On Duality of an Optimal Transport Problem with Backward Martingale Constraint

Yan Xu*

July 10, 2019

Abstract

We continue to study the backward optimal transport problem with backward martingale constraint introduced in [9]. Unlike in [9] where the problem is investigated from the *primal* side, here we approach the problem from a *dual* perspective. We establish existence of dual optimizers and absence of a duality gap. Moreover, we show that a first order condition of the dual problem is given precisely by the martingale property of a special transport map. For future reference, we also establish continuity of the value function with respect to the 2-Wasserstein metric.

1 Introduction

In [9], motivated by the classical Kyle(1985) model [10], we study the following optimal transport problem:

$$\text{minimize} \int c(x,y)d\gamma \quad \text{over} \quad \gamma \in \Gamma(\nu), \tag{1}$$

where $c = c(x,y)$ is the *covariance-type* cost function

$$c(x,y) = (x_1 - y_1)(x_2 - y_2), \quad x, y \in \mathbb{R}^2,$$

ν is a given probability measure on $\mathbb{R}^2$, and $\Gamma(\nu)$ is the family of probability measures $\gamma = \gamma(dx,dy)$ on $\mathbb{R}^2 \times \mathbb{R}^2$ that have ν as their y-marginal: $\gamma(\mathbb{R}^2, dy) = \nu(dy)$, and make a martingale out of the canonical process

*Carnegie Mellon University, Department of Mathematical Sciences, 5000 Forbes Avenue, Pittsburgh, PA, 15213-3890, US.. *Email:* yanx1@andrew.cmu.edu

(X,Y): $\mathbb{E}^{\gamma}(Y|X) = X$. This problem differs from the standard single-period martingale optimal transport problem studied in [3],[4], [6], and [7], among others, as the initial marginal $\mu(dx) = \gamma(dx, \mathbb{R}^2)$ is part of the solution. In [9], we establish existence of solutions and provide equivalent characterizations of optimality in terms of the geometry of their supports. For more details and references, we refer to this paper.

Problem (1) is a convex optimization problem , as the objective function is linear and the constraint space is convex. In this sense, it can be regarded as a *primal* formulation, who admits the following *dual* formulation:

$$\text{maximize} \int \phi_G(y)\, d\nu, \quad G \in \mathfrak{M}, \tag{2}$$

where $\mathfrak{M}$ is the family of maximal monotone subsets of $\mathbb{R}^2$, and the function ϕ_G is given by

$$\phi_G(y) = \inf_{x \in G} c(x, y), \quad y \in \mathbb{R}^2.$$

Unlike in [9] where our main focus is on problem (1), the goal of this paper is to establish a complete duality theory (i.e., existence of dual optimizers and absence of a duality gap) in this setup.

Duality principle is a common phenomenon in the study of the optimal transport. In the classical unconstrained case, the Kantorovich duality theorem establishes a strong duality relation for lower semi-continuous cost functions, as well as existence of primal and dual optimizers provided that the value function is finite (see, for instance, [12, Theorem 5.10]). In the case with an additional martingale constraint, a complete duality theory is obtained in [4] for general measurable cost functions.

Our main results are Theorem 2.1 and Theorem 3.1. In Theorem 2.1, we show that the *dual* problem (2) admits a solution. Our arguments rely on the local compactness of the space of non-empty compact sets equipped with the Hausdorff metric. In Theorem 3.1, we prove a strong duality relation, which was obtained in [9] as an immediate corollary of the geometric characterizations of solutions of the *primal* problem (1) (i.e., [9, Theorem 2.2]). Here, our proof instead use a well-known minimax principle due to Aubin and Ekeland [2, Theorem 6.2.7].

Furthermore, in the case where the measure ν is regular, we obtain a first order condition associated with the *dual* problem (2), which turns out to be the martingale property of a particular transport map constructed in [9]. We also prove continuity of the value function with respect to the 2-Wasserstein metric, which will be used frequently in our study for the multi-period case.

2 Dual Problem

We follow closely the notations in [9]. We shall write a point in $\mathbb{R}^4 = \mathbb{R}^2 \times \mathbb{R}^2$ as (x, y), where $x = (x_1, x_2)$ and $y = (y_1, y_2)$ belong to $\mathbb{R}^2$. We denote by X (resp. Y) the cannonical projections of $\mathbb{R}^2 \times \mathbb{R}^2$ Onto the x- (resp. y-) coordinates. A set $G \subset \mathbb{R}^2$ is called *monotone* if

$$c(x, y) \geq 0, \quad x, y \in \mathbb{R}^2,$$

where $c = c(x, y)$ is the *covariacne-type* function given by

$$c(x, y) = (x_1 - y_1)(x_2 - y_2), \quad x, y \in G.$$

We call a monotone set *maximal* if it is not a proper subset of another monotone set, and denote by $\mathfrak{M}$ the collection of maximal monotone sets in $\mathbb{R}^2$.

We denote by $\mathcal{P}_2(\mathbb{R}^2)$ the family of Borel probability measures on $\mathbb{R}^2$ with finite second moment. For $\nu \in \mathcal{P}_2(\mathbb{R}^2)$, we denote by $\Gamma(\nu)$ the family of probability measures $\gamma = \gamma(dx, dy) \in \mathcal{P}_2(\mathbb{R}^2 \times \mathbb{R}^2)$ that have ν as their y-marginal and under which the canonical process (X, Y) becomes a *1-step martingale*:

$$\Gamma(\nu) \triangleq \left\{ \gamma \in \mathcal{P}_2(\mathbb{R}^2 \times \mathbb{R}^2) : \ \gamma(\mathbb{R}^2, dy) = \nu(dy) \text{ and } \mathbb{E}^{\gamma}(Y|X) = X \right\}.$$

The *backward* martingale transport problem, introduced in [9], is to

$$\text{minimize} \quad \int c(x, y)\, d\gamma \quad \text{over} \quad \gamma \in \Gamma(\nu). \tag{3}$$

This is the problem in its *primal* formulation, whose *dual* formulation is to

$$\text{maximize} \quad \int \phi_G \, d\nu \quad \text{over} \quad G \in \mathfrak{M}, \tag{4}$$

where the function ϕ_G is given by

$$\phi_G(y) = \inf_{x \in G} c(x, y), \quad y \in \mathbb{R}^2.$$

In this section, we seek to (1) establish existence of dual maximizer, *without* refering to problem (3); and (2) obtain martingale property of the map defined in Theorem 4.3 of [9] as the first order condition of the dual problem (4). Our arguments for existence rely on local compactness of the space of non-empty compact sets equipped with the Hausdorff metric, which will be discussed in the sequel.

