Computational Methods for Martingale Optimal Transport problems*

Gaoyue Guo† and Jan Obłój‡
Mathematical Institute, University of Oxford
AWB, ROQ, Oxford OX2 6GG, UK

October 24, 2017

Abstract
We establish numerical methods for solving the martingale optimal transport problem (MOT) – a version of the classical optimal transport with an additional martingale constraint on transport’s dynamics. We prove that the MOT value can be approximated using linear programming (LP) problems which result from a discretisation of the marginal distributions combined with a suitable relaxation of the martingale constraint. Specialising to dimension one, we provide bounds on the convergence rate of the above scheme. We also show a stability result under only partial specification of the marginal distributions. Finally, we specialise to a particular discretisation scheme which preserves the convex ordering and does not require the martingale relaxation. We introduce an entropic regularisation for the corresponding LP problem and detail the corresponding iterative Bregman projection. We also rewrite its dual problem as a minimisation problem without constraint and solve it by computing the concave envelope of scattered data.

Keywords: martingale optimal transport, martingale relaxation, duality, robust hedging, linear programming, iterative Bregman projection, entropic regularisation, concave envelope.

1 Introduction
Let \( \mu \) and \( \nu \) be two probability measures on \( \mathbb{R}^d \), with finite first moment and increasing in convex order, and \( c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R} \) be a measurable function. The martingale optimal transport (MOT) problem aims at maximising

\[
\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y) \pi(dx,dy), \tag{1}
\]

*This research is supported by the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 335421. The second author also gratefully acknowledges support from St John’s College, Oxford. Authors are thankful to Guillaume Carlier, Bruno Lévy and Tongseok Lim for insightful discussions and comments.

†E-mail: gaoyue.guo@maths.ox.ac.uk; web: http://www.maths.ox.ac.uk/people/gaoyue.guo
‡E-mail: jan.obloj@maths.ox.ac.uk; web: http://www.maths.ox.ac.uk/people/jan.obloj
among all probability measures $\pi$ such that
\[
\pi[ E \times \mathbb{R}^d ] = \mu[E] \quad \text{and} \quad \pi[\mathbb{R}^d \times E] = \nu[E], \quad \text{for all} \ E \in \mathcal{B}(\mathbb{R}^d),
\]
\[
\int_{\mathbb{R}^d} y \pi_x(dy) = x, \quad \text{for} \ \mu - \text{a.e.} \ x \in \mathbb{R}^d,
\]
where $(\pi_x)_{x \in \mathbb{R}^d}$ denotes the regular conditional probability distribution (r.c.p.d.) of $\pi$ w.r.t. $\mu$. Under quite mild assumptions, there exists an optimiser $\pi^*(\mu, \nu)$ which maximises the integral (1). Similar to the classical optimal transport (OT), see e.g. \[31, 34\], we are able to identify its convex dual. The dual elements are maps $\varphi, \psi : \mathbb{R}^d \rightarrow \mathbb{R}$ and $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$ satisfying the constraint
\[
\varphi(x) + \psi(y) + h(x)(y - x) \geq c(x, y), \quad \text{for all} \ x, y \in \mathbb{R}^d.
\]
The dual objective is to minimise
\[
\int_{\mathbb{R}^d} \varphi(x) \mu(dx) + \int_{\mathbb{R}^d} \psi(x) \nu(dx)
\]
overall $(\varphi, \psi, h)$ satisfying the inequality (3). Interestingly, the dual problem also has a financial interpretation as a robust hedging problem. Indeed, the left–hand side of (3) represents a portfolio consisting of underlying assets and European option instruments and the inequality means that the portfolio dominates, or super–replicates, the “exotic” payoff $c$. The dual problem thus represents the minimal super–replication cost, see the seminal paper \[24\] of Hobson for further discussions. Beiglböck, Henry-Labordère & Penkner \[3\] first studied MOT in a discrete time setup and established Kantorovich duality, i.e. that the primal and the dual problems have the same value. In continuous time the analogous duality was investigated in a stream of papers, see e.g. \[21, 19, 20, 28, 22\]. MOT continues to be an active field of research and we mention works on characterising the optimisers for a MOT, see e.g. \[4, 23\] or on the attainment for the dual problem, see \[5\].

Naturally, both from the theoretical and from the applied point of view, it is important to solve MOT problems explicitly. Hobson & Neuberger \[27\] and Hobson & Klimmek \[26\] focused on the particular case of $\mathbb{R}^d = 1$ and $c(x, y) = \pm |x - y|$, and showed that the optimiser $\pi^*(\mu, \nu)$ can be obtained in a semi–closed form. Another trivial but important observation is that, when $\mu$ and $\nu$ have finite supports, i.e.
\[
\mu(dx) = \sum_{1 \leq i \leq m} \alpha_i \delta_{x_i}(dx) \quad \text{and} \quad \nu(dy) = \sum_{1 \leq j \leq n} \beta_j \delta_{y_j}(dy),
\]
the MOT problem reduces to the following linear programming (LP) problem
\[
\sup_{(p_{i,j})_{1 \leq i \leq m, 1 \leq j \leq n} \in \mathbb{R}^{mn}_+} \sum_{1 \leq i \leq m, 1 \leq j \leq n} p_{i,j} c(x_i, y_j)
\]
s.t.
\[
\sum_{1 \leq j \leq n} p_{k,j} = \alpha_k, \quad \text{for} \ k = 1, \cdots, m,
\]
\[
\sum_{1 \leq i \leq m} p_{i,k} = \beta_k, \quad \text{for} \ k = 1, \cdots, n,
\]
\[
\sum_{1 \leq j \leq n} p_{k,j} y_j = \alpha_k x_k, \quad \text{for} \ k = 1, \cdots, m.
\]
Recently, Juillet [29] proved that if
\[ d = 1 \quad \text{and} \quad c(x, y) = \varphi(x) \psi(y) \quad \text{or} \quad c(x, y) = h(x - y), \]
where \( \varphi, \psi, h : \mathbb{R} \to \mathbb{R} \) are assumed to satisfy the conditions in Remark 2.10 of [29],
then the optimiser \( \pi^*(\mu, \nu) \) is Lipschitz w.r.t \( (\mu, \nu) \) under a topology of Wasserstein type. In particular,
\[
(\mu^n, \nu^n) \to (\mu, \nu) \implies \int_{\mathbb{R}^2} c(x, y) \, d\pi^*(\mu^n, \nu^n) \to \int_{\mathbb{R}^2} c(x, y) \, d\pi^*(\mu, \nu).
\]
Hence, the approximation scheme becomes tractable by solving a LP problem related to \( (\mu^n, \nu^n) \) which have finite support and are “close” to \( (\mu, \nu) \) under the topology above. The LP approach was pioneered in fact in Davis, Obłój & Raval [18], where instead of full marginal constraint \( \nu \), only finitely many constraints were given which, thanks to convexity, led to optimisers with finite support. In continuous time, MOT for many cost functionals could be solved in an explicit, or semi-explicit form, using methods from stochastic optimal control, see e.g. [7, 12, 21], and from Skorokhod embeddings, see e.g. [10, 14, 15, 18, 25].

However, in contrast to the theory and applications recalled above, numerical methods for MOT are close to non-existent. In fact, the martingale condition renders any of the usual OT approximation techniques unusable. This paper fills in this important gap. We develop a numerical method for solving a one-step MOT problem in \( \mathbb{R}^d \) in a systematic way. Our approximation for the primal problem relies on a suitable discretisation and solving a LP to which we apply the entropic regularisation known from OT, see [6]. The scheme converges and in dimension one, \( d = 1 \), we obtain the convergence rate. We also develop numerical methods for the dual problem. Together, these allow to sandwich the true value function. Our investigation involves a number of novel results and techniques which, we believe, are of independent interest.

The paper is organised as follows. In the rest of Section 1 we formulate the relaxed MOT problem and present its Kantorovich duality. Next, in Section 2 we introduce an approximating optimisation problem with a relaxed martingale condition. We investigate its dependence on the marginal distributions and obtain a numerical scheme by discretising the marginals. We show the scheme converges and provide some numerical examples. In Section 3 we study its convergence rate in the case of two marginals on real line. As a by-product, another optimisation problem is considered, where the marginal constraints are replaced by the constraints on expectation for a finite number of call options. We show a stability result, i.e. the optimisation problem converges to some MOT problem as more and more call options are quoted. In Section 4 we focus on the discretised marginals, where the MOT problem is indeed an LP problem. We introduce an entropic regularisation for the LP problem and adapt the iterative Bregman projection initiated by Bregman [9] to the martingale case. Finally, we reformulate its dual problem as a minimisation problem \( \min_{\mathbb{R}^n} J \), where \( J : \mathbb{R}^n \to \mathbb{R} \) denotes some convex functional and is determined by computing the concave envelope of scattered data.

1.1 Relaxation of the Martingale Optimal Transport problem

Let \( \mathcal{X} \subseteq \mathbb{R}^d \) be a closed set and denote by \( \Pi \) the set of probability measures supported on \( \mathcal{X} \) and admitting a finite first moment. Let \( \Lambda \) be the space of Lipschitz
functions on $\mathcal{X}$ and, given $\psi \in \Lambda$, denote by $\text{Lip}(\psi)$ its Lipschitz constant on $\mathcal{X}$. For each $L > 0$, let $\Lambda_L \subseteq \Lambda$ be the subspace of functions $\psi$ with $\text{Lip}(\psi) \leq L$. When we want to stress the dependence on $\mathcal{X}$, or when considering different sets $\mathcal{X}$, we include $\mathcal{X}$ in the notation, e.g. we write $\Pi(\mathcal{X})$, $\Lambda_L(\mathcal{X})$ or $\text{Lip}_\mathcal{X}(\psi)$.

Let $X = (X_k)_{1 \leq k \leq N}$ be the coordinate process on the $N$–product space $\mathcal{X}^N$, i.e. $X_k(x) := x_k$ for all $x = (x_k)_{1 \leq k \leq N} \in \mathcal{X}^N$, and $F = (\mathcal{F}_k)_{1 \leq k \leq N}$ be its natural filtration, i.e. $\mathcal{F}_k := \sigma(X_1, \cdots, X_k)$. Note that the restriction of $\mathcal{F}_k$ to $\mathcal{X}^k$ is simply $\mathcal{B}(\mathcal{X}^k)$, the Borel $\sigma$–algebra on $\mathcal{X}^k$. Given a vector $\mu = (\mu_k)_{1 \leq k \leq N} \in \Pi^N$, define the collection of transport plans with marginals $\mu$ by

$$\Pi(\mu) := \left\{ \pi \in \Pi(\mathcal{X}^N) : \pi \circ X_k^{-1} = \mu_k, \text{ for } k = 1, \cdots, N \right\},$$

where $\pi \circ X_k^{-1}$ denotes push forward of $\pi$ via $X_k$. The MOT involves optimisation over $\pi \in \Pi(\mu)$ under which $X$ is a martingale, as seen for $N = 2$ in [2]. However, as we will see below, when considering stability of, and approximation to, the MOT problem it is convenient to relax the martingale condition. To this end, for a fixed $\varepsilon \in \mathbb{R}_+$ and a vector $\mu \in \Pi^N$, we let $M_\varepsilon(\mu)$ be the set of $\pi \in \Pi(\mu)$ such that

$$\left| \mathbb{E}_\pi[X_{k+1}|\mathcal{F}_k] - X_k \right| \leq \varepsilon, \quad \pi - \text{a.s.}, \quad \text{for } k = 1, \cdots, N - 1,$$

which may be also rewritten as

$$\mathbb{E}_\pi[h(X_1, \cdots, X_k) \cdot (X_{k+1} - X_k)] \leq \varepsilon \|h\|_\infty, \quad \text{for all } h \in L^\infty(\mathcal{X}^k, \mathcal{B}(\mathcal{X}^k); \mathbb{R}^d).$$

In particular, $M_0(\mu)$ is the set of martingale transport plans. Strassen [33] showed that $M_0(\mu) \neq \emptyset$ if and only if measures $\mu_k$, $k = 1, \ldots, N$ are increasing in convex order. In order to generalise this characterisation to $M_\varepsilon(\mu) \neq \emptyset$ we need the following definition.

**Definition 1.1.** $\mu$, $\nu \in \Pi$ are said to be increasing in $\varepsilon$–convex order, denoted by $\mu \preceq_{\varepsilon} \nu$, if

$$\int_{\mathbb{R}^d} \left( \sup_{z : |z| \leq \varepsilon} \psi(z) \right) \mu(dx) \geq \int_{\mathbb{R}^d} \psi(x) \nu(dx)$$

for all concave functions $\psi \in \Lambda$. Denote further by $\Pi_{\varepsilon}^N \subseteq \Pi^N$ the collection of vectors $\mu$ such that $\mu_k \preceq_{\varepsilon} \mu_{k+1}$ for $k = 1, \cdots, N - 1$.

The following theorem is a generalisation of Strassen’s theorem. Its proof uses original results of Strassen [33] and is presented in the Appendix.

**Theorem 1.2.** Let $\mu \in \Pi^N$. Then $M_\varepsilon(\mu) \neq \emptyset$ if and only if $\mu \in \Pi_{\varepsilon}^N$.

Given $\mu \in \Pi_{\varepsilon}^N$ and a measurable cost function $c : \mathcal{X}^N \rightarrow \mathbb{R}$, we consider the following relaxation of the Martingale Optimal Transport problem

$$P_{\varepsilon}^\ast(\mu) := \sup_{\pi \in M_\varepsilon(\mu)} \mathbb{E}_\pi[c(X)].$$

Similarly to the MOT problem, [5] admits a dual formulation. Let $H$ be the set of progressive measurable processes $h = (h_k)_{1 \leq k \leq N-1}$ taking values in $\mathbb{R}^d$, i.e. [5].
of \( k \), then there exists an optimiser \( \mu^* \) such that for all \( x \in \Xi^N \)

\[
\sum_{k=1}^{N} \psi_k(x_k) + \sum_{k=1}^{N-1} h_k(x_1, \ldots, x_k) \cdot (x_{k+1} - x_k) - \varepsilon \sum_{k=1}^{N-1} |h_k(x_1, \ldots, x_k)| \geq c(x). \tag{6}
\]

The dual problem to (5) is defined by

\[
D^c_{\varepsilon}(\mu) := \inf_{(\psi, \mathbf{h}) \in \mathcal{D}_\varepsilon} \left[ \sum_{k=1}^{N} \psi_k d\mu_k \right]. \tag{7}
\]

It is easy to see that \( M_{\varepsilon}(\mu) \) is weakly compact since it is a closed subset of \( \Pi(\mu) \).

