOMETRY AND ROBUST OPTION PRICING \*

JOSHUA ZOEN-GIT HIEW<sup>†</sup>, TONGSEOK LIM<sup>‡</sup>, BRENDAN PASS<sup>§</sup>, AND MARCELO  
CRUZ DE SOUZA<sup>¶</sup>

**Abstract.** This paper addresses the problem of robust option pricing within the framework of Vectorial Martingale Optimal Transport (VMOT). We investigate the geometry of VMOT solutions for  $N$ -period market models and demonstrate that, when the number of underlying assets is  $d = 2$  and the payoff is sub- or supermodular, the extremal model reduces to a single-factor structure in the first period. This structural result allows for a significant dimension reduction, transforming the problem into a more tractable format. We prove that this reduction is specific to the two-asset case and provide counterexamples showing it generally fails for  $d \geq 3$ . Finally, we exploit this monotonicity to develop a reduced-dimension Sinkhorn algorithm. Numerical experiments demonstrate that this structure-preserving approach reduces computational time by approximately 99% compared to standard methods while improving accuracy.

**Key words.** Robust finance, Hedging, Martingale, Optimal transport, Duality, Monotonicity, Sinkhorn Algorithm, Dimension Reduction, Infinite-dimensional linear programming

**AMS subject classifications.** 90Bxx, 90Cxx, 60Dxx, 60Gxx, 49Mxx, 65Kxx

**1. Introduction.** We denote  $[n] := \{1, 2, \dots, n\}$  for  $n \in \mathbb{N}$ , and let  $\mathcal{P}(\Omega)$  denote the set of all probability measures (distributions) over a set  $\Omega$ . Fix  $d, N \in \mathbb{N}$ . Let  $\vec{\mu} = (\vec{\mu}_1, \dots, \vec{\mu}_N)$  such that  $\vec{\mu}_t = (\mu_{t,1}, \dots, \mu_{t,d})$  for  $t \in [N]$ , denoting vectors of probability measures (called marginals) on  $\mathbb{R}$ . Throughout the paper, we assume that all distributions have finite first moments.

We consider the following space of *Vectorial Martingale Transports* (VMT) from  $\vec{\mu}_t$  to  $\vec{\mu}_{t+1}$ ,  $t = 1, \dots, N - 1$  (see [26, 27]):

$$\text{VMT}(\vec{\mu}) := \{ \pi \in \mathcal{P}(\mathbb{R}^{Nd}) \mid \pi = \text{Law}(X), X = (X_1, \dots, X_N), \\ \mathbb{E}_\pi[X_{t+1} | X_1, \dots, X_t] = X_t, \text{Law}(X_{t,i}) = \mu_{t,i}, \text{ for all } t \in [N], i \in [d] \},$$

where  $X = (X_1, \dots, X_N) \in \mathbb{R}^{Nd}$  are random vectors.

For a distribution  $\pi \in \mathcal{P}(\mathbb{R}^{Nd})$ , we denote by  $\pi^t \in \mathcal{P}(\mathbb{R}^d)$  its  $t^{\text{th}}$  period marginal, i.e., if  $\pi = \text{Law}(X)$ , then  $\pi^t = \text{Law}(X_t)$ . We also denote by  $\Pi(\vec{\mu}_t)$  the set of couplings of the marginals  $\mu_{t,i}$ , i.e.,

$$\Pi(\vec{\mu}_t) := \{ \sigma \in \mathcal{P}(\mathbb{R}^d) \mid \sigma = \text{Law}(X_t), \text{Law}(X_{t,i}) = \mu_{t,i}, \forall i \in [d] \}.$$

Clearly, if  $\pi \in \text{VMT}(\vec{\mu})$ , then  $\pi^t \in \Pi(\vec{\mu}_t)$ .

---

\*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.

<sup>†</sup>Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton AB Canada ([joshuazo@ualberta.ca](mailto:joshuazo@ualberta.ca)).

<sup>‡</sup>Mitchell E. Daniels, Jr. School of Business, Purdue University, West Lafayette, Indiana 47907, USA ([lim336@purdue.edu](mailto:lim336@purdue.edu)).

<sup>§</sup>Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton AB Canada ([pass@ualberta.ca](mailto:pass@ualberta.ca)).

<sup>¶</sup>Financial System Monitoring Department, Central Bank of Brazil, Fortaleza, CE, Brazil ([marcelo.souza@bcb.gov.br](mailto:marcelo.souza@bcb.gov.br)).

In the case  $N = 2$ , we denote the marginals as  $\vec{\mu} = (\mu_1, \dots, \mu_d)$  and  $\vec{\nu} = (\nu_1, \dots, \nu_d)$ , where  $\mu_i = \text{Law}(X_i)$  and  $\nu_i = \text{Law}(Y_i)$  for  $i \in [d]$ .

Let  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  be two probability measures. We say they are in *convex order*, denoted by  $\mu \preceq_c \nu$ , if and only if:

$$\int f d\mu \leq \int f d\nu \quad \text{for every convex function } f.$$

It is known that the set  $\text{VMT}(\vec{\mu})$  is nonempty if and only if every pair of marginals  $\mu_{t,i}, \mu_{t+1,i}$  is in *convex order* [23]. Thus, we will always assume  $\mu_{t,i} \preceq_c \mu_{t+1,i}$  for all  $t \in [N-1]$  and  $i \in [d]$ .

Let a measurable function  $c : \mathbb{R}^{Nd} \rightarrow \mathbb{R}$  denote the cost or option payoff function. We define the vectorial martingale optimal transport (VMOT) problem as:

$$(1.1) \quad \text{maximize } \mathbb{E}_\pi[c(X)] \quad \text{over } \pi \in \text{VMT}(\vec{\mu}).$$

A solution  $\pi$  to (1.1) will be called a vectorial martingale optimal transport, or VMOT.

To ensure that the problem (1.1) is well-defined, we make the following assumptions throughout the paper. When considering a VMOT problem with a cost function  $c$ , we assume that the marginals satisfy the following condition: there exist continuous functions  $v_{t,i} \in L^1(\mu_{t,i})$ ,  $t \in [N]$  and  $i \in [d]$ , such that:

$$|c(x_1, \dots, x_N)| \leq \sum_{t=1}^N \sum_{i=1}^d v_{t,i}(x_{t,i}).$$

Note that this ensures:

$$|\mathbb{E}_\pi[c(X)]| \leq \sum_t \sum_i \mathbb{E}_{\mu_{t,i}}[v_{t,i}(X_{t,i})] < \infty \quad \text{for any } \pi \in \text{VMT}(\vec{\mu}).$$

It is standard to show that this in turn implies that the problem (1.1) admits an optimizer whenever  $c$  is upper-semicontinuous [33].

The conditions of the VMOT problem are inspired by [2, 11, 13, 17, 21, 20], which originate from a closely related problem in mathematical finance: the martingale optimal transport (MOT) problem. Instead of considering the coupling of a vector of martingales from given one-dimensional marginals, the MOT problem considers the coupling of a single martingale with a given marginal distribution, which may be multi-dimensional.

From a financial perspective, each sequence of random variables  $(X_{1,i}, \dots, X_{N,i})$  represents an asset price process at  $N$  future times  $0 < t_1 < \dots < t_N$ . Assuming a zero interest rate, each martingale measure  $\pi \in \text{VMT}(\vec{\mu})$  represents the risk-neutral probability under which  $(X_1, \dots, X_N) \in \mathbb{R}^{Nd}$  is an  $\mathbb{R}^d$ -valued  $N$ -period martingale. [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. The cost function  $c(x_{1,1}, \dots, x_{1,d}, \dots, x_{N,1}, \dots, x_{N,d})$  represents an option whose payoff is path-dependent on the asset prices  $(X_1, \dots, X_N)$ .

The key difference between VMOT and classical multi-dimensional MOT lies in the treatment of joint distributions among different assets at each period. While the joint distribution between different assets (dimensions) is given in multi-dimensional MOT, VMOT does not assume a predetermined joint distribution. As a result, there

is no guarantee that a coupling of the marginals at one period satisfies the convex order condition relative to any feasible coupling at the next period, and this must be enforced separately.

Additionally, VMOT is more suited for modeling real-world financial markets. In classical multi-dimensional martingale optimal transport, properly recovering the joint distribution requires observing a set of rainbow options [32]. However, rainbow options are typically traded in the over-the-counter (OTC) market, making them harder to observe and prone to mispricing due to their lower liquidity. On the other hand, VMOT relies only on European vanilla options, which are actively traded, making them easier to observe and less prone to mispricing. This makes the VMOT framework more aligned with real-world market dynamics and more robust to potential market frictions.

With this in mind, we consider the set of all possible laws  $\text{VMT}(\vec{\mu})$  given the marginal information  $\vec{\mu}$ . The max/min value in (1.1) can then be interpreted as the upper/lower arbitrage-free price bound for the option  $c$  among all risk-neutral measures consistent with the market data. We define (1.1) as a maximization problem, though it can also be framed as a minimization problem by simply changing  $c$  to  $-c$ .

Beyond its financial relevance, VMOT exhibits rich geometric properties similar to classical OT. We extend the observations in [12, 26] and find that for the class of supermodular payoff functions, when there are two underlying assets ( $d = 2$ ), the maximum arbitrage-free price is attained when the joint distribution of the underlying assets at the first maturity time is co-monotone. This finding suggests a single-factor market structure for extreme pricing scenarios. Moreover, the monotonicity structure in VMOT helps reduce computational complexity: since the dependence structure of assets at the first time is known explicitly, only the dependence structure at subsequent times, as well as the martingale coupling structure across time periods, need to be computed.

This paper is organized as follows. Section 2 formally defines the VMOT problem and presents our main results, along with a conjecture for extending the monotonicity structure to higher dimensions. Section 3 defines the dual formulation of the VMOT problem and provides the proof of the main theorem. Section 4 presents a counterexample to the conjecture in higher dimensions, along with other related examples in the VMOT problem. We provide a counterexample to show that this dimension reduction is sharp and does not extend to  $d \geq 3$ . Finally, Section 5 demonstrates how our main result helps reduce computational complexity for the VMOT problem. The Appendix provides miscellaneous proofs.

**2. Main Theorem and Conjecture.** Before we state the main theorem, we introduce the concept of supermodularity and irreducibility, which will be crucial in our analysis.

**DEFINITION 2.1.** For  $a, b \in \mathbb{R}^d$ , we define  $a \vee b$  as the componentwise maximum of  $a, b$  and define  $a \wedge b$  as 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  $\beta : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  be a function. Then submodularity and supermodularity of  $\beta$  are defined, for all  $a, b \in \mathbb{R}^d$ , as:

$$\begin{aligned} \beta(a) + \beta(b) &\geq \beta(a \vee b) + \beta(a \wedge b), \\ \beta(a) + \beta(b) &\leq \beta(a \vee b) + \beta(a \wedge b), \end{aligned}$$

respectively. A function is called strictly submodular or strictly 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 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 2.2.