2.1 Existence of Dual Optimizer

2.1.1 Some preliminaries

Let (X, d) be a metric space and let $\mathcal{C}(X)$ be the space of closed subsets of X:

$$\mathcal{C}(X) \triangleq \{A \subset X : \ X \text{ is closed}\}.$$

For $x \in X$ and $A \in \mathcal{C}(X)$, we write $d(x, A) = \min_{y \in A} d(x, y)$. A sequence $(A_n) \subset \mathcal{C}(X)$ is said to be *Wijsman convergent* to a set $A \in \mathcal{C}(X)$, denoted $A_n \xrightarrow{\mathrm{W}} A$, if for each $x \in X$,

$$d(x, A_n) \to d(x, A).$$

For $A, B \in \mathcal{C}(X)$, the *Hausdorff distance* between A and B is defined by

$$\delta_H(A, B) \triangleq \max\{\sup_{x \in A} d(x, B), \sup_{y \in B} d(y, A)\}.$$

Notice that

$$\delta_H(A, B) = \sup_{x \in X} |d(x, A) - d(x, B)|.$$

In this sense, Hausdorff convergence is to uniform convergence as Wijsman convergence is to pointwise convergence.

When restricting our attention to $\mathcal{K}(X)$, be the space of non-empty compact subsets of X:

$$\mathcal{K}(X) \triangleq \{A \subset X : \ A \text{ is non-empty and compact}\}.$$

the Hausdorff distance δ_H becomes a metric. It is well-known that $(\mathcal{K}(X), \delta_H)$ is a compact Polish space if (X, d) is so (see, for instance, [11, Proposition 2.4.15 and Proposition 2.4.17]). In particular, when X is compact, convergence in Hausdorff distance is equivalent to the so-called *Kuratowski* convergence. Namely, a sequence $(A_n) \subset \mathcal{K}(X)$ is said to converge to a set A in *Kuratowski* sense, denoted $A_n \xrightarrow{\kappa} A$, if

(1) every cluster point of a sequence (x_n) with $x_n \in A_n$ belongs to A, and

(2) for all $x \in A$, there are $x_n \in A_n$ such that $x_n \to x$.

2.1.2 Main Results and Proofs

We are now ready to prove the following theorem:

Theorem 2.1. *There is a solution to problem* (4).

We divide the proof into some lemmas.

Lemma 2.2. *Let (G_n) be a sequence of maximal monotone subsets of $\mathbb{R}^2$ and let $G \subset \mathbb{R}^2$ be non-empty. Suppose $G_n \xrightarrow{W} G$. Then G is also a maximal monotone set.*

Proof. Let $x, y \in G$. For each n, we can find $x_n, y_n \in G_n$ such that

$$d(x, G_n) = d(x, x_n) \quad \text{and} \quad d(y, G_n) = d(y, y_n).$$

As (G_n) is Wijsman convergent to G, we deduce that $x_n \to x$ and $y_n \to y$. Thus,

$$c(x, y) = \lim_{n\to\infty} c(x_n, y_n) \geq 0.$$

This shows that G is a monotone set.

Next, suppose G is *not* maximal. Then there exist $x \in \mathbb{R}^2 \setminus G$ such that $G \cup \{x\}$ is monotone. As $G_n \xrightarrow{\mathrm{W}} G$, we can assume that there exist positive numbers r and R such that

$$r \leq d(x, G_n) \leq R, \quad \text{for all } n.$$

Since each G_n is a maximal monotone set, it intersects with the diagnal set $D(x) \triangleq \left\{z = (z_1, z_2) \in \mathbb{R}^2 : \ z_1 - x_1 = x_2 - z_2\right\}$ at a unique point x_n. In particular, (x_n) is a bounded sequence and thus admits a subsequence, not relabelled, that converges to some point $x' \in D(x)$. As $\limsup_n d(G_n, x') \leq \lim_{n\to\infty} d(x_n, x') = 0$, we deduce that

$$x' \in G \cap D(x) = \{x\}.$$

This gives us a contradiction. □

A sequence $(A_n) \subset \mathbb{R}^2$ is said to *escape to infinity* if for any $R > 0$, there is $n \in \mathbb{N}$ such that $A_n \cap B_R = \emptyset$, where B_R is the closed ball centered at the origin with radius R.

Lemma 2.3. *Let $(G_n) \subset \mathfrak{M}$ be a maximising sequence for problem* (4)*:*

$$\lim_{n\to\infty} \int \phi_{G_n} \, d\nu = \inf_{G\in\mathfrak{M}} \int \phi_G \, d\nu.$$

Then (G_n) dose not escape to infinity.

Proof. Suppose not. For each $k \in \mathbb{N}$, we can find n_k large enough such that $G_{n_k} \cap B_{2k} = \emptyset$. Note that if $y \in B_k$ and if G is a maximal monotone set such that $G \cap B_{2k} = \emptyset$, then

$$\phi_G(y) = \inf_{x \in G} c(x, y) \le -\frac{k^2}{2}.$$

Therefore, we have

$$\begin{aligned}\int \phi_{G_{n_k}}(y)\, d\nu &\le \int \phi_{G_{n_k}}(y) 1_{\{y \in B_k\}}\, d\nu \\ &\le -\frac{k^2}{2}\nu(B_k) \to -\infty, \quad k \to \infty.\end{aligned}$$

This is a contradiction. $\square$

Lemma 2.4. *Let (G_n) and G be maximal monotone sets in $\mathbb{R}^2$. Suppose that $G_n \xrightarrow{W} G$. Then, for every y,*

$$\limsup_n \phi_{G_n}(y) \le \phi_G(y). \tag{5}$$

Proof. Fix $y \in \mathbb{R}^2$ and assume that $\phi_G(y) > -\infty$. By extracting a subsequence, not relabelled, we may assume that

$$\limsup_n \phi_{G_n}(y) = \lim_{n \to \infty} \phi_{G_n}(y).$$

For each $k \in \mathbb{N}$, find $x^k \in G$ such that $\phi_G(y) \le c(x^k, y) \le \phi_G(y) + \frac{1}{k}$. As $G_n \xrightarrow{\mathrm{W}} G$, we can find $x^{n_k} \in G_{n_k}$ such that

$$|x^{n_k} - x^k| \le \frac{1}{k(1 + |x^k| + |y|)}, \quad k \in \mathbb{N}.$$

It follows that

$$\begin{aligned}\phi_G(y) &\ge -|c(x^k, y) - c(x^{n_k}, y)| + c(x^{n_k}, y) - \frac{1}{k} \\ &\ge c(x^{n_k}, y) - \frac{3}{k} \ge \phi_{G_{n_k}}(y) - \frac{3}{k}.\end{aligned}$$

This readily implies (5). The case where $\phi_G(y) = -\infty$ is similar. $\square$

