DUAL ATTAINMENT FOR THE MARTINGALE TRANSPORT PROBLEM

MATHIAS BEIGLBÖCK, TONGSEOK LIM, AND JAN OBŁÓJ

Abstract. We investigate existence of dual optimizers in one-dimensional martingale optimal transport problems. While [5] established such existence for weak (quasi-sure) duality, [2] showed existence for the natural stronger (pointwise) duality may fail even in regular cases. We establish that (pointwise) dual maximizers exist when $y \mapsto c(x, y)$ is convex, or equivalent to a convex function. It follows that when marginals are compactly supported, the existence holds when the cost $c(x, y)$ is twice continuously differentiable in y. Further, this may not be improved as we give examples with $c(x, \cdot) \in C^{2-\varepsilon}$, $\varepsilon > 0$, where dual attainment fails. Finally, when measures are compactly supported, we show that dual optimizers are Lipschitz if c is Lipschitz.

Keywords: martingale optimal transport, Kantorovich duality, dual attainment, robust mathematical finance.

1. Introduction

In recent years, there has been a significant interest in optimal transport problems where the transport plan is constrained to be a martingale. Referred to as martingale optimal transport (MOT), they were introduced by [2, 9] to study the mathematical finance question of computing model–independent no–arbitrage price bounds, see [12] for a survey, and have been studied in many papers since, e.g. [14, 8, 7, 16]. They are however of much wider mathematical interest. Mirroring classical optimal transport, they have important consequences for the study of martingale inequalities, see e.g. [4, 11, 18]. In continuous time, they are intimately linked with the Skorokhod embedding problem, see [17] for an overview of the latter, and have already led to new contributions to this well established field, see [1].

Most papers on MOT either study the structure and geometry of optimisers or investigate a form of general Kantorovich duality. Duality is of particular importance for mathematical finance: the primal problem corresponds to option pricing while the dual offers robust hedging strategies. However the latter poses a challenge: as already shown in [2], the dual problem in MOT does not admit an optimiser in general. One way to recover the dual attainment is relaxing the duality and considering not pointwise but weaker, quasi–sure, inequalities, as shown in [5]. Our aim here instead is to identify sufficient conditions on the problem under which a suitably nice dual optimiser exists. This has immediate applications in robust

The project has been generously supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 335421. The first and the second author gratefully acknowledge support from the Austrian Science Foundation (FWF) through grant Y782 and the third author also acknowledge the support of St John's College Oxford.

mathematical finance, where pointwise inequalities are more natural. Equally importantly, we believe, this problem is of intrinsic mathematical interest. In fact, answering such questions is an important prerequisite for the future development of the field and understanding geometry of primal optimisers, or existence of Brenier–type MOT plans in multiple dimensions. So far results in this direction are limited to dimension 1 and 2 [14, 13, 3] and more recently [10]. However, the methods and results of [10] would allow to provide a satisfactory answer to this central question, conditionally on the existence of dual maximisers.

To present in more detail the questions we want to study, we need to introduce some notation Let $\Omega := \mathbb{R} \times \mathbb{R}$ be the canonical space and (X, Y) the canonical process, i.e. $X(x, y) = x$ and $Y(x, y) = y$ for all $(x, y) \in \Omega$. We also denote by $P_\mathbb{R}$ and P_Ω the collections of all probability measures on $\mathbb{R}$ and Ω, respectively. For fixed $\mu, \nu \in P_\mathbb{R}$ with finite first moments, we consider the following subsets of P_Ω

$$\Pi(\mu,\nu) := \big\{\mathbb{P} \in P_\Omega : X \sim_\mathbb{P} \mu, Y \sim_\mathbb{P} \nu\big\}, \tag{1.1}$$

$$\mathrm{MT}(\mu,\nu) := \big\{\mathbb{P} \in \Pi(\mu,\nu) : \mathbb{E}^\mathbb{P}[Y|X] = X,\ \mathbb{P}-a.s.\big\}. \tag{1.2}$$

The set $\Pi(\mu,\nu)$ is non-empty as it contains the product measure $\mu\otimes\nu$. By a classical result of Strassen [19], $\mathrm{MT}(\mu,\nu)$ is non-empty if and only if $\mu \preceq \nu$ in convex order:

$$\mu(\xi) \le \nu(\xi) \text{ for all convex function } \xi, \text{ where } \mu(\xi) := \int \xi(x)\mu(dx). \tag{1.3}$$

Throughout we assume that $c : \Omega \to \mathbb{R}$ is a Borel–measurable cost function with $c(x, y) \le a(x) + b(y)$ for some $a \in L^1(\mu)$ and $b \in L^1(\nu)$. Then $\mathbb{E}^\mathbb{P}[c(X, Y)]$ is a well-defined scalar in $\mathbb{R} \cup \{-\infty\}$. The martingale optimal transport problem, as introduced in [2] in the present discrete-time case and in [9] in continuous time, is defined by the following primal problem:

$$\mathbf{P} \;:=\; \mathbf{P}(c) := \inf_{\mathbb{P}\in\mathrm{MT}(\mu,\nu)} \mathbb{E}^\mathbb{P}[c(X,Y)]. \tag{1.4}$$

Its dual is given by

$$\mathbf{D} := \mathbf{D}(c) := \sup_{(f,g)\in D_c} \big\{\nu(g) - \mu(f)\big\}, \tag{1.5}$$

where

$$\begin{aligned} D_c := \big\{(f,g) :\ & f^- \in L^1(\mu),\ g^+ \in L^1(\nu), \text{ and for some } h \in L^\infty(\mu), \\ & g(y) - f(x) - h(x)\cdot(y-x) \le c(x,y) \quad \forall x \in \mathbb{R}, \forall y \in \mathbb{R}\big\}. \end{aligned}$$

We assume that $\mathbf{P}(c)$ is finite.

In mathematical finance, the cost c has the interpretation of the payoff of an exotic derivative and $\mathbf{P}(c)$ gives its lower no-arbitrage price. A triplet (f, g, h) on the dual side corresponds to a robust sub-hedging strategy for c: both f and g are bought through trading European options and $h(x)(y - x)$ corresponds to the payoff from buying $h(x)$ stocks at time zero.

The basic duality $\mathbf{P} = \mathbf{D}$ between the primal and dual problems was established in [2] under the assumption that c is lower-semicontinuous and bounded from below. (In fact we will invoke a more sophisticated duality result from [5] which does not require lower-semicontinuity, see Theorem 3.2 below.)

In [2], the authors also provided a simple example, based on the cost function $c(x, y) = -|y-x|$, where the dual problem is not attained. Our aim here is to study fundamental reasons why dual attainment may fail and to provide sufficient conditions for it to hold and for dual optimiser to have further desirable regularity and integrability properties. We note that [5] showed dual attainment may be recovered if we weaken the dual formulation and require inequalities to hold quasi-surely, i.e. almost surely for any $\mathbb{P} \in \mathrm{MT}(\mu,\nu)$. However this is not entirely satisfying in view of the financial and other, mentioned above, applications.

2. Main Results

We start by defining the crucial notion of a solution for the dual problem (1.5).

Definition 2.1. *Let $\mu \preceq \nu$ be in convex order and let $c(x,y)$ be a cost function. We say that a triple of functions $f : \mathbb{R} \to \mathbb{R}\cup\{+\infty\}$, $g : \mathbb{R} \to \mathbb{R}\cup\{-\infty\}$, $h : \mathbb{R} \to \mathbb{R}$ is a dual maximizer, or a solution for the dual problem* (1.5)*, if f is finite μ-a.s., g is finite ν-a.s., and for any minimizer $\mathbb{P}^* \in \mathrm{MT}(\mu,\nu)$ for the martingale optimal transport problem* (1.4)*, the following holds:*

$$g(y) - f(x) - h(x)\cdot(y-x) \le c(x,y) \quad \forall x \in \mathbb{R}, \forall y \in \mathbb{R}, \tag{2.1}$$
$$g(y) - f(x) - h(x)\cdot(y-x) = c(x,y) \quad \mathbb{P}^*\text{-}a.a.\,(x,y). \tag{2.2}$$

In fact, if (2.1)–(2.2) hold for some $\mathbb{P} \in \mathrm{MT}(\mu,\nu)$, then $\mathbb{P}$ is a minimizer of the martingale transport problem (1.4) and moreover (2.1)–(2.2) hold for all minimizers of (1.4), c.f. [5, Corollary 7.8].