An application of the Min–Max theorem allows then to establish the Kantorovich duality between \( [3] \) and \( [7] \) under suitable mild assumptions on \( c \). We omit the proof as it simply repeats the arguments in [3] where the result was shown for \( \varepsilon = 0 \).

**Theorem 1.3.** Assume \( \mu \in \Pi^2_{\varepsilon} \) and \( c \) is upper semicontinuous and satisfies

\[
\sup_{x \in \Xi^N} \frac{c(x)}{1 + |x|} < +\infty.
\]

Then there exists an optimiser \( \pi^* \) for \( P^c_{\varepsilon}(\mu) \), i.e. \( \pi^* \in M_{\varepsilon}(\mu) \) and \( P^c_{\varepsilon}(\mu) = \mathbb{E}_n[\pi^*] \). Moreover, the duality \( P^c_{\varepsilon}(\mu) = D^c_{\varepsilon}(\mu) \) holds.

**Remark 1.4.** From the financial point of view, the left–hand side of (6) stands for a super–replication of \( c \) by trading dynamically in the underlying assets and statically in a range of Vanilla options. \( \varepsilon \) denotes a constant transaction cost rate and the corresponding term represents transaction costs resulting from the trading of the underlying assets. As Inequality (6) holds under any market scenario, the dual problem (7) denotes the minimal super-replication cost.

In what follows, unless otherwise specified, we take \( \Xi = \mathbb{R}^d \). We also simplify the notations as follows:

- we drop the superscript \( c \) when the cost is fixed, i.e. we write \( P_{\varepsilon}(\mu) \) for \( P^c_{\varepsilon}(\mu) \), \( D_{\varepsilon}(\mu) \) for \( D^c_{\varepsilon}(\mu) \);
- we drop the subscript \( \varepsilon \) when \( \varepsilon = 0 \), e.g. we write \( \leq \) for \( \leq_0 \), \( M(\mu) \) for \( M_0(\mu) \), \( P(\mu) \) for \( P_0(\mu) \) etc.;
- when \( N = 2 \), we let \( X \equiv X_1, Y \equiv X_2, \mu \equiv \mu_1 \) and \( \nu \equiv \mu_2 \).

We finish the preliminaries by introducing a Wasserstein-type metric on \( \Pi^N \). For \( \mu, \nu \in \Pi \), their Wasserstein distance is defined by

\[
\mathcal{W}(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \mathbb{E}_\pi[|X - Y|] = \sup_{\psi \in \Lambda_1} \left\{ \int \psi d\mu - \int \psi d\nu \right\}, \tag{8}
\]

see Chapter 1.2 of [31] for more details. We recall that \( \Pi \), equipped with the metric \( \mathcal{W} \), is a Polish space and for any \( (\mu^n)_{n \geq 1} \subset \Pi \) and \( \mu \in \Pi \),

\[
\mu^n \overset{\mathcal{W}}{\to} \mu \quad \text{if and only if} \quad \begin{cases} (i) & \mu^n \text{ converges weakly to } \mu, \\ (ii) & \int |x| d\mu^n \longrightarrow \int |x| d\mu. \end{cases}
\]

5
We endow $\Pi^N$ with the metric $W^\boxplus$ defined by, for all $\mu = (\mu_k)_{1 \leq k \leq N}, \nu = (\nu_k)_{1 \leq k \leq N} \in \Pi^N$,

$$W^\boxplus(\mu, \nu) := \sum_{k=1}^{N} W(\mu_k, \nu_k).$$

It is easy to see that $\Pi^N$ is a Polish space w.r.t. $W^\boxplus$ and $\Pi^N_{\boxplus} \subseteq \Pi^N$ is closed.

2 Numerical scheme for $P(\mu)$

We develop a unified framework for computing $P(\mu)$ numerically. As mentioned in the introduction, $P(\mu)$ reduces to an LP problem once $\mu_k$ have finite supports for $k = 1, \ldots, N$. We exploit this observation and adopt a discrete approximation approach. In this section we establish convergence results. In Section 3 we will study the convergence rates and in Section 4 we propose efficient methods to solve the LP problem for discrete measures.

2.1 Convergence of relaxed MOT problems

Our first main result, Theorem 2.1, shows that $P(\mu)$ may be approximated by considering the relaxed problem $P_{\varepsilon}(\mu^n)$ for discrete approximations $\mu^n$ of $\mu$. This provides the main insight into our proposed numerical scheme for MOT problems.

**Theorem 2.1.** Let $c \in \Lambda(X^N)$ and $\mu \in \Pi^\boxplus$. Then, for any sequence $(\mu^n)_{n \geq 1} \in \Pi^N$ converging to $\mu$ under $W^\boxplus$, one has

$$\lim_{n \to \infty} P_{\rho_n}(\mu^n) = P(\mu) \quad \text{with} \quad \rho_n := W^\boxplus(\mu^n, \mu).$$

As noted before, it is natural to try to approximate $P(\mu)$ by $P(\mu^n)$ with finitely supported measures $\mu^n$ since the latter amounts to a linear program. However, unlike in the classical optimal transport, continuity of $\mu \to P(\mu)$ is no longer clear. In fact, one has to consider suitably tailored discretisation, see Section 3.2, to even ensure that $M(\mu^n)$ is non-empty. Theorem 2.1 shows that relaxation of the martingale constraint allows to solve this problem and establish the above convergence result. In Section 3 we use the Kantorovich duality from Theorem 1.3 to study the convergence rate in Theorem 2.1 when $d = 1$. The rest of this section is devoted to the proof of Theorem 2.1.

**Proposition 2.2.** Let $\mu \in \Pi^\boxplus$. Then for any $\nu \in \Pi^N$, one has $\nu \in \Pi^\boxplus_{\varepsilon + \rho}$, where $\rho := W^\boxplus(\mu, \nu)$. If in addition $c \in \Lambda(X^N)$, then

$$P_{\varepsilon}(\mu) \leq P_{\varepsilon + \rho}(\nu) + \text{Lip}(c) \rho.$$

**Proof.** Put

$$\rho := \sum_{k=1}^{N} \rho_k, \quad \text{with} \quad \rho_k = W(\mu_k, \nu_k).$$

Take an arbitrary $\pi \in M_{\varepsilon}(\mu)$. It follows from Theorem 1 of Skorokhod [32] that, there exists an enlarged probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ which supports random variables
\( \tilde{X}_k \) and \( Z_k \) for \( k = 1, \ldots, N \) such that
\[
\tilde{\mathbb{P}} \circ (\tilde{X}_1, \ldots, \tilde{X}_N)^{-1} = \pi,
\]
where \( \rho = \rho_0 \).
\( Z_1, \ldots, Z_N \) and \((\tilde{X}_1, \ldots, \tilde{X}_N)\) are mutually independent,
\( \tilde{\mathbb{P}} \circ Z_k^{-1} \) is a standard normal distribution, for \( k = 1, \ldots, N \).

For every \( k = 1, \ldots, N \), let \( \pi_k \) be the optimal transport plan realising the Wasserstein distance between \( \mu_k \) and \( \nu_k \), i.e. \( \pi_k \in \Pi(\mu_k, \nu_k) \) and
\[
\rho_k = \int_{\mathcal{X}^2} |x_k - x_k'| \pi_k(dx_k, dx_k').
\]
It follows from Lemma 5.1 that, there exist measurable functions \( f_k : \mathcal{X}^2 \rightarrow \mathcal{X} \) such that
\[
\tilde{\mathbb{P}} \circ (\tilde{X}_k, X'_k)^{-1} = \pi_k, \text{ where } X'_k := f_k(\tilde{X}_k, Z_k).
\]
In particular \( \tilde{\mathbb{P}} \circ (X'_k)^{-1} = \nu_k \) for \( k = 1, \ldots, N \). Furthermore, one has for all \( h \in L^\infty(\mathcal{X}^k, \mathcal{B}(\mathcal{X}^k); \mathbb{R}^d) \)
\[
\mathbb{E}_\tilde{\mathbb{P}}[h(X'_1, \ldots, X'_k) \cdot (X'_{k+1} - X'_{k})] \leq \mathbb{E}_\tilde{\mathbb{P}}[h(X'_1, \ldots, X'_k) \cdot (X_{k+1} - \tilde{X}_{k})] + \mathbb{E}_\tilde{\mathbb{P}}[h(X'_1, \ldots, X'_k) \cdot (X'_{k} - \tilde{X}_{k})] \\
\leq (\rho_k + \rho_{k+1}) \|h\|_\infty + \mathbb{E}_\tilde{\mathbb{P}}[h(f_1(\tilde{X}_1, Z_1), \ldots, f_k(\tilde{X}_k, Z_k)) \cdot (\tilde{X}_{k+1} - \tilde{X}_{k})] \\
\leq \rho \|h\|_\infty + \int_{\mathcal{X}^k} \mathbb{E}_{\pi}[h(f_1(X_1, z_1), \ldots, f_k(X_k, z_k)) \cdot (X_{k+1} - X_k)]G_k(dz_1, \ldots, dz_k) \\
\leq (\varepsilon + \rho) \|h\|_\infty,
\]
where \( G_k \) denotes the joint distribution of \( Z_1, \ldots, Z_k \). Set \( \pi' := \tilde{\mathbb{P}} \circ (X'_1, \ldots, X'_N)^{-1} \).
Therefore, \( \pi' \in M_{e + \rho}(\nu) \) and \( \nu \in \Pi^\infty_{\varepsilon + \rho} \). To conclude the proof, notice that
\[
\mathbb{E}_\pi[c(X)] - P_{e + \rho}(\nu) \leq \mathbb{E}_\pi[c(X)] - \mathbb{E}_{\pi'}[c(X)] \\
= \mathbb{E}_\tilde{\mathbb{P}}[c(\tilde{X}_1, \ldots, \tilde{X}_N) - c(X'_1, \ldots, X'_N)] \\
\leq \text{Lip}(c) \sum_{k=1}^N \mathbb{E}_\tilde{\mathbb{P}}[|X'_{k} - \tilde{X}_{k}|] = \text{Lip}(c)\rho.
\]
Maximising w.r.t \( \pi \in M_e(\mu) \), it follows that \( P_e(\mu) \leq P_{e + \rho}(\nu) + \text{Lip}(c)\rho \).

In consequence, one has the immediate corollary below.

**Corollary 2.3.** Let \( c \in \Lambda(\mathcal{X}^N) \) and \( (\mu^n)_{n \geq 1} \subseteq \Pi^N \) be a sequence converging to \( \mu \in \Pi^\infty \) under \( W^\oplus \). Then, for any \((\delta_n)_{n \geq 1} \subseteq \mathbb{R}^+ \) converging to zero,
\[
P(\mu) \leq P_{\rho_n + \delta_n}(\mu^n) + \text{Lip}(c)\rho_n \\
\leq P_{2(\rho_n + \delta_n)}(\mu) + 2\text{Lip}(c)\rho_n,
\]
where \( \rho_n := W^\oplus(\mu^n, \mu) \).

**Proof.** The first inequality follows by taking \( \varepsilon = 0, \nu = \mu^n \) and \( \rho = \rho_n \). As for the second one, interchanging \( \mu \) and \( \mu^n \), it suffices to take \( \varepsilon = \rho_n + \delta_n \) and \( \rho = \rho_n \).
Remark 2.4. Here $\delta_n$ may be used to capture the error resulting from the numerical discretisation of $\mu$.

**Proposition 2.5.** Let $c \in \Lambda(\mathcal{X}^n)$. 
(i) For every fixed $\varepsilon \in \mathbb{R}_+$, the map

$$
\Pi^\varepsilon \ni \mu \mapsto P_\varepsilon(\mu) \in \mathbb{R}
$$

is upper semicontinuous under $\mathcal{W}^\oplus$.

(ii) For every fixed $\mu \in \Pi^\varepsilon$, the map

$$
\mathbb{R}_+ \ni \varepsilon \mapsto P_\varepsilon(\mu) \in \mathbb{R}
$$

is non-decreasing, continuous and concave.

Before proving the above proposition, let us remark that, together with Corollary 2.3, it yields an instant proof of our main result above.

**Proof of Proposition 2.5.** (i) We establish a slightly stronger property. Take a sequence $\varepsilon_n \to \varepsilon \geq 0$ and $(\mu^n)_{n \geq 1} \subseteq \Pi^\varepsilon$ with limit $\mu$, and $\pi_n \in M_\varepsilon(\mu^n)$ such that

$$
\limsup_{n \to \infty} P_\varepsilon(\mu^n) = \limsup_{n \to \infty} E_{\pi_n}[c(X)].
$$

Without loss of generality, we may assume that $\limsup_{n \to \infty} E_{\pi_n}[c] = \lim_{n \to \infty} E_{\pi_n}[c]$, and further by Lemma 2.6 that $(\pi_n)_{n \geq 1}$ admits a convergent subsequence, denoted again by itself, with some limit $\pi \in \Pi(\mu)$. For every $k = 1, \cdots , N-1$ and $h \in L^\infty(\mathcal{X}^k, \mathcal{B}(\mathcal{X}^k); \mathbb{R}^d)$, one has

$$
E_{\pi_n}[h(X_1, \cdots, X_k) \cdot (X_{k+1} - X_k)] \leq \varepsilon_n \|h\|_\infty,
$$

which implies $\pi \in M_\varepsilon(\mu)$ by the dominated convergence theorem. Combining the continuity of $c$, it follows from the dominated convergence theorem that

$$
\limsup_{n \to \infty} P_\varepsilon(\mu^n) = \lim_{n \to \infty} E_{\pi_n}[c] = E_\pi[c] \leq P_\varepsilon(\mu).
$$

(ii) First notice that $\varepsilon \mapsto P_\varepsilon(\mu)$ is non-decreasing by definition. Next, let us prove the concavity. Given $\varepsilon, \varepsilon' \in \mathbb{R}_+$ and $\alpha \in [0, 1]$, it remains to show

$$(1 - \alpha)P_\varepsilon(\mu) + \alpha P_{\varepsilon'}(\mu) \leq P_{\varepsilon_\alpha}(\mu), \quad \text{where } \varepsilon_\alpha := (1 - \alpha)\varepsilon + \alpha \varepsilon'.
$$

This indeed follows from the fact that $(1 - \alpha)\pi + \alpha \pi' \in M_\varepsilon(\mu)$ for all $\pi \in M_\varepsilon(\mu)$ and $\pi' \in M_{\varepsilon'}(\mu)$. Hence the map restricted to $(0, +\infty)$ is continuous. Finally, the reasoning in (i) above, with $\mu^n = \mu$ and $\varepsilon_n \to 0$, gives $\lim_{n \to \infty} P_{\varepsilon_n}(\mu) \leq P(\mu)$ which combined with the obvious reverse inequality gives right continuity at $\varepsilon = 0$. 

