universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„(Bi)causal Optimal Transport"

verfasst von / submitted by

Lukas Anzeletti BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 821
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Mathematik
Betreut von / Supervisor:	Univ.-Prof. Dipl.-Ing. Dr. techn. Mathias Beiglböck

Abstract

In the first part of this thesis, we will prove duality and existence of optimal couplings for the (bi)causal optimal transport problem with looser conditions than the current theory allows for. Subsequently, we will give a characterization of the support of cost functions for which the optimal (bi)causal value vanishes by decomposing the underlying measures and using already existing general transport theory. This characterization will prove useful when it comes to counterexamples where the causal and bicausal problem do not agree. In the end, we will use recent theory in the area of weak transport to establish a Kantorovich-Rubenstein type result for the causal case, considering only two-step stochastic processes.

Kurzfassung

Der erste Teil der Arbeit beschäftigt sich mit Dualität und Existenz von optimalen Couplings für das (bi)kausale Transportproblem. Dabei werden weniger restriktive Annahmen als in ähnlichen bereits existierenden Resultaten benötigt. Anschließend wird, mithilfe einer Zerlegung des jeweiligen Maßes, eine Charakterisierung des Trägers einer Kostenfunktion gegeben, deren optimale Transportkosten verschwinden. Diese Charakterisierung stellt sich als hilfreich bei Gegenbeispielen heraus, bei denen das kausale und bikausale Problem nicht übereinstimmt. Den Abschluss der Arbeit bildet eine Art Kantorovich-Rubenstein Dualität für kausalen Transport. Dies wird ermöglicht durch eine Brücke zwischen kausalem und schwachem Transport. Dadurch können vor kurzem bewiesene Dualitätsresultate für schwachen Transport genutzt werden.

Contents

1 Introduction and outline

In the usual optimal transport problem we are confronted with a source measure μ on some space X and a target measure ν on some space Y. The goal is to find couplings between these two measures (i.e. measures on $X \times Y$, such that their projection on X is equal to μ and their projection on Y is equal to ν), which give us the minimal amount of cost. This notion of cost arises from a function defined on $X \times Y$, which heuristically tells us the cost of moving mass from some point $x \in X$ to some point $y \in Y$.

Throughout this thesis we will deal with discrete-time stochastic processes and transport plans between them. In this setup one only allows transport plans which do not have to "look into the future" of the first process in order to assign mass to the second one. Couplings which have this property are called causal and if this property holds true in "both directions" (interchanging what we consider the first and the second process) they are called bicausal.

At the beginning, we will generalize Theorem 2.5 in [BBLZ17] on the duality and existence of optimal couplings by omitting the continuity assumption on the stochastic kernels. The proof is based on Kantorovich duality (see e.g. Theorem 5.9 in [Vil16]) and the fact that we can see causal transport as a special case of the usual transport problem with some additional linear constraints.

The next chapters are inspired by Theorem 2.21 in [Kel84], which characterizes the structure of subsets of $X \times Y$ which have mass zero with respect to all couplings. We will derive a similar result for the bicausal transport problem by decomposing couplings and recursively applying Theorem 2.21 in [Kel84]. This characterization will prove useful when it comes to various examples in which the bicausal optimal transport value differs from the causal one. By adapting Theorem 2.6 in [BBLZ17], we can also give a characterization in the causal case, although it will be less illustrative.

At the end of the thesis, we will build a bridge between causal transport and the recent theory of weak transport. Using this bridge and already established duality theory for weak transport (see [BBP19]), we will prove a Kantorovich-Rubenstein type result for the causal transport problem.

To ease notation we will often work on $\mathbb{R}^N$. Notice that we can replace $\mathbb{R}^N$ by S^N for an abstract Polish space S. The only difference is that the proof of Lemma 5.1 will be less constructive.

2 Definitions and notations

The purpose of this chapter is to introduce notations and some concepts that may not be familiar to the reader who does not have a lot of measure-theoretic background. Nevertheless the concept of universal measurability and analytic sets/measurability will not be discussed, but will be used every now and then. For a detailed explanation on these topics see [BS96].

Remark 1. *Let X be a Polish space and $\mathcal{A}$ its Borel sigma-algebra. We denote the set of all probability measures on the space $(X, \mathcal{A})$ by $\mathcal{P}(X)$. Recall that the space $\mathcal{P}(X)$ is again a Polish space if endowed with the weak topology. By this construction $\mathcal{P}(\mathcal{P}(X))$ is again a Polish space as well. If not mentioned otherwise we always endow $\mathcal{P}(X)$ with the weak topology.*

Definition 1. Let $\mathcal{X}, \mathcal{Y}$ be Polish spaces. We call a collection of probability measures in $\mathcal{P}(\mathcal{Y})$ parametrized by $x \in \mathcal{X}$ a stochastic kernel and denote it by $q_x(dy)$. We call it measurable (resp. continuous) if $q : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $x \mapsto q_x(dy)$ is measurable (resp. continuous).

Definition 2. Let $(X, \mathcal{A}), (Y, \mathcal{B})$ be measurable spaces and let $f : X \to Y$ be a measurable map. For a measure μ on X we denote its f-pushforward measure by $f_{\#}\mu$. The measure $f_{\#}\mu$ is given by

$$(f_{\#}\mu)(B) := \mu(f^{-1}(B)), \text{ for } B \in \mathcal{B}.$$

Theorem 2.1. ***(Disintegration on product spaces)***
Let $X = X_1 \times X_2$ be a Polish space, $\mu \in \mathcal{P}(X)$ and let $\pi_i : X \to X_i$ be the natural projection for $i = 1, 2$. Identifying $\pi_1^{-1}(x_1)$, for $x_1 \in X_1$, with X_2, there exist a collection of probability measures $\{\mu_{x_1}\}_{x_1 \in X_1}$ in $\mathcal{P}(X_2)$, such that

$$\mu(A \times B) = \int_A \mu_{x_1}(B) d((\pi_1)_{\#}\mu)(x_1)$$

for $A \subset X_1, B \subset X_2$ measurable. The collection of probability measures $\{\mu_{x_1}\}_{x_1 \in X_1}$ is unique $(\pi_1)_{\#}\mu$-a.s.

Definition 3. Let X be a topological space and $f : X \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$. We call f lower semicontinuous if $\{x \in X : f(x) \leqslant c\}$ is closed for all $c \in \overline{\mathbb{R}}$. We will abbreviate this by f being l.s.c.

By abuse of notation we will denote both $(x_1, ..., x_N) \mapsto x_1$ and $(x_1, ..., x_N, y_1, ..., y_N) \mapsto (x_1, ..., x_N)$ by p^1. For a measure $\mu \in \mathcal{P}(\mathbb{R}^N)$ and the first of the above functions we use the notation $\mu^1 := (p^1)_{\#}(\mu)$.

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$, where $\mathbb{R}^N$ is endowed with the filtration generated from the coordinate processes (i.e. $\mathcal{F}_t$ is the smallest sigma-algebra such that $(x_1, ..., x_N) \mapsto (x_1, ..., x_t)$ is $\mathcal{F}_t$-measurable). We denote the set of all transport plans between μ and ν by

$$\Pi(\mu, \nu) = \{\gamma \in \mathcal{P}(\mathbb{R}^N \times \mathbb{R}^N) : p^1_{\#}\gamma = \mu, p^2_{\#}\gamma = \nu\}.$$

Definition 4. We call $\gamma \in \Pi(\mu,\nu)$ a causal transport plan, if $x \mapsto \gamma_x(B)$ is $\mathcal{F}_t$-measurable for all $t \leqslant N$ and for all $B \in \mathcal{F}_t$ and we denote the set of all causal transport plans between μ and ν by $\Pi_c(\mu,\nu)$. If it also holds true that $e_{\#}\gamma \in \Pi_c(\nu,\mu)$, for $e(x,y) := (y,x)$, we call γ a bicausal coupling and we denote the set of all bicausal couplings between μ and ν by $\Pi_{bc}(\mu,\nu)$. Notice that the coupling $\gamma = \mu \otimes \nu$ is bicausal.

The idea we have in mind here is that for two stochastic processes, knowing the future of one of them does not provide any additional information about the status quo of the other process. So for two stochastic processes X and Y causality tells us that, knowing X up to some point, Y up to this point is independent of the future of X. To be more precise this means that for $1 \leqslant t \leqslant N$ and $B_i \in \mathcal{F}_i$, for $1 \leqslant i \leqslant t$,

$$\mathbb{P}(Y_1 \in B_1, ..., Y_t \in B_t | X_1, ..., X_N) = \mathbb{P}(Y_1 \in B_1, ..., Y_t \in B_t | X_1, ..., X_t).$$

Throughout the thesis we will use the notion of a Bochner integral. We will give a very brief overview of the kind of Bochner integral we need. For a thorough construction see for example [Coh13].

Remark 2. *Let $(X,\mathcal{A},\mu)$ be a measure space and E a separable Banach space endowed with its Borel sigma-algebra. We call $f : X \to E$ Bochner integrable if it is measurable and $x \mapsto \|f(x)\|$ is integrable.*

For a simple function $f = \sum_{i=1}^n \mathbb{1}_{A_i} a_i$ we define $\int f d\mu$ to be $\sum_{i=1}^n a_i \mu(A_i)$. For an arbitrary Bochner integrable function we define the integral by the usual approximation by simple functions. (see [Coh13])

We will often need a special case of a Bochner integral. Let $\mathcal{X}$ be a Polish space endowed with its Borel sigma-algebra $\mathcal{A}$. Let the Banach space E be the space of bounded signed measures on $\mathcal{X}$ with the total variation norm, which we denote by $Ba(\mathcal{X})$. In this setup $f : \mathcal{P}(\mathcal{X}) \to Ba(\mathcal{X})$, $\hat{x} \mapsto \hat{x}$ is integrable with respect to any $\mu \in \mathcal{P}(\mathcal{P}(\mathcal{X}))$. We will also refer to $\int \hat{z}\mu(d\hat{z})$ as $mean(\mu)$.

Lemma 2.2. *Let $A \in \mathcal{A}$ and $\mu \in \mathcal{P}(\mathcal{P}(\mathcal{X}))$. Then we have that*

$$\left(\int \hat{x}\mu(d\hat{x}) \right)(A) = \int \hat{x}(A)\mu(d\hat{x}).$$

Proof. Let $f : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R}) \subset Ba(\mathcal{X})$, $\hat{x} \mapsto \hat{x}$. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of simple functions with $f(\hat{x}) = \lim_n f_n(\hat{x})$ for all $\hat{x} \in \mathcal{P}(\mathbb{R})$, such that $\|f_n(\hat{x})\| \leqslant \|f(\hat{x})\|$ for all n and for all $\hat{x} \in \mathcal{P}(\mathcal{X})$.

Then we have that

$$\begin{aligned}
\left(\int \hat{x} d\mu \right)(A) &= \lim_n \left(\int f_n(\hat{x}) d\mu \right)(A) \\
&= \lim_n \int f_n(\hat{x})(A) d\mu \\
&= \int \lim_n f_n(\hat{x})(A) d\mu \\
&= \int \hat{x}(A)\mu(d\hat{x}).
\end{aligned}$$

The first equality holds true, as convergence w.r.t. total variation implies strong convergence and the third equality follows from dominated convergence. $\square$

For a metric space X we denote the set of 1-Lipschitz functions mapping to $\mathbb{R}$ by $Lip_1(X)$. If it is clear which space X is meant, we just write Lip_1.

For a Polish metric space X we denote by $\mathcal{P}_1(X)$ the set of probability measures with finite first moment (i.e. $\mathcal{P}_1(X) := \{\mu \in \mathcal{P}(X) : \int d(x, x_0)\mu(dx) < \infty \ \forall x_0 \in x\}$). The space $\mathcal{P}_1(X)$ is again Polish. The underlying topology can be characterized in the following way: A series of measures $\mu_n \in \mathcal{P}_1(X)$ converges to μ if it converges weakly and the series of their first moments converge to the first moment of μ. A complete metric can be given by the first order Wasserstein-distance

$$W(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int d(x, y)\pi(dx, dy).$$

Let

$$\mathcal{C}_{lin}(\mathbb{R}) := \{f \in \mathcal{C}(\mathbb{R}) : \exists C, x_0 \in \mathbb{R} \ s.th. \ |f(x)| \leqslant C(1 + d(x_0, x)) \ \forall x \in \mathbb{R}\}.$$

Notice that for $\phi \in \mathcal{C}_{lin}(\mathbb{R})$ and $p \in \mathcal{P}_1(\mathcal{P}_1(\mathbb{R}))$ we have that

$$\int \phi(x)(\textstyle\int \hat{x}p(d\hat{x}))(dx) = \int \hat{x}(\phi)p(d\hat{x}). \tag{2.1}$$

In particular $\int \hat{x}p(d\hat{x}) \in \mathcal{P}_1(\mathbb{R})$ because both sides in (2.1) are finite. To see this, first notice that for arbitrary $\hat{x}_0 \in \mathcal{P}_1(\mathbb{R})$

$$\int W(\hat{x}, \hat{x}_0)p(d\hat{x}) < \infty.$$

Let $C \in \mathbb{R}$ such that $|\phi(x)| \leqslant C(1 + |x|)$ for all $x \in \mathbb{R}$. Therefore, choosing $\hat{x}_0 = \delta_0$ and using Kantorovich duality we have that

$$\begin{aligned}
\infty > \int W(\hat{x}, \hat{x}_0)p(d\hat{x}) &= \int \sup_{\psi \in Lip_1} \left(\int \psi(x)\hat{x}(dx) - \int \psi(x)\hat{x}_0(dx) \right) p(d\hat{x}) \\
&\geqslant \int \left(\int (1 + |x|)\hat{x}(dx) \right) p(d\hat{x}) - 1 \\
&\geqslant \frac{1}{C} \int \int |\phi(x)|\hat{x}(dx)p(d\hat{x}) - 1.
\end{aligned}$$

Definition 5. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. For $0 < p < \infty$ we denote the vector space $\mathcal{L}^p(\Omega, \mathcal{A}, \mu) := \{f : \Omega \to \mathbb{R} \text{ measurable} : \int |f|^p d\mu < \infty\}$ by $\mathcal{L}^p(\mu)$.

3 Duality for (bi)causal transport

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$ and c some l.s.c. cost function on $\mathbb{R}^N \times \mathbb{R}^N$. The main result in this chapter will be to establish attainment and duality of

$$\inf_{\pi \in \Pi_c(\mu,\nu)} \int c d\pi \text{ and } \inf_{\pi \in \Pi_{bc}(\mu,\nu)} \int c d\pi. \tag{3.1}$$

We will refer to the first expression in (3.1) as (Pc) and to the second one as (Pbc). Our main results in this chapter (Theorem 3.2 and 3.3) generalize Theorem 2.5 and Corollary 3.3 in [BBLZ17]. Indeed, compared to Theorem 2.5 and Corollary 3.3 in [BBLZ17], we do not impose the assumption that μ and ν have continuous kernels. We rely on arguments similar to the ones used in [BBLZ17]. Additionally we make use of the following: for Polish spaces X and Y and finitely many Borel measurable functions $f_i : X \to Y$ we can find a finer topology on X, with the same Borel sets as the original topology, such that all f_i are continuous.