- i)  $\{a \vee b, a \wedge b\}$  is called a 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  $\mu$  is supported on  $A$ , i.e.,  $\mu(A) = 1$ .
- iv) Given a vector of probability measures  $\vec{\mu} = (\mu_1, \dots, \mu_d)$  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  $\mu_1, \dots, \mu_d$  as its marginals (i.e.,  $\chi_{\vec{\mu}} \in \Pi(\vec{\mu})$ ), is called the monotone coupling of  $\vec{\mu}$ .

Note that  $\chi_{\vec{\mu}} = (F_{\mu_1}^{-1}, F_{\mu_2}^{-1}, \dots, F_{\mu_d}^{-1})_{\#} \mathcal{L}_{[0,1]}$ ,<sup>1</sup> where each  $F_{\mu_i}^{-1}$  denotes the generalized inverse of the cumulative distribution function of  $\mu_i$  (i.e., the quantile function), defined by  $F^{-1}(y) = \inf\{x \in \mathbb{R} : F(x) \geq y\}$  for  $y \in [0, 1]$ , and  $\mathcal{L}_{[0,1]}$  denotes the uniform probability measure on  $[0, 1]$ .

We now discuss *irreducibility* between two probability measures. A pair of measures  $(\mu, \nu)$  on  $\mathbb{R}$  in convex order is said to be *irreducible* if the set  $I := \{x \in \mathbb{R} \mid \int |x - y| d(\nu - \mu)(y) > 0\}$  is connected and contains the full mass of  $\mu$ . We say that a sequence of measures  $(\mu_t)_{t \in [N]}$  is *irreducible* if each consecutive pair  $(\mu_t, \mu_{t+1})$  is irreducible.

We note that the irreducibility condition imposed on the sequence of marginals holds for many pairs of probability distributions  $\mu_t \preceq_c \mu_{t+1}$  on the real line in convex order. Moreover, if a pair of measures is not irreducible, it can be perturbed in an arbitrarily small way to make the pair irreducible. For further details on irreducibility, we refer the reader to [5].

**2.1. Main Theorem.** We first present the main theorem of this paper, which establishes the existence of a VMOT with specific properties for the first-period marginals.

THEOREM 2.3. *Let  $d = 2$  and  $N \geq 2$ . Consider the cost function  $c(x_1, \dots, x_N) = \sum_{t=1}^N c_t(x_{t,1}, x_{t,2})$ , where each  $c_t$  is a supermodular function. Assume the following conditions:*

- i) *For each  $i = 1, 2$ ,  $(\mu_{t,i})_{t \in [N]}$  is an irreducible sequence of marginals on  $\mathbb{R}$ .*
- ii) *The cost function  $c : \mathbb{R}^{2N} \rightarrow \mathbb{R}$  is upper-semicontinuous, and there exist continuous functions  $v_{t,i} \in L^1(\mu_{t,i})$  such that  $|c(x)| \leq \sum_{t=1}^N \sum_{i=1}^2 v_{t,i}(x_{t,i})$  for all  $x \in \mathbb{R}^{2N}$ .*
- iii) *The second moments of  $\mu_{1,1}$  and  $\mu_{1,2}$  are finite.*

Then:

- i) *There exists a VMOT  $\pi$  such that its first period marginal  $\pi^1$  is the monotone coupling of  $\mu_{1,1}$  and  $\mu_{1,2}$ .*
- ii) *If  $c_1$  is strictly supermodular, then every VMOT  $\pi$  satisfies that its first period marginal  $\pi^1$  is the monotone coupling of  $\mu_{1,1}$  and  $\mu_{1,2}$ .*

Using the proof of Theorem 2.3 and Proposition 3.3 iii) in Section 3.3, we also obtain the following mirror statement. A set  $A \subseteq \mathbb{R}^2$  is called *anti-monotone* if the

<sup>1</sup>Given a measurable map  $F : \mathcal{X} \rightarrow \mathcal{Y}$  and a measure  $\mu$  on  $\mathcal{X}$ , the push-forward of  $\mu$  by  $F$ , denoted by  $F_{\#}\mu$ , is a measure on  $\mathcal{Y}$  satisfying  $F_{\#}\mu(A) = \mu(F^{-1}(A))$  for every  $A \subseteq \mathcal{Y}$ .

set  $\{x = (x_1, x_2) \in \mathbb{R}^2 \mid (-x_1, x_2) \in A\}$  is monotone. A measure  $\mu \in \mathcal{P}(\mathbb{R}^2)$  is called anti-monotone if  $\mu$  is supported on an anti-monotone set.

**COROLLARY 2.4.** *Let  $d = 2, N \geq 2$  and  $c(x_1, \dots, x_N) = \sum_{t=1}^N c_t(x_{t,1}, x_{t,2})$ , where each  $c_t$  is submodular. Assume the same conditions as in Theorem 2.3. Then:*

- i) *There exists a VMOT  $\pi$  such that its first period marginal  $\pi^1$  is the anti-monotone coupling of  $\mu_1$  and  $\mu_2$ .*
- ii) *If  $c_1$  is strictly submodular, then every VMOT  $\pi$  satisfies that its first period marginal  $\pi^1$  is the anti-monotone coupling of  $\mu_{1,1}$  and  $\mu_{1,2}$ .*

**Remark 2.5.** The (anti-)monotonicity of the first period implies that there exists a random variable  $Z$  such that  $X_{1,1} = f_1(Z)$  and  $X_{1,2} = f_2(Z)$ , where  $f_1, f_2$  are deterministic functions. A financial interpretation may be given as follows: for an  $N$ -period cap<sup>2</sup> whose payoff depends on two underlying assets, Theorem 2.3 and Corollary 2.4 describe the existence of an extreme market model that attains the extremal of the price bounds. Specifically, this market model implies that the two assets at the first period  $(X_{1,1}, X_{1,2})$  are controlled by a single factor  $Z$ . Furthermore, if  $c_1$  is strictly submodular or supermodular, then every extremal market model exhibits this property.

**2.2. The Conjecture.** Theorem 2.3 establishes that VMOT exhibits a monotone structure for  $d = 2$  and  $N \geq 2$  when the cost function is supermodular. This result highlights the critical role of supermodular costs in ensuring monotone structural properties within VMOT, a concept analogous to its well-established role in classical multi-marginal optimal transport problems.

This raises a natural question: can these structural properties extend to VMOT in higher dimensions, specifically for  $d \geq 3$ ? We propose the following conjecture:

**CONJECTURE 2.6.** *Let  $d \geq 2, N \geq 2$ , and assume the following conditions:*

- i) *The sequence of marginals  $(\mu_{t,i})_{t \in [N]}$  is irreducible for each  $i \in [d]$ , and in convex ordering, i.e.,  $\mu_{t,i} \preceq_c \mu_{t+1,i}$  for all  $t \in [N-1]$  and  $i \in [d]$ .*
- ii) *The cost function  $c(x_1, \dots, x_N) = \sum_{t=1}^N c_t(x_{t,1}, \dots, x_{t,d})$  is upper semicontinuous with supermodular  $c_t$  for all  $t \in [N]$ , and  $c$  is absolutely bounded by the sum of continuous functions  $v_{t,i}(x_{t,i}) \in L^1(\mu_{t,i})$  for each  $t \in [N]$  and  $i \in [d]$ .*
- iii) *The second moments of  $\mu_{1,i}$  are finite for all  $i \in [d]$ .*

*Then:*

- i) *There exists a VMOT  $\pi$  for the problem (1.1) such that its first period marginal  $\pi^1$  is the monotone coupling of  $\mu_1 = (\mu_{1,1}, \dots, \mu_{1,d})$ .*
- ii) *If  $c_1$  is strictly supermodular, then every VMOT  $\pi$  satisfies that its first period marginal  $\pi^1 = \chi_{\mu_1}$ , where  $\chi_{\mu_1}$  is the monotone coupling of  $\mu_1$ .*

This conjecture is based on the intuition that supermodularity may induce monotonicity in the first period. However, the martingale constraints in subsequent periods reduce the likelihood that monotonicity will be maintained across all periods. Theorem 2.3 confirms the validity of the conjecture for two assets with an arbitrary number of periods  $N$ .

**2.3. Motivation and Heuristic.** To understand both the main theorem and the conjecture, we draw motivation from the classical multi-dimensional optimal transport setting. In this framework, the monotonicity of the optimal measure is guaranteed for supermodular costs, regardless of the dimension  $d$ :

<sup>2</sup>Option payoffs of the form  $c(x_1, \dots, x_N) = \sum_{t=1}^N c_t(x_t)$  are sometimes referred to as an  $N$ -period cap, where  $x_t = (x_{t,1}, \dots, x_{t,d})$  denotes the prices of  $d$  assets at the  $t^{\text{th}}$  maturity.

**THEOREM 2.7.** [9, 28] *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  $\chi_{\vec{\mu}}$  is the unique maximizer if  $c$  is strictly supermodular.*

The monotonicity of optimizers in the classical optimal transport problem for supermodular costs is well-known [9, 28]. Results asserting higher-dimensional deterministic solutions, such as those by Brenier [7] (for two marginals) and Gangbo-Świąch [15] (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 [25, 14, 8] (for two marginals), [16, 30, 24, 31] (for several marginals), can be seen as higher-dimensional analogues of this monotonicity. Our conjecture can be interpreted as a vectorial martingale transport version of these results. Indeed, note that if  $N = 2$  and each  $\mu_i$  is a Dirac mass (corresponding to the case where the first time is the present), the VMOT problem reduces to the classical (multi-marginal) optimal transport problem on the  $\nu_i$ 's.

**Heuristic.** Given marginals  $\vec{\mu} = (\vec{\mu}_1, \dots, \vec{\mu}_N)$  such that  $\vec{\mu}_t = (\mu_{t,1}, \dots, \mu_{t,d})$  for  $t \in [N]$  and a cost function  $c(x_1, \dots, x_N) = \sum_{t=1}^N c_t(x_t)$  where each  $c_t$  is a supermodular function, in light of Theorem 2.7, the ideal situation would be that for the marginals  $\pi^t$  of each period of  $\pi \in \text{VMT}(\vec{\mu})$ , we have the monotone coupling  $\pi^t = \chi_{\vec{\mu}_t}$ , such that  $\mathbb{E}_{\pi}[c(X)] = \sum_{t=1}^N \mathbb{E}_{\chi_{\vec{\mu}_t}}[c_t(X)]$ . However, the martingale constraint imposed on  $\pi$  reduces the possibility of such an ideal situation.