Proof of Theorem 2.1. We proceed by a diagonalization argument. Let (G_n) be a maximising sequence for the dual problem. From Lemma 2.3, we know there exists $R > 0$ such that

$$G_n \cap B_R \neq \emptyset, \quad \forall n \in \mathbb{N}.$$

From our discussion in section 2.1.1, we know the space $(\mathcal{K}(B_{R+1}), \delta_H)$ is a compact metric space. Thus, there is a subsequence $(s_1(n)) \subset \mathbb{N}$ and a non-empty compact set $\widehat{G}_1$ such that

$$B_{R+1} \cap G_{s_1(n)} \xrightarrow{\delta_H} \widehat{G}_1.$$

Next, as $(\mathcal{K}(B_{R+2}), \delta_H)$ is a compact metric space, we can extract a further subsequence $(s_2(n)) \subset (s_1(n))$ and find a non-empty compact set $\widehat{G}_2$ such that

$$B_{R+2} \cap G_{s_2(n)} \xrightarrow{\delta_H} \widehat{G}_2.$$

In particular, we have $\widehat{G}_2 \cap B_{R+1} = \widehat{G}_1$.

We can continue this procedure to obtain

- a nested sequence of indices: $\cdots \subset (s_k(n)) \subset (s_{k-1}(n)) \subset \cdots \subset (s_1(n))$, and
- a sequence of non-empty compact sets $(\widehat{G}_k)$ with $\widehat{G}_k \subset B_{R+k}$ for each k,

such that

$$\begin{aligned} &B_{R+k} \cap G_{s_k(n)} \xrightarrow{\delta_H} \widehat{G}_k \\ &\widehat{G}_{k+1} \cap B_{R+k} = \widehat{G}_k, \quad k \in \mathbb{N}. \end{aligned}$$

Hereafter, we write G_n in place of $G_{s_n(n)}$ for simplicity. Define $\widehat{G} = \bigcup_{k=1}^{\infty} \widehat{G}_k$. For any k, we have

$$B_{R+k} \cap G_n \xrightarrow{\delta_H} B_{R+k} \cap \widehat{G}, \quad \text{as } n \to \infty$$

In particular, (G_n) is Wijsman convergent to $\widehat{G}$. We then deduce from Lemma 2.3 that $\widehat{G}$ is a maximal monotone set.

As $\phi_{\widehat{G}}$ and ϕ_{G_n} are non-positive, the result now follows from Lemma 2.4 and the reverse Fatou lemma. □

Remark 2.5. An unpleasant fact of problem (4) is that the objective is non-linear and the admissible set $\mathfrak{M}$ is non-convex. To remedy the situation, [9] considers the family Φ_M that contains functions dominated by element of $\mathfrak{M}$. More precisely, $\phi \in \Phi_M$ if and only if there is $G \in \mathfrak{M}$ such that $\phi \leq \phi_G$. In this relaxed formulation, the problem becomes linear and the constraint set becomes convex. In particualr, existence of optimizer can be obtained through Komlos lemma (see, for instance, [5, Lemma A1.1]) and a characterization of Φ_M (see [9, Lemma 2.1]).

2.2 Martingale Property as First Order Condition

This section is compliment to section 4 of [9]. We start by introducing the notion of optimal map. Let $Y = (Y_1, Y_2)$ be 2-dimensional random variable having a finite second moment: $Y \in \mathcal{L}^2 = \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. As usual, we identify random variables that differ only on a set of measure zero. The optimal *map* problem is to

$$\text{minimize} \quad \mathbb{E}\left(c(X,Y)\right) \quad \text{over} \quad X \in \mathcal{X}(Y) \tag{6}$$

for the same cost function $c(x,y) = (x_1 - y_1)(x_2 - y_2)$ and the constraint

$$\mathcal{X}(Y) \triangleq \left\{ X = (X_1, X_2) \in \mathcal{L}^2 : \ X \text{ is } Y\text{-measurable and } \mathbb{E}\left(Y \middle| X\right) = X \right\}.$$

We denote $\nu = \text{Law}(Y)$ and observe that $\text{Law}(X,Y) \in \Gamma(\nu)$ for every $X \in \mathcal{X}(Y)$. Thus, optimal *plan* problem (3) may be viewed as a Kantorovich-type relaxation of optimal *map* problem (6).

We follow notations used in section 4 of [9] related to a function $\phi = \phi_G$, where $G \in \mathfrak{M}$. In particular, $D^c \triangleq (D^c_1, D^c_2)$ stands for the differential operator associated with the cost function $c = c(x,y)$:

$$D^c_1\phi(y) \triangleq y_1 - \frac{\partial \phi}{\partial y_2}(y), \quad D^c_2\phi(y) \triangleq y_2 - \frac{\partial \phi}{\partial y_1}(y), \quad y \in \operatorname{dom} \nabla\phi,$$

where $\operatorname{dom}\nabla\phi$ is the set of points where ϕ is differentiable. We denote by $E^G = E^G_1 \cup E^G_2$ the union of the vertical and horizontal line segments of G:

$$\begin{aligned} E^G_i(t) &= \left\{x = (x_1, x_2) \in G : \ x_i = t\right\}, \quad t \in \mathbb{R}, \\ \mathcal{T}^G_i &= \left\{t \in \mathbb{R} : \ E^G_i(t) \text{ has more than one point}\right\}, \\ E^G_i &= \bigcup_{t \in \mathcal{T}^G_i} E^G_i(t), \quad i = 1, 2, \end{aligned} \tag{7}$$

Clearly, the sets $(\mathcal{T}_i^G)$ are countable at most. Define a map by

$$X = Y 1_{\{Y \in E\}} + D^c \phi(Y) 1_{\{Y \notin E\}}. \tag{8}$$

The goal of this section is to obtain the martingale property of (X, Y): $\mathbb{E}\,(Y \mid X) = X$, as the first order condition of the dual problem (4). We need some more notations. Define

$$\begin{aligned}
\mathrm{Arg}_G(y) &\triangleq \arg\min_{x \in G} c(x,y) = \{x \in G:\ \phi_G(y) = c(x,y)\}\,,\\
\mathrm{dom}\,\mathrm{Arg}_G &\triangleq \left\{y \in \mathbb{R}^2:\ \mathrm{Arg}_G(y) \neq \emptyset\right\},\\
\widehat{\mathrm{dom}}\,\mathrm{Arg}_G &\triangleq \{y \in \mathrm{dom}\,\mathrm{Arg}_G:\ \mathrm{Arg}_G(y) \text{ is a singleton}\}\,.
\end{aligned}$$

For $y \in \mathrm{dom}\,\phi$, we call a sequence $(x^n) \subset G$ a minimizing sequence of y if $\phi(y) = \lim_n c(x^n, y)$. We denote by S the collection of points y in $\mathrm{dom}\,\mathrm{Arg}_G \setminus G$ that have an unbounded minimizing sequence. Hereafter, we shall simply write

$$\mathrm{Arg} = \mathrm{Arg}_G, \quad E_i = E_i^G, \quad E_i(t) = E_i^G(t), \quad \mathcal{T}_i = \mathcal{T}_i^G.$$

as the set G is fixed.