Lemma 3.1. *Denote the usual topology on $\mathbb{R}^N$ by τ. Let $\tilde{\tau} \supseteq \tau$ such that their corresponding Borel sets are equal. Let $t < N$ and assume that $(x_1, ..., x_t, x_{t+1}, ..., x_N) \mapsto \mu_{x_1,...,x_t}$ is continuous with respect to $\tilde{\tau}$. For $g \in \mathcal{C}_b(\mathbb{R}^N, \tau)$, the function defined by*

$$(x_1, ..., x_N) \to g(x_1, ..., g_N) - \int g(x_1, ..., x_t, \tilde{x}_{t+1}, ..., \tilde{x}_N)\mu_{x_1,...,x_t}(d\tilde{x}_{t+1}, ..., d\tilde{x}_N)$$

belongs to $\mathcal{C}_b(\mathbb{R}^N, \tilde{\tau})$.

Proof. If $\|g\|_\infty = 0$ the statement is clearly true, so we assume that $\|g\|_\infty \neq 0$. As $\tilde{\tau} \supseteq \tau$, it suffices to check the continuity of

$$(x_1, ..., x_N) \mapsto \int g(x_1, ..., x_t, \tilde{x}_{t+1}, ..., \tilde{x}_N)\mu_{x_1,...,x_t}(d\tilde{x}_{t+1}, ..., d\tilde{x}_N).$$

It suffices to show sequential continuity as $(\mathbb{R}^N, \tilde{\tau})$ is a Polish space and therefore we can describe its topology by a metric. Let $(x_1^n, ..., x_N^n)$ converge to $(y_1, ..., y_N)$ w.r.t. $\tilde{\tau}$ and therefore also w.r.t. τ. We denote $(x_1^n, ..., x_t^n)$ by x^n and $(y_1, ..., y_t)$ by y.

Let $\varepsilon > 0$. By assumption, μ_{x^n} converges weakly to μ_y, so the sequence is tight by Prokhorov's Theorem. Hence, there exists $K \subset (\mathbb{R}^{N-t}, \tau)$ compact such that $\sup_n \mu_{x^n}(K^{\mathsf{c}}) < \frac{\varepsilon}{6\|g\|_\infty}$. Notice that $B = \{x^n : n \in \mathbb{N}\} \cup \{y\}$ is sequentally compact and therefore also compact in $(\mathbb{R}^N, \tilde{\tau})$. In particular B is compact in $(\mathbb{R}^N, \tau)$. Its projection $\tilde{B}$ on $(\mathbb{R}^t, \tau)$ is compact as well. So $\tilde{B} \times K \subset (\mathbb{R}^N, \tau)$ is compact as well. Note that if $(x_1^n, ..., x_N^n)$ converges to $(y_1, ..., y_N)$ in $(\mathbb{R}^N, \tilde{\tau})$, $(x_1^n, ..., x_t^n)$ converges to $(y_1, ..., y_t)$ in $(\mathbb{R}^t, \tau)$.

We then get for n large enough that

$$\left| \int g(x^n, z)\mu_{x^n}(dz) - \int g(y, z)\mu_y(dz) \right| \leqslant \alpha + \beta + \gamma,$$

where

$$\alpha = \left| \int_K g(x^n, z) - g(y, z)\mu_{x^n}(dz) \right|, \beta = \left| \int_{K^c} g(x^n, z) - g(y, z)\mu_{x^n}(dz) \right|,$$

$$\gamma = \left| \int g(y, z)\mu_{x^n}(dz) - \int g(y, z)\mu_y(dz) \right|.$$

Each of the terms α, β and γ is smaller than $\varepsilon/3$. This is true for α because for n large enough $\sup_{n>n_0, z\in K} |g(x^n, z) - g(y, z)| < \varepsilon/3$, for β because $\sup_n \mu_{x^n}(K^c) < \frac{\varepsilon}{6\|g\|_\infty}$ and for γ because $g \in \mathcal{C}_b(\mathbb{R}^N)$ and $\mu_{x^n} \to \mu_y$.

□

With the previous Lemma we are ready to prove a duality result for the causal transport problem. The proof of Theorem 3.2 resembles the proof of Theorem 2.1 in [Zae15].

For $\mu \in \mathcal{P}(\mathbb{R}^N)$, $\nu \in \mathcal{P}(\mathbb{R}^N)$ and $c : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ measurable, let

$$A(\mu, \nu, c) := \{(\phi, \psi) : \phi \in \mathcal{L}^1(\mu), \psi \in \mathcal{L}^1(\nu), \phi(x) + \psi(y) \leqslant c(x, y)\ \forall (x, y) \in \mathbb{R}^N \times \mathbb{R}^N\}.$$

Theorem 3.2. *Let $c : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ be l.s.c. and bounded from below. Then*

$$\inf_{\gamma \in \Pi_c(\mu,\nu)} \int c d\gamma = \sup_{\substack{(\phi,\psi)\in A(\mu,\nu,c-f) \\ f\in\mathbb{F}}} \left(\int \phi d\mu + \int \psi d\nu \right) \tag{3.2}$$

for

$$\mathbb{F} = \left\{ \begin{array}{c} F : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R} : F(x_1, ..., x_N, y_1, ...y_N) = \\ \sum_{t<N} h_t(y_1, ..., y_t)[g_t(x_1, ..., x_N) \\ - \int g_t(x_1, ...x_t, \bar{x}_{t+1}, ..., \bar{x}_N)\mu_{x_1,...,x_t}(d\bar{x}_{t+1}, ..., d\bar{x}_N)], \\ g_t \in \mathcal{C}_b(\mathbb{R}^N), h_t \in \mathcal{C}_b(\mathbb{R}^t)\ \forall\ t < N \end{array} \right\}$$

and the infimum on the left hand side of (3.2) is attained.

Proof. Let τ be the usual topology on $\mathbb{R}^N$. By Prop. 2.80 in [Dob14] there exists a topology $\tilde{\tau}$ on $\mathbb{R}^N$ such that $\mathbb{R}^N$ endowed with $\tilde{\tau}$ is still a Polish space, $(x_1, ..., x_t) \to \mu_{x_1,...,x_t}$ is continuous for all $t < N$ and all Borel sets are preserved. By Lemma 3.1 all $f \in \mathbb{F}$ are continuous w.r.t. $(\mathbb{R}^N \times \mathbb{R}^N, \tilde{\tau} \times \tau)$. By Proposition 2.3 in [BBLZ17] we know that

$$\Pi_c(\mu, \nu) = \Pi(\mu, \nu) \cap \bigcap_{f\in\mathbb{F}} \phi_f^{-1}(0), \tag{3.3}$$

where $\phi_f(\pi) = \int f d\pi$. For f continuous and bounded, ϕ_f is continuous due to the definition of weak convergence. As $\Pi(\mu, \nu)$ is compact, $\Pi_c(\mu, \nu)$ is compact and attainment follows by lower semicontinuity of $\gamma \mapsto \int c d\gamma$ (see Theorem 4.1 in [Vil16]).

Notice that we have

$$\sup_{\substack{(\phi,\psi)\in A(\mu,\nu,c-f) \\ f\in\mathbb{F}}} \left(\int \phi d\mu + \int \psi d\nu \right) = \sup_{f\in\mathbb{F}} \sup_{(\phi,\psi)\in A(\mu,\nu,c-f)} \left(\int \phi d\mu + \int \psi d\nu \right)$$

$$= \sup_{f\in\mathbb{F}} \inf_{\pi\in\Pi(\mu,\nu)} \int c - f d\pi. \tag{3.4}$$

In the second line we used Kantorovich duality (see Theorem 5.9 in [Vil16]), which is possible as we equipped $\mathbb{R}^N \times \mathbb{R}^N$ with $\tilde{\tau} \times \tau$ and therefore f is continuous and bounded. Hence $c - f$ is l.s.c.

Now, we can interchange supremum and infimum, because $\Pi(\mu, \nu) \subset \mathcal{P}(\mathbb{R}^N \times \mathbb{R}^N)$ is compact, $\mathbb{F}$ is convex and w.r.t. $\tilde{\tau} \times \tau$, $c - f$ is l.s.c. and therefore $\pi \to \int c - f d\pi$ is as well by Lemma 4.3. in [Vil16] . These are all the conditions necessary to apply Theorem 2.3. in [Zae15]. So (3.4) is equal to

$$\inf_{\pi \in \Pi(\mu,\nu)} \sup_{f \in \mathbb{F}} \int c - f d\pi. \tag{3.5}$$

For $\pi \notin \Pi_c(\mu, \nu)$ we can choose an $f \in \mathbb{F}$ s.th. $\int f d\pi < 0$ as $\mathbb{F}$ is stable under scalar multiplication. Choosing αf for $\alpha \to \infty$ gives us that it is sufficient to consider the infimum over all causal couplings. By (3.3) the integral of f with respect to a causal coupling vanishes, so we get that (3.5) is equal to

$$\inf_{\pi \in \Pi_c(\mu,\nu)} \sup_{f \in \mathbb{F}} \int c - f d\pi = \inf_{\pi \in \Pi_c(\mu,\nu)} \int c d\pi.$$

□

The proof of the bicausal equivalent to Theorem 3.2 is very similar to the proof of Theorem 3.2, so we will omit it and just state the result.

Theorem 3.3. *Let $c : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ be l.s.c. and bounded from below. Then*

$$\inf_{\gamma \in \Pi_{bc}(\mu,\nu)} \int c d\gamma = \sup_{\substack{(\phi,\psi) \in A(\mu,\nu,c-f) \\ f \in \mathbb{F}'}} \left(\int \phi d\mu + \int \psi d\nu \right) \tag{3.6}$$

for

$$\mathbb{F}' = \left\{ \begin{array}{c} F : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R} : F(x_1, ..., x_N, y_1, ...y_N) = \\ \sum_{t<N} h_t(y_1, ..., y_t)[g_t(x_1, ..., x_N) - \\ \int g_t(x_1, ...x_t, \bar{x}_{t+1}, ..., \bar{x}_N) \mu_{x_1,...,x_t}(d\bar{x}_{t+1}, ..., d\bar{x}_N)] + \\ \sum_{t<N} h'_t(x_1, ..., x_t)[g'_t(y_1, ..., y_N) - \\ \int g'_t(y_1, ...y_t, \bar{y}_{t+1}, ..., \bar{y}_N) \nu_{y_1,...,y_t}(d\bar{y}_{t+1}, ..., d\bar{y}_N)] \\ g_t, g'_t \in \mathcal{C}_b(\mathbb{R}^N), h_t, h'_t \in \mathcal{C}_b(\mathbb{R}^t) \ \forall \ t < N \end{array} \right\}$$

and the infimum on the left hand side of (3.6) is attained.

4 Bicausal 0-sets for N=2

In the following two chapters, for a function $f : X \to Y$, we will use the notation

$$Gr(f) := \{(x, f(x)) : x \in X\}.$$

Let $\mathcal{X}$, $\mathcal{Y}$ be Polish spaces and $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$. By Theorem 2.21 in [Kel84] we know that if, for nonnegative f,

$$\sup_{\pi \in \Pi(\mu,\nu)} \int f d\pi = 0$$

then there exist functions g and h with $f \leqslant g \oplus h$ and $\mu(g) = \nu(h) = 0$.

In this chapter we will derive a similar result in the case of bicausal couplings.

Lemma 4.1. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$. Let $\mathcal{X}$ and $\mathcal{Y}$ be closed subsets of $\mathbb{R}^N$. Then*

$$D := \{(x_1, y_1, \pi) : x_1 \in p^1(\mathcal{X}), y_1 \in p^1(\mathcal{Y}), \pi \in \Pi(\mu_{x_1}, \nu_{y_1})\}$$

is an analytic set.

Proof. As we already did before, we can equip $\mathbb{R}$ with a topology $\tilde{\tau}$, maintaining Borel-sets, such that $\mathbb{R}$ is still Polish and $x_1 \mapsto \mu_{x_1}$ and $y_1 \mapsto \nu_{y_1}$ are continuous. Let $(x_1^n, y_1^n, \pi_n)_{n\in\mathbb{N}} \in \{(x_1, y_1, \pi) : x_1, y_1 \in (\mathbb{R}, \tilde{\tau}), \pi \in \Pi(\mu_{x_1}, \nu_{y_1})\}$ converge to some (x_1, y_1, π). As $\pi_n \to \pi$ we have that, for $f \in \mathcal{C}_b(\mathbb{R})$,

$$\lim_{n\to\infty} \int f d(p^1_{\#}\pi_n) = \lim_{n\to\infty} \int f \circ p^1 d\pi_n = \int f \circ p^1 d\pi = \int f d(p^1_{\#}\pi).$$

Hence $p^1_{\#}\pi = \mu_{x_1}$ and by the same considerations $p^2_{\#}\pi = \nu_{y_1}$. Therefore $\{(x_1, y_1, \pi) : x_1, y_1 \in (\mathbb{R}, \tilde{\tau}), \pi \in \Pi(\mu_{x_1}, \nu_{y_1})\}$ is closed, in particular it is Borel and also analytic.

As the projections of closed sets are analytic sets and their product is again analytic, we have that $p^1(\mathcal{X}) \times p^1(\mathcal{Y}) \times \mathcal{P}(\mathbb{R}^{N-1} \times \mathbb{R}^{N-1})$ is analytic. Hence

$$\begin{aligned} D = \{(x_1, y_1, \pi) : x_1, y_1 \in \mathbb{R}, &\pi \in \Pi(\mu_{x_1}, \nu_{y_1})\} \\ &\cap \left(p^1(\mathcal{X}) \times p^1(\mathcal{Y}) \times \mathcal{P}(\mathbb{R}^{N-1} \times \mathbb{R}^{N-1})\right) \end{aligned}$$

is also analytic as a finite intersection of analytic sets is analytic.

□

Remark 3. *Note that $D := \{(x_1, y_1, \pi) : x_1, y_1 \in \mathbb{R}, \pi \in \Pi_{bc}(\mu_{x_1}, \nu_{y_1})\}$ is not necessarily closed for $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$, for $N > 2$, even if $x_1 \mapsto \mu_{x_1}$ and $y_1 \mapsto \nu_{y_1}$ are continuous. To see this let $x_1^n \to x_1$, $y_1^n \to y_1$ and*

$$\mu_{x_1^n} = \frac{1}{2}\delta_{(1/n,1)} + \frac{1}{2}\delta_{(-1/n,-1)}, \ \nu_{y_1^n} = \frac{1}{2}\delta_{(1,1)} + \frac{1}{2}\delta_{(-1,-1)},$$

where $\mu_{x_1^n} \to \mu = \frac{1}{2}\delta_{(0,1)} + \frac{1}{2}\delta_{(0,-1)}$. *Then* $\pi_n = \frac{1}{2}\delta_{(1/n,1,1,1)} + \frac{1}{2}\delta_{(-1/n,-1,-1,-1)} \in \Pi_{bc}(\mu_{x_1^n}, \nu_{y_1^n})$ *and converges weakly to* $\pi = \frac{1}{2}\delta_{(0,1,1,1)} + \frac{1}{2}\delta_{(0,-1,-1,-1)}$, *which is not bicausal.*

Lemma 4.2. *Let* $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$. *Let* $D := \{(x_1, y_1, \pi) : x_1, y_1 \in \mathbb{R}, \pi \in \Pi(\mu_{x_1}, \nu_{y_1})\}$. *Let* $f : \mathbb{R}^2 \times \mathbb{R}^2 \to \overline{\mathbb{R}}$ *be upper semianalytic and nonnegative. Then*

$$\sup_{\pi\in\Pi_{bc}(\mu,\nu)} \int f d\pi = \sup_{\gamma\in\Pi(\mu^1,\nu^1)} \int \sup_{\lambda\in\Pi(\mu_{x_1},\nu_{y_1})} \int f d\lambda d\gamma. \tag{4.1}$$

Proof. Let $\tilde{f} : \mathbb{R} \times \mathbb{R} \times \mathcal{P}(\mathbb{R}^2) \to \overline{\mathbb{R}}, (x_1, y_1, \lambda) \mapsto \int f d\lambda$. We can use Proposition 7.48 in [BS96] to see that $\tilde{f}$ is upper semianalytic, choosing, using the notation of Proposition 7.48 in [BS96], X to be $\mathbb{R} \times \mathbb{R} \times \mathcal{P}(\mathbb{R}^2)$ and Y to be $\mathbb{R} \times \mathbb{R}$. Let $D_{(x_1,y_1)} := \{\lambda \in \mathcal{P}(\mathbb{R} \times \mathbb{R}) : (x_1, y_1, \lambda) \in D\}$. By Proposition 7.47 in [BS96] $(x_1, y_1) \mapsto \sup_{\pi\in D_{(x_1,y_1)}} \tilde{f}(x_1, y_1, \pi)$ is upper semianalytic as well as D is an analytic set by Lemma 4.1. In particular $(x_1, y_1) \mapsto \sup_{\pi\in D_{(x_1,y_1)}} \int f d\pi$ is universally measurable. This combined with nonnegativity shows that the integral on the right hand side of (4.1) is well defined.

