Dual Attainment in Multi-Period Multi-Asset Martingale Optimal Transport and Its Computation

Charlie Che[*†1], Tongseok Lim[*‡2], and Yue Sun[*§3]

[1]Quantitative Trading & Research, JPMorganChase, New York, NY 10017, USA
[2]Mitch Daniels School of Business, Purdue University, West Lafayette, Indiana 47907, USA
[3]Global Technology Applied Research, JPMorganChase, New York, NY 10001, USA

February 2, 2026

Abstract

We establish dual attainment for the multimarginal, multi-asset martingale optimal transport (MOT) problem, a fundamental question in the mathematical theory of model-independent pricing and hedging in quantitative finance. Our main result proves the existence of dual optimizers under mild regularity and irreducibility conditions, extending previous duality and attainment results from the classical and two-marginal settings to arbitrary numbers of assets and time periods. This theoretical advance provides a rigorous foundation for robust pricing and hedging of complex, path-dependent financial derivatives. To support our analysis, we present numerical experiments that demonstrate the practical solvability of large-scale discrete MOT problems using the state-of-the-art primal-dual linear programming (PDLP) algorithm. In particular, we solve multi-dimensional (or vectorial) MOT instances arising from the robust pricing of worst-of autocallable options, confirming the accuracy and feasibility of our theoretical results. Our work advances the mathematical understanding of MOT and highlights its relevance for robust financial engineering in high-dimensional and model-uncertain environments.

1 Introduction

Martingale optimal transport (MOT) has emerged as a central topic in mathematical finance and probability, providing a rigorous framework for model-independent pricing and hedging of financial derivatives under marginal and martingale constraints. The MOT problem generalizes classical optimal transport (OT) by requiring the transport plan to respect the martingale property, which is fundamental for risk-neutral valuation in quantitative finance. While the classical OT problem, introduced by Monge and Kantorovich [27, 20], has a rich mathematical theory and wide-ranging applications [34, 31], the martingale extension introduces significant new challenges, both theoretically and computationally.

A key question in MOT is the existence of dual optimizers, the so-called *dual attainment* property. Dual attainment is not only of intrinsic mathematical interest, but also has direct implications for robust hedging: dual optimizers correspond to explicit sub- and super-hedging strategies using tradable instruments [2, 3, 5, 4]. While duality and dual attainment have been extensively studied in the classical OT and single-asset, two-period MOT settings [3, 5, 4, 13], the general case of multi-marginal, multi-asset MOT remains much less understood. Theoretical advances in this direction are crucial for the robust pricing and hedging of complex,

---

[*]Authors are listed in alphabetical order by last name.

[†]charlie.che@jpmchase.com

[‡]lim336@purdue.edu

[§]yue.sun@jpmorgan.com

path-dependent financial products, which often depend on the joint evolution of multiple assets over several time periods.

This paper is devoted to the theoretical development of dual attainment for the multimarginal, multi-asset MOT problem. We establish the existence of dual optimizers under mild regularity and irreducibility conditions, extending previous results to arbitrary numbers of assets and time periods. Our approach builds on and generalizes the techniques developed for the single-asset, two-period case, and provides a rigorous foundation for robust pricing and hedging in high-dimensional, path-dependent settings. The main results contribute to the mathematical understanding of MOT and open the door to further applications in quantitative finance.

While the primary focus of this work is theoretical, we complement our analysis with a numerical demonstration. Specifically, we apply a state-of-the-art large-scale linear programming solver, PDLP [1, 18], to solve discrete instances of the multimarginal MOT problem. To our knowledge, this is the first demonstration of PDLP applied to large-scale multimarginal MOT, and the results provide supporting evidence for the practical relevance and accuracy of our theoretical findings. The numerical experiments, including a case study on a worst-of autocallable option, are intended to illustrate the feasibility of solving high-dimensional MOT problems and to validate the dual attainment results in realistic financial scenarios.

Contributions. The primary contribution of this paper is a rigorous proof of dual attainment for the multimarginal, multi-asset martingale optimal transport (MOT) problem under general conditions. We also present numerical illustrations based on PDLP as supporting evidence; however, the development of computational methodologies is not a focus of this work and serves only to corroborate the theoretical results. Overall, our findings advance the mathematical theory of MOT and provide a foundation for robust pricing and hedging in complex financial markets.

Organization. The remainder of the paper is organized as follows. Section 1.1 provides a review of the related literature on optimal transport, martingale optimal transport, and their connections to mathematical finance. Section 1.2 discusses key applications of MOT in quantitative finance, motivating the theoretical developments of this work. Section 2 introduces the mathematical formulation of the multimarginal martingale optimal transport problem, including the market setting, the VMOT problem, and its dual. Section 3 develops our main theoretical results, establishing duality and dual attainment in the multi-marginal, multi-asset setting, and highlights the key ideas underlying the proofs. Section 4 describes the numerical methodology, including the PDLP algorithm and the discrete MOT formulation, and Section 5 reports numerical experiments that support the theoretical findings, with a focus on a worst-of autocallable option case study. Section 6 concludes the paper and outlines directions for future research. Finally, Section 7 contains detailed proofs of the main theoretical results.

1.1 Related Literature

The classical OT problem has been extensively studied, with foundational contributions by Monge, Kantorovich, and later Villani [34] and Santambrogio [31]. The extension to martingale constraints was pioneered in [2, 3], with further developments on duality and dual attainment in [5, 4, 13]. Multi-marginal and multi-asset MOT has been investigated in [10, 28, 15, 21], but dual attainment in the general setting remains an open and challenging problem. Applications in quantitative finance include model calibration [25, 17, 16], robust pricing and hedging [19, 13], and the extraction of risk-neutral marginals from market data using the Breeden–Litzenberger formula [6]. Moreover, recent advances in large-scale linear programming, such as PDLP [1, 18], have enabled practical solution of high-dimensional MOT problems, which we adopt as our computational approach in this paper.

1.2 Applications in Quantitative Finance

The martingale optimal transport (MOT) framework has found increasing relevance in quantitative finance, both as a tool for model calibration and for robust pricing and hedging of complex derivatives. In this section,

we review key application domains, highlight recent advances, and illustrate how the multi-marginal, multi-asset MOT approach enables new capabilities for practitioners.

1.2.1 Model Calibration via Martingale Optimal Transport

A major application of MOT in finance is the calibration of market models to observed data, particularly volatility surfaces and marginal distributions inferred from vanilla option prices. Traditional calibration methods often rely on parametric models, which may fail to capture the full range of market-implied dynamics. MOT provides a non-parametric alternative, allowing practitioners to fit models that are consistent with observed marginals while imposing the martingale property required by risk-neutral pricing.

Recent works have leveraged MOT for volatility surface fitting and stochastic volatility model calibration [25, 17, 16]. For example, Demarch and Labordere formulated volatility surface fitting as a MOT problem using entropic regularization, while Guyon addressed the joint calibration of SPX and VIX smiles. Guo, Loeper, and Wang extended these ideas to a general setting, enabling calibration with arbitrary cost functionals and constraints. The flexibility of the MOT framework allows for the incorporation of additional market observable information, such as intermediate marginals or observable non vanilla instruments, to further tighten calibration and pricing bounds [22, 32].

1.2.2 Robust Pricing and Hedging

Beyond calibration, MOT has emerged as a powerful tool for robust pricing and hedging of exotic derivatives. The pioneering work of Hobson [19] used the Skorokhod embedding problem to derive tight, model-independent price bounds for complex derivatives. In practice, however, price bounds generated by classical optimal transport are often too wide to be useful, motivating the imposition of additional constraints such as martingality and observable market data.

The dual formulation of the MOT problem is particularly significant for hedging, as dual attainment corresponds to the existence of optimal sub- or super-hedging strategies using tradable instruments. Establishing dual attainment thus provides theoretical justification for robust hedging approaches and informs the construction of practical hedging portfolios.

Most research to date has focused on single-asset, two-period payoffs due to their theoretical tractability. However, many popular structured products in the market are multi-asset and path-dependent. Extending dual attainment results to the vectorial, multi-marginal setting, as in this work, provides a foundation for robust pricing and hedging of such exotics.

1.2.3 Multi-Asset Path-Dependent Products

The MOT framework is especially well-suited for pricing and hedging multi-asset, path-dependent derivatives, such as autocallables, worst-of options, and basket products. These instruments often depend on the joint evolution of several underlyings over multiple time periods, and their payoffs are sensitive to both marginal distributions and the dependence structure among assets.

A key advantage of the MOT approach is its minimal reliance on assumptions about asset correlations or joint dynamics. By only requiring consistency with observed marginals and the martingale property, MOT produces price bounds that reflect genuine model uncertainty. These bounds can be further tightened by incorporating additional market information, such as prices of liquid correlation-sensitive products or intermediate marginals. For example, the pricing bounds generated from the vectorial multi-marginal MOT framework can be interpreted as the range of model uncertainty due to unknown correlations. By introducing constraints based on market prices of pairwise correlation products (e.g., call vs. call or put vs. put), practitioners can further refine these bounds. This flexibility opens the door to robust pricing and risk management for a wide variety of exotic products.

In summary, the martingale optimal transport framework offers a robust foundation for a wide range of financial applications, from model calibration to pricing and hedging of complex derivatives. In the following

section, we turn to the numerical implementation and results that demonstrate the practical effectiveness of these methods.

2 Problem Formulation

This section introduces the mathematical framework for martingale optimal transport in multi-asset, multi-period markets. We first describe the market setting and the available marginal information, then formulate the vectorial martingale optimal transport (VMOT) problem, and finally present its dual formulation together with its financial interpretation.

2.1 Market Setting and Marginal Information

Consider a financial market with d underlying assets, whose price processes are denoted by $(X_{t,i})_{t \geq 0}$ for each asset $i \in [d] := \{1, 2, \ldots, d\}$. We do not assume that the joint law of all asset prices—often referred to as the *market model*—is known, since such information cannot be fully inferred from observable market data.

Instead, following the classical argument of Breeden and Litzenberger [6], we assume that the market reveals the marginal distribution of each asset price at any fixed maturity $t > 0$. That is, for each t and i, the distribution $\mathrm{Law}(X_{t,i}) = \mu_{t,i} \in \mathcal{P}(\mathbb{R})$ is observed, where $\mathcal{P}(\mathcal{X})$ denotes the set of probability measures on a space $\mathcal{X}$. We consider a finite collection of maturities $0 < T_1 < T_2 < \cdots < T_N$, and for notational convenience write $X_{t,i} := X_{T_t,i}$ and $X_t := (X_{t,1}, \ldots, X_{t,d})$. Throughout the paper, we assume $(X_t)_{t \in [N]}$ is an $\mathbb{R}^d$-valued martingale, consistent with the standard risk-neutral framework in mathematical finance. While the full joint distribution $\mathrm{Law}(X_1, \ldots, X_N)$ remains unspecified, we assume that the collection of Nd marginal distributions $\mu_{t,i} := \mathrm{Law}(X_{t,i})$ is fixed and known. This motivates the following admissible class.

Definition 2.1 (Vectorial Martingale Transports). *Assume the marginal measures $(\mu_{t,i})_{t,i}$ have finite first moments. Let $\mu_t := (\mu_{t,1}, \ldots, \mu_{t,d})$ and $\mu := (\mu_1, \ldots, \mu_N)$. The set of* vectorial martingale transports *is defined by*

$$\mathrm{VMT}(\mu) := \Big\{ \pi \in \mathcal{P}(\mathbb{R}^{Nd}) \,\Big|\, \pi = \mathrm{Law}(X), \quad \mathbb{E}_\pi[X_{t+1} \mid X_t] = X_t, \tag{2.1}$$

$$\mathrm{Law}(X_{t,i}) = \mu_{t,i} \text{ for all } t \in [N], \ i \in [d] \Big\}.$$

2.2 Vectorial Martingale Optimal Transport (VMOT)

Given a measurable cost (or payoff) function $c : \mathbb{R}^{Nd} \to \mathbb{R}$, the vectorial martingale optimal transport (VMOT) problem is defined as

$$\min_{\pi \in \mathrm{VMT}(\mu)} \mathbb{E}_\pi[c(X)] \qquad \text{or} \qquad \max_{\pi \in \mathrm{VMT}(\mu)} \mathbb{E}_\pi[c(X)]. \tag{2.2}$$

In a financial interpretation, $c(X)$ represents the payoff of a path-dependent derivative determined by the entire price trajectory $X = (X_t)_{t=1}^N$. Since the true market model π is unknown, all martingale measures consistent with the observed marginal information must be considered. The minimum and maximum values in (2.2) therefore correspond to the lower and upper arbitrage-free bounds for the derivative price.

A key distinction between VMOT and classical optimal transport lies in the martingale constraint $\mathbb{E}_\pi[X_{t+1} \mid X_t] = X_t$. This constraint imposes that, for each asset i and time t, the marginal distributions satisfy the *convex order* condition

$$\mu_{t,i} \preceq_c \mu_{t+1,i},$$

meaning that $\mu_{t,i}(f) \leq \mu_{t+1,i}(f)$ for all convex functions f on $\mathbb{R}$, where $\mu(f) := \int f(x) d\mu(x)$. Strassen's theorem [33] guarantees that this condition is both necessary and sufficient for $\mathrm{VMT}(\mu)$ to be nonempty. Accordingly, we assume $\mu_{t,i} \preceq_c \mu_{t+1,i}$ for all $t < N$ and $i \in [d]$ throughout the paper.

