*The Annals of Applied Probability*
2019, Vol. 29, No. 6, 3311–3347
https://doi.org/10.1214/19-AAP1481


# COMPUTATIONAL METHODS FOR MARTINGALE OPTIMAL TRANSPORT PROBLEMS[1]

BY GAOYUE GUO AND JAN OBŁÓJ

*University of Oxford*

We develop computational methods for solving the *martingale optimal transport* (MOT) problem—a version of the classical optimal transport with an additional martingale constraint on the transport's dynamics. We prove that a general, multi-step multi-dimensional, MOT problem can be approximated through a sequence of *linear programming* (LP) problems which result from a discretization of the marginal distributions combined with an appropriate relaxation of the martingale condition. Further, we establish two generic approaches for discretising probability distributions, suitable respectively for the cases when we can compute integrals against these distributions or when we can sample from them. These render our main result applicable and lead to an implementable numerical scheme for solving MOT problems. Finally, specialising to the one-step model on real line, we provide an estimate of the convergence rate which, to the best of our knowledge, is the first of its kind in the literature.

**1. Introduction.** The *optimal transport* (OT) problem is concerned with transferring mass from one location to another in such a way as to optimise a given criterion. Rephrased mathematically, and for simplicity considering the one-dimensional case, we are given two probability distributions $\mu$ and $\nu$ on $\mathbb{R}$ and seek to minimise

$$\int_{\mathbb{R}^2} c(x,y)\mathbb{P}(dx,dy), \tag{1}$$

among all probability measures $\mathbb{P}$, also known as *transport plans*, such that

$$\mathbb{P}[E\times\mathbb{R}]=\mu[E] \quad \text{and} \quad \mathbb{P}[\mathbb{R}\times E]=\nu[E] \quad \text{for all } E\in\mathcal{B}(\mathbb{R}), \tag{2}$$

where $c:\mathbb{R}^2\to\mathbb{R}$ is a measurable cost function. Theoretical advances in the last fifty years characterise existence, uniqueness, representation and smoothness properties of optimisers in a variety of different settings (see, e.g., [35, 38]), and applications are abundant throughout most of the applied sciences including biomedical sciences, geography and data science. Accordingly, numerical techniques for the

Received December 2018.

[1] Support of the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 335421 is gratefully acknowledged.

*MSC2010 subject classifications.* Primary 49M25, 60H99; secondary 90C08.

*Key words and phrases.* Martingale optimal transport, martingale relaxation, robust hedging, duality, discretization of measure, linear programming.

OT are of great importance and have rapidly developed into an important and separate field of applied mathematics:

1. In the absolutely continuous case, that is, $\mu(dx) = \rho(x)\,dx$ and $\nu(dy) = \sigma(y)\,dy$, Benamou and Brenier proposed in [6] a numerical scheme for the quadratic distance function $c(x, y) = (x - y)^2$ using an equivalent formulation arising from fluid mechanics.
2. In the purely discrete case, that is, $\mu(dx) = \sum_{i=1}^m \alpha_i \delta_{x_i}(dx)$ and $\nu(dy) = \sum_{j=1}^n \beta_j \delta_{y_j}(dy)$, the OT problem reduces to a *linear programming* (LP) problem and can be computed using the iterative Bregman projection; see Benamou et al. [7].
3. In the semidiscrete case, that is, $\mu(dx) = \rho(x)\,dx$ and $\nu(dy) = \sum_{j=1}^n \beta_j \delta_{y_j}(dy)$, Lévy [32] adopted a computational geometry approach to the cost $c(x, y) = (x - y)^2$ and solved the OT problem by means of Laguerre's tessellations.

Recently, an additional constraint has been taken into account, which leads to the so-called *martingale optimal transport* (MOT) problem. This optimization problem was motivated by, and contributed to, the so-called model-independent, or robust, pricing of exotic options in mathematical finance, perspective which has gained significant momentum in the wake of financial crisis. More precisely, the two given measures $\mu$ and $\nu$ describe the initial and final distributions of stock prices and can be recovered from market prices of traded call/put options. Calibrated market models are thus identified by martingales with these prescribed marginals, that is, transport plans $\mathbb{P}$ which further satisfy

$$\int_{\mathbb{R}} y \mathbb{P}_x(dy) = x \quad \text{for } \mu\text{-a.e. } x \in \mathbb{R}, \tag{3}$$

where $(\mathbb{P}_x)_{x \in \mathbb{R}}$ denotes the disintegration of $\mathbb{P}$ with respect to $\mu$. The MOT problem aims at maximising[2] the integral (1) overall $\mathbb{P}$, still named transport plans, satisfying the constraints (2) and (3), and it corresponds to the model-independent price for option $c$. This methodology was presented by Beiglböck et al. [3], to which we refer for a more detailed discussion, see also [18, 20, 30] for the continuous time setting. It is also worth mentioning that some concrete MOT problems for particular payoffs have been investigated, by means of stochastic control or Skorokhod embedding techniques, in a stream of papers going back to Hobson [27]; see, for example, [8, 10, 12–14, 25, 28, 29].

Given the active theoretical interest in MOT problems, as well as their importance for applications in mathematical finance, it becomes increasingly important to develop numerical techniques and computational methods for these problems. A natural starting point is given by a simple, but important, observation that, for the

[2]The maximization formulation is more adapted to financial applications. We refer to $c$ as a reward function or payoff, which is commonly accepted in the finance jargon.

purely discrete case stated above, the MOT problem is equivalent to the following LP problem:

$$\max_{(p_{i,j})_{1\le i\le m,1\le j\le n}\in\mathbb{R}_+^{mn}} \sum_{i=1}^{m}\sum_{j=1}^{n} p_{i,j}c(x_i,y_j) \quad \text{s.t.}$$

$$\sum_{j=1}^{n} p_{i,j} = \alpha_i \quad \text{for } i = 1,\dots,m,$$

$$\sum_{i=1}^{m} p_{i,j} = \beta_j \quad \text{for } j = 1,\dots,n,$$

$$\sum_{j=1}^{n} p_{i,j} y_j = \alpha_i x_i \quad \text{for } i = 1,\dots,m.$$

Such LP formulation, for a convex $c$, was pioneered in Davis et al. [15], where instead of the marginal constraint $\nu$, only a finite number of expectation constraints are given. The setting reflects the actual market data, and hence is directly motivated by the applications. It was considered by Henry–Labordère [26] who proposed the first numerical approach, based on the dual problem (see (17) below), and illustrated its performance on a number of practically relevant problems. To adopt this approach in general, we could hope to approximate the MOT problem for $(\mu,\nu)$ with the above LP problem for finitely supported $(\mu^n,\nu^n)$ which are 'close' to $(\mu,\nu)$. Unfortunately, this naive idea hits two important obstacles on the way to its implementation and a proof of convergence. First, there are no general continuity results of the MOT problem as a function of its input $(\mu,\nu)$. To the best of our knowledge, the only exception is Juillet [31] who proved that, if $c(x,y)=\varphi(x)\psi(y)$ or $c(x,y)=h(x-y)$, where $\varphi,\psi,h:\mathbb{R}\to\mathbb{R}$ are assumed to satisfy the conditions of Remark 2.10 in [31], then there exists an optimiser $\mathbb{P}^*(\mu,\nu)$ which is Lipschitz with respect to $(\mu,\nu)$ under a topology of Wasserstein type. We extend his result to more general payoffs $c$ in Proposition 4.7 under additional convergence of second moments. Further extensions to continuity under Wasserstein metric of order one and beyond one dimension remain as challenging open problems[3]. Second, even if $(\mu,\nu)$ admits a martingale transport plan, in dimension $d>1$ it may be difficult to construct a discrete approximation $(\mu^n,\nu^n)$ which also does so; see Remark 3.3 below. In fact, the martingale condition, which seems harmless, renders any of the usual OT techniques unusable, for example, stability results, tools from PDE and computational geometry. To the best of our knowledge, and in contrast to the OT, numerical methods for MOT problems are close to nonexistent so far, relative to the theory and applications.

[3] *Added at the proofs stage*: Remarkably, the former appears to have been achieved in two independent recent works [2, 39].

This paper fills in this important gap. We provide an approximation approach for solving systematically $N$-period MOT problems on $\mathbb{R}^d$, with $N \geq 2$ and $d \geq 1$. Our approximation of the original problem relies on a discretization of the marginal distributions coupled with a suitable relaxation of the martingale constraint leading to a sequence of LP problems. This sequence converges and, specialising to $N = 2$ and $d = 1$, we obtain the convergence speed. Our investigation involves a number of novel results and techniques which we believe are of independent interest. In particular, we compute explicitly the constants in [19] for the convergence rate of empirical measures to the limit in Glivenko–Cantelli's theorem.

The paper is organised as follows. In the rest of this Introduction, we clarify the framework and notation under which we work. Section 2 contains all the main theoretical results: we introduce the *relaxed martingale optimal transport* (relaxed MOT), show the convergence of approximating LP problems to the MOT problem and provide a bound on the convergence rate in dimension one. In Section 3, we consider possible implementations of our method. This requires approximating a probability measure $\mu$ by discrete measures $\mu^n$ and being able to compute, or bound, the Wasserstein distance between $\mu^n$ and $\mu$. We develop two generic approaches to achieve this, and then present several numerical examples which illustrate our methods and provide heuristic insights into the structure of optimisers, including a conjecture in [22]. Section 4 contains all the related proofs. Section 5 concludes the paper and points to potential future work.

1.1. *Preliminaries.* For a given set $E$, we denote by $E^k$ its $k$-fold product. If $E$ is Polish, then $\mathcal{B}(E)$ denotes its Borel $\sigma$-field and $\mathcal{P}(E)$ is the set of probability measures on $(E, \mathcal{B}(E))$ which admit a finite first moment. As is common when studying the OT, we formulate our problem on the canonical space, which plays an important role in the analysis. Let $\Omega := \mathbb{R}^d$ with its elements denoted by $\mathsf{x} = (x_1, \dots, x_d)$ and $\mathcal{P} := \mathcal{P}(\Omega)$. Throughout, we endow $\mathbb{R}^d$ with the $\ell_1$ norm $|\cdot|$, that is, $|\mathsf{x}| := \sum_{i=1}^d |x_i|$. Define $\Lambda$ to be the space of Lipschitz functions on $\mathbb{R}^d$ and, given $f \in \Lambda$, denote by $\mathrm{Lip}(f)$ its Lipschitz constant on $\mathbb{R}^d$. For each $L > 0$, let $\Lambda_L \subset \Lambda$ be the subspace of functions $f$ with $\mathrm{Lip}(f) \leq L$. We consider the coordinate process $(\mathsf{S}_k)_{1 \leq k \leq N}$, that is, $\mathsf{S}_k(\mathsf{x}_1, \dots, \mathsf{x}_N) := \mathsf{x}_k$ for all $(\mathsf{x}_1, \dots, \mathsf{x}_N) \in \Omega^N$, and its natural filtration $(\mathcal{F}_k)_{1 \leq k \leq N}$, that is, $\mathcal{F}_k := \sigma(\mathsf{S}_1, \dots, \mathsf{S}_k)$. From the financial point of view, $\Omega^N$ models the collection of all possible trajectories for the price evolution of $d$ stocks, where $N$ is the number of trading dates.

Given a vector of probability measures $\boldsymbol{\mu} = (\mu_k)_{1 \leq k \leq N} \in \mathcal{P}^N$, define the set of transport plans with the marginal distributions $\mu_1, \dots, \mu_N$ by

$$\mathcal{P}(\boldsymbol{\mu}) := \{\mathbb{P} \in \mathcal{P}(\Omega^N) : \mathbb{P} \circ \mathsf{S}_k^{-1} = \mu_k, \text{ for } k = 1, \dots, N\},$$

where $\mathbb{P} \circ \mathsf{S}_k^{-1}$ denotes the push forward of $\mathbb{P}$ via the map $\mathsf{S}_k : \Omega^N \to \Omega$. In particular, the Wasserstein distance (of order 1) between $\mu$ and $\nu \in \mathcal{P}$ is given by

$$
\begin{aligned}
\mathcal{W}(\mu, \nu) &:= \inf_{\mathbb{P} \in \mathcal{P}(\mu,\nu)} \mathbb{E}_{\mathbb{P}}\big[|\mathsf{S}_1 - \mathsf{S}_2|\big] \\
&= \sup_{f \in \Lambda_1} \left\{ \int_{\mathbb{R}^d} f(\mathsf{x})\mu(d\mathsf{x}) - \int_{\mathbb{R}^d} f(\mathsf{x})\nu(d\mathsf{x}) \right\},
\end{aligned}
\tag{4}
$$

where the second equality follows by Kantorovich's duality. We recall that $\mathcal{P}$, equipped with the metric $\mathcal{W}$, is a Polish space. Further, for any $(\mu^n)_{n \geq 1} \subset \mathcal{P}$ and $\mu \in \mathcal{P}$, $\mathcal{W}(\mu^n, \mu) \to 0$ holds if and only if

$$
\mu^n \xrightarrow{\mathcal{L}} \mu \quad \text{and} \quad \int_{\mathbb{R}^d} |\mathsf{x}|\mu^n(d\mathsf{x}) \longrightarrow \int_{\mathbb{R}^d} |\mathsf{x}|\mu(d\mathsf{x}),
$$

where $\mathcal{L}$ represents the weak convergence of probability measures; see the monograph of Rachev and Rüschendorf [35] for more details. To facilitate our analysis in the sequel, we endow $\mathcal{P}^N$ with the product metric $\mathcal{W}^{\oplus}$ defined by $\mathcal{W}^{\oplus}(\boldsymbol{\mu}, \boldsymbol{\nu}) := \sum_{k=1}^N \mathcal{W}(\mu_k, \nu_k)$, for all $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathcal{P}^N$. It follows that $\mathcal{P}^N$ is Polish with respect to $\mathcal{W}^{\oplus}$. We close this introduction by listing some notation used in the following.

*Notation.*

- $\mathsf{0} := (0, \ldots, 0)$, $\mathsf{1} := (1, \ldots, 1) \in \mathbb{R}^d$, and to stress the unidimensional case, we write $\mathsf{x} \equiv x$ and $\mathsf{S}_k \equiv S_k$ for $d = 1$; see, for example, Section 3.3.
- $\mathbb{L}^0(\Omega^k; \mathbb{R}^d)$ is the set of measurable functions from $\Omega^k$ to $\mathbb{R}^d$. Denote by $\mathbb{L}^\infty(\Omega^k; \mathbb{R}^d) \subset \mathbb{L}^0(\Omega^k; \mathbb{R}^d)$ the subset of (uniformly) bounded functions, and by $\mathcal{C}_b(\Omega^k; \mathbb{R}^d) \subset \mathbb{L}^\infty(\Omega^k; \mathbb{R}^d)$ the subset of continuous bounded functions.
- For simplicity purposes, we adopt the abbreviations below whenever the context is clear:

$$
\int f\, d\mu \equiv \int_{\mathbb{R}^d} f(\mathsf{x})\mu(d\mathsf{x}),
$$

$$
(p_{i_1,\ldots,i_N}) \equiv (p_{i_1,\ldots,i_N})_{i_1 \in I_1,\ldots,i_N \in I_N},
$$

$$
\sum_{i_1,\ldots,i_N} \equiv \sum_{i_1 \in I_1,\ldots,i_N \in I_N}.
$$

**2. Main results.** Our computational method relies on the convergence result stated in Theorem 2.2. To introduce the result, we need the notion of $\varepsilon$-*approximating martingale measure*.

DEFINITION 2.1. For any $\varepsilon \geq 0$, a probability measure $\mathbb{P} \in \mathcal{P}(\Omega^N)$ is said to be an $\varepsilon$-approximating martingale measure if for each $k = 1, \ldots, N - 1$,

$$
\mathbb{E}_{\mathbb{P}}\big[\big|\mathbb{E}_{\mathbb{P}}[\mathsf{S}_{k+1}|\mathcal{F}_k] - \mathsf{S}_k\big|\big] \leq \varepsilon,
\tag{5}
$$

or equivalently, in view of the monotone class theorem,

$$
\mathbb{E}_{\mathbb{P}}\big[h(\mathsf{S}_1, \ldots, \mathsf{S}_k) \cdot (\mathsf{S}_{k+1} - \mathsf{S}_k)\big] \leq \varepsilon \|h\|_\infty \quad \text{for all } h \in \mathcal{C}_b\big(\Omega^k; \mathbb{R}^d\big),
\tag{6}
$$

where $\|h\|_\infty := \max(\|h^{(1)}\|_\infty, \dots, \|h^{(d)}\|_\infty)$ and $\|h^{(i)}\|_\infty := \sup_{(\mathsf{x}_1,\dots,\mathsf{x}_k)\in\Omega^k} \times |h^{(i)}(\mathsf{x}_1, \dots, \mathsf{x}_k)|$ for $i = 1, \dots, d$.

Given $\varepsilon \geq 0$, let $\mathcal{M}_\varepsilon(\boldsymbol{\mu}) \subset \mathcal{P}(\boldsymbol{\mu})$ be the subset containing all $\varepsilon$-approximating martingale measures. Then $\mathcal{M}_\varepsilon(\boldsymbol{\mu})$ is convex and closed with respect to the weak topology by (6), and thus compact. For a measurable function $c : \Omega^N \to \mathbb{R}$, the relaxed MOT problem is defined by

$$\mathsf{P}_\varepsilon(\boldsymbol{\mu}) := \sup_{\mathbb{P}\in\mathcal{M}_\varepsilon(\boldsymbol{\mu})} \mathbb{E}_\mathbb{P}\big[c(\mathsf{S}_1, \dots, \mathsf{S}_N)\big], \tag{7}$$

where we set by convention $\mathsf{P}_\varepsilon(\boldsymbol{\mu}) := -\infty$ whenever $\mathcal{M}_\varepsilon(\boldsymbol{\mu}) = \varnothing$. Denote further by $\mathcal{P}_\varepsilon^\preceq \subset \mathcal{P}^N$ the collection of measures $\boldsymbol{\mu}$ such that $\mathcal{M}_\varepsilon(\boldsymbol{\mu}) \neq \varnothing$. We note that every $\mathbb{P} \in \mathcal{M}_0(\boldsymbol{\mu})$ is a martingale measure, that is, $(\mathsf{S}_k)_{1\leq k\leq N}$ is a martingale under $\mathbb{P}$, and $\mathsf{P}_0(\boldsymbol{\mu})$ is the MOT problem. In the rest of the paper, for simplicity, we drop the subscript $\varepsilon$ when $\varepsilon = 0$, for example, $\mathcal{P}^\preceq \equiv \mathcal{P}_0^\preceq$, $\mathcal{M}(\boldsymbol{\mu}) \equiv \mathcal{M}_0(\boldsymbol{\mu})$, $\mathsf{P}(\boldsymbol{\mu}) \equiv \mathsf{P}_0(\boldsymbol{\mu})$, etc.

As previously mentioned, $\mathsf{P}(\boldsymbol{\mu})$ reduces to an LP problem once the marginals $\mu_k$ have finite support for $k = 1, \dots, N$. We now couple this observation with a suitable relaxation of the martingale constraint to obtain a unified framework for computing $\mathsf{P}(\boldsymbol{\mu})$ numerically.