By Proposition 5.1 in [BBLZ17] we have that

$$\sup_{\pi\in\Pi_{bc}(\mu,\nu)} \int f d\pi = \sup_{\substack{\gamma\in\Pi(\mu^1,\nu^1) \\ \lambda:\mathbb{R}\times\mathbb{R}\to\Pi(\mu_\cdot,\nu_\cdot) univ.meas.}} \int \int f d\lambda d\gamma.$$

Now we can use Proposition 7.50 in [BS96], again because D is an analytic set. So for arbitrarily chosen $\varepsilon > 0$, we find a universally measurable function $\lambda_\varepsilon : \mathbb{R}^2 \to \mathcal{P}(\mathbb{R}^2)$ with $Gr(\lambda_\varepsilon) \subset D$ such that

$$\tilde{f}(x_1, y_1, \lambda_\varepsilon(x_1, y_1)) > \sup_{\pi\in\Pi(\mu_{x_1},\nu_{y_1})} \tilde{f}(x_1, y_1, \pi) - \varepsilon$$

for all (x_1, y_1) for which $\sup_{\pi\in\Pi(\mu_{x_1},\nu_{y_1})} \tilde{f}(x_1, y_1, \pi) < \infty$ and

$$\tilde{f}(x_1, y_1, \lambda_\varepsilon(x_1, y_1)) > 1/\varepsilon \text{ otherwise.}$$

Therefore, if

$$\pi(\{(x_1, y_1) : \sup_{\gamma\in\Pi(\mu_{x_1},\nu_{y_1})} \tilde{f}(x_1, y_1, \gamma) < \infty\}) = 1 \; \forall \; \gamma \in \Pi(\mu^1, \nu^1), \tag{4.2}$$

we have that

$$\begin{aligned}\sup_{\substack{\pi\in\Pi(\mu^1,\nu^1) \\ \lambda:\mathbb{R}\times\mathbb{R}\to\Pi(\mu_\cdot,\nu_\cdot) univ.meas.}} \int \int f d\lambda d\pi &\geqslant \sup_{\pi\in\Pi(\mu^1,\nu^1)} \int \tilde{f}(x_1, y_1, \lambda_\varepsilon(x_1, y_1)) d\pi \\ &\geqslant \sup_{\pi\in\Pi(\mu^1,\nu^1)} \int \sup_{\gamma\in\Pi(\mu_{x_1},\nu_{y_1})} \tilde{f}(x_1, y_1, \gamma) d\pi - \varepsilon.\end{aligned}$$

Letting ε go to zero gives us that the LHS in (4.1) is greater or equal than the RHS in this case. If (4.2) does not hold true both expressions in (4.1) are equal to ∞ as we can make $\tilde{f}(x_1, y_1, \lambda_\varepsilon(x_1, y_1))$ arbitrarily large on a set of positive measure for some coupling π.

To see that the LHS in (4.1) is smaller or equal than the RHS assume that there exists a bicausal coupling $\gamma = \bar{\gamma}(dx_1, dy_1)\gamma_{x_1,y_1}(dx_2, dy_2)$ such that $\int f d\gamma$ is strictly greater than the RHS. By Proposition 5.1 in [BBLZ17] we have that $\bar{\gamma} \in \Pi(\mu^1, \nu^1)$ and $\gamma_{x_1,y_1} \in \Pi(\mu_{x_1}, \nu_{y_1})$. Choosing $\bar{\gamma}$ and γ_{x_1,y_1} as couplings on the RHS immediately leads to a contradiction.

□

Theorem 4.3. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$. Let $f : \mathbb{R}^2 \times \mathbb{R}^2 \to \overline{\mathbb{R}}$ be upper semianalytic and nonnegative. Then the following are equivalent:*

1. $\sup_{\pi \in \Pi_{bc}(\mu,\nu)} \int f d\pi = 0$
2. *There exist nonnegative functions g, h, g^{x_1,y_1} and h^{x_1,y_1} with $g = 0$ μ^1-a.s, $h = 0$ ν^1-a.s., $g^{x_1,y_1} = 0$ μ_{x_1}-a.s. and $h^{x_1,y_1} = 0$ ν_{y_1}-a.s. such that*
$$f(x_1, x_2, y_1, y_2) \leqslant g(x_1) + h(y_1) + g^{x_1,y_1}(x_2) + h^{x_1,y_1}(y_2)$$
for all $(x_1, x_2, y_1, y_2) \in \mathbb{R}^2 \times \mathbb{R}^2$.

Proof. 1. $\Longrightarrow$ 2.: Using equation (4.1), by Theorem 2.21 in [Kel84] this implies that there exist functions g and h with $g = 0$ μ^1-a.s. and $h = 0$ ν^1-a.s. such that $\sup_{\lambda \in \Pi(\mu_{x_1}, \nu_{y_1})} \int f d\lambda \leqslant g(x_1) + h(y_1)$ for all $(x_1, y_1) \in \mathbb{R}^2$. We set the functions equal to ∞ for all points on which they do not vanish. Repeating the same argument, for $(x_1, y_1) \in A := \{(x_1, y_1) : g(x_1) + h(y_1) = 0\}$, we get functions g^{x_1,y_1} and h^{x_1,y_1} with $g^{x_1,y_1} = 0$ μ_{x_1}-a.s. and $h^{x_1,y_1} = 0$ ν_{y_1}-a.s. such that $f(x_1, x_2, y_1, y_2) \leqslant g^{x_1,y_1}(x_2) + h^{x_1,y_1}(y_2)$ for all $(x_2, y_2) \in \mathbb{R}^2$ and $(x_1, y_1) \in A$. Therefore we get that
$$f(x_1, x_2, y_1, y_2) \leqslant g(x_1) + h(y_1) + g^{x_1,y_1}(x_2) + h^{x_1,y_1}(y_2)$$
for all $(x_1, x_2, y_1, y_2) \in \mathbb{R}^2 \times \mathbb{R}^2$.

2. $\Longrightarrow$ 1.: Take $\pi \in \Pi_{bc}(\mu, \nu)$. By Proposition 5.1 in [BBLZ17] it is of the form $\pi(dx_1, dx_2, dy_1, dy_2) = \bar{\pi}(dx_1, dy_1)\pi_{x_1,y_1}(dx_2, dy_2)$ with $\bar{\pi}(dx_1, dy_1) \in \Pi(\mu^1, \nu^1)$ and $\pi_{x_1,y_1}(dx_2, dy_2) \in \Pi(\mu_{x_1}, \nu_{y_1})$. So we have
$$\int f d\pi \leqslant \int \int g \oplus h \oplus g^{x_1,y_1} \oplus h^{x_1,y_1} d\pi_{x_1,y_1} d\bar{\pi} = 0.$$

□

5 Bicausal 0-sets for general N

For the following Lemma let $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$ and let $\tilde{D}_N^M$ be a countable dense subset of $\mathcal{C}_b([-M, M]^N)$, endowed with the topology induced by $\|\cdot\|_\infty$, which is possible as $\mathcal{C}_b(K)$ is separable for K compact. For every $\tilde{g} \in \tilde{D}_N^M$ we can choose a continuous extension g to $\mathbb{R}^N$, such that $\|g\|_\infty \leqslant \|\tilde{g}\|_\infty$ by Tietze's extension Theorem. Let D_N^M be the union of these g and $D_N = \bigcup_{M \in \mathbb{N}} D_N^M$. Let

$$\tilde{\mathbb{F}} = \left\{ \begin{array}{c} F : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R} : F(x_1, ..., x_N, y_1, ...y_N) = \\ \sum_{t<N} h_t(y_1, ..., y_t)[g_t(x_1, ..., x_N) \\ - \int g_t(x_1, ...x_t, \bar{x}_{t+1}, ..., \bar{x}_N)\mu_{x_1,...,x_t}(d\bar{x}_{t+1}, ..., d\bar{x}_N)] + \\ \sum_{t<N} h'_t(x_1, ..., x_t)[g'_t(y_1, ..., y_N) \\ - \int g'_t(y_1, ...y_t, \bar{y}_{t+1}, ..., \bar{y}_N)\nu_{y_1,...,y_t}(d\bar{y}_{t+1}, ..., d\bar{y}_N)] \\ g_t, g'_t \in D_N, h_t, h'_t \in D_t \ \forall \ t < N \end{array} \right\}.$$

Lemma 5.1.

$$\pi \in \Pi_{bc}(\mu, \nu) \iff \int f d\pi = 0 \textit{ for all } f \in \tilde{\mathbb{F}}.$$

Proof. $\implies$: Clear by the characterization of bicausal couplings in Proposition 2.3 in [BBLZ17].
$\impliedby$: Again by Proposition 2.3 in [BBLZ17] and by symmetry it is sufficient to show that, for fixed bicausal π and $t < N$,

$$\int \tilde{h}_t(y_1, ..., y_t)\Big(\tilde{g}_t(x_1, ..., x_N) \\ - \int \tilde{g}_t(x_1, ..., x_t, \bar{x}_{t+1}, ..., \bar{x}_N)\mu_{x_1,...,x_t}(d\bar{x}_{t+1}, ..., d\bar{x}_N)d\pi\Big) = 0$$

for all $\tilde{h}_t \in \mathcal{C}_b(\mathbb{R}^t)$ and $\tilde{g}_t \in \mathcal{C}_b(\mathbb{R}^N)$.

Let $\tilde{h}_t \in \mathcal{C}_b(\mathbb{R}^t)$ and $\tilde{g}_t \in \mathcal{C}_b(\mathbb{R}^N)$ be arbitrary and let $\gamma := \max\{\|\tilde{g}_t\|, \|\tilde{h}_t\|\}$.

Due to Lusin's Theorem we find a compact set $K_t \subset \mathbb{R}^N$ such that $(x_1, ..., x_N) \mapsto \mu_{x_1,...,x_t}$ is continuous on K_t and $\pi(K_t^\mathsf{c}) < \varepsilon$. Choose a hypercube $\hat{K}$ with $K_t \subseteq \hat{K}$. By Prokhorov's Theorem, the compactness of K_t gives us that $\{\mu_{x_1,...,x_t} : (x_1, ..., x_t) \in K_t\}$ is tight and we can find a hypercube $\tilde{K}$ such that $\mu_{x_1,...,x_t}(\tilde{K}^\mathsf{c}) < \varepsilon$ for all $(x_1, ..., x_t) \in K_t$. Let $K = \hat{K} \cup \tilde{K}$. Now we can choose $g_t \in D_N$ and $h_t \in D_t$, such that $\|\tilde{g}_t - g_t\| < \varepsilon$ and $\|\tilde{h}_t - h_t\| < \varepsilon$ on K. W.l.o.g. $\|g_t\| \leqslant \|\tilde{g}_t\|$ and $\|h_t\| \leqslant \|\tilde{h}_t\|$.

Due to our assumption we have that

$$\int \tilde{h}_t\Big(\tilde{g}_t - \int \tilde{g}_t d\mu_{x_1,...,x_t}\Big)d\pi = \\ \int \tilde{h}_t \left(\tilde{g}_t - \int \tilde{g}_t d\mu_{x_1,...,x_t}\right) d\pi - \int h_t \left(g_t - \int g_t d\mu_{x_1,...,x_t}\right) d\pi.$$

Notice that

$$\left|\int_{K_t^c} \tilde{h}_t \left(\tilde{g}_t - \int \tilde{g}_t d\mu_{x_1,...,x_t}\right) d\pi - \int_{K_t^c} h_t \left(g_t - \int g_t d\mu_{x_1,...,x_t}\right) d\pi\right| \leqslant \int_{K_t^c} 4\gamma^2 d\pi \leqslant 4\varepsilon\gamma^2.$$

We also have that

$$\left|\int_{K_t} \tilde{h}_t \tilde{g}_t - h_t g_t d\pi\right| \leqslant 2\varepsilon\gamma, \tag{5.1}$$

using that

$$\|\tilde{h}_t \tilde{g}_t - h_t g_t\| \leqslant \|\tilde{h}_t \tilde{g}_t - \tilde{h}_t g_t\| + \|\tilde{h}_t g_t - h_t g_t\| \leqslant \|\tilde{h}_t\| \|\tilde{g}_t - g_t\| + \|g_t\| \|\tilde{h}_t - h_t\|.$$

Due to the tightness of $\{\mu_{x_1,...,x_t} : (x_1, ..., x_t) \in K_t\}$ we have that

$$\left|\int_{K_t} h_t \int_{\tilde{K}^c} g_t d\mu_{x_1,...,x_t} - \tilde{h}_t \int_{\tilde{K}^c} \tilde{g}_t d\mu_{x_1,...,x_t} d\pi\right| \leqslant 2\varepsilon\gamma^2.$$

Notice that $\int_{\tilde{K}} g_t - \tilde{g}_t d\mu_{x_1,...,x_t} \leqslant \varepsilon$ and by the same argument as in (5.1) we get

$$\left|\int_{K_t} h_t \int_{\tilde{K}} g_t d\mu_{x_1,...,x_t} - \tilde{h}_t \int_{\tilde{K}} \tilde{g}_t d\mu_{x_1,...,x_t} d\pi\right| \leqslant 2\varepsilon\gamma.$$

All in all we have

$$\left|\int \tilde{h}_t \left(\tilde{g}_t - \int \tilde{g}_t d\mu_{\tilde{x}_1,...,x_t}\right) d\pi\right| \leqslant \varepsilon(6\gamma^2 + 4\gamma).$$

Therefore it is equal to zero, by the arbitrary choice of ε. □

Lemma 5.2. *The set $B_N := \{\pi \in \Pi_{bc}(\mu, \nu) : \mu, \nu \in \mathcal{P}(\mathbb{R}^N)\}$ is Borel.*