A simple but important observation is that if (f,g,h) is a dual maximizer, then g can always be replaced by a "better" candidate g^* induced by (f,h), as follows:

$$g^*(y) := \inf_{x\in\mathbb{R}} \big(f(x) + h(x)\cdot(y-x) + c(x,y) \big). \tag{2.3}$$

Observe that then $g \le g^*$ while (2.1)–(2.2) still holds with g^*. The minus signs on f and h in (2.1) and (2.2) were chosen to define g^* by (2.3). In this paper, unless stated otherwise we will always assume that $g = g^*$. Now we state our main theorem.

Definition 2.2. *Let J be an interval and μ be a positive measure on $\mathbb{R}$. We say that $c(x,y)$ is semiconvex in $y \in J$ μ-uniformly in x, if there exists a Borel function $u : J \to \mathbb{R}$ such that*

$$\textit{for } \mu\textit{-a.e. } x,\ y \mapsto c(x,y) + u(y) \textit{ is continuous and convex on } J. \tag{2.4}$$

In this case, we say that u is a y-convexifier on J for c.

Theorem 2.3. *Let $\mu \preceq \nu$ be in convex order and let $J := \mathrm{conv}(\mathrm{supp}(\nu))$. Suppose that there exists a y-convexifier u on J for c. If J is not compact, then further suppose that $y \mapsto c(x,y) + u(y)$ is of linear growth on J. Then there exists a dual maximizer in the sense of Definition 2.1.*

Corollary 2.4. *In the setting of Theorem 2.3, if $c \in C^{0,2}$ – that is $\frac{\partial^2 c}{\partial y^2}$ exists and is continuous on Ω – and if ν is compactly supported then there exists a dual maximizer.*

Note that Definition 2.1 is made in a pointwise sense, that is we do not require $f \in L^1(\mu)$, $g \in L^1(\nu)$, nor $h \in L^\infty(\mu)$. But as already observed in [3, 5], this classical integrability assumption is too restrictive for the existence of dual maximizer. But by using the "extended notion of integrability" introduced in [5], this pointwise dual maximizer (f,g,h) may still be viewed as dual maximizer in the generalized sense.

To ensure integrability in the classical sense, further assumptions are required, as summarised in the following result.

Theorem 2.5. *Suppose the assumptions in Theorem 2.3 hold and that further c is Lipschitz on $J \times J$ and u is Lipschitz on J. Then there exists a dual maximizer (f, g, h) such that f and g are Lipschitz on J and h is bounded on J. In particular, $\mu(f) + \nu(g) = \mathbb{E}^{\mathbb{P}^*}[c(X, Y)]$ for any solution $\mathbb{P}^*$ to the problem* (1.4).

Remark 2.6. In Theorem 2.5 if c and u are Lipschitz with Lipschitz constant L, then f and g can be taken to have Lipschitz constants $19L$ and $17L$ respectively on J while $|h|$ is bounded by $18L$ on J, as computed in the proof. Furthermore, the Lipschitz assumption on c can be somewhat weakened, see Remark 3.8.

We close this section with a discussion of how, and in what sense, the above results are sharp. Examples which support this discussion are presented after the proofs in Section 4. First, we note that the linear growth condition in Theorem 2.3 can not be removed. Indeed, Example 4.1 gives a cost function which violates the linear growth condition together with marginals $\mu \preceq \nu$ for which the dual maximizers fail to exist. Second, the convexity condition (2.4) on J cannot be relaxed to just local convexity around x, as shown in Example 4.2, and this even for compactly supported marginals. Third, the $C^{0,2}$ regularity in Corollary 2.4, is optimal in the sense that for any given $\varepsilon > 0$ we can construct a cost function $c \in C^{2-\varepsilon}$ and compactly supported, convex–ordered marginals $\mu \preceq \nu$ for which a dual maximizer satisfying (2.1)–(2.2) does not exist. This is carried out in Example 4.3. Finally, in Example 4.4 we show the necessity of semiconvexity for the regularity in Theorem 2.5 by showing that there exist 1-Lipschitz cost c for which there is a dual maximizer (f, g, h) but $g \notin L^1(\nu)$, even when (μ, ν) are compactly supported and irreducible (see Definition 3.1 below).

3. Proofs

To establish Theorem 2.3, we prove Propositions 3.4 and 3.5 below. The key idea is to consider the martingale optimal transport problem on its irreducible components (see Definition 3.1 below and [3, Appendix A]). It is known, see Theorem 3.2 below, that on each irreducible component the dual problem admits a maximizer. Using the semiconvexity assumption on the cost function, we can show that these maximizers are appropriately bounded, such that it is possible to glue them together to obtain global maximizers of the dual problem.

Let μ, ν be probability measures on $\mathbb{R}$ which are in convex order. It was shown in [3] and [5], see Proposition 3.6 below, that there is a canonical decomposition of μ, ν into irreducible pairs $(\mu_i, \nu_i)_{i \in \mathbb{N}}$ such that for each irreducible pair (μ_i, ν_i), the dual problem attains a solution. For a probability measure μ on $\mathbb{R}$, we define its potential function by

$$u_\mu : \mathbb{R} \to \mathbb{R}, \quad u_\mu(x) := \int |x - y| \, d\mu(y).$$

Definition 3.1. *Let $\mu \preceq \nu$ be in convex order and let $I := \{x : u_\mu(x) < u_\nu(x)\}$. We say that (μ, ν) is irreducible on the domain I if I is an open interval and μ is concentrated on I.*

We recall the following result from [5] (which requires our standing assumption that $\mathbf{P}(c) \in \mathbb{R}$ and $c(x,y) \le a(x) + b(y)$ for some $a \in L^1(\mu)$ and $b \in L^1(\nu)$).

Theorem 3.2. [5, Theorem 6.2] *Let $\mu \preceq \nu$ be irreducible on the domain I. Then a dual maximizer exists.*

We remark that the definition of dual maximizer in the above theorem is slightly different from ours. However, when a primal minimizer $\mathbb{P}^*$ exists, the above result implies existence of a dual maximizer in our sense. This motivates the following general existence result, the proof of which is given at the end of this section.

Proposition 3.3. *Let $\mu \preceq \nu$ be in convex order. Assume $\mathbf{P}(c) \in \mathbb{R}$ and $c(x,y) \le a(x) + b(y)$ for some $a \in L^1(\mu)$ and $b \in L^1(\nu)$, and suppose that there exists a Borel function u such that $y \mapsto c(x,y) + u(y)$ is continuous for μ-a.e. x. Then a minimizer to the primal problem* (1.4) *exists.*

The next proposition asserts that for any dual maximizer (f,g,h) in Theorem 3.2, if the function $y \mapsto c(x,y)$ is convex for each $x \in I$, then the lower envelope function g has a "desirable shape" modulo an affine function. We first deal with the bounded domain case.

Proposition 3.4. *Let $\mu \preceq \nu$ be irreducible on a bounded domain $I =]a,b[$ and assume that (f,g,h) is a dual maximizer in Theorem 3.2. Suppose that there exist $A \in \mathbb{R} \cup \{-\infty\}$, $B \in \mathbb{R} \cup \{\infty\}$ such that $A \le a < b \le B$ and that for μ-a.e. x,*

$$y \mapsto c(x,y) \text{ is continuous and convex on } [A,B]. \tag{3.1}$$

Then we can find an affine function $L(y) = L(x) + \nabla L \cdot (y - x)$ such that $\big(\tilde f(x), \tilde g(y), \tilde h(x)\big) := \big(f(x) - L(x), g(y) - L(y), h(x) - \nabla L\big)$ is a dual maximizer, and furthermore

$$\tilde g(y) \le 0 \text{ on }]a,b[, \tag{3.2}$$
$$\tilde g(y) \ge 0 \text{ on } [A,a] \cup [b,B]. \tag{3.3}$$
$$\text{If } \nu(a) > 0 \text{ then } \tilde g(a) = 0. \text{ If } \nu(b) > 0 \text{ then } \tilde g(b) = 0. \tag{3.4}$$

Proof. **Step 1.** Let us begin by recalling some terminology from [3]: for a set $\Gamma \subseteq \mathbb{R} \times \mathbb{R}$, denote X_Γ as its projection to the first coordinate space $\mathbb{R}$, and Y_Γ to the second. We will also write $\Gamma_x = \{y : (x,y) \in \Gamma\}$.