4

2.3 Dual Problem and Financial Interpretation

The VMOT problem is an infinite-dimensional linear program and therefore admits a natural dual formulation. When (2.2) is a minimization problem, the dual problem takes the form

$$\sup_{(\phi,h)\in\Psi} \mu(\phi), \tag{2.3}$$

where $\phi = (\phi_1,\ldots,\phi_N)$ with $\phi_t = (\phi_{t,1},\ldots,\phi_{t,d})$, each $\phi_{t,i} : \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$, and

$$\mu(\phi) := \sum_{t=1}^{N}\sum_{i=1}^{d} \int \phi_{t,i}\, d\mu_{t,i}.$$

The process $h = (h_1,\ldots,h_N)$, with $h_N \equiv 0$, consists of functions $h_{t,i} : \mathbb{R}^{td} \to \mathbb{R}$ representing dynamic trading strategies. The admissible class Ψ consists of all (ϕ,h) such that $\phi_{t,i} \in L^1(\mu_{t,i})$, $h_{t,i}$ is bounded, and the following pathwise inequality holds:

$$\sum_{t=1}^{N}\sum_{i=1}^{d} \phi_{t,i}(x_{t,i}) + h_{t,i}(x_1,\ldots,x_t)\,(x_{t+1,i} - x_{t,i}) \leq c(x), \tag{2.4}$$

for all $x = (x_1,\ldots,x_N) \in \mathbb{R}^{Nd}$.

For any $\pi \in \text{VMT}(\mu)$ and $(\phi,h) \in \Psi$, the martingale property implies

$$\mu(\phi) = \mathbb{E}_\pi\left[\sum_{t=1}^{N}\sum_{i=1}^{d} \phi_{t,i}(X_{t,i}) + h_{t,i}(X_1,\ldots,X_t)\,(X_{t+1,i} - X_{t,i})\right]. \tag{2.5}$$

Thus the dual value represents the cost of a semi-static hedging strategy composed of European options ϕ and dynamic trading strategies h. The inequality (2.4) enforces that the portfolio subhedges the payoff c pathwise. When (2.2) is a maximization problem, the dual problem becomes $\inf_{(\phi,h)\in\Psi} \mu(\phi)$, with the inequality reversed (and $\phi_{t,i}$ taking on their values in $\mathbb{R} \cup \{+\infty\}$).

Having established the market framework, the VMOT formulation, and its dual interpretation, we now turn to our main theoretical results concerning duality and dual attainment, which form the foundation for robust pricing and hedging in multi-asset, multi-period markets.

3 Main Theoretical Results

In this section, we present the central theoretical contributions of the paper. We first recall the duality result for the vectorial martingale optimal transport (VMOT) problem, then discuss the challenges surrounding dual attainment and state our main theorem, and finally give an intuitive explanation of the proof strategy, emphasizing the role of irreducibility and local L^1 bounds.

3.1 Duality in Martingale Optimal Transport

Under mild assumptions on the cost function c and the marginal distributions, the primal and dual optimal values coincide (see, e.g., [13, 35]):

$$P(c) := \inf_{\pi\in\text{VMT}(\mu)} \mathbb{E}_\pi[c(X)] \tag{3.1}$$

$$= \sup_{(\phi,h)\in\Psi} \mu(\phi) =: D(c).$$

Furthermore, the primal problem is known to admit a minimizer; that is, there exists a martingale transport plan $\pi \in \text{VMT}(\mu)$ such that $\mathbb{E}_\pi[c(X)] = P(c)$. Such martingale transports represent extremal market models for asset price evolution, optimizing the expected payoff of the derivative c.

5

3.2 Dual Attainment: Existence and Challenges

In contrast to the primal problem, whose solvability follows from standard compactness arguments under mild regularity conditions on c, proving *dual attainment*—existence of an optimal dual pair (ϕ, h)—is considerably more delicate. This difficulty already appears in classical optimal transport, as exemplified by Brenier's work [7]. In the martingale setting, the additional martingale constraint significantly complicates matters. As shown in [3, 4, 5], even in the single-asset, two-period case $(d, N) = (1, 2)$, dual attainment may fail within the natural class Ψ, and positive results typically rely on careful analysis of limiting convex potentials.

This issue is not merely technical. In financial terms, dual optimizers correspond to optimal sub-/super-hedging strategies and encode important structural information about primal solutions, i.e., extremal market models. For this reason, the dual attainment problem has attracted substantial attention, though most of the literature focuses on the single-asset, two-period framework. Notable exceptions include [10, 28, 30], which treat multi-period single-asset settings, and [23, 24, 26, 15, 21, 29], which study vector-valued martingale transports mostly in two-period models.

3.3 Main Result: Dual Attainment for Multi-Asset, Multi-Period MOT

Our main theoretical result establishes dual attainment for the martingale optimal transport problem with an arbitrary number of time steps N and assets d. This extension is particularly relevant for financial applications in which derivative payoffs depend on the full price path of multiple underlyings.

Theorem 3.1. *Let $(\mu_{t,i})_{t \in [N]}$ be an irreducible sequence of marginal distributions on $\mathbb{R}$ for each $i \in [d]$. Suppose $c : \mathbb{R}^{Nd} \to \mathbb{R}$ is a lower semicontinuous cost function 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 optimizer, that is, a family of functions

$$(\phi, h) = (\phi_{t,i}, h_{t,i})_{t \in [N], \, i \in [d]}$$

satisfying the pathwise inequality (2.4), and such that, for every primal optimizer $\pi \in \mathrm{VMT}(\mu)$ solving (2.2), we have the pathwise equality

$$\sum_{t=1}^{N} \sum_{i=1}^{d} \phi_{t,i}(x_{t,i}) + h_{t,i}(x_1, \ldots, x_t)\,(x_{t+1,i} - x_{t,i}) = c(x) \quad \pi\text{-a.s.} \tag{3.2}$$

The dual optimizer need not belong to the class Ψ.

The pair (ϕ, h) in Theorem 3.1 is often called a *dual maximizer*, as it solves the dual problem (2.3) associated with the primal minimization problem (2.2). The identity (3.2) shows that the semi-static portfolio built from (ϕ, h) *replicates* the derivative c: for any primal minimizer π, the portfolio payoff agrees with $c(X)$ for π-almost every price path. The inequality (2.4) ensures that, away from primal minimizers, the same portfolio provides a pathwise subhedge of c. By symmetry, Theorem 3.1 also yields the existence of a dual *minimizer* for the maximization problem in (2.2) when c is upper semicontinuous; in that case, the resulting portfolio superhedges the payoff and replicates c along primal maximizers.

It is crucial to note that a dual optimizer is not guaranteed to lie in Ψ. Previous work has shown that the dual problem (2.3) is in general not attained within Ψ, even for $(d, N) = (1, 2)$ (see [3, 5]), unless c satisfies additional regularity assumptions [4]. While Ψ is a natural domain for formulating the dual problem, it is relatively "narrow" due to the lack of compactness in infinite dimensions. Proving dual attainment is therefore substantially harder than proving duality (3.1), which often follows from standard tools in functional analysis and convex duality. In particular, although the dual optimizer (ϕ, h) in Theorem 3.1 may lie outside Ψ, each component $\phi_{t,i}$ is real-valued $\mu_{t,i}$-a.s., and each $h_{t,i}$ is real-valued as well. Since the marginals $(\mu_{t,i})_{t,i}$ are fixed in the VMOT problem, all functions appearing in the theorem can be regarded as essentially real-valued measurable functions, with no further regularity imposed.

3.4 Intuitive Explanation and Approximating Dual Maximizers

Upgrading the duality relation (3.1) to full dual attainment requires a refined approximation argument. The key idea, originating in [3, 5] for the single-asset, two-period case, is to construct a sequence of "nice" dual candidates and then extract a pointwise limit using local compactness and irreducibility.

We call a sequence $(\phi_n, h_n)_{n \in \mathbb{N}}$ an *approximating dual maximizer* if, for each n:

- $\phi_{t,i,n}$ is real-valued, continuous, and integrable with respect to $\mu_{t,i}$, i.e., $\phi_{t,i,n} \in L^1(\mu_{t,i})$;

- $h_{t,i,n}$ is bounded and continuous (with the convention $h_{N,i,n} \equiv 0$);

- the following pathwise inequality holds:

$$\sum_{t=1}^{N} \sum_{i=1}^{d} \phi_{t,i,n}(x_{t,i}) + h_{t,i,n}(\bar{x}_t) \, \Delta x_{t,i} \; \leq \; c(x) \quad \text{for all } x \in \mathbb{R}^{Nd}, \tag{3.3}$$

  where $\bar{x}_t := (x_1, \ldots, x_t) \in \mathbb{R}^{dt}$ and $\Delta x_{t,i} := x_{t+1,i} - x_{t,i}$;

- and the dual values converge increasingly to $D(c)$:

$$\mu(\phi_n) := \sum_{t=1}^{N} \sum_{i=1}^{d} \int \phi_{t,i,n}(x_{t,i}) \, d\mu_{t,i}(x_{t,i}) \nearrow D(c) \quad \text{as } n \to \infty. \tag{3.4}$$

Adapting the strategy of [3, 5] to the multi-asset, multi-period setting, the next proposition provides pointwise convergence of a subsequence of (ϕ_n, h_n), which is a crucial step toward Theorem 3.1.

Proposition 3.2. *Under the assumptions of Theorem 3.1, there exists an approximating dual maximizer $(\phi_n, h_n)_{n \in \mathbb{N}}$ such that, for each $t \in [N]$ and $i \in [d]$, the sequence $\phi_{t,i,n}$ converges $\mu_{t,i}$-a.s. to a real-valued function $\phi_{t,i}$ as $n \to \infty$.*

The proof of Proposition 3.2 relies on a structural property of the marginals, namely irreducibility in convex order, together with a local L^1 bound. We briefly recall the relevant notion (see [5] for details). A pair of probability measures $\mu \preceq_c \nu$ (with finite first moments) on $\mathbb{R}$ is called *irreducible* if the open set

$$I := \{x \in \mathbb{R} : u_\mu(x) < u_\nu(x)\}$$

is a connected interval and $\mu(I) = \mu(\mathbb{R})$, where

$$u_\mu(x) := \int_{\mathbb{R}} |x - y| \, d\mu(y)$$

is the potential function of μ. In this case, we define the *domain* (I, J) of (μ, ν) by letting J be the smallest interval with $\nu(J) = \nu(\mathbb{R})$. Then J is the union of I and any endpoints of I that are atoms of ν, and one has $I = \text{int}(J)$. The interval J may be bounded or unbounded, and may take the form $(a, b]$, $[a, b)$, (a, b), or $[a, b]$, with $I = (a, b)$ in all cases. Intuitively, irreducibility expresses that ν is "spread out" from μ in a regular way. It is a natural and generic property: most pairs of distributions in convex order satisfy irreducibility, and pairs that do not can typically be perturbed slightly to obtain it.

Let $(I_{t,i}, J_{t,i})$ denote the domain of the irreducible pair $\mu_{t,i} \preceq_c \mu_{t+1,i}$ for $t \in [N-1]$, with the conventions $J_{0,i} := J_{1,i}$ and $I_{N,i} := J_{N-1,i}$. We have $J_{t,i} \subset I_{t+1,i}$ for all $t \in [N-1]$ and $i \in [d]$. The next lemma provides a local L^1 bound on the approximating potentials.

Lemma 3.3. *Under the assumptions of Theorem 3.1, there exists an approximating dual maximizer $(\phi_n, h_n)_{n \in \mathbb{N}}$. Moreover, for each $t \in [N-1]$ and $i \in [d]$, there exists an increasing sequence of compact intervals $(J_{t,i,k})_{k \in \mathbb{N}}$ such that $\bigcup_{k \in \mathbb{N}} J_{t,i,k} = J_{t,i}$ and*

$$\sup_n \left\| \phi_{t,i,n} - \int \phi_{t,i,n} \, d\mu_{t,i,k} \right\|_{L^1(\mu_{t,i,k})} \; \leq \; C_k, \tag{3.5}$$

where $\mu_{t,i,k} := \frac{1}{\mu_{t,i}(J_{t-1,i,k})} \mu_{t,i}\big|_{J_{t-1,i,k}}$ is the normalized restriction of $\mu_{t,i}$ to $J_{t-1,i,k}$, and C_k depends only on k (not on n).

The local L^1 bound in Lemma 3.3 is a key ingredient for extracting pointwise convergent subsequences and thus for proving Proposition 3.2 and Theorem 3.1. Beyond its technical role, it is also of independent interest in the analysis of martingale optimal transport with general cost functions. Detailed proofs of Proposition 3.2 and Lemma 3.3 are given in Section 7.

These duality and dual attainment results establish a rigorous foundation for robust pricing and hedging in multi-asset, multi-period markets. In the following sections, we show how they translate into practical tools for model calibration, risk management, and the pricing of complex path-dependent derivatives.

4 Numerical Methods and Implementation

As discussed in Section 1.2, the martingale optimal transport (MOT) framework has emerged as a powerful tool in quantitative finance, enabling both model calibration to observed market marginals and robust pricing and hedging of complex, multi-asset derivatives. In particular, MOT offers a non-parametric, model-independent methodology that naturally accommodates path-dependence and requires only minimal assumptions on asset dynamics. Moreover, it allows practitioners to incorporate additional market information—such as intermediate marginals or correlation-sensitive products—to further tighten price bounds. These theoretical advances, particularly in the multi-marginal and multi-asset setting, underscore the importance of scalable and accurate numerical methods. Motivated by these applications, we now turn to the numerical implementation that make the solution of high-dimensional MOT problems feasible in practice.

In this section, we detail the computational strategies employed to solve the multi-marginal MOT problem, emphasizing the challenges posed by high-dimensionality and path-dependent constraints. We introduce the primal-dual linear programming (PDLP) approach, describe its algorithmic enhancements, and present the discrete formulation of the MOT problem suitable for large-scale numerical resolution.

4.1 Computational Challenges

The numerical resolution of the martingale optimal transport (MOT) problem has garnered considerable attention in recent literature. In the multi-marginal case, the direct solution of the underlying linear programming (LP) formulation of MOT suffers from the curse of dimensionality, as the computational complexity grows exponentially with the number of time periods and the number of dimensions of the process.