Proof. For $h \in \mathcal{C}_b(\mathbb{R}^t)$ and $g \in \mathcal{C}_b(\mathbb{R}^N)$ let $\phi^t_{g,h} : \mathcal{P}(\mathbb{R}^N \times \mathbb{R}^N) \to \overline{\mathbb{R}}$,

$$\pi \mapsto \int h(y_1, ..., y_t)\Big(g(x_1, ..., x_N) - \int g(x_1, ..., x_t, \tilde{x}_{t+1}, ..., \tilde{x}_N) d(p^1_\# \pi)_{x_1,...,x_t}(\tilde{x}_{t+1}, ..., \tilde{x}_N)\Big) d\pi$$

and $\psi^t_{g,h} : \mathcal{P}(\mathbb{R}^N \times \mathbb{R}^N) \to \overline{\mathbb{R}}$,

$$\pi \mapsto \int h(x_1, ..., x_t)\Big(g(y_1, ..., y_N) - \int g(y_1, ..., y_t, \tilde{y}_{t+1}, ..., \tilde{y}_N) d(p^2_\# \pi)_{y_1,...,y_t}(\tilde{y}_{t+1}, ..., \tilde{y}_N)\Big) d\pi.$$

By Lemma 5.1 we know that

$$B_N = \bigcap_{t<N} \bigcap_{g \in D_N} \bigcap_{h \in D_t} (\phi^t_{g,h})^{-1}(0) \cap \bigcap_{t<N} \bigcap_{g \in D_N} \bigcap_{h \in D_t} (\psi^t_{g,h})^{-1}(0).$$

By Proposition 7.29 in [BS96] both $\phi^t_{g,h}$ and $\psi^t_{g,h}$ are measurable functions. Hence B_N is Borel as it is a countable intersection of Borel sets. □

Lemma 5.3. *Let $\mathcal{X}$ and $\mathcal{Y}$ be closed subsets of $\mathbb{R}^N$. Then $D = \{(x_1, y_1, \pi) : x_1 \in p^1(\mathcal{X}), y_1 \in p^1(\mathcal{Y}), \pi \in \Pi_{bc}(\mu_{x_1}, \nu_{y_1})\}$ is an analytic set.*

Proof. Using the set B_N from the previous Lemma we get that

$$D = \big(p^1(\mathcal{X}) \times p^1(\mathcal{Y}) \times B_{N-1}\big) \\ \cap \{(x_1, y_1, \pi) : x_1 \in p^1(\mathcal{X}), y_1 \in p^1(\mathcal{Y}), \pi \in \Pi(\mu_{x_1}, \nu_{y_1})\}.$$

Note that $p^1(\mathcal{X})$ and $p^1(\mathcal{Y})$ are both analytic sets, as they are projections of Borel sets. By Lemma 5.2, Lemma 4.1 and the fact that finite products and finite intersections of analytic sets are still analytic we get that D is an analytic set. □

Lemma 5.4. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$. Let $D = \{(x_1, y_1, \pi) : x_1 \in \mathbb{R}, y_1 \in \mathbb{R}, \pi \in \Pi_{bc}(\mu_{x_1}, \nu_{y_1})\}$. Let $f : \mathbb{R}^N \times \mathbb{R}^N \to \overline{\mathbb{R}}$ be upper semianalytic and nonnegative. Then*

$$\sup_{\pi \in \Pi_{bc}(\mu,\nu)} \int f d\pi = \sup_{\gamma \in \Pi(\mu^1,\nu^1)} \int \sup_{\lambda \in \Pi_{bc}(\mu_{x_1},\nu_{y_1})} \int f d\lambda d\gamma. \tag{5.2}$$

Proof. By Proposition 5.1 in [BBLZ17] we have that π is a bicausal coupling of μ and ν iff $\pi = \bar{\pi}\pi_{x_1,y_1}$ with $\bar{\pi} \in \Pi(\mu^1, \nu^1)$ and $\pi_{x_1,y_1} \in \Pi(\mu_{x_1}, \nu_{y_1})$ $\bar{\pi}$-a.s. Hence

$$\sup_{\pi \in \Pi_{bc}(\mu,\nu)} \int f d\pi = \sup_{\substack{\gamma \in \Pi(\mu^1,\nu^1) \\ \lambda : \mathbb{R}\times\mathbb{R} \to \Pi_{bc}(\mu_\cdot,\nu_\cdot) univ.meas.}} \int \int f d\lambda d\gamma.$$

Let $\tilde{f} : \mathbb{R} \times \mathbb{R} \times \mathcal{P}(\mathbb{R}^{N-1} \times \mathbb{R}^{N-1}) \to \overline{\mathbb{R}}$ be defined by

$$(x_1, y_1, \lambda) \mapsto \int f(x_1, \tilde{x}_2, ..., \tilde{x}_N, y_1, \tilde{y}_2, ..., \tilde{y}_N)\lambda(d\tilde{x}_2, ..., d\tilde{x}_N, d\tilde{y}_2, ..., d\tilde{y}_N).$$

We can use Proposition 7.48 in [BS96] to see that $\tilde{f}$ is upper semianalytic, choosing, using the notation of Proposition 7.48 in [BS96], X to be $\mathbb{R} \times \mathbb{R} \times \mathcal{P}(\mathbb{R}^{N-1} \times \mathbb{R}^{N-1})$ and Y to be $\mathbb{R}^{N-1} \times \mathbb{R}^{N-1}$. Let $D_{(x_1,y_1)} := \{\lambda \in \mathcal{P}(\mathbb{R}^{N-1} \times \mathbb{R}^{N-1}) : (x_1, y_1, \lambda) \in D\}$. By Proposition 7.47 in [BS96], $(x_1, y_1) \mapsto \sup_{\lambda \in D_{(x_1,y_1)}} \tilde{f}(x_1, y_1, \lambda)$ is upper semianalytic as well, as D is an analytic set by Lemma 5.3. In particular $(x_1, y_1) \mapsto \sup_{\lambda \in D_{(x_1,y_1)}} \int f d\lambda = \sup_{\lambda \in \Pi(\mu_{x_1},\nu_{y_1})} \int f d\lambda$ is universally measurable. This combined with nonnegativity shows that the integral on the right hand side of (5.2) is well defined.

By Proposition 7.50 in [BS96] (again using that D is an analytic set), for arbitrarily chosen $\varepsilon > 0$, we find a universally measurable function $\lambda_\varepsilon : \mathbb{R}^2 \to \mathcal{P}(\mathbb{R}^{N-1} \times \mathbb{R}^{N-1})$ with $Gr(\lambda) \subseteq D$ such that

$$\tilde{f}(x_1, y_1, \lambda_\varepsilon(x_1, y_1)) > \sup_{\pi \in \Pi_{bc}(\mu_{x_1},\nu_{y_1})} \tilde{f}(x_1, y_1, \pi) - \varepsilon$$

for all (x_1, y_1) for which $\sup_{\pi \in \Pi_{bc}(\mu_{x_1},\nu_{y_1})} \tilde{f}(x_1, y_1, \pi) < \infty$ and

$$\tilde{f}(x_1, y_1, \lambda_\varepsilon(x_1, y_1)) > 1/\varepsilon$$

otherwise.

Therefore, if

$$\pi(\{(x_1, y_1) : \sup_{\gamma\in\Pi_{bc}(\mu_{x_1},\nu_{y_1})} \tilde{f}(x_1, y_1, \gamma) < \infty\}) = 1 \text{ for all } \pi \in \Pi(\mu^1, \nu^1), \tag{5.3}$$

then

$$\sup_{\substack{\pi\in\Pi(\mu^1,\nu^1)\\ \lambda:\mathbb{R}\times\mathbb{R}\to\Pi_{bc}(\mu_\cdot,\nu_\cdot) univ.meas.}} \int\int f d\lambda d\pi \geqslant \sup_{\pi\in\Pi(\mu^1,\nu^1)} \int \tilde{f}(x_1, y_1, \lambda_\varepsilon(x_1, y_1))d\pi$$

$$\geqslant \sup_{\pi\in\Pi(\mu^1,\nu^1)} \int \sup_{\gamma\in\Pi_{bc}(\mu_{x_1},\nu_{y_1})} \tilde{f}(x_1, y_1, \gamma)d\pi - \varepsilon.$$

Letting ε go to zero gives us that the LHS in (4.1) is greater or equal than the RHS in this case. In the case of (5.3) not holding true both expressions in (4.1) are equal to ∞ as we can make $\tilde{f}(x_1, y_1, \lambda_\varepsilon(x_1, y_1))$ arbitrarily large on a set of positive measure for some coupling π.

Assume there exists a coupling $\gamma \in \Pi_{bc}(\mu, \nu)$, such that the LHS in (5.2) is strictly greater. Choosing $\bar{\gamma}$ and γ_{x_1,y_1} immediately leads to a contradiction.

□

Theorem 5.5. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$. Let $f : \mathbb{R}^N \times \mathbb{R}^N \to \overline{\mathbb{R}}$ be upper semianalytic and nonnegative. Then the following are equivalent:*

1. $\sup_{\pi\in\Pi_{bc}(\mu,\nu)} \int f d\pi = 0$
2. *There exist nonnegative functions $g, h, g^{x_1,y_1}, h^{x_1,y_1}, ..., g^{x_1,...,x_{N-1},y_1,...,y_{N-1}}$, $h^{x_1,...,x_{N-1},y_1,...,y_{N-1}}$ with $g = 0$ μ^1-a.s, $h = 0$ ν^1-a.s, $g^{x_1,...,x_i,y_1,...,y_i} = 0$ $\mu_{x_1,...,x_i}$-a.s. and $h^{x_1,...,x_i,y_1,...,y_i}$ $\nu_{y_1,...,y_i}$-a.s for $i < N$ such that*

$$f(x_1, ...x_N, , y_1, ..., y_N) \leqslant g(x_1) + h(y_1) + \sum_{i=1}^{N-1} g^{x_1,...,x_i,y_1,...,y_i}(x_{i+1}) + h^{x_1,...,x_i,y_1,...,y_i}(y_{i+1}) \tag{5.4}$$

for all $(x_1, ..., x_N, y_1, ..., y_N) \in \mathbb{R}^N \times \mathbb{R}^N$.

Proof. 1. $\implies$ 2.: We will do an induction over N. The implication holds true by Theorem 4.3 for $N = 2$. Assume it holds true for $N - 1$. Using equation (5.2), by Theorem 2.21 in [Kel84] this implies that there exist functions g and h with $g = 0$ μ^1-a.s. and $h = 0$ ν^1-a.s. such that $\sup_{\lambda\in\Pi_{bc}(\mu_{x_1},\nu_{y_1})} \int f d\lambda \leqslant g(x_1)+h(y_1)$ for all $(x_1, y_1) \in \mathbb{R}^2$. We set the functions equal to ∞ for all points on which they do not vanish. By the induction hypothesis, for pairs (x_1, y_1) for which $g(x_1) + h(y_1)$ vanishes, there exist functions $g^{x_1,...,x_i,y_1,...,y_i}$ and $h^{x_1,...,x_i,y_1,...,y_i}$ for $i < N$ such that

$$f(x_1, ..., x_N, y_1, ..., y_N) \leqslant \sum_{i=1}^{N-1} g^{x_1,x_2,...,x_i,y_1,...,y_i}(x_{i+1}) + h^{x_1,x_2,...,x_i,y_1,...,y_i}(y_{i+1})$$

for all $(x_2, ..., x_N, y_2, ..., y_N) \in \mathbb{R}^{N-1} \times \mathbb{R}^{N-1}$.

For all other points the inequality (5.4) holds true because for these points (x_1, y_1) we either set $g(x_1) = \infty$ or $h(y_1) = \infty$.

2. $\implies$ 1.: We prove this direction by induction over N as well. For $N = 2$ the implication is true due to Theorem 4.3. By Lemma 5.4 we have that

$$\begin{aligned}\sup_{\pi\in\Pi_{bc}(\mu,\nu)}\int f d\pi &= \sup_{\bar\pi\in\Pi(\mu^1,\nu^1)}\int \sup_{\lambda\in\Pi_{bc}(\mu_{x_1},\nu_{y_1})}\int f d\lambda d\bar\pi \\ &\leqslant \sup_{\bar\pi\in\Pi(\mu^1,\nu^1)}\int g(x_1)+h(y_1) \\ &+ \sup_{\lambda\in\Pi_{bc}(\mu_{x_1},\nu_{y_1})}\int \sum_{i=1}^{N-1} g^{x_1,\dots,x_i,y_1,\dots,y_i}(x_{i+1}) \\ &+ h^{x_1,\dots,x_i,y_1,\dots,y_i}(y_{i+1}) d\lambda d\bar\pi.\end{aligned} \tag{5.5}$$

Therefore for all (x_1, y_1) for which

$$f(x_1, ..., x_N, y_1, ..., y_N) \leqslant \sum_{i=1}^{N-1} g^{x_1,\dots,x_i,y_1,\dots,y_i}(x_{i+1}) + h^{x_1,\dots,x_i,y_1,\dots,y_i}(y_{i+1}) \tag{5.6}$$

for all $(x_2, ..., x_N, y_2, ..., y_N) \in \mathbb{R}^{N-1} \times \mathbb{R}^{N-1}$, we have by the induction hypothesis that $\sup_{\lambda\in\Pi_{bc}(\mu_{x_1},\nu_{y_1})} \int f d\lambda = 0$. As g and h vanish μ^1 respectively ν^1 a.s., (5.6) holds true $\bar\pi$-a.s. Combining this with (5.5) we get the desired result. □

6 Counterexamples

By Theorem 2.9 in [BBLZ17], the causal and bicausal problem coincide in the case that the starting measure μ is the product of its marginals and the cost function has a separable structure (i.e. is of the form $c(x_1,...,x_N,y_1,...,y_N) = \sum_{i=1}^N c_i(x_i,y_i)$). The following two examples show that one cannot drop either of the two assumptions, using the results from the last chapters. Although there are already counterexamples in [BBLZ17], we will give some other examples which are more intuitive. In the following examples we denote the Lebesgue measure on $[0,1]$ by λ.

Example 1. *Consider the non-separable cost function $c = \mathbb{1}_{C^c}$ for $C = \{(x_1,x_2,x_3,x_1) : x_1,x_2,x_3 \in [0,1]\}$ and $\mu = \nu = \lambda \otimes \lambda$. Then*

$$\inf_{\pi\in\Pi_c(\mu,\nu)} \int \mathbb{1}_{C^c} d\pi = 0, \ \ but \inf_{\pi\in\Pi_{bc}(\mu,\nu)} \int \mathbb{1}_{C^c} d\pi = 1. \tag{6.1}$$

The first equality in (6.1) holds true as we can choose the causal coupling $\pi = f_\#(\lambda\otimes\lambda\otimes\lambda)$ for $f(x_1,x_2,x_3) := (x_1,x_2,x_3,x_1)$ which is supported on C. Heuristically π is causal, as the third component is independent both from the first and the second component. The idea why π is not bicausal is that knowing the third component, the fourth component tells us exactly what the first component should be. To be more precise on why the coupling is causal we can use the characterization from Proposition 2.3 in [BBLZ17]. It is easy to verify that we can decompose π into $\bar{\pi}(dx_1,dy_1)\pi_{x_1,y_1}(dx_2,dy_2)$ with $\bar{\pi} = \lambda\otimes\lambda$ and $\pi_{x_1,y_1} = \lambda\otimes\delta_{x_1}$. Hence $\bar{\pi} \in \Pi(\mu^1,\nu^1)$, $(p^1)_\#(\pi_{x_1,y_1}) = \mu_{x_1} = \lambda$ and for $A \subset [0,1]$ measurable

$$\int_{y_2} \mathbb{1}_A \pi_{y_1}(dy_2) = \int_{x_1}\int_{y_2} \mathbb{1}_A \pi_{x_1,y_1}(dy_2)\pi_{y_1}(dx_1) = \lambda(A) = \nu_{y_1}(A).$$

By Proposition 2.3 in [BBLZ17] this implies that π is a causal coupling between μ and ν.