Lemma 2.6. *The set S is contained in $\partial\,\mathrm{dom}\,\phi$ and has at most two elements.*

Proof. If $y \in \mathrm{int}\,\mathrm{dom}\,\phi$, then the lines $\{z : z_1 = y_1\}$ and $\{z : z_2 = y_2\}$ intersect G at x^1 and x^2, and any minimizing sequence of y will be contained in the segment of G bounded by x^1 and x^2. Thus, $S \subset \partial\,\mathrm{dom}\,\phi$.

From Lemma B.2 of [9], we have that $\mathrm{int}\,\mathrm{dom}\,\phi = (a_1, b_1) \times (a_2, b_2)$, where $-\infty \le a_i < b_i \le \infty$ and (a_i, b_i) is the interior of the projection of G on the x_i-coordinate. Without loss of generality we can assume that $a_1 > -\infty$. Let $y^0, y^1 \in S$ be distinct and assume that $y_1^0 = y_1^1 = a_1$, $y_2^0 < y_2^1 \le b_2$. Let $x^i \in \mathrm{Arg}(y^i), i = 0, 1$. Note that

$$\begin{aligned}
\phi(y^1) \le c(x^0, y^1) &= (x_1^0 - y_1^1)(x_2^0 - y_2^1)\\
&= (x_1^0 - y_1^0)(x_2^0 - y_2^0) + (x_1^0 - y_1^0)(y_2^0 - y_2^1)\\
&< c(x^0, y^0) = \phi(y^0)
\end{aligned}$$

If (z^n) is an unbounded minimizing sequence of y^1, then

$$\lim_{n \to \infty} z_1^n = a_1 \text{ and } \lim_{n \to \infty} z_2^n = -\infty.$$

It follows that

$$\begin{aligned}
\phi(y^0) &\leq \limsup_{n\to\infty} c(z^n, y^0) \\
&= \limsup_{n\to\infty} \left\{(z_1^n - y_1^1)(z_2^n - y_2^1) + (z_1^n - y_1^1)(y_2^1 - y_2^0)\right\} \\
&= \lim_{n\to\infty} c(z^n, y^1) = \phi(y^1),
\end{aligned}$$

and we obtain a contradiction. $\square$

Let $\mathcal{D}$ be the family of graphs of strictly decreasing functions $h = h(t)$ defined on closed intervals of $\mathbb{R}$ such that both h and its inverse h^{-1} are Lipschitz functions:

$$\frac{1}{K}(t-s) \leq h(s) - h(t) \leq K(t-s), \quad s < t,$$

for some constant $K = K(h) > 0$.

Definition 2.7. A Borel probability measure μ on $\mathbb{R}^2$ is $\mathcal{D}$-*regular* if $\mu(D) = 0$, $D \in \mathcal{D}$.

The standing assumption of this section is that $\nu = \mathrm{Law}(Y)$ is $\mathcal{D}$-regular. We now state the main result:

Theorem 2.8. *X given by (8) is a martingale map: for $i = 1, 2$,*

$$\int h(D^c\phi(y))(D_i^c\phi(y) - y_i)\, d\nu = 0, \quad h \in C_b(\mathbb{R}^2).$$

Proof. From [9, Theorem 2.2], we have $\nu(\mathrm{dom}\,\mathrm{Arg}) = 1$. By [9, Theorem B.12],

$$\mathrm{dom}\,\mathrm{Arg} = E \cup D \cup (\widehat{\mathrm{dom}}\,\mathrm{Arg} \cap \mathrm{dom}\,\nabla\phi),$$

where $D = \bigcup_{n=1}^{\infty} D_n$, $D_n \in \mathcal{D}$. Let ∂E be the set of relative boundary points of each line segments that constitute E:

$$\partial E = \bigcup_{i=1}^{2} \bigcup_{t\in\mathcal{T}_i} \partial E_i(t).$$

Clearly, ∂E is at most countable. In view of [9, Lemma B.8], the set of points $y \in \widehat{\mathrm{dom}}\,\mathrm{Arg} \cap \mathrm{dom}\,\nabla\phi$ such that $D^c\phi(y) \in \partial E$ is a countable union of sets in $\mathcal{D}$, and thus has ν measure zero. Thus, we may assume $D^c\phi(y) \in G \setminus E$ if $y \in \widehat{\mathrm{dom}}\,\mathrm{Arg} \cap \mathrm{dom}\,\nabla\phi$.

Let h be a bounded continuous function on G. For each $n \in \mathbb{N}$, let $h_n \in C_b(\mathbb{R}^2)$ be such that

(1) $h_n = 0$ outside $B_{n+\frac{1}{n}}$.

(2) $h_n = h$ in $G \setminus (E \cup \bigcup_{x\in\partial E} B_{\frac{1}{n}}(x))$.

(3) $h_n = 0$ on E.

Clearly, h_n converges pointwise to $h1_{\{G\setminus E\}}$.

For $\varepsilon > 0, n \in \mathbb{N}$, define

$$\phi_{n,\varepsilon}(y) = \inf\{c(x, y) + \varepsilon h_n(x)(x_1 - y_1)\}.$$

Note that $\phi_{n,\varepsilon}(y) = \inf_{x\in G^{n,\varepsilon}} c(x, y)$, where $G^{n,\varepsilon} = \{(x_1, x_2 + \varepsilon h_n(x)) : x \in G\}$. Consider the following two cases.

Fix $n \in \mathbb{N}$. Let $y \in \widehat{\text{dom}}\,\text{Arg}\cap\text{dom}\,\nabla\phi$. In this case, $\text{Arg}(y)$ is a singleton with the unique element $x^0 = D^c\phi(y)$. For each $\varepsilon > 0$, we can find $x^{n,\varepsilon} \in G$ such that

$$\phi_{n,\varepsilon}(y) + \varepsilon^2 \geq c(x^{n,\varepsilon}, y) + \varepsilon h_n(x^{n,\varepsilon})(x_1^{n,\varepsilon} - y_1) \geq \phi(y) + \varepsilon h_n(x^{n,\varepsilon})(x_1^{n,\varepsilon} - y_1).$$

On the other hand,

$$\phi_{n,\varepsilon}(y) \leq c(x^0, y) + \varepsilon h_n(x^0)(x_1^0 - y_1) = \phi(y) + \varepsilon h_n(x^0)(x_1^0 - y_1).$$