Now let $G \subseteq \mathbb{R} \times \mathbb{R}$ be the "contact set" induced by the dual optimizer (f,g,h)

$$G := \{(x,y) \,:\, g(y) - f(x) - h(x) \cdot (y - x) = c(x,y)\} \tag{3.5}$$

so that we have

$$g(y) - f(x) - h(x) \cdot (y - x) \le c(x,y) \quad \forall x \in \mathbb{R}, \forall y \in \mathbb{R}, \tag{3.6}$$
$$g(y) - f(x) - h(x) \cdot (y - x) = c(x,y) \quad \text{on} \quad G. \tag{3.7}$$

Note that f,h are real-valued on X_G and g is real-valued on Y_G. Now the fact that (μ,ν) is irreducible and $\mathbb{P}(G) = 1$ for some martingale measure $\mathbb{P}$ (in fact, for all optimal martingale measures) implies the following: for every $z \in I$, there exists $x_z \in X_G$ and $a_z, b_z \in \overline{I}$ such that $(x_z, a_z), (x_z, b_z) \in G$ and $a_z < z < b_z$. In particular the family $]a_z, b_z[_{z \in I}$ is an open cover of I, hence we can find a countable sequence $\{x_n\} \subseteq X_G$ (not necessarily disjoint) and $\{a_n\}$, $\{b_n\}$ such that

$(x_n, a_n), (x_n, b_n) \in G$ and $\{]a_n, b_n[_{n\geq 1}\}$ is an open cover of I. Furthermore, by rearrangement we can get that

$$]l_n, r_n[\cap]a_{n+1}, b_{n+1}[\neq \emptyset \quad \forall n, \quad \text{and} \tag{3.8}$$

$$l_n \searrow a, \quad r_n \nearrow b \quad \text{as} \quad n \to \infty \tag{3.9}$$

where $l_n := \min_{i\leq n} a_i$, $r_n := \max_{i\leq n} b_i$. We can assume that the case $l_n \leq a_{n+1} < b_{n+1} \leq r_n$ does not occur, since if it occurs then we may simply discard the respective x_{n+1} from the sequence.

For convenience, we record the following simple fact:

$$\text{For each } n, \text{ there exist } 1 \leq i, j \leq n \text{ such that (3.7) holds for } (x_i, l_n) \text{ and } (x_j, r_n). \tag{3.10}$$

Step 2. For each n, define

$$v_{x_n}(y) = c(x_n, y) + f(x_n) + h(x_n) \cdot (y - x_n), \tag{3.11}$$

$$g_n(y) = \inf_{i\leq n} \big(v_{x_i}(y)\big), \tag{3.12}$$

$$L_n(y) = \frac{g_n(r_n) - g_n(l_n)}{r_n - l_n} \cdot (y - l_n) + g_n(l_n). \tag{3.13}$$

We claim that, for each n,

$$g_n(y) \leq L_n(y) \quad \forall y \in [l_n, r_n], \tag{3.14}$$

$$g_n(y) \geq L_n(y) \quad \forall y \in [A, l_n] \cup [r_n, B]. \tag{3.15}$$

The claim is obvious for $n = 1$ since $v_{x_1}(y)$ is convex. Suppose that the claim is true for n. Then there are three cases. To deal with them, the following fact which comes from (3.6), (3.7), shall be useful:

$$v_{x_m}(a_n) \geq v_{x_n}(a_n) = g_n(a_n), \quad v_{x_m}(b_n) \geq v_{x_n}(b_n) = g_n(b_n) \quad \forall m, n \in \mathbb{N}. \tag{3.16}$$

Case 1 : $a_{n+1} \leq l_n < r_n \leq b_{n+1}$.
First, since $g_{n+1}(y) = \min\big(g_n(y), v_{x_{n+1}}(y)\big)$ and since $g_{n+1}(a_{n+1}) = v_{x_{n+1}}(a_{n+1})$, $g_{n+1}(b_{n+1}) = v_{x_{n+1}}(b_{n+1})$ by (3.16), by convexity of $v_{x_{n+1}}(y)$ we see that

$$g_{n+1}(y) \leq v_{x_{n+1}}(y) \leq L_{n+1}(y) \quad \forall y \in [a_{n+1}, b_{n+1}]. \tag{3.17}$$

This establishes the claim in (3.14) for $n + 1$.
For the second claim, fix $i \in \{1, 2, \ldots, n + 1\}$. Then there exists $y_i \in [a_{n+1}, b_{n+1}]$ such that $(x_i, y_i) \in G$. Then (3.17) implies

$$v_{x_i}(y_i) = g_{n+1}(y_i) \leq L_{n+1}(y_i). \tag{3.18}$$

Meanwhile, (3.16) gives

$$v_{x_i}(a_{n+1}) \geq v_{x_{n+1}}(a_{n+1}) = g_{n+1}(a_{n+1}), \quad v_{x_i}(b_{n+1}) \geq v_{x_{n+1}}(b_{n+1}) = g_{n+1}(b_{n+1}). \tag{3.19}$$

Now by (3.18), (3.19) and convexity of $v_{x_i}(y)$, we deduce that

$$v_{x_i}(y) \geq L_{n+1}(y) \quad \forall y \in [A, a_{n+1}] \cup [b_{n+1}, B]. \tag{3.20}$$

As (3.20) holds for every i, this verifies the claim (3.15) for $n + 1$, completing the inductive step.

Case 2 : $l_n \le a_{n+1} \le r_n \le b_{n+1}$.
First, since $v_{x_{n+1}}(a_{n+1}) = g_{n+1}(a_{n+1}) \le g_n(a_{n+1})$, by the induction hypothesis

$$v_{x_{n+1}}(a_{n+1}) \le L_n(a_{n+1}). \tag{3.21}$$

Also note that by (3.10), we have

$$v_{x_{n+1}}(r_n) \ge g_n(r_n) = L_n(r_n). \tag{3.22}$$

Now the convexity of $v_{x_{n+1}}$ implies

$$v_{x_{n+1}}(b_{n+1}) \ge L_n(b_{n+1}). \tag{3.23}$$

Note that $L_n(l_n) = g_n(l_n) = g_{n+1}(l_n)$ and $g_{n+1}(b_{n+1}) = v_{x_{n+1}}(b_{n+1})$. Hence by (3.23),

$$\lambda g_{n+1}(l_n) + (1-\lambda) g_{n+1}(b_{n+1}) \ge \lambda L_n(l_n) + (1-\lambda) L_n(b_{n+1}) \quad \forall \lambda \in [0,1]. \tag{3.24}$$

As $g_{n+1}(y) = \min\big(g_n(y), v_{x_{n+1}}(y)\big)$, the induction hypothesis and convexity of $v_{x_{n+1}}(y)$ along with (3.21), (3.24) imply the first claim (3.14) for $n+1$.
For the second claim, fix $i \in \{1, 2, \dots, n+1\}$. Then there exists $y_i \in [l_n, b_{n+1}]$ such that $(x_i, y_i) \in G$, and the first claim (3.14) for $n+1$ gives $v_{x_i}(y_i) \le L_{n+1}(y_i)$. On the other hand, $v_{x_i}(l_n) \ge g_n(l_n) = g_{n+1}(l_n)$ and $v_{x_i}(b_{n+1}) \ge v_{x_{n+1}}(b_{n+1}) = g_{n+1}(b_{n+1})$. Hence by convexity of $v_{x_i}(y)$, we deduce that

$$v_{x_i}(y) \ge L_{n+1}(y) \quad \forall y \in [A, l_n] \cup [b_{n+1}, B]. \tag{3.25}$$

As (3.25) holds for every i, this verifies the claim (3.15) for $n+1$, completing the induction. The third case $a_{n+1} \le l_n \le b_{n+1} \le r_n$ can be treated in the same way.
Step 3. We claim that there exists $M > 0$ such that

$$\sup_{y \in [a,b], n \in \mathbb{N}} |L_n(y)| \le M. \tag{3.26}$$

To prove (3.26), choose $M > 0$ such that $|v_{x_1}(y)| \le M$ on $[a,b]$. Then $L_n(l_n) = g_n(l_n) \le v_{x_1}(l_n) \le M$ and $L_n(r_n) = g_n(r_n) \le v_{x_1}(r_n) \le M$. Hence, as L is linear, $L_n(a_1) \le M$ and $L_n(b_1) \le M$. On the other hand, by Step 2, $-M \le v_{x_1}(a_1) = g_n(a_1) \le L_n(a_1)$ and $-M \le v_{x_1}(b_1) = g_n(b_1) \le L_n(b_1)$. This implies (3.26). In particular, there exists a subsequence of L_n (which we denote as L_k) such that $L_k(y)$ uniformly converges to an affine function as $k \to \infty$, say $L(y)$ on every compact interval in $\mathbb{R}$. Now we claim that, for $v_x(y) := c(x,y) + f(x) + h(x)\cdot(y-x)$ and $g(y) = \inf_{x \in X_G} \big(v_x(y)\big)$,

$$g(y) \le L(y) \text{ on }]a,b[, \tag{3.27}$$

$$g(y) \ge L(y) \text{ on } [A,a] \cup [b,B]. \tag{3.28}$$

First it is easy to see (3.27) as follows: if $y \in]a,b[$ then for all large k we have $y \in]l_k, r_k[$, thus by Step 2, $g(y) \le g_k(y) \le L_k(y)$. By taking $k \to \infty$, we see that $g(y) \le L(y)$, proving (3.27).