Conventional methods for solving MOT include entropy-regularized approaches, most notably the Sinkhorn algorithm, and neural network (NN) based methods. The Sinkhorn algorithm has been successfully applied to the primal formulation of optimal transport problems by exploiting Bregman projections to solve the regularized linear program [11]. De March [12] proposes a modified Sinkhorn algorithm for bimarginal MOT, which replaces the standard Bregman projections with an approximate projection procedure that iteratively enforces the martingale constraints through a fixed-point update mechanism, thereby achieving adequate accuracy while mitigating the numerical instabilities inherent in the exact projection computations. However, significant challenges arise when extending this approach to the multi-marginal MOT framework. Moreover, the Bregman projections essential to the Sinkhorn algorithm are problematic to compute under the martingale constraints; indeed, even in the bimarginal case, these projections are solved only approximately, which compromises accuracy. Additionally, the iterative nature of Sinkhorn renders it susceptible to significant numerical instability when handling the intricate constraints imposed by the martingale requirement.

Neural network methods, which have been developed to tackle the dual formulation of the MOT problem [14], offer a flexible approach to handle high-dimensional data. However, these methods currently do not offer theoretical guarantees regarding convergence or precision, which limits their reliability in scenarios where rigorous accuracy is paramount.

Given these challenges, there is a need for scalable, robust algorithms that can efficiently handle the high-dimensional, sparse and path-dependent structure of multi-marginal MOT problems.

4.2 PDLP Algorithm and Enhancements

To address these computational challenges, we employ the primal-dual linear programming (PDLP) framework [1, 18], which is well-suited for large-scale linear programs. PDLP is a first-order method that combines a saddle-point reformulation of the LP with adaptive algorithmic enhancements. It reformulates the standard LP as a saddle-point problem and iteratively updates primal and dual variables using matrix-vector products, avoiding explicit matrix storage and factorization. Its inherent matrix-free nature and reliance solely on matrix-vector products make PDLP particularly well-suited to the high-dimensional and sparse settings often encountered in MOT problems.

In brief, PDLP begins by rewriting the standard LP formulation

$$
\begin{aligned}
&\min_{x \in \mathbb{R}^n} \; c^\top x, \\
&\text{s.t. } Ax = b, \quad x \geq 0,
\end{aligned}
\tag{4.1}
$$

into an equivalent saddle-point (or primal-dual) problem:

$$
\min_{x \in X} \max_{y \in Y} \; L(x,y) \quad \text{with} \quad L(x,y) = c^\top x - y^\top Ax + b^\top y,
\tag{4.2}
$$

where $X = \mathbb{R}^n_{\geq 0}$ and $Y = \mathbb{R}^m$. The *primal-dual hybrid gradient* (PDHG) scheme [8] is then applied by iteratively performing the updates

$$
x_{k+1} = \text{proj}_X \left(x_k - \tau \left(c - A^\top y_k \right) \right),
\tag{4.3}
$$

$$
y_{k+1} = y_k + \sigma \left(b - A \left(2x_{k+1} - x_k \right) \right),
\tag{4.4}
$$

where τ and σ denote the primal and dual step sizes, respectively.

PDLP enhances the baseline PDHG method by incorporating several key features:

- **Adaptive Step Size.** An adaptive heuristic adjusts the step size dynamically to guarantee that the step condition

$$
\tau, \sigma \leq \frac{\|z_{k+1} - z_k\|_\omega^2}{2 \left(y_{k+1} - y_k \right)^\top A (x_{k+1} - x_k)}
$$

 is satisfied. This avoids overly conservative fixed-step estimates (e.g., $\tau = \sigma = 1/\|A\|^2$) and accelerates convergence.

- **Adaptive Restarts.** Periodic restarts based on the normalized duality gap (which remains finite even when the standard duality gap is unbounded) ensure that progress is regularly "reset" to counteract self-inhibiting tailing-off, thereby leading to empirical linear convergence.

- **Primal Weight Updates.** During restarts, a weight parameter is updated to balance the progress in the primal and dual variables. In effect, the update aims to equalize the weighted distances to optimality in both spaces.

- **Presolving and Diagonal Preconditioning.** Prior to the main iterations, problem data are simplified using presolve routines and then rescaled via a diagonal preconditioner. Such scaling helps to alleviate numerical imbalances, thereby reducing the effective condition number of the data matrix.

From a computational complexity perspective, recent analyses have established rigorous bounds for the PDLP algorithm in both special and general cases. While PDLP achieves particularly strong complexity guarantees for totally unimodular constraint matrices [18], it is important to note that total unimodularity of the constraint matrix A is a property specific to the classical bimarginal optimal transport (OT) problem. In contrast, the constraint matrices encountered in multimarginal OT and in bi- and multi-marginal martingale optimal transport (MOT) problems generally lack total unimodularity due to the more complex structure of the marginal and martingale constraints.

9

For general linear programs, PDLP requires at most $O\left(\frac{L^2 R^2}{\epsilon^2}\right)$ matrix-vector multiplications to reach an ϵ-optimal solution [1], where L is a Lipschitz constant related to the problem data, R is a bound on the feasible region, and ϵ is the desired accuracy. Although this worst-case bound is less favorable than in the unimodular case, PDLP remains highly effective in practice for large, sparse, and structured problems such as those encountered in multi-marginal MOT. Its matrix-free implementation and scalability make it a robust choice for high-dimensional optimal transport computations.

This general complexity result underscores the practical viability of PDLP for MOT applications, even when the underlying linear program does not possess special structure. The algorithm's ability to exploit sparsity and avoid explicit matrix factorizations is particularly advantageous in the high-dimensional, path-dependent setting of multi-marginal MOT.

4.3 Discrete Multimarginal MOT Formulation

For numerical implementation, we discretize the multi-marginal MOT problem over a finite grid of asset prices and time steps. Specifically, we consider a MOT problem that spans multiple assets and time steps, defined over a discrete support. The discrete formulation is given by the following linear program:

Problem 4.1 (discrete multimarginal MOT)**.**

$$
\begin{aligned}
\min_{\pi \geq 0} \quad & \langle C, \pi \rangle \\
subject\ to \quad & \mathrm{proj}_{(t,k)}(\pi) = \mu_{t,k}, \quad \forall t \in [T],\ k \in [d], \\
& \mathbb{E}_\pi\left[s_{t+1,k} - s_{t,k} \mid \mathcal{F}_t\right] = 0, \quad \forall t \in [T-1],\ k \in [d],
\end{aligned}
\tag{multiMOT}
$$

where $C = [C_{i_{1,1},\ldots i_{T,d}}]_{i_{1,1}\in[n_{1,1}],\ldots,i_{T,d}\in[n_{T,d}]}$ denotes the non-negative cost tensor with rank $m = Td$, $\mu_{t,k} = [\mu_{t,k,i}]_{i\in[n_{t,k}]}$ is the probability vector representing the marginal of the k-th process at the t-th time step.

To account for discretization error, we relax the martingale constraints by introducing a tolerance parameter corresponding to the grid size:

Problem 4.2 (discrete multimarginal MOT with relaxed martingale constraints)**.**

$$
\begin{aligned}
\min_{\pi \geq 0} \quad & \langle C, \pi \rangle \\
subject\ to \quad & \mathrm{proj}_{(t,k)}(\pi) = \mu_{t,k}, \quad \forall t \in [T],\ k \in [d], \\
& \mathbb{E}_\pi\left[s_{t+1,k} - s_{t,k} \mid \mathcal{F}_t\right] \leq \frac{\Delta_{t,k}}{2}, \quad \forall t \in [T-1],\ k \in [d], \\
& \mathbb{E}_\pi\left[s_{t+1,k} - s_{t,k} \mid \mathcal{F}_t\right] \geq -\frac{\Delta_{t,k}}{2}, \quad \forall t \in [T-1],\ k \in [d],
\end{aligned}
\tag{multiMOTrelaxed}
$$

where $C = [C_{i_{1,1},\ldots i_{T,d}}]_{i_{1,1}\in[n_{1,1}],\ldots,i_{T,d}\in[n_{T,d}]}$ denotes the non-negative cost tensor with rank $m = Td$, $\mu_{t,k} = [\mu_{t,k,i}]_{i\in[n_{t,k}]}$ is the probability vector representing the marginal of the k-th process at the t-th time step.

This discrete formulation enables the use of efficient large-scale linear programming solvers, such as the PDLP implementation in NVIDIA cuOpt, to compute optimal transport plans and corresponding dual certificates under realistic market constraints. Leveraging this computational framework, we now present the results of our numerical experiments, including solution quality metrics, illustrative case studies, and an analysis of both primal and dual solutions.

5 Numerical Experiments and Results

In this section, we present the results of our numerical experiments for the MOT problem. We begin by describing the metrics used to evaluate solution quality, detail the construction of marginal data from market information, analyze the computed primal and dual solutions, and provide a case study on a worst-of autocallable option. We conclude with a discussion of the implications and limitations of our findings.

5.1 Metrics for solution quality

To rigorously assess the quality and reliability of the computed solutions to the MOT problem, we employ a suite of quantitative metrics that capture both optimality and feasibility aspects. These metrics are designed to provide a comprehensive evaluation of the numerical performance of the PDLP approach under the high-dimensional and path-dependent constraints characteristic of the MOT framework.

The primary metric is the *primal objective value*, which represents the expected cost (or payoff) under the optimal transport plan for the given cost tensor. Complementing this, the *dual objective value* aggregates the contributions from the dual variables associated with the marginal constraints, weighted by the prescribed marginals. The difference between these two values, known as the *duality gap*, serves as a certificate of optimality: a small gap indicates that the computed solution is close to optimal, while a larger gap may signal numerical issues or suboptimal convergence.

Feasibility with respect to the problem constraints is evaluated through measures of *primal infeasibility* and *dual infeasibility*. Primal infeasibility quantifies the deviation of the computed transport plan from the prescribed marginal distributions, while dual infeasibility measures the extent to which the dual variables fail to satisfy the dual constraints imposed by the cost structure and the martingale conditions.

To summarize these deviations, we report the ℓ_p-norms of the infeasibility vectors, including ℓ_1, ℓ_2, and ℓ_∞ norms. These norms provide insight into the aggregate, average, and worst-case violations, respectively, and are useful for benchmarking solver performance across different problem instances and parameter regimes.

For clarity, the key metrics are:

- **Primal objective value** (V_p): Expected cost or payoff under the optimal transport plan.

- **Dual objective value** (V_d): Aggregate value from dual variables, weighted by the prescribed marginals.

- **Duality gap** (G): Difference between primal and dual objective values, indicating solution optimality.

- **Primal infeasibility** (δ^p): Deviation of the computed transport plan from the prescribed marginal distributions.

- **Dual infeasibility** (δ^d): Extent to which dual variables fail to satisfy dual constraints.

- **ℓ_p-norms of infeasibility**: Aggregate (ℓ_1), average (ℓ_2), and worst-case (ℓ_∞) violations of constraints.

Together, these metrics enable a transparent and robust evaluation of both the accuracy and feasibility of the computed solutions, facilitating meaningful comparisons with alternative methods and guiding further algorithmic improvements.

5.2 Case study: worst-of autocallable option

To showcase the practical relevance of our approach, we consider a worst-of autocallable option referencing the S&P 500 and NASDAQ 100 indices. Autocallable options are a prominent class of structured products in finance, widely traded in equity-linked markets due to their attractive risk-return profiles and embedded path-dependent features. These instruments offer periodic opportunities for early redemption (autocall), contingent on the performance of one or more underlying assets relative to specified barriers. Their popularity stems from the ability to generate enhanced yields in low-volatility environments, provide partial downside protection, and facilitate tailored risk exposures for both retail and institutional investors. The liquidity and market depth of autocallable products, especially those linked to major equity indices, make them a natural choice for robust pricing and risk management studies. Moreover, the worst-of variant, where the payoff depends on the least-performing asset in a basket, is particularly relevant for risk transfer and capital protection strategies.

Definition 5.1 (Multi-Asset Autocall Payoff)**.** *Let $\boldsymbol{S}(t) := (S_1(t), S_2(t), \ldots, S_d(t))$ for $t \in [0, T]$ denote the vector of stochastic processes representing the prices of d underlying assets. Consider a sequence of observation dates $0 < t_1 < t_2 < \cdots < t_m = T$, where barrier and knock-out conditions are evaluated and*

coupons are accrued. Let K be the strike price, B_{KI} the knock-in barrier, B_{KO} the knock-out barrier, and C_j the coupon accrued over $[t_{j-1}, t_j]$ for $j \in [m]$ (with $t_0 := 0$).

Define the worst-performing asset at time t as

$$X(t) := \min_{i \in [d]} S_i(t).$$

Define the knock-out stopping time as

$$\tau := \inf \left\{ j \in [m] : X(t_j) \geq B_{\mathrm{KO}} \right\},$$

with the convention that $\inf \varnothing = +\infty$.

The payoff at time t_k for $k \in [m]$ of a worst-of multi-asset autocall is given by

$$
f_k \left(\boldsymbol{S}(t_1), \ldots, \boldsymbol{S}(t_k) \right) =
\begin{cases}
\sum_{j=1}^{\tau} C_j, & \text{if } k = \tau < m, \\
\min\{X(T) - K, 0\}, & \text{if } k = m \leq \tau \text{ and } X(T) \leq B_{\mathrm{KI}}, \\
\sum_{j=1}^{m} C_j, & \text{if } k = m \leq \tau \text{ and } X(T) > B_{\mathrm{KI}}, \\
0, & \text{otherwise, i.e., } k < \min\{m, \tau\}.
\end{cases}
$$

In our specific application, we consider a worst-of autocallable structure linked to two major equity indices: the S&P 500 (SPX) and the NASDAQ 100 (NDX), corresponding to $d = 2$ underlying assets. The product features a three-year maturity, with an inception date $t_{\mathrm{inc}} = -\frac{5}{24}$ years and three annual monitoring dates: $t_1 = t_{\mathrm{inc}} + 1$, $t_2 = t_{\mathrm{inc}} + 2$, and $t_3 = t_{\mathrm{inc}} + 3$ years. The knock-out barrier is set at $B_{\mathrm{KO}} = 120\%$ of the initial level, and the knock-in barrier at $B_{\mathrm{KI}} = 60\%$. The strike price is $K = 100\%$. The coupon is a fixed rate, with an annualized rate of 8%, accrued at each period, i.e., $C_j = 8\% \times (t_j - t_{j-1})$ for $j \in \{1, 2, 3\}$. We perform present value calculations as of $t_0 = 0$, assuming a constant discount rate of 1%.[1]

It is important to note that the payoff function described above is defined *per unit notional*. In practical applications, the final payout to the investor is obtained by multiplying the computed payoff by the notional amount of the contract.