THEOREM 2.2. *Fix $\boldsymbol{\mu} \in \mathcal{P}^\preceq$. Let $(\boldsymbol{\mu}^n)_{n\geq 1} \subset \mathcal{P}^N$ be a sequence converging to $\boldsymbol{\mu}$ under $\mathcal{W}^\oplus$. Then, for all $n \geq 1$, $\boldsymbol{\mu}^n \in \mathcal{P}_{r_n}^\preceq$ with $r_n := \mathcal{W}^\oplus(\boldsymbol{\mu}^n, \boldsymbol{\mu})$. Assume further $c$ is Lipschitz.*

(i) *For any sequence $(\varepsilon_n)_{n\geq 1}$ converging to zero such that $\varepsilon_n \geq r_n$ for all $n \geq 1$, one has*

$$\lim_{n\to\infty} \mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n) = \mathsf{P}(\boldsymbol{\mu}).$$

(ii) *For each $n \geq 1$, $\mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n)$ admits an optimiser $\mathbb{P}_n \in \mathcal{M}_{\varepsilon_n}(\boldsymbol{\mu}^n)$, that is, $\mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n) = \mathbb{E}_{\mathbb{P}_n}[c]$. The sequence $(\mathbb{P}_n)_{n\geq 1}$ is tight and every limit point must be an optimiser for $\mathsf{P}(\boldsymbol{\mu})$. In particular, $(\mathbb{P}_n)_{n\geq 1}$ converges weakly whenever $\mathsf{P}(\boldsymbol{\mu})$ has a unique optimiser.*

REMARK 2.3. (i) By Strassen's theorem [37], $\boldsymbol{\mu} \in \mathcal{P}^\preceq$ if and only if $\mu_k \preceq \mu_{k+1}$ for $k = 1, \dots, N-1$, or namely, $\int f\, d\mu_k \leq \int f\, d\mu_{k+1}$ holds for all convex functions $f \in \Lambda$ and $k = 1, \dots, N-1$. In addition, it follows by definition that $\mathcal{P}_\varepsilon^\preceq \subset \mathcal{P}^N$ is convex and closed under $\mathcal{W}^\oplus$, and $\mathcal{M}(\boldsymbol{\mu}) \subset \mathcal{M}_\varepsilon(\boldsymbol{\mu})$ for all $\varepsilon \geq 0$.

(ii) As noted before, a natural idea is to try to approximate $\mathsf{P}(\boldsymbol{\mu})$ by $\mathsf{P}(\boldsymbol{\mu}^n)$ with finitely supported measures $\mu_1^n, \dots, \mu_N^n$ since the latter amounts to an LP problem. For the classical OT, the continuous dependency of the optimization problem on $\boldsymbol{\mu}$ can be derived either from the primal problem, or from its dual formulation. However, the additional martingale constraint means the usual OT arguments no longer work. The continuity of $\boldsymbol{\mu} \mapsto \mathsf{P}(\boldsymbol{\mu})$ remains an open question in general. For $d = 1$, a partial result is shown in [31] and we extend it in Proposition 4.7 below. Additionally, one has to consider suitable approximations (see Section 4.2) to even ensure that $\mathcal{M}(\boldsymbol{\mu}^n)$ is nonempty. This becomes involved for $d > 1$. Theorem 2.2 shows that, a further relaxation of the martingale constraint allows to avoid both issues and to establish the desired convergence result. We also remark that the distance $r_n$ does not admit a closed-form expression and its numerical estimation could be costly. Thanks to Theorem 2.2, we may use in practice any upper bound $\varepsilon_n \geq r_n$ converging to zero.

(iii) Finally, we point out the Lipschitz assumption can be slightly weakened. Let $E \subseteq \mathbb{R}^d$ be a closed subset such that $\operatorname{supp}(\mu_k^n) \subseteq E$ for all $n \geq 1$ and $k = 1, \dots, N$. Then it suffices to assume in Theorem 2.2 that $c$, restricted to $E^N$, is Lipschitz.

We now show that $\mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n)$ is equivalent to an LP problem. Hence, with a slight abuse of language, we always refer to $\mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n)$ as the approximating LP problem of $\mathsf{P}(\boldsymbol{\mu})$.

COROLLARY 2.4. *Let $\boldsymbol{\mu}^n = (\mu_k^n)_{1\leq k\leq N}$ be chosen such that each $\mu_k^n$ has finite support, that is,*

$$\mu_k^n(d\mathsf{x}) = \sum_{i_k \in I_k} \alpha_{i_k}^k \delta_{\mathsf{x}_{i_k}^k}(d\mathsf{x}),$$

*where $I_k = \{1, \dots, n(k)\}$ labels the support $\operatorname{supp}(\mu_k^n)$. Denote by $p = (p_{i_1,\dots,i_N})_{i_1\in I_1,\dots,i_N\in I_N}$ the elements of $\mathbb{R}_+^D$ with $D := \Pi_{k=1}^N n(k)$, then $\mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n)$ can be rewritten as an LP problem.*

PROOF. By assumption, every element $\mathbb{P} \in \mathcal{M}_{\varepsilon_n}(\boldsymbol{\mu}^n)$ can be identified by some $p \in \mathbb{R}_+^D$. Therefore, $\mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n)$ turns to be the optimization problem below

$$\begin{aligned}
&\max_{p\in\mathbb{R}_+^D} \sum_{i_1,\dots,i_N} p_{i_1,\dots,i_N} c\big(\mathsf{x}_{i_1}^1, \dots, \mathsf{x}_{i_N}^N\big) \\
&\quad \text{s.t.} \sum_{i_1,\dots,i_{k-1},i_{k+1},\dots,i_N} p_{i_1,\dots,i_N} = \alpha_{i_k}^k \quad \text{for } i_k \in I_k \text{ and } k = 1, \dots, N, \\
&\quad \sum_{i_1,\dots,i_k} \Bigg| \sum_{i_{k+1},\dots,i_N} p_{i_1,\dots,i_N}\big(\mathsf{x}_{i_{k+1}}^{k+1} - \mathsf{x}_{i_k}^k\big)\Bigg| \leq \varepsilon_n \quad \text{for } k = 1, \dots, N-1.
\end{aligned} \tag{8}$$

(8) is not a LP formulation, however, by adding slack variables

$$\big(\delta_{i_1,\dots,i_k,j}^k\big)_{i_1\in I_1,\dots,i_k\in I_k, j\in J} \in \mathbb{R}_+^{D_k}$$

with $J := \{1, \ldots, d\}$ and $D_k := d\Pi_{r=1}^k n(r)$, (8) is equivalent to the following $\mathsf{LP}$ problem:

$$\max_{p\in\mathbb{R}_+^D,\delta^1\in\mathbb{R}_+^{D_1},\ldots,\delta^{N-1}\in\mathbb{R}_+^{D_{N-1}}} \sum_{i_1,\ldots,i_N} p_{i_1,\ldots,i_N} c(\mathsf{x}_{i_1}^1, \ldots, \mathsf{x}_{i_N}^N)$$

$$\text{s.t.} \sum_{i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_N} p_{i_1,\ldots,i_N} = \alpha_{i_k}^k \quad \text{for } i_k \in I_k \text{ and } k = 1, \ldots, N,$$

$$-\delta_{i_1,\ldots,i_k,j}^k \leq \sum_{i_{k+1},\ldots,i_N} p_{i_1,\ldots,i_N}(x_{i_{k+1},j}^{k+1} - x_{i_k,j}^k) \leq \delta_{i_1,\ldots,i_k,j}^k \quad \text{for } i_k \in I_k,$$

$$j \in J \text{ and } k = 1, \ldots, N,$$

$$P \sum_{i_1,\ldots,i_k,j} \delta_{i_1,\ldots,i_k,j}^k \leq \varepsilon_n \quad \text{for } k = 1, \ldots, N-1,$$

where we recall $\mathsf{x}_{i_k}^k = (x_{i_k,1}^k, \ldots, x_{i_k,d}^k)$. $\square$

Having obtained a general convergence result, we next turn to the ensuing problem on the convergence rate of $\mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n)$. We provide an estimation of the convergence rate for the one-step model on real line. To the best of our knowledge, the error bound below is the first of its kind in the literature.

THEOREM 2.5. *Let $N = 2$ and $d = 1$, or equivalently, $\boldsymbol{\mu} = (\mu, \nu)$ and $c : \mathbb{R}^2 \to \mathbb{R}$. In addition to the conditions of Theorem 2.2, we assume that $\sup_{(x,y)\in\mathbb{R}^2} |\partial_{yy}^2 c(x, y)| < \infty$ and $\nu$ has a finite second moment. Then there exists $C > 0$ such that*

$$\left|\mathsf{P}_{\varepsilon_n}(\mu^n, \nu^n) - \mathsf{P}(\mu, \nu)\right| \leq C \inf_{R>0} \lambda_n(R) \quad \textit{for all } n \geq 1,$$

*where $\lambda_n : (0, \infty) \to \mathbb{R}$ is given by*

$$\lambda_n(R) := (R+1)\varepsilon_n + \int_{(-\infty,-R)\cup(R,\infty)} (|y| - R)^2 \nu(dy).$$

*In particular, the convergence rate is proportional to $\varepsilon_n$ if* $\operatorname{supp}(\nu)$ *is bounded.*

We postpone the proofs of Theorems 2.2 and 2.5 to Section 4, and end this section with a discussion about how Theorem 2.2 is applied to solve other constrained $\mathsf{OT}$ problems arising in finance.

REMARK 2.6. In general, the distributions $\mu_1, \ldots, \mu_N$ will not be fully specified by the market when $d \geq 2$. For $k = 1, \ldots, N$, let $\mathsf{S}_k := (S_k^{(1)}, \ldots, S_k^{(d)})$, where $S_k^{(i)}$ stands for the price of the $i$th stock at time $k$. Then, in practice, only prices of call options $(S_k^{(i)} - K)^+$, or put options $(K - S_k^{(i)})^+$, for a finite set of strikes

$K$ are liquidly available in the market. Even assuming call options are quoted for all possible strikes $K$ only yields the distributions $\mu_{k,i}$ of $S_k^{(i)}$. Therefore, this leads to a modified optimization problem. Denote $\vec{\mu}_k := (\mu_{k,1}, \dots, \mu_{k,d})$ and $\vec{\boldsymbol{\mu}} := (\vec{\mu}_k)_{1\le k\le N}$, and let $\mathcal{M}_\varepsilon(\vec{\boldsymbol{\mu}})$ be the set of $\varepsilon$-approximating martingale measures $\mathbb{P}$ satisfying $\mathbb{P} \circ (S_k^{(i)})^{-1} = \mu_{k,i}$, for $k = 1, \dots, N$ and $i = 1, \dots, d$. Then we define the optimization problem by

$$\mathsf{P}_\varepsilon(\vec{\boldsymbol{\mu}}) := \sup_{\mathbb{P}\in\mathcal{M}_\varepsilon(\vec{\boldsymbol{\mu}})} \mathbb{E}_{\mathbb{P}}\big[c(\mathsf{S}_1, \dots, \mathsf{S}_N)\big]. \tag{9}$$

The problem (9), with $\varepsilon = 0$, was first introduced by Lim and called *multi-martingale optimal transport* in [33]. Although this paper focuses on the numerical computation of $\mathsf{P}(\boldsymbol{\mu})$, we emphasise that Theorem 2.2 admits an immediate extension to approximate $\mathsf{P}_0(\vec{\boldsymbol{\mu}})$.

## 3. A numerical scheme for $\mathsf{P}(\boldsymbol{\mu})$: Probability discretization.

Motivated by Theorem 2.2 and Corollary 2.4, we next develop a numerical scheme to compute $\mathsf{P}(\boldsymbol{\mu})$ based on a suitable discretization of the marginal distributions. The key is to select a suitable sequence $(\boldsymbol{\mu}^n)_{n\ge1}$ such that, for $k = 1, \dots, N$,

(a) $\mu_k^n$ is supported on a finite set $\{\mathsf{x}_{i_k}^k : i_k \in I_k\}$,

(b) the weights $\mu_k^n[\{\mathsf{x}_{i_k}^k\}]$ can be either computed explicitly or approximated with a precision that is known *a priori*,

(c) an upper bound for $\mathcal{W}(\mu_k^n, \mu_k)$ is easy to obtain.

Posed as above, the problem is intimately linked to the optimal quantization for probability measures whose goal is to best approximate a given probability measure $\mu \in \mathcal{P}$ by a discrete measure with a given number of supporting points. For the given $\mu$, its $n^d$-quantization $\mu^n$ related to $(\mathsf{x}_i)_{1\le i\le n^d} \subset \mathbb{R}^d$ and $(E_i)_{1\le i\le n^d}$ is defined by $\mu^n(d\mathsf{x}) := \sum_{i=1}^{n^d} \mu[E_i]\delta_{\mathsf{x}_i}(d\mathsf{x})$, where $(E_i)_{1\le i\le n^d}$ is a $\mu$-partition, that is, $\mu[E_i \cap E_j] = 0$ for all $i \ne j$ and $\mu[\bigcup_{1\le i\le n^d} E_i] = 1$. Accordingly, the $n^d$-optimal quantization of $\mu$ is the solution to

$$\inf\left\{\sum_{i=1}^{n^d} \int_{E_i} |\mathsf{x} - \mathsf{x}_i|\mu(d\mathsf{x})\right\}, \tag{10}$$

where the inf is taken over all $(\mathsf{x}_i)_{1\le i\le n^d}$ and $\mu$-partitions $(E_i)_{1\le i\le n^d}$. We state the convergence result from Graf and Luschgy [23]; see also [17, 24].

THEOREM 3.1 (Graf and Luschgy). *For each $n \ge 1$, the inf in (10) can be achieved by an $n^d$-optimal quantiser $(\mathsf{x}_i^*)_{1\le i\le n^d}$ and $(E_i^*)_{1\le i\le n^d}$. Let $\mu_*^n$ be the corresponding optimiser, then $\lim_{n\to\infty} n\mathcal{W}(\mu_*^n, \mu)$ exists and is finite.*

It follows with $A_\mu := \lim_{n\to\infty} n\mathcal{W}(\mu_*^n, \mu)$ that there exists $n_\mu \ge 1$ such that

$$\mathcal{W}(\mu_*^n, \mu) \le (A_\mu + 1)/n \quad \text{for all } n \ge n_\mu.$$

Despite their theoretical appeal, in practice, the use of optimal quantisers is problematic as the key quantities above, such as $A_\mu$ and $n_\mu$ are in general unknown. Similarly, in general, the quantities $\mu_*^n[\{\mathsf{x}_i^*\}]$ are hard to compute exactly or approximate with a prescribed accuracy. To overcome these difficulties, we adopt two different discretization methods, both of which can be implemented in practice. Our first method, which we call *deterministic discretization*, applies when we are given the marginals $\mu_1, \dots, \mu_N$ in the sense of being able to compute integrals against them. This is the case, for example, when $\mu_1, \dots, \mu_N$ have known density functions. The second method, called *random discretization*, applies when we are able to sample from the marginals. Throughout Section 3, we need the following integrability condition, which corresponds to the market price of a power option being finite.

ASSUMPTION 3.2 ($\theta$th-moment). There exist $\theta > 1$ and $M_\theta < \infty$ such that

$$\int_{\mathbb{R}^d} |\mathsf{x}|^\theta \mu_N(d\mathsf{x}) \le M_\theta.$$

Note that, by Jensen's inequality, the above conditions implies $\int_{\mathbb{R}^d} |\mathsf{x}|^\theta \mu_k(d\mathsf{x}) \le M_\theta$ for $k = 1, \dots, N$. Further, whenever we consider a generic measure $\mu$ below, we will also assume it satisfies Assumption 3.2.

3.1. *Deterministic discretization.* We devise a simple discretization procedure which has the same asymptotic efficiency as the optimal quantization when $\mathrm{supp}(\mu)$ is bounded. We assume here that $\mu$ is known in the sense that the probabilities $\mu[E]$ are known for all $E \in \mathcal{B}(\mathbb{R}^d)$. We start with this idealised setting and then consider the case of known densities, which allows to compute $\mu[E]$ with a certain accuracy.

*Step* 1: *Truncation.* For $R > 0$, let $\mathsf{B}_R \subset \mathbb{R}^d$ denote the box defined by

$$\mathsf{B}_R := \{\mathsf{x} = (x_1, \dots, x_d) : |x_i| \le R, \text{ for } i = 1, \dots, d\}.$$

Then one has $\{\mathsf{x} \in \mathbb{R}^d : |\mathsf{x}| \le R\} \subset \mathsf{B}_R \subset \{\mathsf{x} \in \mathbb{R}^d : |\mathsf{x}| \le dR\}$. Take $R$ such that $\mu[\mathsf{B}_R] > 0$, and truncate $\mu$ into a probability measure $\mu_R(d\mathsf{x}) := \mathbb{1}_{\mathsf{B}_R}(\mathsf{x})\mu(d\mathsf{x}) + \mu[\mathsf{B}_R^c]\delta_0(d\mathsf{x})$, where $\mathsf{B}_R^c := \mathbb{R}^d \setminus \mathsf{B}_R$. Clearly, $\mu_R$ is supported on $\mathsf{B}_R$. Consider a random variable $X$ drawn from $\mu$ and observe that $\mathbb{1}_{\mathsf{B}_R}(X)X$ is distributed according to $\mu_R$. We have, by the definition of Wasserstein distance,

$$\mathcal{W}(\mu_R, \mu) \le \mathbb{E}\big[|\mathbb{1}_{\mathsf{B}_R}(X)X - X|\big] = \int_{\mathsf{B}_R^c} |\mathsf{x}|\mu(d\mathsf{x}) \le M_\theta / R^{\theta-1}, \tag{11}$$

which yields in particular $\lim_{R\to\infty} \mathcal{W}(\mu_R, \mu) = 0$.

*Step* 2: *Discretization*. In what follows, $n \geq 2$ is an integer. Denote by $\Omega_n \subset \mathbb{R}^d$ the countable subspace consisting of elements $\mathsf{q}/n$ for all $\mathsf{q} = (q_1, \dots, q_d) \in \mathbb{Z}^d$. For each $\mathsf{q} \in \mathbb{Z}^d$, we denote by $V(\mathsf{q}/n) \subset \mathbb{R}^d$ the subset of $\mathsf{x} = (x_1, \dots, x_d)$ such that $\lfloor n\mathsf{x} \rfloor = \mathsf{q}$, that is, $\lfloor nx_i \rfloor = q_i$ for $i = 1, \dots, d$, where for $a \in \mathbb{R}$, $\lfloor a \rfloor \in \mathbb{Z}$ is the largest integer less or equal to $a$. We construct a probability measure $\mu^{(n)}$ whose support is included in $\Omega_n$ by $\mu^{(n)}[\{\mathsf{q}/n\}] := \mu[V(\mathsf{q}/n)]$. Then $\mu^{(n)} \in \mathcal{P}$ satisfies for all $f \in \Lambda$,

$$\int f\, d\mu^{(n)} = \sum_{\mathsf{q} \in \mathbb{Z}^d} f(\mathsf{q}/n)\mu^{(n)}[\{\mathsf{q}/n\}] = \int f^{(n)}\, d\mu, \tag{12}$$

where $f^{(n)} : \mathbb{R}^d \to \mathbb{R}$ is defined by $f^{(n)}(\mathsf{x}) := f(\lfloor n\mathsf{x} \rfloor / n)$. This implies in view of (4) that

$$\mathcal{W}(\mu^{(n)}, \mu) = \sup_{f \in \Lambda_1} \left| \int f\, d\mu^{(n)} - \int f\, d\mu \right| \leq \sup_{f \in \Lambda_1} \int |f^{(n)} - f|\, d\mu \leq d/n,$$

where the second inequality is by (12). Notice that, if $\mathrm{supp}(\mu)$ is bounded, then so is $\mathrm{supp}(\mu^{(n)})$, and the distance $\mathcal{W}(\mu^{(n)}, \mu)$ is of order $1/n$, which is the same as for $\mathcal{W}(\mu^n_*, \mu)$.