Next, suppose that there exists $(x,y) \in G$ with $a < y < b$. Then again for all large k we have $v_x(y) = g(y) \le g_k(y) \le L_k(y)$, thus $v_x(y) \le L(y)$. On the other hand, by (3.6), (3.7) we have $v_x(l_k) \ge g_k(l_k)$ and $v_x(r_k) \ge g_k(r_k)$, thus $v_x(a) \ge L(a)$ and $v_x(b) \ge L(b)$ by letting $k \to \infty$. By convexity of v_x, this implies that $v_x(y) \ge L(y)$ on $[A,a] \cup [b,B]$.

If there is no y such that $a < y < b$ and $(x,y) \in G$, this means that $G_x = \{a,b\}$, i.e. $(x,a), (x,b) \in G$. Then without loss of generality we may simply include this x in the sequence $\{x_n\}$ defined in Step 1, say we put $x = x_1$. This implies that $l_n = a$

and $r_n = b$ for all n. Thus $v_x(a) = L(a)$ and $v_x(b) = L(b)$. By convexity of v_x, this implies that $v_x(y) \geq L(y)$ on $[A, a] \cup [b, B]$. Hence, for any $x \in X_G$ we deduce that $v_x(y) \geq L(y)$ on $[A, a] \cup [b, B]$, therefore (3.28) follows.

If $\nu(a) > 0$ then there exists $x_a \in I$ such that $(x_a, a) \in G$. Then again we may include x_a in the sequence $\{x_n\}$ defined in Step 1. This implies that $l_n = a$ for all large n, thus $g(a) = v_{x_a}(a) = L(a)$. Similarly if $\nu(b) > 0$ then $g(b) = L(b)$.

Finally, we see that $\big(\tilde f(x), \tilde g(y), \tilde h(x)\big) := \big(f(x) - L(x), g(y) - L(y), h(x) - \nabla L\big)$ satisfies (3.6), (3.7), (3.2), (3.3) and (3.4), concluding the proof. $\square$

Next, we deal with the half-infinite domain case.

Proposition 3.5. *Let $\mu \preceq \nu$ be irreducible on a half-infinite domain I and assume that (f, g, h) is a dual maximizer in Theorem 3.2. Suppose that there exists an interval J such that $\overline{I} \subseteq J$ and for μ-a.e. x,*

$y \mapsto c(x, y)$ is continuous and convex on J, and

there exists an affine function L_x such that $c(x, y) \leq L_x(y)$ for all $y \in I$.

Then we can find an affine function L such that $\big(\tilde f(x), \tilde g(y), \tilde h(x)\big) := \big(f(x) - L(x), g(y) - L(y), h(x) - \nabla L\big)$ is a dual maximizer, and furthermore

$$\tilde g(y) \leq 0 \text{ on } I, \tag{3.29}$$

$$\tilde g(y) \geq 0 \text{ on } J \setminus I, \tag{3.30}$$

$$\text{if } \nu(a) > 0 \text{ then } \tilde g(a) = 0, \text{ where } \{a\} = \partial I. \tag{3.31}$$

Proof. Without loss of generality we assume that $a = 0$, $I =]0, \infty[$, and $J = [A, \infty[$ for some $A \in [-\infty, 0]$. Recall the Step 1 in the proof of Proposition 3.4 and note that now $l_n \searrow 0$, $r_n \nearrow +\infty$ as $n \to \infty$. Also recall definitions (3.11) – (3.13). Altering the triple (f, g, h) by an appropriate affine function and using the condition of linear growth and convexity satisfied by the cost, we can assume that

$$v_{x_1}(y) \text{ is decreasing on } [A, \infty[,\ v(0) = 0 \text{ and } \lim_{y \to \infty} v_{x_1}(y) = b > -\infty. \tag{3.32}$$

Now we claim that, for each n,

$$g_n(y) \text{ is decreasing on } [A, l_n], \text{ and } g_n(y) \leq g_n(l_n) \text{ on } [l_n, \infty[. \tag{3.33}$$

Note that the claim (3.33) is obviously true for $n = 1$ by the assumption (3.32). Suppose the claim is true for n. We will show that the claim is also true for $n + 1$. To see this, note that as $b_{n+1} \geq l_n$, using (3.6), (3.7) and the induction hypothesis (3.33), we see that

$$v_{x_{n+1}}(b_{n+1}) \leq g_n(b_{n+1}) \leq g_n(l_n), \text{ while } v_{x_{n+1}}(l_n) \geq g_n(l_n). \tag{3.34}$$

(If $b_{n+1} = \infty$, then instead of b_{n+1} we may argue with arbitrarily large c_{n+1} satisfying $(x_{n+1}, c_{n+1}) \in G$.) By (3.34) and convexity of $v_{x_{n+1}}(y)$, we see that $v_{x_{n+1}}(y)$ is decreasing on $[A, l_n]$. As $g_{n+1}(y) = \min\big(g_n(y), v_{x_{n+1}}(y)\big)$ and $g_n(y)$ is decreasing on $[A, l_n]$, we see that

$$g_{n+1}(y) \text{ is decreasing on } [A, l_n]. \tag{3.35}$$

In particular, for any $y \in [l_{n+1}, l_n]$ we have $g_{n+1}(l_{n+1}) \geq g_{n+1}(y)$. For $y \geq l_n$, we see that $g_{n+1}(l_{n+1}) \geq g_{n+1}(l_n) = g_n(l_n)$ by (3.10), and $g_n(l_n) \geq g_n(y) \geq g_{n+1}(y)$ by (3.33). Hence

$$g_{n+1}(y) \leq g_{n+1}(l_{n+1}) \text{ on } [l_{n+1}, \infty[. \tag{3.36}$$

Therefore, (3.33) is proved for all n.
We have observed that the sequence $\{g_n(l_n)\}$ is increasing. Note that $g_n(l_n) \le v_{x_1}(l_n) \le v_{x_1}(0) = 0$ for all n, thus $\{g_n(l_n)\}$ converges to a constant, say, C. Then we claim

$$g(y) \ge C \text{ on } [A, 0], \tag{3.37}$$

$$g(y) \le C \text{ on }]0, +\infty[, \tag{3.38}$$

where, as before, $v_x(y) := c(x,y) + f(x) + h(x) \cdot (y - x)$ and $g(y) := \inf_{x \in X_G} \big(v_x(y)\big)$.
To see this, fix $x > 0$. Then there exists n such that $l_n < x$. Arguing as above, we see that $v_x(y)$ is decreasing on $[A, l_n]$ and $v_x(l_n) \ge g_n(l_n)$ for all n. Hence for any $y \le 0$ we see that $v_x(y) \ge g_n(l_n)$. Letting $n \to \infty$ we conclude

$$g(y) \ge C \quad \text{for all } y \in [A, 0]. \tag{3.39}$$

Now for $y > 0$, there exists n such that $l_n < y$. Then by (3.33), we see that $g(y) \le g_n(y) \le g_n(l_n)$ for all large n, thus by taking $n \to \infty$ we conclude

$$g(y) \le C \quad \text{for all } y > 0. \tag{3.40}$$

If $\nu(0) > 0$ then there is $x \in X_G$ with $(x, 0) \in G$, and we may simply put this x into the sequence $\{x_n\}$ by letting $x = x_1$. Then every l_n simply becomes 0 and $\{g_n(l_n)\}$ becomes the constant sequence C. Hence, $g(0) = C$. Finally, altering the triple (f, g, h) by the constant function $-C$, we can assume that $C = 0$. This proves the proposition. □