At each monitoring date, the reference level is determined by the worst-performing underlying between SPX and NDX. For example, suppose at $t_1 = t_{\mathrm{inc}} + 1$ years, SPX is at 125% and NDX is at 130% of their respective initial levels. The worst-performing asset is SPX, which exceeds the 120% knock-out barrier, triggering the autocall feature. The product is then redeemed early, and the holder receives the notional plus the accrued coupon up to that date.

If neither asset breaches the knock-out barrier at any monitoring date, the product continues to maturity. At $t_3 = t_1 + 3$ years, if the worst-performing asset is below the knock-in barrier (60%), the holder receives a down-and-in put payoff, $\min\{X(T) - K, 0\}$, reflecting the loss relative to the strike. If the worst-performing asset remains above the knock-in barrier, the holder receives the full coupon accrued over the three years.

This structure provides a realistic and challenging test case for the martingale optimal transport framework, as the payoff is highly path-dependent and sensitive to joint asset dynamics. The dimensionality of the problem is $dm = 2 \times 3 = 6$ for the two assets and three monitoring dates, making it tractable for numerical solution via large-scale linear programming. In the following sections, we present numerical results illustrating dual attainment and optimal transport plans for this worst-of autocallable payoff, computed using PDLP solvers under the prescribed marginal and martingale constraints.

5.3 Marginal data construction

To construct the marginal distributions required for the multi-period, multi-asset martingale optimal transport (MOT) problem, we utilize market data in the form of vanilla option prices for the S&P 500 (SPX) and

---

[1]In practice, the *inception date* refers to the date when the product is issued and the initial asset levels are fixed, while the *evaluation date* is the date at which the product is being valued or analyzed, typically for risk management or pricing purposes. The separation allows for the modeling of products that have already been running for some time, and for the use of current market data in valuation.
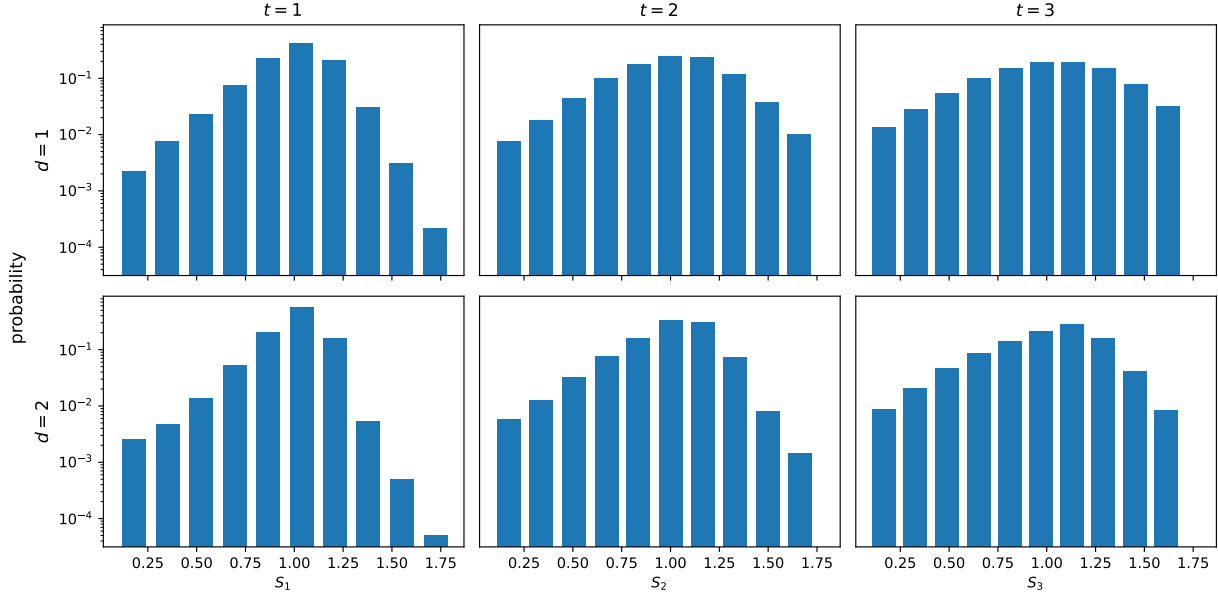
Figure 1: Discrete marginal distributions used as input to the 2-dimensional, 3-time-step MOT problem. Asset prices are scaled by their respective spot prices at t_0.

NASDAQ 100 (NDX) indices, evaluated as of December 5, 2025 (t_0). These option prices are used to infer the risk-neutral marginal distributions of the underlying asset prices at the three monitoring dates specified in the autocallable payoff definition (see Definition 5.1).

For each asset and each monitoring date, the marginal distribution is discretized over a support of $n = 10$ points, chosen to adequately capture the range of plausible price scenarios implied by the market. The support points are scaled by the respective spot prices at the evaluation date (t_0), ensuring comparability and normalization across assets and time steps.

The construction of these marginals from market data follows a standard procedure in quantitative finance. First, a smooth implied volatility surface is calibrated to the observed vanilla option prices for each asset and maturity. Then, using the Breeden–Litzenberger formula [6], the risk-neutral probability density function is extracted by taking the second derivative of the call price with respect to strike. The resulting densities are discretized to form the input marginals for the MOT problem.

The resulting discrete marginal distributions for SPX and NDX at all three monitoring dates are visualized in Figure 1. These marginals serve as input constraints for the MOT problem, enabling robust pricing and hedging of the worst-of autocallable payoff under the prescribed market information.

5.4 Discussion of results

In our numerical experiments, we employ the PDLP solver implemented within NVIDIA cuOpt [9], running on GPU hardware, to solve the discrete multi-marginal MOT problems. Leveraging GPU acceleration enables efficient large-scale computation and significantly reduces solution times for high-dimensional instances. The solver is configured with an optimality tolerance of 10^{-12}, ensuring that the primal and dual infeasibilities, as well as the optimality gap, are all maintained below this threshold in both absolute and relative terms. This stringent tolerance guarantees that the computed solutions are not only highly accurate but also robust with respect to the complex constraints inherent in the MOT framework.

The primal solution, represented by the optimal transport plan π, encodes the joint probability distribution over asset prices and time steps that minimizes (or maximizes) the expected payoff under the specified

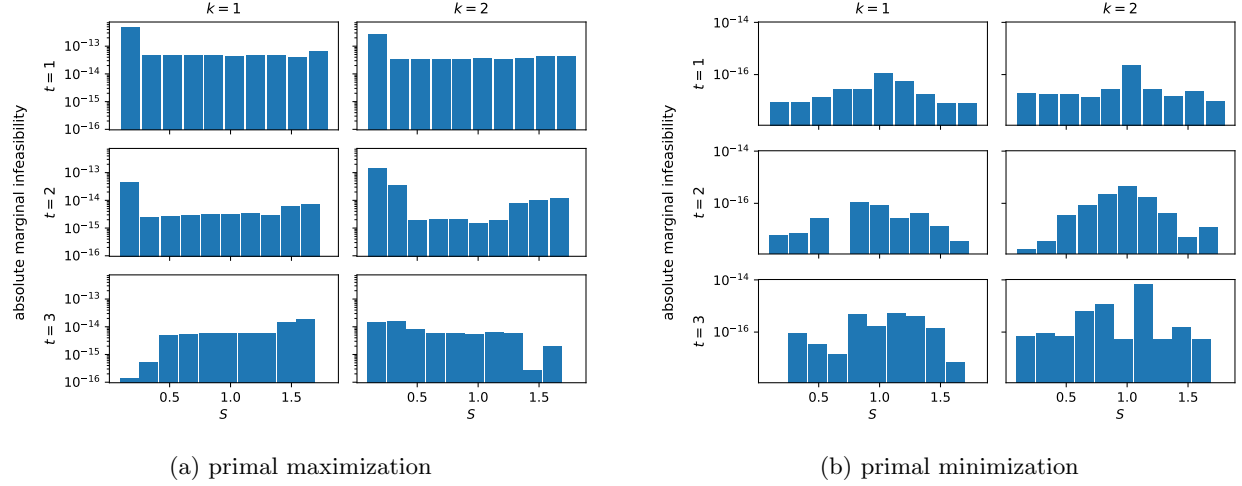(a) primal maximization

(b) primal minimization

Figure 2: Absolute infeasibility with respect to the marginal constraints in the primal solution of the 2-dimensional, 3-time-step MOT problem. Results are shown for both maximization and minimization directions, computed using the PDLP solver in NVIDIA cuOpt with an optimality tolerance of 10^{-12}. The low levels of infeasibility confirm the solver's ability to enforce marginal constraints with high precision.

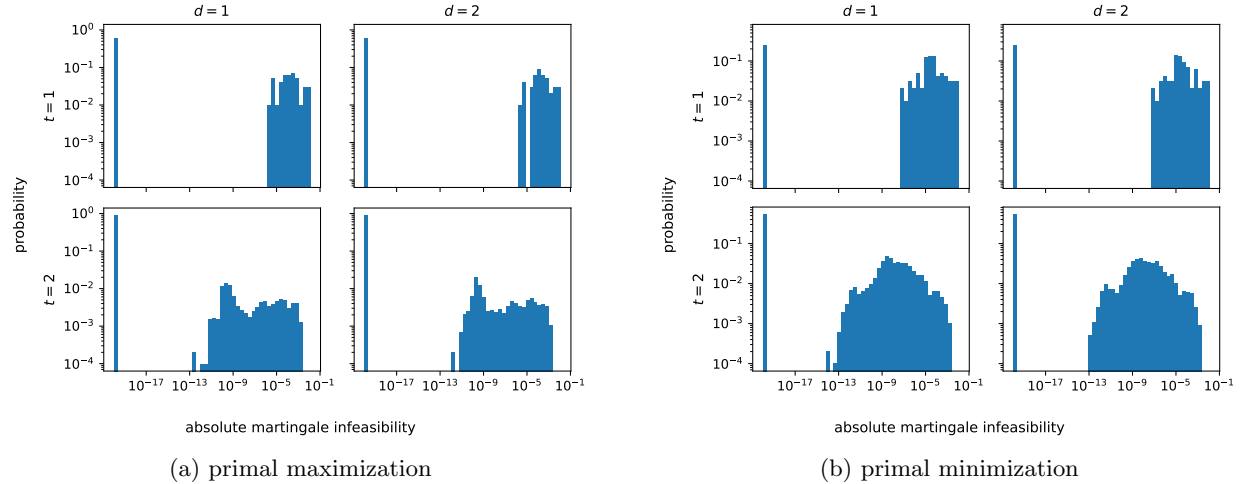(a) primal maximization

(b) primal minimization

Figure 3: Absolute infeasibility with respect to the martingale constraints in the primal solution of the 2-dimensional, 3-time-step MOT problem. The PDLP solver in cuOpt maintains infeasibility below the grid discretization error for both optimization directions, demonstrating robust enforcement of path-dependent martingale constraints.

cost function, subject to the marginal and martingale constraints. Figures 2 and 3 illustrate the infeasibility of the primal solution with respect to the marginal and martingale constraints, respectively, for both the maximization and minimization problems. The results demonstrate that the PDLP solver achieves high accuracy, with infeasibility levels well within the prescribed tolerances. In particular, the relaxed martingale constraints are satisfied within the grid discretization error, confirming the effectiveness of the presolve and preconditioning steps in the numerical pipeline.

The dual solution, comprising the variables ϕ and h, provides valuable insight into the shadow prices associated with the marginal and martingale constraints. These dual variables can be interpreted as sensitiv-

14

(a) primal maximization
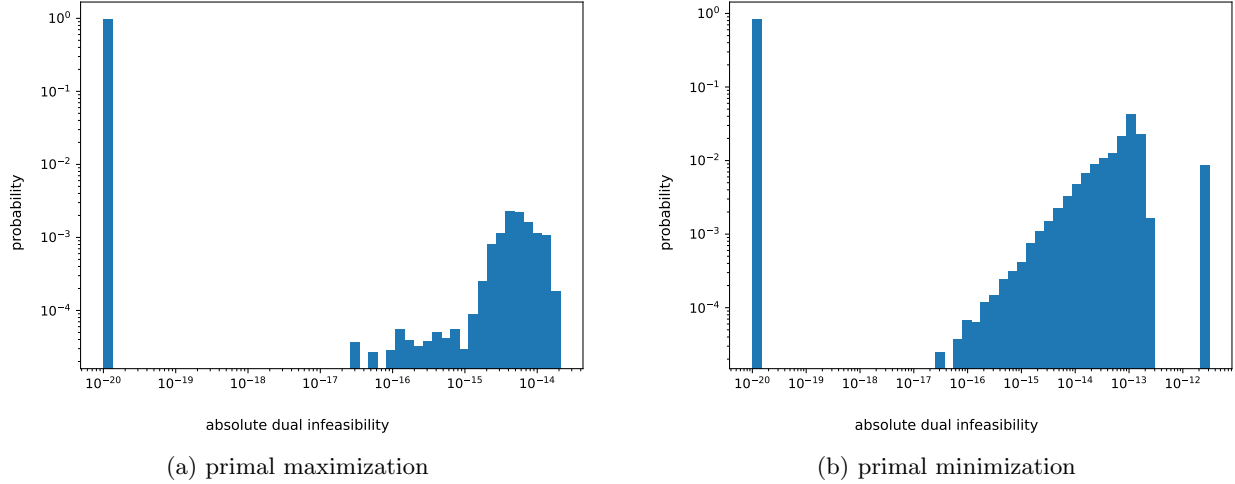

(b) primal minimization

Figure 4: Absolute infeasibility with respect to the dual constraints in the primal solution of the 2-dimensional, 3-time-step MOT problem. The results highlight the tightness of the dual bounds achieved by the PDLP solver in cuOpt, with infeasibility consistently below the prescribed tolerance.