*Step* 3: *Choice of the parameters*. Replacing $\mu$ by $\mu_R$ in *Step 2*, one has

$$\mathcal{W}(\mu_R^{(n)}, \mu) \leq \mathcal{W}(\mu_R^{(n)}, \mu_R) + \mathcal{W}(\mu_R, \mu) \leq d/n + M_\theta / R^{\theta - 1}.$$

It follows from Young's inequality that $|a|^\gamma + |b|^\theta \geq \gamma^{1/\gamma}\theta^{1/\theta}|ab|$ for all $a$, $b \in \mathbb{R}$, where $\gamma > 1$ is the conjugate number of $\theta$, that is, $1/\theta + 1/\gamma = 1$. Setting respectively $a = (d/n)^{1/\gamma}$ and $b = (M_\theta / R^{\theta-1})^{1/\theta}$, it holds that $d/n + M_\theta / R^{\theta-1} \geq (\gamma d)^{1/\gamma}(\theta M_\theta)^{1/\theta}/(Rn)^{1/\gamma}$, and the equality can be achieved for $R^{\theta-1} = \theta M_\theta n / \gamma d$. Since the cardinal of $\mathrm{supp}(\mu_R^{(n)})$ is proportional to $(Rn)^d$ which determines the number of variables in the corresponding LP problem, setting $R = R_n := (\theta M_\theta n / \gamma d)^{1/(\theta-1)}$ leads to an optimal upper bound for a fixed computational complexity, that is, $\mathcal{W}(\mu_{R_n}^{(n)}, \mu) \leq \gamma d/n$. Replacing $\mu$ respectively by $\mu_k$ for $k = 1, \dots, N$, we obtain $\boldsymbol{\mu}^n = (\mu_k^n)_{1 \leq k \leq N}$ following the above steps with $\mu_k^n := \mu_{k,R_n}^{(n)}$. Then Theorem 2.2 yields $\lim_{n\to\infty} \mathsf{P}_{N\gamma d/n}(\boldsymbol{\mu}^n) = \mathsf{P}(\boldsymbol{\mu})$.

REMARK 3.3. In general $\boldsymbol{\mu}^n$ may no longer belong to $\mathcal{P}^{\preceq}$, even if $\boldsymbol{\mu} \in \mathcal{P}^{\preceq}$. When $d = 1$, an explicit discretization preserving the increasing convex order is given in Section 4.2. In a recent parallel work Alfonsi et al. [1] have investigated methods of constructing $\boldsymbol{\mu}^n$ such that $\boldsymbol{\mu}^n \in \mathcal{P}^{\preceq}$.

The above analysis allows us to construct approximating measures $\mu^n$ assuming the values $\mu[V(\mathsf{q}/n)]$ are known for all $\mathsf{q}/n \in \Omega_n$. This may be possible, for example, when $\mu$ is atomic, but in general we need to argue how to approximate well such values. We do this for measures which admit a density function, that is, $\mu(d\mathsf{x}) = \rho(\mathsf{x})\,d\mathsf{x}$. In this case, a simple point estimate $\rho(\mathsf{x}_\mathsf{q})/n^d$, for some $\mathsf{x}_\mathsf{q} \in V(\mathsf{q}/n)$, provides a natural candidate to approximate $\mu^{(n)}[\{\mathsf{q}/n\}]$. However, to use Theorem 2.2, we need to bound the Wasserstein distance between the resulting measure and $\mu^{(n)}$ in an explicit and nonasymptotic manner.

As before, we truncate $\mu$ to $\mathsf{B}_R$ and set $R$ to be an integer $m$ for simplicity. Let $\mu_m^{(n)}$ and $\tilde{\mu}_m^{(n)}$ be supported on $\Omega_n \cap \mathsf{B}_m$ and defined as follows: If $\mathsf{0} \neq \mathsf{q}/n \in \mathsf{B}_m$, then

$$\mu_m^{(n)}\big[\{\mathsf{q}/n\}\big] := \int_{V(\mathsf{q}/n)} \rho(\mathsf{x})\,d\mathsf{x} \quad \text{and} \quad \tilde{\mu}_m^{(n)}\big[\{\mathsf{q}/n\}\big] := \rho(\mathsf{x}_\mathsf{q})/n^d,$$

where $\mathsf{x}_\mathsf{q}^{(n)} \equiv \mathsf{x}_\mathsf{q} \in V(\mathsf{q}/n)$ are chosen arbitrarily, and

$$\mu_m^{(n)}\big[\{\mathsf{0}\}\big] := 1 - \sum_{\mathsf{q}/n\neq\mathsf{0}} \mu_m^{(n)}\big[\{\mathsf{q}/n\}\big] \quad \text{and} \quad \tilde{\mu}_m^{(n)}\big[\{\mathsf{0}\}\big] := 1 - \sum_{\mathsf{q}/n\neq\mathsf{0}} \tilde{\mu}_m^{(n)}\big[\{\mathsf{q}/n\}\big],$$

where the sums above are finite as $\mu_m^{(n)}[\{\mathsf{q}/n\}] = \tilde{\mu}_m^{(n)}[\{\mathsf{q}/n\}] = 0$ for $\mathsf{q}/n \notin \mathsf{B}_m$. We point out that in general $\tilde{\mu}_m^{(n)}[\{\mathsf{0}\}]$ may be negative and then $\tilde{\mu}_m^{(n)}$ is a signed measure. However, if $\mathsf{x}_\mathsf{q}$ satisfies $\rho(\mathsf{x}_\mathsf{q}) = \min_{\mathsf{x}\in V(\mathsf{q}/n)} \rho(\mathsf{x})$, or more generally is such that $\rho(\mathsf{x}_\mathsf{q})/n^d \le \int_{V(\mathsf{q}/n)} \rho(\mathsf{x})\,d\mathsf{x}$ holds for every $\mathsf{q}/n \in \Omega_n$, then $\tilde{\mu}_m^{(n)}[\{\mathsf{0}\}] \ge 0$ and $\tilde{\mu}_m^{(n)}$ is a probability measure. In this case, the following proposition provides an upper bound for $\mathcal{W}(\tilde{\mu}_m^{(n)}, \mu_m^{(n)})$.

PROPOSITION 3.4. *Let Assumption 3.2 hold. Suppose that $\rho$ is continuous, or namely, for each $R > 0$, there exists $\kappa_R : [0,\infty) \to \mathbb{R}$ that is nondecreasing such that $\kappa_R(0) = 0$ and*

$$\big|\rho(\mathsf{x}) - \rho(\mathsf{y})\big| \le \kappa_R\big(|\mathsf{x} - \mathsf{y}|\big) \quad \text{for all } \mathsf{x}, \mathsf{y} \in \mathsf{B}_R.$$

*Assume further that $\tilde{\mu}_m^{(n)}[\{\mathsf{0}\}] \ge 0$.*

(i) *If $\mu$ has bounded support, that is,* $\operatorname{supp}(\mu) \subseteq \mathsf{B}_R$ *for some $R > 0$, then*

$$\mathcal{W}\big(\tilde{\mu}_m^{(n)}, \mu\big) \le \varepsilon_n := d/n + 2^d d(R+1)^{d+1}\kappa_{R+1}(d/n) \quad \text{for all } m \ge \lfloor R \rfloor + 1. \tag{13}$$

(ii) *If $\rho$ is uniformly continuous, that is, there exists a uniform $\kappa = \kappa_R$ for all $R > 0$, then*

$$\mathcal{W}\big(\tilde{\mu}_m^{(n)}, \mu\big) \le \varepsilon_{m,n} := d/n + M_\theta/m^{\theta-1} + 2^d d m^{d+1}\kappa(d/n). \tag{14}$$

(iii) *If $\{\mathsf{x}_\mathsf{q}\}_{\mathsf{0}\neq\mathsf{q}/n\in\mathsf{B}_m}$ satisfies $\rho(\mathsf{x}_\mathsf{q}) \le \rho(\mathsf{x})$ for all $\mathsf{x} \in V(\mathsf{q}/n)$, then*

$$\begin{aligned}\mathcal{W}\big(\tilde{\mu}_m^{(n)}, \mu\big) \le \tau_{m,n} :={}& d/n + M_\theta/m^{\theta-1} \\ &+ \inf_{1\le j\le m}\big\{2^d d j^{d+1}\kappa_j(d/n) + 4M_\theta/j^{\theta-1}\big\}.\end{aligned} \tag{15}$$

PROOF. Note that $\tilde{\mu}_m^{(n)}$ is a probability measure and $\mathcal{W}(\mu_m^{(n)}, \mu) \le d/n + M_\theta/m^{\theta-1}$. The following analysis gives an upper bound for $\mathcal{W}(\tilde{\mu}_m^{(n)}, \mu_m^{(n)})$. Under the condition (i), one has $\mu = \mu_m$ as $m \ge \lfloor R \rfloor + 1$. Since $\rho$ is uniformly continuous on $\mathsf{B}_R$, $\tilde{\mu}_m^{(n)}$ becomes a probability measure as $\tilde{\mu}_m^{(n)}[\{0\}] \ge 0$ $n$ is sufficiently large. Further, $\mathcal{W}(\mu_m^{(n)}, \mu) = \mathcal{W}(\mu_m^{(n)}, \mu_m) \le d/n$. Note also $\operatorname{supp}(\mu_m^{(n)})$, $\operatorname{supp}(\tilde{\mu}_m^{(n)}) \subseteq \mathsf{B}_{\lfloor R \rfloor+1}$ by definition. For any $f \in \Lambda_1$, it holds for $m \ge \lfloor R \rfloor + 1$,

$$\begin{aligned}\left|\int f d\tilde{\mu}_m^{(n)} - \int f\, d\mu_m^{(n)}\right| &= \left|\sum_{0 \ne \mathsf{q}/n \in \mathsf{B}_{\lfloor R \rfloor+1}} f(\mathsf{q}/n) \int_{V(\mathsf{q}/n)} \big(\rho(\mathsf{x}) - \rho(\mathsf{x}_\mathsf{q})\big)\, d\mathsf{x}\right| \\ &\le \left|\sum_{0 \ne \mathsf{q}/n \in \mathsf{B}_{\lfloor R \rfloor+1}} |\mathsf{q}/n| \int_{V(\mathsf{q}/n)} \kappa_{R+1}(d/n)\, d\mathsf{x}\right| \\ &\le 2^d d(R+1)^{d+1} \kappa_{R+1}(d/n),\end{aligned}$$

which yields $\mathcal{W}(\tilde{\mu}_m^{(n)}, \mu_m^{(n)}) \le 2^d d(R+1)^{d+1}\kappa_{R+1}(d/n)$ and further (13). As for (ii), we deduce $\mathcal{W}(\tilde{\mu}_m^{(n)}, \mu_m^{(n)}) \le 2^d\, dm^{d+1}\kappa(d/n)$ using the same arguments for $\mathsf{B}_m$, and obtain thus (14). Alternatively, assume that the third condition holds. For each integer $1 \le j \le m$, one has

$$\begin{aligned}&\left|\int f\, d\tilde{\mu}_m^{(n)} - \int f\, d\mu_m^{(n)}\right| \\ &\quad \le \left|\sum_{0 \ne \mathsf{q}/n \in \mathsf{B}_j} |\mathsf{q}/n| \int_{V(\mathsf{q}/n)} \big(\rho(\mathsf{x}) - \rho(\mathsf{x}_\mathsf{q})\big)\, d\mathsf{x}\right| \\ &\quad\quad + \left|\sum_{\mathsf{q}/n \in \mathsf{B}_m \setminus B_j} \left|\mathsf{q}/n\right| \int_{V(\mathsf{q}/n)} \big(\rho(\mathsf{x}) - \rho(\mathsf{x}_\mathsf{q})\big)\, d\mathsf{x}\right| \\ &\quad \le 2^d dj^{d+1}\kappa_j(d/n) + 2\int_{\mathsf{B}_j^c} (d + |\mathsf{x}|)\rho(\mathsf{x})\, d\mathsf{x} \\ &\quad \le 2^d dj^{d+1}\kappa_j(d/n) + 4M_\theta/j^{\theta-1}.\end{aligned}$$

Thus $\mathcal{W}(\tilde{\mu}_m^{(n)}, \mu_m^{(n)}) \le \inf_{1 \le j \le m}\{2^d dj^{d+1}\kappa_j(d/n) + 4M_\theta/j^{\theta-1}\}$ follows and (15) is derived. $\square$

By a straightforward computation, one has $\lim_{m\to\infty}\lim_{n\to\infty}\varepsilon_{m,n} = \lim_{m\to\infty}\lim_{n\to\infty}\tau_{m,n} = 0$. In consequence, there exist suitable sequences $(m_n)_{n\ge1}$ and $(n_m)_{m\ge1}$, such that

$$\lim_{n\to\infty}\varepsilon_{m_n,n} = \lim_{n\to\infty}\tau_{m_n,n} = 0 \quad \text{and} \quad \lim_{m\to\infty}\varepsilon_{m,n_m} = \lim_{m\to\infty}\tau_{m,n_m} = 0.$$

Note that the previous choice $m_n := \lfloor R_n \rfloor$ may not yield $\lim_{n\to\infty}\varepsilon_{\lfloor R_n \rfloor,n} = \lim_{n\to\infty}\tau_{\lfloor R_n \rfloor,n} = 0$ and these sequences have to be computed from $\rho$. However, if

$\rho$ is $L$-Lipschitz, then one has $\varepsilon_{m,n} = d/n + M_\theta/m^{\theta-1} + 2^d d^2 m^{d+1} L/n$ and we deduce that it suffices to take $(m_n)_{n\geq 1}$ or $(n_m)_{m\geq 1}$ such that $\lim_{n\to\infty} m_n^{d+1}/n = 0$ or $\lim_{m\to\infty} m^{d+1}/n_m = 0$.

Putting everything together, taking respectively $\mu_k$ in the place of $\mu$ for $k = 1,\dots,N$, the above procedures yield a vector of measures $\tilde{\boldsymbol{\mu}}_m^{(n)} = (\tilde{\mu}_{k,m}^{(n)})_{1\leq k\leq N}$. Under the conditions in Proposition 3.4, we have $\lim_{n\to\infty} \mathsf{P}_{N\varepsilon_{m_n,n}}(\tilde{\boldsymbol{\mu}}_{m_n}^{(n)}) = \lim_{n\to\infty} \mathsf{P}_{N\tau_{m_n,n}}(\tilde{\boldsymbol{\mu}}_{m_n}^{(n)}) = \mathsf{P}(\boldsymbol{\mu})$ or $\lim_{m\to\infty} \mathsf{P}_{N\varepsilon_{m,n_m}}(\tilde{\boldsymbol{\mu}}_m^{(n_m)}) = \lim_{m\to\infty} \mathsf{P}_{N\tau_{m,n_m}}(\tilde{\boldsymbol{\mu}}_m^{(n_m)}) = \mathsf{P}(\boldsymbol{\mu})$.

REMARK 3.5. We note that an alternative construction for $\mu_m^{(n)}$ and $\tilde{\mu}_m^{(n)}$, based on a renormalisation, is also possible. This might be desirable, in particular, if it is hard to choose $\mathsf{x}_\mathsf{q}$ is such a way that $\tilde{\mu}_m^{(n)}[\{\mathsf{0}\}] \geq 0$. Namely, we could set, for $\mathsf{q}/n \in \mathsf{B}_m$,

$$\mu_m^{(n)}[\{\mathsf{q}/n\}] := \frac{\int_{V(\mathsf{q}/n)} \rho(\mathsf{x})\,d\mathsf{x}}{\sum_{\mathsf{q}'/n\in\mathsf{B}_m} \int_{V(\mathsf{q}'/n)} \rho(\mathsf{x})\,d\mathsf{x}} \quad \text{and}$$
$$\tilde{\mu}_m^{(n)}[\{\mathsf{q}/n\}] := \frac{\rho(\mathsf{x}_\mathsf{q})}{\sum_{\mathsf{q}'/n\in\mathsf{B}_m} \rho(\mathsf{x}_{\mathsf{q}'})},$$

and $\mu_m^{(n)}[\{\mathsf{q}/n\}] = \tilde{\mu}_m^{(n)}[\{\mathsf{q}/n\}] = 0$ otherwise. Then $\mu_m^{(n)}$ and $\tilde{\mu}_m^{(n)}$ are probability measures supported on $\Omega_n \cap \mathsf{B}_m$ and using entirely analogous reasoning as above, one can obtain the bounds for $\mathcal{W}(\mu_m^{(n)}, \mu)$ and $\mathcal{W}(\tilde{\mu}_m^{(n)}, \mu_m^{(n)})$.

3.2. *Random discretization.* We consider now a different discretization procedure, which applies to the case where one has a black box to generate independent random variables according to $\mu$. Provided a sequence of i.i.d. $\mu$-distributed random variables $(X_n)_{n\geq 1}$, define the empirical measure $\hat{\mu}^n$ by

$$\hat{\mu}^n(d\mathsf{x}) := \sum_{i=1}^n \frac{1}{n}\delta_{X_i}(d\mathsf{x}).$$

By definition $\hat{\mu}^n$ is a random measure, and following Glivenko–Cantelli's theorem (see, e.g., Fournier and Guillin [19]), $\lim_{n\to\infty} \mathcal{W}(\hat{\mu}^n, \mu) = 0$ almost surely and $\lim_{n\to\infty} \mathbb{E}[\mathcal{W}(\hat{\mu}^n, \mu)] = 0$. Construct random measures $\hat{\mu}_k^n$ by replacing $\mu$ by $\mu_k$ for $k = 1,\dots,N$ and set $\hat{\boldsymbol{\mu}}^n := (\hat{\mu}_k^n)_{1\leq k\leq N}$. Compared to Theorem 2.2, we now obtain a stochastic convergence result.

PROPOSITION 3.6. *Let the conditions of Theorem 2.2 hold. Given a sequence $(\varepsilon_m)_{m\geq 1} \subset (0,\infty)$ converging to zero, one has $\lim_{m\to\infty}\lim_{n\to\infty} \mathsf{P}_{\varepsilon_m}(\hat{\boldsymbol{\mu}}^n) = \mathsf{P}(\boldsymbol{\mu})$ almost surely. Further, for any subsequence $(\hat{n}_m)_{m\geq 1}$ such that $\sum_{m\geq 1} \mathbb{E}[\mathcal{W}^{\oplus}(\hat{\boldsymbol{\mu}}^{\hat{n}_m}, \boldsymbol{\mu})]/\varepsilon_m < \infty$, it holds almost surely $\lim_{m\to\infty} \mathsf{P}_{\varepsilon_m}(\hat{\boldsymbol{\mu}}^{\hat{n}_m}) = \mathsf{P}(\boldsymbol{\mu})$.*

PROOF. For each fixed $\varepsilon_m > 0$, one has the inequality below by Corollary 4.3,

$$\begin{aligned}(16)\qquad &\left|\mathsf{P}_{\varepsilon_m}(\hat{\boldsymbol{\mu}}^n) - \mathsf{P}(\boldsymbol{\mu})\right| \le \mathrm{Lip}(c)\varepsilon_m + \mathsf{P}_{2\varepsilon_m}(\boldsymbol{\mu}) - \mathsf{P}(\boldsymbol{\mu}) \\ &\quad \text{whenever } \mathcal{W}^{\oplus}(\hat{\boldsymbol{\mu}}^n, \boldsymbol{\mu}) \le \varepsilon_m,\end{aligned}$$

where we recall that $\mathrm{Lip}(c)$ is the Lipschitz constant of $c$. Applying Glivenko–Cantelli's theorem, one obtains $\lim_{n\to\infty} |\mathsf{P}_{\varepsilon_m}(\hat{\boldsymbol{\mu}}^n) - \mathsf{P}(\boldsymbol{\mu})| \le \mathrm{Lip}(c)\varepsilon_m + \mathsf{P}_{2\varepsilon_m}(\boldsymbol{\mu}) - \mathsf{P}(\boldsymbol{\mu}) =: \delta_m$ almost surely. Using Proposition 4.4, we have $\lim_{m\to\infty} \mathsf{P}_{2\varepsilon_m}(\boldsymbol{\mu}) = \mathsf{P}(\boldsymbol{\mu})$ and the first asserted convergence result follows. For any $\delta > 0$, there exists $m_\delta$ such that $\varepsilon_m \le \delta$ for all $m \ge m_\delta$. Taking $\hat{n}_m$ as in the statement, we have

$$\begin{aligned}\sum_{m\ge m_\delta} \mathbb{P}\left[\left|\mathsf{P}_{\varepsilon_m}(\hat{\boldsymbol{\mu}}^{\hat{n}_m}) - \mathsf{P}(\boldsymbol{\mu})\right| > \delta\right] &\le \sum_{m\ge m_\delta} \mathbb{P}\left[\mathcal{W}^{\oplus}(\hat{\boldsymbol{\mu}}^{\hat{n}_m}, \boldsymbol{\mu}) > \varepsilon_m\right] \\ &\le \sum_{m\ge m_\delta} \mathbb{E}\left[\mathcal{W}^{\oplus}(\hat{\boldsymbol{\mu}}^{\hat{n}_m}, \boldsymbol{\mu})\right]/\varepsilon_m,\end{aligned}$$

which by Borel–Cantelli's lemma, implies that $\lim_{m\to\infty} \mathsf{P}_{\varepsilon_m}(\hat{\boldsymbol{\mu}}^{\hat{n}_m}) = \mathsf{P}(\boldsymbol{\mu})$ almost surely. □

Roughly speaking, given $\varepsilon_m > 0$, it suffices to focus on the LP problems $\mathsf{P}_{\varepsilon_m}(\hat{\boldsymbol{\mu}}^n)$ such that $\mathcal{W}^{\oplus}(\hat{\boldsymbol{\mu}}^n, \boldsymbol{\mu}) \le \varepsilon_m$ occurs with high probability. To this end, and in order to choose suitable sequences $(\varepsilon_m)_{m\ge 1}$ and $(\hat{n}_m)_{m\ge 1}$, we need to quantify $\mathbb{E}[\mathcal{W}^{\oplus}(\hat{\boldsymbol{\mu}}^n, \boldsymbol{\mu})]$. Fortunately, Theorem 1 in [19] provides such an estimation under Assumption 3.2, albeit with some cases omitted, for example, $d = 1, 2$ and $\theta = 2$. For the sake of completeness, we state this result as Lemma 3.7 by taking all the cases into account.