Together, we have $\lim_{\varepsilon\to0} \phi_{n,\varepsilon} = (y) = \lim_{\varepsilon\to0} c(x^{n,\varepsilon}, y) = \phi(y)$, and

$$-\varepsilon + h_n(x^{n,\varepsilon})(x_1^{n,\varepsilon} - y_1) \leq \frac{1}{\varepsilon}(\phi_{n,\varepsilon}(y) - \phi(y)) \leq h_n(x^0)(x_1^0 - y_1).$$

In view of Lemma 2.6, we may assume that $(x^{n,\varepsilon})_{\varepsilon>0}$ is bounded. As $y \in \widehat{\text{dom}}\,\text{Arg}$, we have $\lim_{\varepsilon\to0} x^{n,\varepsilon} = x^0$ along a subsequence. By dominated convergence theorem, we have

$$\lim_{\varepsilon\to0} \int_{E^c} \frac{1}{\varepsilon}(\phi_{n,\varepsilon}(y) - \phi(y))\, d\nu = \int_{E^c} h_n(D^c\phi(y))(D_1^c\phi(y) - y_1)\, d\nu. \tag{9}$$

Next, let $y \in E$. By [9, Lemma B.1], $\phi(y) = 0$. For each $\varepsilon > 0$, we can find $x^{n,\varepsilon} \in G$ such that

$$\phi_{n,\varepsilon}(y) - \varepsilon h_n(x^{n,\varepsilon})(x_1^{n,\varepsilon} - y_1) + \varepsilon^2 \geq c(x^{n,\varepsilon}, y) \geq \phi(y) = 0.$$

Taking $\varepsilon \to 0$, we have $\lim_{\varepsilon\to0} c(x^{n,\varepsilon}, y) = 0$. If $(x^{n,\varepsilon})_{\varepsilon>0}$ is bounded, then along a subsequence, $x^{n,\varepsilon} \to x^n \in E$ as ε goes to zero. If $(x^{n,\varepsilon})_{\varepsilon>0}$ is unbounded, then there is $\varepsilon_n > 0$ such that $x^{n,\varepsilon} \in E \cap \partial \operatorname{dom} \phi$ for $0 < \varepsilon < \varepsilon_n$. Combining these two possibilities, we deduce that

$$\lim_{\varepsilon \to \infty} \frac{1}{\varepsilon} \phi_{n,\varepsilon}(y) = \lim_{\varepsilon \to \infty} h_n(x^{n,\varepsilon})(x_1^{n,\varepsilon} - y_1) = 0.$$

Apply dominated convergence theorem, we have

$$\lim_{\varepsilon \to \infty} \frac{1}{\varepsilon} \int_E \phi_{n,\varepsilon}(y)\, d\nu = 0. \tag{10}$$

Now, by optimality of ϕ, we have

$$\int \phi\, d\nu \geq \int \phi_{n,\varepsilon}\, d\nu = \int_E \phi_{n,\varepsilon}\, d\nu + \int_{E^c} \phi_{n,\varepsilon}\, d\nu$$

In view of (9) and (10), we have

$$\begin{aligned} \int_{E^c} h_n(D^c\phi(y))(D_1^c\phi(y) - y_1)\, d\nu &= \lim_{\varepsilon \to 0} \int_{E^c} \frac{1}{\varepsilon}(\phi_{n,\varepsilon}(y) - \phi(y))\, d\nu \\ &\leq - \lim_{\varepsilon \to \infty} \frac{1}{\varepsilon} \int_E \phi_{n,\varepsilon}(y)\, d\nu = 0 \end{aligned}$$

Replacing h by $-h$, we obtain

$$\int_{E^c} h_n(D^c\phi(y))(D_1^c\phi(y) - y_1)\, d\nu = 0, \quad n \in \mathbb{N}.$$

The result now follows by sending n to infinity and applying the dominated convergence theorem. $\square$

3 Duality: A Minimax Argument

This section is devoted to establishing the following strong duality relation:

Theorem 3.1. *Let $\nu \in \mathcal{P}_2(\mathbb{R}^2)$. Then*

$$\min_{\gamma \in \Gamma(\nu)} \int c(x,y)\, d\gamma = \max_{G \in \mathfrak{M}} \int \phi_G(y)\, d\nu. \tag{11}$$

In [9], this relation is obtained as an immediate corollary of the first order condition of the primal problem (3). Here, our proof relies on a well-know minimax principle due to Aubin and Ekeland (see, for instance, [2, Chapter 6.2.7]). We state the theorem here for convenience of the readers.

Theorem 3.2. *Let X be a convex subset of a topological vector space, and Y be a convex subset of a vector space. Assume $f : X \times Y \to \mathbb{R}$ satisfies the following conditions:*

1. *For every $y \in Y$, the map $x \mapsto f(x, y)$ is lower semi-continuous and convex.*
2. *There exists y_0 such that $x \mapsto f(x, y_0)$ is inf-compact. That is, the set $\{x \in X : f(x, y_0) \leq a\}$ is relatively compact for each $a \in \mathbb{R}$.*
3. *For every $x \in X$, the map $y \mapsto f(x, y)$ is convex.*

Then, we have

$$\inf_{x \in X} \sup_{y \in Y} f(x, y) = \sup_{y \in Y} \inf_{x \in X} f(x, y).$$

We start with some notations. We denote by proj_x (resp. proj_y) the projection of $\mathbb{R}^2 \times \mathbb{R}^2$ onto its x-coordinates (resp. y-coordinates). For a measuable map $T : \mathbb{R}^2 \to \mathbb{R}^2$ and a probability measure $\mu \in \mathcal{P}_2(\mathbb{R}^2)$, the *push forward* of μ by T is given by

$$T_{\#}\mu(A) = \mu(T^{-1}(A)), \quad A \in \mathcal{B}(\mathbb{R}^2).$$

For $\nu \in \mathcal{P}_2(\mathbb{R}^2)$, the *support* of ν, denote by $\mathrm{supp}\,\nu$, is the smallest closed set that has ν-measure 1. We denote by $\Pi(\nu)$ the family of probability measures on $\mathbb{R}^2 \times \mathbb{R}^2$ with y-marginal given by ν, and by $\Pi_c(\nu)$ the elements of $\Pi(\nu)$ that have compactly supported x-marginal. If $E \subset \mathbb{R}^2$ is measurable, we denote by $\Pi_E(\nu)$ the elements of $\Pi(\nu)$ whose x-marginal has support contained in E. In summary,

$$\begin{aligned}
\Pi(\nu) &= \left\{\gamma \in \mathcal{P}_2(\mathbb{R}^2 \times \mathbb{R}^2) : \ \gamma(\mathbb{R}^2, dy) = \nu(dy)\right\} \\
\Pi_c(\nu) &= \{\gamma \in \Pi(\nu) : \ \mathrm{supp}\{(\mathrm{proj}_x)_{\#}\pi\} \text{ is compact}\} \\
\Pi_E(\nu) &= \{\gamma \in \Pi(\nu) : \ \mathrm{supp}\{(\mathrm{proj}_x)_{\#}\pi\} \subset E\}\,.
\end{aligned}$$