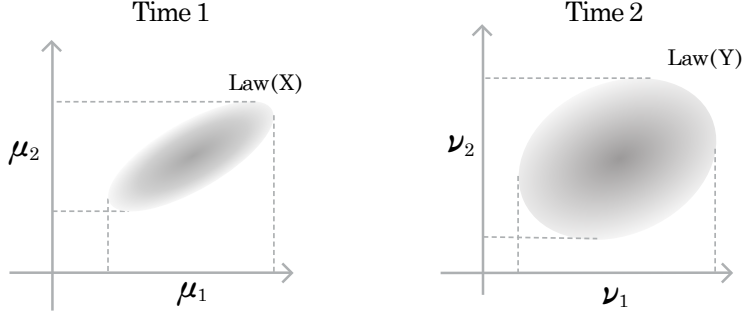
Even if  $\mu_{t,i} \preceq_c \mu_{t+1,i}$  for all  $i \in [d]$ , the monotone couplings  $\chi_{\vec{\mu}_t}$  and  $\chi_{\vec{\mu}_{t+1}}$  may not satisfy the convex order in general. As an example, consider the following case:

*Example 2.8.* Let  $d = 2$ ,  $\mu_1 = \mu_2$  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 the uniform probability on  $l_1 = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = x_2, x_1 \in [-1, 1]\}$ , and  $\chi_{\vec{\nu}}$  is uniform on  $l_2 = \{(y_1, y_2) \in \mathbb{R}^2 \mid y_2 = \frac{2}{3}y_1, y_1 \in [-3, 3]\}$ . Then  $\chi_{\vec{\mu}}$  and  $\chi_{\vec{\nu}}$  cannot be in convex order because it is necessary for  $l_1$  to be contained in the convex hull of  $l_2$  to satisfy  $\chi_{\vec{\mu}} \preceq_c \chi_{\vec{\nu}}$ .

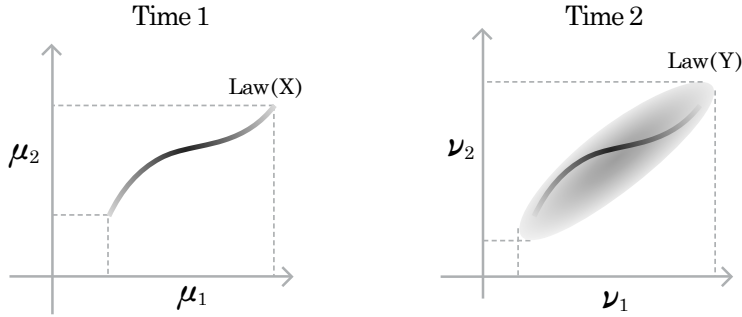
The intuition behind this is that the convex order condition in each of the marginals is not sufficient to guarantee that the convex hull of the monotone set induced by the second-period marginals contains the first-period monotone set. This demonstrates that the ideal situation is infeasible in general. Nevertheless, it is plausible that a VMOT  $\pi$  can still couple the marginals  $\vec{\mu}_1$  monotonically.<sup>3</sup> This leads us to our heuristic behind the conjecture, as illustrated in Figure 1.

*Remark 2.9.* [12] showed that the conjecture holds for  $N = 2$  and general  $d$  when the cost function is of the quadratic form (2.3), and the marginals  $\mu_i, \nu_i$  satisfy a special condition known as the linear increment of marginals. This is the case when, for example,  $\mu_i, \nu_i$  are Gaussians with increasing variance. In fact, [12] showed that under these conditions,  $\pi^1$  of a VMOT  $\pi$  is distributed on a straight line in  $\mathbb{R}^d$ . This implies that the assumptions on the cost and marginals are very restrictive, and the conclusion cannot be extended to general marginals. See [26] for a related discussion. In this paper, we will prove that the conjecture is correct when  $d = 2$  (without any particular condition on the marginals) but incorrect in general when  $d \geq 3$ . It is unfortunate that we cannot find a general sufficient condition for the marginals  $\mu_{t,i}$  other than the condition given by [12] to ensure the conjecture holds for general  $d$ .

<sup>3</sup>For any  $\gamma \in \Pi(\vec{\mu})$ , there exists  $\pi \in \text{VMT}(\vec{\mu}, \vec{\nu})$  such that  $\pi^1 = \gamma$ . One can simply take the disintegration kernels  $\pi^{1,i}$  of some  $\pi_i \in \Pi^M(\mu_i, \nu_i)$ , where  $\Pi^M(\mu_i, \nu_i)$  denotes the set of all martingale couplings, and set  $\pi = \gamma(dx_1, \dots, dx_d) \otimes \pi^{1,1}(dy_1) \otimes \dots \otimes \pi^{1,d}(dy_d)$ . Similar arguments can be made for general  $N$ .



(a) Generic marginals of a two-period vectorial martingale transport.



(b) A martingale transport with a monotone first-time marginal.

FIG. 1. Conjecture 2.6 asserts that the martingale transport in Figure 1b can be superior to the one in Figure 1a for the problem (1.1) due to the supermodularity of  $c_1, c_2$ . We overlap  $\text{Law}(X)$  with  $\text{Law}(Y)$  in Figure 1b to emphasize that they must be in convex order.

**2.4. Examples of Supermodular Costs.** To illustrate the importance of supermodular costs in finance, we consider the following cost function:

$$(2.1) \quad c(x, y) = \sum_{1 \leq i, j \leq d} (a_{ij}x_i x_j + b_{ij}x_i y_j + c_{ij}y_i y_j),$$

which represents the assets' mutual covariances at two future times. For any  $\pi \in \text{VMT}(\bar{\mu}, \bar{\nu})$ ,  $\mathbb{E}_\pi[X_i Y_j] = \mathbb{E}_\pi[\mathbb{E}_\pi[X_i Y_j | X]] = \mathbb{E}_\pi[X_i \mathbb{E}_\pi[Y_j | X]] = \mathbb{E}_\pi[X_i X_j]$  by the martingale constraint, while  $\mathbb{E}_\pi[X_i^2] = \int_{\mathbb{R}} x^2 d\mu_i(x)$ ,  $\mathbb{E}_\pi[Y_j^2] = \int_{\mathbb{R}} y^2 d\nu_j(y)$  are fixed by the marginal constraints. Hence, we can reduce the cost (2.1) to the following form:

$$(2.2) \quad c(x, y) = \sum_{1 \leq i < j \leq d} (a_{ij}x_i x_j + b_{ij}y_i y_j).$$

We assume  $a_{ij} \geq 0$ ,  $b_{ij} \geq 0$  to ensure that the cost function is supermodular.

In particular, for  $d = 2$ , this becomes

$$c(x, y) = ax_1 x_2 + by_1 y_2,$$

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  $\pi$ .



To motivate the study of the cost function of the form (2.2), consider a portfolio consisting of assets with prices  $Y_1, \dots, Y_d$  at the terminal maturity  $t_2$ , with weights  $w_1, \dots, w_d > 0$ , so that the value of the portfolio at maturity is  $\sum_{i=1}^d w_i Y_i$ . The variance is a commonly used measure of the portfolio's risk:

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

In the VMOT framework, the term  $\left( \sum_{i=1}^d w_i \mathbb{E}[Y_i] \right)^2$  does not depend on  $\pi$ , since the individual distributions are assumed to be known. Maximizing the variance is thus equivalent to maximizing  $\mathbb{E}_{\pi^Y} \left[ \left( \sum_i w_i Y_i \right)^2 \right] = \int_{\mathbb{R}^d} \left( \sum_i w_i y_i \right)^2 d\pi^Y(y)$ , which is a classical (multi-marginal) optimal transport (OT) problem with cost function:

$$(2.3) \quad c(x, y) = c(y) = \left( \sum_{i=1}^d w_i y_i \right)^2.$$

Notice that the optimization problems with  $c$  as in (2.2) and  $c$  as in (2.3) are equivalent, with each  $a_{ij} = 0$  and  $b_{ij} = w_i w_j$  in the VMOT problem, again due to the given marginal information.

Similar to OT, the VMOT problem reflects a worst-case risk scenario in which, in addition to the distributions  $\nu_i$  at some future time  $t_2$ , the distributions  $\mu_i$  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  $t_2$ , incorporating the extra information from the knowledge of distributions  $\mu_i$  at  $t_1$ . This leads to the VMOT problem with cost (2.3). As a result, the VMOT approach can yield tighter bounds on the portfolio's variance than the classical OT approach, even though the cost functions depend only on the terminal prices  $Y$ . The reason is that only couplings  $\pi^2$  of the marginals  $\nu_i$  which dominate some coupling  $\pi^1$  of the marginals  $\mu_i$  in convex order can arise as feasible couplings in VMOT, whereas all couplings in  $\Pi(\vec{\nu})$  are allowed as candidate optimizers in the classical OT.

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 the 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 [22]. The VMOT approach, which derives marginal distributions from option prices, resonates with these findings.

Table 1 concludes this section by presenting several supermodular payoff functions for exotic options, some of which are also found in [1]. The  $N$ -period cap with the payoff  $c(x_1, \dots, x_N) = \sum_{t=1}^N c_t(x_t)$ , where  $c_t$  is one of the supermodular functions listed in Table 1, written on two underlying assets, fits into the form of the cost function applicable to our Theorem 2.3.

### 3. Dual Problem, Dual Attainment and Proof of the Main Theorem.

**3.1. Dual Problem.** The VMOT problem is a type of infinite-dimensional linear programming, which admits a corresponding *dual problem*. We introduce the



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	$(K - \max_{1 \leq i \leq d} \{X_i\})^+$
Call on the minimum among $d$ stocks	$(\min_{1 \leq i \leq d} \{X_i\} - K)^+$
Covariance among $d$ stocks	$\sum_{i,j} a_{ij} X_i X_j$

TABLE 1

A list of supermodular derivative payouts with nonnegative coefficients  $a_{ij} \geq 0$ .

space of functions

$$\mathcal{H} = \left\{ \varphi(x) := \sum_{t=1}^N \sum_{i=1}^d u_{t,i}(x_{t,i}) + \sum_{t=1}^{N-1} \sum_{i=1}^d h_{t,i}(x_1, \dots, x_t)(x_{t+1,i} - x_{t,i}) \right. \\ \left. \left| \begin{array}{l} u_{t,i} \in L^1(\mu_{t,i}), \quad t \in [N], i \in [d]; \quad h_{t,i} \in L^\infty(\mathbb{R}^{td}), \quad t \in [N-1], i \in [d] \end{array} \right. \right\}.$$

The space  $\mathcal{H}$  consists of the Lagrange multipliers associated with the marginal and martingale constraints in the dual problem.

We then define the set  $\Xi$  as the subset of  $\mathcal{H}$  satisfying the pointwise inequality:

$$(3.1) \quad \Xi := \{ \varphi \in \mathcal{H} \mid \varphi(x) \geq c(x) \text{ for all } x \in \mathbb{R}^{Nd} \}.$$

For any martingale measure  $\pi \in \text{VMT}(\vec{\mu})$  and  $\varphi \in \mathcal{H}$ , it holds that  $\int \varphi d\pi = \sum_{t,i} \int u_{t,i} d\mu_{t,i}$ . This leads to the dual problem associated with (1.1) as:

$$(3.2) \quad \inf_{\varphi \in \Xi} \sum_{t=1}^N \sum_{i=1}^d \int u_{t,i} d\mu_{t,i}.$$

In this formulation, (1.1) can be referred to as the *primal problem*. In the special case of  $N = 2$ , we denote  $\Xi$  as the set of functions of the form  $(u_{1,i}, u_{2,i}, h_{1,i})_{i=1}^d = (u_i, v_i, h_i)_{i=1}^d$ .