For the second equality in (6.1) *notice that the functions $g = h = g^{x_1,y_1} = 0$ and $h^{x_1,y_1} = \mathbb{1}_{\{x_1\}}$ fulfill the requirements given in Theorem 4.3 for $f = \mathbb{1}_C$ (i.e. $g = 0$ μ^1-a.s., $h = 0$ ν-a.s., $g^{x_1,y_1} = 0$ μ_{x_1}-a.s, $h^{x_1,y_1} = 0$ ν_{y_1}-a.s. and $\mathbb{1}_C \leqslant g \oplus h \oplus g^{x_1,y_1} \oplus h^{x_1,y_1}$) and therefore for all γ bicausal $\int \mathbb{1}_C d\gamma = 0$. Hence $\int \mathbb{1}_{C^c} d\gamma = 1$.*

The next example shows that we can also not drop the assumption that μ is a product of its marginals

Example 2. *Let $\mu = f_\#\lambda$ for $f(x) := (x,x)$, $\nu = \delta_0 \otimes \lambda$ and $c := \mathbb{1}_{\{x_2\neq y_2\}}$. Then*

$$\inf_{\pi\in\Pi_c(\mu,\nu)} \int c d\pi = 0, \tag{6.2}$$

whereas

$$\inf_{\pi\in\Pi_{bc}(\mu,\nu)} \int c d\pi = 1.$$

We can see that equation (6.2) holds true by choosing $\pi = g_{\#}\lambda$ *for* $g(x) = (x, x, 0, x)$.

For the bicausal case notice that the functions $g = h = 0$, $g^{x_1,y_1}(x_2) = 1 - \mathbb{1}_{\{x_1\}}$ *and* $h^{x_1,y_1}(y_2) = \mathbb{1}_{\{x_1\}}$ *fulfill all the requirements from Theorem 4.3, choosing* $f(x_1, x_2, y_1, y_2) = \mathbb{1}_{\{x_2=y_2\}}$. *Therefore*

$$\sup_{\pi\in\Pi_{bc}(\mu,\nu)} \int \mathbb{1}_{\{x_2=y_2\}} d\pi = 0.$$

This gives us that for every bicausal coupling π

$$\int \mathbb{1}_{\{x_2\neq y_2\}} d\pi = 1$$

and therefore we have that

$$\inf_{\pi\in\Pi_{bc}(\mu,\nu)} \int \mathbb{1}_{\{x_2\neq y_2\}} d\pi = 1.$$

7 Causal 0-sets

From now on, for Polish spaces X, Y, $m \in \mathcal{P}(X)$ and $\mu \in \mathcal{P}(X \times Y)$ we will also use the notation

$$\int_X m(dx)\mu_x(dy)$$

for the Bochner integral $\int_X \mu_x(dy)m(dx)$ and in the same way we will also frequently interchange the integrand and the underlying measure in usual integrals, if convenient.

Apart from Theorem 7.4 the results in this chapter and the idea of their proofs resemble [BBLZ17]. Nevertheless we cannot directly use the results from [BBLZ17] as we are interested in obtaining slightly different ones (considerung suprema instead of infima). These considerations will lead to Theorem 7.4, which can be seen as the equivalent to Theorem 5.5, considering causal instead of bicausal couplings.

Lemma 7.1. *Let $\mu \in \mathcal{P}(\mathbb{R}^N)$. Then, for $t \leqslant N$, $f : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ given by $m \mapsto \int m(dx_{t-1})\mu_{x_{t-1}}(dx_t)$ is measurable.*

Proof. For $\mu_0 \in \mathcal{P}(\mathbb{R})$, $g_i \in C_b(\mathbb{R})$ and $\varepsilon > 0$ let

$$N(\mu_0, g_1, ..., g_k, \varepsilon) = \left\{ \mu : \max_{1 \leqslant i \leqslant k} \left| \int g_i d\mu - \int g_i d\mu_0 \right| < \varepsilon \right\}.$$

Recall that $\{N(\mu_0, g_1, ..., g_k, \varepsilon) : \mu_0 \in \mathcal{P}(\mathbb{R}), g_1, ..., g_k \in \mathcal{C}_b(\mathbb{R}), \varepsilon > 0\}$ is a basis for the weak topology. As $\mathcal{P}(\mathbb{R})$ is a Polish space, we can write every open set as a countable union of base elements. We will show that the preimage of every base element is measurable, then it easily follows that every preimage of an element of its generated sigma algebra is also measurable. So we look at

$$\begin{aligned} &f^{-1}(N(\mu_0, g_1, ..., g_k, \varepsilon)) \\ &\quad = \left\{ m \in \mathcal{P}(\mathbb{R}) : \max_{1 \leqslant i \leqslant k} \left| \int_{x_{t-1}} \int_{x_t} g_i(x_t)\mu_{x_{t-1}}(dx_t)m(dx_{t-1}) - \int g_i d\mu_0 \right| < \varepsilon \right\}. \end{aligned}$$

By Proposition 7.29 in [BS96] we have that $\tilde{g}(x_{t-1}) := \int g_i(x_t)\mu_{x_{t-1}}(dx_t)$ is measurable and it is bounded as g_i is bounded. By Corollary 7.29.1 in [BS96] $m \mapsto \int \tilde{g} dm$ is measurable as well. Therefore $f^{-1}(N(\mu_0, g_1, ..., g_k, \varepsilon))$ is measurable.

□

Definition 6. Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$. We call $\pi \in \Pi_c(\mu, \nu)$ causal quasi-Markov, denoted by $\pi \in \Pi_{cqm}(\mu, \nu)$, if for $1 \leqslant t \leqslant N - 1$ we have that

$$\pi_{x_1,...,x_t,y_1,...,y_t}(dx_{t+1}, dy_{t+1}) = \pi_{x_t,y_1,...,y_t}(dx_{t+1}, dy_{t+1}).$$

Remark 4. *The set of causal quasi-Markov couplings is non-empty if μ is a Markov measure, but there may also be causal quasi-Markov couplings if μ is non-Markov. This fixes a little inaccuracy in [BBLZ17].*

If μ is a Markov measure we can take the independent coupling of μ and ν, which is obviously quasi-Markov.

To see that we can have quasi-Markov couplings, even though μ is not Markov, take $N > 2$ and let $\mu = \nu \in \mathcal{P}(\mathbb{R}^N)$ be non-Markov. Then taking the coupling $\pi = f_{\#}\mu$, for

$$f(x_1, ..., x_N) = (x_1, ..., x_N, x_1, ..., x_N),$$

we get that

$$\pi_{x_1,...,x_t,y_1,...,y_t}(dx_{t+1}, dy_{t+1}) = \pi_{y_1,...,y_t}(dx_{t+1}, dy_{t+1}) = \pi_{x_t,y_1,...,y_t}(dx_{t+1}, dy_{t+1}).$$

Theorem 7.2. *Let μ, $\nu \in \mathcal{P}(\mathbb{R}^N)$. Let μ be Markov and c be semiseparable (i.e. $c = \sum_{t=1}^N c_t(x_t, y_1, ..., y_t)$ for c_t measurable and nonnegative). We set $V_N^c = 0$ and define recursively for $t = N, ..., 2$:*

$$\begin{aligned} V_{t-1}^c(y_1, ..., y_{t-1}, m(dx_{t-1})) = \\ \sup_{\gamma \in \Pi(\int_{x_{t-1}} m(dx_{t-1})\mu_{x_{t-1}}(dx_t), \nu_{y_1,...,y_{t-1}}(dy_t))} \int \gamma(dx_t, dy_t)\big(c(x_t, y_1, .., y_t) \\ + V_t^c(y_1, ..., y_t, \gamma_{y_t}(dx_t))\big). \end{aligned} \tag{7.1}$$

If we set

$$V_0^c = \sup_{\gamma \in \Pi(\mu^1, \nu^1)} \int \gamma(dx_1, dy_1) \left(c(x_1, y_1) + V_1^c(y_1, \gamma_{y_1}(dx_1)) \right),$$

we get that

$$V_0^c = \sup_{\pi \in \Pi_{cqm}(\mu,\nu)} \int c d\pi. \tag{7.2}$$

We will refer to the right hand side in (7.2) as value$(Pcqm)$.

Proof. First we show that the sets

$$D_{t-1} = \{(y_1, ..., y_{t-1}, m, \gamma) : \gamma \in \Pi(\int_{x_{t-1}} m(dx_{t-1})\mu_{x_{t-1}}(dx_t), \nu_{y_1,...,y_{t-1}}(dy_t))\}$$

are Borel and therefore also analytic. Let $\tilde{D} = \{(p, q, \gamma) : p, q \in \mathcal{P}(\mathbb{R}), \gamma \in \Pi(q, p)\}$ which is closed by a similar argument as in the proof of Lemma 4.1. Let Φ_{t-1} be defined by

$$(y_1, ..., y_{t-1}, m, \gamma) \mapsto (\nu_{y_1,...,y_{t-1}}(dy_t), \int_{x_{t-1}} m(dx_{t-1})\mu_{x_{t-1}}(dx_t), \gamma).$$

By the way we defined Φ_{t-1} we have that $D_{t-1} = \Phi_{t-1}^{-1}(\tilde{D})$, so it suffices to show that $m \mapsto \int_{x_{t-1}} m(dx_{t-1})\mu_{x_{t-1}}(dx_t)$ is measurable in order to obtain measurability of D_{t-1}, which is true by Lemma 7.1.

Now we will show that all the integrals in the recursive formula are well defined by showing that the integrands are upper semianalytic and therefore universally measurable.

By Corollary 7.27.2 in [BS96] we can choose the kernel $\gamma_{y_t}(dx_t)$ such that $y_t \mapsto \gamma_{y_t}(dx_t)$ is measurable. Therefore $(y_1, ..., y_t, \gamma) \mapsto (y_1, ..., y_t, \gamma_{y_t}(dx_t))$ is measurable as well. If we show that $(y_1, ..., y_t, m) \mapsto V_t^c(y_1, ..., y_t, m)$ is upper semianalytic, we get that $(y_1, ..., y_t, \gamma) \mapsto V_t^c(y_1, ..., y_t, \gamma_{y_t})$ is upper semianalytic because the composition of an upper semianalytic function with a measurable function is upper semianalytic by Lemma 7.30 (3) in [BS96].

We will show that V_t^c is upper semianalytic by reverse induction. First we show that

$$(y_1, ..., y_{N-1}, m, \gamma) \mapsto \int c_N \gamma(dx_N, dy_N) \tag{7.3}$$

is Borel measurable, so in particular upper semianalytic. To see this we can use Proposition 7.29 in [BS96] and its notation, defining $q_\gamma(dx_N, dy_N)$ as $\gamma(dx_N, dy_N)$, which is clearly measurable as the function $q : \mathcal{P}(\mathbb{R}^2) \to \mathcal{P}(\mathbb{R}^2)$ from Definition 1 is the identity. We can use Proposition 7.47 in [BS96] to get that V_{N-1}^c is upper semianalytic, because we already showed that D_{N-1} is an analytic set.

Suppose now that V_t^c is upper semianalytic. Let us look at

$$\begin{aligned}(y_1, ..., y_{t-1}, m, \gamma) \mapsto &\int \gamma(dx_t, dy_t) c_t(x_t, y_1, ..., y_t) \\ &+ \int \nu_{y_1,...,y_{t-1}}(dy_t) V_t^c(y_1, ..., y_t, \gamma_{y_t}).\end{aligned} \tag{7.4}$$

The integrand of the second integral is upper semianalytic by the induction hypothesis and as it is the composition of an upper semianalytic function with a Borel measurable function. Its integral is also upper semianalytic by Proposition 7.48 in [BS96]. The first summand on the right hand side of (7.4) is measurable by the same argument as for (7.3). By Lemma 7.30 (4) in [BS96] their sum is upper semianalytic as well. Again by Proposition 7.47 in [BS96], which we can apply as we showed that D_{t-1} is analytic, we conclude that V_{t-1}^c is upper semianalytic.

We wrote $V_{t-1}^c(y_1, ..., y_{t-1}, m)$ as a supremum of an upper semianalytic function over a fiber of the analytic set D_{t-1}. Hence we can use Proposition 7.50 in [BS96] to get an universally measurable function defined by

$$(y_1, ..., y_{t-1}, m) \mapsto L_{t-1,\varepsilon}^{y_1,...,y_{t-1},m} \in \Pi\Big(\int_{x_{t-1}} m(dx_{t-1}) \mu_{x_{t-1}}(dx_t), \nu_{y_1,...,y_{t-1}}(dy_t)\Big),$$

such that

$$\begin{aligned}V_{t-1}^c(y_1, ..., y_{t-1}, m) - \varepsilon \leqslant \int L_{t-1,\varepsilon}^{y_1,...,y_{t-1},m}&(dx_t, dy_t)[c(x_t, y_1, ..., y_t) \\ &+ V_t^c(y_1, ..., y_t, (L_{t-1,\varepsilon}^{y_1,...,y_{t-1},m})_{y_t})]\end{aligned} \tag{7.5}$$

for all $(y_1, ...y_{t-1}, , m)$ for which $V_{t-1}^c(y_1, ...y_{t-1}, , m) < \infty$ and for all others we can make the RHS in (7.5) greater than $1/\varepsilon$. We will now build a measure that solves the recursion up to an ε margin each step, or gives us arbitrary large values if necessary.