**Lemma 2.6.** Let $(\mu^n)_{n \geq 1} \subseteq \Pi^N$ be a sequence converging to $\mu$ under $\mathcal{W}^\oplus$, and $\pi_n \in \Pi(\mu^n)$. Then there exists a weakly convergent subsequence $(\pi_{n_k})_{k \geq 1}$, and its limit $\pi$ belongs to $\Pi(\mu)$.
2.2 Discretisation of marginal distributions

In this subsection, we introduce several discretisations \( \mu^n = (\mu^n_1, \ldots, \mu^n_N) \) which approximates the original target measures \( \mu \). Since every \( \mu^n \) is supported on a countable or finite set, the computation of \( \mathcal{P}_{d_n}(\mu^n) \) turns to be the resolution of an LP problem, infinite– or finite– dimensional. This allows us to use classical techniques of LP to analyse the approximating problem.

2.2.1 A general discretisation scheme

We detail here a discretisation procedure which is valid in a general setting. For each \( n \geq 1 \), we take a fixed \( \Delta_n > 0 \), and define a sequence of subspaces. Set

\[
\mathcal{X}^{(n)} := \{ q\Delta_n : q \equiv (q_1, \ldots, q_d) \in \mathbb{Z}^d \}.
\]

We define also a vector of probability measures \( \mu^n \) supported on \( \mathcal{X}^{(n)} \) by,

\[
\mu^n_k[\{ q\Delta_n \}] := \mu_k[\{ x = (x_1, \ldots, x_d) \in \mathcal{X} : [x_i/\Delta_n] = q_i \text{ for } i = 1, \ldots, d \}],
\]

where for \( a \in \mathbb{R} \), \( \lfloor a \rfloor \in \mathbb{Z} \) is the largest integer less or equal to \( a \). For every \( \psi \in \Lambda \), define \( \hat{\psi}^{(n)} : \mathcal{X} \rightarrow \mathbb{R} \) by

\[
\hat{\psi}^{(n)}(x) := \psi(\Delta_n [x/\Delta_n]).
\]

Using the definitions of \( \mu^n_k \) and \( \hat{\psi}^{(n)} \), we directly calculate that

\[
\int \hat{\psi}^{(n)} d\mu_k = \sum_{q \in \mathbb{Z}^d} \hat{\psi}^{(n)}(q\Delta_n) \mu^n_k[\{ x \in \mathcal{X} : [x_i/\Delta_n] = q_i \text{ for } i = 1, \ldots, d \}],
\]

\[
= \sum_{q \in \mathbb{Z}^d} \hat{\psi}^{(n)}(q\Delta_n) \mu^n_k[\{ q\Delta_n \}] = \int \psi d\mu^n_k,
\]

which implies in view of (8) that

\[
\mathcal{W}(\mu^n_k, \mu_k) := \sup_{\psi \in \Lambda_1} \left| \int \psi d\mu^n_k - \int \psi d\mu_k \right| \leq \sup_{\psi \in \Lambda_1} \int |\hat{\psi}^{(n)} - \psi| d\mu_k \leq \sqrt{d}\Delta_n.
\]
In general, it follows by Corollary 2.3 that, for any sequence $(\Delta_n)_{n \geq 1}$ converging to zero, one has
\[
\lim_{n \to \infty} P_{\sqrt{n}\Delta_n}(\mu^n) = P(\mu).
\]
In general, $P_{\sqrt{n}\Delta_n}(\mu^n)$ corresponds to an infinite-dimensional LP problem. Before proceeding as above, we may first truncate $\mu_k$ for $k = 1, \cdots, N$. Take an arbitrary $R > 0$, and consider a measure $\pi \in M(\mu)$. Set
\[
\mu_k^{(R)} := \pi \circ (X_k^R)^{-1}, \quad \text{where } X_k^R := X_k 1_{\{|X_k| \leq R\}}.
\]
Clearly $\mu_k^{(R)}$ has a bounded support, and one has as $R \to +\infty$ that
\[
\mathcal{W}(\mu_k^{(R)}, \mu_k) \leq \mathbb{E}_{\pi}[|X_k^R - X_k|] = \int \mathbb{1}_{|x| > R} \mu_k(dx) \to 0.
\]
Let $\mu^{(R)} := (\mu_k^{(R)})_{1 \leq k \leq N}$ and $\rho_R := \mathcal{W}(\mu^{(R)}, \mu)$, then one obtains that
\[
\lim_{R \to +\infty} P_{\rho_R}(\mu^{(R)}) = P(\mu).
\]
Notice that, in general $\mu^n$ and $\mu^{(R)}$ may no longer belong to $\Pi^\leq$, even if $\mu \in \Pi^\leq$. However, when $d = 1$, explicit discretisations which preserve the increasing convex order are well known, see Section 3.2. More generally, in a recent parallel work, Alfonsi, Corbetta & Jourdin [1] investigate methods of constructing $\mu^n$ such that $\mu^n \in \Pi^\leq$. In the rest of Section 2.2, we provide some examples in the case of $N = 2$, where we may find more specialised discretisations. Recall that, for the sake of simplicity, we write $\mu \equiv \mu_1, \nu \equiv \mu_2, X \equiv X_1$ and $Y \equiv X_2$. Further, for any discretised $(\mu^n, \nu^n)$, we write $d_n \equiv \mathcal{W}(\mu^n, \nu^n, (\mu, \nu))$.

### 2.2.2 Some specific discretisation schemes

We first focus on the case $d = 1$ and assume that $\mu$ and $\nu$ have bounded supports. Without loss of generality, let $\text{supp}(\mu) \subset [-1, 1]$ and $\text{supp}(\nu) \subset [-2, 2)$. Using the discretisation given in Section 2.2.1, taking $\Delta_n = 1/n$ and setting further $\alpha^n_k := \mu^n [\{k/n\}]$ and $\beta^n_k := \nu^n [\{k/n\}]$, one has $\alpha^n_k = \mu(\{k/n, (k+1)/n\})$ and $\beta^n_k = \nu(\{k/n, (k+1)/n\})$ for $k \in \mathbb{Z}$. Then $d_n \leq 1/n$, and the LP problem $P_{1/n}(\mu^n; \nu^n)$ may be written as

\[
\begin{align*}
\sup_{(p_{i,j})} & \sum_{-n \leq i < n, -2n \leq j < 2n} \sum_{-n \leq i < n, -2n \leq j < 2n} p_{i,j} c(i/n, j/n) \\
\text{s.t.} & \sum_{-2n \leq j < 2n} p_{k,j} = \alpha^n_k, \quad \text{for } k = -n, \cdots, n - 1, \\
& \sum_{-n \leq i < n} p_{i,k} = \beta^n_k, \quad \text{for } k = -2n, \cdots, 2n - 1, \\
& \sum_{-2n \leq j < 2n} p_{k,j} / n \geq (k - 1) \alpha^n_k / n, \quad \text{for } k = -n, \cdots, n - 1, \\
& \sum_{-2n \leq j < 2n} p_{k,j} / n \leq (k + 1) \alpha^n_k / n, \quad \text{for } k = -n, \cdots, n - 1.
\end{align*}
\]
Example 2.7. We start with an example from as follows. Let $\mu, \nu$ be the uniform distributions given by

$$
\mu(dx) = \frac{1}{2}1_{\{x \in [-1,1]\}}dx \quad \text{and} \quad \nu(dy) = \frac{1}{4}1_{\{y \in [-2,2]\}}dy.
$$

It follows that the parameters are given by

$$
\alpha_k^n = \frac{1}{2n}, \quad \text{for} \quad k = -n, \cdots, n-1,
$$

$$
\beta_k^n = \frac{1}{4n}, \quad \text{for} \quad k = -2n, \cdots, 2n-1.
$$

Taking $c(x, y) = |x - y|$, it follows from Hobson & Neuberger [27] that, there exist functions $\xi_\pm(x) = x \pm 1$ such that the corresponding optimisers $\pi^*$ may be written as

$$
\pi^*(dx, dy) = \mu(dx) \otimes \left\{ \frac{1}{2} \delta_{\xi_+(x)}(dy) + \frac{1}{2} \delta_{\xi_-(x)}(dy) \right\}.
$$

A straightforward computation yields that $P(\mu, \nu) = 1$. Solving the corresponding LP problem, we recover the theoretical value as well as the correct form of the optimal transport plan, see Figure 1.

![Figure 1: Computations for Example 2.7](image)

The left pane show the optimal transport with $n = 50$ and the right pane shows the convergence of the value with $n$.

Example 2.8. We keep $c(x, y) = |x - y|$ but consider $\mu, \nu$ with unbounded support. Specifically, consider Gaussian marginals $\mu = \mathcal{N}(0,1)$ and $\nu = \mathcal{N}(0,2)$. Clearly, $\alpha_k^n$ and $\beta_k^n$ do not have closed expressions, which makes the previous discretisation costly. Accordingly, we approximate $\mu$ (resp. $\nu$) by a simple random walk. We consider a sequence of i.i.d. random variables $(Z_n)_{n \geq 1}$ with $P[Z_n = 1] = P[Z_n = -1] = 1/2$. Then, it follows by central limit theorem that, as $n \to \infty$

$$
\frac{\sum_{i=1}^n Z_i}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mu \quad \text{and} \quad \frac{\sum_{i=1}^n Z_i}{\sqrt{n/2}} \xrightarrow{\mathcal{L}} \nu.
$$

Let $\mu^n$ and $\nu^n$ be respectively the law of $(\sum_{i=1}^n Z_i)/\sqrt{n}$ and $(\sum_{i=1}^n Z_i)/\sqrt{n/2}$, then a straightforward computation yields

$$
\int_{\mathcal{X}} x^2 \mu^n(dx) = \int_{\mathcal{X}} x^2 \mu(dx) \quad \text{and} \quad \int_{\mathcal{X}} y^2 \nu^n(dy) = \int_{\mathcal{X}} y^2 \nu(dy),
$$

11
which implies \( \lim_{n \to \infty} \mathcal{W}^\Pi((\mu^n, \nu^n), (\mu, \nu)) = 0 \). In view of the dual formulation (8) and Stein’s method in [2], one has further

\[
\mathcal{W}(\mu^n, \mu) \leq \frac{5 \mathbb{E}^\mathbb{P}[|Z|^3]}{\sqrt{n}} = \frac{5}{\sqrt{n}}
\]

Similarly one obtains \( \mathcal{W}(\nu^n, \nu) \leq 5\sqrt{2}/\sqrt{n} \), and further \( d_n \leq 5/\sqrt{n} + 5\sqrt{2}/\sqrt{n} \). Therefore

\[
\lim_{n \to \infty} P_{5/\sqrt{n} + 5\sqrt{2}/\sqrt{n}}(\mu^n, \nu^n) = P(\mu, \nu).
\]

The results of our LP solver are presented in Figures 3 and 4. We note that the optimiser distributes mass from \( x \) to \( \{\xi_+(x), \eta_+(x)\} \) for two increasing functions \( \xi_+ \) and \( \eta_+ \), which agrees with the theoretical results in Hobson & Neuberger [27].

**Remark 2.9.** With the distributions \( \mu \) and \( \mu^n \) of Example 2.8, for any function \( f \in \Lambda \), one has

\[
\mathcal{W}(\mu^n \circ f^{-1}, \mu \circ f^{-1}) = \sup_{\psi \in \Lambda_1} \left\{ \int \psi \circ f d\mu^n - \int \psi \circ f d\mu \right\} \leq \text{Lip}(f) \mathcal{W}(\mu^n, \mu),
\]

where \( \mu^n \circ f^{-1} \) and \( \mu \circ f^{-1} \) denote the pushforward measures of \( \mu \) and \( \mu^n \). In particular, for every pair \((\mu \circ f^{-1}, \nu \circ g^{-1}) \in \Pi^\geq \) with \( f, g \in \Lambda \), one has \( \mathcal{W}^\Pi((\mu^n \circ f^{-1}, \nu^n \circ g^{-1}), (\mu \circ f^{-1}, \nu \circ g^{-1})) \leq \max(\text{Lip}(f), \text{Lip}(g)) d_n \) and

\[
\lim_{n \to \infty} P_{d_n}(\mu^n \circ f^{-1}, \nu^n \circ g^{-1}) = P(\mu \circ f^{-1}, \nu \circ g^{-1}).
\]

**Example 2.10.** Consider now log-normal marginals, which are image measures of Gaussian \( \mu, \nu \) from Example 2.8 but under non Lipschitz transformations. Let \( f(x) = e^{x-1/2} \) and \( g(x) = e^{x-2} \). Then \( \mu \circ f^{-1} \) and \( \nu \circ g^{-1} \) are two log-normal distributions increasing in convex order. Let us estimate next \( \mathcal{W}^\Pi((\mu^n \circ f^{-1}, \nu^n \circ g^{-1}), (\mu \circ f^{-1}, \nu \circ g^{-1})) \). Without loss of generality, we only treat \( \mathcal{W}(\mu^n \circ f^{-1}, \mu \circ f^{-1}) \). Let \( \pi^n \) be the optimal transport plan realising \( \mathcal{W}(\mu^n, \mu) \), i.e. \( \pi^n \in \Pi(\mu^n, \mu) \) and \( \mathcal{W}(\mu^n, \mu) = \mathbb{E}_{\pi^n}[|X - Y|] \). Then it follows by definition that \( \mathcal{W}(\mu^n \circ f^{-1}, \mu \circ f^{-1}) \leq \)