LEMMA 3.7 ([19]). *Let Assumption 3.2 hold. There exists $C(\theta, d) > 0$ such that $\mathbb{E}[\mathcal{W}^{\oplus}(\hat{\boldsymbol{\mu}}^n, \boldsymbol{\mu})] \le \chi(n)$ for all $n \ge 1$, where*

$$\chi(n) := \mathrm{N}C(\theta, d) \begin{cases} n^{1/\theta - 1} & \text{if } d = 1 \text{ and } 1 < \theta < 2, \\ (1 + \log n)n^{-1/2} & \text{if } d = 1 \text{ and } \theta = 2, \\ n^{-1/2} & \text{if } d = 1 \text{ and } \theta > 2, \\ n^{1/\theta - 1} & \text{if } d = 2 \text{ and } 1 < \theta < 2, \\ \left(1 + (\log n)^2\right)n^{-1/2} & \text{if } d = 2 \text{ and } \theta = 2, \\ (1 + \log n)n^{-1/2} & \text{if } d = 2 \text{ and } \theta > 2, \\ n^{1/\theta - 1} & \text{if } d \ge 3 \text{ and } 1 < \theta < d/(d-1), \\ (1 + \log n)n^{-1/d} & \text{if } d \ge 3 \text{ and } \theta = d/(d-1), \\ n^{-1/d} & \text{if } d \ge 3 \text{ and } \theta > d/(d-1). \end{cases}$$

*In consequence, one has $\mathbb{P}[\mathcal{W}^{\oplus}(\hat{\boldsymbol{\mu}}^n, \boldsymbol{\mu}) > \varepsilon_m] \le \chi(n)/\varepsilon_m$.*

We remark that the constants $C(\theta, d)$ are not explicitly specified in [19]. To implement our scheme, we need to determine $\hat{n}_m$ and for this we have to compute explicitly $C(\theta, d)$. This is possible, following largely the arguments in [19], but tedious and is postponed to the Appendix.

3.3. *Numerical examples.* We discuss now some concrete MOT problems to illustrate how Theorem 2.2, together with our discretization schemes, can be applied. Most of our examples admit either a closed-form optimiser or an analytical characterization thereof, which allow us to verify the numerical results. We recall that for $d = 1$, we write simply $\mathsf{x} = x$ and $\mathsf{S}_k = S_k$.

EXAMPLE 3.8. Beiglböck and Juillet studied in [4] a specific MOT problem in the case of $N = 2$, $d = 1$ and $c(x, y) = h(x - y)$, where $h : \mathbb{R} \to \mathbb{R}$ has a strictly convex derivative. Let $(\mu, \nu) \in \mathcal{P}^{\preceq}$. Theorem 1.7 of [4] shows that, if $\mu$ has a density $\rho$, then there exist two measurable functions $\xi_{\pm} : \mathbb{R} \to \mathbb{R}$ such that the unique optimiser $\mathbb{P}^* \in \mathcal{M}(\mu, \nu)$ for $\mathsf{P}(\mu, \nu)$ is supported on $\xi_{\pm}$, that is,

$$\mathbb{P}^*(dx, dy) = \mu(dx) \otimes \left\{ \frac{x - \xi_-(x)}{\xi_+(x) - \xi_-(x)} \delta_{\xi_+(x)}(dy) + \frac{\xi_+(x) - x}{\xi_+(x) - \xi_-(x)} \delta_{\xi_-(x)}(dy) \right\},$$

where $\xi_-(x) \le x \le \xi_+(x)$, $\xi_+(x) < \xi_+(x')$ and $\xi_-(x') \notin (\xi_-(x), \xi_+(x))$ for all $x$, $x' \in \mathbb{R}$ with $x < x'$. We want to illustrate numerically the above result. Let $\rho$ be a truncated Gamma function defined by

$$\rho(x) := \mathbb{1}_{[0,1]}(x) x^{3/2} e^{-x} / C, \quad \text{where } C := \int_0^1 x^{3/2} e^{-x} \, dx > 1/5.$$

Next, construct $\nu(dx) = \sigma(y)\, dy$ by

$$\begin{aligned} \sigma(y) &:= \rho(y/2)/6 + 4\rho(2y)/3 \\ &= \mathbb{1}_{[0,2]}(y)(y/2)^{3/2} e^{-y/2}/C + \mathbb{1}_{[0,1/2]}(y)(2y)^{3/2} e^{-2y}/C. \end{aligned}$$

By the construction, one has $(\mu, \nu) \in \mathcal{P}^{\preceq}$, $\operatorname{supp}(\mu) = [0, 1]$ and $\operatorname{supp}(\nu) = [0, 2]$. Further, one can verify that $\rho$ and $\sigma$ are $L$-Lipschitz on $\operatorname{supp}(\mu)$ and $\operatorname{supp}(\nu)$ with $L = 7$. Applying the discretization of Section 3.1 to $\mu$ and $\nu$, we obtain $\tilde{\mu}^{(n)}$ and $\tilde{\nu}^{(n)}$, supported on $\{i/n : 0 \le i < n\}$ and $\{j/n : 0 \le j < 2n\}$, and defined by

$$\tilde{\mu}^{(n)}[\{i/n\}] := \begin{cases} 1 - \sum_{k=1}^{n-1} \rho(x_k)/n & \text{if } i = 0, \\ \rho(x_i)/n & \text{if } 1 \le i < n, \end{cases}$$

$$\tilde{\nu}^{(n)}[\{j/n\}] := \begin{cases} 1 - \sum_{k=1}^{2n-1} \sigma(y_k)/n & \text{if } j = 0, \\ \sigma(y_j)/n & \text{if } 1 \le j < 2n, \end{cases}$$

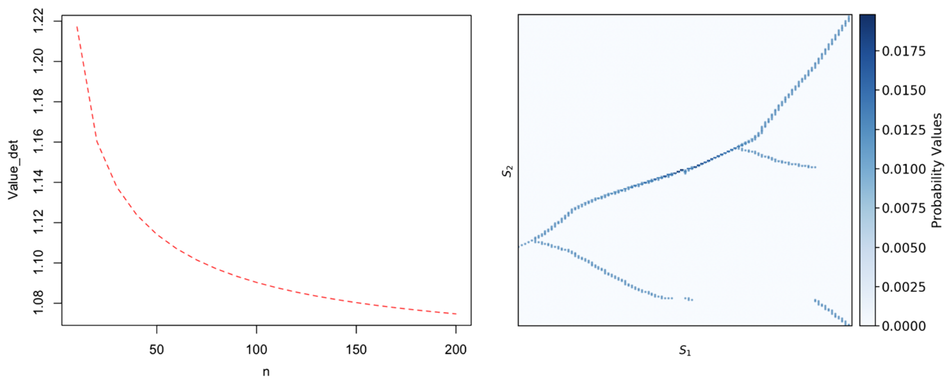


FIG. 1. *Computations for Example 3.8. The first pane shows the values* $\mathsf{P}_{\varepsilon_n}(\tilde{\mu}^{(n)}, \tilde{\nu}^{(n)})$ *for* $10 \le n \le 200$. *The second pane draws the heat map of the optimiser for* $n = 100$.

where $x_i \in [i/n, (i+1)/n)$ and $y_j \in [j/n, (j+1)/n)$ for $i = 1, \dots, n-1$ and $j = 1, \dots, 2n-1$. It follows from Proposition 3.4 that $\mathcal{W}^{\oplus}((\tilde{\mu}^{(n)}, \tilde{\nu}^{(n)}), (\mu, \nu)) \le (3L+2)/n =: \varepsilon_n$. Then the corresponding LP problem is as follows:

$$\max_{(p_{i,j}) \in \mathbb{R}_+^{2n^2}} \sum_{i=0}^{n-1} \sum_{j=0}^{2n-1} p_{i,j} h\big((i-j)/n\big) \quad \text{s.t.}$$

$$\sum_{j=0}^{2n-1} p_{i,j} = \alpha_i^n \quad \text{for } i = 0, \dots, n-1,$$

$$\sum_{i=0}^{n-1} p_{i,j} = \beta_j^n \quad \text{for } j = 0, \dots, 2n-1,$$

$$\sum_{i=0}^{n-1} \left| \sum_{j=0}^{2n-1} p_{i,j} j/n - \alpha_i^n i/n \right| \le \varepsilon_n,$$

where $\alpha_i^n := \tilde{\mu}^{(n)}[\{i/n\}]$ and $\beta_j^n := \tilde{\nu}^{(n)}[\{j/n\}]$. Taking $h(x) := e^x$, we solve the LP problem using Gurobi solver and present the results in Figure 1. The left pane exhibits the values $\mathsf{P}_{\varepsilon_n}(\tilde{\mu}^{(n)}, \tilde{\nu}^{(n)})$ for $10 \le n \le 200$, which shows numerically the convergence of $\mathsf{P}_{\varepsilon_n}(\tilde{\mu}^{(n)}, \tilde{\nu}^{(n)})$. The right pane displays the heat map of the optimiser $(p_{i,j}^*)$ for $n = 100$. We see that the strictly positive weights $p_{i,j}^*$ are concentrating around two curves that satisfy the conditions of $\xi_\pm$ stated above.

For comparison, we adopt now the random discretization developed in Section 3.2. We sample, using an accept–reject algorithm, two sequences of i.i.d. random variables $(X_i)_{1 \le i \le n}$ and $(Y_j)_{1 \le j \le n}$ from $\mu$ and $\nu$, respectively. Let $\mathcal{X}$ and $\mathcal{Y}$ be two sets containing respectively all values taken by $X_i$ and $Y_j$, and we relabel $\mathcal{X}$ and $\mathcal{Y}$ by $\mathcal{X} := \{X^1, X^2, \dots, X^{\#\mathcal{X}}\}$ and $\mathcal{Y} := \{Y^1, Y^2, \dots, Y^{\#\mathcal{Y}}\}$, where $\#\mathcal{X}$,

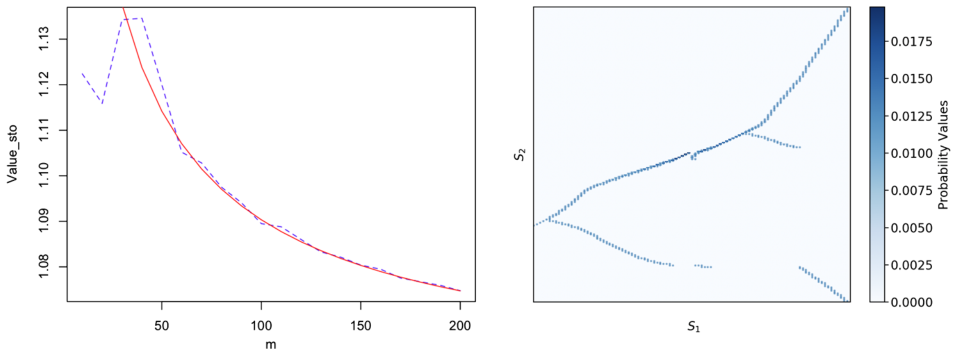


FIG. 2. *Computations for Example 3.8. The first pane shows the values* $\mathsf{P}_{\varepsilon_m}(\hat{\mu}^{\hat{n}_m}, \hat{\nu}^{\hat{n}_m})$ (*dashed line*) *for* $10 \le m \le 200$. *The second pane draws the heat map of the optimiser for* $m = 100$.

$\#\mathcal{Y} \le n$ denote the cardinal of $\mathcal{X}$ and $\mathcal{Y}$. Define further $\hat{\mu}^n(dx) := \sum_{i=1}^{\#\mathcal{X}} \hat{\alpha}_i^n \delta_{X^i}(dx)$ and $\hat{\nu}^n(dy) := \sum_{j=1}^{\#\mathcal{Y}} \hat{\beta}_i^n \delta_{Y^j}(dy)$, where

$$\hat{\alpha}_i^n := \frac{\#\mathcal{X}_i}{n} \quad \text{for } i = 1, \dots, \#\mathcal{X} \quad \text{and} \quad \hat{\beta}_j^n := \frac{\#\mathcal{Y}_j}{n} \quad \text{for } j = 1, \dots, \#\mathcal{Y},$$

with $\mathcal{X}_i := \{X_k \in \mathcal{X} : X_k = X^i\}$ and $\mathcal{Y}_j := \{X_k \in \mathcal{Y} : X_k = Y^j\}$. The LP problem $\mathsf{P}_{\varepsilon_m}(\hat{\mu}^n, \hat{\nu}^n)$ is given by

$$\max_{(p_{i,j}) \in \mathbb{R}_+^{\#\mathcal{X}\#\mathcal{Y}}} \sum_{i=1}^{\#\mathcal{X}} \sum_{j=1}^{\#\mathcal{Y}} p_{i,j} h\big(X^i - Y^j\big) \quad \text{s.t.}$$

$$\sum_{j=1}^{\#\mathcal{Y}} p_{i,j} = \hat{\alpha}_i^n \quad \text{for } i = 1, \dots, \#\mathcal{X},$$

$$\sum_{i=1}^{\#\mathcal{X}} p_{i,j} = \hat{\beta}_j^n \quad \text{for } j = 1, \dots, \#\mathcal{Y},$$

$$\sum_{i=1}^{\#\mathcal{X}} \left| \sum_{j=1}^{\#\mathcal{Y}} p_{i,j} Y^j - \hat{\alpha}_i^n X^i \right| \le \varepsilon_m.$$

Notice that $\nu$ admits a finite $\theta$th-moment for all $\theta > 1$. With $\theta = 3$, one has $\chi(n) = 2C(3,1)n^{-1/2}$, where $C(3,1)$ is defined in Proposition 3.7. We set $\hat{n}_m := \lfloor m^r \rfloor$ so that $\sum_{m \ge 1} \chi(\hat{n}_m)/\varepsilon_m < \infty$ whenever $r > 4$, and hence $\lim_{m\to\infty} \mathsf{P}_{\varepsilon_m}(\hat{\mu}^{\hat{n}_m}, \hat{\nu}^{\hat{n}_m}) = \mathsf{P}(\mu, \nu)$ holds almost surely. Taking $r = 4.1$, we compute $\mathsf{P}_{\varepsilon_m}(\hat{\mu}^{\hat{n}_m}, \hat{\nu}^{\hat{n}_m})$ and present the results in Figure 2. The blue line in the left pane shows the convergence of $\mathsf{P}_{\varepsilon_m}(\hat{\mu}^{\hat{n}_m}, \hat{\nu}^{\hat{n}_m})$ in $m$ while the red line reproduces the convergence of $\mathsf{P}_{\varepsilon_m}(\tilde{\mu}^{(m)}, \tilde{\nu}^{(m)})$ from the first pane of Figure 1. While the random discretization displays some instability for small $m$, for $m \ge 50$ we we find that two lines are very close. We note that for the same $\varepsilon_m$, the number of variables for $\mathsf{P}_{\varepsilon_m}(\tilde{\mu}^{(m)}, \tilde{\nu}^{(m)})$ is proportional to $m^2$ while it is of order $m^{2r} \gg m^2$ for $\mathsf{P}_{\varepsilon_m}(\hat{\mu}^{\hat{n}_m}, \hat{\nu}^{\hat{n}_m})$. In the right pane, the heat map of the optimiser for $m = 100$ is drawn and matches closely that of Figure 1.

EXAMPLE 3.9. Motivated by the model-independent pricing, we consider a stock with three trading dates, that is, $N = 3$ and $d = 1$. We take the Black–Scholes model, that is, $\mu_k(dx) = \rho_k(x)\,dx$ with

$$\rho_k(x) := \mathbb{1}_{(0,\infty)}(x)\frac{\exp(-(\log(x) + 2^{k-4})^2/2^{k-2})}{x\sqrt{2^{k-2}\pi}} \quad \text{for } k = 1, 2, 3,$$

and consider Lookback and Asian options, that is, $c(x, y, z) := \max(x, y, z) - z$ and $c(x, y, z) := ((x + y + z)/3 - \lambda z)^+$ with $\lambda \geq 0$. Notice that all $\rho_k$ are $L$-Lipschitz on $\mathbb{R}$ with $L = 12$, and have finite $\theta$-moments for all $\theta > 1$ with

$$\int_{\mathbb{R}} |x|^\theta \rho_k(x)\,dx \leq \int_{\mathbb{R}} |x|^\theta \rho_3(x)\,dx = e^{\theta(\theta-1)} =: M_\theta.$$

As all $\mu_k$ have unbounded support, we employ the full procedure of approximation in Section 3.1. Let $\tilde{\mu}^{(n)}_{k,m}$ be supported on $\{i/n : 0 \leq i < mn\}$, and defined by

$$\tilde{\mu}^{(n)}_{k,m}[\{i/n\}] := \begin{cases} 1 - \sum_{j=1}^{mn-1} \rho_k(x_{k,j})/n & \text{if } i = 0, \\ \rho_k(x_{k,i})/n & \text{if } 1 \leq i < mn, \end{cases}$$

where $x_{k,i} := \operatorname{argmin}_{x\in[i/n,(i+1)/n]} \rho_k(x)$ for $i = 1, \ldots, mn - 1$. Proposition 3.4 implies that

$$\begin{aligned} &\mathcal{W}^{\oplus}\big((\tilde{\mu}^{(n)}_{1,m}, \tilde{\mu}^{(n)}_{2,m}, \tilde{\mu}^{(n)}_{3,m}), (\mu_1, \mu_2, \mu_3)\big) \\ &\quad \leq 3\Big(1/n + M_\theta/m^{\theta-1} + \inf_{1\leq j\leq m}\{2j^2L/n + 4M_\theta/j^{\theta-1}\}\Big). \end{aligned}$$

Taking $j = m = m_n := \lfloor (n(\theta - 1)M_\theta/L)^{1/(\theta+1)} \rfloor$ and setting $\mu^n_k := \tilde{\mu}^{(n)}_{k,m_n}$, one has

$$\begin{aligned} &\mathcal{W}^{\oplus}\big((\mu^n_1, \mu^n_2, \mu^n_3), (\mu_1, \mu_2, \mu_3)\big) \\ &\quad \leq 3\big(1/n + M_\theta/m_n^{\theta-1} + 2m_n^2L/n + 4M_\theta/m_n^{\theta-1}\big) := \varepsilon_n, \end{aligned}$$

where $\lim_{n\to\infty} \varepsilon_n = 0$. Numerical solutions to the $\mathsf{LP}$ problems for Lookback and Asian options with $\lambda = 2$, corresponding to the above discretization, are presented in Figure 3. In Figure 4, we exhibit the heat maps of the optimisers projected respectively on $(S_1, S_2)$ and $(S_2, S_3)$ for $n = 100$. The two panes above are for the Lookback option, where conditioning on $S_1$, $S_2$ takes two values, while conditioning on $S_2$, $S_3$ may take up to four values. The two panes below are for the Asian option. It appears that $(S_1, S_2)$ is concentrated on the boundary of a quadrilateral polygon, and $(S_2, S_3)$ is on the boundaries of two disjoint quadrilateral polygons.