We are now ready to show the existence of dual optimizers for the martingale optimal transport problem in Theorem 2.3. In particular we no longer assume the irreducibility of (μ, ν). Note that if (μ, ν) is irreducible on the domain $I = \mathbb{R}$ then Theorem 2.3 simply follows from Theorem 3.2. Otherwise, (μ, ν) can be decomposed into at most countably many irreducible components, and any martingale $\mathbb{P} \in \mathrm{MT}(\mu, \nu)$ is decomposed accordingly. More precisely we recall:

Proposition 3.6. [3, Theorem A.4] *Let $\mu \preceq \nu$ and let $(I_k)_{k \ge 1}$ be the open connected components of the set $\{x : u_\mu(x) < u_\nu(x)\}$. Set $I_0 = \mathbb{R} \setminus \cup_{k \ge 1} I_k$ and $\mu_k = \mu|_{I_k}$ for $k \ge 0$, so that $\mu = \sum_{k \ge 0} \mu_k$. Then, there exists a unique decomposition $\nu = \sum_{k \ge 0} \nu_k$ such that*

$$\mu_0 = \nu_0 \qquad \text{and} \qquad \mu_k \preceq \nu_k \quad \text{for all} \quad k \ge 1,$$

and this decomposition satisfies $I_k = \{x : u_{\mu_k}(x) < u_{\nu_k}(x)\}$ for all $k \ge 1$. Moreover, any $\mathbb{P} \in \mathrm{MT}(\mu, \nu)$ admits a unique decomposition $\mathbb{P} = \sum_{k \ge 0} \mathbb{P}_k$ such that $\mathbb{P}_k \in \mathrm{MT}(\mu_k, \nu_k)$ for all $k \ge 0$.

Note that as $\mu_0 = \nu_0$, $\mathbb{P}_0$ must be the identity martingale. We can now give the proof of our first main result.

Proof of Theorem 2.3. Notice that by definition of the dual maximizer and the assumption on the cost, we can assume that $y \mapsto c(x, y)$ is continuous and convex on $J := \mathrm{conv}(\mathrm{supp}(\nu))$. Let $\mathbb{P}^*$ be any minimizer in $\mathrm{MT}(\mu, \nu)$ for the problem (1.4). Then $\mathbb{P}^*_k$ is a minimizer in $\mathrm{MT}(\mu_k, \nu_k)$. For each $k \ge 1$, choose a set $G_k \subseteq \mathbb{R} \times \mathbb{R}$ and a triple (f_k, g_k, h_k) provided by Proposition 3.4 if I_k is bounded, or by Proposition 3.5 if I_k is half-infinite. We need to define G_0 and (f_0, g_0, h_0) for I_0. As $\mathbb{P}^*_0$ is the identity map, of course we take $G_0 := \{(x, x) : x \in I_0\}$. For each $x \in I_0$ define $f_0(x) = -c(x, x)$, and choose $h_0(x)$ in such a way that the convex function

$v_x(y) := c(x,y) + f_0(x) + h_0(x) \cdot (y-x)$ satisfies $v(x) = v'(x) = 0$ (more precisely 0 belongs to the subdifferential of v_x at x). Define $g_0(y) = \inf_{x \in I_0}\{v_x(y)\}$ so in particular $g_0 = 0$ on I_0. Finally, define

$$\begin{aligned} f(x) &= f_k(x) \quad \text{if} \quad x \in X_{G_k}, \\ h(x) &= h_k(x) \quad \text{if} \quad x \in X_{G_k}, \\ g(y) &= \inf_{k \geq 0} g_k(y). \end{aligned}$$

Let $G = \cup_{k \geq 0} G_k$. Obviously $\mathbb{P}^*(G) = 1$. Now observe that the properties (3.2), (3.3), (3.4), (3.29), (3.30), (3.31) verified in Proposition 3.4, 3.5 imply that the triples $(f_k, g_k, h_k)_{k \geq 0}$ are compatible, that is, the duality (3.6), (3.7) holds for G and (f, g, h). This completes the proof. $\square$

Remark 3.7. The linear growth assumption of the function $y \mapsto c(x,y)$ is required only for those x in the half-infinite irreducible domain of μ, ν as in Proposition 3.5.

Proof of Theorem 2.5. We will say that a function f is L-Lipschitz on D if $|f(x) - f(y)| \leq L|x-y|$ for all $x, y \in D$. Assume c is L-Lipschitz on $J \times J$ and u is L-Lipschitz on J, and let $K = 2L$.

Consider (μ', ν') which is an irreducible component of (μ, ν) and let $I =]a, b[$ (possibly unbounded) be the domain of (μ', ν'). That is, $(\mu', \nu') = (\mu_k, \nu_k)$ for some $k \geq 1$ in Proposition 3.6. Let $\tilde{c}(x,y) := c(x,y) + u(y)$, and let (f, g, h) be a dual maximizer for the cost $\tilde{c}$ and (μ', ν'), satisfying the conclusion of (i) Proposition 3.4 for bounded I, or (ii) Proposition 3.5 for half-unbounded I (for $I = \mathbb{R}$ we have $\nu' = \nu$ and a dual optimizer is given by Theorem 3.2). Recall that there exists a regular set $G \subseteq I \times \overline{I}$ on which every solution $\mathbb{P}^*$ to the problem (1.4) with (μ', ν') is concentrated, so that the duality (3.6)–(3.7) holds with (f, g, h), $\tilde{c}$, and G.

Define $v_x(y) = f(x) + h(x) \cdot (y - x) + \tilde{c}(x,y)$ and note that as $y \mapsto \tilde{c}(x,y)$ is K-Lipschitz and convex on $\overline{I}$, we have

$$\frac{dv_x}{dy}(b-) - \frac{dv_x}{dy}(a+) \leq 2K. \tag{3.41}$$

Step 1. First we establish the Lipschitz property of g – in the case of bounded I and then for unbounded I.

Case 1: I is bounded. Then $\partial I = \{a, b\}$ are real-valued. Proposition 3.4 tells us that $g(a) \geq 0$, $g(b) \geq 0$ while there is $y \in [a, b]$ such that $g(y) \leq 0$, since $g(y) \leq 0$ whenever $(x, y) \in G$ for some x. As $g(y) = \inf_{x \in X_G}\{v_x(y)\}$, if $(x, y) \in G$ then we have $v_x(a) \geq g(a) \geq 0$, $v_x(b) \geq g(b) \geq 0$ while $v_x(y) = g(y) \leq 0$. With (3.41) this implies that v_x is $2K$-Lipschitz on $[a, b]$ for any $x \in X_G$, hence g is also $2K$-Lipschitz on $[a, b]$. Proposition 3.4 also tells us that $g \leq 0$ on $[a, b]$, $g \geq 0$ on $J \backslash (a, b)$.

Case 2: I is unbounded. In this case we do not have two "pillars" $\{a, b\}$, and this forces us to look into the structure of G more carefully. Recall the Step 1 in the proof of Proposition 3.4. Now let us define

$$s_n = \frac{v_{x_n}(b_n) - v_{x_n}(a_n)}{b_n - a_n}. \tag{3.42}$$

Altering the triple (f,g,h) by an appropriate affine function, we can assume that $f(x_1) = h(x_1) = 0$ and hence

$$v_{x_1} \text{ is } K\text{-Lipschitz and convex.} \tag{3.43}$$

We want to control s_n for all n and the key is, as usual, (3.16). Now we claim that

$$-3K \leq s_n \leq 3K \quad \text{for every } n. \tag{3.44}$$

Note that the claim (3.44) is obviously true for $n = 1$ by the assumption (3.32). Suppose the claim is true for all $k = 1, 2, ..., n$. We will show that the claim is also true for $n+1$. As the situation is symmetric we will only show $-3K \leq s_n$. To see this, we separate into two cases: $a_{n+1} < l_n$, or $l_n \leq a_{n+1} < r_n < b_{n+1}$ (recall (3.8)).

(Case 1 : $a_{n+1} < l_n$) Then in particular $a_{n+1} < a_1$. Now observe that by (3.32) v_{x_1} has slope nowhere smaller than $-K$, and $v_{x_{n+1}}(a_{n+1}) \leq v_{x_1}(a_{n+1})$, $v_{x_{n+1}}(a_1) \leq v_{x_1}(a_1)$. In conjunction with the assumption on the cost $\tilde{c}$, this implies that

$$v'_{x_{n+1}}(a_{n+1}-) \geq -3K$$

i.e. the left-handed limit of the slope of convex function $v_{x_{n+1}}$ is not smaller than $-3K$. Of course, by convexity of $y \mapsto \tilde{c}(x,y)$ this implies $-3K \leq s_{n+1}$.