---

**Figure 2:** Computations for Example 2.8. The left pane show the optimal transport with \( n = 150 \) and the right pane shows the convergence of the value with \( n \).
$\mathbb{E}_{\pi^n}[|f(X) - f(Y)|]$. Taking an arbitrary $R > 0$, rewrite the expectation on the right as

$$
\mathbb{E}_{\pi^n}[|f(X) - f(Y)|] \leq \mathbb{E}_{\pi^n}[|f(X) - f(Y)|1_{\{|X| > R\}}] + \mathbb{E}_{\pi^n}[|f(X) - f(Y)|1_{\{|Y| > R\}}] + \mathbb{E}_{\pi^n}[|f(X) - f(Y)|1_{\{|X| \leq R, |Y| \leq R\}}].
$$

It follows by Hölder’s inequality that

$$
\mathbb{E}_{\pi^n}[|f(X) - f(Y)|1_{\{|X| > R\}}] \leq \left(\mathbb{E}_{\pi^n}[|f(X)|^2]^{1/2} + \mathbb{E}_{\pi^n}[|f(Y)|^2]^{1/2}\right) \mathbb{E}_{\pi^n}[|X| > R]^{1/2}
$$

and similarly

$$
\mathbb{E}_{\pi^n}[|f(X) - f(Y)|1_{\{|Y| > R\}}] \leq e^{-1/2} \left(\frac{e^{2/\sqrt{n}} + e^{-2/\sqrt{n}}}{2}\right)^{n/2} + e \pi^n[|X| > R]^{1/2}.
$$

As for the last term, we have further

$$
\mathbb{E}_{\pi^n}[|f(X) - f(Y)|1_{\{|X| \leq R, |Y| \leq R\}}] \leq \max_{x \in [-R, R]} |f'(x)| \times \mathbb{E}_{\pi^n}[|X - Y|]
$$

$$
= \max_{x \in [-R, R]} |f'(x)| \times \mathcal{W}(\mu^n, \mu)
$$

$$
\leq \frac{5e^{R-1/2}}{\sqrt{n}}.
$$

One obtains finally

$$
\mathbb{E}_{\pi^n}[|f(X) - f(Y)|] \leq 2e^{-1/2} \left(\frac{e^{2/\sqrt{n}} + e^{-2/\sqrt{n}}}{2}\right)^{n/2} + e \left(\frac{5}{\sqrt{n}} + \pi^n[|Y| > R - 1]\right)^{1/2}
$$

$$
+ \frac{5e^{R-1/2}}{\sqrt{n}}
$$

$$
\leq 8e^{1/2} \left(\frac{5}{\sqrt{n}} + \pi^n[|Y| > R - 1]\right)^{1/2} + \frac{5e^{R-1/2}}{\sqrt{n}}
$$

$$
\leq 8e^{1/2} \left(\frac{5}{\sqrt{n}} + \sqrt{\frac{2}{\pi}} e^{-(R-1)^2/2}\right)^{1/2} + \frac{5e^{R-1/2}}{\sqrt{n}}.
$$

Optimising the above expression w.r.t. $R$, there exists some constant $C$ such that

$$
\mathcal{W}^\mathcal{E}(\mu^n \circ f^{-1}, \nu^n \circ g^{-1}, (\mu \circ f^{-1}, \nu \circ g^{-1})) \leq C \sqrt{\frac{\log n}{n}}.
$$

**Example 2.11.** Our last example considers a two-dimensional setting. Denote, for any $R > 0$, by $B_R$ the the disc with radius $R$, and by $\partial B_R$ its circle. Let $\mu$ and $\nu$ be the uniform probability distributions supported respectively on $B_1$ and $\partial B_2$.

We consider the distance cost $c(x, y) = -|x - y|$. Then it follows from Theorem 1.2 in Lim [30] that the optimiser is given by

$$
\pi^*(\mu, \nu) = \mu(dx_1, dx_2) \otimes \left\{ \frac{2 + R}{4} \delta_{2x_1, 2x_2/2}(dy_1, dy_2) + \frac{2 - R}{4} \delta_{-2x_1, -2x_2/2}(dy_1, dy_2) \right\},
$$

13
where \( r \equiv r(x_1, x_2) := \sqrt{x_1^2 + x_2^2} \). This yields further

\[
P(\mu, \nu) = -\mathcal{E}_{\pi^*}(\mu, \nu)[|X - Y|] = -\int_{B_1} \frac{4 - r^2}{2} dx_1 dx_2 = -\frac{7}{4}\pi.
\]

We adapt in this example another discretisation. Let us represent the points on the plane by means of the polar coordinate system. Hence

\[
B_1 = \left\{ (r \cos \theta, r \sin \theta) : r \in (0, 1] \text{ and } \theta \in (0, 2\pi) \right\},
\]

\[
\partial B_2 = \left\{ (2 \cos \theta, 2 \sin \theta) : \theta \in (0, 2\pi) \right\}.
\]

Define for \( 1 \leq i, j, k \leq n \)

\[
\alpha_{i,j}^n := \mu \left\{ (r \cos \theta, r \sin \theta) : r \in \left( \frac{i-1}{n}, \frac{i}{n} \right] \text{ and } \theta \in \left( \frac{2(j-1)\pi}{n}, \frac{2j\pi}{n} \right] \right\} = \frac{2i-1}{n^3},
\]

\[
\beta_k^n := \nu \left\{ (2 \cos \theta, 2 \sin \theta) : \theta \in \left( \frac{2(k-1)\pi}{n}, \frac{2k\pi}{n} \right) \right\} = \frac{1}{n}.
\]

We consider the probability measures \( \mu^n \) and \( \nu^n \) given by

\[
\mu^n \left\{ \left( \frac{i}{n} \cos \frac{2j\pi}{n}, \frac{i}{n} \sin \frac{2j\pi}{n} \right) \right\} = \alpha_{i,j}^n,
\]

\[
\nu^n \left\{ \left( \frac{2k\pi}{n}, \frac{2k\pi}{n} \right) \right\} = \beta_k^n.
\]

Then a straightforward computation yields \( W^{(\pm)}((\mu^n, \nu^n), (\mu, \nu)) \leq 4\sqrt{2}/n \). Therefore, the LP problem \( P_{4\sqrt{2}/n}(\mu^n, \nu^n) \) could be written as

\[
\sup_{(p_{i,j,k})_{1 \leq i,j,k \leq n} \in \mathbb{R}^{n^3}_+} \sum_{1 \leq i,j,k \leq n} p_{i,j,k} \left( \left( \frac{i}{n} \cos \frac{2j\pi}{n}, \frac{i}{n} \sin \frac{2j\pi}{n} \right), \left( \frac{2k\pi}{n}, \frac{2k\pi}{n} \right) \right)
\]

s.t.

\[
\sum_{1 \leq k \leq n} p_{i,j,k} = \alpha_{i,j}^n, \quad \text{for } i, j = 1, \ldots, n,
\]

\[
\sum_{1 \leq i,j \leq n} p_{i,j,k} = \beta_k^n, \quad \text{for } k = 1, \ldots, n,
\]

\[
\left| \sum_{1 \leq k \leq n} p_{i,j,k} \frac{2k\pi}{n} - \alpha_{i,j}^n \frac{i}{2n} \cos \frac{2j\pi}{n} \right| \leq \frac{2\sqrt{2}}{n}, \quad \text{for } i, j = 1, \ldots, n,
\]

\[
\left| \sum_{1 \leq k \leq n} p_{i,j,k} \frac{2k\pi}{n} - \alpha_{i,j}^n \frac{i}{2n} \sin \frac{2j\pi}{n} \right| \leq \frac{2\sqrt{2}}{n}, \quad \text{for } i, j = 1, \ldots, n.
\]

## 3 Analysis of convergence rate

In this section, we provide a convergence rate in the case of \( d = 1 \) and \( N = 2 \). This requires the duality for the relaxed optimisation problem \([5]\) which we establish, in a general setting.
3.1 Convergence rates for a general discretisation scheme

Recall that, we still write \( \mu \equiv \mu_1, \nu \equiv \mu_2, X \equiv X_1 \) and \( Y \equiv X_2 \). Corollary 2.3 with \( \rho_n := \mathcal{W}_2^\mathcal{D}((\mu^n, \nu^n), (\mu, \nu)) \), gives

\[
|P_{\rho_n}(\mu^n, \nu^n) - P(\mu, \nu)| \leq (P_{2\rho_n}(\mu, \nu) - P(\mu, \nu)) + 2\text{Lip}(c)\rho_n.
\]

It follows that to estimate the convergence rate in Theorem 2.1, we have to understand the asymptotic behaviour of \( P_\varepsilon(\mu, \nu) - P(\mu, \nu) \) as \( \varepsilon \searrow 0 \). To emphasise the dependency on \( c \), we write \( P_\varepsilon(\mu, \nu) \equiv P_\varepsilon^c(\mu, \nu) \). Clearly, for any \( c_1 \) and \( c_2 \) one has

\[
P_{\varepsilon}^{c_1+c_2}(\mu, \nu) \leq P_{\varepsilon}^{c_1}(\mu, \nu) + P_{\varepsilon}^{c_2}(\mu, \nu).
\]

Then we have the following result.

**Theorem 3.1.** Let \( c \in C^2(\mathbb{R}^2) \) satisfy Lip\((c) \leq L \) and \( \sup_{(x,y) \in \mathbb{R}^2} |\partial^2_{xy} c(x,y)| \leq L \) for some \( L > 0 \). If \( \nu \) has a finite second moment, then one has

\[
|P^c_{\rho_n}(\mu^n, \nu^n) - P^c(\mu, \nu)| \leq 2L \left( \inf_{R \in \mathbb{R}_+} \lambda_n(R) + 10\rho_n \right)
\]

where

\[
\lambda_n(R) := 9R\rho_n + \int_{(R,\infty)} (y - R)^2 \nu(dy) + \int_{(-\infty,-R)} (y + R)^2 \nu(dy).
\]

**Proof.** Let \( c_L(x,y) := c(x,y) - Ly^2/2 \). Then the map \( y \mapsto c_L(x,y) \) is concave for all \( x \in \mathbb{R} \). Define further, for every \( R \in \mathbb{R}_+ \),

\[
c_L^R(x,y) := \begin{cases} c_L(x,-R) + (y+R)\partial_y c_L(x,-R) & \text{if } y \in (-\infty,-R] \\ c_L(x,y) & \text{if } y \in (-R,R] \\ c_L(x,R) + (y-R)\partial_y c_L(x,R) & \text{if } y \in (R,\infty), \end{cases}
\]

Then \( y \mapsto c_L^R \) is still concave and \( c_L^R \) is Lipschitz with Lip\((c_L^R) \leq L_R := L(R+1) \). It follows from Theorem 2.4 and Remark 2.5 in Beiglböck, Lim & Oblój [1] that, there exists an optimiser \((\varphi^*, \psi^*, h^*) \in D\) for the dual problem \( D_L^c(\mu, \nu) \) such that \( \varphi^*, \psi^* \in \Lambda_{10L_R} \) and \( ||h^*||_\infty \leq 9L_R \). Set \( \varepsilon = 2dn \), then one has

\[
P_{\varepsilon}^{c_L^R}(\mu, \nu) - P_{\varepsilon}^{c_L}(\mu, \nu) = D_{\varepsilon}^{c_L^R}(\mu, \nu) - D_{\varepsilon}^{c_L}(\mu, \nu)
\]

\[
= D_{\varepsilon}^{c_L^R}(\mu, \nu) - \left[ \int \varphi^* d\mu + \int \psi^* d\nu \right]
\]

\[
\leq \left[ \int (\varphi^* + \varepsilon|h^*|) d\mu + \int \psi^* d\nu \right] - \left[ \int \varphi^* d\mu + \int \psi^* d\nu \right]
\]

\[
= \varepsilon \int |h^*| d\mu
\]

\[
\leq 9\varepsilon L_R.
\]

In addition, a straightforward computation yields \( |c_L(x,y) - c_L^R(x,y)| \leq 1_{\{y>R\}} L(y-R)^2 + 1_{\{y<-R\}} L(y+R)^2 \), which implies that

\[
|P_{\varepsilon}^{c_L^R}(\mu, \nu) - P_{\varepsilon}^{c_L}(\mu, \nu)| \leq P_{\varepsilon}^{c_L^R - c_L}(\mu, \nu)
\]

\[
\leq L \left( \int_{(R,\infty)} (y-R)^2 \nu(dy) + \int_{(-\infty,-R)} (y+R)^2 \nu(dy) \right).
\]
Hence,

\[
|P^r_\varepsilon(\mu, \nu) - P^c(\mu, \nu)| = |P^c_\varepsilon(\mu, \nu) - P^c(\mu, \nu)|
\]

\[
\leq |P^c_\varepsilon(\mu, \nu) - P^c_R(\mu, \nu)| + |P^c_\varepsilon(\mu, \nu) - P^c_L(\mu, \nu)| + |P^c_R(\mu, \nu) - P^c_L(\mu, \nu)|
\]

\[
\leq 9 \varepsilon L (R + 1) + 2L \left( \int_{[R, +\infty)} (y - R)^2 \nu(dy) + \int_{(-\infty, -R)} (y + R)^2 \nu(dy) \right),
\]

which concludes the proof by taking \( \varepsilon = 2\rho_n \).

\[\square\]

**Remark 3.2.** (i) If \( \nu \) has bounded support, then Theorem 3.1 holds whenever \( \varepsilon \in C(\mathbb{R}^2) \).