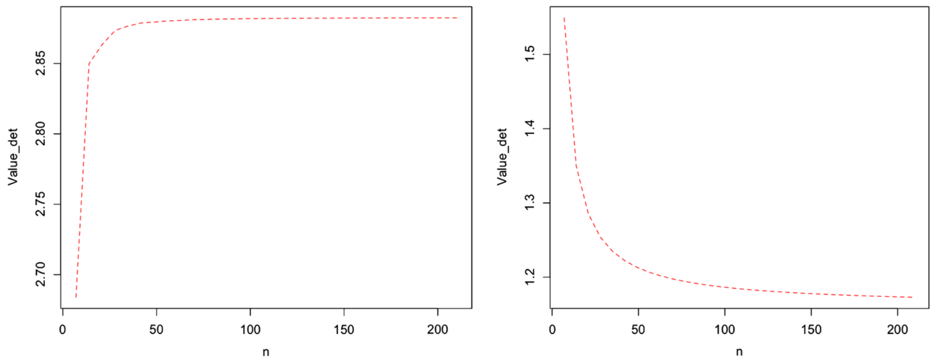


FIG. 3. *Computations for Example 3.9. The two panes show the values* $\mathsf{P}_{\varepsilon_n}(\mu_1^n, \mu_2^n, \mu_3^n)$ *for* $10 \le n \le 200$. *The left pane stands for the Lookback option and the right one for the Asian option.*

EXAMPLE 3.10. Recently, the geometry of MOT problems in general dimensions has been studied; see, for example, [16, 22, 34]. We provide here numerical evidence in two dimensions which casts doubt on Conjecture 2 of [22].

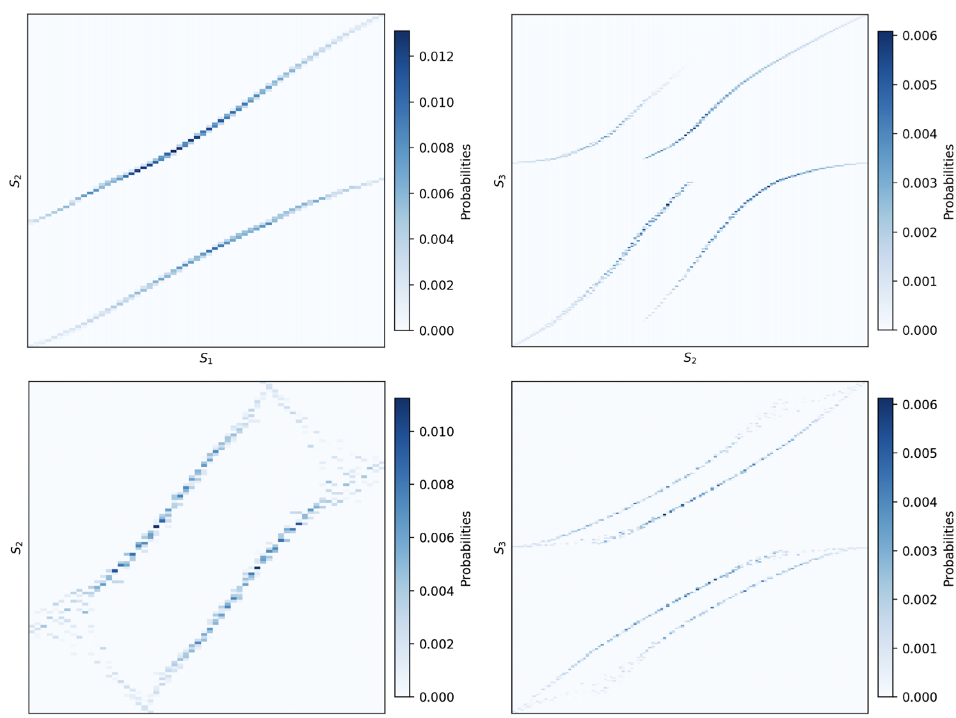


FIG. 4. *Computations for Example 3.9. The four panes exhibit the heat maps of the optimiser projected on* $(S_1, S_2)$ *and* $(S_2, S_3)$ *for* $n = 100$. *The top two correspond to the Lookback option and the bottom two correspond to the Asian option.*

Take $N = d = 2$ and $c(\mathsf{x}, \mathsf{y}) := -\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$ for all $\mathsf{x} = (x_1, x_2)$, $\mathsf{y} = (y_1, y_2) \in \mathbb{R}^2$. Given $(\mu, \nu) \in \mathcal{P}^{\preceq}$, the conjecture is stated as follows: If $\mu$ admits a density on $\mathbb{R}^2$, then the support of $\mathbb{P}^*_{\mathsf{x}}$ contains at most three points for $\mu$-a.e. $\mathsf{x} \in \mathbb{R}^2$, where $\mathbb{P}^* \in \mathcal{M}(\mu, \nu)$ is the optimiser for $\mathsf{P}(\mu, \nu)$ and $(\mathbb{P}^*_{\mathsf{x}})_{\mathsf{x} \in \mathbb{R}^2}$ is the regular conditional disintegration of $\mathbb{P}^*$ with respect to $\mu$. Let $\mu(d\mathsf{x}) = \rho(\mathsf{x})\, d\mathsf{x}$ and $\nu(d\mathsf{y}) = \sigma(\mathsf{y})\, d\mathsf{y}$ be identified by $\rho(\mathsf{x}) := \mathbb{1}_{[-1,1]^2}(\mathsf{x})/4$ and

$$\begin{aligned}\sigma(\mathsf{y}) := {} & \frac{2 - y_1}{4}\mathbb{1}_{[1,2]\times[-1,1]}(\mathsf{y}) + \frac{2 + y_1}{4}\mathbb{1}_{[-2,-1]\times[-1,1]}(\mathsf{y}) \\ & + \frac{2 - y_2}{4}\mathbb{1}_{[-1,1]\times[1,2]}(\mathsf{y}) \\ & + \frac{2 + y_2}{4}\mathbb{1}_{[-1,1]\times[-2,-1]}(\mathsf{y}).\end{aligned}$$

Note that $\mu, \nu$ have bounded support and the deterministic discretization $\mu^{(n)}$ and $\nu^{(n)}$ of Section 3.1 may be computed explicitly. We obtain $\mu^{(n)}[\{(i/n, j/n)\}] = 1/4n^2$ for $i, j = -n, \dots, n-1$ and, for $i', j' = -2n, \dots, 2n-1$,

$$\begin{aligned}&\nu^{(n)}[\{(i'/n, j'/n)\}] \\ &\quad = \begin{cases} (4n + 2i' + 1)/8n^3 & \text{if } -2n \le i' \le -n-1, -n \le j' \le n-1, \\ (4n + 2j' + 1)/8n^3 & \text{if } -n \le i' \le n-1, -2n \le j' \le -n-1, \\ (4n - 2j' - 1)/8n^3 & \text{if } -n \le i' \le n-1, n \le j' \le 2n-1, \\ (4n - 2i' - 1)/8n^3 & \text{if } n \le i' \le 2n-1, -n \le j' \le n-1, \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

With $\varepsilon_n := 4/n \ge \mathcal{W}^{\oplus}((\mu^{(n)}, \nu^{(n)}), (\mu, \nu))$, we obtain the LP problem $\mathsf{P}_{\varepsilon_n}(\mu^{(n)}, \nu^{(n)})$. For comparison, we also consider an approximation based on the random discretization. As in Example 3.8, we denote $\hat{\mu}^n$, $\hat{\nu}^n$ the corresponding empirical measures and $\mathsf{P}_{\varepsilon_m}(\hat{\mu}^n, \hat{\nu}^n)$ the LP problem. As $\nu$ has a bounded support, with $C(3, 2)$ defined in Proposition 3.7, one has

$$\mathbb{E}\big[\mathcal{W}^{\oplus}\big((\hat{\mu}^n, \hat{\nu}^n), (\mu, \nu)\big)\big] \le 2C(3, 2)(1 + \log n)n^{-1/2} =: \chi(n).$$

Setting $\hat{n}_m := \lfloor m^r \rfloor$ with $r = 4.1$, one has $\sum_{m \ge 1} \chi(\hat{n}_m)/\varepsilon_m < \infty$ and $\lim_{m\to\infty} \mathsf{P}_{\varepsilon_m}(\hat{\mu}^{\hat{n}_m}, \hat{\nu}^{\hat{n}_m}) = \mathsf{P}(\mu, \nu)$ almost surely. Solving $\mathsf{P}_{\varepsilon_n}(\mu^{(n)}, \nu^{(n)})$ and $\mathsf{P}_{\varepsilon_m}(\hat{\mu}^{\hat{n}_m}, \hat{\nu}^{\hat{n}_m})$, the values are plotted in Figure 5, where the convergence is illustrated. Note that the complexity of $\mathsf{P}_{\varepsilon_m}(\hat{\mu}^{\hat{n}_m}, \hat{\nu}^{\hat{n}_m})$ is of order $m^{2r}$, which is the same as in Example 3.8, however, the complexity of $\mathsf{P}_{\varepsilon_n}(\mu^{(n)}, \nu^{(n)})$ is of order $n^4$ which is the square of that in the one-dimensional case. In Figure 6, we draw the heat map of the optimisers projected on $(S_1^{(1)}, S_1^{(2)}, S_2^{(1)})$, where we recall that $\mathsf{S}_1 = (S_1^{(1)}, S_1^{(2)})$ and $\mathsf{S}_2 = (S_2^{(1)}, S_2^{(2)})$. As $\mu$ and $\nu$ are invariant by the

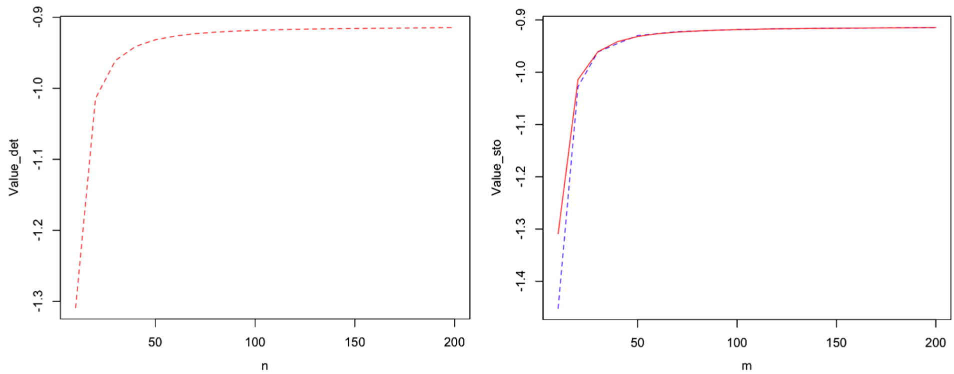


FIG. 5. *Computations for Example 3.10. The left pane shows the values of* $\mathsf{P}_{\varepsilon_n}(\mu^{(n)}, \nu^{(n)})$ *for* $10 \le n \le 200$ *and the right pane shows the values* $\mathsf{P}_{\varepsilon_m}(\hat{\mu}^{\hat{n}_m}, \hat{\nu}^{\hat{n}_m})$ *(dashed line) for* $10 \le m \le 200$.

map $\mathbb{R}^2 \ni (x, y) \mapsto (y, x) \in \mathbb{R}^2$, $(S_1^{(1)}, S_1^{(2)}, S_2^{(1)})$ and $(S_1^{(2)}, S_1^{(1)}, S_2^{(2)})$ are indistinguishable in law under the optimiser. The areas highlighted in red correspond to the values of $\mathsf{S}_1$ which are transferred into more than three points. These clearly appear to have positive mass in disagreement with Conjecture 2 in [22].

EXAMPLE 3.11. To show the universality of our method, we consider in the last example an MOT problem in $\mathbb{R}^3$, that is, $N = 2$ and $d = 3$. Let $c(\mathsf{x}, \mathsf{y}) := (\sum_{i=1}^3 \lambda_i |x_i - y_i| - K)^+$ for all $\mathsf{x} = (x_1, x_2, x_3)$, $\mathsf{y} = (y_1, y_2, y_3) \in \mathbb{R}^3$, where $K > 0$, $\lambda_i \ge 0$ and $\sum_{i=1}^3 \lambda_i = 1$. Here, $c$ represents the payoff of a basket option written on three forward start options with strike $K$. We construct $(\mu, \nu) \in \mathcal{P}^{\preceq}$ in the following way. Let $\rho : \mathbb{R}^3 \to [0, \infty)$ be an $L$-Lipschitz density function with a finite $\theta$-moment for some $\theta > 1$ and denote $\mu(d\mathsf{x}) = \rho(\mathsf{x})\, d\mathsf{x}$. We define next $\nu$ as

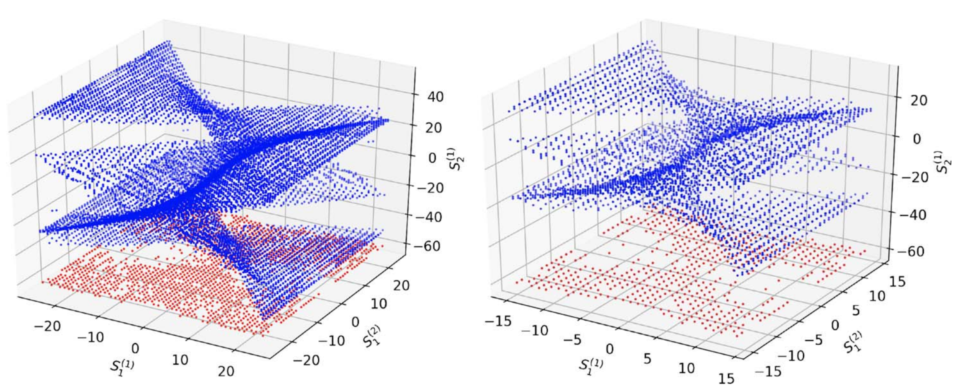


FIG. 6. *Computations for Example 3.10. The left pane shows the heat map of the optimiser projected on* $(S_1^{(1)}, S_1^{(2)}, S_2^{(1)})$ *for* $n = 100$ *and the right pane shows the heat map of the optimisers projected on* $(S_1^{(1)}, S_1^{(2)}, S_2^{(1)})$ *for* $m = 100$.

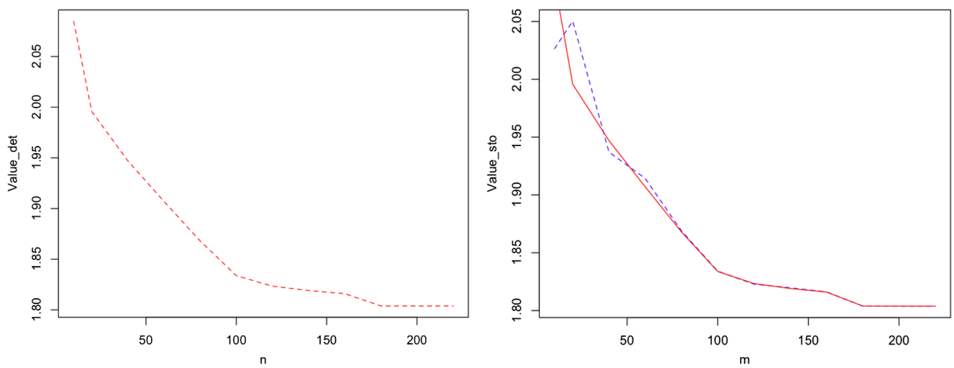


FIG. 7. *Computations for Example 3.11. The left pane shows the values* $\mathsf{P}_{\varepsilon_n}(\mu^n, \nu^n)$ *for* $10 \le n \le 220$ *and the right pane shows the values* $\mathsf{P}_{\varepsilon_m}(\hat{\mu}^{\hat{n}_m}, \hat{\nu}^{\hat{n}_m})$ *(dashed line) for* $10 \le m \le 220$.

the convolution of $\mu$ with a standard normal distribution, that is, $\nu(d\mathsf{y}) = \sigma(\mathsf{y})\,d\mathsf{y}$, where

$$\sigma(\mathsf{y}) := \int_{\mathbb{R}^3} \rho(\mathsf{y}-\mathsf{x}) \frac{1}{(2\pi)^{3/2}} \exp\left(-\frac{x_1^2 + x_2^2 + x_3^2}{2}\right) d\mathsf{x}.$$

Then it turns that $\sigma$ is $L$-Lipschitz and $\nu$ admits finite $\theta$-moment. We are now under the same conditions as Example 3.9. Taking $\lambda_1 = 1/2$, $\lambda_1 = 1/3$, $\lambda_3 = 1/6$, $K = 1$ and

$$\rho(\mathsf{x}) := \mathbb{1}_{[-1,1]^3}(\mathsf{x}) \frac{|x_1| + |x_2 x_3|}{C(1 + x_1^2 + 2x_2^2 + 3x_3^2)}$$
$$\text{where } C := \int_{[-1,1]^3} \frac{|x_1| + |x_2 x_3|}{1 + x_1^2 + 2x_2^2 + 3x_3^2} d\mathsf{x},$$

one has $L = 7/C$ and further

$$\int_{\mathbb{R}^3} |\mathsf{y}|^2 \sigma(\mathsf{y})\,d\mathsf{y} \le \frac{3}{C}\left(\frac{9}{2} + \frac{8}{\sqrt{2\pi}}\right) := M_2 \quad \text{and} \quad \chi(n) := 2C(2,3)n^{-1/3},$$

where $C(2,3)$ is given in Proposition 3.7. We carry out the same discretization procedure, deterministic as in Example 3.9 and random as in Examples 3.8 and 3.10, and solve the corresponding $\mathsf{LP}$ problems. The resulting value functions are displayed in Figure 7.

**4. Proofs.** Section 4 is devoted to the proofs of Theorems 2.2 and 2.5. Similar to the usual $\mathsf{MOT}$, the relaxed problem $\mathsf{P}_\varepsilon(\boldsymbol{\mu})$ admits a dual formulation given by

$$\mathsf{D}_\varepsilon(\boldsymbol{\mu}) := \inf_{(H,\psi)\in\mathcal{D}_\varepsilon} \left[\sum_{k=1}^N \int \psi_k \, d\mu_k\right], \tag{17}$$

where $\mathcal{H}$ is the set of $\mathbb{F}$-adapted processes $H = (H_k)_{1\le k\le N-1}$ taking values in $\mathbb{R}^d$, that is, $H_k \in \mathbb{L}^\infty(\Omega^k; \mathbb{R}^d)$ for $k = 1, \dots, N-1$ and $\mathcal{D}_\varepsilon \subset \mathcal{H} \times \Lambda^N$ denotes the collection of pairs $(H = (H_k)_{1\le k\le N-1}, \psi = (\psi_k)_{1\le k\le N})$ such that for $(\mathsf{x}_1, \dots, \mathsf{x}_N) \in \Omega^N$

$$\sum_{k=1}^{N-1} H_k(\mathsf{x}_1,\dots,\mathsf{x}_k)\cdot(\mathsf{x}_{k+1}-\mathsf{x}_k) - \varepsilon\sum_{k=1}^{N-1}\|H_k\|_\infty + \sum_{k=1}^{N}\psi_k(\mathsf{x}_k) \ge c(\mathsf{x}_1,\dots,\mathsf{x}_N). \tag{18}$$

Recall that $\mathcal{M}_\varepsilon(\boldsymbol{\mu}) \subset \mathcal{P}(\boldsymbol{\mu})$ is convex and compact. An application of the min-max theorem allows to establish the Kantorovich duality between (7) and (17) in Theorem 4.1 below. The proof largely repeats the reasoning in [3] where the result was shown for $\varepsilon = 0$, but it is nonetheless included in the Appendix.