For a measurable function $h : \mathbb{R}^2 \to \mathbb{R}^2$, we write

$$\psi_h(y) \triangleq \inf_{x \in \mathbb{R}^2} \{c(x, y) + \langle h(x), y - x\rangle\}, \quad y \in \mathbb{R}^2,$$

where $\langle x, y\rangle = \sum_{i=1}^2 x_i y_i$ denotes the scalar product of $\mathbb{R}^2$. We also define a functional $\mathcal{F} : \Pi_c(\nu) \times C(\mathbb{R}^2, \mathbb{R}^2) \to \mathbb{R}$ by

$$\mathcal{F}(\pi, h) \triangleq \int c(x, y) + \langle h(x), y - x\rangle\, d\pi.$$

where $C(\mathbb{R}^2, \mathbb{R}^2)$ (resp. $C_b(\mathbb{R}^2, \mathbb{R}^2)$) denotes the space of continuous (resp. bounded continuous) functions on $\mathbb{R}^2$ with values in $\mathbb{R}^2$.

We divide the proof of Theorem 3.1 into a few lemmas.

Lemma 3.3. *For $\pi \in \Pi_c(\nu)$, we have*

$$\sup_{h \in C_b(\mathbb{R}^2,\mathbb{R}^2)} \mathcal{F}(\pi, h) = \sup_{h \in C_{\nu,\psi}(\mathbb{R}^2,\mathbb{R}^2)} \mathcal{F}(\pi, h),$$

where $C_{\nu,\psi}(\mathbb{R}^2, \mathbb{R}^2) = \big\{h \in C(\mathbb{R}^2, \mathbb{R}^2) : \ \psi_h$ is ν-integrable$\big\}$.

Proof. First, as $\pi \in \Pi_c(\nu)$, the inequality

$$\sup_{h \in C_b(\mathbb{R}^2,\mathbb{R}^2)} \mathcal{F}(\pi, h) \geq \sup_{h \in C_{\nu,\psi}(\mathbb{R}^2,\mathbb{R}^2)} \mathcal{F}(\pi, h)$$

holds trivially. On the other hand, let B_r be a ball centered at the origin with radius $r > 0$, that contains the support of the x-marginal of π. For $h \in C_b(\mathbb{R}^2, \mathbb{R}^2)$, define

$$\widehat{h}(x) \triangleq h(x)1_{\{x \in B_r\}} + \big[h(\frac{rx}{|x|}) - x + \frac{rx}{|x|}\big]1_{\{x \in \mathbb{R}^2 \setminus B_r\}}.$$

Then, $\widehat{h}$ is a continuous function such that $\psi_{\widehat{h}}$ is ν-integrable. In particular, $\mathcal{F}(\pi, h) = \mathcal{F}(\pi, \widehat{h})$ and the result follows. □

Lemma 3.4. *For $h \in C_{\nu,\psi}(\mathbb{R}^2, \mathbb{R}^2)$, the map $\pi \mapsto \mathcal{F}(\pi, h)$ is lower semi-continuous on $\Pi(\nu)$ under the weak convergence of measures.*

Proof. We will use the notation $a \wedge b$ to denote the minimum between a and b. Let (π_n) be a sequence in $\Pi(\nu)$, $\pi \in \Pi(\nu)$ and suppose that π_n converges weakly to π. We compute

$$\begin{aligned}
\mathcal{F}(\pi, h) - \int \psi_h(y) d\nu &= \int \big[c(x, y) + \langle h(x), y - x\rangle - \psi_h(y)\big]\, d\pi \\
&= \lim_{m\to\infty} \int \big(c(x, y) + \langle h(x), y - x\rangle - \psi_h(y)\big) \wedge m \, d\pi \\
&= \lim_{m\to\infty} \lim_{n\to\infty} \int \big(c(x, y) + \langle h(x), y - x\rangle - \psi_h(y)\big) \wedge m \, d\pi^n \\
&\leq \liminf_n \int \big(c(x, y) + \langle h(x), y - x\rangle - \psi_h(y)\big)\, d\pi \\
&= \liminf_n \mathcal{F}(\pi^n, h) - \int \psi_h(y) d\nu,
\end{aligned}$$

where we have used the monotone convergence theorem in the third line. □

We proceed to the proof of Theorem (3.2).

Proof of Theorem (3.2). Recall the functional $\mathcal{F} : \Pi_c(\nu) \times \mathcal{C}(\mathbb{R}^2, \mathbb{R}^2) \to \mathbb{R}$ is given by

$$\mathcal{F}(\pi, h) \triangleq \int c(x, y) + \langle h(x), y - x\rangle \, d\pi.$$

Clearly, it is linear in both of its arguments.

Note that both $\Pi_c(\nu)$ and $C_{\nu,\psi}(\mathbb{R}^2, \mathbb{R}^2)$ are convex sets. With the choice of $h_0(x) = -x$, we may apply the inequality $-\frac{1}{8}a^2 - 8b^2 \le 2ab$ to obtain

$$\{\pi \in \Pi_c(\nu) : \ \mathcal{F}(\pi, h_0) \le a\} \subset \left\{\pi \in \Pi(\nu) : \ \int |x|^2 d\pi \le 32\big(a + \int |y|^2 d\nu\big)\right\}$$

Under the weak convergence of measures, the set on the right-hand-side is closed and tight, and thus compact by Prokhorov Theorem. Thus, the map $\pi \mapsto \mathcal{F}(\pi, h_0)$ is *inf-compact*. By Lemma 3.4, the map $\pi \mapsto \mathcal{F}(\pi, h)$ is lower semi-continuous on $\Pi(\nu)$ under the weak convergence of measures, for each $h \in C_{\nu,\psi}(\mathbb{R}^2, \mathbb{R}^2)$. Therefore, we are in the position to apply Theorem 3.2 and obtain

$$\inf_{\pi\in\Pi_c(\nu)} \sup_{h\in C_{\nu,\psi}(\mathbb{R}^2,\mathbb{R}^2)} \mathcal{F}(\pi, h) = \sup_{h\in C_{\nu,\psi}(\mathbb{R}^2,\mathbb{R}^2)} \inf_{\pi\in\Pi_c(\nu)} \mathcal{F}(\pi, h). \tag{12}$$

The martingale property of a probability measure $\pi \in \mathcal{P}_2(\mathbb{R}^2 \times \mathbb{R}^2)$ (i.e., $\mathbb{E}^\pi(Y|X) = X$) is equivalent to

$$\sup_{h\in C_b(\mathbb{R}^2,\mathbb{R}^2)} \int \langle h(x), y - x\rangle \, d\pi = 0.$$