(ii) We may extend the above analysis to more general functions \( \varepsilon \). Assume that \( \varepsilon \in C^2(\mathbb{R}^2) \) is with linear growth, i.e. \( |c(x, y)| \leq L(1 + |x| + |y|) \) for some \( L > 0 \), then for every \( \varepsilon \geq 1 \), there exists a function \( \varepsilon \in C^2(\mathbb{R}^2) \) such that

\[
c_R(x, y) = \begin{cases} 
  c(x, y) & \text{if } |x| \leq R \text{ and } |y| \leq R \\
  0 & \text{if } |x| > R + 1 \text{ or } |y| > R + 1
\end{cases}
\]

and

\[
\|c_R\|_\infty \leq \sup_{(x,y) \in [-R,R] \times [-R,R]} |c(x, y)| \leq L(1 + 2R).
\]

Then clearly, \( c_R \) satisfies the conditions of Theorem 3.1 as its support is compact, and moreover,

\[
|c(x, y) - c_R(x, y)| \leq \frac{1_{\{\max(|x|, |y|) > R\} \cdot 2L(1 + |x| + |y|)}}{R},
\]

\[
\leq \frac{\max(|x|, |y|)}{R} \cdot 4L(|x| + |y|),
\]

\[
\leq \frac{8L(|x|^2 + |y|^2)}{R},
\]

which implies that

\[
|P^r_\varepsilon(\mu, \nu) - P^c_R(\mu, \nu)| \leq \frac{8L}{R} \left( \int_{\mathbb{R}} |x|^2 \mu(dx) + \int_{\mathbb{R}} |y|^2 \nu(dy) \right) =: \frac{L'}{R}.
\]

We obtain finally that

\[
|P^r_\varepsilon(\mu, \nu) - P^c(\mu, \nu)|
\]

\[
\leq |P^r_\varepsilon(\mu, \nu) - P^c_R(\mu, \nu)| + |P^c_\varepsilon(\mu, \nu) - P^c_R(\mu, \nu)| + |P^c_R(\mu, \nu) - P^c(\mu, \nu)|
\]

\[
\leq |P^c_\varepsilon(\mu, \nu) - P^c_R(\mu, \nu)| + 2L'/R.
\]

For a fixed \( R \), using Theorem 3.1 for the first term, we deduce a bound on the convergence rate for \( \varepsilon \). The result can then be optimised over \( R \).
3.2 Convergence rates for a specific discretisation scheme

In this section, we explore a tailored discretisation in dimension one which preserves the convex order. Specifically, given \((\mu, \nu) \in \Pi^\leq\), define two sequence of measures supported on \(\{k/n\}_{k \in \mathbb{Z}}\) as follows:

\[
\mu^n[\{k/n\}] := \int_{(k-1)/n, k/n} (nx + 1 - k)\mu(dx) + \int_{k/n, (k+1)/n} (1 + k - nx)\mu(dx) \quad (9)
\]

\[
\nu^n[\{k/n\}] := \int_{(k-1)/n, k/n} (nx + 1 - k)\nu(dx) + \int_{k/n, (k+1)/n} (1 + k - nx)\nu(dx). \quad (10)
\]

Note also that 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 agreeing with that of \(\mu\) in those points:

\[
\int_{\mathbb{R}} \left| \frac{k}{n} - x \right| \mu(dx) = \int_{\mathbb{R}} \left| \frac{k}{n} - x \right| \mu^n(dx), \quad \text{for all } k \in \mathbb{Z}.
\]

Then we have the following result.

**Theorem 3.3.** Let the conditions of Theorem 3.1 hold. Then \((\mu^n, \nu^n) \in \Pi^\leq\) for all \(n \geq 1\), and

\[
|P^c(\mu^n, \nu^n) - P^c(\mu, \nu)| \leq 2L \left( \inf_{R \in \mathbb{R}^+} \lambda_n(R) + \frac{11}{n} \right),
\]

where

\[
\lambda_n(R) := \frac{10R}{n} + \left( \int_{(R, \infty)} (y - R)^2\nu(dy) + \int_{(-\infty, -R)} (y + R)^2\nu(dy) \right).
\]

In particular, if \(\nu\) has a bounded support, then the convergence is \(O(1/n)\).

**Remark 3.4.** Let \(\nu\) have a bounded support \(J\). For any continuous function \(c\) and any \(\varepsilon > 0\), there exists \(c_\varepsilon \in C^2(\mathbb{R}^2)\) such that

\[
\sup_{(x, y) \in J \times J} |c(x, y) - c_\varepsilon(x, y)| \leq \varepsilon.
\]

Hence, with a slight modification, the above theorem applies for the class of continuous functions on \(\mathbb{R}^2\).

The rest of Section 3.2 is devoted to the proof of Theorem 3.3. We start with Proposition 3.5.

**Proposition 3.5.** Let \((\mu^n, \nu^n)_{n \geq 1}\) be defined by (9) and (10) all \(n \geq 1\). Then \((\mu^n, \nu^n), (\mu, \mu^n), (\nu, \nu^n) \in \Pi^\leq\) and \(W^\leq((\mu^n, \nu^n), (\mu, \nu)) \leq 2/n\) for all \(n \geq 1\).

**Proof.** For any measurable function \(\psi : \mathbb{R} \to \mathbb{R}\), define \(\hat{\psi}_n : \mathbb{R} \to \mathbb{R}\) by

\[
\hat{\psi}_n(x) := (1 + |nx| - nx)\psi((nx)/n) + (nx - |nx|)\psi((1 + |nx|)/n),
\]

Then it follows from a straightforward computation that, see also Dolinsky & Soner [19], one has

\[
\int \psi d\mu^n = \int \hat{\psi}_n d\mu \quad \text{and} \quad \int \psi d\nu^n = \int \hat{\psi}_n d\nu.
\]
Take $\psi \equiv 1$, then $\hat{\psi}_n \equiv 1$, and further $\mu^n$ and $\nu^n$ are well defined probability measures. Moreover, taking $\psi(x) = |x|$, it is clear that $\hat{\psi}_n = \psi$ and thus

\[
\int \psi d\mu^n = \int \hat{\psi}_n d\mu < +\infty,
\]

\[
\int \psi d\nu^n = \int \hat{\psi}_n d\nu < +\infty.
\]

To prove $(\mu^n, \nu^n), (\mu, \mu^n), (\nu, \nu^n) \in \Pi^\leq$, it suffices to test for $\psi(x) = (x - K)^+$. It follows easily that $\hat{\psi}_n$ is convex and $\hat{\psi}_n \geq \psi$ by computation. This implies that $(\mu^n, \nu^n), (\mu, \mu^n), (\nu, \nu^n) \in \Pi^\leq$. To end the proof, we need the dual formulation of the Wasserstein metric. For each $\psi \in \Lambda_1$, one has

\[
\left| \int \psi d\mu^n - \int \psi d\mu \right| \leq \int |\hat{\psi}_n - \psi| d\mu \leq 1/n.
\]

Hence it follows from $\mathfrak{S}$ that $\mathcal{W}(\mu^n, \mu) \leq 1/n$, which concludes the proof. \hfill $\square$

**Proof of Theorem 3.3.** First, it follows from Proposition 3.5 that $(\mu^n, \nu^n) \in \Pi^\leq$ for all $n \geq 1$. Next, one has

\[
|P^{c}(\mu^n, \nu^n) - P^{c}(\mu, \nu)| = |P^{cl}(\mu, \nu) - P^{cl}(\mu, \nu)| \leq |P^{cl}(\mu^n, \nu^n) - P^{cl}(\mu^n, \nu^n)| + |P^{cl}(\mu^n, \nu^n) - P^{cl}(\mu, \nu)| + |P^{cl}(\mu, \nu) - P^{cl}(\mu, \nu)|
\]

As for the first and third terms, it follows by the same arguments in the proof of Theorem $3.3$ that

\[
|P^{cl}(\mu^n, \nu^n) - P^{cl}(\mu^n, \nu^n)| \leq L \left( \int_{(R, +\infty)} (y - R)^2 \nu^n(dy) + \int_{(-\infty, -R)} (y + R)^2 \nu^n(dy) \right),
\]

\[
|P^{cl}(\mu, \nu) - P^{cl}(\mu, \nu)| \leq L \left( \int_{(R, +\infty)} (y - R)^2 \nu(dy) + \int_{(-\infty, -R)} (y + R)^2 \nu(dy) \right).
\]

Combining with the construction of $\nu^n$, we obtain that

\[
|P^{cl}(\mu^n, \nu^n) - P^{cl}(\mu^n, \nu^n)| \leq L \left( \int_{(R, +\infty)} (y - R)^2 \nu(dy) + \int_{(-\infty, -R)} (y + R)^2 \nu(dy) + \frac{1}{n^2} \right).
\]

It remains to estimate the second term. Using again Theorem 2.4 of [5], $D^R_{cl}(\mu^n, \nu^n)$ is attained by $(\varphi^n, \psi^n, h^n) \in \mathcal{D}$, where $\varphi^n, \psi^n \in \Lambda_{10L_R}$. Hence,

\[
P^{cl}(\mu, \nu) - P^{cl}(\mu^n, \nu^n) = D^{cl}(\mu, \nu) - D^{cl}(\mu^n, \nu^n)
\]

\[
\leq \left[ \int \varphi^n d\mu + \int \psi^n d\nu \right] - \left[ \int \varphi^n d\mu^n + \int \psi^n d\nu^n \right]
\]

\[
\leq \left[ \int \varphi^n d\mu - \int \varphi^n d\mu^n \right] + \left[ \int \psi^n d\nu - \int \psi^n d\nu^n \right]
\]

\[
\leq 10L_R \mathcal{W}^\leq((\mu, \nu), (\mu^n, \nu^n)).
\]

Interchanging $(\mu, \nu)$ and $(\mu^n, \nu^n)$ and repeating the above reasoning, one has

\[
|P^{cl}(\mu, \nu) - P^{cl}(\mu^n, \nu^n)| \leq 10L_R \mathcal{W}^\leq((\mu, \nu), (\mu', \nu')) \leq \frac{20L(R + 1)}{n},
\]

which concludes the proof. \hfill $\square$
Example 3.6. We go back to Example 2.7, i.e. \( c(x, y) = |x - y| \) and

\[
\mu(dx) = \frac{1}{2}1_{\{x \in [-1, 1]\}}dx \quad \text{and} \quad \nu(dy) = \frac{1}{4}1_{\{y \in [-2, 2]\}}dy.
\]

One obtains by a direct calculation that

\[
\alpha_n^n = \frac{1}{4n}, \quad \alpha_k^n = \frac{1}{2n}, \quad \text{for} \quad -n < k < n, \quad \alpha_n^n = \frac{1}{4n},
\]

\[
\beta_{-2n}^n = \frac{1}{8n}, \quad \beta_k^n = \frac{1}{4n}, \quad \text{for} \quad -2n < k < 2n, \quad \beta_{2n}^n = \frac{1}{8n},
\]

which yields the LP problem \( P(\mu_n, \nu_n) \) as follows

\[
\sup_{(p_{i,j})_{-n \leq i \leq n, -2n \leq j \leq 2n \in \mathbb{R}_+^{(2n+1) \times (4n+1)}}} \sum_{-n \leq i \leq n, -2n \leq j \leq 2n} p_{i,j} c(i/n, j/n)
\]

s.t. \( \sum_{-2n \leq j \leq 2n} p_{k,j} = \alpha_k^n, \quad \text{for} \quad k = -n, \ldots, n, \)

\( \sum_{-n \leq i \leq n} p_{i,k} = \beta_k^n, \quad \text{for} \quad k = -2n, \ldots, 2n, \)

\( \sum_{-2n \leq j \leq 2n} p_{k,j} = \beta_k^n, \quad \text{for} \quad k = -n, \ldots, n. \)

Solving the LP problem, we obtain the optimal transport plan which is displayed in Figure 3 together with the error decay on a logarithmic scale.

Figure 3: Computations for Example 3.6. The left pane show the optimal transport with \( n = 50 \) and the right pane shows the decay of the error on a logarithmic scale.

Here we find easily that the plot on the right is linear, and the convergence of the order \( 1/n \).

3.3 Stability of MOT with respect to the marginals

As already recalled, the MOT problem is an abstraction of the so-called model-independent pricing problem. Specification of the marginal distributions corresponds to the knowledge of call or put prices for all strikes \( K \in \mathbb{R}_+ \). However, in practice, only finitely many options are quoted for any given maturity. In this section, we consider the optimisation problem which corresponds to such a setting
and study its convergence as strikes become dense in $\mathbb{R}_+$. Throughout this section, we assume $X = \mathbb{R}_+$.

Let $K = (K_j)_{0 \leq j \leq n}$ denote the vector of strikes with $0 = K_0 < \cdots < K_n$, and $C_i = (C_{i,j})_{0 \leq j \leq n} \in \mathbb{R}^n_+$ denote the corresponding prices of call options, for $i = 1, 2$. Define

$$M(K, C_1, C_2) := \left\{ \pi \in \Pi(\mathbb{R}^2_+) : (X, Y) \text{ is a } \pi \text{- martingale and } \mathbb{E}_\pi[(X - K_j)^+] = C_{1,j} \text{ and } \mathbb{E}_\pi[(Y - K_j)^+] = C_{2,j} \text{ for } j = 0, \cdots, n \right\}.$$ 

Let $\mathcal{A}(K)$ be the collection $(C_1, C_2)$ such that $M(K, C_1, C_2) \neq \emptyset$, then it is clear that $\mathcal{A}(K)$ is convex by definition, see Davis & Hobson [17] for a complete characterisation of $\mathcal{A}(K)$. For any $(C_1, C_2) \in \mathcal{A}(K)$, the prescribed optimisation problem is given by

$$P(K, C_1, C_2) := \sup_{\pi \in M(K, C_1, C_2)} \mathbb{E}_\pi[c(X, Y)]. \quad (11)$$

Remark 3.7. (i) For the sake of clarity, we assume that the call options with different maturities have the same set of strikes.

(ii) Similarly, we may provide the corresponding dual problem which can be interpreted as the minimal cost of super replications using the underlying assets and call options.