The dual problem has a concrete financial interpretation. Specifically, the inequality in (3.1) represents a semi-static super-replicating portfolio for the path-dependent payoff  $c$ . Here,  $u_{t,i}(x_{t,i})$  corresponds to the European options written on each individual underlying asset, with maturities at time  $t$ , and  $h_{t,i}$  represents the position held in the  $i^{\text{th}}$  asset for delta-hedging between times  $t$  and  $t+1$ .

If the problem (1.1) is a minimization problem, then the dual problem corresponds to an optimal subhedging problem, where the inequality in (3.1) is reversed.

The value in (3.2) represents the minimum cost required to construct a superhedging portfolio. Thus, the goal of the dual problem is to identify the optimal (least-cost) superhedging portfolio.

**3.2. Dual Attainment.** It has been established that the dual problem (3.2) generally does not admit a solution within the class  $\Xi$ , even in the simple case  $d = 1$  and  $N = 2$ , i.e., when the option  $c$  depends on a single asset (see [3, 5]). This issue arises unless certain regularity conditions are imposed on the payoff function  $c$  [4]. Consequently, a generalized notion of *dual attainment*, which refers to the solvability of the dual problem, has been introduced for multi-period problems. This was studied

in [29] for  $N \geq 2, d = 1$ , and in [27] for  $N \geq 2, d \geq 2$ . More specifically, [27] provides the following result on dual attainment.

**THEOREM 3.1 ([27]).** *Let  $(\mu_{t,i})_{t \in [N]}$  be an irreducible sequence of marginals on  $\mathbb{R}$  for each  $i \in [d]$ . Let  $c : \mathbb{R}^{Nd} \rightarrow \mathbb{R}$  be an upper-semicontinuous cost such that  $|c(x)| \leq \sum_{t=1}^N \sum_{i=1}^d v_{t,i}(x_{t,i})$  for some continuous functions  $v_{t,i} \in L^1(\mu_{t,i})$ . Then there exists a dual minimizer, consisting of two sets of functions  $(u_{t,i})_{t \in [N], i \in [d]}$  and  $(h_{t,i})_{t \in [N-1], i \in [d]}$ , that satisfies the following pathwise constraint:*

$$(3.3) \quad \sum_{t=1}^N \sum_{i=1}^d u_{t,i}(x_{t,i}) + \sum_{t=1}^{N-1} \sum_{i=1}^d h_{t,i}(x_1, \dots, x_t)(x_{t+1,i} - x_{t,i}) = c(x) \quad \pi - a.e. x$$

for every VMOT  $\pi$  that solves the primal problem (1.1).

**3.3. Proof of the Main Theorem.** Before proceeding with the proof of the main theorem, we first present some useful properties of supermodular functions.

**DEFINITION 3.2.** *For a proper function  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ , its convex conjugate  $f^*$  is defined as the following convex lower semi-continuous function:*

$$(3.4) \quad f^*(y) = \sup_{x \in \mathbb{R}^d} (x \cdot y - f(x)), \quad y \in \mathbb{R}^d.$$

It is well-known that  $f^{**} = (f^*)^*$  is the largest convex lower semi-continuous function satisfying  $f^{**} \leq f$ . We refer to  $f^{**}$  as the *convex biconjugate* of  $f$ .

**PROPOSITION 3.3.**

- i) *If  $\beta$  is a submodular function on  $\mathbb{R}^d$ , then its convex conjugate  $\beta^*$  is supermodular.*
- ii) *If  $\beta$  is supermodular on  $\mathbb{R}^2$ , then its convex conjugate  $\beta^*$  is submodular.*
- iii) *If  $\beta$  is submodular or supermodular on  $\mathbb{R}^2$ , then its convex biconjugate  $\beta^{**}$  is also submodular or supermodular, respectively.*

The proof of Proposition 3.3 is provided in the appendix.

With the dual attainment theorem and the proposition about the convex biconjugate of submodular and supermodular functions, we now proceed with the proof of Theorem 2.3.

*of Theorem 2.3.* Theorem 3.1 guarantees the existence of an optimal dual solution  $(u_{t,i})_{t \in [N], i \in [d]}$  and  $(h_{t,i})_{t \in [N-1], i \in [d]}$  that satisfies:

$$\sum_{t=1}^N \sum_{i=1}^d u_{t,i}(x_{t,i}) + \sum_{t=1}^{N-1} \sum_{i=1}^d h_{t,i}(x_1, \dots, x_t)(x_{t+1,i} - x_{t,i}) \geq \sum_{t=1}^N c_t(x_t),$$

for all  $x \in \mathbb{R}^{Nd}$ , with equality  $\pi$ -almost every  $x$  for any VMOT  $\pi$ .

Rearranging terms, we obtain a linear function of  $x_N$  on the right-hand side:

$$\sum_{i=1}^d u_{N,i}(x_{N,i}) - c_N(x_N) \geq \sum_{t=1}^{N-1} \left( c_t(x_t) - \sum_{i=1}^d u_{t,i}(x_{t,i}) \right) - \sum_{t=1}^{N-1} \sum_{i=1}^d h_{t,i}(x_1, \dots, x_t)(x_{t+1,i} - x_{t,i}),$$

for all  $x \in \mathbb{R}^{Nd}$ , with equality  $\pi$ -almost every  $x$ . Let

$$\beta_N(x_N) := \sum_{i=1}^d u_{N,i}(x_{N,i}) - c_N(x_N).$$

Since the convex biconjugate of a function is the supremum of all affine functions below it, we have

$$(3.5) \quad \beta_N^{**}(x_N) \geq \sum_{t=1}^{N-1} \left( c_t(x_t) - \sum_{i=1}^d u_{t,i}(x_{t,i}) \right) - \sum_{t=1}^{N-1} \sum_{i=1}^d h_{t,i}(x_1, \dots, x_t) (x_{t+1,i} - x_{t,i}),$$

for all  $x \in \mathbb{R}^{Nd}$ , with equality  $\pi$ -almost every  $x$ . In particular, by letting  $x_N = x_{N-1}$ , we obtain

$$\beta_N^{**}(x_{N-1}) \geq \sum_{t=1}^{N-1} \left( c_t(x_t) - \sum_{i=1}^d u_{t,i}(x_{t,i}) \right) - \sum_{t=1}^{N-2} \sum_{i=1}^d h_{t,i}(x_1, \dots, x_t) (x_{t+1,i} - x_{t,i}),$$

for all  $(x_1, \dots, x_{N-1}) \in \mathbb{R}^{(N-1)d}$ .

Next, disintegrating any VMOT  $\pi$  with respect to  $x_1, \dots, x_{N-1}$ , thereby representing

$\pi(dx_1, \dots, dx_N) = \kappa(x_1, \dots, x_{N-1}, dx_N) \otimes \pi^{1, \dots, N-1}(dx_1, \dots, dx_{N-1})$ , we integrate (3.5) with respect to the martingale kernel  $\kappa(x_1, \dots, x_{N-1}, dx_N)$ , and obtain

$$\begin{aligned} \sum_{t=1}^{N-1} \left( c_t(x_t) - \sum_{i=1}^d u_{t,i}(x_{t,i}) \right) - \sum_{t=1}^{N-2} \sum_{i=1}^d h_{t,i}(x_1, \dots, x_t) (x_{t+1,i} - x_{t,i}) \\ = \int \beta_N^{**}(x_N) \kappa(x_1, \dots, x_{N-1}, dx_N) \\ \geq \beta_N^{**}(x_{N-1}) \quad \pi^{1, \dots, N-1}\text{-a.e. } (x_1, \dots, x_{N-1}), \end{aligned}$$

since  $\int h_{N-1,i}(x_1, \dots, x_{N-1}) (x_{N,i} - x_{N-1,i}) \kappa(x_1, \dots, x_{N-1}, dx_N) = 0$  by the martingale property, while the last inequality is from Jensen's inequality for the convex function  $\beta_N^{**}$ . We conclude:

$$\beta_N^{**}(x_{N-1}) \geq \sum_{t=1}^{N-1} \left( c_t(x_t) - \sum_{i=1}^d u_{t,i}(x_{t,i}) \right) - \sum_{t=1}^{N-2} \sum_{i=1}^d h_{t,i}(x_1, \dots, x_t) (x_{t+1,i} - x_{t,i})$$

for all  $(x_1, \dots, x_{N-1}) \in \mathbb{R}^{(N-1)d}$ , with equality  $\pi^{1, \dots, N-1}$ -almost every  $(x_1, \dots, x_{N-1})$ .

By inductively defining the following function

$$\beta_t(x_t) := \beta_{t+1}^{**}(x_t) + \sum_{i=1}^d u_{t,i}(x_{t,i}) - c_t(x_t),$$

for  $t = N-1, N-2, \dots, 2$  and continuing this process, we deduce

$$\beta_2^{**}(x_1) - c_1(x_1) \geq - \sum_{i=1}^d u_{1,i}(x_{1,i}),$$

for all  $x_1 \in \mathbb{R}^d$ , with equality  $\pi^1$ -almost every  $x_1$ .

This implies that  $\pi^1$  solves the optimal transport problem with cost  $\beta_2^{**}(x_1) - c_1(x_1)$  and marginals  $\mu_{11}, \dots, \mu_{1d}$ .

As each  $c_t$  is supermodular,  $\beta_N$  is submodular. Assuming  $d = 2$  and  $c_1$  is strictly supermodular, by Proposition 3.3 iii), each  $\beta_t^{**}$  is submodular, making  $\beta_2^{**}(x_1) - c_1(x_1)$  strictly submodular. By Theorem 2.7,  $\pi^1 = \chi_{(\mu_{11}, \mu_{12})}$  must be the monotone coupling of  $\mu_{11}, \mu_{12}$ , proving part ii).

To prove part i), fix  $\delta > 0$  and choose a VMOT  $\pi$  for the perturbed cost  $c_\delta(x_1, \dots, x_N) = c_1(x_1) + \delta x_{11}x_{12} + \sum_{t=2}^N c_N(x_N)$ .<sup>4</sup> Note that the mixed derivative of  $\delta x_1x_2$  is  $\delta$ , implying that  $c_1(x) + \delta x_1x_2$  is strictly supermodular. Let  $\pi$  be a VMOT for the cost  $c_\delta$  for a fixed  $\delta > 0$ . By part ii),  $\pi^1 = \chi_{(\mu_{11}, \mu_{12})}$  is the monotone coupling of  $\mu_{11}, \mu_{12}$ . Let  $\pi^2, \dots, \pi^N \in \mathcal{P}(\mathbb{R}^{(N-1)d})$  denote the  $(X_2, \dots, X_N)$ -marginal of  $\pi$ . Now since  $\pi$  is a VMOT,  $\pi^2, \dots, \pi^N$  must maximize  $\mathbb{E}_\gamma[\sum_{i=2}^N c_i(X_i)]$  among all couplings  $\gamma \in \bigcup_{(\mu_2, \dots, \mu_N) \in \mathcal{A}} \Pi(\mu_2, \dots, \mu_N)$ , where  $(\mu_2, \dots, \mu_N) \in \mathcal{A}$  means that  $\mu_t \in \Pi(\mu_{t,1}, \mu_{t,2})$  for  $t = 2, \dots, N$  and satisfies the convex order  $\chi_{(\mu_{11}, \mu_{12})} \preceq_c \mu_2 \preceq_c \dots \preceq_c \mu_N$ . This implies that  $\pi$  remains a VMOT for the cost  $c_\delta(x_1, \dots, x_N)$  for all  $\delta > 0$ . Letting  $\delta \searrow 0$ , we conclude that  $\pi$  remains a VMOT for the original cost  $c(x_1, \dots, x_N)$ .  $\square$

**4. Counterexamples and Related Examples.** Theorem 2.3 and the conjecture 2.6 propose structural properties for VMOT under supermodular costs. However, these properties face limitations in certain scenarios. In this section, we present four examples to explore these boundaries and other related aspects of the VMOT framework.