We then deduce from Lemma 3.3, and (12) that

$$\begin{aligned}
\inf_{\pi\in\Gamma(\nu)} \int c(x, y) \, d\pi &= \inf_{\pi\in\Pi(\nu)} \left\{ \int c(x, y) d\pi + \sup_{h\in\mathcal{C}_b(\mathbb{R}^2,\mathbb{R}^2)} \int \langle h(x), y - x\rangle \, d\pi \right\} \\
&\le \inf_{\pi\in\Pi_c(\nu)} \sup_{h\in C_{\nu,\psi}(\mathbb{R}^2,\mathbb{R}^2)} \mathcal{F}(\pi, h) \\
&= \sup_{h\in C_{\nu,\psi}(\mathbb{R}^2,\mathbb{R}^2)} \inf_{\pi\in\Pi_c(\nu)} \mathcal{F}(\pi, h)
\end{aligned}$$

For $h \in C_{\nu,\psi}(\mathbb{R}^2, \mathbb{R}^2)$, Lemma A.1 shows that

$$\begin{aligned}
\inf_{\pi\in\Pi_c(\nu)} \mathcal{F}(\pi, h) &\le \limsup_n \inf_{\pi\in\Pi_{B_n}(\nu)} \mathcal{F}(\pi, h) \\
&= \limsup_n \int \inf_{x\in B_n} \{c(x, y) + \langle h(x), y - x\rangle\} \, d\nu \\
&= \int \psi_h(y) \, d\nu,
\end{aligned}$$

where in the last line we use the monotone convergence theorem. All in all, we have shown that

$$\inf_{\pi\in\Gamma(\nu)}\int c(x,y)\,d\pi \le \sup_{h\in C_{\nu,\psi}(\mathbb{R}^2,\mathbb{R}^2)}\int \psi_h(y)\,d\nu \le \sup_{h\in C(\mathbb{R}^2,\mathbb{R}^2)}\int \psi_h(y)\,d\nu$$

To finish the proof, it suffices to observe that, for $h\in C(\mathbb{R}^2,\mathbb{R}^2)$, the function ψ_h belongs to the class Φ_M. Indeed, for any $t\in[0,1]$ and $y^0,y^1\in\mathbb{R}^2$, we can take $x=(1-t)y^0+ty^1$ in the infimum of ψ_h to obtain

$$(1-t)\psi_h(y^0)+t\psi_h(y^1)\le t(1-t)c(y^0,y^1).$$

The result now follows from [9, Lemma 2.1]. □

4 Continuity of Value Function

In this final section, we establish continuity of the value function:

$$V(\nu)\triangleq\max_{G\in\mathfrak{M}}\int \phi_G(y)\,d\nu=\min_{\gamma\in\Gamma(\nu)}\int c(x,y)\,d\gamma,$$

under the Wasserstein distance:

$$W_2(\mu,\nu)\triangleq\inf_{\pi\in\Pi(\mu,\nu)}\left\{\int |x-y|^2\,d\pi\right\}^{\frac{1}{2}},$$

where $\Pi(\mu,\nu)$ is the family of probability measures on $\mathbb{R}^2\times\mathbb{R}^2$ with x-marginal μ and y-marginal ν. We recall the following equivalent characterizations of convergence in W_2 (see [1, Theorem 2.7 and Proposition 2.4]): for ν_n,ν in $\mathcal{P}_2(\mathbb{R}^2)$,

- $\nu_n\xrightarrow{W_2}\nu$;
- $\nu_n\to\nu$ weakly and $\int|x|^2\,d\nu_n\to\int|x|^2\,d\nu$;
- $\int f(x)\,d\nu_n\to\int f(x)\,d\nu$ for every continuous function $f=f(x)$ with quadratic growth:

$$|f(x)|\le K(1+|x|^2),\quad x\in\mathbb{R}^2.$$

This time, we will adopt the primal formulation and write $J(\gamma)=\int c(x,y)\,d\gamma$. As a result,

$$V(\nu)=\min_{\gamma\in\Gamma(\nu)}J(\gamma).$$

Theorem 4.1. *Let $(\nu_n) \subset \mathcal{P}_2(\mathbb{R}^2)$ be a sequence of probability measures and let $\nu \in \mathcal{P}_2(\mathbb{R}^2)$. Suppose ν_n converges to ν under W_2. Then*

$$\lim_{n\to\infty} V(\nu_n) = V(\nu).$$

Proof. Let $(\nu_{n_k}) \subset (\nu_n)$ be a subsequence such that the limit $\lim_{k\to\infty} V(\nu_{n_k})$ exists. For each k, let $\gamma_{n_k} \in \Gamma(\nu_{n_k})$ be an optimal plan. By Jensen's inequality, martingale property, and the W_2-convergence of (ν_{n_k}), we have

$$\int |x|^2 + |y|^2 \, d\gamma_{n_k} \le 2 \sup_k \int |y|^2 \, d\nu_{n_k} < \infty. \tag{13}$$

It follows that the sequence (γ_{n_k}) is tight. Along a further subsequence, not relabelled, we have γ_{n_k} converges weakly to some probability measure γ.

Clearly, γ has y-marginal given by ν. By Skorodhod representation theorem (see [13, Theorem 17.3], there are random variables (X^{n_k}, Y^{n_k}), $k \in \mathbb{N}$, and (X, Y), defined on the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$, with values in $\mathbb{R}^2 \times \mathbb{R}^2$, such that

1. $(X^{n_k}, Y^{n_k}) \to (X, Y)$, λ-almost everywhere, and
2. the laws of (X^{n_k}, Y^{n_k}) and (X, Y) under λ are given by γ_{n_k} and γ, respectively.

Here, λ is the Lebesgue measure on the interval $[0,1]$. From (13), we deduce that the sequence (X^{n_k}, Y^{n_k}) is uniformly integrable. Therefore, for any bounded continuous function $h \in C_b(\mathbb{R}^2, \mathbb{R}^2)$, we have

$$\begin{aligned} \int \langle h(x), y - x \rangle \, d\gamma &= \mathbb{E}^\lambda(\langle h(X), Y - X \rangle) \\ &= \lim_{k\to\infty} \mathbb{E}^\lambda(\langle h(X^{n_k}), Y^{n_k} - X^{n_k} \rangle) = 0. \end{aligned}$$

Hence, $\gamma \in \Gamma(\nu)$.