We study the asymptotic behaviour of the upper bound $P(K, C_1, C_2)$ with respect to the market information. We assume that there is a couple of measures $(\mu, \nu) \in \Pi^2$ which describe the true risk neutral dynamics of the stock prices. We consider sequences $(K^n, C^n_1, C^n_2)$ and, for simplicity, assume $K^n$ has exactly $n + 1$ elements: $0 = K_0^n < \cdots < K_n^n$. It will be clear from the proofs that this can be generalised as long as the analogue of the following assumption holds:

Assumption 3.8. As $n \to \infty$, $(K^n)_{n \geq 1}$ satisfies

$$\Delta K^n := \max_{1 \leq j \leq n} \left( K^n_j - K^n_{j-1} \right) \to 0 \quad \text{and} \quad K^n_n \to +\infty.$$ 

We assume the partial market information is consistent with $(\mu, \nu)$, i.e.

$$C^n_{1,j} = \int_{\mathbb{R}_+} (x - K^n_j)^+ \mu(dx) \quad \text{and} \quad C^n_{2,j} = \int_{\mathbb{R}_+} (x - K^n_j)^+ \nu(dx), \quad \text{for } j = 0, \cdots, n,$$

and note that, by definition, one has that

$$P(K^n, C^n_1, C^n_2) \geq P(\mu, \nu), \quad \text{for all } n \geq 1.$$ 

Theorem 3.9. Let the conditions of Theorem 1.3 hold. Then for any sequence $(K^n)_{n \geq 1}$ satisfying Assumption 3.8, one has

$$\lim_{n \to \infty} P(K^n, C^n_1, C^n_2) = P(\mu, \nu).$$

Proof. It suffices to prove that, for any subsequence $(K^{n_k})_{k \geq 1}$, one may find a further subsequence $(K^{n_{k_l}})_{l \geq 1}$ such that

$$\lim_{l \to \infty} P(K^{n_{k_l}}, C^{n_{k_l}}_1, C^{n_{k_l}}_2) = P(\mu, \nu).$$
It follows by definition that there exists a sequence \((\pi_{nk})_{k \geq 1}\) such that
\[
\pi_{nk} \in M(K_{n_k}^a, C_{1_k}^a, C_{2_k}^a) \quad \text{and} \quad \lim_{k \to \infty} \left( \mathbb{P}(K_{n_k}^a, C_{1_k}^a, C_{2_k}^a) - \mathbb{E}_{\pi_{nk}}[c] \right) = 0.
\]
By a straightforward computation, one has
\[
\sup_{k \geq 1} \left( \mathbb{E}_{\pi_{nk}}[X] + \mathbb{E}_{\pi_{nk}}[Y] \right) = \int_{\mathbb{R}_+} x \mu(dx) + \int_{\mathbb{R}_+} y \nu(dy) < +\infty,
\]
which implies that the sequence \((\pi_{nk})_{k \geq 1}\) is tight. Denote by \((\pi_{nk_i})_{i \geq 1}\) the convergent subsequence with limit \(\pi\), then it follows by the dominated convergence theorem that
\[
\mathbb{E}_{\pi}[Y|X] = X, \quad \pi - \text{a.s.}
\]
In addition, one has by Lemma 3.10 that
\[
\pi \circ X^{-1} = \mu \quad \text{and} \quad \pi \circ Y^{-1} = \nu,
\]
which implies that \(\pi \in M(\mu, \nu)\). The proof is fulfilled by using again the dominated convergence theorem
\[
\mathbb{P}(\mu, \nu) \leq \lim_{l \to \infty} \mathbb{P}(K_{n_k l}, C_{1_k l}^a, C_{2_k l}^a) = \lim_{l \to \infty} \mathbb{E}_{\pi_{nk_l}}[c] = \mathbb{E}_{\pi}[c] \leq \mathbb{P}(\mu, \nu).
\]

\(\Box\)

**Lemma 3.10.** Let \(\mu\) be a probability measure on \(\mathbb{R}_+\). Let \((\mu^n)_{n \geq 1}\) be a weakly convergent sequence of probability distributions such that
\[
\int_{\mathbb{R}_+} (x - K_j^n)^+ \mu^n(dx) = \int_{\mathbb{R}_+} (x - K_j^n)^+ \mu(dx), \quad \text{for } j = 0, \cdots, n,
\]
where \((K^n)_{n \geq 1}\) is a sequence satisfying Assumption 3.8 Then \(\lim_{n \to \infty} \mu^n = \mu\).

**Proof.** Set for all \(K \in \mathbb{R}_+\)
\[
f_n(K) := \int_{\mathbb{R}_+} (x - K)^+ \mu^n(dx) \quad \text{and} \quad f(K) := \int_{\mathbb{R}_+} (x - K)^+ \mu(dx).
\]
Since \(|(x - a)^+ - (x - b)^+| \leq |a - b|\), then the functions \(f_n, f \in A_1(\mathbb{R}_+)\). In addition, for every \(K \in \bigcup_{n \geq 1} K^n\), one has \(f_n(K) = f(K)\) for \(n\) large enough, which implies that \(f_n\) converges uniformly to \(f\) as \(n \to \infty\) \(K^n\) is dense in \(\mathbb{R}_+\). Let \(\psi \in C^2(\mathbb{R}_+)\) with bounded support. Then it follows by integration by parts formula that
\[
\psi(x) = \int_{\mathbb{R}_+} \psi''(K)(x - K)^+ dK - \psi'(0)x, \quad \text{for all } x \in \mathcal{X}.
\]
Hence
\[
\left| \int_{\mathbb{R}_+} \psi(x) \mu^n(dx) - \int_{\mathbb{R}_+} \psi(x) \mu(dx) \right| = \left| \int_{\mathbb{R}_+} (\psi(x) - \psi'(0)x) \mu^n(dx) - \int_{\mathbb{R}_+} (\psi(x) - \psi'(0)x) \mu(dx) \right| = \left| \int_{\mathbb{R}_+} \psi''(K) f_n(K) dK - \int_{\mathbb{R}_+} \psi''(K) f(K) dK \right|.
\]
We may conclude in view of Fubini’s Theorem. \(\Box\)
In the rest of Section 3.3, we quantify the convergence rate of \( P(K, C^n_1, C^n_2) \) to \( P(\mu, \nu) \). For technical reasons, we assume that

\[
\int_{\mathbb{R}^+} y^p \nu(dy) = V < +\infty
\]

and consider the subset

\[
M_V(K, C_1, C_2) := \{ \pi \in M(K, C_1, C_2) : \mathbb{E}[Y^p] = V \},
\]

where \( p > 1 \) and \( V > 0 \) are fixed in this section. This restriction comes from the quoted Power option \( Y^p \), where \( V \) denotes its market price. We let

\[
P^V(K, C_1, C_2) := \sup_{P \in M_V(K, C_1, C_2)} \mathbb{E}[c(X, Y)].
\]

and note that, by definition,

\[
P(\mu, \nu) \leq P^V(K^n, C^n_1, C^n_2) \leq P(K^n, C^n_1, C^n_2),
\]

for all \( n \geq 1 \), which yields the convergence of \( P^V(K^n, C^n_1, C^n_2) \) to \( P(\mu, \nu) \) for any sequence \( (K^n)_{n \geq 1} \) satisfying Assumption 3.8. Let \( q > 1 \) be the conjugate number of \( p \), i.e. \( 1/p + 1/q = 1 \). Then the following characterises the "distance" between the \( M_V(K^n, C^n_1, C^n_2) \) and \( M(\mu, \nu) \).

**Proposition 3.11.** For each \( n \geq 1 \) and any \( \pi \in M_V(K^n, C^n_1, C^n_2) \), set \( \mu^n = \pi \circ X^{-1} \) and \( \nu^n = \pi \circ Y^{-1} \). Then there exists a constant \( C > 0 \), depending only on \( V \) and \( p \), such that

\[
\mathcal{W}^p((\mu^n, \nu^n), (\mu, \nu)) \leq CK_n \left( \sqrt{\Delta K_n} + (K_n^p)^{-p/2q} \right).
\]

**Proof.** It suffices to treat \( \mathcal{W}(\mu^n, \mu) \). Let \( R = K_n^p \) and \( \mu_R \) (resp. \( \mu^n_R \)) be the trunctated distribution of \( \mu \) (resp. \( \mu^n \)). Indeed, let \( Z \) (resp. \( Z^n \)) denotes some random variable of law \( \mu \) (resp. \( \mu^n \)), then \( \mu_R \) (resp. \( \mu^n_R \)) be the law of \( Z_R := R \wedge Z \) (resp. \( Z^n_R = R \wedge Z^n \)). Then one has for all \( K \in [0, R] \)

\[
\left| \int (x-K)^+ d\mu^n_R - \int (x-K)^+ d\mu_R \right| = \left| \mathbb{E} \left[ (Z^n_R - K)^+ \right] - \mathbb{E} \left[ (Z_R - K)^+ \right] \right|
\]

\[
\leq \left| \mathbb{E} \left[ (Z^n - K)^+ \right] - \mathbb{E} \left[ (Z - K)^+ \right] \right| + \mathbb{E} \left[ |Z^n - Z^n_R| \right] + \mathbb{E} \left[ |Z - Z_R| \right]
\]

\[
\leq \left| \int (x-K)^+ d\mu^n - \int (x-K)^+ d\mu \right| + 2\mathbb{E} \left[ Z^n 1_{\{Z > R\}} \right] + 2\mathbb{E} \left[ Z^n 1_{\{Z^n > R\}} \right].
\]

Notice that for each \( K \in [0, R] \) there exists \( 1 \leq j < n \) such that \( K \in [K^n_j, K^n_{j+1}] \), thus

\[
\int (x-K)^+ d\mu^n - \int (x-K)^+ d\mu \leq \int (x-K^n_j)^+ d\mu^n - \int (x-K^n_{j+1})^+ d\mu
\]

\[
= \int (x-K^n_j)^+ d\mu - \int (x-K^n_{j+1})^+ d\mu \leq \Delta K^n.
\]

Hence

\[
\left| \int (x-K)^+ d\mu^n - \int (x-K)^+ d\mu \right| \leq \Delta K^n.
\]
In addition,
\[
\mathbb{E} \left[ Z^n 1_{\{Z^n > R\}} + Z 1_{\{Z > R\}} \right] \leq \mathbb{E} \left[ \frac{(Z^n)^p}{R^{p-1}} 1_{\{Z^n > R\}} + \frac{Z^p}{R^{p-1}} 1_{\{Z > R\}} \right] \leq \frac{2V}{R^{p-1}},
\]
which yields by (12) that
\[
\left| \int (x - K)^+ d\mu^n_R - \int (x - K)^+ d\mu_R \right| \leq \Delta K^n + \frac{4V}{(K^n_R)^{p/q}}.
\]
It follows by Lemma 3.13 that there exists some \( C > 0 \) such that
\[
\mathcal{W}(\mu^n_R, \mu_R) \leq C K^n_n \left( \sqrt{\Delta K^n} + (K^n_n)^{-p/2q} \right).
\]
It remains to estimate \( \mathcal{W}(\mu^n, \mu^n_R) \) and \( \mathcal{W}(\mu, \mu_R) \). It follows by definition
\[
\mathcal{W}(\mu^n, \mu^n_R) \text{ (resp. } \mathcal{W}(\mu, \mu_R) \text{) } \leq \frac{2V}{(K^n_n)^{p/q}},
\]
which yield the required inequalities by the triangle inequality. \( \square \)

**Remark 3.12.** Notice that, in order to ensure \( \mathcal{W}^\mathbb{E}(\mu^n, \nu^n), (\mu, \nu) \) converge to zero, we need \( p > 3 \) and \( K^n_n \sqrt{\Delta K^n} \to 0 \) as \( n \to \infty \).

**Lemma 3.13.** Let \( \mu \) and \( \nu \) be two probability measures supported on \([0, R]\) for some fixed \( R > 0 \). Assume that there exists some \( \varepsilon > 0 \) such that
\[
\left| \int_{\mathbb{R}_+} (x - K)^+ \mu(dx) - \int_{\mathbb{R}_+} (x - K)^+ \nu(dx) \right| \leq \varepsilon, \text{ for all } K \in [0, R].
\]
Then \( \mathcal{W}(\mu, \nu) \leq R \sqrt{2\varepsilon} \).

**Proof.** Let \( \rho(\cdot, \cdot) \) denote the Prokhorov distance, i.e.
\[
\rho(\mu, \nu) := \inf \{ \varepsilon > 0 : F_\mu(x - \varepsilon) - \varepsilon \leq F_\nu(x) \leq F_\mu(x + \varepsilon) + \varepsilon \text{ for all } x \in \mathbb{R} \},
\]
where \( F_\mu \) (resp. \( F_\nu \)) denotes the cumulative distribution function of \( \mu \) (resp. \( \nu \)). For any \( 0 < \delta < \rho(\mu, \nu) \), one has some \( K \in [0, R] \) such that \( F_\mu(K - \delta) - F_\nu(K) > \delta \) or \( F_\nu(K) - F_\mu(K + \delta) > \delta \). Take the first case without loss of generality, and it yields
\[
\int_{K-\delta}^K (F_\mu(x) - F_\nu(x)) dx \geq \int_{K-\delta}^K (F_\mu(K - \delta) - F_\nu(K)) dx > \delta^2.
\]
On the other hand,
\[
\int_{K-\delta}^K (F_\mu(x) - F_\nu(x)) dx = \left| \int_{K-\delta}^{+\infty} (F_\mu(x) - F_\nu(x)) dx - \int_{-\infty}^K (F_\mu(x) - F_\nu(x)) dx \right| = \left| \int_{\mathbb{R}_+} (x - K + \delta)^+ (\mu - \nu)(dx) - \int_{\mathbb{R}_+} (x - K)^+ (\mu - \nu)(dx) \right|.
\]

23
It follows by assumption that $\delta^2 < 2\varepsilon$. That is, $\rho(\mu, \nu) \leq \sqrt{2\varepsilon}$. Recall the dual expressions of $\rho$ and $\mathcal{W}$

$$
\rho(\mu, \nu) = \sup_{\psi \in \Lambda_1: \|\psi\|_{\infty} \leq 1} \left\{ \int_{\mathbb{R}^+} \psi d\mu - \int_{\mathbb{R}^+} \psi d\nu \right\},
$$
$$
\mathcal{W}(\mu, \nu) = \sup_{\psi \in \Lambda_1} \left\{ \int_{\mathbb{R}^+} \psi d\mu - \int_{\mathbb{R}^+} \psi d\nu \right\}.
$$

For every $\psi \in \Lambda_1$, one has $\|((\psi - \psi(0))/R\|_{\infty} \leq 1$ and thus

$$
\int_{\mathbb{R}^+} \psi d\mu - \int_{\mathbb{R}^+} \psi d\nu = \int_{\mathbb{R}^+} (\psi - \psi(0)) d\mu - \int_{\mathbb{R}^+} (\psi - \psi(0)) d\nu \leq R\rho(\mu, \nu),
$$

which implies $\mathcal{W}(\mu, \nu) \leq R\sqrt{2\varepsilon}$.

Finally, we obtain the following theorem.