Suppose that c is bounded from above and therefore also that every V_t^c is finite, in particular V_0^c. We choose an ε optimizer $\gamma^{0,\varepsilon}(dx_1, dy_1)$. Then we take $y_1 \mapsto \gamma_{y_1}^{1,\varepsilon}(dx_2, dy_2) := L_{1,\varepsilon}^{y_1,(\gamma^{0,\varepsilon})_{y_1}}$, which is universally measurable as it is the composition of two universally measurable functions (see Proposition 7.44 in [BS96]). Suppose that $(y_1, ..., y_{t-1}) \mapsto \gamma^{t-1,\varepsilon}_{y_1,...,y_{t-1}}(dx_t, dy_t)$ is universally measurable. We define

$$(y_1, ..., y_t) \mapsto \gamma^{t,\varepsilon}_{y_1,...,y_t}(dx_{t+1}, dy_{t+1}) := L^{y_1,...,y_t,(\gamma^{t-1,\varepsilon}_{y_1,...,y_{t-1}})_{y_t}}_{t,\varepsilon}(dx_{t+1}, dy_{t+1}),$$

which is universally measurable, because we can write it as the composition of the universally measurable functions

$$\begin{aligned}(y_1, ..., y_t) &\mapsto (y_1, ..., y_t, \gamma^{t-1,\epsilon}_{y_1,...,y_{t-1}}),\\ (y_1, ..., y_t, m(dx_t, dy_t)) &\mapsto (y_1, ..., y_t, m_{y_t}) \text{ and}\\ (y_1, ..., y_t, m(dx_t)) &\mapsto L^{y_1,...,y_t,m}_{t,\varepsilon}(dx_{t+1}, dy_{t+1}).\end{aligned}$$

By definition we have that

$$\gamma^{t,\varepsilon}_{y_1,...,y_t}(dx_{t+1}, dy_{t+1}) \in \Pi\left(\int_{x_t} (\gamma^{t-1,\varepsilon}_{y_1,...,y_{t-1}})_{y_t}(dx_t)\mu_{x_t}(dx_{t+1}), \nu_{y_1,...,y_t}(dy_{t+1})\right)$$

and the integral w.r.t. $\gamma^{t,\varepsilon}_{y_1,...,y_t}$ attains $V^c_t(y_1, ..., y_t, (\gamma^{t-1,\varepsilon}_{y_1,...,y_{t-1}})_{y_t})$ up to an ε margin. Let

$$(x_t, y_1, ..., y_t) \mapsto \Gamma^{x_t,y_1,...,y_t}_{t,\varepsilon}(dx_{t+1}, dy_{t+1}) := \mu_{x_t}(dx_{t+1})(\gamma^{t,\varepsilon}_{y_1,...,y_t})_{x_{t+1}}(dy_{t+1}), \tag{7.6}$$

which is again measurable by Proposition 7.44 and Proposition 7.45 in [BS96]. By Proposition 7.45 in [BS96] successive composition of these kernels gives us a unique Borel measure Γ_ε such that

$$\begin{aligned}\Gamma_\varepsilon(dx_1, ..., dx_N, dy_1, ...,dy_N) = \gamma_{0,\varepsilon}(dx_1, dy_1)\Gamma^{x_1,y_1}_{1,\varepsilon}(dx_2, dy_2)...\\ \Gamma^{x_t,y_1,...,y_t}_{t,\varepsilon}(dx_{t+1}, dy_{t+1})...\Gamma^{x_{N-1},y_1,...,y_{N-1}}_{N-1,\varepsilon}(dx_N, dy_N).\end{aligned}$$

By construction Γ_ε is causal quasi-Markov and we will show that $\Gamma_\varepsilon \in \Pi(\mu, \nu)$ as well. The first projection is equal to μ as by the definition of $\Gamma^{x_t,y_1,...,y_t}_{t,\epsilon}$ in (7.6), after integrating the y_i out, we are left with

$$\Gamma_\varepsilon(dx_1, ..., dx_N) = \gamma_{0,\varepsilon}(dx_1)\mu_{x_1}(dx_2)...\mu_{x_{N-1}}(dx_N)$$

and by definition $\gamma^{0,\varepsilon} \in \Pi(\mu^1, \nu^1)$.

Now we will also show that $\Gamma_\epsilon(dy_1, ..., dy_N) = \nu(dy_1, ..., dy_N)$. The definition of $\gamma_{0,\varepsilon}$ gives us that $\Gamma_\varepsilon(dy_1) = \nu(y_1)$ and therefore also

$$\begin{aligned}\Gamma_\varepsilon(dy_1, dy_2) &= \int_{x_1,x_2} \gamma^{0,\varepsilon}(dx_1, dy_1)\mu_{x_1}(dx_2)(\gamma^{1,\varepsilon}_{y_1})_{x_2}(dy_2)\\ &= \int_{x_1,x_2} \gamma^{0,\varepsilon}_{y_1}(dx_1)\nu(dy_1)\mu_{x_1}(dx_2)(\gamma^{1,\varepsilon}_{y_1})_{x_2}(dy_2)\\ &= \nu(dy_1)\int_{x_2}\left(\int_{x_1} \gamma^{0,\varepsilon}_{y_1}(dx_1)\mu_{x_1}(dx_2)\right)(\gamma^{1,\varepsilon}_{y_1})_{x_2}(dy_2) \qquad (7.7)\\ &= \nu(dy_1)\nu_{y_1}(dy_2).\end{aligned}$$

The last equality is true, as the expression inside the brackets in the penultimate line is the first marginal of $\gamma^{1,\varepsilon}_{y_1}$. Inductively we will show that we can write $\Gamma_\varepsilon(dy_1, ..., dy_N)$ in the following way:

$$\begin{aligned}\Gamma_\varepsilon(dy_1, ..., dy_N) &= \nu(dy_1)\nu_{y_1}(dy_2)...\nu_{y_1,...,y_{N-2}}(dy_{N-1})\\ &\int_{x_N}\left(\int_{x_{N-1}}(\gamma^{N-2,\varepsilon}_{y_1,...,y_{N-2}})_{y_{N-1}}(dx_{N-1})\mu_{x_{N-1}}(dx_N)\right)(\gamma^{N-1,\varepsilon}_{y_1,...,y_{N-1}})_{x_N}(dy_N)\end{aligned} \tag{7.8}$$

Then we can conclude our argument as (7.8) is equal to

$$\begin{aligned}&\nu(dy_1)...\nu_{y_1,...,y_{N-2}}(dy_{N-1})\int_{x_N}\gamma^{N-1,\varepsilon}_{y_1,...,y_{N-1}}(dx_N, dy_N)\\ &= \nu(dy_1)...\nu_{y_1,...,y_{N-2}}(dy_{N-1})\int_{x_N}(\gamma^{N-1,\varepsilon}_{y_1,...,y_{N-1}})_{y_N}(dx_N)\nu_{y_1,...,y_{N-1}}(dy_N)\\ &= \nu(dy_1)...\nu_{y_1,...,y_{N-2}}(dy_{N-1})\nu_{y_1,...,y_{N-2},y_{N-1}}(dy_N).\end{aligned}$$

By (7.7) we can write it in the way like in 7.8 for $N = 2$. The induction step works as well as

$$\begin{aligned}\Gamma_\varepsilon(dy_1, ..., dy_N, dy_{N+1}) &= \nu(dy_1)...\nu_{y_1,...,y_{N-2}}(dy_{N-1})\\ &\int_{x_N,x_{N+1}}\left(\int_{x_{N-1}}(\gamma^{N-2,\varepsilon}_{y_1,...,y_{N-2}})_{y_{N-1}}(dx_{N-1})\mu_{x_{N-1}}(dx_N)\right)(\gamma^{N-1,\varepsilon}_{y_1,...,y_{N-1}})_{x_N}(dy_N)\\ &\quad \mu_{x_N}(dx_{N+1})(\gamma^{N,\varepsilon}_{y_1,...,y_N})_{x_{N+1}}(dy_{N+1})\\ &= \nu(dy_1)...\nu_{y_1,...,y_{N-2}}(dy_{N-1})\\ &\quad\int_{x_N,x_{N+1}}\gamma^{N-1,\varepsilon}_{y_1,...,y_{N-1}}(dx_N, dy_N)\mu_{x_N}(dx_{N+1})(\gamma^{N,\varepsilon}_{y_1,...,y_N})_{x_{N+1}}(dy_{N+1})\\ &= \nu(dy_1)...\nu_{y_1,...,y_{N-2}}(dy_{N-1})\int_{x_N,x_{N+1}}(\gamma^{N-1,\varepsilon}_{y_1,...,y_{N-1}})_{y_N}(dx_N)\\ &\qquad \nu_{y_1,...,y_{N-1}}(dy_N)\mu_{x_N}(dx_{N+1})(\gamma^{N,\varepsilon}_{y_1,...,y_N})_{x_{N+1}}(dy_{N+1})\\ &= \nu(dy_1)...\nu_{y_1,...,y_{N-2}}(dy_{N-1})\nu_{y_1,...,y_{N-1}}(dy_N)\\ &\quad\int_{x_{N+1}}\left(\int_{x_N}(\gamma^{N-1,\varepsilon}_{y_1,...,y_{N-1}})_{y_N}(dx_N)\mu_{x_N}(dx_{N+1})\right)(\gamma^{N,\varepsilon}_{y_1,...,y_N})_{x_{N+1}}(dy_{N+1}).\end{aligned}$$

By the arguments above we know that Γ_ε is a causal quasi-Markov coupling and we designed it to be ε optimal at every step. Hence

$$V^c_0 - N\varepsilon \leqslant \int c d\Gamma_\varepsilon \leqslant \sup_{\pi\in\Pi_{cqm}(\mu,\nu)}\int c d\pi.$$

As ε was chosen arbitrarily we have that $V^c_0 \leqslant value(Pcqm)$. Now we show the reverse inequality. Let $\gamma \in \Pi_{cqm}(\mu, \nu)$. Notice that

$$\gamma_{y_1,...,y_t}(dx_{t+1}) = \int_{x_t}\gamma_{y_1,...,y_t}(dx_t)\gamma_{y_1,...,y_t,x_t}(dx_{t+1}) = \int_{x_t}\gamma_{y_1,...,y_t}(dx_t)\mu_{x_t}(dx_{t+1}),$$

where the last inequality holds true due the γ being causal quasi-Markov and Proposition 2.3 in [BBLZ17]. Also by Proposition 2.3 in [BBLZ17] we have that $\gamma_{y_1,...,y_t}(dy_{t+1}) = \nu_{y_1,...,y_t}(dy_{t+1})$ and therefore

$$\gamma_{y_1,...,y_t}(dx_{t+1}, dy_{t+1}) \in \Pi\left(\int \gamma_{y_1,...,y_t}(dx_t)\mu_{x_t}(dx_{t+1}), \nu_{y_1,...,y_t}(y_{t+1})\right). \tag{7.9}$$

Furthermore for $1 \leqslant t \leqslant N$

$$\begin{aligned}\int c_t d\gamma &= \int \gamma(dy_1, ..., dy_{t-1})\gamma_{y_1,...,y_{t-1}}(dx_t, dy_t)c_t(x_t, y_1, ..., y_t)\\ &= \int \gamma(dy_1)\gamma_{y_1}(dy_2)\gamma_{y_1,y_2}(dy_3)...\gamma_{y_1,...,y_{t-1}}(dx_t, dy_t)c_t(x_t, y_1, ..., y_t)\\ &= \int \gamma(dx_1, dy_1)\gamma_{y_1}(dx_2, dy_2)...\gamma_{y_1,...,y_{t-1}}(dx_t, dy_t)c_t(x_t, y_1, ..., y_t).\end{aligned}$$

As we assumed c to be of the form $c = \sum_{t=1}^{N} c_t(x_t, y_1, ..., y_t)$ we have that

$$\int c d\gamma = \int \gamma(dx_1, dy_1)\left[c_1 + \int \gamma_{y_1}(dx_2, dy_2)\left[c_2 + \int \gamma_{y_1,y_2}(dx_3, dy_3)\left[c_3 + \int ...\right]\right]\right].$$

Combining this with (7.9) we get that $value(Pcqm) \leqslant V_0^c$.

Let us treat the case that c is possibly not bounded from above. For $M \in \mathbb{N}$ let $c_M := \sum_{t=1}^{N} c_t^M$, where $c_t^M := c_t \wedge M$. By the considerations up to this point, we have that

$$V_0^{c_M} = \sup_{\pi\in\Pi_{cqm}(\mu,\nu)} \int c_M d\pi. \tag{7.10}$$

Using monotone convergence and equation (7.10) we have that

$$\begin{aligned}\sup_{\pi\in\Pi_{cqm}(\mu,\nu)} \int c d\pi = \sup_{\pi\in\Pi_{cqm}(\mu,\nu)} \sup_M \int c_M d\pi &= \sup_M \sup_{\pi\in\Pi_{cqm}(\mu,\nu)} \int c_M d\pi\\ &= \sup_M V_0^{c_M}.\end{aligned}$$

Inductively applying monotone convergence we get that this is equal to

$$\begin{aligned}&\sup_M \sup_\gamma \int d\gamma\left[c_1^M + \sup_{\gamma_1}\int d\gamma_1\left[c_2^M + \sup_{\gamma_2}\int d\gamma_2\left[c_3^M + ...\right]\right]\right]\\ &= \sup_\gamma \int d\gamma\left[c_1 + \sup_M \sup_{\gamma_1}\int d\gamma_1\left[c_2^M + \sup_{\gamma_2}\int d\gamma_2\left[c_3^M + ...\right]\right]\right] = ...\\ &= V_0^c.\end{aligned}$$

□

Theorem 7.3. *Let μ be Markov and c be semiseparable. Then $value(Pc) = value(Pcqm)$.*

Proof. Let $\gamma \in \Pi_c(\mu,\nu)$ and

$$\tilde{\gamma} = \gamma(dx_1, dy_1)\gamma_{x_1,y_1}(dx_2, dy_2)\gamma_{x_2,y_1,y_2}(dx_3, dy_3)...\gamma_{x_{N-1},y_1,...,y_{N-1}}(dx_N, dy_N).$$

We will show that $\tilde{\gamma}$ is causal quasi-Markov and that the integrals of a semiseparable cost function with respect to γ and $\tilde{\gamma}$ coincide.

By definition $\tilde{\gamma}(dx_1, dy_1) \in \Pi(\mu^1, \nu^1)$. We also have that

$$\begin{aligned}\tilde{\gamma}_{x_1,...,x_t,y_1,...,y_t}(dx_{t+1}) &= \gamma_{x_t,y_1,...,y_t}(dx_{t+1}) \\ &= \int_{x_1,...,x_{t-1}} \gamma_{x_t,y_1,...,y_t}(dx_1,...,dx_{t-1})\gamma_{x_1,...,x_t,y_1,...,y_t}(dx_{t+1}) \\ &= \int_{x_1,...,x_{t-1}} \gamma_{x_t,y_1,...,y_t}(dx_1,...,dx_{t-1})\mu_{x_t}(dx_{t+1}) = \mu_{x_t}(dx_{t+1}). \qquad (7.11)\end{aligned}$$

The last line in (7.11) holds true due to Proposition 2.3 in [BBLZ17] and μ being Markov. Therefore $p^1(\tilde{\gamma}) = \mu$.

Let $H := H(x_t, y_1, ..., y_t)$ be nonnegative and measurable. If we show that $\int H d\gamma = \int H d\tilde{\gamma}$ for every such H we get that $p^2(\tilde{\gamma}) = \nu$ and for a cost function with semiseparable structure the integrals w.r.t. γ and $\tilde{\gamma}$ coincide. Notice that $p^2(\tilde{\gamma}) = \nu$ gives us in particular that $\gamma_{y_1,...,y_t}(dy_{t+1}) = \nu_{y_1,...,y_t}(dy_{t+1})$. Combining this with (7.11) and the fact that $\tilde{\gamma}_{x_1,...,x_t,y_1,...,y_t} = \gamma_{x_t,y_1,...,y_t} = \tilde{\gamma}_{x_t,y_1,...,y_t}$ gives us that $\tilde{\gamma} \in \Pi_{cqm}(\mu,\nu)$ by Proposition 2.3 in [BBLZ17].

We have that $\int H d\gamma = \int H d\tilde{\gamma}$ holds true as

$$\begin{aligned}\int H d\tilde{\gamma} &= \int H\gamma(dx_1, dx_2, dy_1, dy_2)\gamma_{x_2,y_1,y_2}(dx_3, dy_3)... \\ &= \int H\gamma(dx_2, dy_1, dy_2)\gamma_{x_2,y_1,y_2}(dx_3, dy_3)... \\ &= \int H\gamma(dx_2, dx_3, dy_1, dy_2, dy_3)\gamma_{x_3,y_1,y_2,y_3}(dx_4, dy_4)... \\ ... &= \int H\gamma(dx_{t-1}, dy_1, ..., dy_{t-1})\gamma_{x_{t-1},y_1,...,y_{t-1}}(dx_t, dy_t) = \int H d\gamma.\end{aligned}$$

□

Theorem 7.4. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$ and c be semiseparable. If*

$$\sup_{\gamma \in \Pi_c(\mu,\nu)} \int c d\pi = 0,$$

then we have, for all $(x_1, x_2, y_1, y_2) \in \mathbb{R}^2 \times \mathbb{R}^2$, that

$$c(x_1, x_2, y_1, y_2) \leqslant g(x_1) + h(y_1) + g^{x_1,y_1}(x_2) + h^{x_1,y_1}(y_2),$$

where g (resp. h) is equal to zero μ_1-a.s. (resp. ν_1-a.s.), $h^{x_1,y_1} = 0$ ν_{y_1}-a.s. and $g^{x_1,y_1} = 0$ a.s. w.r.t. the measure $\int_{x_1} \gamma_{y_1}(dx_1)\mu_{x_1}(dx_2)$ for every $\gamma \in \Pi(\mu_1, \nu_1)$. In particular $g^{x_1,y_1} = 0$ μ_{x_1}-a.s.