THEOREM 4.1. *Let $\boldsymbol{\mu} \in \mathcal{P}_\varepsilon^{\preceq}$. If $c$ is upper semicontinuous with linear growth, then there exists an optimiser $\mathbb{P}^*$ for $\mathsf{P}_\varepsilon(\boldsymbol{\mu})$, that is, $\mathbb{P}^* \in \mathcal{M}_\varepsilon(\boldsymbol{\mu})$ and $\mathsf{P}_\varepsilon(\boldsymbol{\mu}) = \mathbb{E}_{\mathbb{P}^*}[c]$. Moreover, there is no duality gap, that is, $\mathsf{P}_\varepsilon(\boldsymbol{\mu}) = \mathsf{D}_\varepsilon(\boldsymbol{\mu})$.*

For $\varepsilon = 0$, the left-hand side of (18) stands for a super-replication of the payoff $c$ by trading dynamically in the underlying assets and statically in a range of Vanilla options. More precisely, $H_k(\mathsf{S}_1, \dots, \mathsf{S}_k)$ denotes the number of shares held by the trader at time $k$. Vanilla options allow the holder to receive the cash flow equal to $\psi_k(\mathsf{S}_k)$ at time $k$ for $k = 1, \dots, N$, and their market price is given as the integral of $\psi_k$ with respect to $\mu_k$, where $\mu_1, \dots, \mu_N$ represent the market-implied distributions of $\mathsf{S}_1, \dots, \mathsf{S}_N$. When $d = 1$, as observed by Breeden and Litzenberger [9], $\mu_k$ are uniquely determined from the observed prices of call/put options for all possible strikes. In consequence, the expression in brackets on the right-hand side of (17) represents the cost of pursuing a super-hedging strategy $(H, \psi)$ and $\mathsf{D}(\boldsymbol{\mu})$ is equal to the minimal super-hedging price of $c$.

4.1. *Convergence of relaxed* MOT *problems.* Theorem 2.2 shows that $\mathsf{P}(\boldsymbol{\mu})$ can be approximated by considering a sequence of relaxed MOT problems, which provides the main insight into our proposed scheme for solving MOT problems. The proof of Theorem 2.2 is divided into the proofs of Corollary 4.3, Proposition 4.4 and Lemma 4.5.

PROPOSITION 4.2. *Let $\boldsymbol{\mu} \in \mathcal{P}_\varepsilon^{\preceq}$. Then for any $\boldsymbol{\nu} \in \mathcal{P}^N$, one has $\boldsymbol{\nu} \in \mathcal{P}_{\varepsilon+r}^{\preceq}$ with $r := \mathcal{W}^{\oplus}(\boldsymbol{\mu}, \boldsymbol{\nu})$. If we assume in addition that $c$ is Lipschitz with Lipschitz constant* $\mathrm{Lip}(c)$, *then* $\mathsf{P}_\varepsilon(\boldsymbol{\mu}) \le \mathsf{P}_{\varepsilon+r}(\boldsymbol{\nu}) + \mathrm{Lip}(c)r$.

PROOF. Set $r_k := \mathcal{W}(\mu_k, \nu_k)$ for $k = 1, \dots, N$ and one has by definition $r = \sum_{k=1}^N r_k$. Take an arbitrary $\mathbb{P} \in \mathcal{M}_\varepsilon(\boldsymbol{\mu})$. It follows from Theorem 1 of Skorokhod [36] that there exists an enlarged probability space $(E, \mathcal{E}, \mathcal{Q})$ which supports random variables $X_k$ and $Z_k$ taking values in $\mathbb{R}^d$ for $k = 1, \dots, N$ such that

$$\mathcal{Q} \circ (X_1, \dots, X_N)^{-1} = \mathbb{P},$$
$$Z_1, \dots, Z_N \text{ and } (X_1, \dots, X_N) \text{ are mutually independent},$$
$$\mathcal{Q} \circ Z_k^{-1} \text{ is a standard normal distribution on } \mathbb{R}^d, \text{ for } k = 1, \dots, N.$$

For $k = 1, \dots, N$, let $\mathbb{P}_k$ be the optimal transport plan realising the Wasserstein distance between $\mu_k$ and $\nu_k$, that is, $\mathbb{P}_k \in \mathcal{P}(\mu_k, \nu_k)$ and $r_k = \mathbb{E}_{\mathbb{P}_k}[|\mathsf{S}_1 - \mathsf{S}_2|]$. From Lemma A.1, there exist measurable functions $f_k : \Omega^2 \to \mathbb{R}^d$ such that $\mathcal{Q} \circ (X_k, Y_k)^{-1} = \mathbb{P}_k$ with $Y_k := f_k(X_k, Z_k)$, for $k = 1, \dots, N$, which yields in particular $\mathcal{Q} \circ Y_k^{-1} = \nu_k$. Furthermore, one has, for all $h \in \mathcal{C}_b(\Omega^k; \mathbb{R}^d)$,

$$\begin{aligned}
&\mathbb{E}_{\mathcal{Q}}\big[h(Y_1, \dots, Y_k) \cdot (Y_{k+1} - Y_k)\big] \\
&\quad = \mathbb{E}_{\mathcal{Q}}\big[h(Y_1, \dots, Y_k) \cdot (Y_{k+1} - X_{k+1})\big] + \mathbb{E}_{\mathcal{Q}}\big[h(Y_1, \dots, Y_k) \cdot (X_{k+1} - X_k)\big] \\
&\qquad + \mathbb{E}_{\mathcal{Q}}\big[h(Y_1, \dots, Y_k) \cdot (X_k - Y_k)\big] \\
&\quad \le (r_k + r_{k+1})\|h\|_\infty + \mathbb{E}_{\mathcal{Q}}\big[h\big(f_1(X_1, Z_1), \dots, f_k(X_k, Z_k)\big) \cdot (X_{k+1} - X_k)\big] \\
&\quad = (r_k + r_{k+1})\|h\|_\infty + \int_{\Omega^k} \mathbb{E}_{\mathbb{P}}\big[h\big(f_1(\mathsf{S}_1, \mathsf{x}_1), \dots, f_k(\mathsf{S}_k, \mathsf{x}_k)\big) \cdot (\mathsf{S}_{k+1} - \mathsf{S}_k)\big] \\
&\qquad \times \mathcal{N}_k(d\mathsf{x}_1, \dots, d\mathsf{x}_k) \\
&\quad \le (\varepsilon + r)\|h\|_\infty,
\end{aligned}$$

where $\mathcal{N}_k$ denotes the joint distribution of $Z_1, \dots, Z_k$. Therefore, $\mathbb{E}_{\mathbb{P}'}[h(\mathsf{S}_1, \dots, \mathsf{S}_k) \cdot (\mathsf{S}_{k+1} - \mathsf{S}_k)] \le (\varepsilon + r)\|h\|_\infty$ holds for all $h \in \mathcal{C}_b(\Omega^k; \mathbb{R}^d)$, where $\mathbb{P}' := \mathcal{Q} \circ (Y_1, \dots, Y_N)^{-1}$. In view of the monotone class theorem, this is equivalent to

$$\mathbb{E}_{\mathbb{P}'}\big[\big|\mathbb{E}_{\mathbb{P}'}[\mathsf{S}_{k+1}|\mathcal{F}_k] - \mathsf{S}_k\big|\big] \le \varepsilon + r.$$

Hence, $\mathbb{P}' \in \mathcal{M}_{\varepsilon+r}(\boldsymbol{\nu})$ and $\boldsymbol{\nu} \in \mathcal{P}^{\preceq}_{\varepsilon+r}$. To conclude the proof, notice that

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}}\big[c(\mathsf{S}_1, \dots, \mathsf{S}_N)\big] - \mathsf{P}_{\varepsilon+r}(\boldsymbol{\nu}) \\
&\quad \le \mathbb{E}_{\mathbb{P}}\big[c(\mathsf{S}_1, \dots, \mathsf{S}_N)\big] - \mathbb{E}_{\mathbb{P}'}\big[c(\mathsf{S}_1, \dots, \mathsf{S}_N)\big] \\
&\quad = \mathbb{E}_{\mathcal{Q}}\big[c(X_1, \dots, X_N) - c(Y_1, \dots, Y_N)\big] \le \mathrm{Lip}(c) \sum_{k=1}^N \mathbb{E}_{\mathcal{Q}}\big[|X_k - Y_k|\big] \\
&\quad = \mathrm{Lip}(c) r,
\end{aligned}$$

which yields $\mathsf{P}_\varepsilon(\boldsymbol{\mu}) \le \mathsf{P}_{\varepsilon+r}(\boldsymbol{\nu}) + \mathrm{Lip}(c) r$ since $\mathbb{P} \in \mathcal{M}_\varepsilon(\boldsymbol{\mu})$ is arbitrary. $\square$

In consequence, the corollary below follows immediately.

COROLLARY 4.3. *Let $(\boldsymbol{\mu}^n)_{n\geq 1}$ and $(\varepsilon_n)_{n\geq 1}$ be the sequences in Theorem 2.2. Then*

$$\mathsf{P}(\boldsymbol{\mu}) \leq \mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n) + \mathrm{Lip}(c)\varepsilon_n \leq \mathsf{P}_{2\varepsilon_n}(\boldsymbol{\mu}) + 2\mathrm{Lip}(c)\varepsilon_n \quad \textit{for all } n \geq 1.$$

PROOF. Taking $\varepsilon = 0$, $\boldsymbol{\nu} = \boldsymbol{\mu}^n$ and $r = r_n$, one has $\mathsf{P}(\boldsymbol{\mu}) \leq \mathsf{P}_{r_n}(\boldsymbol{\mu}^n) + \mathrm{Lip}(c)r_n \leq \mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n) + \mathrm{Lip}(c)\varepsilon_n$, where $\mathsf{P}_{r_n}(\boldsymbol{\mu}^n) \leq \mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n)$ follows by definition. The second inequality follows with the same arguments but interchanging $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^n$. $\square$

To complete the proof of Theorem 2.2, it remains to show $\mathsf{P}_{2\varepsilon_n}(\boldsymbol{\mu}) \to \mathsf{P}(\boldsymbol{\mu})$ as $n \to \infty$.

PROPOSITION 4.4. *Let $c$ be Lipschitz.*

(i) *For every fixed $\varepsilon \geq 0$, the map $\mathcal{P}^{\preceq}_{\varepsilon} \ni \boldsymbol{\mu} \mapsto \mathsf{P}_{\varepsilon}(\boldsymbol{\mu}) \in \mathbb{R}$ is upper semicontinuous under $\mathcal{W}^{\oplus}$.*

(ii) *For every fixed $\boldsymbol{\mu} \in \mathcal{P}^{\preceq}$, the map $[0, \infty) \ni \varepsilon \mapsto \mathsf{P}_{\varepsilon}(\boldsymbol{\mu}) \in \mathbb{R}$ is nondecreasing, continuous and concave.*

Before proving Proposition 4.4, let us remark that, together with Corollary 4.3 and Lemma 4.5, it yields an instant proof of our main result.

PROOF OF THEOREM 2.2. (i) We have $\mathcal{M}_{\varepsilon_n}(\boldsymbol{\mu}^n) \neq \varnothing$ from Proposition 4.2. Corollary 4.3 yields $-\mathrm{Lip}(c)\varepsilon_n \leq \mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n) - \mathsf{P}(\boldsymbol{\mu}) \leq (\mathsf{P}_{2\varepsilon_n}(\boldsymbol{\mu}) - \mathsf{P}(\boldsymbol{\mu})) + \mathrm{Lip}(c)\varepsilon_n$ for all $n \geq 1$, and Proposition 4.4 gives $\lim_{n\to\infty} \mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n) = \mathsf{P}(\boldsymbol{\mu})$.

(ii) By Theorem 4.1, we know the existence of optimiser $\mathbb{P}_n$ for all $n \geq 1$. As $\mathbb{P}_n \in \mathcal{M}_{\varepsilon_n}(\boldsymbol{\mu}^n) \subset \mathcal{P}(\boldsymbol{\mu}^n)$, it follows from Lemma 4.5 that, $(\mathbb{P}_n)_{n\geq 1}$ is tight and every limit point must belong to $\mathcal{P}(\boldsymbol{\mu})$. Take an arbitrary convergent subsequence $(\mathbb{P}_{n_k})_{k\geq 1}$ with limit $\mathbb{P}$, then $\mathbb{P} \in \mathcal{P}(\boldsymbol{\mu})$. Repeating the proof of Proposition 4.4(i), one deduces that $\mathbb{P} \in \mathcal{M}(\boldsymbol{\mu})$ and $\mathbb{P}$ is thus an optimiser for $\mathsf{P}(\boldsymbol{\mu})$. If $\mathsf{P}(\boldsymbol{\mu})$ has a unique optimiser, then every convergent subsequence of $(\mathbb{P}_n)_{n\geq 1}$ has the same limit, which shows that $(\mathbb{P}_n)_{n\geq 1}$ converges weakly as $\mathcal{P}(\Omega^N)$ is Polish. $\square$

PROOF OF PROPOSITION 4.4. (i) We establish a slightly stronger property. Take two sequences $(\varepsilon_n)_{n\geq 1} \subset [0, \infty)$ and $(\boldsymbol{\mu}^n)_{n\geq 1} \subset \mathcal{P}^{\preceq}_{\varepsilon_n}$ with limits $\varepsilon$ and $\boldsymbol{\mu}$. Let $\mathbb{P}_n \in \mathcal{M}_{\varepsilon_n}(\boldsymbol{\mu}^n)$ satisfy $\limsup_{n\to\infty} \mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n) = \limsup_{n\to\infty} \mathbb{E}_{\mathbb{P}_n}[c(\mathsf{S}_1, \dots, \mathsf{S}_N)]$. Up to passing to a subsequence, we may assume that $\limsup_{n\to\infty} \mathbb{E}_{\mathbb{P}_n}[c] = \lim_{n\to\infty} \mathbb{E}_{\mathbb{P}_n}[c]$, and further by Lemma 4.5 that $(\mathbb{P}_n)_{n\geq 1}$ converges in Wasserstein sense to some limit $\mathbb{P} \in \mathcal{P}(\boldsymbol{\mu})$. For every $k = 1, \dots, N-1$ and $h \in \mathcal{C}_b(\Omega^k; \mathbb{R}^d)$, one has $\mathbb{E}_{\mathbb{P}_n}[h(\mathsf{S}_1, \dots, \mathsf{S}_k) \cdot (\mathsf{S}_{k+1} - \mathsf{S}_k)] \leq \varepsilon_n \|h\|_\infty$ for all $n \geq 1$, and hence for the limiting measures, which implies $\mathbb{P} \in \mathcal{M}_{\varepsilon}(\boldsymbol{\mu})$. Similarly, Lipschitz continuity of $c$ gives

$$\limsup_{n\to\infty} \mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n) = \lim_{n\to\infty} \mathbb{E}_{\mathbb{P}_n}\big[c(\mathsf{S}_1, \dots, \mathsf{S}_N)\big] = \mathbb{E}_{\mathbb{P}}\big[c(\mathsf{S}_1, \dots, \mathsf{S}_N)\big] \leq \mathsf{P}_{\varepsilon}(\boldsymbol{\mu}).$$

(ii) We first prove the concavity. Given $\varepsilon$, $\varepsilon' \geq 0$ and $\alpha \in [0,1]$, it remains to show $(1-\alpha)\mathsf{P}_\varepsilon(\boldsymbol{\mu}) + \alpha\mathsf{P}_{\varepsilon'}(\boldsymbol{\mu}) \leq \mathsf{P}_{\varepsilon_\alpha}(\boldsymbol{\mu})$, where $\varepsilon_\alpha := (1-\alpha)\varepsilon + \alpha\varepsilon'$. This indeed follows from the fact that $(1-\alpha)\mathbb{P} + \alpha\mathbb{P}' \in \mathcal{M}_{\varepsilon_\alpha}(\boldsymbol{\mu})$ for all $\mathbb{P} \in \mathcal{M}_\varepsilon(\boldsymbol{\mu})$ and $\mathbb{P}' \in \mathcal{M}_{\varepsilon'}(\boldsymbol{\mu})$. In particular, the map restricted to $(0,\infty)$ is continuous. Finally, the reasoning in (i) above, with $\boldsymbol{\mu}^n = \boldsymbol{\mu}$ and $\varepsilon_n \to 0$, gives $\lim_{n\to\infty} \mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}) = \mathsf{P}(\boldsymbol{\mu})$ which combined with the obvious reverse inequality yields the right continuity at $\varepsilon = 0$. □

LEMMA 4.5. *Let $(\boldsymbol{\mu}^n)_{n\geq 1} \subset \mathcal{P}^N$ be a sequence converging to $\boldsymbol{\mu} \in \mathcal{P}^N$ under $\mathcal{W}^\oplus$, and $\mathbb{P}_n \in \mathcal{P}(\boldsymbol{\mu}^n)$ for all $n \geq 1$. Then there exists a subsequence $(\mathbb{P}_{n_k})_{k\geq 1}$ converging in Wasserstein metric on $\mathcal{P}(\Omega^N)$ and its limit $\mathbb{P}$ belongs to $\mathcal{P}(\boldsymbol{\mu})$.*

PROOF. Taking the compact $E_R := \{(\mathsf{x}_1, \ldots, \mathsf{x}_N) \in \Omega^N : |\mathsf{x}_k| \leq R,\ \text{for } k = 1, \ldots, N\}$. One has

$$\mathbb{P}_n\big[E_R^c\big] \leq \sum_{k=1}^N \int_{\mathbb{R}^d} \mathbb{1}_{[R,\infty)}\big(|\mathsf{x}|\big)\mu_k^n(d\mathsf{x}) \leq \frac{1}{R}\sup_{n\geq 1}\left\{\sum_{k=1}^N \int_{\mathbb{R}^d} |\mathsf{x}|\mu_k^n(d\mathsf{x})\right\}.$$

Further, the convergence under $\mathcal{W}^\oplus$ implies that

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} |\mathsf{x}|\mu_k^n(d\mathsf{x}) = \int_{\mathbb{R}^d} |\mathsf{x}|\mu_k(d\mathsf{x}) \quad \text{for } k = 1, \ldots, N,$$

which yields the tightness of $(\mathbb{P}_n)_{n\geq 1}$ since $\lim_{R\to\infty}\{\sup_{n\geq 1} \mathbb{P}_n[E_R^c]\} = 0$. This implies there exists a weakly convergent subsequence, which we still denote $(\mathbb{P}_n)_{n\geq 1}$ and let $\mathbb{P}$ be its limit. Notice that the projection map $\mathsf{S}_k$ is continuous, then $\mu_k^n = \mathbb{P}_n \circ \mathsf{S}_k^{-1}$ also converges weakly to $\mathbb{P} \circ \mathsf{S}_k^{-1}$ for $k = 1, \ldots, N$, which shows $\mathbb{P} \in \mathcal{P}(\boldsymbol{\mu})$. Finally, the convergence of $\mathbb{P}_n$ to $\mathbb{P}$ in the Wasserstein sense follows from the convergence of first moment. □

4.2. *Convergence rate analysis*: $N = 2$ *and* $d = 1$. This section concerns the estimation of the convergence rate for the one-step model in dimension one. The duality in Theorem 4.1 plays an important role and is used repeatedly. To the best of our knowledge, the error bound in Theorem 2.5 is the first of its kind in the literature. Fix a pair $(\mu, \nu) \in \mathcal{P}^{\preceq}$. Throughout Section 4.2, we stress the dependence on $c$ and write $\mathsf{P}_\varepsilon^c(\mu,\nu) \equiv \mathsf{P}_\varepsilon(\mu,\nu)$, $\mathcal{D}_\varepsilon^c \equiv \mathcal{D}_\varepsilon$ and $\mathsf{D}_\varepsilon^c(\mu,\nu) \equiv \mathsf{D}_\varepsilon(\mu,\nu)$. Clearly, for any $c_1$ and $c_2 : \mathbb{R}^2 \to \mathbb{R}$, it holds that $\mathsf{P}_\varepsilon^{c_1+c_2}(\mu,\nu) \leq \mathsf{P}_\varepsilon^{c_1}(\mu,\nu) + \mathsf{P}_\varepsilon^{c_2}(\mu,\nu)$. In view of Corollary 4.3, one has

$$\big|\mathsf{P}_{\varepsilon_n}^c\big(\mu^n,\nu^n\big) - \mathsf{P}^c(\mu,\nu)\big| \leq \big(\mathsf{P}_{2\varepsilon_n}^c(\mu,\nu) - \mathsf{P}^c(\mu,\nu)\big) + \mathrm{Lip}(c)\varepsilon_n,$$

where $\varepsilon_n \geq \mathcal{W}^\oplus((\mu^n,\nu^n),(\mu,\nu))$ for all $n \geq 1$. For the purpose of estimating the difference $|\mathsf{P}_{\varepsilon_n}^c(\mu^n,\nu^n) - \mathsf{P}^c(\mu,\nu)|$, we need to understand the asymptotic behavior of $\mathsf{P}_\varepsilon^c(\mu,\nu) - \mathsf{P}^c(\mu,\nu)$ as $\varepsilon$ goes to 0, which is shown in the proof of Theorem 2.5.