Optimization Direction		Maximization	Minimization
Optimal Primal Objective Value		0.187880	0.062208
Optimal Dual Objective Value		0.187880	0.062208
Duality Gap		-7.7577×10^{-14}	-1.1971×10^{-13}
Primal Infeasibility	ℓ_1	4.1653×10^{-12}	1.3085×10^{-14}
	ℓ_2	1.6503×10^{-12}	7.0355×10^{-15}
	ℓ_∞	1.4217×10^{-12}	6.8279×10^{-15}
Dual Infeasibility	ℓ_1	7.0409×10^{-11}	3.7987×10^{-8}
	ℓ_2	7.7958×10^{-13}	2.8302×10^{-10}
	ℓ_∞	2.0761×10^{-14}	3.0800×10^{-12}

Table 1: Summary of numerical results for the 2-dimensional, 3-time-step MOT problem solved using the PDLP solver in NVIDIA cuOpt. The table reports optimal primal and dual objective values, duality gaps, and infeasibility norms for both maximization and minimization. All metrics are on the order of or below 10^{-8}, with the majority well below 10^{-10}, confirming the accuracy and robustness of the computed solutions.

ities of the optimal value to perturbations in the input marginals and the martingale conditions, respectively. Figure 4 presents the infeasibility of the dual solution, further corroborating the tightness of the computed bounds and the robustness of the solver.

Table 1 summarizes the key numerical results, including the optimal primal and dual objective values, duality gaps, and infeasibility norms for both maximization and minimization directions. The duality gaps are on the order of 10^{-13} or smaller, demonstrating that the solutions achieve optimality to within nearly machine precision. Primal and dual infeasibility norms are similarly small, reflecting the solver's ability to enforce the complex constraints of the MOT problem in a high-dimensional, path-dependent setting.

Overall, the computed primal and dual solutions validate the theoretical guarantees of the PDLP approach and demonstrate its practical effectiveness for large-scale, multi-asset, multi-period MOT problems. The use of GPU-accelerated computation via NVIDIA cuOpt was essential for handling the high-dimensional, path-dependent structure of these problems, enabling robust and scalable solution of MOT instances that would be infeasible on standard CPU-based solvers. These results provide a foundation for further applications in robust pricing, risk management, and model calibration in financial engineering and related fields.

6 Conclusion and Future Work

In this work, we have developed a comprehensive framework for dual attainment in the multi-marginal, multi-asset martingale optimal transport (MOT) problem. By extending duality and existence results to arbitrary numbers of assets and time periods, we have provided both theoretical and practical foundations for robust pricing and hedging of complex financial derivatives. Our analysis demonstrates that, under mild regularity and irreducibility conditions, dual optimizers exist and can be constructed, enabling the replication and sub-/super-hedging of path-dependent payoffs in multi-dimensional settings.

On the computational side, we have shown that the primal-dual linear programming (PDLP) approach, implemented within NVIDIA cuOpt and executed on GPU hardware, is highly effective for solving large-scale discrete MOT problems. Leveraging GPU acceleration was essential for handling the high-dimensional, path-dependent structure of these problems, enabling solution times and scalability that would be infeasible on standard CPU-based solvers. Our numerical experiments confirm that the PDLP solver achieves near-optimality and enforces intricate marginal and martingale constraints with high precision, even for challenging products such as worst-of autocallable options.

The martingale optimal transport framework thus offers a powerful and flexible tool for quantitative finance, supporting model calibration, robust pricing, and risk management in environments with limited model assumptions and rich market data. The demonstrated effectiveness of GPU-accelerated computation further expands the practical applicability of MOT to real-world, large-scale financial problems.

Future work   There remain several promising directions for further research in the study of multi-marginal, multi-asset martingale optimal transport. One avenue is to investigate how additional structural properties of the payoff function may be leveraged to refine theoretical bounds and enhance the robustness of pricing results. Another important direction is the development of more efficient computational techniques, particularly those that exploit problem-specific features to improve scalability and reduce dimensionality in large-scale numerical implementations. Advancing these aspects could lead to significant improvements in both the theoretical understanding and practical applicability of the MOT framework in high-dimensional financial settings. We leave these and related questions for future work.

7 Proofs of Main Results

The assumption $|c(x)| \leq \sum_{t,i} v_{t,i}(x_{t,i})$ for some continuous $v_{t,i} \in L^1(\mu_{t,i})$ ensures $P(c) = D(c)$ in (3.1) (see e.g. [35]). Clearly, a dual optimizer exists for $c(x)$ if and only if so does for $\tilde{c}(x) := c(x) - \sum_{t,i} v_{t,i}$. Thus by replacing c with $\tilde{c}$, from now on we can (and will) assume that $c \leq 0$ throughout the proofs.

Proof of Lemma 3.3. Let $(\phi_n, h_n)_{n \in \mathbb{N}}$ be an approximating dual maximizer satisfying (3.3), (3.4), where $D(c) = P(c) \in \mathbb{R}$ and $c \leq 0$ without loss of generality. Denote $\phi_{t,n}^{\oplus}(x_t) = \sum_{i=1}^{d} \phi_{t,i,n}(x_{t,i})$, $h_{t,n}(\bar{x}_t) = \big(h_{t,1,n}(\bar{x}_t), ..., h_{t,d,n}(\bar{x}_t)\big)$. Define

$$\chi_{t,n}(x_t) := \sup_{x_1, ..., x_{t-1}} \sum_{s=1}^{t-1} \big(\phi_{s,n}^{\oplus}(x_s) + h_{s,n}(\bar{x}_s) \cdot \Delta x_s\big) \tag{7.1}$$

with the convention $\chi_{1,n} = \chi_{N+1,n} \equiv 0$. Notice $\chi_{t,n}$ is a convex function on $\mathbb{R}^d$, since it is a supremum of affine functions of x_t. We claim

$$\chi_{t,n} \leq \chi_{t+1,n} - \phi_{t,n}^{\oplus} \quad \text{for all } t \in [N] \text{ and } n \in \mathbb{N}. \tag{7.2}$$

This inequality can be shown as follows:

$$\chi_{t+1,n}(x_{t+1}) = \sup_{x_1,\dots,x_t} \sum_{s=1}^{t} \left(\phi_{s,n}^{\oplus}(x_s) + h_{s,n}(\bar{x}_s) \cdot \Delta x_s \right)$$

$$\geq \sup_{\substack{x_1,\dots,x_{t-1} \\ x_t = x_{t+1}}} \sum_{s=1}^{t-1} \left(\phi_{s,n}^{\oplus}(x_s) + h_{s,n}(\bar{x}_s) \cdot \Delta x_s \right) + \phi_{t,n}^{\oplus}(x_{t+1})$$

$$= \chi_{t,n}(x_{t+1}) + \phi_{t,n}^{\oplus}(x_{t+1}),$$

while the inequality $\chi_{N,n} \leq -\phi_{N,n}^{\oplus}$ directly follows from $c \leq 0$. Let $\mu_t^{\otimes} = \mu_{t,1} \otimes \cdots \otimes \mu_{t,d}$ denote the product measure. We claim the following bound

$$\int \chi_{t,n} \, d(\mu_t^{\otimes} - \mu_{t-1}^{\otimes}) \leq C \quad \text{for all } t = 2, \dots, N \text{ and } n \in \mathbb{N}, \tag{7.3}$$

where C is independent of n throughout the proof.

To show (7.3), observe that repeated application of (7.2) gives

$$\mu_t^{\otimes}(\chi_{t,n}) \leq \mu_t^{\otimes}(\chi_{t+1,n}) - \mu_t^{\otimes}(\phi_{t,n}^{\oplus})$$

$$\leq \mu_{t+1}^{\otimes}(\chi_{t+1,n}) - \mu_t^{\otimes}(\phi_{t,n}^{\oplus})$$

$$\leq \mu_{t+1}^{\otimes}(\chi_{t+2,n}) - \mu_{t+1}^{\otimes}(\phi_{t+1,n}^{\oplus}) - \mu_t^{\otimes}(\phi_{t,n}^{\oplus})$$

$$\leq \cdots \leq -\sum_{s=t}^{N} \mu_s^{\otimes}(\phi_{s,n}^{\oplus}),$$

where the second inequality is due to $\mu_t^{\otimes} \preceq_c \mu_{t+1}^{\otimes}$ and convexity of χ. Similarly,

$$\mu_{t-1}^{\otimes}(\chi_{t,n}) \geq \mu_{t-1}^{\otimes}(\chi_{t-1,n}) + \mu_{t-1}^{\otimes}(\phi_{t-1,n}^{\oplus})$$

$$\geq \mu_{t-2}^{\otimes}(\chi_{t-1,n}) + \mu_{t-1}^{\otimes}(\phi_{t-1,n}^{\oplus})$$

$$\geq \mu_{t-2}^{\otimes}(\chi_{t-2,n}) + \mu_{t-2}^{\otimes}(\phi_{t-2,n}^{\oplus}) + \mu_{t-1}^{\otimes}(\phi_{t-1,n}^{\oplus})$$

$$\geq \cdots \geq \sum_{s=1}^{t-1} \mu_s^{\otimes}(\phi_{s,n}^{\oplus}).$$

The two inequalities, along with $\mu_{t-1}^{\otimes}(\chi_{t,n}) \leq \mu_t^{\otimes}(\chi_{t,n})$ due to $\mu_{t-1}^{\otimes} \preceq_c \mu_t^{\otimes}$, imply

$$\int \chi_{t,n} \, d(\mu_t^{\otimes} - \mu_{t-1}^{\otimes}) \leq -\sum_{s=1}^{N} \mu_s^{\otimes}(\phi_{s,n}^{\oplus}) = -\mu(\phi_n) \tag{7.4}$$

which, in conjunction with (3.4), yields (7.3) as claimed.

We can now obtain local uniform boundedness of $\{\chi_{t,n}\}_n$ using (7.3) and [21, Proposition A.1]. We need to meet the proposition's second condition. For this, fix any $a \in I_1 := I_{1,1} \times \cdots \times I_{1,d}$, and let $L_{2,n} : \mathbb{R}^d \to \mathbb{R}$ be an affine function satisfying $L_{2,n} \leq \chi_{2,n}$ and $L_{2,n}(a) = \chi_{2,n}(a)$. By linearity, we have $L_{2,n}(x_2) = \nabla L_{2,n}(x_1) \cdot (x_2 - x_1) + L_{2,n}(x_1)$. This allows us to modify the given approximating dual maximizer $(\phi_n, h_n)_{n \in \mathbb{N}}$ by replacing $\phi_{1,n}^{\oplus}(x_1)$ with $\phi_{1,n}^{\oplus}(x_1) - L_{2,n}(x_1)$, $\phi_{2,n}^{\oplus}(x_2)$ with $\phi_{2,n}^{\oplus}(x_2) + L_{2,n}(x_2)$, and $h_{1,n}(x_1)$ with $h_{1,n}(x_1) - \nabla L_{2,n}(x_1)$ ($\nabla L_{2,n}$ is constant and does not depend on x_1). Notice this yields $\chi_{2,n} \geq 0$ and $\chi_{2,n}(a) = 0$. We can continue subtracting appropriate linear functions $L_{t,n}$, $t = 2, \dots, N$, and achieve

$$\chi_{t,n} \geq 0 \quad \text{and} \quad \chi_{t,n}(a) = 0 \quad \text{for all } n \in \mathbb{N} \text{ and } t \in [N]. \tag{7.5}$$

Note that the modifications have no effect on the value $\mu(\phi_n)$.