(Case 2 : $l_n \leq a_{n+1} < r_n < b_{n+1}$) Then there exists $k \in \{1, 2, ..., n\}$ such that $a_k \leq a_{n+1} < b_k < b_{n+1}$. Now observe that the convexity of $y \mapsto \tilde{c}(x,y)$ and the relations $v_{x_k}(a_{n+1}) \geq v_{x_{n+1}}(a_{n+1})$, $v_{x_k}(b_k) \leq v_{x_{n+1}}(b_k)$ clearly imply that

$$s_k \leq s_{n+1}.$$

Hence by induction we have $-3K \leq s_{n+1}$, therefore the claim (3.44) is proved.

Now we will show uniform Lipschitz property of v_x for any $x \in X_G$. Fix $x \in X_G$. As $I =]a,b[$ is unbounded, either $a = -\infty$ or $b = \infty$ (or both). In particular, there exists y such that $(x,y) \in G$ and $n \in \mathbb{N}$ such that $a_n < y < b_n$. Recall $(x_n, a_n), (x_n, b_n) \in G$ and so we have

$$v_x(a_n) \geq v_{x_n}(a_n), \quad v_x(y) \leq v_{x_n}(y), \quad v_x(b_n) \geq v_{x_n}(b_n). \tag{3.45}$$

Define

$$s_x^- = \frac{v_x(y) - v_x(a_n)}{y - a_n}, \quad s_x^+ = \frac{v_x(b_n) - v_x(y)}{b_n - y}.$$

By (3.45) and (3.44) we observe

$$s_x^- \leq \frac{v_{x_n}(y) - v_{x_n}(a_n)}{y - a_n} \leq 3K, \quad s_x^+ \geq \frac{v_{x_n}(b_n) - v_{x_n}(y)}{b_n - y} \geq -3K,$$

and by we have $0 \leq s_x^+ - s_x^- \leq 2K$. This implies that $-5K \leq s_x^- \leq s_x^+ \leq 5K$, and again convexity of $y \mapsto \tilde{c}(x,y)$ and (3.41) imply

$$-7K \leq v'_x(y) \leq 7K \quad \forall y \in I.$$

In view of (2.3), we conclude that g is $7K$-Lipschitz on $\overline{I}$ as desired.

Let us comment further on the case of half-infinite domain I. In order to obtain in addition the sign-changing property of g, as described in Proposition 3.5, we need to alter the triple (f,g,h) by an affine function to satisfy the asymptotic property of v_{x_1} as in (3.32), instead of (3.43). This can be done at the cost of having g being $8K$-Lipschitz on $\overline{I}$.

Step 2. Now we will prove the regularity properties of a dual optimizer for $\tilde{c}$ and (μ, ν) described in Theorem 2.5, where irreducibility of (μ, ν) is not assumed.

In the proof of Theorem 2.3 we showed that there is a dual maximizer (f, g, h) to the problem (1.4) where $g := \inf_{k \geq 0} g_k$. In view of Step 1, we conclude that g is $8K$-Lipschitz on J.

Next, observing the duality relation

$$g(y) - \tilde{c}(x, y) \leq f(x) + h(x) \cdot (y - x) \quad \forall x \in J, \forall y \in J, \tag{3.46}$$

note that we can replace f, h by $\tilde{f}$, $\tilde{h}$ respectively, as follows: define $H : J \times J \to \mathbb{R}$ by the upper concave envelope in y variable

$$H(x, y) := \operatorname{conc}[g(\cdot) - \tilde{c}(x, \cdot)](y).$$

Then we define $\tilde{f}(x) := H(x, x)$ and $\tilde{h}(x) := \frac{\partial H(x,y)}{\partial y}\big|_{y=x}$. More precisely, $\tilde{h}(x)$ is an element of the superdifferential of the concave function $y \to H(x, y)$ at x, and there exists a measurable choice of such an $\tilde{h}$. Now in view of (3.46), it is clear that $(\tilde{f}, g, \tilde{h})$ is a dual maximizer. Observe that since $y \mapsto g(y) - \tilde{c}(x, y)$ is $9K$-Lipschitz, it is immediate that $|\tilde{h}| \leq 9K$ on J. Then, to see that $\tilde{f}$ is Lipschitz, note that since $x \mapsto g(y) - \tilde{c}(x, y)$ is L-Lipschitz, by definition of H we have

$$|H(x, y) - H(x', y)| \leq L|x - x'| \quad \forall x, x', y \in J. \tag{3.47}$$

On the other hand, since the concave envelope of a Lipschitz function is Lipschitz,

$$|H(x, y) - H(x, y')| \leq 9K|y - y'| \quad \forall x, y, y' \in J.$$

These inequalities immediately imply that, for any $x, x' \in J$,

$$|\tilde{f}(x) - \tilde{f}(x')| = |H(x, x) - H(x', x')| \leq 19L|x - x'|.$$

Finally, recall that $\tilde{c}(x, y) = c(x, y) + u(y)$ where u is L-Lipschitz, we replace g with $\tilde{g} = g - u$ which is $17L$-Lipschitz on J. Then $(\tilde{f}, \tilde{g}, \tilde{h})$ is a dual maximizer satisfying the conclusion of Theorem 2.5. □

Remark 3.8. A close look at the above proof shows that the Lipschitz assumption on c is only used in (3.41) and (3.47). Hence, the Lipschitz assumption for c on $J \times J$ in Theorem 2.5 can be weakened as follows: there exists $K, L \geq 0$ such that

(1) (3.41) holds for every irreducible domain $I =]a, b[$ of (μ, ν), and
(2) $|c(x, y) - c(x', y)| \leq L|x - x'|$ for all $x, x', y \in J$.

We conclude this section with the proof of Proposition 3.3.

Proof of Proposition 3.3. Considering the irreducible decomposition of (μ, ν), it is enough to prove the proposition when (μ, ν) are irreducible. Under the assumption there exists a dual optimizer (f, g, h) (in the sense of [5]), so in particular $g(y) - f(x) - h(x)(y - x) \leq c(x, y)$. Let $d(x, y) = c(x, y) - [g(y) - f(x) - h(x)(y - x)]$, which is nonnegative.

Now a result of Jacod and Mémin [15, Proposition 2.4] tells us that if c is bounded and $y \mapsto c(x, y)$ is continuous, then $\mathbb{P} \in \Pi(\mu, \nu) \mapsto \int c \, d\mathbb{P}$ is continuous. Note that while $y \mapsto d(x, y)$ is not continuous, $y \mapsto d(x, y) + u(y) + g(y)$ is continuous. From this, the technique in [5, Remark 7.9] applies and one can change the topology of $\mathbb{R}$ to a somewhat finer (but still Polish) topology under which $y \mapsto d(x, y)$ is continuous, while the compactness of $\mathrm{MT}(\mu, \nu)$ remains unchanged (since the new topology leads to the same Borel structure; see [6, Remark 2.4]).

From this we can show that $\mathbb{P} \mapsto \langle d, \mathbb{P}\rangle := \int d(x,y)\, d\mathbb{P}$ is lower-semicontinuous. For a contradiction, suppose that $\mathbb{P}_n \to \mathbb{P}$, $\langle d, \mathbb{P}\rangle \in \mathbb{R}$ but there exist a subsequence $\{\mathbb{P}_k\}$ of $\{\mathbb{P}_n\}$ and $c > 0$ such that $\langle d, \mathbb{P}_k\rangle + 3c \leq \langle d, \mathbb{P}\rangle$ for all k. Choose a sufficiently large N (by monotone convergence) such that $|\langle d, \mathbb{P}\rangle - \langle d \wedge N, \mathbb{P}\rangle| < c$. By Jacod and Mémin, for all large k we have $|\langle d \wedge N, \mathbb{P}\rangle - \langle d \wedge N, \mathbb{P}_k\rangle| < c$. As $\langle d, \mathbb{P}_k\rangle \geq \langle d \wedge N, \mathbb{P}_k\rangle$ we have $\langle d, \mathbb{P}\rangle - \langle d, \mathbb{P}_k\rangle \leq 2c$, a contradiction. The case $\langle d, \mathbb{P}\rangle = \infty$ is easier.