PROOF OF THEOREM 2.5. Set $L := \max(\mathrm{Lip}(c), \sup_{(x,y)\in\mathbb{R}^2} |\partial^2_{yy} c(x,y)|) < \infty$ and introduce $c_L(x,y) := c(x,y) - Ly^2/2$. Then, for each $x \in \mathbb{R}$, the map $y \mapsto c_L(x,y)$ is concave. Further, let us truncate $c_L$ by an affine function with respect to $y$. Namely, define, for every $R \geq 0$,

$$c_L^R(x,y) := \begin{cases} c_L(x,-R) + (y+R)\partial_y c_L(x,-R) & \text{if } y \leq -R, \\ c_L(x,y) & \text{if } -R < y \leq R, \\ c_L(x,R) + (y-R)\partial_y c_L(x,R) & \text{otherwise.} \end{cases}$$

It follows by a straightforward computation that, $y \mapsto c_L^R(x,y)$ is concave and $\mathrm{Lip}(c_L^R) \leq L_R := L(R+1)$. In view of Remark 2.6 in Beiglböck et al. [5], there exists an optimiser $(H^*, \varphi^*, \psi^*) \in \mathcal{D}^{c_L^R}$ for the dual problem $\mathsf{D}^{c_L^R}(\mu,\nu)$ such that $\|H^*\|_\infty \leq 18L_R$ and $\varphi^*, \psi^* \in \Lambda_{19L_R}$, that is, $H^*(x)(y-x) + \varphi^*(x) + \psi^*(y) \geq c_L^R(x,y)$ for all $(x,y) \in \mathbb{R}^2$ and

$$\mathsf{D}^{c_L^R}(\mu,\nu) = \int \varphi^*\, d\mu + \int \psi^*\, d\nu.$$

For each $\varepsilon \geq 0$, it follows from the duality $\mathsf{P}_\varepsilon^{c_L^R}(\mu,\nu) = \mathsf{D}_\varepsilon^{c_L^R}(\mu,\nu)$ that

$$\begin{aligned} \left|\mathsf{P}_\varepsilon^{c_L^R}(\mu,\nu) - \mathsf{P}^{c_L^R}(\mu,\nu)\right| &= \left|\mathsf{D}_\varepsilon^{c_L^R}(\mu,\nu) - \mathsf{D}^{c_L^R}(\mu,\nu)\right| \\ &= \left|\mathsf{D}_\varepsilon^{c_L^R}(\mu,\nu) - \left[\int \varphi^*\, d\mu + \int \psi^*\, d\nu\right]\right| \\ &\leq \left|\left[\int (\varphi^* + \varepsilon\|H^*\|_\infty)\, d\mu + \int \psi^*\, d\nu\right]\right. \\ &\qquad \left. - \left[\int \varphi^*\, d\mu + \int \psi^*\, d\nu\right]\right| \\ &= \varepsilon \int \|H^*\|_\infty\, d\mu \leq 18\varepsilon L_R, \end{aligned}$$

where the third inequality holds as $(H^*, \varphi^* + \varepsilon\|H^*\|_\infty, \psi^*) \in \mathcal{D}_\varepsilon^{c_L^R}$. In addition, one has by construction $|c_L(x,y) - c_L^R(x,y)| \leq \mathbb{1}_{(-\infty,-R)\cup(R,\infty)}(y) L(|y|-R)^2$, which implies that

$$\left|\mathsf{P}_\varepsilon^{c_L^R}(\mu,\nu) - \mathsf{P}_\varepsilon^{c_L}(\mu,\nu)\right| \leq L \int_{(-\infty,-R)\cup(R,\infty)} \big(|y|-R\big)^2 \nu(dy).$$

Therefore,

$$\begin{aligned} &\left|\mathsf{P}_\varepsilon^{c}(\mu,\nu) - \mathsf{P}^{c}(\mu,\nu)\right| \\ &\quad = \left|\mathsf{P}_\varepsilon^{c_L}(\mu,\nu) - \mathsf{P}^{c_L}(\mu,\nu)\right| \\ &\quad \leq \left|\mathsf{P}_\varepsilon^{c_L}(\mu,\nu) - \mathsf{P}_\varepsilon^{c_L^R}(\mu,\nu)\right| + \left|\mathsf{P}_\varepsilon^{c_L^R}(\mu,\nu) - \mathsf{P}^{c_L^R}(\mu,\nu)\right| \end{aligned}$$

$$
\begin{aligned}
&+\left|\mathsf{P}^{c_L^R}(\mu,\nu)-\mathsf{P}^{c_L}(\mu,\nu)\right| \\
&\le 18\varepsilon L(R+1)+2L\int_{(-\infty,-R)\cup(R,\infty)}\big(|y|-R\big)^2\nu(dy),
\end{aligned}
$$

which fulfills the proof by setting $\varepsilon = 2\varepsilon_n$. $\square$

REMARK 4.6. If $\nu$ is supported on some closed subset $E \subseteq \mathbb{R}$, then Theorem 2.5 still holds by assuming that $c$ is Lipschitz on $E^2$ and $\sup_{(x,y)\in E^2}|\partial^2_{yy}c(x,y)| < \infty$. In addition, it is worth mentioning that the above analysis can be extended to more general functions $c$. Let $c$ be continuous and with linear growth, that is, $|c(x,y)| \le L(1+|x|+|y|)$ for some $L > 0$. Then for every $R \ge 1$, there exists a function $c_R \in \mathcal{C}^2(\mathbb{R}^2)$ such that $\sup_{(x,y)\in\mathsf{B}_R}|c(x,y)-c_R(x,y)| \le 1/R$, $c_R(x,y) = 0$ for $(x,y)\notin \mathsf{B}_{R+1}$ and $\|c_R\|_\infty \le \sup_{(x,y)\in\mathsf{B}_R}|c(x,y)| \le L(1+2R)$. Further, one has $|c(x,y)-c_R(x,y)| \le 1/R + 8L(|x|^2+|y|^2)/R$, which implies that

$$
\begin{aligned}
\left|\mathsf{P}^c_\varepsilon(\mu,\nu)-\mathsf{P}^{c_R}_\varepsilon(\mu,\nu)\right| &\le 1/R + 8L\biggl(\int_{\mathbb{R}}|x|^2\mu(dx)+\int_{\mathbb{R}}|y|^2\nu(dy)\biggr)\Big/R \\
&=: L'/R.
\end{aligned}
$$

Hence, we obtain using the same reasoning $|\mathsf{P}^c_\varepsilon(\mu,\nu)-\mathsf{P}^c(\mu,\nu)| \le |\mathsf{P}^{c_R}_\varepsilon(\mu,\nu) - \mathsf{P}^{c_R}(\mu,\nu)| + 2L'/R$. Then $c_R$ satisfies the conditions of Theorem 2.5. For every $R \ge 1$, using Theorem 2.5 we deduce a bound on the difference $|\mathsf{P}^{c_R}_\varepsilon(\mu,\nu) - \mathsf{P}^{c_R}(\mu,\nu)|$. The result can then be optimised over all $R \ge 1$.

Following the proof of Theorem 2.5, we provide below a stability result for the map $\mathcal{P}^{\preceq} \ni (\mu,\nu) \mapsto \mathsf{P}(\mu,\nu) \in \mathbb{R}$.

PROPOSITION 4.7. *Let $\mathcal{P}^{\preceq}_2 \subset \mathcal{P}^{\preceq}$ be the subset of $(\mu,\nu)$ with $\nu$ having a finite second moment. If $c$ satisfies the conditions of Theorem 2.5, then there exists $C > 0$ such that*

$$
\begin{aligned}
&\left|\mathsf{P}^c(\mu',\nu')-\mathsf{P}^c(\mu,\nu)\right| \le C\inf_{R>0}\tilde{\lambda}(R)+\frac{L}{2}\left|\int_{\mathbb{R}}y^2\big(\nu'(dy)-\nu(dy)\big)\right| \\
&\qquad \text{for all } (\mu,\nu),\big(\mu',\nu'\big)\in\mathcal{P}^{\preceq}_2,
\end{aligned}
$$

*where $\tilde{\lambda} : (0,\infty) \to \mathbb{R}$ is defined by*

$$
\begin{aligned}
\tilde{\lambda}(R) := {}&(R+1)\mathcal{W}^{\oplus}\big((\mu',\nu'),(\mu,\nu)\big) \\
&+\int_{(-\infty,-R)\cup(R,\infty)}\big(|y|-R\big)^2\big(\nu'(dy)+\nu(dy)\big).
\end{aligned}
$$

*For any sequence $((\mu^n,\nu^n))_{n\ge1} \subset \mathcal{P}^{\preceq}_2$ satisfying $\lim_{n\to\infty}\mathcal{W}^{\oplus}((\mu^n,\nu^n),(\mu,\nu)) = 0$, one has $\lim_{n\to\infty}\mathsf{P}^c(\mu^n,\nu^n) = \mathsf{P}^c(\mu,\nu)$ if $\lim_{n\to\infty}\int y^2\nu^n(dy) = \int y^2\nu(dy)$.*

PROOF. Similar to Theorem 2.5, the key is also the duality. First, one has

$$\big|\mathsf{P}^{c}(\mu',\nu')-\mathsf{P}^{c}(\mu,\nu)\big|\le\big|\mathsf{P}^{c_L}(\mu',\nu')-\mathsf{P}^{c_L}(\mu,\nu)\big|+\frac{L}{2}\bigg|\int_{\mathbb{R}}y^2\big(\nu'(dy)-\nu(dy)\big)\bigg|$$

and $|\mathsf{P}^{c_L}(\mu',\nu')-\mathsf{P}^{c_L}(\mu,\nu)|\le|\mathsf{P}^{c_L}(\mu',\nu')-\mathsf{P}^{c_L^R}(\mu',\nu')|+|\mathsf{P}^{c_L^R}(\mu',\nu')-\mathsf{P}^{c_L^R}(\mu,\nu)|+|\mathsf{P}^{c_L^R}(\mu,\nu)-\mathsf{P}^{c_L}(\mu,\nu)|$, where $c_L, c_L^R:\mathbb{R}^2\to\mathbb{R}$ are defined as same as in the proof of Theorem 2.5. Repeating the arguments in the proof of Theorem 2.5, it holds that

$$\big|\mathsf{P}^{c_L}(\mu',\nu')-\mathsf{P}^{c_L^R}(\mu',\nu')\big|\le L\int_{(-\infty,-R)\cup(R,\infty)}\big(|y|-R\big)^2\nu'(dy),$$
$$\big|\mathsf{P}^{c_L}(\mu,\nu)-\mathsf{P}^{c_L^R}(\mu,\nu)\big|\le L\int_{(-\infty,-R)\cup(R,\infty)}\big(|y|-R\big)^2\nu(dy).$$

It remains to estimate $|\mathsf{P}^{c_L^R}(\mu',\nu')-\mathsf{P}^{c_L^R}(\mu,\nu)|$. Recall that, in view of Remark 2.6 of [5], $\mathsf{D}^{c_L^R}(\mu',\nu')$ is attained by $(H',\varphi',\psi')\in\mathcal{D}^{c_L^R}$, where $\varphi',\psi'\in\Lambda_{19L_R}$ with $L_R=L(R+1)$. Therefore,

$$\begin{aligned}
&\mathsf{P}^{c_L^R}(\mu,\nu)-\mathsf{P}^{c_L^R}(\mu',\nu')\\
&\quad=\mathsf{D}^{c_L^R}(\mu,\nu)-\mathsf{D}^{c_L^R}(\mu',\nu')\\
&\quad\le\bigg[\int\varphi'\,d\mu+\int\psi'\,d\nu\bigg]-\bigg[\int\varphi'\,d\mu'+\int\psi'\,d\nu'\bigg]\\
&\quad=\bigg[\int\varphi'\,d\mu-\int\varphi'\,d\mu'\bigg]+\bigg[\int\psi'\,d\nu-\int\psi'\,d\nu'\bigg]\\
&\quad\le 19L_R\mathcal{W}^{\oplus}\big((\mu',\nu'),(\mu,\nu)\big).
\end{aligned}$$

Interchanging $(\mu,\nu)$ and $(\mu',\nu')$ and using again the above reasoning, one has $|\mathsf{P}^{c_L^R}(\mu',\nu')-\mathsf{P}^{c_L^R}(\mu,\nu)|\le 19L_R\mathcal{W}^{\oplus}((\mu',\nu'),(\mu,\nu))$, which concludes the proof. □

We now consider a specific discretization introduced by Dolinsky and Soner [18]. We define two sequences of measures supported on $(k/n)_{k\in\mathbb{Z}}$ as follows:

$$\begin{aligned}
\mu^n\big[\{k/n\}\big]&=\int_{[(k-1)/n,(k+1)/n)}\big(1-|nx-k|\big)\mu(dx),\\
\nu^n\big[\{k/n\}\big]&=\int_{[(k-1)/n,(k+1)/n)}\big(1-|ny-k|\big)\nu(dy).
\end{aligned}\tag{19}$$

In the potential theoretic terms of Chacon [11], $\mu^n$ may be defined as the unique measure supported on $(k/n)_{k\in\mathbb{Z}}$ with its potential function agreeing with that of $\mu$

in those points, that is,

$$\int_{\mathbb{R}} |k/n - x|\mu(dx) = \int_{\mathbb{R}} |k/n - x|\mu^n(dx) \quad \text{for all } k \in \mathbb{Z}.$$

Then we have the following result.

PROPOSITION 4.8.

(i) *With the notation of* (19), *one has* $(\mu^n, \nu^n)$, $(\mu, \mu^n)$, $(\nu, \nu^n) \in \mathcal{P}^{\preceq}$ *and* $\mathcal{W}^{\oplus}((\mu^n, \nu^n), (\mu, \nu)) \le 2/n$ *for all* $n \ge 1$.

(ii) *Let the conditions of Theorem* 2.5 *hold. Then there exists* $C > 0$ *such that* $|\mathsf{P}^c(\mu^n, \nu^n) - \mathsf{P}^c(\mu, \nu)| \le C \inf_{R>0} \tilde{\lambda}_n(R)$, *where* $\tilde{\lambda}_n : (0, \infty) \to \mathbb{R}$ *is given by*

$$\tilde{\lambda}_n(R) := \frac{R+1}{n} + \int_{(-\infty,-R)\cup(R,\infty)} (|y| - R)^2 \nu(dy).$$

PROOF. (i) For any continuous $f : \mathbb{R} \to \mathbb{R}$, define $f^{(n)} : \mathbb{R} \to \mathbb{R}$ by

$$f^{(n)}(x) := (1 + \lfloor nx \rfloor - nx) f(\lfloor nx \rfloor/n) + (nx - \lfloor nx \rfloor) f((1 + \lfloor nx \rfloor)/n).$$

Then it follows from a straightforward computation that $\int f\, d\mu^n = \int f^{(n)}\, d\mu$ and $\int f\, d\nu^n = \int f^{(n)}\, d\nu$. Take $f \equiv 1$, then $f^{(n)} \equiv 1$, and further $\mu^n$ and $\nu^n$ are well-defined probability measures. Moreover, taking $f(x) = |x|$, it is clear that $f^{(n)} = f$, and thus $\int f\, d\mu^n = \int f^{(n)}\, d\mu < \infty$ and $\int f\, d\nu^n = \int f^{(n)}\, d\nu < \infty$. To prove $(\mu^n, \nu^n)$, $(\mu, \mu^n)$, $(\nu, \nu^n) \in \mathcal{P}^{\preceq}$, it suffices to test for $f(x) = (x - K)^+$. It follows easily that $f^{(n)}$ is convex and $f^{(n)} \ge f$ by computation. This implies that $(\mu^n, \nu^n)$, $(\mu, \mu^n)$, $(\nu, \nu^n) \in \mathcal{P}^{\preceq}$. To end the proof, we notice that $|\int f\, d\mu^n - \int f\, d\mu| \le \int |f^{(n)} - f|\, d\mu \le 1/n$, which yields $\mathcal{W}(\mu^n, \mu) \le 1/n$ by (4).

(ii) It suffices to apply Proposition 4.7 with $\mu' := \mu^n$ and $\nu' := \nu^n$. Using the construction of $\nu^n$, one has

$$\int_{(-\infty,-R)\cup(R,\infty)} (|y| - R)^2 \nu^n(dy) \le \int_{(-\infty,-R)\cup(R,\infty)} (|y| - R)^2 \nu(dy) + \frac{1}{n^2},$$

$$\left| \int_{\mathbb{R}} y^2 (\nu^n(dy) - \nu(dy)) \right| \le \frac{1}{4n^2},$$

which concludes the proof. □

**5. Summary and possible extensions.** We believe that our paper offers an important and pioneering contribution to computational methods for MOT problems. Our first main result, Theorem 2.2, establishes an approximation result of a general MOT problem $\mathsf{P}(\boldsymbol{\mu})$ via LP problems by discretising the marginal distributions and relaxing the martingale condition. Further, we introduce two kinds of approximations: a deterministic one, $\boldsymbol{\mu}^n$, and a stochastic one, $\hat{\boldsymbol{\mu}}^n$. We investigate $\mathcal{W}^{\oplus}(\boldsymbol{\mu}^n, \boldsymbol{\mu})$ and $\mathbb{E}[\mathcal{W}^{\oplus}(\hat{\boldsymbol{\mu}}^n, \boldsymbol{\mu})]$ such that the computation of suitable LP problems

$\mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n)$ and $\mathsf{P}_{\varepsilon_m}(\boldsymbol{\mu}^{\hat{n}_m})$ can be carried out. In addition, we provide some numerical examples for illustration.

Our second main result, Theorem 2.5, provides an estimation on the convergence rate for the one-dimensional case. This result, in particular, allows us to deduce a complete scheme for calculating $\mathsf{P}(\mu, \nu)$ to a given precision.

As a relatively immediate, but practically relevant, extension of our setup, for the computation of $\mathsf{P}(\vec{\boldsymbol{\mu}})$ defined in (9), Theorem 2.2 can be easily extended to show $\lim_{n\to\infty} \mathsf{P}_{\varepsilon_n}(\vec{\boldsymbol{\mu}}^n) = \mathsf{P}(\vec{\boldsymbol{\mu}})$, where the sequence $(\varepsilon_n)_{n\geq 1}$ converges to zero and satisfies

$$\varepsilon_n \geq \sum_{k=1}^{N}\sum_{i=1}^{d} \mathcal{W}(\mu_{k,i}^n, \mu_{k,i}) \quad \text{for all } n \geq 1.$$

Further investigation of this setup, which is of practical relevance, is an ongoing work.

Last, but not least, we point out that solving efficiently the LP problems $\mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n)$ and $\mathsf{P}_{\varepsilon_m}(\boldsymbol{\mu}^{\hat{n}_m})$ is also an interesting avenue of research and may attract the attention from practitioners. We notice that $\mathsf{P}_{\varepsilon_n}(\boldsymbol{\mu}^n)$ and $\mathsf{P}_{\varepsilon_m}(\boldsymbol{\mu}^{\hat{n}_m})$ are in fact LP problems with a particular structure, that is, the constraints are given by a sparse matrix, and some existing algorithms can be extended to their setup:

- If $N = 2$ and $d = 1$, the iterative Bregman projection in [7] can be applied to solve $\mathsf{P}_{\varepsilon_n}(\mu^n, \nu^n)$ with an additional entropic regularization.
- If $N = 2$, the stochastic averaged gradient approach (see, e.g., Genevay et al. [21]) may deal with $\mathsf{P}_{\varepsilon_n}(\mu^n, \nu^n) = \mathsf{D}_{\varepsilon_n}(\mu^n, \nu^n)$ by the duality.