17

Next, let $\{\epsilon_k\}_k$ be a positive decreasing sequence tending to zero as $k \to \infty$, and write $I_{t,i} =]a_{t,i}, b_{t,i}[$ where $-\infty \le a_{t,i} < b_{t,i} \le +\infty$. Then we define the compact interval $J_{t,i,k} := [c_{t,i,k}, d_{t,i,k}]$ for $t \in [N-1]$, $i \in [d]$ and $k \in \mathbb{N}$ as follows:

$$\text{If } a_{t,i} > -\infty, \text{ then define } c_{t,i,k} \text{ by the following rule:} \tag{7.6}$$

$$\mu_{t+1,i}(a_{t,i}) = 0 \Rightarrow c_{t,i,k} := a_{t,i} + \epsilon_k; \quad \mu_{t+1,i}(a_{t,i}) > 0 \Rightarrow c_{t,i,k} := a_{t,i};$$

$$\text{If } b_{t,i} < +\infty, \text{ then define } d_{t,i,k} \text{ by the following rule:}$$

$$\mu_{t+1,i}(b_{t,i}) = 0 \Rightarrow d_{t,i,k} := b_{t,i} - \epsilon_k; \quad \mu_{t+1,i}(b_{t,i}) > 0 \Rightarrow d_{t,i,k} := b_{t,i};$$

$$\text{If } a_{t,i} = -\infty, \text{ then } c_{t,i,k} := -1/\epsilon_k;$$

$$\text{If } b_{t,i} = +\infty, \text{ then } d_{t,i,k} := +1/\epsilon_k.$$

Set $J_{0,i,k} := J_{1,i,k}$. For example, if $\mu_{t+1,i}(a_{t,i}) = 0$ and $\mu_{t+1,i}(b_{t,i}) > 0$, then $J_{t,i,k} = [a_{t,i} + \epsilon_k, b_{t,i}]$. Let ϵ_1 be so small so that $\mu_{t,i}(J_{t,i,1}) > 0$, $\mu_{t+1,i}(J_{t,i,1}) > 0$ for every t, i. Observe that $J_{t,i,k} \nearrow J_{t,i}$ as $k \to \infty$. Let $J_{t,k} := J_{t,1,k} \times J_{t,2,k} \times ... \times J_{t,d,k}$. Then [21, Proposition A.1] with (7.3) and (7.5) yields that there exists $M_k \ge 0$ for each $k \in \mathbb{N}$ such that

$$0 \le \sup_{n \in \mathbb{N}} \sup_{x_t \in J_{t-1,k}} \chi_{t,n}(x_t) \le M_k. \tag{7.7}$$

Next, another repeated application of (7.2) gives

$$C \ge -\sum_{s=1}^{N} \mu_s^{\otimes}(\phi_{s,n}^{\oplus})$$

$$\ge \mu_N^{\otimes}(\chi_{N,n}) - \sum_{s=1}^{N-1} \mu_s^{\otimes}(\phi_{s,n}^{\oplus})$$

$$\ge \mu_{N-1}^{\otimes}(\chi_{N,n}) - \sum_{s=1}^{N-1} \mu_s^{\otimes}(\phi_{s,n}^{\oplus})$$

$$= \mu_{N-1}^{\otimes}(\chi_{N,n}) - \mu_{N-1}^{\otimes}(\chi_{N-1,n}) + \mu_{N-1}^{\otimes}(\chi_{N-1,n}) - \sum_{s=1}^{N-1} \mu_s^{\otimes}(\phi_{s,n}^{\oplus})$$

$$= \| \chi_{N,n} - \chi_{N-1,n} - \phi_{N-1,n}^{\oplus} \|_{L^1(\mu_{N-1}^{\otimes})} + \mu_{N-1}^{\otimes}(\chi_{N-1,n}) - \sum_{s=1}^{N-2} \mu_s^{\otimes}(\phi_{s,n}^{\oplus})$$

$$\ge \| \chi_{N,n} - \chi_{N-1,n} - \phi_{N-1,n}^{\oplus} \|_{L^1(\mu_{N-1}^{\otimes})} + \mu_{N-2}^{\otimes}(\chi_{N-1,n}) - \sum_{s=1}^{N-2} \mu_s^{\otimes}(\phi_{s,n}^{\oplus})$$

$$\ge \cdots \ge \sum_{s=2}^{N} \| \chi_{s,n} - \chi_{s-1,n} - \phi_{s-1,n}^{\oplus} \|_{L^1(\mu_{s-1}^{\otimes})}$$

where the third and sixth inequality is due to the convexity of χ with $\mu_t^{\otimes} \preceq_c \mu_{t+1}^{\otimes}$, and the fifth equality is by the nonnegativity $\chi_{t,n} - \chi_{t-1,n} - \phi_{t-1,n}^{\oplus} \ge 0$ from (7.2).

On the other hand, the nonpositivity $c \le 0$ in (3.3) yields

$$\sum_{s=1}^{N} \left(\phi_{s,n}^{\oplus}(x_s) + h_{s,n}(\bar{x}_s) \cdot \Delta x_s \right) \le \chi_{N,n}(x_N) + \phi_{N,n}^{\oplus}(x_N) \le 0$$

where the first inequality follows by taking supremum over $x_1, ..., x_{N-1}$ (recall (7.1) and the convention $h_{N,n} \equiv 0$). Integrating with any $\pi \in \text{VMT}(\mu)$ yields $\| \chi_{N,n}(x_N) + \phi_{N,n}^{\oplus} \|_{L^1(\mu_N^{\otimes})} \le -\sum_{s=1}^{N} \mu_s^{\otimes}(\phi_{s,n}^{\oplus}) \le C$. We therefore deduce

$$\| \chi_{t+1,n} - \chi_{t,n} - \phi_{t,n}^{\oplus} \|_{L^1(\mu_t^{\otimes})} \le C \quad \text{for all } n \in \mathbb{N}, t \in [N]. \tag{7.8}$$

18

Now for each $k \in \mathbb{N}$, let $\mu_{t,i,k}$ be the restriction of $\mu_{t,i}$ on $J_{t-1,i,k}$ (where $J_{0,i,k} := J_{1,i,k}$) then normalized to be a probability distribution, as defined in Lemma 3.3. Let $\mu_{t,k}^{\otimes} = \otimes_i \mu_{t,i,k}$, so that $\mu_{t,k}^{\otimes}(J_{t-1,k}) = 1$ where $J_{t-1,k} := \otimes_i J_{t-1,i,k}$. Define

$$v_{t,i,k,n} := \int \phi_{t,i,n}\, d\mu_{t,i,k}, \quad t \in [N], i \in [d], k \in \mathbb{N}, n \in \mathbb{N}.$$

To complete the proof of Lemma 3.3, that is to show for each $k \in \mathbb{N}$

$$\sup_n \| \phi_{t,i,n} - v_{t,i,k,n} \|_{L^1(\mu_{t,i,k})} \leq C, \tag{7.9}$$

observe that (7.7), (7.8) and the fact $J_{t-1,k} \subset J_{t,k}$ together imply

$$\begin{aligned}
&\| \phi_{t,n}^{\oplus} \|_{L^1(\mu_{t,k}^{\otimes})} \\
&\leq \| \chi_{t+1,n} - \chi_{t,n} - \phi_{t,n}^{\oplus} \|_{L^1(\mu_{t,k}^{\otimes})} + \| \chi_{t,n} - \chi_{t+1,n} \|_{L^1(\mu_{t,k}^{\otimes})} \\
&\leq C + M_k =: C
\end{aligned}$$

where we used $\chi_{t+1,n} - \chi_{t,n} - \phi_{t,n}^{\oplus} \geq 0$ to get the bound $\| \chi_{t+1,n} - \chi_{t,n} - \phi_{t,n}^{\oplus} \|_{L^1(\mu_{t,k}^{\otimes})} \leq C$ from (7.8). From this, we obtain the bound

$$\left| \sum_{i=1}^{d} v_{t,i,k,n} \right| \leq \| \phi_{t,n}^{\oplus} \|_{L^1(\mu_{t,k}^{\otimes})} \leq C \text{ for all } n, \tag{7.10}$$

where the first inequality is by Jensen's inequality. Next, because $\phi_{t,n}^{\oplus} \leq M_k$ on $J_{t,k}$ by (7.2), (7.5) and (7.7), by taking supremum over $J_{t,k}$, we have

$$\sum_{i=1}^{d} \sup_{x_{t,i} \in J_{t,i,k}} \phi_{t,i,n}(x_{t,i}) \leq M_k \text{ for all } n.$$

In particular, since $v_{t,i,k,n} \leq \sup_{x_{t,i} \in J_{t,i,k}} \phi_{t,i,n}(x_{t,i})$,

$$\sup_{x_{t,1} \in J_{t,1,k}} \phi_{t,1,n}(x_{t,1}) + \sum_{i=2}^{d} v_{t,i,k,n} \leq M_k. \tag{7.11}$$

Define $\hat{v}_{t,1,k,n} := -\sum_{i=2}^{d} v_{t,i,k,n}$. Since $\phi_{t,n}^{\oplus} \leq M_k$ on $J_{t,k}$, we have

$$C \geq \| M_k - \phi_{t,n}^{\oplus} \|_{L^1(\mu_{t,k}^{\otimes})} = M_k - \int (\phi_{t,1,n} + \sum_{i=2}^{d} v_{t,i,k,n})\, d\mu_{t,1,k}.$$

With (7.11), this implies that $\sup_n \| \phi_{t,1,n} - \hat{v}_{t,1,k,n} \|_{L^1(\mu_{t,1,k})}$ is bounded, and then by (7.10), $\sup_n \| \phi_{t,1,n} - v_{t,1,k,n} \|_{L^1(\mu_{t,1,k})}$ is bounded. This proves (7.9). $\square$

We turn to the proof of Proposition 3.2.

Proof of Proposition 3.2 . Lemma 3.3 and Komlós lemma (which states that every L^1-bounded sequence of real functions contains a subsequence such that the arithmetic means of all its subsequences converge almost everywhere) implies that for each $k \in \mathbb{N}$, there exists a subsequence $\{\phi_{t,i,k,n}\}_n$ of $\{\phi_{t,i,n}\}_n$ such that
(i) $\{\phi_{t,i,k+1,n}\}_n$ is a further subsequence of $\{\phi_{t,i,k,n}\}_n$, and
(ii) $\tilde{\phi}_{t,i,k,n}(x_{t,i}) - \tilde{v}_{t,i,k,n}$ converges $\mu_{t,i,k}$ - a.s. as $n \to \infty$,
where $v_{t,i,k,n} = \int \phi_{t,i,k,n}\, d\mu_{t,i,k}$, $\hat{v}_{t,1,k,n} = -\sum_{i=2}^{d} v_{t,i,k,n}$, $\tilde{v}_{t,1,k,n} = \frac{1}{n}\sum_{m=1}^{n} \hat{v}_{t,1,k,m}$, $\tilde{v}_{t,i,k,n} = \frac{1}{n}\sum_{m=1}^{n} v_{t,i,k,m}$

19

for $i \geq 2$, and $\tilde{\phi}_{t,i,k,n} = \frac{1}{n} \sum_{m=1}^{n} \phi_{t,i,k,m}$. Note that for each k, our choice of a subsequence index can be made identical for every t, i, since there are finitely many indices of t, i. Then we select the diagonal sequence

$$\Phi_{t,i,n} := \phi_{t,i,n,n}$$

and again define

$$w_{t,i,k,n} = \int \Phi_{t,i,n} d\mu_{t,i,k}, \quad \tilde{w}_{t,i,k,n} = \frac{1}{n} \sum_{m=1}^{n} w_{t,i,k,m}, \quad 2 \leq i \leq d,$$

$$\hat{w}_{t,1,k,n} = -\sum_{i=2}^{d} w_{t,i,k,n}, \quad \tilde{w}_{t,1,k,n} = \frac{1}{n} \sum_{m=1}^{n} \hat{w}_{t,1,k,m}, \quad \tilde{\Phi}_{t,i,n}(x_{t,i}) = \frac{1}{n} \sum_{m=1}^{n} \Phi_{t,i,m}(x_{t,i}).$$

We then claim:

$$\tilde{\Phi}_{t,i,n}(x_{t,i}) - \tilde{w}_{t,i,1,n} \text{ converges } \mu_{t,i} - a.s. \text{ for all } t \in [N], i \in [d]. \tag{7.12}$$

Note that the dependence on k has now been removed. To prove the claim, since $\{\Phi_{t,i,n}\}_n$ is a subsequence of $\{\phi_{t,i,k,n}\}_n$ for every $k \in \mathbb{N}$, Komlós lemma implies

$$\tilde{\Phi}_{t,i,n}(x_{t,i}) - \tilde{w}_{t,i,k,n} \text{ converges } \mu_{t,i,k} - a.s. \text{ for all } t \text{ and } i. \tag{7.13}$$

In particular, both $\{\tilde{\Phi}_{t,i,n}(x_{t,i}) - \tilde{w}_{t,i,1,n}\}_n$ and $\{\tilde{\Phi}_{t,i,n}(x_{t,i}) - \tilde{w}_{t,i,k,n}\}_n$ converge $\mu_{t,i,1}$-a.s. as $n \to \infty$, so does the difference $\{\tilde{w}_{t,i,1,n} - \tilde{w}_{t,i,k,n}\}_n$ for each k. Hence (7.12) follows from (7.13). The identity $\sum_{i=1}^{d} \tilde{w}_{t,i,1,n} = 0$ then allows to replace the approximating dual maximizer $(\phi_n, h_n)_n$ by $(\psi_n, \tilde{h}_n)_n$, where $\psi_{t,i,n} := \tilde{\Phi}_{t,i,n}(x_{t,i}) - \tilde{w}_{t,i,1,n}$ and $\tilde{h}_{t,i,n}$ is the corresponding Cesàro mean of the subsequence of $(h_{t,i,n})_n$ chosen consistently with the selection of $\{\Phi_{t,i,n}\}_n$ out of $\{\phi_{t,i,n}\}_n$. □

Finally, we turn to the proof of Theorem 3.1.

Proof of Theorem 3.1. By Proposition 3.2, we have an approximating dual maximizer $(\phi_n, h_n)_n$ such that

$$\{\phi_{t,i,n}\}_n \text{ converges to a function } \phi_{t,i} \mu_{t,i} - a.s. \text{ as } n \to \infty \text{ for each } t \in [N] \text{ and } i \in [d]. \tag{7.14}$$

We shall prove the convergence of the convex functions $\{\chi_{t,n}\}_n$ defined in (7.1).

In the proof of Lemma 3.3, we obtained an approximating dual maximizer (say $(\phi_n^0, h_n^0)_n$) such that the associated $\{\chi_{t,n}^0\}_n$ satisfied the local bound (7.7). As observed in the proof of Proposition 3.2, the approximating dual maximizer $(\phi_n, h_n)_n$ satisfying (7.14) can be taken as a Cesàro mean of a suitable subsequence of $(\phi_n^0, h_n^0)_n$. This implies the upper bound in (7.7) continues to hold

$$\sup_n \chi_{t,n} \leq M_k \quad \text{on} \quad J_{t-1,k}. \tag{7.15}$$

We shall modify $(\phi_n, h_n)_n$ so that (7.5) holds while (7.15) is retained. Let $I_t := I_{t,1} \times \cdots \times I_{t,d}$. Fix any $a \in I_1$ and define $\phi_t^{\oplus} := \sum_i \phi_{t,i}$. First, (7.14) implies that there exists $a_t \in I_t$ for every $t \in [N]$ such that

$$\lim_{n \to \infty} \phi_{t,n}^{\oplus}(a_t) = \phi_t^{\oplus}(a_t) \in \mathbb{R}. \tag{7.16}$$

In view of (7.2) which gives $\phi_{1,n}^{\oplus} \leq \chi_{2,n}$, this implies

$$\inf_n \chi_{2,n}(a_1) > -\infty. \tag{7.17}$$

On the other hand, since $I_1 = \text{int}(J_1)$ and $J_{1,k} \nearrow J_1$ (see (7.6) for the definition of compact intervals $J_{t,i,k}$), for large enough k we have $\{a, a_1\} \subset \text{int}(J_{1,k})$. Now (7.15), (7.17) imply that both $\chi_{2,n}(a)$ and $\nabla\chi_{2,n}(a)$ are uniformly bounded in n, where $\nabla\chi_{2,n}(a) \in \partial\chi_{2,n}(a)$ is a subgradient of the convex function $\chi_{2,n}$ at a.