Proof. Combining Theorem 7.2 with Theorem 7.3, we can recursively apply Theorem 2.21 from [Kel84], as we already did in the proof of Theorem 4.3, to obtain this result.

To see that the last line of the Theorem is indeed true, assume that there exists a set $A \subset \mathbb{R}$ with $\mu^1(A) > 0$ such that $\mu_{x_1}(g^{x_1,y_1}) > 0$ for $x_1 \in A$. Let $\gamma = \mu^1 \otimes \nu^1$. Then

$$\left(\int_{x_1} \gamma_{y_1}(dx_1)\mu_{x_1}(dx_2)\right)(g^{x_1,y_1}) = \int_{x_1} \mu(dx_1)\mu_{x_1}(g^{x_1,y_1}) > 0,$$

which contradicts that $g^{x_1,y_1} = 0$ a.s. with respect to the measure $\int_{x_1} \gamma_{y_1}(dx_1)\mu_{x_1}(dx_2)$. □

8 Weak transport

Let X be a Polish space with a compatible metric d. We denote by $\Psi(X)$ the set of all continuous functions $\phi : X \to \mathbb{R}$, which are bounded from below and satisfy that

$$|\phi(x)| \leqslant a + bd(x, x_0),$$

for all $x \in X$, for some $a > 0$, $b > 0$ and some $x_0 \in X$.

In this chapter we will consider cost functions, which may also depend on the chosen coupling. Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ for Polish spaces X, Y. Consider a l.s.c. cost function

$$C : X \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$$

which is convex in the second component and bounded from below. Then we call

$$\inf_{\pi \in \Pi(\mu,\nu)} \int C(x, \pi_x)\mu(dx)$$

the weak transport problem between μ and ν.

We will frequently encounter the Polish space $\mathcal{P}_1(\mathbb{R} \times \mathcal{P}_1(\mathbb{R}))$ in this chapter. A complete metric on this space can be given by the 1-Wasserstein metric w.r.t. the metric $d((x, \hat{x}), (y, \hat{y})) = |x - y| + W(\hat{x}, \hat{y})$. For more details see [BBEP20].

More concretely, in Theorem 8.8, we will establish duality for the cost function

$$C : \mathbb{R} \times \mathcal{P}_1(\mathbb{R}) \times \mathcal{P}_1(\mathbb{R} \times \mathcal{P}_1(\mathbb{R})) \to \mathbb{R},$$
$$C(x, \hat{x}, p) = \int |x - y| p(dy) + W\left(\hat{x}, \int \hat{y} p(d\hat{y})\right),$$

which is the cost function we consider for the remainder of the chapter. This duality in Theorem 8.8 will prove to be useful in the next chapter in order to make a connection to causal transport.

Definition 7. Let $(X, X^*, \langle\cdot,\cdot\rangle)$ be a dual pair and let $f : X \to \overline{\mathbb{R}}$ be a proper convex function (i.e. f is convex, $f > -\infty$ and there exists an $x \in X$ s.th. $f(x) < +\infty$). Then we call $f^* : X^* \to \overline{\mathbb{R}}$, defined by

$$f^*(x^*) = \sup_{x \in X} \{\langle x^*, x\rangle - f(x)\},$$

the conjugate function of f.

Lemma 8.1. *Let $\phi : \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ be convex and l.s.c. Then*

$$\int \phi(\mu)\alpha(d\mu) \geqslant \phi\left(\int \mu\alpha(d\mu)\right)$$

for $\alpha \in \mathcal{P}_1(\mathcal{P}_1(\mathbb{R}))$.

Proof. Notice that $\mathcal{P}_1(\mathbb{R}) \subset \mathcal{M}_1(\mathbb{R})$ is closed and convex. Here $\mathcal{M}_1(\mathbb{R})$ denotes the the space of signed measures with finite first moment endowed with the initial topology with respect to the family of functions of the form $\mu \mapsto \int f d\mu$, for $f \in \mathcal{C}_{lin}(\mathbb{R})$. So we are looking at the dual pair $(\mathcal{M}_1(\mathbb{R}), \mathcal{C}_{lin}(\mathbb{R}), \langle \cdot, \cdot \rangle)$, where $\langle m, f \rangle := \int f dm$. Restricted to $\mathcal{P}_1(\mathbb{R})$ this topology coincides with the topology on $\mathcal{P}_1(\mathbb{R})$ arising from the Wasserstein-distance (see Definition 6.7 in [Vil16]). Let $\tilde{\phi} : \mathcal{M}_1(\mathbb{R}) \to \overline{\mathbb{R}}$ be defined by

$$\begin{cases} \tilde{\phi}(\hat{x}) = \phi(\hat{x}) & \hat{x} \in \mathcal{P}_1(\mathbb{R}), \\ \tilde{\phi}(\hat{x}) = +\infty & \hat{x} \notin \mathcal{P}_1(\mathbb{R}). \end{cases}$$

By definition $\tilde{\phi}$ is clearly convex. Notice that $\{\hat{x} \in \mathcal{M}_1(\mathbb{R}) : \tilde{\phi}(\hat{x}) \leqslant c\} = \{\hat{x} \in \mathcal{P}_1(\mathbb{R}) : \phi(\hat{x}) \leqslant c\}$ is closed for $c \in \mathbb{R}$ by the lower semicontinuity of ϕ and therefore $\tilde{\phi}$ is l.s.c. as well. So we can use Theorem 2.3.3 in [Zal02] in order to get $\tilde{\phi} = ((\tilde{\phi})^*)^*$ and therefore

$$\tilde{\phi}(\mu) = \sup_{f \in \mathcal{C}_{lin}(\mathbb{R})} \left(\int f d\mu - (\tilde{\phi})^*(f) \right). \tag{8.1}$$

Using (8.1) we get, for $\alpha \in \mathcal{P}_1(\mathcal{P}_1(\mathbb{R}))$, that

$$\begin{aligned} \int \phi(\mu)\alpha(d\mu) &= \int \tilde{\phi}(\mu)\alpha(d\mu) \\ &= \int \sup_{f \in \mathcal{C}_{lin}(\mathbb{R})} \left(\int f d\mu - (\tilde{\phi})^*(f) \right) \alpha(d\mu) \\ &\geqslant \sup_{f \in \mathcal{C}_{lin}(\mathbb{R})} \left(\int \left(\int f d\mu - (\tilde{\phi})^*(f) \right) \alpha(d\mu) \right) \\ &= \sup_{f \in \mathcal{C}_{lin}(\mathbb{R})} \left(\int f(x)(\textstyle\int \mu\alpha(d\mu))(dx) - (\tilde{\phi})^*(f) \right) \\ &= \tilde{\phi} \left(\int \mu\alpha(d\mu) \right) \\ &= \phi \left(\int \mu\alpha(d\mu) \right). \end{aligned}$$

□

We will need the following two results characterizing convex functions, in which we denote the pointwise supremum of all convex functions, which are dominated by ϕ, by $\bar{\phi}$. It is easily seen that $\bar{\phi}$ is convex.

Lemma 8.2. *Let $\phi : \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ be convex and 1-Lipschitz. Then*

$$\phi(\hat{x}) = \inf_{\alpha \in \mathcal{P}_1(\mathcal{P}_1(\mathbb{R})),\ mean(\alpha) = \hat{x}} \int \phi d\alpha.$$

Proof. One inequality can directly be seen by choosing $\alpha = \delta_{\hat{x}}$.

For the reverse inequality let $\alpha \in \mathcal{P}_1(\mathcal{P}_1(\mathbb{R}))$ with $mean(\alpha) = \hat{x}$, we can use the version of Jensen's inequality from Lemma 8.1 to get

$$\phi(\hat{x}) = \phi \left(\int \hat{z}\alpha(d\hat{z}) \right) \leqslant \int \phi(\hat{z})\alpha(d\hat{z}).$$

Passing over to the infimum we get the result. □

Corollary 8.3. *For $\phi : \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ 1-Lipschitz we have that*

$$\bar{\phi}(\hat{x}) = \inf_{\alpha \in \mathcal{P}_1(\mathcal{P}_1(\mathbb{R})),\ mean(\alpha)=\hat{x}} \int \phi(\hat{z})\alpha(d\hat{z}). \tag{8.2}$$

Proof. It is easily seen that the function on the right hand side is convex and that it is dominated by ϕ as we can again choose $p = \delta_{\hat{x}}$. Hence the right hand side in (8.2) is dominated by the left hand side. For the reverse inequality let ψ be a convex function with $\psi \leqslant \phi$ and α with $mean(\alpha) = \hat{x}$. Then, due to Lemma 8.1,

$$\psi(\hat{x}) = \psi\left(\int \hat{z}\alpha(d\hat{z})\right) \leqslant \int \psi(\hat{z})\alpha(d\hat{z}) \leqslant \int \phi(\hat{z})\alpha(d\hat{z}).$$

Again we can pass over to the infimum on the right side. □

Lemma 8.4. *The function*

$$C : \mathbb{R} \times \mathcal{P}_1(\mathbb{R}) \times \mathcal{P}_1(\mathbb{R} \times \mathcal{P}_1(\mathbb{R})) \to \mathbb{R},$$
$$C(x, \hat{x}, p) = \int |x - y| p(dy) + W\left(\hat{x}, \int \hat{y} p(d\hat{y})\right)$$

is 1*-Lipschitz in* p.

Proof. Let $p, q \in \mathcal{P}_1(\mathbb{R} \times \mathcal{P}_1(\mathbb{R}))$. We assume w.l.o.g. that $\int |x-y|p(dy) > \int |x-y|q(dy)$.
For $\phi \in Lip_1(\mathbb{R})$ the function $g_\phi : \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}, \hat{x} \mapsto \hat{x}(\phi)$ is also 1-Lipschitz, because

$$\begin{aligned}\left|\int \phi(x)\hat{x}(dx) - \int \phi(y)\hat{y}(dy)\right| &\leqslant \sup_{f \in Lip_1(\mathbb{R})} \left(\int f(x)\hat{x}(dx) - \int f(y)\hat{y}(dy)\right) \\ &= W(\hat{x}, \hat{y}).\end{aligned}$$

Using (2.1) we have that

$$W\left(\int \hat{x} p(d\hat{x}), \int \hat{y} q(d\hat{y})\right) = \sup_{\phi \in Lip_1(\mathbb{R})} \left(\int g_\phi(\hat{x}) p(d\hat{x}) - \int g_\phi(\hat{y}) q(d\hat{y})\right). \tag{8.3}$$

Applying (8.3) and the fact that g_ϕ is 1-Lipschitz for ϕ 1-Lipschitz we get

$$\begin{aligned}|C(x,\hat{x},p) - C(x,\hat{x},q)| &\leqslant \int |x-y|p(dy) - \int |x-y|q(dy) \\ &\quad + \left|W\left(\hat{x}, \int \hat{y}p(d\hat{y})\right) - W\left(\hat{x}, \int \hat{y}q(d\hat{y})\right)\right| \\ &\leqslant \int |x-y|p(dy) - \int |x-y|q(dy) + W\left(\int \hat{y}p(d\hat{y}), \int \hat{y}q(d\hat{y})\right) \\ &\leqslant \sup_{\phi \in Lip_1(\mathbb{R})} \left(\int |x-y| + g_\phi(\hat{y})p(dy,d\hat{y}) - \int |x-y| + g_\phi(\hat{y})q(dy,d\hat{y})\right) \\ &\leqslant \sup_{f \in Lip_1(\mathbb{R} \times \mathcal{P}_1(\mathbb{R}))} \left(\int f dp - \int f dq\right) \\ &= W(p,q).\end{aligned}$$

□

Lemma 8.5 and Theorem 8.8 and the idea of their proof resemble Theorem 2.11 in [GRST17].

Lemma 8.5. *For all* $\phi \in \Psi(\mathbb{R} \times \mathcal{P}_1(\mathbb{R})) \cap Lip_1$ *and for all* $(x, \hat{x}) \in \mathbb{R} \times \mathcal{P}_1(\mathbb{R})$ *we have that*

$$\hat{Q}\phi(x,\hat{x}) := \inf_{p \in \mathcal{P}_1(\mathbb{R}\times\mathcal{P}_1(\mathbb{R}))} \left\{ \int \phi(y,\hat{y})p(dy,d\hat{y}) + \int |x-y|p(dy) + W\Big(\hat{x}, \int \hat{y}p(d\hat{y})\Big) \right\} \\ = Q\hat{\phi}(x,\hat{x}),$$

where $\hat{\phi}$ *denotes the supremum of all functions which are dominated by* ϕ *and convex in the second component and* $Qf(x,\hat{x}) := \inf_{\hat{z}\in\mathcal{P}_1(\mathbb{R})}\{f(x,\hat{z}) + W(\hat{x},\hat{z})\}$.

Proof. First we will show that

$$\hat{Q}\phi(x,\hat{x}) = \inf_{\hat{z}\in\mathcal{P}_1(\mathbb{R})}\{g(x,\hat{z}) + W(\hat{x},\hat{z})\},$$

where

$$g(x,\hat{z}) := \inf_{p\in\mathcal{P}_1(\mathbb{R}\times\mathcal{P}_1(\mathbb{R}))} \left\{ \int \phi dp + \int |x-y|dp(y), \int \hat{y}p(d\hat{y}) = \hat{z} \right\}$$

Then we conclude by showing that $g = \hat{\phi}$.

"$\geqslant$": Let $p \in \mathcal{P}_1(\mathbb{R}\times\mathcal{P}_1(\mathbb{R}))$ be arbitrary. Let $\lambda_p = \int \hat{y}p(d\hat{y})$. Then

$$\int \phi dp + \int |x-y|p(dy) + W(\hat{x},\lambda_p) \geqslant g(x,\lambda_p) + W(\hat{x},\lambda_p) \\ \geqslant \inf_{\hat{z}\in\mathcal{P}_1(\mathbb{R})}\{g(x,\hat{z}) + W(\hat{x},\hat{z})\}.$$

Therefore we can also pass over to the infimum over all p.

"$\leqslant$": Let $\hat{z} \in \mathcal{P}_1(\mathbb{R})$ be arbitrary. Then we have

$$\begin{aligned} &\inf_p \left\{ \int \phi dp + \int |x-y|p(dy), \int \hat{y}dp = \hat{z} \right\} + W(\hat{x},\hat{z}) \\ &= \inf_p \left\{ \int \phi dp + \int |x-y|p(dy) + W\left(\hat{x}, \int \hat{y}p(d\hat{y})\right), \int \hat{y}dp = \hat{z} \right\} \\ &\geqslant \inf_p \left\{ \int \phi dp + \int |x-y|p(dy) + W\left(\hat{x}, \int \hat{y}p(d\hat{y})\right) \right\} \\ &= \hat{Q}\phi(x,\hat{x}). \end{aligned}$$

Therefore we can also pass over to the infimum over all $\hat{z}$.