We believe that extending the above algorithms to multiple steps and higher dimensions is an important and challenging problem.

## APPENDIX: SUPPLEMENTARY PROOFS

PROOF OF THEOREM 4.1. The existence of $\mathbb{P}^*$ is a consequence the compactness of $\mathcal{M}_\varepsilon(\boldsymbol{\mu})$. As for the duality, we prove a slightly stronger result. Let $\overline{\mathcal{H}} \subset \mathcal{H}$ be the subset of $H = (H_k)_{1\leq k\leq N-1}$ such that $H_k \in \mathcal{C}_b(\Omega^k; \mathbb{R}^d)$ for $k = 1, \ldots, N-1$. Define the minimization problem:

$$\overline{\mathsf{D}}_\varepsilon(\boldsymbol{\mu}) := \inf_{(H,\psi)\in\overline{\mathcal{D}}_\varepsilon} \left[\sum_{k=1}^{N} \int_{\mathbb{R}^d} \psi_k(\mathsf{x})\mu_k(d\mathsf{x})\right] \quad \text{where } \overline{\mathcal{D}}_\varepsilon := \mathcal{D}_\varepsilon \cap \big(\overline{\mathcal{H}} \times \Lambda^N\big).$$

Then, by definition, $\mathsf{P}_\varepsilon(\boldsymbol{\mu}) \leq \mathsf{D}_\varepsilon(\boldsymbol{\mu}) \leq \overline{\mathsf{D}}_\varepsilon(\boldsymbol{\mu})$. Define the function $\Phi : \mathcal{P}(\boldsymbol{\mu}) \times \overline{\mathcal{H}} \to \mathbb{R}$ by

$$\Phi(\mathbb{P}, H) := \mathbb{E}_{\mathbb{P}}\left[c(\mathsf{S}_1, \ldots, \mathsf{S}_N) - \sum_{k=1}^{N-1} H_k(\mathsf{S}_1, \ldots, \mathsf{S}_k) \cdot (\mathsf{S}_{k+1} - \mathsf{S}_k)\right]$$

$$+\varepsilon\sum_{k=1}^{N-1}\|H_k\|_\infty.$$

Since $\Phi(\cdot, H)$ is continuous and concave for all $H\in\overline{\mathcal{H}}$ and $\Phi(\mathbb{P},\cdot)$ is continuous and convex for all $\mathbb{P}\in\mathcal{P}(\boldsymbol{\mu})$, then it holds that $\sup_{\mathbb{P}\in\mathcal{P}(\boldsymbol{\mu})}\inf_{H\in\overline{\mathcal{H}}}\Phi(\mathbb{P},H)=\inf_{H\in\overline{\mathcal{H}}}\sup_{\mathbb{P}\in\mathcal{P}(\boldsymbol{\mu})}\Phi(\mathbb{P},H)$ in view of the min-max theorem as $\mathcal{P}(\boldsymbol{\mu})$ is convex and compact. Hence,

$$\begin{aligned}\overline{\mathsf{D}}_\varepsilon(\boldsymbol{\mu}) &= \inf_{H\in\overline{\mathcal{H}}}\inf_{\psi\in\Lambda^N:\sum_{k=1}^N\psi(\mathsf{x}_k)\ge c(\mathsf{x}_1,\dots,\mathsf{x}_N)-\sum_{k=1}^{N-1}(H_k(\mathsf{x}_1,\dots\mathsf{x}_k)\cdot(\mathsf{x}_{k+1}-\mathsf{x}_k)-\varepsilon\|H_k\|_\infty)}\\ &\quad\times\sum_{k=1}^N\int_{\mathbb{R}^d}\psi_k(\mathsf{x})\mu_k(d\mathsf{x})\\ &= \inf_{H\in\overline{\mathcal{H}}}\sup_{\mathbb{P}\in\mathcal{P}(\boldsymbol{\mu})}\Phi(\mathbb{P},H)=\sup_{\mathbb{P}\in\mathcal{P}(\boldsymbol{\mu})}\inf_{H\in\overline{\mathcal{H}}}\Phi(\mathbb{P},H)\\ &= \sup_{\mathbb{P}\in\mathcal{M}_\varepsilon(\boldsymbol{\mu})}\inf_{H\in\overline{\mathcal{H}}}\Phi(\mathbb{P},H)\le\sup_{\mathbb{P}\in\mathcal{M}_\varepsilon(\boldsymbol{\mu})}\mathbb{E}_{\mathbb{P}}\big[c(\mathsf{S}_1,\dots,\mathsf{S}_N)\big]=\mathsf{P}_\varepsilon(\boldsymbol{\mu}),\end{aligned}$$

where the second equality follows from the classical duality of Kantorovich, and the fourth equality is by the fact $\inf_{H\in\overline{\mathcal{H}}}\Phi(\mathbb{P},H)=-\infty$ once $\mathbb{P}\notin\mathcal{M}_\varepsilon(\boldsymbol{\mu})$. □

PROOF OF PROPOSITION 3.7. As $\mathcal{W}^{\oplus}(\hat{\boldsymbol{\mu}}^n,\boldsymbol{\mu})=\sum_{k=1}^N\mathcal{W}(\hat{\mu}_k^n,\mu_k)$ with every $\mu_k$ satisfying Assumption 3.2, it suffices to deal with $\mathcal{W}(\hat{\mu}_1^n,\mu_1)$. For notational simplicity, we write $\hat{\mu}^n\equiv\hat{\mu}_1^n$ and $\mu\equiv\mu_1$. In the rest of the proof, we refer to [19]. Combining Lemmas 5 and 6 together with the proof of Theorem 1, it holds that

$$\mathbb{E}\big[\mathcal{W}(\hat{\mu}^n,\mu)\big]\le 24(M_\theta+1)d^{(1-\theta)/2}2^\theta\sum_{i\ge0}2^i\sum_{j\ge0}2^{-j}\min\big(\varepsilon_i,2^{dj/2}(\varepsilon_i/n)^{1/2}\big)$$

with $\varepsilon_i:=2^{-\theta i}$.

For every $\varepsilon\in(0,1)$, it follows by a straightforward computation that

$$\sum_{j\ge0}2^{-j}\min\big(\varepsilon,2^{dj/2}(\varepsilon/n)^{1/2}\big)\le\frac{9}{2\log2}\begin{cases}\min\big(\varepsilon,(\varepsilon/n)^{1/2}\big) & \text{if } d=1,\\ \min\big(\varepsilon,(\varepsilon/n)^{1/2}\log(2+\varepsilon n)\big) & \text{if } d=2,\\ \min\big(\varepsilon,\varepsilon(\varepsilon n)^{-1/d}\big) & \text{if } d>2.\end{cases}$$

Next, we substitute $\varepsilon$ by $\varepsilon_i$ and distinguish different $d$. If $d=1$, then

$$\sum_{i\ge0}2^i\min\big(\varepsilon_i,(\varepsilon_i/n)^{1/2}\big)\le\begin{cases}2\sqrt{2}n^{1/\theta-1}/\big((2^{1-\theta/2}-1)(1-2^{1-\theta})\big) & \text{if } \theta<2,\\ 4(1+\log n)n^{-1/2} & \text{if } \theta=2,\\ n^{-1/2}/\big(1-2^{1-\theta/2}\big) & \text{if } \theta>2.\end{cases}$$

If $d=2$, then

$$\sum_{i\ge0}2^i\min\big(\varepsilon_i,(\varepsilon_i/n)^{1/2}\log(2+\varepsilon_i n)\big)$$

$$\leq \begin{cases} 7n^{1/\theta-1}/(2^{1-\theta/2}-1)^2 & \text{if } \theta < 2, \\ 6(1+(\log n)^2)n^{-1/2} & \text{if } \theta = 2, \\ (1+\log n)n^{-1/2}/(1-2^{1-\theta/2}) & \text{if } \theta > 2. \end{cases}$$

If $d > 2$, then

$$\sum_{i \geq 0} 2^i \min(\varepsilon_i, \varepsilon_i(\varepsilon_i n)^{-1/d})$$

$$\leq \begin{cases} 3n^{1/\theta-1}/((2^{1-\theta(1-1/d)}-1)(1-2^{1-\theta})) & \text{if } \theta < d/(d-1), \\ 6(1+\log n)n^{-1/d} & \text{if } \theta = d/(d-1), \\ n^{-1/d}/(1-2^{1-\theta(1-1/d)}) & \text{if } \theta > d/(d-1). \end{cases}$$

The proof is completed with $C(\theta, d)$ being the product of the corresponding coefficients. $\square$

LEMMA A.1. *With the same conditions and notation of Proposition 4.2, there exist measurable functions $f_1, \ldots, f_N : \Omega^2 \to \mathbb{R}^d$ such that $\mathcal{Q} \circ (X_k, Y_k)^{-1} = \mathbb{P}_k$, where $Y_k := f_k(X_k, Z_k)$ for $k = 1, \ldots, N$.*

PROOF. Without loss of generality, we only prove for $k = 1$. Further, we drop the subscript without any danger of confusion, that is, $X \equiv X_1$, $Z \equiv Z_1$, $\mu \equiv \mu_1$, $\mathbb{P} \equiv \mathbb{P}_1$, etc. Disintegrating with respect to $\mu$, one has $\mathbb{P}(d\mathsf{x}, d\mathsf{y}) = \mu(d\mathsf{x}) \otimes \lambda_{\mathsf{x}}(d\mathsf{y})$, where $(\lambda_{\mathsf{x}}(d\mathsf{y}))_{\mathsf{x}\in\mathbb{R}^d}$ denotes the regular conditional probability distribution (r.c.p.d.). Hence, the above claim is equivalent to the existence of a measurable function $f : \Omega^2 \to \mathbb{R}^d$ satisfying, for $\mu$-a.e. $\mathsf{x} \in \mathbb{R}^d$,

$$\mathcal{Q}[f(\mathsf{x}, Z) \in A | X = \mathsf{x}] = \lambda_{\mathsf{x}}(A) \quad \text{for all } A \in \mathcal{B}(\mathbb{R}^d),$$

or namely, $f(\mathsf{x}, \cdot)$ transfers the law of $Z$ to $\lambda_{\mathsf{x}}$ for $\mu$-a.e. $\mathsf{x} \in \mathbb{R}^d$. We first prove this claim for the case of $d = 1$, that is, $\mathsf{x} = x$ and then conclude for the general case.

(i) Let $F$ and $G_x$ be respectively the cumulative distribution functions of $Z$ and $\lambda_x$, and define the right-continuous inverse by $G_x^{-1}(t) := \inf\{y \in \mathbb{R} : G_x(y) > t\}$. Define further $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) := G_x^{-1} \circ F(y)$, then $f$ is measurable by the definition of r.c.p.d. Moreover, it follows by Villani [38], pages 19–20, that for $\mu$-a.e. $x \in \mathbb{R}$ one has $\mathcal{Q}[Y \in A | X = x] = \lambda_x(A)$ for all $A \in \mathcal{B}(\mathbb{R})$, which concludes the proof.

(ii) Now let us treat the general case. Recall that $\mathsf{x} = (x_1, \ldots, x_d)$, $\mathsf{y} = (y_1, \ldots, y_d)$ and $\mathsf{z} = (z_1, \ldots, z_d)$. We proceed as follows.

*Step 1*: Take the marginal distributions on the first coordinate for $Z$ and $\lambda_{\mathsf{x}}$, denoted by $F_1(z_1)\,dz_1$ and $\lambda^1_{\mathsf{x}}(dy_1)$, where we note that $Z$ admits a density function on $\mathbb{R}^d$. Then repeat the the procedure of (i), and construct the measurable map $f_1(\mathsf{x}, \cdot)$ which may transfers the law $F_1(z_1)\,dz_1$ to the other one $\lambda^1_{\mathsf{x}}(dy_1)$.

*Step 2*: Next take the marginals on the first two coordinates for $Z$ and $\lambda_{\mathsf{x}}$, $F_2(z_1, z_2)\,dz_1\,dz_2$ and $\lambda^2_{\mathsf{x}}(dy_1, dy_2)$, and disintegrate them with respect to the first one. This yields

$$F_2(z_1, z_2)\,dz_1\,dz_2 := F_1(z_1)\,dz_1 \otimes F_{z_1,2}(z_2)\,dz_2 \quad \text{and}$$
$$\lambda^2_{\mathsf{x}}(dy_1, dy_2) := \lambda^1_{\mathsf{x}}(dy_1) \otimes \lambda^2_{\mathsf{x},y_1}(dy_2).$$

For each $z_1$, set $y_1 = f_1(\mathsf{x}, z_1)$, and define $f_{1,2}(\mathsf{x}, z_1, \cdot)$ according to (i), which transfers thus $F_{z_1,2}(z_2)\,dz_2$ to $\lambda^2_{\mathsf{x}, f_1(\mathsf{x},z_1)}(dy_2)$.

*Step 3*: We repeat the construction of *Step 2* by adding coordinates one after the other and defining $f_{1,2,3}(\mathsf{x}, z_1, z_2, \cdot)$, etc. After $N$ steps, this produces the required map $f(\mathsf{x}, z)$ which transports the law of $Z$ to $\lambda_{\mathsf{x}}$. $\square$

**Acknowledgments.** The authors thank Guillaume Carlier, Bruno Lévy, Tongseok Lim, Terry Lyons and Pietro Siorpaes for insightful discussions and comments. The second author also gratefully acknowledges support from St John's College, Oxford.

## REFERENCES

[1] ALFONSI, A., CORBETTA, J. and JOURDAIN, B. (2019). Sampling of one-dimensional probability measures in the convex order and computation of robust option price bounds. *Int. J. Theor. Appl. Finance* **22** 1950002, 41. MR3951948

[2] BACKHOFF-VERAGUAS, J. and PAMMER, G. (2019). Stability of martingale optimal transport and weak optimal transport. Available at arXiv:1904.04171.

[3] BEIGLBÖCK, M., HENRY-LABORDÈRE, P. and PENKNER, F. (2013). Model-independent bounds for option prices—A mass transport approach. *Finance Stoch.* **17** 477–501. MR3066985

[4] BEIGLBÖCK, M. and JUILLET, N. (2016). On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.* **44** 42–106. MR3456332

[5] BEIGLBÖCK, M., LIM, T. and OBŁÓJ, J. (2019). Dual attainment for the martingale transport problem. *Bernoulli* **25** 1640–1658. MR3961225

[6] BENAMOU, J.-D. and BRENIER, Y. (2000). A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math.* **84** 375–393. MR1738163

[7] BENAMOU, J.-D., CARLIER, G., CUTURI, M., NENNA, L. and PEYRÉ, G. (2015). Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.* **37** A1111–A1138. MR3340204

[8] BONNANS, J. F. and TAN, X. (2013). A model-free no-arbitrage price bound for variance options. *Appl. Math. Optim.* **68** 43–73. MR3072239

[9] BREEDEN, D. T. and LITZENBERGER, R. H. (1978). Prices of state-contingent claims implicit in option prices. *J. Bus.* **51** 621–51.

[10] BROWN, H., HOBSON, D. and ROGERS, L. C. G. (2001). Robust hedging of barrier options. *Math. Finance* **11** 285–314. MR1839367

[11] CHACON, R. V. (1977). Potential processes. *Trans. Amer. Math. Soc.* **226** 39–58. MR0501374

[12] CLAISSE, J., GUO, G. and HENRY-LABORDÈRE, P. (2018). Some results on Skorokhod embedding and robust hedging with local time. *J. Optim. Theory Appl.* **179** 569–597. MR3865340

[13] COX, A. M. G. and OBŁÓJ, J. (2011). Robust hedging of double touch barrier options. *SIAM J. Financial Math.* **2** 141–182. MR2772387

[14] COX, A. M. G. and OBŁÓJ, J. (2011). Robust pricing and hedging of double no-touch options. *Finance Stoch.* **15** 573–605. MR2833100

[15] DAVIS, M., OBŁÓJ, J. and RAVAL, V. (2014). Arbitrage bounds for prices of weighted variance swaps. *Math. Finance* **24** 821–854. MR3274933

[16] DE MARCH, H. and TOUZI, N. (2019). Irreducible convex paving for decomposition of multidimensional martingale transport plans. *Ann. Probab.* **47** 1726–1774. MR3945758

[17] DELATTRE, S., FORT, J.-C. and PAGÈS, G. (2004). Local distortion and $\mu$-mass of the cells of one dimensional asymptotically optimal quantizers. *Comm. Statist. Theory Methods* **33** 1087–1117. MR2054859

[18] DOLINSKY, Y. and SONER, H. M. (2014). Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Related Fields* **160** 391–427. MR3256817

[19] FOURNIER, N. and GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738. MR3383341

[20] GALICHON, A., HENRY-LABORDÈRE, P. and TOUZI, N. (2014). A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.* **24** 312–336. MR3161649

[21] GENEVAY, A., CUTURI, M., PEYRÉ, G. and BACH, F. (2018). Stochastic optimization for large-scale optimal transport. In *NIPS'16 Proceedings of the 30th International Conference on Neural Information Processing Systems* 3440–3448.

[22] GHOUSSOUB, N., KIM, Y.-H. and LIM, T. (2019). Structure of optimal martingale transport plans in general dimensions. *Ann. Probab.* **47** 109–164. MR3909967

[23] GRAF, S. and LUSCHGY, H. (2000). *Foundations of Quantization for Probability Distributions*. *Lecture Notes in Math.* **1730**. Springer, Berlin. MR1764176

[24] GRAF, S., LUSCHGY, H. and PAGÈS, G. (2008). Distortion mismatch in the quantization of probability measures. *ESAIM Probab. Stat.* **12** 127–153. MR2374635

[25] GUO, G. (2018). A stability result on optimal Skorokhod embedding. Available at arXiv:1701.08204.

[26] HENRY-LABORDÈRE, P. (2013). Automated option pricing: Numerical methods. *Int. J. Theor. Appl. Finance* **16** 1350042, 27. MR3157553

[27] HOBSON, D. (1998). Robust hedging of the lookback option. *Finance Stoch.* **2** 329–347.

[28] HOBSON, D. and KLIMMEK, M. (2015). Robust price bounds for the forward starting straddle. *Finance Stoch.* **19** 189–214. MR3292129

[29] HOBSON, D. and NEUBERGER, A. (2012). Robust bounds for forward start options. *Math. Finance* **22** 31–56. MR2881879

[30] HOU, Z. and OBŁÓJ, J. (2018). Robust pricing-hedging dualities in continuous time. *Finance Stoch.* **22** 511–567. MR3816548

[31] JUILLET, N. (2016). Stability of the shadow projection and the left-curtain coupling. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1823–1843. MR3573297

[32] LÉVY, B. (2015). A numerical algorithm for $L_2$ semi-discrete optimal transport in 3D. *ESAIM Math. Model. Numer. Anal.* **49** 1693–1715. MR3423272

[33] LIM, T. (2017). Multi-martingale optimal transport. Available at arXiv:1611.01496.

[34] OBŁÓJ, J. and SIORPAES, P. (2017). Structure of martingale transports in finite dimensions. Available at arXiv:1702.08433.

[35] RACHEV, S. T. and RÜSCHENDORF, L. (1998). *Mass Transportation Problems. Vol. II. Applications. Probability and Its Applications*. Springer, New York. MR1619171

[36] SKOROHOD, A. V. (1976). On a representation of random variables. *Teor. Veroyatn. Primen.* **21** 645–648. MR0428369

[37] STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Stat.* **36** 423–439. MR0177430

[38] VILLANI, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454

[39] WIESEL, J. (2019). Continuity of the martingale optimal transport problem on the real line. Available at arXiv:1905.04574.

MATHEMATICAL INSTITUTE
UNIVERSITY OF OXFORD
AWB, ROQ, WOODSTOCK ROAD
OXFORD OX2 6GG
UNITED KINGDOM
E-MAIL: gaoyue.guo@maths.ox.ac.uk
jan.obloj@maths.ox.ac.uk