**Theorem 3.14.** Let the conditions of Theorem 3.1 hold and $p > 3$. If $\lim_{n \to \infty} K_n^\alpha \sqrt{\Delta K_n} = 0$, then the convergence rate is given as below

$$
0 \leq \mathcal{P}^V(K_n, C_1^n, C_2^n) - \mathcal{P}(\mu, \nu) \leq O\left(\frac{1}{n^2} + \sqrt{\rho_n}\right),
$$

where

$$
\rho_n := K_n^\alpha \left(\sqrt{\Delta K_n} + (K_n^\alpha)^{-p/2q}\right).
$$

**Proof.** It follows from Proposition 3.11 that, there exists some $C > 0$ such that for any $\pi \in M_V(K_n, C_1^n, C_2^n)$, one has

$$
\mathcal{W}^\oplus((\pi \circ X^{-1}, \pi \circ Y^{-1}), (\mu, \nu)) \leq CK_n^\alpha \left(\sqrt{\Delta K_n} + (K_n^\alpha)^{-p/2q}\right).
$$

Repeating the reasoning in the proof of Theorem 3.3 and using $p > 3$, we obtain for every $R > 0$

$$
\mathcal{P}(\pi \circ X^{-1}, \pi \circ Y^{-1}) - \mathcal{P}(\mu, \nu) \leq 2L \left(\int_{\mathbb{R}^+} y^3 \nu(dy)/R + 1/n^2 + 10(R + 1)\rho_n\right).
$$

The proof is fulfilled by optimising the above inequality w.r.t. $\pi \in M_V(K_n, C_1^n, C_2^n)$ and $R > 0$.

**4 Numerical algorithms for the MOT–LP problem**

We focus now on solving the LP problem corresponding to the MOT problem with finitely supported measures. For simplicity we restrict ourselves to $d = 1$:

$$
\max_{(p_{i,j})_{1 \leq i \leq m, 1 \leq j \leq n} \in \mathbb{R}_{+}^{mn}} \sum_{1 \leq i \leq m, 1 \leq j \leq n} p_{i,j} c_{i,j}
$$

s.t. \[ \sum_{1 \leq j \leq n} p_{k,j} = \alpha_k, \quad \text{for} \ k = 1, \cdots, m, \]

\[ \sum_{1 \leq i \leq m} p_{i,k} = \beta_k, \quad \text{for} \ k = 1, \cdots, n, \]

\[ \sum_{1 \leq j \leq n} p_{k,j} y_j = \alpha_k x_k, \quad \text{for} \ k = 1, \cdots, m, \]

24
where \( \alpha = (\alpha_i)_{1 \leq i \leq m} \) and \( \beta = (\beta_k)_{1 \leq k \leq n} \) stand for two marginal distributions
\[
\mu(dx) = \sum_{1 \leq i \leq m} \alpha_i \delta_{x_i}(dx) \quad \text{and} \quad \nu(dy) = \sum_{1 \leq j \leq n} \beta_j \delta_{y_j}(dy)
\]
that are increasing in convex order. Without loss of generality, we assume here
\[
\alpha_i \beta_j > 0, \quad \text{for} \quad i = 1, \ldots, m, \quad j = 1, \ldots, n,
\]
\[
x_1 < \cdots < x_m \quad \text{and} \quad y_1 < \cdots < y_n,
\]
\[
y_1 < x_1 < x_m < y_n.
\]
This is an LP problem with specific linear constraints. We provide two numerical algorithms based respectively on the primal and dual problems.

### 4.1 Primal problem: iterative Bregman projection

We consider in this subsection the primal problem. Set for any \( p \in \mathbb{R}^{mn}_+ \)
\[
E(p) := \sum_{1 \leq i \leq m, 1 \leq j \leq n} p_{i,j} c_{i,j}
\]
and
\[
E_\varepsilon(p) := E(p) - \varepsilon \sum_{1 \leq i \leq m, 1 \leq j \leq n} p_{i,j} (\log(p_{i,j}) - 1),
\]
where we set by default \( 0 \times \log(0) := 0 \). Then it follows by definition that \( E(p) \leq E_\varepsilon(p) \) for all \( p \in \mathbb{R}^{mn}_+ \). Recall that \( M(\mu, \nu) \) is the set of martingale transport plans, then one has the following proposition.

**Proposition 4.1.** For any given vectors (resp. marginals) \( \alpha \) (resp. \( \mu \)) and \( \beta \) (resp. \( \nu \)), there exists some \( C > 0 \) such that
\[
0 \leq \max_{p \in M(\mu, \nu)} E_\varepsilon(p) - \max_{p \in M(\mu, \nu)} E(p) \leq C \varepsilon.
\]

**Proof.** It follows by a straightforward computation that \( p \mapsto E_\varepsilon(p) \) is continuous and strictly concave on \( \mathbb{R}^{mn}_+ \), which implies that there exists a unique \( p^*_\varepsilon \in M(\mu, \nu) \) such that \( \max_{p \in M(\mu, \nu)} E_\varepsilon(p) = E_\varepsilon(p^*_\varepsilon) \). Hence
\[
\max_{p \in M(\mu, \nu)} E_\varepsilon(p) - \max_{p \in M(\mu, \nu)} E(p) \leq E_\varepsilon(p^*_\varepsilon) - E(p^*_\varepsilon) \leq C \varepsilon,
\]
where
\[
C := - \min_{p \in M(\mu, \nu)} \sum_{1 \leq i \leq m, 1 \leq j \leq n} p_{i,j} \left( \log(p_{i,j}) - 1 \right) \leq \log(mn) + 1.
\]

Define \( q \in \mathbb{R}^{mn}_+ \) by \( q_{i,j} = e^{c_{i,j}/\varepsilon} \), then one has \( E_\varepsilon(p) := \varepsilon \text{KL}(p|q) \), where
\[
\text{KL}(p|q) := \sum_{1 \leq i \leq m, 1 \leq j \leq n} p_{i,j} \left[ 1 - \log \left( \frac{p_{i,j}}{q_{i,j}} \right) \right].
\]
Next, let us follow Bregman’s idea in [9] and separate the linear constraints given by $M(\mu, \nu)$. Define by $C_1$, $C_2$ and $C_3 \subset \mathbb{R}_{+}^{mn}$ the subsets of matrices $p = (p_{i,j})_{1 \leq i \leq m, 1 \leq j \leq n}$ as follows

\begin{align*}
C_1 & := \left\{ p \in \mathbb{R}_{+}^{mn} : \sum_{1 \leq j \leq n} p_{k,j} = \alpha_k, \text{ for } k = 1, \ldots, m \right\}, \\
C_2 & := \left\{ p \in \mathbb{R}_{+}^{mn} : \sum_{1 \leq i \leq m} p_{i,k} = \beta_k, \text{ for } k = 1, \ldots, n \right\}, \\
C_3 & := \left\{ p \in \mathbb{R}_{+}^{mn} : \sum_{1 \leq j \leq n} p_{k,j} y_j = \alpha_k x_k, \text{ for } k = 1, \ldots, m \right\}.
\end{align*}

Then one has $M(\mu, \nu) = C_1 \cap C_2 \cap C_3$ and

$$
\max_{p \in M(\mu, \nu)} E_\varepsilon(p) = \varepsilon \max_{p \in C_1 \cap C_2 \cap C_3} \text{KL}(p|q).
$$

Notice that $p \mapsto \text{KL}(p|q)$ is strictly concave, then we may apply the iterative Bregman projection as follows. Define the sequence $\{C_l\}_{l \geq 1}$ by $C_{l+3} := C_l$ for all $l \geq 1$. Let $p^{(0)} = q$, and set for $l \geq 1$

$$
p^{(l)} := \text{Proj}_{C_l}(p^{(l-1)}) := \text{argmax}_{p \in C_l} \text{KL}(p|p^{(l-1)}).
$$

The next theorem follows from Theorem 3 of Bregman [9].

**Theorem 4.2.** With the above notations, the sequence $(p^{(l)})_{l \geq 1}$ constructed above is convergent and

$$
\lim_{l \to \infty} p^{(l)} = \text{argmax}_{p \in C_1 \cap C_2 \cap C_3} \text{KL}(p|q).
$$

For each projection, let us compute $p^{(l)}$ by introducing a Lagrangian multiplier. We distinguish three cases:

(i) $C_l = C_1$. We consider the following Lagrangian formulation

$$
L(p, \lambda) := \text{KL}(p|p^{(l-1)}) + \sum_{i=1}^{m} \lambda_i \left( \sum_{j=1}^{n} p_{i,j} - \alpha_i \right).
$$

Setting

$$
\partial p_{i,j} L(p, \lambda) = \partial \lambda_i L(p, \lambda) = 0, \text{ for } i = 1, \ldots, m, j = 1, \ldots, n,
$$

we obtain

\begin{align*}
p_{i,j}^{(l)} &= e^{\lambda_i p_{i,j}^{(l-1)}}, \text{ for } i = 1, \ldots, m, j = 1, \ldots, n, \\
\sum_{j=1}^{n} p_{i,j}^{(l)} &= \alpha_i, \text{ for } i = 1, \ldots, m,
\end{align*}

which yields

$$
p_{i,j}^{(l)} = \frac{\alpha_i p_{i,j}^{(l-1)}}{\sum_{j=1}^{n} p_{i,j}^{(l-1)}}, \text{ for } i = 1, \ldots, m, j = 1, \ldots, n.
$$
(ii) \( C_l = C_2 \). Using the same argument, we get
\[
p_{i,j}^{(l)} = \frac{\beta_j p_{i,j}^{(l-1)}}{\sum_{i=1}^{m} p_{i,j}^{(l-1)}}, \quad \text{for } i = 1, \ldots, m, j = 1, \ldots, n.
\]

(iii) \( C_l = C_3 \). We consider also
\[
L(p, \lambda) := KL(p|p_{l-1}) + \sum_{i=1}^{m} \lambda_i \left( \sum_{j=1}^{n} p_{i,j} y_j - \alpha_i x_i \right)
\]
and obtain the equations
\[
p_{i,j} = e^{\lambda y_j p_{i,j}^{(l-1)}}, \quad \text{for } i = 1, \ldots, m, j = 1, \ldots, n,
\]
\[
\sum_{j=1}^{n} y_j p_{i,j} = x_i \alpha_i, \quad \text{for } i = 1, \ldots, m.
\]
Define for \( i = 1, \ldots, m \) the functions \( f_i : \mathbb{R} \to \mathbb{R} \) by
\[
f_i(z) := \sum_{j=1}^{n} y_j e^{zy_j p_{i,j}^{(l-1)}} - x_i \alpha_i.
\]

Then the following proposition shows that \( f_i \) has a unique root on \( \mathbb{R} \).

**Proposition 4.3.** With the notations above, there exists a unique \( \lambda_{i}^{(l)} \in \mathbb{R} \) such that \( f_i(\lambda_{i}^{(l)}) = 0 \).

**Proof.** Notice first that \( p_{i,j}^{(l-1)} > 0 \) by the iteration for \( i = 1, \ldots, m \) and \( j = 1, \ldots, n \). Then the map \( z \to y_j e^{zy_j p_{i,j}^{(l-1)}} =: g_{i,j}(z) \) is strictly increasing whenever \( y_j \neq 0 \), and satisfies further
\[
g_{i,j}(+\infty) = +\infty \quad \text{and} \quad g_{i,j}(-\infty) = 0, \quad \text{if } y_j > 0,
\]
\[
g_{i,j}(+\infty) = 0 \quad \text{and} \quad g_{i,j}(-\infty) = -\infty, \quad \text{if } y_j < 0.
\]
Hence \( f_i \in C^\infty(\mathbb{R}) \) is strictly increasing. Next we distinguish three cases:

(i) If \( y_n \leq 0 \), then one has \( f_i(+) > 0 \) and \( f_i(-) = -\infty \), which implies that \( f_i \) has a unique root on \( \mathbb{R} \);

(ii) If \( y_1 \geq 0 \), then one has \( f_i(+) = +\infty \) and \( f_i(-) < 0 \), which implies that \( f_i \) has a unique root on \( \mathbb{R} \);

(iii) Otherwise, then one has \( f_i(+) = +\infty \) and \( f_i(-) = -\infty \), which implies that \( f_i \) has a unique root on \( \mathbb{R} \).

As \( f_i \) is strictly increasing and \( f'_i(z) > 0 \) for all \( z \in \mathbb{R} \), we may apply Newton’s method to search numerically for the root \( \lambda_{i}^{(l)} \).

**Proposition 4.4.** Let \( z_0 = 0 \) and set for all \( k \geq 0 \)
\[
z_{k+1} := z_k - \frac{f_i(z_k)}{f'_i(z_k)}
\]
Then \( \lim_{k \to \infty} z_k = \lambda_{i}^{(l)} \) and the convergence is quadratic.
4.2 Dual problem: computation of concave envelope

We propose now another numerical method based on the dual formulation as follows

\[ D(\mu, \nu) := \inf_{(\psi_1, \ldots, \psi_m, h_1, \ldots, h_m) \in \mathbb{R}^n \times \mathbb{R}^m} \left\{ \sum_{1 \leq i \leq m} \alpha_i (c_i - \psi)(x_i) + \sum_{1 \leq j \leq n} \beta_j \psi_j \right\}, \]

s.t. \( \varphi_i + \psi_j + h_i (y_j - x_i) \geq c_{i,j} \), for \( i = 1, \ldots, m \), \( j = 1, \ldots, n \).

We first show that the above problem can be rewritten as another minimisation problem \( \inf_{\psi \in \mathbb{R}^n} \mathcal{J}(\psi) \) without constraint, where \( \mathcal{J} : \mathbb{R}^n \rightarrow \mathbb{R} \) denotes some convex function. We need to introduce the concave envelope of discrete version. Given the data \((\psi_j)_{1 \leq j \leq n}\) at the points \((y_j)_{1 \leq j \leq n}\), denote by \( \psi : [y_1, y_n] \rightarrow \mathbb{R} \) its linear interpolation. In order for \( \psi^c \) to be the discrete concave envelope of \( \psi \), we require that \( \psi(y) \leq \psi^c(y) \) for all \( y \in [y_1, y_n] \) and \( \psi^c \) is concave. With confusion, \( \psi^c \) is also said to be the concave envelope of \((\psi_j)_{1 \leq j \leq n}\). Then one has the following proposition.

**Proposition 4.5.** The above problem \( D(\mu, \nu) \) is equivalent to the following minimisation problem

\[ D'(\mu, \nu) := \inf_{(\psi_1, \ldots, \psi_n) \in \mathbb{R}^n} \left\{ \sum_{1 \leq i \leq m} \alpha_i (c_i - \psi)(x_i) + \sum_{1 \leq j \leq n} \beta_j \psi_j \right\}, \]

where \( (c_i - \psi)^c \) denotes the concave envelope of \((c_{i,j} - \psi_j)_{1 \leq j \leq n}\).