**4.1. Counterexamples to the Strict Supermodularity of the Cost.** In Theorem 2.3, if  $c_1$  is not strictly supermodular, then the first marginal  $\pi^1$  of a VMOT  $\pi$  is not necessarily monotone. We now provide a counterexample to demonstrate this. Recall that  $\text{conv}(A)$  denotes the convex hull of a set  $A$  in a vector space.

*Example 4.1.* The strict supermodularity of  $c_1$  is necessary for part ii) of Theorem 2.3. 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  $\psi_2(y_2) = \psi_1^*(y_2) = \frac{2}{3}|y_2|^{\frac{3}{2}}$ . The inequality  $\psi_1(y_1) + \psi_2(y_2) \geq y_1y_2$  holds for all  $y_1, y_2 \in \mathbb{R}$ , and the set  $\Gamma_{\{\psi_1, \psi_2\}}$  is defined as:

$$\begin{aligned} \Gamma_{\{\psi_1, \psi_2\}} &:= \{(y_1, y_2) \in \mathbb{R}^2 \mid \psi_1(y_1) + \psi_2(y_2) = y_1y_2\} \\ &= \{(y_1, y_2) \mid y_2 = \psi_1'(y_1)\} \\ &= \{(y_1, y_2) \mid y_2 = |y_1|^2 \text{ if } y_1 \geq 0, y_2 = -|y_1|^2 \text{ if } y_1 \leq 0\}. \end{aligned}$$

Let  $z = (-1, 1)$  and  $w = (1, -1)$ , and define the measure  $\pi^1 = \frac{1}{2}\delta_z + \frac{1}{2}\delta_w \in \mathcal{P}(\mathbb{R}^2)$ . Now, choose martingale kernels  $\kappa_z, \kappa_w \in \mathcal{P}(\mathbb{R}^2)$  such that:

$$\int_{\mathbb{R}^2} x \kappa_z(dx) = z, \quad \int_{\mathbb{R}^2} x \kappa_w(dx) = w, \quad \kappa_z(\Gamma_{\{\psi_1, \psi_2\}}) = \kappa_w(\Gamma_{\{\psi_1, \psi_2\}}) = 1.$$

This choice is possible because  $\text{conv}(\Gamma_{\{\psi_1, \psi_2\}}) = \mathbb{R}^2$ .

We then define a martingale measure  $\pi$  such that the first marginal is  $\pi^1$  and the kernel is  $\{\kappa_z, \kappa_w\}$ . Now take the cost function  $c(x, y) = y_1y_2$ , with  $\phi_1 = \phi_2 = h_1 = h_2 = 0$ , and note that  $(\phi_i, \psi_i, h_i)_{i=1,2}$  and  $\pi$  jointly satisfy the optimality conditions (3.1) and (3.3). This implies that  $\pi$  is a VMOT in the class  $\text{VMT}(\mu_1, \mu_2, \nu_1, \nu_2)$ , where  $\mu_1, \mu_2, \nu_1, \nu_2$  are the one-dimensional marginals of  $\pi$ . However, by construction, the first marginal  $\pi^X = \frac{1}{2}\delta_z + \frac{1}{2}\delta_w$  is not monotone.

**4.2. Counterexamples to the conjecture.** We demonstrate that the conjecture fails when  $d > 2$ . Consider the following examples:

*Example 4.2.* Conjecture 2.6 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 (1.1) with the cost  $c = c(y) = y_1y_2 + y_2y_3 + y_3y_1$

<sup>4</sup>The finite second moment assumption is used here to ensure the integrability of  $x_{11}x_{12}$ .

and the marginals  $\vec{\mu}, \vec{\nu}$ , its first time marginal  $\pi^1$  fails to be the monotone coupling of  $\vec{\mu} = (\mu_1, \mu_2, \mu_3)$ .

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. Then, define  $\beta_0 : \mathbb{R}^3 \rightarrow \mathbb{R} \cup \{+\infty\}$  as:

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

It is clear 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  $\text{conv}(\mathcal{Y})$ , where

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

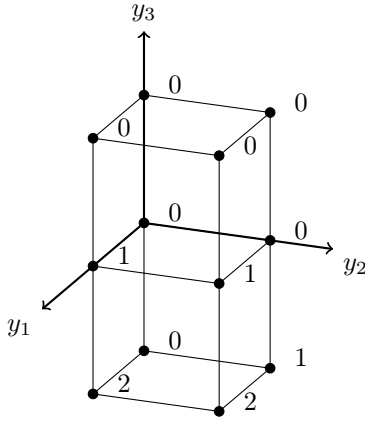


FIG. 2.  $\beta_0$  values on the vertices.

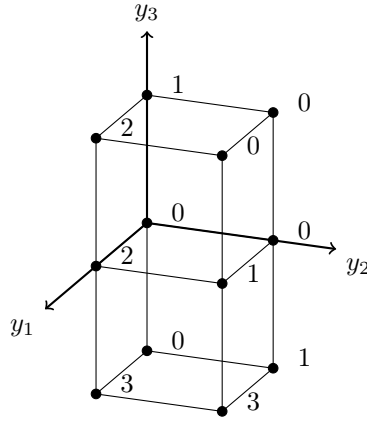


FIG. 3.  $\beta$  values on the vertices.

Define  $\psi_1, \psi_2, \psi_3 : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$  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  $c(y) = y_1 y_2 + y_2 y_3 + y_3 y_1$ . We have

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

Notice  $\beta \geq \beta_0$ , and  $\beta = \beta_0$  on  $\mathcal{Z} := \{n_0^-, n_2, n_{12}^+\}$ . Observed that  $\beta_0 = \beta_0^{**}$  on  $\mathcal{Z}$  (in fact also on  $\mathcal{Y}$ ), hence  $\beta = \beta^{**}$  on  $\mathcal{Z}$  as well. We 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  $\{u, u'\} \subseteq \text{conv}(\mathcal{Z})$ , 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  $\pi^1 := \frac{1}{2}(\delta_u + \delta_{u'})$  as the first time marginal of  $\pi$ . Then we take the martingale kernel  $\kappa_1$  as the unique probability measure supported on  $\mathcal{Z}$  with its barycenter  $x$  in  $\text{conv}(\mathcal{Z})$ . Now define  $\pi = \pi^1 \otimes \kappa_1$  (note that we only need  $\pi_1$  for  $x = u, u'$ ), and let  $\vec{\mu} := (\mu_1, \mu_2, \mu_3)$  be the 1D marginals of  $\pi^1$ , and let  $\vec{\nu} := (\nu_1, \nu_2, \nu_3)$  be the 1D marginals of  $\pi^2 = \frac{1}{2}(\kappa_u + \kappa_{u'})$ . We now claim that  $\pi$  is a VMOT solving the problem (1.1) 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. 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  $\sum_{i=1}^3 \phi_i = \frac{L_1+L_2}{2} = \frac{1}{2}x_1 + \frac{1}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}$ . Now take  $h(x) = (h_1(x), h_2(x), h_3(x))$  as

$$(4.1) \quad h(x) = \begin{cases} \nabla L_1 = (0, 0, 0) & \text{if } L_1(x) > L_2(x), \\ \nabla L_2 = (1, 1, -1) & \text{if } L_1(x) < L_2(x), \\ \nabla \frac{L_1+L_2}{2} = (\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}) & \text{if } L_1(x) = L_2(x). \end{cases}$$

In order to prove the optimality of  $\pi$  and  $(\phi, \psi, h)$  simultaneously, we need to confirm the optimality conditions (3.1) and (3.3). To see (3.1), observe

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

Observe further that (3.3) follows by the fact that on  $\mathcal{Z}$ ,  $L_3 \leq L_1 = L_2 = \beta$ , such that the above inequalities become equalities 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 problems respectively with the cost  $c$  and the marginals  $\vec{\mu}, \vec{\nu}$ .

Finally, take any  $\gamma \in \text{VMT}(\vec{\mu}, \vec{\nu})$ , such that its first time marginal  $\gamma^1 = \frac{1}{2}(\delta_{\bar{u}} + \delta_{\bar{u}'})$  is monotone. We claim that  $\gamma$  cannot be optimal. If  $\gamma$  were optimal, it must satisfy the optimality condition (3.3) 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 implies the strict inequality

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

This shows that  $\gamma$  cannot satisfy (3.3), and hence cannot be a VMOT for (1.1).

#### 4.3. Example for Tighter Bounds for Single-Period Costs in VMOT.

This example demonstrates how VMOT leverages its martingale structure to provide tighter bounds for costs that depend solely on a single period, outperforming classical OT in this setting.

*Example 4.3.* Let  $N = 2$  and  $d = 2$ , and consider  $\mu_1 = \mu_2 = \nu_1$ , which are uniform distributions on  $[0.5, 1]$ , while  $\nu_2$  is uniform on  $[0, 1.5]$ . The cost function is  $c(y_1, y_2) = e^{y_1 + y_2}$ , depending solely on the second period. Notably,  $c(y_1, y_2)$  is strictly supermodular.

By Theorem 2.3, the first-period marginal coupling  $\pi^1$  is monotone under any VMOT  $\pi$ . Since  $\mu_1 = \mu_2$ , the monotone coupling of these identical measures is realized by the identity map. Therefore, we conclude  $X_1 = X_2$  almost surely under  $\pi$ . Moreover, as  $\mu_1 = \nu_1$ , the martingale constraint implies  $X_1 = X_2 = Y_1$  almost surely under  $\pi$ .

Let  $(u_1, u_2, v_1, v_2, h_1, h_2)$  be a dual minimizer described in Theorem 3.1. Then it holds

$$u_1(x_1) + u_2(x_2) + v_1(y_1) + v_2(y_2) + h_1(x_1, x_2)(y_1 - x_1) + h_2(x_1, x_2)(y_2 - x_2) \geq c(y_1, y_2),$$

where equality holds  $\pi$ -a.s.. Since  $X_1 = X_2 = Y_1 = x$  a.s., the above dual constraint simplifies to

$$u(x) + v(y) + h(x)(y - x) \geq c(x, y),$$

where  $Y_2 = y$ ,  $u(x) = u_1(x) + u_2(x) + v_1(x)$ ,  $v(y) = v_2(y)$ , and  $h(x) = h_2(x, x)$ . This simplified dual constraint implies that the VMOT problem reduces to a standard two-period MOT in  $d = 1$ .