Next, we show that γ is an optimal plan. Consider the set

$$C = \left\{ \big((x^0, y^0), (x^1, y^1)\big) \in \mathbb{R}^4 \times \mathbb{R}^4 \;\middle|\; \begin{aligned} &(1-t)c(x^0, y^0) + tc(x^1, y^1) \\ &\qquad \le t(1-t)c(y^0, y^1), \quad \forall t \in [0,1] \end{aligned} \right\}.$$

By [9, Theorem 2.2], we deduce from optimality of each γ_{n_k} that $\operatorname{supp} \gamma_{n_k} \otimes \gamma_{n_k} \subset C$. Continuity of $c = c(x, y)$ implies the set C is closed. As $\gamma_{n_k} \otimes \gamma_{n_k}$ converges weakly to $\gamma \otimes \gamma$, we have

$$\gamma \otimes \gamma(C) \ge \limsup_k \gamma_{n_k} \otimes \gamma_{n_k}(C) = 1.$$

This shows that γ is optimal.

Now, as ν_{n_k} converges to ν under W_2, we have $\lim_{k\to\infty} \mathbb{E}^\lambda(|Y^{n_k}|^2) = \mathbb{E}^\lambda(|Y|^2)$. It follows that Y^{n_k} converges to Y in $\mathcal{L}_2$. On the other hand, let

$$\begin{aligned} W^{n_k} &= \mathbb{E}^\lambda(Y^{n_k} - Y|X^{n_k}) \\ Z^{n_k} &= \mathbb{E}^\lambda(Y|X^{n_k}). \end{aligned}$$

By Jensen's inequality, W^{n_k} converges to 0 in $\mathcal{L}_2$, and in particular, λ-almost everywhere along a further subsequence. Therefore, as $k \to \infty$,

$$Z^{n_k} = X^{n_k} - W^{n_k} \to X, \quad \lambda\text{-a.e.} \tag{14}$$

It is easy to see that the family of random variables $(|Z^{n_k}|^2)$ is uniformly integrable. Together with (14), this implies that Z^{n_k} converges to X in $\mathcal{L}_2$. We then deduce that X^{n_k} converges to X in $\mathcal{L}_2$, which in turn implies

$$\gamma_{n_k} \xrightarrow{W_2} \gamma, \quad \text{as} \quad k \to \infty.$$

As $c = c(x, y)$ is a continuous function with quadratic growth, we conclude that

$$\lim_{k\to\infty} V(\nu^{n_k}) = \lim_{k\to\infty} J(\gamma_{n_k}) = J(\gamma) = V(\nu).$$

The result now follows from arbitrariness of (ν_{n_k}).

$\square$

A A measurable selection result

We denote by Σ_1^1 the family of analytic sets in $\mathbb{R}^d$, and write $\sigma(\Sigma_1^1)$ to represent the σ-algebra generated by Σ_1^1. A subset $A \in \mathbb{R}^d$ is called *universally measurable* if it is μ-measurable for any Borel probability measure μ. Given a set $C \subset \mathbb{R}^d \times \mathbb{R}^d$, we recall a *uniformization* of C is a function s with domain $D = \text{proj}_x(C)$, such that $s(x) \in C_x$ for every $x \in D$, where $C_x \triangleq \{y : (x, y) \in C\}$ is the x-section of C.

Lemma A.1. *Let $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, $E \subset \mathbb{R}^d$ be a closed set, and $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a Borel measurable function such that the partial minimization*

$$y \mapsto \inf_{x\in E} f(x, y)$$

is Borel measurable. Then,

$$\inf_{\pi\in\Pi(E,\nu)} \int f(x, y)\, d\pi = \int \inf_{x\in E} f(x, y)\, d\nu,$$

where $\Pi(E, \nu) = \{\pi \in \Pi(\nu) : (\text{proj}_x)_\# \pi \subset E\}$.

Proof. Let $\varepsilon > 0$. Define a function ϕ by $\phi(y) = \inf_{x \in E} f(x, y)$. By assumption, ϕ is Borel. Therefore, the set

$$B_\varepsilon = \left\{(x, y) \in E \times \mathbb{R}^d : \ f(x, y) \le \phi(y) + \varepsilon\right\}$$

is a Borel set such that no y-section of empty. By the Uniformization Theorem of Von Neumann and Jankov (see [8, Theorem 18.1]), there exits a uniformization function s_ε of B_ε that is $\sigma(\Sigma^1_1)$-measurable. By the Lusin's Theorem (see [8, Theorem 29.7]), s_ε is universally measurable. In particular, it is ν-measurable.

Now, define a probability measurable π_ε by

$$\pi_\varepsilon(A) = \nu((s_\varepsilon, \mathbf{I})^{-1}(A)), \quad A \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d),$$

where $\mathbf{I}$ is the identity function on $\mathbb{R}^d$. Clearly, π_ε has y-marginal given by ν, and x-marginal given by $\mu = (s_\varepsilon)_\# \nu$. In particular, $\mu(E) = 1$. Therefore, $\operatorname{supp} \mu \subset E$. We then deduce that

$$\inf_{\pi \in \Pi(E, \nu)} \int f(x, y)\, d\pi \le \int f(x, y)\, d\pi_\varepsilon \le \int \phi(y)\, d\nu + \varepsilon.$$

The result now follows as $\varepsilon > 0$ is arbitrary. ☐

References

[1] Luigi Ambrosio and Nicola Gigli. *A User's Guide to Optimal Transport*, pages 1–155. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[2] J.P. Aubin and I. Ekeland. *Applied Nonlinear Analysis*. Dover Books on Mathematics Series. Dover Publications, 2006.

[3] Mathias Beiglbck and Nicolas Juillet. On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.*, 44(1):42–106, 01 2016.

[4] Mathias Beiglbck, Marcel Nutz, and Nizar Touzi. Complete duality for martingale optimal transport on the line. *Ann. Probab.*, 45(5):3038–3074, 09 2017.

[5] Freddy Delbaen and Walter Schachermayer. *A General Version of the Fundamental Theorem of Asset Pricing (1994)*, pages 149–205. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[6] Nassif Ghoussoub, Young-Heon Kim, and Tongseok Lim. Structure of optimal martingale transport plans in general dimensions. *Ann. Probab.*, 47(1):109–164, 01 2019.

[7] Pierre Henry-Labordère and Nizar Touzi. An explicit martingale version of the one-dimensional brenier theorem. *Finance and Stochastics*, 20(3):635–668, Jul 2016.

[8] A. Kechris. *Classical Descriptive Set Theory*. Graduate Texts in Mathematics. Springer New York, 2012.

[9] Dmitry Kramkov and Yan Xu. An optimal transport problem with backward martingale constraints motivated by insider trading. *arXiv e-prints*, page arXiv:1906.03309, Jun 2019.

[10] Albert S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985.

[11] S.M. Srivastava. *A Course on Borel Sets*. Graduate Texts in Mathematics. Springer New York, 1998.

[12] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg.

[13] David Williams. *Probability with Martingales*. Cambridge University Press, 1991.