**Proof.** For every \((\psi_j)_{1 \leq j \leq n} \in \mathbb{R}^n\), let \( \varphi_i := (c_i - \psi)^c(x_i) \) and \( h_i \) be the left derivative of \( (c_i - \psi)^c \) at \( x_i \). Then it follows by definition that

\[ c_{i,j} - \psi_j \leq (c_i - \psi)^c(y_j) \leq \varphi_i + h_i (y_j - x_i), \]

for \( j = 1, \ldots, n \), which implies that \((\varphi_i)_{1 \leq i \leq m}, (\psi_j)_{1 \leq j \leq n}, (h_i)_{1 \leq i \leq m}\) is an admissible triplet, and further \( D(\mu, \nu) \leq D'(\mu, \nu) \). Conversely, for each triplet \((\varphi, \psi, h)\) satisfying the constraint for \( D(\mu, \nu) \), one has

\[ \varphi_i + h_i (y_j - x_i) \geq c_{i,j} - \psi_j, \]

which yields by the linearity that

\[ \varphi_i + h_i (y - x_i) \geq (c_i - \psi)(y), \]

for all \( y \in [y_1, y_n] \) and further

\[ \varphi_i \geq (c_i - \psi)^c(x_i), \]

for \( i = 1, \ldots, m \), which fulfils the proof. \( \square \)

Define the map \( \mathcal{J} : \mathbb{R}^n \rightarrow \mathbb{R} \) by

\[ \mathcal{J}(\psi_1, \ldots, \psi_n) := \sum_{1 \leq i \leq m} \alpha_i (c_i - \psi)(x_i) + \sum_{1 \leq j \leq n} \beta_j \psi_j, \]

then \( \mathcal{J} \) is convex and \( D(\mu, \nu) = \inf_{(\psi_1, \ldots, \psi_n) \in \mathbb{R}^n} \mathcal{J}(\psi_1, \ldots, \psi_n) \in \mathbb{R} \). It remains to solve a minimisation problem for a convex function where the computation of concave envelope is involved.
A lot of numerical methods for computing the concave envelope are achieved in the literature. Here we will make use of the particular structure of \( J \). Denote by \( c_i^\psi := c_i - \psi \). It follows from Theorem 2.35 of Dacorogna [10] that,

\[
(c_i^\psi)^c(y) = \sup \left\{ \rho c_i^\psi(y') + (1 - \rho) c_i^\psi(y'') : y = \rho y' + (1 - \rho) y'', \ 0 \leq \rho \leq 1 \right\}.
\]

By definition, there exist \([y_{j'},y_{j'+1}]\) and \([y_{j''},y_{j''+1}]\) such that \( y'_i \in \left[ y_{j'},y_{j'+1} \right] \) and \( y''_i \in \left[ y_{j''},y_{j''+1} \right] \), which yields some \( \rho', \rho'' \in [0,1] \) such that

\[
c_i^\psi(y'_i) = \rho' c_i^\psi(y'_{j'}) + (1 - \rho') c_i^\psi(y'_{j'+1}),
\]

\[
c_i^\psi(y''_i) = \rho'' c_i^\psi(y''_{j''}) + (1 - \rho'') c_i^\psi(y''_{j''+1}).
\]

Hence

\[
(c_i^\psi)^c(y) = \sup \left\{ \sum_{1 \leq k \leq 4} \rho_k(c_{i,j_k} - \psi_{j_k}) : \sum_{1 \leq k \leq 4} \rho_k y_{j_k} = y, \ \sum_{1 \leq k \leq 4} \rho_k = 1, \ 0 \leq \rho_k \leq 1 \right\}.
\]

Then we may use the convex program in Boyd & Vandenberghe [8, Section 6.5.5] to build the concave envelope of \( c_i^\psi \).

Given a numerical solver for \( J \), we may solve the minimisation problem by the following sub-gradient method. Letting \( (\gamma_n)_{n \geq 0} \) be a sequence of positive constants, we proceed as following. Set \( (\psi^0, \cdots , \psi^0_n) := 0 \) and for all \( k \geq 0 \), define

\[
(\psi^{k+1}, \cdots , \psi^{k+1}_n) := (\psi^k, \cdots , \psi^n) + \gamma_k \nabla J_k,
\]

where \( \nabla J_k \) is the sub-gradient of \( J \) at \( (\psi^k_1, \cdots , \psi^k_n) \) and can be taken as a finite differences approximation.

5 Appendix

Proof of Theorem 1.3. It suffices to argue for \( N = 2 \), the general result then follows by composition of disintegration kernels. Write \( \mu = \mu_1 \) and \( \nu = \mu_2 \). Set

\[
\Pi^c(\mu) := \left\{ \pi \in \Pi(\mathbb{R}^2) : \pi \circ X^{-1} = \mu \ \text{and} \ |E_\pi[Y|X] - X| \leq \varepsilon, \ \pi - \text{a.s.} \right\}.
\]

Then it follows from Theorem 7 of Strassen [33] that, there exists \( \pi \in \Pi^c(\mu) \) such that \( \pi \circ Y^{-1} = \nu \) if and only if

\[
\int_\mathbb{R}^2 \psi(x) \nu(dx) \leq \sup_{\pi \in \Pi^c(\mu)} \int_\mathbb{R}^2 \psi(y) \pi(dx, dy)
\]

holds for all \( \psi \in \Lambda \), which is equivalent to, by a measurable selection argument,

\[
\int_\mathbb{R} \psi d\nu \leq \int_\mathbb{R} \psi d\mu
\]

for all \( \psi \in \Lambda \), where

\[
\psi^\varepsilon(x) := \sup_{\eta \in \Pi^c(x)} \int \psi d\eta \ \text{and} \ \Pi^c(x) := \left\{ \eta \in \Pi : \left| \int_\mathbb{R} \eta(dy) - x \right| \leq \varepsilon \right\}.
\]
Let us simplify further $\psi_x$. It follows by definition that,

$$
\psi_x(x) = \sup_{\eta \in \Pi^x} \int \psi d\eta = \sup_{z : |z-x| \leq \varepsilon} \sup_{\{\eta : \int_{\mathbb{R}^d} g \, d\eta = z\}} \int \psi d\eta = \sup_{z : |z-x| \leq \varepsilon} \psi^c(z),
$$

where $\psi^c$ denotes the concave envelope of $\psi$. Therefore, $M_\varepsilon(\mu, \nu) \neq \emptyset$ if and only if

$$
\int_{\mathbb{R}} \psi(x) \nu(dx) \leq \int_{\mathbb{R}} \sup_{z : |z-x| \leq \varepsilon} \psi^c(x) \mu(dx)
$$

holds for all $\psi \in \Lambda$, which concludes the proof. \hfill \Box

**Lemma 5.1.** With the same conditions and notations of Proposition 2.2, there exist measurable functions $f_1, \ldots, f_m : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ such that for $k = 1, \ldots, m$

$$
\tilde{P} \circ (\hat{X}_k, X'_k)^{-1} = \pi_k, \quad \text{where } X'_k := f_k(\hat{X}_k, Z_k).
$$

**Proof.** Without loss of generality, we only prove for $k = 1$, and denote for the sake of simplicity $x_1 \equiv x$ and $x_1' \equiv x'$. Disintegrating w.r.t. the first coordinate $x$, one has $\pi_1(dx, dx') = \mu_1(dx) \otimes \lambda_2(dx')$, where $(\lambda_x(dx'))_{x \in \mathbb{R}^d}$ denotes the r.c.p.d., or namely, for each $x \in \mathbb{R}^d$, $\lambda_x$ is a probability measure on $\mathbb{R}^d$. The above claim is equivalent to the existence of a measurable function $f_1 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ satisfying, for $\mu$-a.s. $x \in \mathbb{R}^d$

$$
\tilde{P} \left[ f(x, Z_1) \in A \middle| \hat{X}_1 = x \right] = \lambda_x(A), \quad \text{for all } A \subseteq \mathbb{R}^d,
$$

or namely, $f_1(x, \cdot)$ transfers the law of $Z_1$ to $\lambda_x$ for $\mu$-a.s. $x \in \mathbb{R}^d$. To prove the above claim, we first show for the case $d = 1$, and then conclude for the general case.

(i) Let $F$ and $G_x$ be respectively the cumulative distribution functions of $Z_1$ and $\lambda_x$, and define further the right-continuous inverse by

$$
G_x^{-1}(t) := \inf \{ y \in \mathbb{R} : G_x(y) > t \}.
$$

Define $f_1(x, x') := G_x^{-1} \circ F(x')$, then $f_1$ is clearly measurable by the definition of r.c.p.d., and moreover in view of Villani [34], page 19-20, one has for $\mu$-a.s. $x \in \mathbb{R}$

$$
\tilde{P} \left[ X'_1 \in A \middle| \hat{X}_1 = x \right] = \lambda_x(A), \quad \text{for all } A \subseteq \mathbb{R},
$$

which concludes the claim above.

(ii) Now let us turn to the general case. Set $x' = (x_1', \ldots, x_d')$ and $y' = (y_1', \ldots, y_d')$.

**Step 1:** Take the marginal law on the first variable for $Z_1$ and $\lambda_x$, which gives probability measures $\partial_1 F_1(x_1') dx_1'$ and $\lambda_x^1(dy_1')$. Then construct the above map $f_1(x, \cdot)$ which may transfer $\partial_1 F_1(x_1') dx_1'$ to $\lambda_x^1(dy_1')$.

**Step 2:** Now take the marginal on the first two variables and disintegrate it with respect to the first variable. This gives probability measures $\partial_{x_1, x_2} F_2(x_1', x_2') dx_1' dx_2' := \partial_1 F_1(x_1') dx_1' \otimes F_2(x_1', x_2') dx_2'$ and $\lambda^2_x(dy_1', y_2') := \lambda_x^1(dy_1') \otimes \lambda^2_x(y_2')$. Then, for each $x_1'$, set $y_1' = f_1(x, x_1')$, and define $f_{1,2}(x, x_1', \cdot)$ using the above formula which relates $F_{x_1,2}(x_2') dx_2'$ to $\lambda^2_x(y_1', dy_2')$.

**Step 3:** Repeat the construction of Step 2 by adding variables one after the other and defining $f_{1,2,3}(x, x_1', x_2', \cdot)$ etc. After $m$ steps, this produces the required map $f(x, \cdot)$ which transports the law of $Z_1$ to $\lambda_x$. \hfill \Box
References

[1] A. Alfonsi, J. Corbetta, and B. Jourdain. Sampling of probability measures in the convex order and approximation of martingale optimal transport problems. arXiv:1709.05287, 2017.

[2] A. D. Barbour. Stein’s method for diffusion approximations. *Probab. Theory Related Fields*, 84(3):297–322, 1990.

[3] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices—a mass transport approach. *Finance Stoch.*, 17(3):477–501, 2013.

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

[5] M. Beiglböck, T. Lim, and J. Obłój. Dual attainment for the martingale transport problem. arXiv:1705.04273, 2017.

[6] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):A1111–A1138, 2015.

[7] J. F. Bonnans and X. Tan. A model-free no-arbitrage price bound for variance options. *Appl. Math. Optim.*, 68(1):43–73, 2013.

[8] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004.

[9] L. M. Brègman. A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming. *Ž. Vyčisl. Mat. i Mat. Fiz.*, 7:620–631, 1967.

[10] H. Brown, D. Hobson, and L. C. G. Rogers. Robust hedging of barrier options. *Math. Finance*, 11(3):285–314, 2001.

[11] R. V. Chacon. Potential processes. *Trans. Amer. Math. Soc.*, 226:39–58, 1977.

[12] J. Claisse, G. Guo, and P. Henry-Labordère. Robust hedging of options on local time. arXiv:1511.07230, 2016.

[13] A. M. G. Cox and J. Obłój. Robust hedging of double touch barrier options. *SIAM J. Financial Math.*, 2:141–182, 2011.

[14] A. M. G. Cox and J. Obłój. Robust pricing and hedging of double no-touch options. *Finance Stoch.*, 15(3):573–605, 2011.

[15] A. M. G. Cox and J. Wang. Root’s barrier: Construction, optimality and applications to variance options. *Ann. Appl. Probab.*, 23(3):859–894, March 2013.

[16] B. Dacorogna. *Direct methods in the calculus of variations*, volume 78 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2008.

[17] M. Davis and D. Hobson. The range of traded option prices. *Math. Finance*, 17(1):1–14, 2007.

[18] M. Davis, J. Obłój, and V. Raval. Arbitrage bounds for prices of weighted variance swaps. *Math. Finance*, 24(4):821–854, 2014.

[19] Y. Dolinsky and H. M. Soner. Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Related Fields*, 160(1-2):391–427, 2014.
[20] Y. Dolinsky and H. M. Soner. Martingale optimal transport in the Skorokhod space. *Stochastic Process. Appl.*, 125(10):3893–3931, 2015.

[21] A. Galichon, P. Henry-Labordère, and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.*, 24(1):312–336, 2014.

[22] G. Guo, X. Tan, and N. Touzi. Tightness and duality of martingale transport on the Skorokhod space. *Stochastic Process. Appl.*, 127(3):927–956, 2017.

[23] P. Henry-Labordère, X. Tan, and N. Touzi. An explicit martingale version of the one-dimensional Brenier’s theorem with full marginals constraint. *Stochastic Process. Appl.*, 126(9):2800–2834, 2016.

[24] D. Hobson. Robust hedging of the lookback option. *Finance Stoch.*, 2(4):329–347, 1998.

[25] D. Hobson and M. Klimmek. Model-independent hedging strategies for variance swaps. *Finance Stoch.*, 16(4):611–649, 2012.

[26] D. Hobson and M. Klimmek. Robust price bounds for the forward starting straddle. *Finance Stoch.*, 19(1):189–214, 2015.

[27] D. Hobson and A. Neuberger. Robust bounds for forward start options. *Math. Finance*, 22(1):31–56, 2012.

[28] Z. Hou and J. Obłój. On robust pricing–hedging duality in continuous time. arXiv 1503.02822v2, 2015.

[29] N. Juillet. Stability of the shadow projection and the left-curtain coupling. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(4):1823–1843, 2016.

[30] T. Lim. Optimal martingale transport between radially symmetric marginals in general dimensions. arXiv:1412.3530, 2016.

[31] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. I. Probability and its Applications* (New York). Springer-Verlag, New York, 1998. Theory.

[32] A. V. Skorokhod. On a representation of random variables. *Teor. Verojatnost. i Primenen.*, 21(3):645–648, 1976.

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

[34] C. Villani. *Optimal transport, old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.