The cost function  $c(x, y) = e^{x+y}$  satisfies the martingale Spence-Mirrlees condition  $c_{xyy} > 0$ , ensuring that the optimal coupling for the second period is the left monotone coupling. [3, 18] This results in a tighter upper bound for  $\int c(y_1, y_2) d\pi$  compared to the monotone coupling between  $\nu_1$  and  $\nu_2$  in classical OT. Specifically, the optimal VMOT cost is approximately 5.1124, while the classical OT cost is 5.2669.

This example demonstrates how VMOT incorporates additional information, such as martingale constraints, to provide sharper bounds on the no-arbitrage prices (costs), offering insights into scenarios where multi-period dynamics influence optimal couplings.

**4.4. Example: Relaxing the Martingale Constraints: Marginal Martingale Transport.** We consider a relaxation of the VMOT problem, introducing the Marginal Martingale Optimal Transport (MMOT) framework. This framework simplifies the constraints by requiring marginal martingale conditions, explores a simpler coupling condition compared to the full vectorial martingale transport. The MMOT problem is defined as:

$$\text{maximize } \mathbb{E}_\pi[c(X)] \text{ over } \pi \in \text{MMT}(\vec{\mu}),$$

where the feasible set MMT (Marginal Martingale Transport) consists of couplings  $\pi$  satisfying:

$$\begin{aligned} \text{MMT}(\vec{\mu}) &:= \{\pi \in \mathcal{P}(\mathbb{R}^{Nd}) \mid \pi = \text{Law}(X), X = (X_1, \dots, X_N), \\ &\quad \mathbb{E}_\pi[X_{t+1,i} | X_{1,i}, \dots, X_{t,i}] = X_{t,i}, \text{Law}(X_{t,i}) = \mu_{t,i}, \text{ for all } t \in [N], i \in [d]\}. \end{aligned}$$

*Example 4.4.* To demonstrate the solution structure, we use the same setting as the previous example:  $\mu_1 = \mu_2 = \nu_1$ , which are uniform distributions on  $[0.5, 1]$ , and  $\nu_2$  is uniform on  $[0, 1.5]$ . The cost function is  $c(y_1, y_2) = e^{y_1 + y_2}$ .

To construct an optimal coupling  $\pi \in \text{MMT}$ , let  $\gamma_1$  and  $\gamma_2$  be any chosen martingale couplings of  $\mu_1$  with  $\nu_1$  and  $\mu_2$  with  $\nu_2$ , respectively. ( $X_1 = Y_1$  under  $\gamma_1$ .) Disintegrate these couplings:

$$\gamma_1 = \nu_1 \otimes \kappa_1^2, \quad \gamma_2 = \nu_2 \otimes \kappa_2^2,$$



where  $\kappa_1^2$  and  $\kappa_2^2$  are conditional probability kernels. Define the coupling:

$$\pi = \pi_2 \otimes (\kappa_1^2 \otimes \kappa_2^2),$$

where  $\pi_2$  is the monotone coupling of  $\nu_1$  and  $\nu_2$ .

This construction ensures  $\mathbb{E}_\pi[Y_i | X_i] = X_i$ , satisfying the marginal martingale constraints. Since the cost  $c(y_1, y_2)$  depends only on the second period, the expected cost becomes:

$$\mathbb{E}_\pi[c] = \mathbb{E}_{\pi_2}[c].$$

Thus,  $\pi_2$ , being the monotone coupling, maximizes the cost and guarantees the optimality for the MMOT problem.

Note that since  $\chi_{\vec{\mu}} \not\preceq_c \chi_{\vec{\nu}}$ , this example also illustrates that the MMOT problem and the VMOT problem can exhibit fundamentally different optimality properties.

**5. Numerical implementation.** Throughout this section, we assume that  $d = 2$  and the cost function  $c$  is of the form  $c(x_1, \dots, x_N) = \sum_{t=1}^N c_t(x_{t,1}, x_{t,2})$  where each  $c_t$  is supermodular. We develop a variant of the VMOT problem that reduces its dimensionality taking advantage of our main result. We then employ a general version of the Sinkhorn algorithm, developed in [19] to compute the solutions for two examples in both general and reduced versions. As expected, the lower dimensional reduced version performs much better.

**5.1. A hybrid version of the VMOT Problem.** Theorem 2.3 motivates the following modified problem:

$$(5.1) \quad \text{maximize } \mathbb{E}_\pi[c(X)] \quad \text{over } \pi \in \text{VMT}(\chi, \vec{\mu}_2, \dots, \vec{\mu}_N).$$

where

$$(5.2) \quad \begin{aligned} \text{VMT}(\chi, \vec{\mu}_2, \dots, \vec{\mu}_N) := & \{ \pi \in \mathcal{P}(\mathbb{R}^{2N}) \mid \pi = \text{Law}(X), X = (X_1, \dots, X_N), \\ & \mathbb{E}_\pi[X_t \mid X_1, \dots, X_{t-1}] = X_{t-1}, \text{Law}(X_1) = \chi, \text{Law}(X_{t,i}) = \mu_{t,i}, \\ & \text{for all } t = 2, \dots, N, i = 1, 2 \}, \end{aligned}$$

with some notation overload. This formulation assumes that we know the joint distribution of  $X_1$ . By Theorem 2.3, the VMOT problem (1.1) is equivalent to the above problem in which  $\chi \in \mathcal{P}(\mathbb{R}^2)$  is the monotone coupling of  $\vec{\mu}_1$ ,  $\chi = \chi_{\vec{\mu}_1}$ <sup>5</sup>. For fixed  $\chi_{\vec{\mu}_1}$ , we call  $T$  the monotone map such that  $\mu_{1,2} = T_{\#}\mu_{1,1}$ , and  $\tilde{c}(x_{1,1}, x_2, \dots, x_N) = c(x_{1,1}, T(x_{1,1}), \dots, x_N)$ . With this, we can rewrite (5.1) as

$$(5.3) \quad \text{maximize } \mathbb{E}_\pi[\tilde{c}(X_{1,1}, X_2, \dots, X_N)] \quad \text{over } \pi \in \text{VMT}(\chi_{\vec{\mu}_1}, \vec{\mu}_2, \dots, \vec{\mu}_N).$$

This problem is simpler than the original (1.1), due to the dimensional reduction in going from  $X_1 \in \mathbb{R}^2$  to  $X_{1,1}$  which is only 1 dimensional. We define the simplified space of functions

$$\begin{aligned} \tilde{\mathcal{H}} = & \left\{ \tilde{\varphi}(x_{1,1}, x_2, \dots, x_N) := u_{1,1}(x_{1,1}) + h_{1,1}(x_{1,1})(x_{2,1} - x_{1,1}) + h_{1,2}(x_{1,1})(x_{2,2} - T(x_{1,1})) \right. \\ & + \sum_{t=2}^N \sum_{i=1}^2 u_{t,i}(x_{t,i}) + \sum_{t=2}^{N-1} \sum_{i=1}^2 h_{t,i}(x_{1,1}, x_2, \dots, x_t)(x_{t+1,i} - x_{t,i}) \\ & \left. \left| \begin{array}{l} u_{t,i} \in L^1(\mu_{t,i}), \quad h_{t,i} \in L^\infty(\mathbb{R}^{td}), \quad t \in [N], \quad i = 1, 2 \end{array} \right. \right\}. \end{aligned}$$

<sup>5</sup>Notice that in this case (5.2) refines the original definition in the sense that it defines a smaller search set that still contains the solution, suggesting an improvement in numerical efficiency.

The dual problem to 5.3 is formulated as

$$(5.4) \quad \inf_{(u,h) \in \tilde{\Xi}} \int u_{1,1} d\mu_{1,1} + \sum_{t=2}^N \sum_{i=1}^2 \int u_{t,i} d\mu_{t,i}$$

where  $\tilde{\Xi}$  consists of functions  $\tilde{\varphi} \in \tilde{\mathcal{H}}$  such that  $\tilde{\varphi}(x_{1,1}, x_2, \dots, x_N) \geq \tilde{c}(x_{1,1}, x_2, \dots, x_N)$ . As with the primal problem, this dual is simpler than the original (3.2), since we replace the two potential functions  $u_{1,1}, u_{1,2}$  with the single  $u_{1,1}$ , while the functions  $h_{1,i}$  now take a scalar argument  $x_{1,1}$  in place of  $x_1$ .

We can treat both the original VMT condition and the modified VMT condition as linear constraints over the set of couplings. For the original VMT condition,  $\pi \in \text{VMT}(\vec{\mu})$  if and only if  $\pi$  has the correct marginals and

$$(5.5) \quad \int_{\mathbb{R}^{Nd}} \sum_{t=1}^{N-1} \sum_{i=1}^2 h_{t,i}(x_1, \dots, x_t)(x_{t+1,i} - x_{t,i}) d\pi(x_1, \dots, x_t) = 0,$$

for any  $h_{i,t}$ , while for the modified VMT condition,  $\pi \in \text{VMT}(\chi, \vec{\mu}_2, \dots, \vec{\mu}_N)$  if and only if  $\pi$  has the correct marginals and

$$(5.6) \quad \int_{\mathbb{R}^{Nd}} h_{1,1}(x_{1,1})(x_{2,1} - x_{1,1}) + h_{1,2}(x_{1,1})(x_{2,2} - T(x_{1,1})) \\ + \sum_{t=2}^{N-1} \sum_{i=1}^2 h_{t,i}(x_{1,1}, x_2, \dots, x_t)(x_{t+1,i} - x_{t,i}) d\pi(x_{1,1}, x_2, \dots, x_t) = 0,$$

for any  $h_{i,t}$ .

**5.2. Entropic regularization and Sinkhorn algorithm.** We next consider an entropic regularization of the VMOT problem, which enables efficient computation via a generalized Sinkhorn algorithm. Given  $\vec{\mu} = (\mu_{t,i})_{t,i}$ ,  $\hat{\mu} = \otimes_{t,i} \mu_{t,i}$ , and  $\epsilon > 0$ , we solve

$$(5.7) \quad \text{maximize} \quad \int c d\pi - \epsilon H_{\hat{\mu}}(\pi) \quad \text{over } \pi \in \text{VMT}(\vec{\mu}),$$

where  $H_{\hat{\mu}}(\pi) = \int \log\left(\frac{d\pi}{d\hat{\mu}}\right) d\pi$  is the relative entropy. The corresponding dual is

$$(5.8) \quad \text{minimize} \quad \sum_{t,i} \int u_{t,i} d\mu_{t,i} + \epsilon \int \exp\left(\frac{c-\varphi}{\epsilon}\right) d\hat{\mu},$$

where  $\varphi(x) := \sum_{t=1}^N \sum_{i=1}^2 u_{t,i}(x_{t,i}) + \sum_{t=1}^{N-1} \sum_{i=1}^2 h_{t,i}(x_1, \dots, x_t)(x_{t+1,i} - x_{t,i})$  as in (3.2). Similarly, the reduced dimension problem (5.3) and its dual (5.4) may be regularized; this amounts to replacing  $(x_{1,1}, x_{1,2})$  by  $(x_{1,1}, T(x_{1,1}))$ , removing  $u_{1,2}$  and removing one dimension for each  $h_{t,i}$  in (5.7) and (5.8). The regularized VMOT problems (1.1) and (3.2) as well as their modified versions (5.3) and (5.4) fall into the class of optimal transport problems treated in [19]. As  $\epsilon \rightarrow 0$ , solutions to each of these problems converge to solutions of the unregularized versions; therefore, solving them for small  $\epsilon > 0$  yields approximate solutions to the unregularized problems. A general version of the celebrated Sinkhorn algorithm which applies to dual problems of this type was developed in [19]. We will apply this algorithm to approximate solutions of (3.2) and (5.3), illustrating the improved performance of the modified version (5.3).