Hence by taking a subsequence, we can assume that $\{\chi_{2,n}(a)\}_n$ and $\{\nabla\chi_{2,n}(a)\}_n$ both converge. As in the proof of Lemma 3.3, define an affine function $L_{2,n}(y) = \chi_{2,n}(a) + \nabla\chi_{2,n}(a) \cdot (y - a)$, and replace $\phi_{1,n}^\oplus(x_1)$ with $\phi_{1,n}^\oplus(x_1) - L_{2,n}(x_1)$, $\phi_{2,n}^\oplus(x_2)$ with $\phi_{2,n}^\oplus(x_2) + L_{2,n}(x_2)$, and finally $h_{1,n}(x_1)$ with $h_{1,n}(x_1) - \nabla\chi_{2,n}(a)$. This yields $\chi_{2,n}(a) = \nabla\chi_{2,n}(a) = 0$ for all n, while the a.s. convergence of $\phi_{t,i,n}$ and the bound (7.15) are retained. Next, the inequality $\chi_{3,n} \geq \chi_{2,n} + \phi_{2,n}^\oplus$ with $\chi_{2,n} \geq 0$ yields $\inf_n \chi_{3,n}(a_2) > -\infty$. Thus we can repeat the argument and achieve the normalization (7.5), while the a.s. convergence of $\phi_{t,i,n}$ and the bound (7.15) are retained. Thanks to the local bound (7.7) and (7.3), we can now apply [21, Proposition A.1] to deduce the pointwise convergence $\chi_{t,n} \to \chi_t$ on J_{t-1}.

As the limit function $(\phi_{t,i})_{t,i}$ is only well defined $\mu_{t,i}$-a.s., define $\phi_{t,i} := -\infty$ on a $\mu_{t,i}$-null set (which includes $\mathbb{R} \setminus I_{t,i}$), so that they are defined everywhere on $\mathbb{R}$. We now show that there exist functions $h_t = (h_{t,i})_i : \mathbb{R}^{td} \to \mathbb{R}^d$ for all $t \in [N]$ with $h_N \equiv 0$, such that

$$\sum_{t=1}^{N} \left(\phi_t^\oplus(x_t) + h_t(\bar{x}_t) \cdot \Delta x_t\right) \leq c(x) \quad \text{for all } x \in \mathbb{R}^{Nd}. \tag{7.18}$$

For any function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ which is bounded below by an affine function, let $\text{conv}[f] : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ denote the lower semi-continuous convex envelope of f, that is the supremum of all affine functions l satisfying $l \leq f$ (If there is no such l, let $\text{conv}[f] \equiv -\infty$.) We will inductively obtain $h_{N-1}, h_{N-2}, ..., h_1$. Let us rewrite (3.3) as

$$\sum_{t=1}^{N-1} \left(\phi_{t,n}^\oplus(x_t) + h_{t,n}(\bar{x}_t) \cdot \Delta x_t\right) \leq c(x) - \phi_{N,n}^\oplus(x_N). \tag{7.19}$$

Define $H_{N-1,n}(\bar{x}_{N-1}, x_N) = \text{conv}[c(\bar{x}_{N-1}, \cdot) - \phi_{N,n}^\oplus(\cdot)](x_N)$. We then have

$$\sum_{t=1}^{N-1} \left(\phi_{t,n}^\oplus(x_t) + h_{t,n}(\bar{x}_t) \cdot \Delta x_t\right) \leq H_{N-1,n}(\bar{x}_{N-1}, x_N) \leq c(x) - \phi_{N,n}^\oplus(x_N)$$

because the left hand side is affine in x_N. If we let $x_N = x_{N-1}$, we get

$$\sum_{t=1}^{N-2} \left(\phi_{t,n}^\oplus(x_t) + h_{t,n}(\bar{x}_t) \cdot \Delta x_t\right) \leq H_{N-1,n}(\bar{x}_{N-1}, x_{N-1}) - \phi_{N-1,n}^\oplus(x_{N-1}). \tag{7.20}$$

We see that (7.19) and (7.20) exhibit a similar structure. With the convention $H_{N,n}(x) := c(x)$, this allows us to inductively deduce, backward in t,

$$\sum_{s=1}^{t} \left(\phi_{s,n}^\oplus(x_s) + h_{s,n}(\bar{x}_s) \cdot \Delta x_s\right) \leq H_{t+1,n}(\bar{x}_{t+1}, x_{t+1}) - \phi_{t+1,n}^\oplus(x_{t+1}) \tag{7.21}$$

for $t = 1, ..., N - 1$, where

$$H_{t,n}(\bar{x}_t, x_{t+1}) := \text{conv}[H_{t+1,n}(\bar{x}_t, \cdot) - \phi_{t+1,n}^\oplus(\cdot)](x_{t+1}) \tag{7.22}$$

with an abuse of notation $H_{t+1,n}(\bar{x}_t, x_{t+1}) := H_{t+1,n}(\bar{x}_{t+1}, x_{t+1})$.

Now by dropping the index n, we analogously define

$$H_t(\bar{x}_t, x_{t+1}) := \text{conv}[H_{t+1}(\bar{x}_t, \cdot) - \phi_{t+1}^\oplus(\cdot)](x_{t+1}) \tag{7.23}$$

with $H_N(x) := c(x)$. Next, since the lim sup of convex functions is convex, in conjunction with the almost sure convergence (7.14), we have

$$\limsup_{n\to\infty} H_{N-1,n}(\bar{x}_{N-1}, x_N) = \limsup_{n\to\infty} \text{conv}[c(\bar{x}_{N-1}, \cdot) - \phi_{N,n}^\oplus(\cdot)](x_N)$$
$$\leq \text{conv}[\limsup_{n\to\infty} \left(c(\bar{x}_{N-1}, \cdot) - \phi_{N,n}^\oplus(\cdot)\right)](x_N)$$
$$\leq \text{conv}[c(\bar{x}_{N-1}, \cdot) - \phi_N^\oplus(\cdot)](x_N)$$
$$= H_{N-1}(\bar{x}_{N-1}, x_N).$$

This allows us to inductively deduce, for $t = 1, ..., N-1$,

$$
\begin{aligned}
\limsup_{n \to \infty} H_{t,n}(\bar{x}_t, x_{t+1}) &= \limsup_{n \to \infty} \text{conv}[H_{t+1,n}(\bar{x}_t, \cdot) - \phi^\oplus_{t+1,n}(\cdot)](x_{t+1}) \\
&\leq \text{conv}[\limsup_{n \to \infty} (H_{t+1,n}(\bar{x}_t, \cdot) - \phi^\oplus_{t+1,n}(\cdot))](x_{t+1}) \\
&\leq \text{conv}[H_{t+1}(\bar{x}_t, \cdot) - \phi^\oplus_{t+1}(\cdot)](x_{t+1}) \\
&= H_t(\bar{x}_t, x_{t+1}).
\end{aligned}
$$

Let us discuss continuity of the convex function $x_{t+1} \mapsto H_t(\bar{x}_t, x_{t+1})$. The following inequality from (7.23)

$$
H_t(\bar{x}_t, x_{t+1}) \leq H_{t+1}(\bar{x}_{t+1}, x_{t+1}) - \phi^\oplus_{t+1}(x_{t+1})
$$

becomes $H_{N-1}(\bar{x}_{N-1}, x_N) \leq c(x) - \phi^\oplus_N(x_N)$ when $t = N-1$. Then the $\mu^\otimes_N$-a.s. finiteness of $\phi^\oplus_N$ implies, by convexity, $H_{N-1}(\bar{x}_{N-1}, x_N) < \infty$ if $x_N \in J_{N-1}$. Backward induction in t then gives $H_t(\bar{x}_t, x_{t+1}) < \infty$ if $x_{t+1} \in J_t$. This implies that if there exists $y_0 \in \mathbb{R}^d$ such that $H_t(\bar{x}_t, y_0) > -\infty$, then $y \mapsto H_t(\bar{x}_t, y)$ is real-valued and continuous in J_t. This being observed, now (7.21), (7.22) gives

$$
\phi^\oplus_{1,n}(x_1) + h_{1,n}(x_1) \cdot \Delta x_1 \leq H_{1,n}(x_1, x_2) \leq H_{2,n}(x_1, x_2, x_2) - \phi^\oplus_{2,n}(x_2).
$$

Letting $x_2 = x_1$ gives $\phi^\oplus_{1,n}(x_1) \leq H_{1,n}(x_1, x_1)$. Taking $\limsup$ yields

$$
\phi^\oplus_1(x_1) \leq H_1(x_1, x_1) \quad \text{and} \quad H_1(x_1, x_2) \leq H_2(x_1, x_2, x_2) - \phi^\oplus_2(x_2).
$$

Define $A_t := \{x_t \in \mathbb{R}^d \,|\, \phi^\oplus_t(x_t) \in \mathbb{R}\}$ for each $t \in [N]$, and note that $A_t \subset I_t$. Since $x_2 \mapsto H_1(x_1, x_2)$ is continuous in J_1 for every $x_1 \in A_1$, the subdifferential $\partial H_1(x_1, \cdot)(x_2)$ is nonempty, convex and compact for every $x_2 \in I_1 = \text{int}(J_1)$. This allows us to choose a measurable function $h_1 : A_1 \to \mathbb{R}^d$ satisfying $h_1(x_1) \in \partial H_1(x_1, \cdot)(x_1)$. Then for $x_1 \in A_1$, we have

$$
\begin{aligned}
\phi^\oplus_1(x_1) + h_1(x_1) \cdot (x_2 - x_1) &\leq H_1(x_1, x_1) + h_1(x_1) \cdot (x_2 - x_1) \\
&\leq H_1(x_1, x_2) \\
&\leq H_2(x_1, x_2, x_2) - \phi^\oplus_2(x_2).
\end{aligned}
$$

In particular, for $x_1 \in A_1$ and $x_2 \in A_2$, it holds $H_2(x_1, x_2, x_2) > -\infty$. Hence again we can choose $h_2 : A_1 \times A_2 \to \mathbb{R}^d$ that satisfies $h_2(x_1, x_2) \in \partial H_2(x_1, x_2, \cdot)(x_2)$. Then for every $x_1 \in A_1$ and $x_2 \in A_2$, we have

$$
\begin{aligned}
&\phi^\oplus_1(x_1) + \phi^\oplus_2(x_2) + h_1(x_1) \cdot (x_2 - x_1) + h_2(x_1, x_2) \cdot (x_3 - x_2) \\
&\leq H_2(x_1, x_2, x_2) + h_2(x_1, x_2) \cdot (x_3 - x_2) \\
&\leq H_2(x_1, x_2, x_3) \\
&\leq H_3(x_1, x_2, x_3, x_3) - \phi^\oplus_3(x_3).
\end{aligned}
$$

By induction in t, we obtain $h_t : A_1 \times \cdots \times A_t \to \mathbb{R}^d$ (with $h_N \equiv 0$), satisfying (7.18) as desired. We may define $h_t = 0$ in $\mathbb{R}^{td} \setminus A_1 \times \cdots \times A_t$, noting that the left hand side of (7.18) is $-\infty$ if $x_t \notin A_t$ for some t.

The last step is to show the following claim: for any functions $h_t : \mathbb{R}^{td} \to \mathbb{R}^d$, $t = 1, ..., N$ with $h_N \equiv 0$ satisfying (7.18) (whose existence was just shown), and for any minimizer π^* for the VMOT problem (2.2), it holds

$$
\sum_{t=1}^N \left(\phi^\oplus_t(x_t) + h_t(\bar{x}_t) \cdot \Delta x_t \right) = c(x), \quad \pi^* - a.s.. \tag{7.24}
$$

In other words, every minimizer π^* is concentrated on the contact set

$$
\Gamma := \left\{ x \in \mathbb{R}^{Nd} \,\bigg|\, \sum_{t=1}^N \left(\phi^\oplus_t(x_t) + h_t(\bar{x}_t) \cdot \Delta x_t \right) = c(x) \right\}
$$

whenever $\{h_t\}_t$ satisfies (7.18). This will complete the proof of the theorem.

The proof of this claim was provided in [5] for the single-asset, two-period case $(d, N) = (1, 2)$, and we build upon that approach to extend to the general case. To begin, recall $\phi_{t,i,n} \to \phi_{t,i}$ $\mu_{t,i}$-a.s. and $\chi_{t,n} \to \chi_t$ in J_{t-1} where $\chi_{t,n}$ is defined in (7.1) with χ_t being its limit. For any $\pi \in \mathrm{VMT}(\mu)$ (not necessarily an optimizer), we have $c \in L^1(\pi)$ by the assumption of Theorem 3.1. We claim:

$$\limsup_{n \to \infty} \int \sum_{t=1}^{N} \big(\phi_{t,n}^{\oplus}(x_t) + h_{t,n}(\bar{x}_t) \cdot \Delta x_t\big) d\pi \tag{7.25}$$

$$\leq \int \sum_{t=1}^{N} \big(\phi_t^{\oplus}(x_t) + h_t(\bar{x}_t) \cdot \Delta x_t\big) d\pi.$$

To see how the claim implies (7.24), let π^* be any minimizer for (2.2) (which exists by the assumption on c). Then $D(c) = P(c) = \int c \, d\pi^*$, hence

$$P(c) = \lim_{n \to \infty} \int \sum_{t=1}^{N} \big(\phi_{t,n}^{\oplus}(x_t) + h_{t,n}(\bar{x}_t) \cdot \Delta x_t\big) d\pi^*$$

$$\leq \int \sum_{t=1}^{N} \big(\phi_t^{\oplus}(x_t) + h_t(\bar{x}_t) \cdot \Delta x_t\big) d\pi^*$$

$$\leq \int c(x) \, d\pi^* = P(c)$$

by Fatou's lemma, yielding equality throughout. Notice that this implies (7.24).