Hence there exists a minimizer $\mathbb{P}^*$ for $d(x,y)$. Finally, since $\int [g(y) - f(x) - h(x)(y-x)] d\mathbb{P}$ is well-defined and independent of the choice of $\mathbb{P} \in \mathrm{MT}(\mu, \nu)$ (see [5]), we conclude that $\mathbb{P}^*$ is also a minimizer for $c(x,y)$. $\square$

4. Some insightful examples

Example 4.1. This example shows that a growth assumption on the cost in Theorem 2.3, when the marginals μ, ν have unbounded support, is necessary for dual attainment. More precisely, we will construct a cost function which is of quadratic growth in y at $+\infty$ and marginals μ, ν having finite second moments, and show that there is no dual optimizer.

To this end, we begin with a simple construction of a dual optimizer. Let $x_n = n$, $n = 1, 2, 3, \ldots$. For each n, simply set $f(n) = h(n) = 0$ for all n, and define the cost $c(n,y)$ to be

$$\begin{aligned} c(n,y) &= y^2 && \text{if } y \geq n-1, \\ c(n,y) &= (n-1)y && \text{if } y \leq n-1. \end{aligned}$$

Define $g(y) = \inf_{n \in \mathbb{N}} c(n,y)$ and notice that then $g(y) = y^2$ on $[0, \infty[$, and

$$g(y) = -\infty \quad \text{on} \quad]-\infty, 0[. \tag{4.1}$$

Now observe that the triple (f, g, h) supports the set $G = \{(n, n+1), (n, n-1) : n \in \mathbb{N}\}$ in view of (3.6) – (3.7). This implies that any martingale measure $\mathbb{P}$ with $\mathbb{P}(G) = 1$ is optimal in $\mathrm{MT}(\mu, \nu)$ where μ, ν are the marginals of $\mathbb{P}$ having finite second moments. For definiteness, let us construct one such $\mathbb{P}$. Pick a probability measure μ with finite second moment and $\mathrm{supp}(\mu) = \mathbb{N}$. For each $n \in \mathbb{N}$, define a martingale measure $\mathbb{P}$ whose disintegration $(\mathbb{P}_n)_n$ w.r.t. μ is as follows:

$$\mathbb{P}_n = \frac{1}{2}(\delta_{n-1} + \delta_{n+1}) \quad \forall n = 1, 2, \ldots.$$

Let ν be the second marginal of $\mathbb{P}$ and note that ν also has finite second moment. Then as noted, $\mathbb{P}$ is optimal in $\mathrm{MT}(\mu, \nu)$.

Now let $\mu^* = \frac{1}{2}\delta_{-1} + \frac{1}{2}\mu$, $\nu^* = \frac{1}{2}\delta_{-1} + \frac{1}{2}\nu$, and $\mathbb{P}^* = \frac{1}{2}\delta_{(-1,-1)} + \frac{1}{2}\mathbb{P}$. Then $\mathbb{P}^* \in \mathrm{MT}(\mu^*, \nu^*)$ and it is clear that $\mathbb{P}^*$ is optimal (no matter how $y \mapsto c(-1, y)$ is defined, but for definiteness let $c(-1, y) = 0$). We will show that there is no dual optimizer. For a contradiction, suppose that (f^*, g^*, h^*) is a dual optimizer. Then we claim that (f^*, g^*, h^*) is essentially the same as the above (f, g, h); they should differ only by an affine function due to the structure of $\mathbb{P}$. To see this, as usual let $v_n(y) = f^*(n) + h^*(n)(y - n) + c(n, y)$. Then (3.6) – (3.7) gives, for every $n \in \mathbb{N}$,

$$\begin{aligned} &v_n(n+1) \leq v_{n+1}(n+1), \quad v_n(n) \geq v_{n+1}(n), \quad v_n(n+2) \geq v_{n+1}(n+2) \quad \iff \\ &h^*(n) \leq f^*(n+1) - f^*(n) \leq h^*(n+1), \quad f^*(n+1) - f^*(n) \leq 2h^*(n) - h^*(n+1). \end{aligned}$$

Observe that this implies $h^*(n) = f^*(n+1) - f^*(n)$, and this in turn implies that $h^* = C$ is constant on $\mathbb{N}$ and $f^*(n) = Cn + D$ for some constant D, thus

$f^*(n) + h^*(n)(y-n) = Cy + D$. In view of (4.1), this implies that $g^*(-1) = -\infty$, a contradiction to the fact that $g^*(-1)$ must be real-valued.

Example 4.2. In this example, we show that if the convexity assumption on $y \mapsto c(x,y)$ holds only locally around x, then the dual maximizer can fail to exist.

Let $y_0 = 0$, $y_n = \sum_{k=1}^n \frac{1}{k^2}$, $x_n = (y_{n-1} + y_n)/2$, and $x_\infty = y_\infty = \sum_{k=1}^\infty \frac{1}{k^2}$. Define the cost function by

$$c(x_\infty, y) = 0, \quad \text{and} \quad c(x_n, y) = \begin{cases} 0 \quad \text{if} \quad y \in [y_{n-1}, y_n], \\ y - y_{n-1} \quad \text{if} \quad y \le y_{n-1}, \\ -y + y_n \quad \text{if} \quad y \ge y_n. \end{cases}$$

Let μ be any probability measure whose support is $\{x_n\}_{1 \le n \le \infty}$, and construct a martingale measure $\mathbb{P}$ whose disintegration $(\mathbb{P}_x)_x$ w.r.t. μ is as follows:

$$\mathbb{P}_{x_n} = \frac{1}{2}(\delta_{y_{n-1}} + \delta_{y_n}) \quad \forall n = 1, 2, \dots \quad \text{and} \quad \mathbb{P}_{x_\infty} = \delta_{x_\infty},$$

and define ν as the second marginal of $\mathbb{P}$. Notice that then $\mathbb{P}$ is the unique element in $\mathrm{MT}(\mu, \nu)$. Now we will show that this (optimal) $\mathbb{P}$ does not allow a dual maximizer, that is, there does not exist a triple (f, g, h) which satisfies the following:

$$g(y) \le c(x_n, y) + f(x_n) + h(x_n) \cdot (y - x_n) \quad \forall n \in \mathbb{N} \cup \{\infty\}, \ \forall y \in \mathbb{R}, \tag{4.2}$$

$$g(y_n) = c(x_n, y_n) + f(x_n) + h(x_n) \cdot (y_n - x_n) \quad \forall n \ge 1, \tag{4.3}$$

$$g(y_{n-1}) = c(x_n, y_{n-1}) + f(x_n) + h(x_n) \cdot (y_{n-1} - x_n) \quad \forall n \ge 1, \tag{4.4}$$

$$g(y_\infty) = c(x_\infty, y_\infty) + f(x_\infty) + h(x_\infty) \cdot (y_\infty - x_\infty) = f(x_\infty). \tag{4.5}$$

Recall that once such a (f, g, h) exists, then we can redefine g as follows:

$$g(y) := \inf_{n \in \mathbb{N} \cup \{\infty\}} \big(c(x_n, y) + f(x_n) + h(x_n) \cdot (y - x_n) \big). \tag{4.6}$$

We claim that, if we have such a (f, g, h), then we must have

$$g(y_\infty) = -\infty,$$

which is a contradiction to (4.5). To see this, for convenience let us define

$$v_x(y) := c(x, y) + f(x) + h(x) \cdot (y - x). \tag{4.7}$$

Then by (4.2), (4.3), (4.4) we must have

$$v_{x_n}(y_n) = v_{x_{n+1}}(y_n), \quad \text{and} \tag{4.8}$$

$$v_{x_n}(y_{n-1}) \le v_{x_{n+1}}(y_{n-1}), \quad \forall n \ge 1. \tag{4.9}$$

Notice that these with the definition of $c(x_n, y)$ immediately implies

$$h(x_n) \ge h(x_{n+1}) + 1, \quad \forall n \ge 1. \tag{4.10}$$

Also notice that $g(y)$ is a piecewise linear function on $[0, y_\infty[$, and in fact $g(y) = f(x_n) + h(x_n) \cdot (y - x_n)$ on $[y_{n-1}, y_n]$. Hence by (4.10) and the fact $\sum_n \frac{1}{n} = \infty$ and the concavity of g, we see that

$$\begin{aligned} g(y_\infty) &= \lim_{y \nearrow y_\infty} g(y) = g(0) + \sum_{n=1}^\infty \big(g(y_n) - g(y_{n-1}) \big) \\ &= g(0) + \sum_{n=1}^\infty \frac{h(x_n)}{n^2} \le g(0) + \sum_{n=1}^\infty \frac{h(x_1) - (n-1)}{n^2} = -\infty \end{aligned}$$

a contradiction to the fact that $g(y_\infty)$ must be real-valued.