The generalized Sinkhorn algorithm alternates updates of the dual potentials  $(u_{t,i})$  and the martingale multipliers  $(h_{t,i})$ . Each update enforces the marginal and martingale constraints respectively while keeping other variables fixed. The updates for  $u_{t,i}$  admit closed-form exponential formulas, while those for  $h_{t,i}$  are implicit and are solved numerically. Note that  $h_{t,i}$  can be solved uniquely. Full derivations of the update rules are provided in Appendix B.

Take the  $d = 2$ ,  $N = 2$  VMOT problem with each marginal discretized on  $n$  points as an illustrative example; exploiting the monotone first-period coupling lowers the discrete unknowns from  $n^4$  to  $n^3$  in the reduced formulation. This dimensional reduction effect carries to the dual after entropic regularization, making each Sinkhorn step cheaper. Empirically, for our three-period experiment in Table 3, the CPU time is reduced by 99%.

**5.3. Examples.** We illustrate the performance of the Sinkhorn-based method and the dimensional reduction effect on two instances of the VMOT problem: a two-period case and a three-period case. All numerical calculations are performed in Python on a 13th Gen Intel(R) Core(TM) i7-13620H 2.40 GHz notebook.

*Two-period case..* Let  $U(a, b)$  denote the uniform measure supported between  $a$  and  $b$ . Each marginal  $\mu_{1,1} = U(-0.1, 0.1)$ ,  $\mu_{1,2} = U(-0.1, 0.1)$ ,  $\mu_{2,1} = U(-0.4, 0.4)$ ,  $\mu_{2,2} = U(-0.3, 0.3)$  is discretized on 100 uniform grid points. The cost function is  $c(x_1, x_2) = x_{1,1}x_{1,2} + x_{2,1}x_{2,2}$ , for which each period is super-modular. The reduced Sinkhorn formulation achieves an optimal value of 0.02542 in 2.52 seconds, while the complete version yields 0.01906 in 327.32 seconds; note that the reduction in computation time is roughly two orders of magnitude, consistent with the heuristics described above for  $n = 100$ . We can see the comparison in Table 2. Note that it is not surprising that the reduced method achieves a higher optimal value, as it has an exact optimal coupling between  $\mu_{1,1}$  and  $\mu_{1,2}$  while the coupling deduced from the full Sinkhorn method is only approximately optimal (see the first column in Figure 4 for a comparison of these couplings). Therefore, the total cost coming from the  $x_{1,1}x_{1,2}$  term in the cost function is necessarily higher in the reduced problem. Note as well that the optimal values reported in Table 2 do not include the entropy  $H_{\hat{\mu}}(\pi)$  of the coupling.

TABLE 2  
*Two-period VMOT problem ( $d = 2$ ,  $N = 2$ ). Each marginal discretized on 100 grid points.*

Method	Optimal value	Time (s)
Full Sinkhorn	0.01906	327.32
Reduced Sinkhorn	0.02542	2.52

We include the heat map plot in Figure 4 to demonstrate the induced calculated coupling between different marginals. Lighter colors represent higher probability and vice versa. The upper row is the plot for the reduced Sinkhorn method, while the bottom row is the complete Sinkhorn method. The first column is the plot for the marginals  $\mu_{1,1}, \mu_{1,2}$ , that is, the coupling for the first period, while the second column is the plot for the marginal  $\mu_{2,1}, \mu_{2,2}$ , i.e., the second period coupling. We see that all couplings are concentrated around the diagonal. For the first period, the complete Sinkhorn method demonstrates a slight deviation from the diagonal due to the error introduced by the entropic regularization, while the reduced Sinkhorn method concentrates right on the diagonal, which is the true coupling based on our previous result. This helps provide a better numerical result with less computational time.

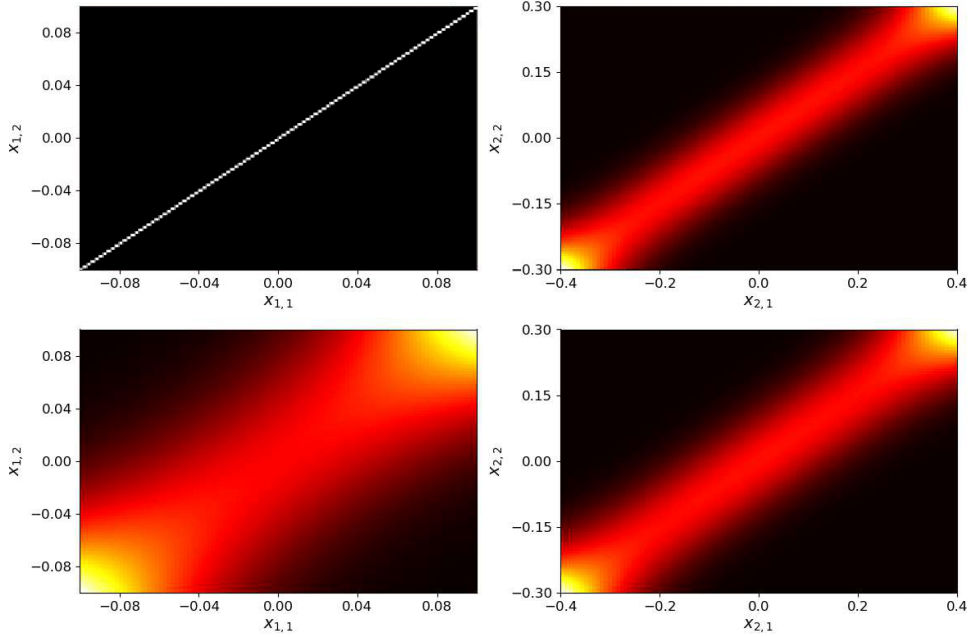


FIG. 4. Heat maps of the induced couplings between different marginals obtained from the reduced (top row) and complete (bottom row) Sinkhorn algorithms. Lighter colors indicate higher probability density.

*Three-period case.* To test scalability, we add a third period with marginals  $\mu_{3,1} = U(-0.6, 0.6)$ ,  $\mu_{3,2} = U(-0.7, 0.7)$  discretized on 30 grid points each, and  $\mu_{1,1}, \mu_{1,2}, \mu_{2,1}, \mu_{2,2}$  on 15 grid points each. The cost function is extended as  $c(x_1, x_2, x_3) = x_{1,1}x_{1,2} + x_{2,1}x_{2,2} + x_{3,1}x_{3,2}$ . The reduced method attains an optimal value of 0.006885 in 57.28 seconds, while the full formulation yields 0.006766 in 15966.32 seconds. In this example, we see that the reduction in computation time is very significant.

TABLE 3  
Three-period VMOT problem ( $d = 2$ ,  $N = 3$ ). Marginals  $\mu_{1,1}, \mu_{1,2}, \mu_{2,1}, \mu_{2,2}$  discretized on 15 grid points and  $\eta_1, \eta_2$  on 30 grid points.

Method	Optimal value	Time (s)
Full Sinkhorn	0.006766	15966.32
Reduced Sinkhorn	0.006885	57.28

We also include the heat map in Figure 5 for the three-period case. The top row is the plot for the reduced Sinkhorn method, while the bottom row is the complete Sinkhorn method. The first column is the plot for the marginal  $\mu_{1,1}, \mu_{1,2}$ . The second column is the plot for the marginal  $\mu_{2,1}, \mu_{2,2}$ . The third column is the plot for the marginal  $\mu_{3,1}, \mu_{3,2}$ . The plot also shows that the first period coupling for the reduced Sinkhorn method is significantly better than the complete Sinkhorn method.

#### Appendix A. Proof of Proposition 3.3.

*Proof.* We begin by proving part i). If  $\beta \equiv \infty$ , then  $\beta^* \equiv -\infty$ , and there is

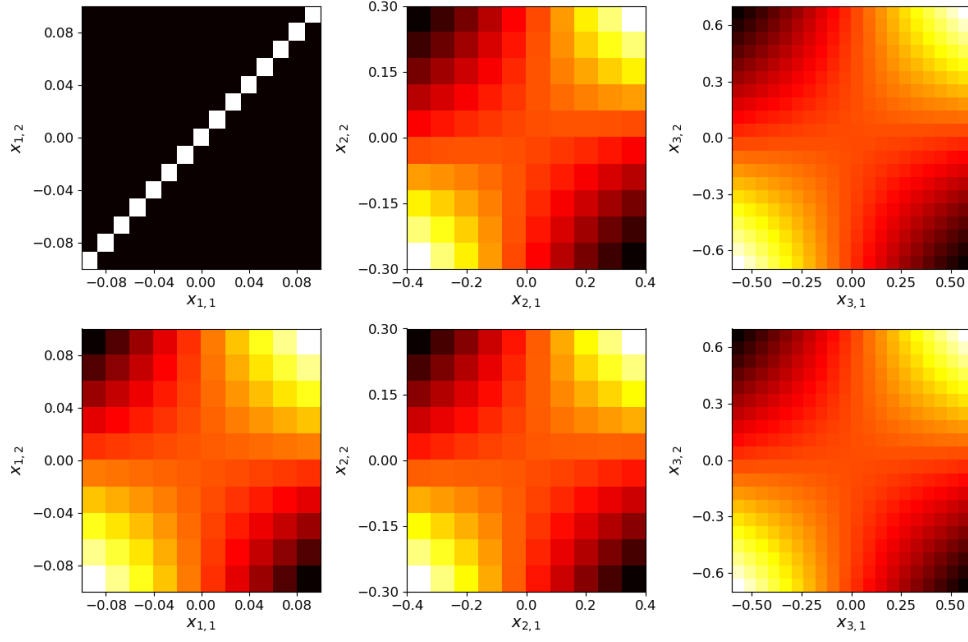


FIG. 5. Heat maps of the induced couplings between different marginals obtained from the Sinkhorn algorithm for the three-period case. Lighter colors indicate higher probability density. The figure compares the complete and reduced formulations discussed in the text.

nothing to prove. If  $\beta \not\equiv \infty$  but  $\beta^*$  is not proper (i.e.,  $\beta^* \equiv \infty$ ), again there is nothing to prove. Thus, we assume that  $\beta$  and  $\beta^*$  are proper.

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

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

Note that  $\beta_R$  is submodular, and for sufficiently large  $R$ ,  $\beta_R$  is proper. Since  $\beta_R$  is compactly supported,  $\beta_R^*$  will be Lipschitz unless  $\beta_R^* \equiv +\infty$ . However, the latter case is excluded because  $\beta^* \geq \beta_R^*$ . Therefore,  $\beta_R^*$  is real-valued everywhere.