To prove (7.25), fix any $\pi \in \mathrm{VMT}(\mu)$. Denoting $\pi = \mathrm{Law}(X)$ where $X = (X_1, ..., X_N)$ is an $\mathbb{R}^d$-valued martingale under π, let $\pi_t := \mathrm{Law}(X_t)$. By following the argument which establishes (7.8), but this time utilizing the convex order $\pi_t \preceq_c \pi_{t+1}$ instead of $\mu_t^{\otimes} \preceq_c \mu_{t+1}^{\otimes}$, by (3.4) and (7.2), we can deduce

$$\| \chi_{t+1,n} - \chi_{t,n} - \phi_{t,n}^{\oplus} \|_{L^1(\pi_t)} \leq C \quad \text{for all } n \in \mathbb{N}, t \in [N]. \tag{7.26}$$

Using $\phi_{t,n}^{\oplus} \to \phi_t^{\oplus}$, $\chi_{t,n} \to \chi_t$ and Fatou's lemma, from (7.26) we deduce

$$\chi_{t+1} - \chi_t - \phi_t^{\oplus} \in L^1(\pi_t), \quad \text{and}$$

$$\limsup_{n \to \infty} \int (\phi_{t,n}^{\oplus} + \chi_{t,n} - \chi_{t+1,n}) \, d\pi_t \leq \int (\phi_t^{\oplus} + \chi_t - \chi_{t+1}) \, d\pi_t,$$

recalling $\chi_{1,n} = \chi_{N+1,n} \equiv 0$ and $\pi_t(J_{t-1}) = 1$. This allows us to deduce

$$\limsup_{n \to \infty} \int \sum_{t=1}^{N} \big(\phi_{t,n}^{\oplus}(x_t) + h_{t,n}(\bar{x}_t) \cdot \Delta x_t\big) d\pi$$

$$= \limsup_{n \to \infty} \int \sum_{t=1}^{N} \big(\phi_{t,n}^{\oplus}(x_t) + \chi_{t,n}(x_t) - \chi_{t+1,n}(x_t)$$

$$- \chi_{t,n}(x_t) + \chi_{t+1,n}(x_t) + h_{t,n}(\bar{x}_t) \cdot \Delta x_t\big) d\pi$$

$$\leq \sum_{t=1}^{N} \int \big(\phi_t^{\oplus} + \chi_t - \chi_{t+1}\big) d\pi_t \tag{7.27}$$

$$+ \limsup_{n \to \infty} \int \sum_{t=1}^{N-1} \big(\chi_{t+1,n}(x_t) - \chi_{t+1,n}(x_{t+1}) + h_{t,n}(\bar{x}_t) \cdot \Delta x_t\big) d\pi,$$

since $\sum_{t=1}^{N} \chi_{t,n}(x_t) = \sum_{t=1}^{N} \chi_{t+1,n}(x_{t+1})$. Now denote $\bar{X}_t = (X_1, ..., X_t)$, $\pi^t := \text{Law}(\bar{X}_t) \in \mathcal{P}(\mathbb{R}^{td})$. Then we can write $\pi^{t+1} = \pi_{\bar{x}_t} \otimes \pi^t$, where $\pi_{\bar{x}_t} \in \mathcal{P}(\mathbb{R}^d)$ is the conditional distribution of X_{t+1} given $\bar{X}_t = \bar{x}_t$ under the martingale law π. Martingale property means that $\int y \, d\pi_{\bar{x}_t}(y) = x_t$. For each t, choose a sequence of functions $\xi_{t,n} : I_t \to \mathbb{R}^d$ satisfying $\xi_{t,n}(x_t) \in \partial \chi_{t+1,n}(x_t)$. Then we compute

$$\int \sum_{t=1}^{N-1} \left(\chi_{t+1,n}(x_t) - \chi_{t+1,n}(x_{t+1}) + h_{t,n}(\bar{x}_t) \cdot \Delta x_t \right) d\pi$$

$$= \int \left(\int \left(\chi_{N,n}(x_{N-1}) - \chi_{N,n}(x_N) + \xi_{N-1,n}(x_{N-1}) \cdot \Delta x_{N-1} \right) d\pi_{\bar{x}_{N-1}}(x_N) \right.$$

$$\left. + \sum_{t=1}^{N-2} \left(\chi_{t+1,n}(x_t) - \chi_{t+1,n}(x_{t+1}) + h_{t,n}(\bar{x}_t) \cdot \Delta x_t \right) \right) d\pi^{N-1}(\bar{x}_{N-1}),$$

where we replaced $h_{N-1,n}(\bar{x}_{N-1})$ by $\xi_{N-1,n}(\bar{x}_{N-1})$ due to the martingale property

$$\int h_{N-1,n}(\bar{x}_{N-1}) \cdot \Delta x_{N-1} \, d\pi_{\bar{x}_{N-1}}(x_N) = \int \xi_{N-1,n}(x_{N-1}) \cdot \Delta x_{N-1} \, d\pi_{\bar{x}_{N-1}}(x_N) = 0.$$

By definition of ξ, we have $\chi_{N,n}(x_{N-1}) - \chi_{N,n}(x_N) + \xi_{N-1,n}(x_{N-1}) \cdot \Delta x_{N-1} \leq 0$. This allows us to disintegrate $\pi^{N-1} = \pi_{\bar{x}_{N-2}} \otimes \pi^{N-2}$ and repeat the same argument

$$\int \sum_{t=1}^{N-1} \left(\chi_{t+1,n}(x_t) - \chi_{t+1,n}(x_{t+1}) + h_{t,n}(\bar{x}_t) \cdot \Delta x_t \right) d\pi$$

$$= \int \cdots \int \left(\chi_{N,n}(x_{N-1}) - \chi_{N,n}(x_N) + \xi_{N-1,n}(x_{N-1}) \cdot \Delta x_{N-1} \right) d\pi_{\bar{x}_{N-1}}(x_N)$$

$$+ \left(\chi_{N-1,n}(x_{N-2}) - \chi_{N-1,n}(x_{N-1}) + \xi_{N-2,n}(x_{N-2}) \cdot \Delta x_{N-2} \right) d\pi_{\bar{x}_{N-2}}(x_{N-1})$$

$$+ \cdots + \left(\chi_{2,n}(x_1) - \chi_{2,n}(x_2) + \xi_{1,n}(x_1) \cdot \Delta x_1 \right) d\pi_{\bar{x}_1}(x_2) d\pi^1(x_1).$$

Since $\chi_{t+1,n}(x_t) - \chi_{t+1,n}(x_{t+1}) + \xi_{t,n}(x_t) \cdot \Delta x_t \leq 0$ for all t, repeated application of Fatou's lemma allows $\limsup$ to continue to penetrate into the innermost integral. Using $\limsup(a_n + b_n) \leq \limsup a_n + \limsup b_n$, this eventually yield

$$\limsup_{n \to \infty} \int \sum_{t=1}^{N-1} \left(\chi_{t+1,n}(x_t) - \chi_{t+1,n}(x_{t+1}) + h_{t,n}(\bar{x}_t) \cdot \Delta x_t \right) d\pi$$

$$\leq \int \cdots \int \left(\chi_N(x_{N-1}) - \chi_N(x_N) + \xi_{N-1}(x_{N-1}) \cdot \Delta x_{N-1} \right) d\pi_{\bar{x}_{N-1}}(x_N)$$

$$+ \left(\chi_{N-1}(x_{N-2}) - \chi_{N-1}(x_{N-1}) + \xi_{N-2}(x_{N-2}) \cdot \Delta x_{N-2} \right) d\pi_{\bar{x}_{N-2}}(x_{N-1})$$

$$+ \cdots + \left(\chi_2(x_1) - \chi_2(x_2) + \xi_1(x_1) \cdot \Delta x_1 \right) d\pi_{\bar{x}_1}(x_2) d\pi^1(x_1)$$

for some $\xi_t(x_t) \in \partial \chi_{t+1}(x_t)$ which is a limit point of the bounded sequence $\{\xi_{t,n}(x_t)\}_n$. Substituting $\xi_t(x_t)$ back to $h_t(\bar{x}_t)$ in the above inequality yields

$$\limsup_{n \to \infty} \int \sum_{t=1}^{N-1} \left(\chi_{t+1,n}(x_t) - \chi_{t+1,n}(x_{t+1}) + h_{t,n}(\bar{x}_t) \cdot \Delta x_t \right) d\pi$$

$$\leq \int \sum_{t=1}^{N-1} \left(\chi_{t+1}(x_t) - \chi_{t+1}(x_{t+1}) + h_t(\bar{x}_t) \cdot \Delta x_t \right) d\pi.$$

Combining the integrals in (7.27) then yields the claim (7.25), hence the theorem. $\qquad\square$

References

[1] David Applegate, Mateo Díaz, Oliver Hinder, Haihao Lu, Miles Lubin, Brendan O'Donoghue, and Warren Schudy. Practical large-scale linear programming using primal-dual hybrid gradient. *Advances in Neural Information Processing Systems*, 34:20243–20257, 2021.

[2] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices: a mass transport approach. *Finance and Stochastics*, 17:477–501, 2013.

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

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

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

[6] D. T. Breeden and R. H. Litzenberger. Prices of state-contingent claims implicit in option prices. *J. Business*, 51(4):621–651, 1978.

[7] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[8] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40:120–145, 2011.

[9] NVIDIA Corporation. cuOpt: GPU-accelerated optimization engine. `https://nvidia.com/cuopt`, 2025. Version 25.08.00.

[10] A. M. G. Cox, J. Obłój, and N. Touzi. The root solution to the multi-marginal embedding problem: an optimal stopping and time-reversal approach. *Probab. Theory Related Fields.*, 173(1-2):211–259, 2019.

[11] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

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

[13] S. Eckstein, G. Guo, T. Lim, and J. Obłój. Robust pricing and hedging of options on multiple assets and its numerics. *SIAM J. Financial Math.*, 12(1):158–188, 2021.

[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] N. Ghoussoub, Y-H Kim, and T. Lim. Structure of optimal martingale transport plans in general dimensions. *Ann. Probab.*, 47(1):109–164, 2019.

[16] I. Guo, G. Loeper, and S. Wang. Calibration of local-stochastic volatility models by optimal transport. *arXiv preprint*, 2019.

[17] J. Guyon. The joint S&P 500/VIX smile calibration puzzle solved. 2020.

[18] Oliver Hinder. Worst-case analysis of restarted primal-dual hybrid gradient on totally unimodular linear programs. *arXiv preprint arXiv:2309.03988*, 2023.

[19] D. Hobson. The skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance 2010*, volume 2003 of *Lecture Notes in Math.*, pages 267–318. Springer, Berlin, 2011.

[20] L. V. Kantorovich. On the translocation of masses. *Dokl. Akad. Nauk. USSR*, 37:199–201, 1942.

[21] T. Lim. Geometry of vectorial martingale optimal transportations and duality. *Mathematical Programming*, 2023. `https://doi.org/10.1007/s10107-023-01954-4`.

[22] E. Lütkebohmert and J. Sester. Tightening robust price bounds for exotic derivatives. *Quantitative Finance*, 19:1797–1815, 2019.

[23] H. De March. Local structure of multi-dimensional martingale optimal transport. *arXiv preprint*, 2018.

[24] H. De March. Quasi-sure duality for multi-dimensional martingale optimal transport. *arXiv preprint*, 2018. `https://arxiv.org/abs/1805.01757`.

[25] H. De March and P. Henry-Labordere. Building arbitrage-free implied volatility: Sinkhorn's algorithm and variants. *arXiv preprint*, 2019. `https://arxiv.org/abs/1902.04456`.

[26] H. De March and N. Touzi. Irreducible convex paving for decomposition of multidimensional martingale transport plans. *Ann. Probab.*, 47(3):1726–1774, 2019.

[27] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, pages 666–704, 1781.

[28] M. Nutz, F. Stebegg, and X. Tan. Multiperiod martingale transport. *Stochastic Process. Appl.*, 130(3):1568–1615, 2020.

[29] J. Obłój and P. Siorpaes. Structure of martingale transports in finite dimensions. *arXiv preprint*, 2017. `https://arxiv.org/abs/1702.08433`.

[30] J. Obłój and P. Spoida. An iterated azéma-yor type embedding for finitely many marginals. *Ann. Probab.*, 45(4):2210–2247, 2017.

[31] F. Santambrogio. *Optimal transport for applied mathematicians. Calculus of variations, PDEs, and modeling*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015.

[32] J. Sester. On intermediate marginals in martingale optimal transportation. *Mathematics and financial economics*, 17:615–654, 2023.

[33] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.

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

[35] D. Zaev. On the monge-kantorovich problem with additional linear constraints. *Mathematical Notes*, 98(5-6):725–741, 2015.

Acknowledgments

Disclaimer

This paper was prepared for informational purposes with contributions from the Global Technology Applied Research center of JPMorgan Chase & Co. This paper is not a product of the Research Department of JPMorgan Chase & Co. or its affiliates. Neither JPMorgan Chase & Co. nor any of its affiliates makes any explicit or implied representation or warranty and none of them accept any liability in connection with this paper, including, without limitation, with respect to the completeness, accuracy, or reliability of the information contained herein and the potential legal, compliance, tax, or accounting effects thereof. This document is not intended as investment research or investment advice, or as a recommendation, offer, or solicitation for the purchase or sale of any security, financial instrument, financial product or service, or to be used in any way for evaluating the merits of participating in any transaction.