Example 4.3. In this example, we show that the C^2 regularity assumption in Theorem 2.3 cannot be weakened to C^r regularity with $r < 2$. More precisely, for any $1 < r < 2$, we construct a cost function $c \in C^r$ and compactly supported marginals $\mu \preceq \nu$ for which the dual attainment fails. This example shall be a slight modification of the previous one. First, let

$$c(x, y) = -|x - y|^r, \tag{4.11}$$

and choose s such that

$$s > 1 \quad \text{and} \quad sr < 2. \tag{4.12}$$

Let $y_0 = 0$, $y_n = \sum_{k=1}^n \frac{1}{n^s}$, $x_n = (y_{n-1} + y_n)/2$, and $x_\infty = y_\infty = \sum_{k=1}^\infty \frac{1}{n^s}$. Define a martingale measure $\mathbb{P}$ and its marginals μ, ν as in Example 4.2. Note that μ, ν are compactly supported since $s > 1$. Again we will show that this (optimal) $\mathbb{P}$ does not allow a dual maximizer, that is, there does not exist a triple (f, g, h) which satisfies (4.2), (4.3), (4.4), (4.5), where g is given as in (4.6). Again we will show that $g(y_\infty) = -\infty$, which is a contradiction to (4.4). To see this, again define v_x as in (4.7) so that we have (4.8), (4.9). Next, let us consider the slope

$$b_n = \frac{v_{x_n}(y_n) - v_{x_n}(y_{n-1})}{y_n - y_{n-1}}. \tag{4.13}$$

In order to estimate b_n, we will first estimate $b_n - b_{n+1}$. For this, as we can modify the (f, g, h) by an affine function, we can assume that $f(x_{n+1}) = h(x_{n+1}) = 0$, thus without loss of generality we can assume that $b_{n+1} = 0$. Now notice that, using (4.8), (4.9), we have the following inequality:

$$\begin{aligned}
b_n - b_{n+1} &\geq \frac{v_{x_{n+1}}(y_n) - v_{x_{n+1}}(y_{n-1})}{y_n - y_{n-1}} \\
&= n^s[-|y_n - x_{n+1}|^r + |y_{n-1} - x_{n+1}|^r] \\
&= n^s\left[-\left(\frac{1}{2(n+1)^s}\right)^r + \left(\frac{1}{2(n+1)^s} + \frac{1}{n^s}\right)^r\right] \\
&= n^s\left[\frac{(2(n+1)^s + n^s)^r - n^{sr}}{2n^{sr}(n+1)^{sr}}\right] \\
&\approx Cn^s \cdot n^{sr} \cdot n^{-2sr} = Cn^{s-sr}. \quad \text{Hence we deduce that}
\end{aligned}$$

$$-b_n = \sum_{k=0}^{n-1}(b_k - b_{k+1}) \gtrapprox Cn^{1+s-sr}. \quad \text{This implies that, since } sr < 2,$$

$$(y_n - y_{n-1})b_n \lessapprox -Cn^{1-sr} \quad \Longrightarrow \quad \sum_{n=1}^\infty (y_n - y_{n-1})b_n = -\infty.$$

Again as in Example 4.2, this tells us that $g(y_\infty) = -\infty$, a contradiction to (4.5).

Example 4.4. In this example we show the necessity of semiconvexity for the regularity in Theorem 2.5 and the Lipschitzness of c alone is not sufficient, by constructing a 1-Lipschitz cost c and a compactly supported, irreducible pair (μ, ν) for which (f, g, h) is a dual maximizer, but $g \notin L^1(\nu)$.

To do this, we take $c(x,y) = -|x-y|$, and let $I =]0,1[$ and $\mu = \mathrm{Leb}\big|_{[0,1]}$. Choose a smooth and strictly concave function $\xi : I \to \mathbb{R}_-$ such that $\xi \le 0$, $\xi(\frac{1}{2}) = 0$, $\lim_{x \to 0^+} \xi(x) = \lim_{x \to 1^-} \xi(x) = -\infty$, and $\int_0^1 \xi(x)\mu(dx) = -\infty$. Now we will construct a probability measure ν where $\nu(I) = 1$ and (μ, ν) are irreducible, and also find a dual maximizer (f, g, h) where $g = \xi$. Then $\int g(x)\nu(dx) \le \int g(x)\mu(dx) = -\infty$, as claimed.

To construct such ν and (f, g, h), observe that for each $x \in I$ there exist unique $f(x), h(x)$ such that the function $v_x(y) := f(x) + h(x) \cdot (y - x) - |x - y|$ satisfies

(1) $v_x(y) \ge \xi(y) \quad \forall x \in I, \forall y \in I$, and
(2) for each $x \in I$, v_x is tangent to ξ at two points, say $y^-(x), y^+(x)$.

Note that then y^-, y^+ are well-defined on I, and $0 < y^-(x) < x < y^+(x) < 1$. Define a probability measure $\mathbb{P}_x := \frac{y^+(x)-x}{y^+(x)-y^-(x)}\delta_{y^-(x)} + \frac{x-y^-(x)}{y^+(x)-y^-(x)}\delta_{y^+(x)}$, and $\mathbb{P} \in P(\mathbb{R}^2)$ by $\mathbb{P}(dx, dy) = \mathbb{P}_x(dy) \cdot \mu(dx)$, i.e. $(\mathbb{P}_x)_x$ is a disintegration of $\mathbb{P}$ with respect to μ. Define ν as the second marginal of $\mathbb{P}$ and note that by definition of $\mathbb{P}$, (μ, ν) are irreducible and are concentrated on I. Now observe that the definition of f, h gives us that $g(y) := \inf_{x \in I}\{v_x(y)\}$ satisfies $g = \xi$ so that $\int g(x)\nu(dx) = -\infty$, and (f, g, h) is a dual maximizer with respect to μ, ν and c.

Although the examples presented so far indicate that the assumptions made in Theorems 2.3 and 2.5 cannot be simply relaxed, we do not claim that they are necessary-and-sufficient condition for the dual attainment under all circumstances. They might be other, possibly substantially different, sets of assumptions which would also imply the dual attainment. We feel that our work clearly shows that the dual attainment and its regularity is indeed an interesting and delicate problem in martingale optimal transport.

References

[1] M. Beiglböck, A. Cox, and M. Huesmann. Optimal transport and Skorokhod embedding. *Invent. Math.*, 208(2):327–400, 2017.

[2] 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.

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

[4] M. Beiglböck and M. Nutz. Martingale inequalities and deterministic counterparts. *Electron. J. Probab.*, 2014.

[5] M. Beiglböck, M. Nutz, and N. Touzi. Complete Duality for Martingale Optimal Transport on the Line. *Ann. Probab.*, 45(5):3038–3074, 2017.

[6] M. Beiglböck and A. Pratelli. Duality for rectified cost functions. *Calc. Var. Partial Differential Equations*, 45(1-2):27–41, 2012.

[7] L. Campi, I. Laachir, and C. Martini. Change of numeraire in the two-marginals martingale transport problem. *Finance Stoch.*, 21(2):471–486, June 2017.

[8] Y. Dolinsky and H. M. Soner. Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Relat. Fields*, 160(1-2):391–427, 2014.

[9] 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.

[10] N. Ghoussoub, Y.-H. Kim, and T. Lim. Structure of optimal martingale transport plans in general dimensions. arXiv:1508.01806, 2016.

[11] P. Henry-Labordère, J. Obłój, P. Spoida, and N. Touzi. The maximum maximum of a martingale with given n marginals. *Ann. Appl. Probab.*, 26(1):1–44, 2016.

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

[13] D. Hobson and M. Klimmek. Robust price bounds for the forward starting straddle. *Finance Stoch.*, 9(1):189–214, April 2015.

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

[15] J. Jacod and J. Mémin. Sur un type de convergence intermédiaire entre la convergence en loi et la convergence en probabilité. In *Séminaire de Probabilités XV*, volume 850 of *Lecture Notes in Math.*, pages 529–546. Springer, Berlin-New York, 1981.

[16] M. Nutz, F. Stebegg, and X. Tan. Multiperiod Martingale Transport. *ArXiv e-prints*, March 2017.

[17] J. Obłój. The Skorokhod embedding problem and its offspring. *Probab. Surv.*, 1:321–390, 2004.

[18] J. Obłój, P. Spoida, and N. Touzi. Martingale inequalities for the maximum via pathwise arguments. In *In Memoriam Marc Yor-Séminaire de Probabilités XLVII*, pages 227–247. Springer, 2015.

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