Next, we show the supermodularity of  $\beta_R^*$ . By the definition of the Legendre transform, for any vectors  $z, w \in \mathbb{R}^d$  and  $\epsilon > 0$ , there exists  $x, y \in \mathbb{R}^d$  such that:

$$\begin{aligned} \beta_R^*(z) &\leq x \cdot z - \beta_R(x) + \epsilon, \\ \beta_R^*(w) &\leq y \cdot w - \beta_R(y) + \epsilon, \end{aligned}$$

and

$$\begin{aligned} \beta_R^*(z \wedge w) &\geq (x \wedge y) \cdot (z \wedge w) - \beta_R(x \wedge y), \\ \beta_R^*(z \vee w) &\geq (x \vee y) \cdot (z \vee w) - \beta_R(x \vee y). \end{aligned}$$

It is a direct calculation to show that:

$$(x \wedge y) \cdot (z \wedge w) + (x \vee y) \cdot (z \vee w) \geq x \cdot z + y \cdot w,$$

we can combine the inequalities to obtain:

$$\begin{aligned}\beta_R^*(z \wedge w) + \beta_R^*(z \vee w) &\geq (x \wedge y) \cdot (z \wedge w) + (x \vee y) \cdot (z \vee w) - \beta_R(x \wedge y) - \beta_R(x \vee y) \\ &\geq x \cdot z + y \cdot w - \beta_R(x) - \beta_R(y) \\ &= \beta_R^*(z) + \beta_R^*(w) - 2\epsilon.\end{aligned}$$

Taking the limit as  $\epsilon \rightarrow 0$ , we conclude that  $\beta_R^*$  is supermodular.

As  $R \rightarrow \infty$ , we have  $\beta_R \searrow \beta$  on  $\mathbb{R}^d$ , and thus  $\beta_R^* \nearrow \beta^*$ . Therefore, the supermodularity of  $\beta^*$  follows.

Now we prove part ii). If  $\beta(x, y)$  is supermodular on  $\mathbb{R}^2$ , then define  $\tilde{\beta}(x, y) := \beta(x, -y)$ . By part i),  $\tilde{\beta}$  is submodular, and thus  $\tilde{\beta}^*$  is supermodular. Hence we have:

$$\begin{aligned}\tilde{\beta}^*(z, -w) &= \sup_{x, y} (xz + y(-w) - \tilde{\beta}(x, y)) \\ &= \sup_{x, y} (xz - yw - \beta(x, -y)) \\ &= \sup_{x, y} (xz + yw - \beta(x, y)) \\ &= \beta^*(z, w),\end{aligned}\quad \square$$

which shows that  $\beta^*(z, w)$  is submodular. Part iii) follows from parts i) and ii).

### Appendix B. Details of the numerical methods.

We illustrate the method when  $N = 2$  and  $d = 2$ . For general  $N$ , the calculations are similar. First, we fix a discretization scheme for the marginals:  $x_{1,1}^i, x_{2,1}^j, x_{1,2}^s, x_{2,2}^t$ .

The dual problem of the entropic regularized VMOT problem with  $N = 2, d = 2$  is

$$\begin{aligned}\min_{u_{1,1}^i, u_{2,1}^j, u_{1,2}^s, u_{2,2}^t, h_1^{i,j}, h_2^{i,j}} &\sum_i u_{1,1}^i \mu_{1,1}^i + \sum_j u_{2,1}^j \mu_{2,1}^j + \sum_s u_{1,2}^s \mu_{1,2}^s + \sum_t u_{2,2}^t \mu_{2,2}^t \\ &+ \epsilon \sum_{i,j,s,t} \exp\left(\frac{1}{\epsilon} (c^{i,j,s,t} - u_{1,1}^i - u_{2,1}^j - u_{1,2}^s - u_{2,2}^t \right. \\ &\quad \left. - h_1^{i,j}(x_{1,2}^s - x_{1,1}^i) - h_2^{i,j}(x_{2,2}^t - x_{2,1}^j))\right) \mu_{1,1}^i \mu_{2,1}^j \mu_{1,2}^s \mu_{2,2}^t.\end{aligned}$$

The generalized Sinkhorn method is an example of the block coordinate descent method [19]. The idea is to divide the set of variables to be optimized into different blocks and to optimize the objective function block by block while keeping all the other blocks fixed.

To implement the numerical algorithm, we first take the partial derivative of the dual functional with respect to each variable and set the derivative equal to 0. The resulting equations are then used to update the variables. We denote the dual functional by  $\Phi(u_{1,1}^i, u_{2,1}^j, u_{1,2}^s, u_{2,2}^t, h_1^{i,j}, h_2^{i,j})$ .

For the variables  $u_{1,1}^i, u_{2,1}^j, u_{1,2}^s, u_{2,2}^t$ , we take  $u_{1,1}^i$  as an example, and the procedure for other variables is similar. If we take the partial derivative of  $\Phi$  with respect to  $u_{1,1}^i$  and we get

$$\begin{aligned}\frac{\partial \Phi}{\partial u_{1,1}^i} &= \mu_{1,1}^i + \sum_{j,s,t} \exp\left(\frac{1}{\epsilon} (c^{i,j,s,t} - u_{1,1}^i - u_{2,1}^j - u_{1,2}^s - u_{2,2}^t \right. \\ &\quad \left. - h_1^{i,j}(x_{1,2}^s - x_{1,1}^i) - h_2^{i,j}(x_{2,2}^t - x_{2,1}^j))\right) \mu_{1,1}^i \mu_{2,1}^j \mu_{1,2}^s \mu_{2,2}^t = 0.\end{aligned}$$

Solving for  $u_{1,1}^i$ , we get the new  $u_{1,1}^{i,k+1}$  after the iteration:

$$u_{1,1}^{i,k+1} = -\epsilon \ln \left( \sum_{j,s,t} \exp \left( \frac{1}{\epsilon} (c^{i,j,s,t} - u_{2,1}^{j,k} - u_{1,2}^{s,k} - u_{2,2}^{t,k} - h_1^{i,j,k}(x_{1,2} - x_{1,1}) - h_2^{i,j,k}(x_{2,2} - x_{2,1})) \right) \mu_{2,1}^j \mu_{1,2}^s \mu_{2,2}^t \right).$$

For  $h_1^{i,j}$ , we differentiate  $\Phi$  and get

$$\begin{aligned} \frac{\partial \Phi}{\partial h_1^{i,j}} = & - \sum_{s,t} \exp \left( \frac{1}{\epsilon} (c^{i,j,s,t} - u_{1,1}^i - u_{2,1}^j - u_{1,2}^s - u_{2,2}^t \right. \\ & \left. - h_1^{i,j}(x_{1,2}^s - x_{1,1}^i) - h_2^{i,j}(x_{2,2}^t - x_{2,1}^j)) \right) (x_{1,2}^s - x_{1,1}^i) \mu_{1,1}^i \mu_{2,1}^j \mu_{1,2}^s \mu_{2,2}^t = 0. \end{aligned}$$

We cannot solve for  $h_1^{i,j}$  analytically. However, we can still simplify the equation. Then the new  $h_1^{i,j}$  after the iteration will be

$$\sum_{s,t} \exp \left( \frac{1}{\epsilon} (c^{i,j,s,t} - u_{1,1}^{i,k} - u_{2,1}^{j,k} - u_{1,2}^{s,k} - u_{2,2}^{t,k} - h_1^{i,j,k+1}(x_{1,2}^s - x_{1,1}^i) - h_2^{i,j,k}(x_{2,2}^t - x_{2,1}^j)) \right) (x_{1,2}^s - x_{1,1}^i) \mu_{1,2}^s \mu_{2,2}^t = 0.$$

A similar procedure applies to  $h_2^{i,j}$ .

## 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-Labordere, and Friedrich Penkner. Model-independent bounds for option prices—a mass transport approach. *Finance and Stochastics*, 17:477–501, 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 Oblój. 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] 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.
- [12] Stephan Eckstein, Gaoyue Guo, Tongseok Lim, and Jan Oblój. Robust pricing and hedging of options on multiple assets and its numerics. *SIAM Journal on Financial Mathematics*, 12(1):158–188, 2021.
- [13] 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.
- [14] Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113 – 161, 1996.
- [15] Wilfrid Gangbo and Andrzej Świąch. 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.
- [16] Henri Heinrich. Problème de monge pour n probabilités. *Comptes Rendus Mathématique*, 334(9):793–795, 2002.



- [17] Pierre Henry-Labordère. *Model-free hedging: A martingale optimal transport viewpoint*. CRC Press, 2017.
- [18] Pierre Henry-Labordère and Nizar Touzi. An explicit martingale version of the one-dimensional brenier theorem. *Finance and Stochastics*, 20:635–668, 2016.
- [19] Joshua Zoen-Git Hiew, Luca Nenna, and Brendan Pass. An ordinary differential equation for entropic optimal transport and its linearly constrained variants. *Numerische Mathematik*, to appear, 2025.
- [20] David Hobson. The skorokhod embedding problem and model-independent bounds for option prices. *Paris-Princeton lectures on mathematical finance 2010*, pages 267–318, 2011.
- [21] David G Hobson. Robust hedging of the lookback option. *Finance and Stochastics*, 2:329–347, 1998.
- [22] Jens Carsten Jackwerth and Mark Rubinstein. Recovering probability distributions from option prices. *The Journal of Finance*, 51(5):1611–1631, 1996.
- [23] Hans G Kellerer. Markov-komposition und eine anwendung auf martingale. *Mathematische Annalen*, 198:99–122, 1972.
- [24] 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.
- [25] Vladimir Levin. Abstract cyclical monotonicity and monge solutions for the general monge–kantorovich problem. *Set-Valued Analysis*, 7(1):7–32, 1999.
- [26] Tongseok Lim. Geometry of vectorial martingale optimal transportations and duality. *Mathematical Programming*, pages 1–35, 2023.
- [27] Tongseok Lim. Replication of financial derivatives under extreme market models given marginals. *arXiv preprint arXiv:2307.00807*, 2023.
- [28] Georg Gunther Lorentz. An inequality for rearrangements. *The American Mathematical Monthly*, 60(3):176–179, 1953.
- [29] Marcel Nutz, Florian Stebegg, and Xiaowei Tan. Multiperiod martingale transport. *Stochastic Processes and their Applications*, 130(3):1568–1615, 2020.
- [30] Brendan Pass. Uniqueness and monge solutions in the multimarginal optimal transportation problem. *SIAM Journal on Mathematical Analysis*, 43(6):2758–2775, 2011.
- [31] Brendan Pass and Adolfo Vargas-Jiménez. Monge solutions and uniqueness in multi-marginal optimal transport via graph theory. *Advances in Mathematics*, 428, 2023.
- [32] Jarno Talponen and Lauri Viitasaari. Note on multidimensional breeden–litzberger representation for state price densities. *Mathematics and Financial Economics*, 8:153–157, 2014.
- [33] Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.