Now we will show that g is indeed equal to $\hat{\phi}$. Let $\hat{z}_1, \hat{z}_2 \in \mathcal{P}_1(\mathbb{R})$, $\varepsilon > 0$ and $\lambda \in [0,1]$. To show convexity in the second argument of g choose an $\varepsilon/2$-optimal measure p_1 for $g(z,\hat{z}_1)$ and an $\varepsilon/2$-optimal measure p_2 for $g(z,\hat{z}_2)$. Then, choosing the measure $\lambda p_1 + (1-\lambda)p_2$, we have that

$$
\begin{aligned}
&g(z, \lambda\hat{z}_1 + (1-\lambda)\hat{z}_2) \\
&\quad \leqslant \lambda \left(\int \phi(x,\hat{x}) + |x-z| p_1(dx, d\hat{x}) \right) + (1-\lambda) \left(\int \phi(x,\hat{x}) + |x-z| p_2(dx, d\hat{x}) \right) \\
&\quad \leqslant \lambda g(z,\hat{z}_1) + (1-\lambda) g(z,\hat{z}_2) + \varepsilon.
\end{aligned}
$$

As ε was chosen arbitrarily this shows that g is convex in the second variable.

As we can choose $p = \delta_x \otimes \delta_{\hat{x}}$ we also have that $g \leqslant \phi$ and therefore $g \leqslant \hat{\phi}$. Now we show that $g \geqslant \hat{\phi}$ as well. By Corollary 8.3 and Lipschitz continuity of ϕ we have that

$$
\begin{aligned}
\hat{\phi}(x,\hat{x}) &= \inf_{p \in \mathcal{P}_1(\mathcal{P}_1(\mathbb{R}))} \left\{ \int \phi(x,\hat{z}) p(d\hat{z}), \int \hat{z} p(d\hat{z}) = \hat{x} \right\} \\
&= \inf_{p \in \mathcal{P}_1(\mathbb{R} \times \mathcal{P}_1(\mathbb{R}))} \left\{ \int \phi(x,\hat{z}) p(dx_0, d\hat{z}), \int \hat{z} p(d\hat{z}) = \hat{x} \right\} \\
&\leqslant \inf_{p \in \mathcal{P}_1(\mathbb{R} \times \mathcal{P}_1(\mathbb{R}))} \left\{ \int \phi(x_0,\hat{z}) + |x - x_0| p(dx_0, d\hat{z}), \int \hat{z} p(d\hat{z}) = \hat{x} \right\} = g(x,\hat{x}).
\end{aligned}
$$

□

Lemma 8.6. *For $\phi \in Lip_1(\mathbb{R} \times \mathcal{P}_1(\mathbb{R}))$ we have that*

$$
Q\phi(x,\hat{x}) := \inf_{\hat{z} \in \mathcal{P}_1(\mathbb{R})} \{\phi(x,\hat{z}) + W(\hat{x},\hat{z})\} = \phi(x,\hat{x}) \tag{8.4}
$$

for all $(x,\hat{x}) \in \mathbb{R} \times \mathcal{P}_1(\mathbb{R})$.

Proof. The LHS in (8.4) is dominated by the RHS as we can choose $\hat{z} = \hat{x}$. To see that the other inequality holds true as well assume that there exists some $\hat{z} \in \mathcal{P}_1(\mathbb{R})$ such that

$$
\phi(x,\hat{z}) + W(\hat{x},\hat{z}) < \phi(x,\hat{x}).
$$

This immediately leads to a contradiction as ϕ is 1-Lipschitz. □

Lemma 8.7. *The function $\hat{\phi}(x,\hat{x})$ is 1-Lipschitz, for $\phi : \mathbb{R} \times \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ 1-Lipschitz.*

Proof. For notational simplicity we assume that $\phi : \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ and we show that $\bar{\phi}$ is 1-Lipschitz.

Let $\hat{x}, \hat{y} \in \mathcal{P}_1(\mathbb{R})$ and $\pi \in \Pi(\hat{x},\hat{y})$ such that $W(\hat{x},\hat{y}) = \int |x-y| \pi(dx,dy)$, which is possible by Theorem 4.1 in [Vil16]. Let λ be the Lebesgue measure on $[0,1]$. We choose $T : \mathbb{R} \times [0,1] \to \mathbb{R}$ such that, for all $x \in \mathbb{R}$, $(y \mapsto T(x,y))_\#(\lambda) = \pi_x$. Then we have that

$$
W(\hat{x},\hat{y}) = \int |x-y| \pi(dx,dy) = \int \int |x - T(x,u)| \lambda(du) \hat{x}(dx).
$$

Let $p \in \mathcal{P}_1(\mathcal{P}_1(\mathbb{R}))$ with $\int \hat{z} p(d\hat{z}) = \hat{x}$ such that

$$
\int \phi(\hat{z}) p(d\hat{z}) - \varepsilon \leqslant \bar{\phi}(\hat{x}). \tag{8.5}
$$

For $u \in [0,1]$ and $T_u := (x \mapsto T(x,u))$ let $q_u := (\hat{z} \mapsto (T_u)_\# \hat{z})_\# p$. Let $q := \int q_u \lambda(du)$. First we verify that $\int \hat{z} q(d\hat{z}) = \hat{y}$. For A measurable we have that

$$
\begin{aligned}
\left(\int \hat{z}q(d\hat{z})\right)(A) &= \int \hat{z}(A)(\textstyle\int q_u\lambda(du))(d\hat{z}) \\
&= \int\int \hat{z}(A)q_u(d\hat{z})\lambda(du) \\
&= \int\int (T_u)_\#(\hat{z})(A)p(d\hat{z})\lambda(du) \\
&= \int\int \hat{z}(\{x : T(x,u) \in A\})p(d\hat{z})\lambda(du) \\
&= \int\int\int \mathbb{1}_{\{x:T(x,u)\in A\}}\hat{z}(dx)p(d\hat{z})\lambda(du) \\
&= \int\int \mathbb{1}_{\{x:T(x,u)\in A\}}\hat{x}(dx)\lambda(du) \\
&= \int\int \mathbb{1}_{\{u:T(x,u)\in A\}}\lambda(du)\hat{x}(dx) \\
&= \int \pi_x(A)\hat{x}(dx) \\
&= \pi(\mathbb{R}\times A) = \hat{y}(A)
\end{aligned}
$$

Moreover, using the Lipschitz continuity of ϕ and (8.5), we have that

$$
\begin{aligned}
\int \phi(\hat{z})q(d\hat{z}) &= \int\int \phi(\hat{z})q_u(d\hat{z})\lambda(du) \\
&= \int\int \phi((T_u)_\#\hat{z})p(d\hat{z})\lambda(du) \\
&\leqslant \int\int \phi(\hat{z}) + W(\hat{z},(T_u)_\#\hat{z})p(d\hat{z})\lambda(du) \\
&\leqslant \bar{\phi}(\hat{x}) + \int\int\int |x-T(x,u)|\hat{z}(dx)p(d\hat{z})\lambda(du) + \varepsilon \\
&= \bar{\phi}(\hat{x}) + \int |x-T(x,u)|\hat{x}(dx)\lambda(du) + \varepsilon \\
&\leqslant \bar{\phi}(\hat{x}) + W(\hat{x},\hat{y}) + \varepsilon.
\end{aligned}
$$

The fourth line holds true as we can choose the coupling π between $\hat{z}$ and $(T_u)_\#\hat{z}$ given by $\pi = f_\#\hat{z}$ for $f(x) = (x, T(x,u))$.

By the arbitrary choice of ε and by passing over to the infimum of all measures $m \in \mathcal{P}_1(\mathcal{P}_1(\mathbb{R}))$ with $mean(m) = \hat{y}$ on the left hand side, we get the desired result.

□

Lemma 8.4 enables us to use an already established duality result on weak transport (see Lemma 5.7 in [BBP19]). Putting this together with Lemma 8.5, Lemma 8.6 and Lemma 8.7 we will be able to prove the following duality Theorem:

Theorem 8.8. *For $\mu, \nu \in \mathcal{P}_1(\mathbb{R} \times \mathcal{P}_1(\mathbb{R}))$ and*

$$C : \mathbb{R} \times \mathcal{P}_1(\mathbb{R}) \times \mathcal{P}_1(\mathbb{R} \times \mathcal{P}_1(\mathbb{R})) \to \mathbb{R},$$
$$C(x, \hat{x}, p) = \int |x - y| p(dy) + W\left(\hat{x}, \int \hat{y} p(d\hat{y})\right),$$

we have that

$$\inf_{\pi \in \Pi(\mu,\nu)} \int C(x, \hat{x}, \pi_{x,\hat{x}}) \mu(dx, d\hat{x}) =$$
$$\sup \left\{ \int \phi d\mu - \int \phi d\nu, \phi \in \Psi(\mathbb{R} \times \mathcal{P}_1(\mathbb{R})) \cap Lip_1 \text{ and conv. in sec. var.} \right\}.$$

Proof. By Lemma 8.5 and the fact that $\hat{\phi} \leqslant \phi$ we have, for $\phi \in \Psi(\mathbb{R} \times \mathcal{P}_1(\mathbb{R}))$, that

$$\int \hat{Q}\phi d\mu - \int \phi d\nu \leqslant \int Q\hat{\phi} d\mu - \int \hat{\phi} d\nu. \tag{8.6}$$

We also notice that $\hat{\phi} \in \Psi(\mathbb{R} \times \mathcal{P}_1(\mathbb{R}))$ as it also satisfies the growth constraint and it is bounded from below if ϕ is. Moreover, by Lemma 8.7 it is also Lipschitz-continuous. Using all these considerations we get that

$$\begin{aligned}
&\inf_{\pi \in \Pi(\mu,\nu)} \int C(x, \hat{x}, \pi_{x,\hat{x}}) \mu(dx, d\hat{x}) = \sup \left\{ \int \hat{Q}\phi d\mu - \int \phi d\nu, \phi \in \Psi(\mathbb{R} \times \mathcal{P}_1(\mathbb{R})) \cap Lip_1 \right\} \\
&\leqslant \sup \left\{ \int Q\hat{\phi} d\mu - \int \hat{\phi} d\nu, \phi \in \Psi(\mathbb{R} \times \mathcal{P}_1(\mathbb{R})) \cap Lip_1 \right\} \\
&\leqslant \sup \left\{ \int Q\psi d\mu - \int \psi d\nu, \psi \in \Psi(\mathbb{R} \times \mathcal{P}_1(\mathbb{R})) \cap Lip_1, \text{ conv. in sec. var.} \right\} \\
&= \sup \left\{ \int \psi d\mu - \int \psi d\nu, \psi \in \Psi(\mathbb{R} \times \mathcal{P}_1(\mathbb{R})) \cap Lip_1, \text{ conv. in sec. var.} \right\} \\
&\leqslant \sup \left\{ \int \hat{Q}\psi d\mu - \int \psi d\nu, \psi \in \Psi(\mathbb{R} \times \mathcal{P}_1(\mathbb{R})) \cap Lip_1, \text{ conv. in sec. var.} \right\} \\
&\leqslant \sup \left\{ \int \hat{Q}\psi d\mu - \int \psi d\nu, \psi \in \Psi(\mathbb{R} \times \mathcal{P}_1(\mathbb{R})) \cap Lip_1, \right\} \\
&= \inf_{\pi \in \Pi(\mu,\nu)} \int C(x, \hat{x}, \pi_{x,\hat{x}}) \mu(dx, d\hat{x}).
\end{aligned}$$

The first and the last equality are due to Lemma 5.7 in [BBP19], the first inequality is due to (8.6), the second inequality due to Lemma 8.7, the second equality due to Lemma 8.6 and the third inequality is true because $\psi = \hat{\psi}$, so we can use Lemma 8.5. ☐

9 Weak transport meets causal transport

In the whole chapter $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$. Using Theorem 8.8 from the last chapter and Theorem 2.6 in [BBLZ17], we will derive duality for causal transport between the measures μ and ν w.r.t. the cost function $c = |x_1 - y_1| + |x_2 - y_2|$ in Theorem 9.2. First though, we need to rewrite the causal problem as a weak problem in the following way, in order to be able to apply the duality results from Chapter 8:

Lemma 9.1. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$. Let $\bar{\mu} := \mu^1 \otimes \delta_{\mu_{x_1}} \in \mathcal{P}(\mathbb{R} \times \mathcal{P}_1(\mathbb{R}))$ and $\bar{\nu} := \nu^1 \otimes \delta_{\nu_{y_1}} \in \mathcal{P}(\mathbb{R} \times \mathcal{P}_1(\mathbb{R}))$. Then we have that*

$$\inf_{\gamma \in \Pi_c(\mu,\nu)} \int |x_1 - y_1| + |x_2 - y_2| d\gamma =$$
$$\inf_{\bar{\gamma} \in \Pi(\bar{\mu},\bar{\nu})} \int \int |x_1 - y_1| \bar{\gamma}_{y_1,\hat{y}}(dx_1) + W\left(\int \hat{x} \bar{\gamma}_{y_1,\hat{y}}(d\hat{x}), \hat{y} \right) \bar{\gamma}(dy_1, d\hat{y}).$$

Proof. Notice that μ is Markov because it is a measure on $\mathbb{R}^2$. Hence we can apply Theorem 2.6 in [BBLZ17], as c has a separable structure, in order to get

$$\inf_{\gamma \in \Pi_c(\mu,\nu)} \int |x_1 - y_1| + |x_2 - y_2| d\gamma =$$
$$\inf_{\bar{\pi} \in \Pi(\mu^1,\nu^1)} \int \bar{\pi}(dx_1, dy_1) \Big(|x_1 - y_1|$$
$$+ \inf_{\pi \in \Pi(\int_{x_1} \bar{\pi}_{y_1}(dx_1) \mu_{x_1}(dx_2), \nu_{y_1}(dy_2))} \int \pi(dx_2, dy_2) |x_2 - y_2| \Big).$$

Also notice that for $\bar{\gamma} \in \Pi(\bar{\mu}, \bar{\nu})$ we have that

$$\bar{\gamma}_{y_1}(dx_1, d\hat{x}) = \int \bar{\gamma}_{y_1,\hat{y}}(dx_1, d\hat{x}) \bar{\gamma}_{y_1}(d\hat{y}) = \bar{\gamma}_{y_1,\nu_{y_1}}(dx_1, d\hat{x}). \tag{9.1}$$

Let $\gamma \in \Pi(\mu^1, \nu^1)$. Choose $\bar{\gamma} \in \Pi(\bar{\mu}, \bar{\nu})$ with $\gamma_{y_1}(dx_1) = \bar{\gamma}_{y_1}(dx_1)$, then

$$\int |x_1 - y_1| + W\left(\int \mu_{x_1}(dx_2) \gamma_{y_1}(dx_1), \nu_{y_1}(dy_2) \right) \gamma(dx_1, dy_1)$$
$$= \int |x_1 - y_1| + W\left(\int \hat{x} \bar{\gamma}_{y_1,\nu_{y_1}}(d\hat{x}), \nu_{y_1}(dy_2) \right) \bar{\gamma}(dx_1, dy_1)$$
$$= \int |x_1 - y_1| \bar{\gamma}_{y_1,\hat{y}}(dx_1) + W\left(\int \hat{x} \bar{\gamma}_{y_1,\hat{y}}(d\hat{x}), \hat{y} \right) \bar{\gamma}_{y_1}(d\hat{y}) \bar{\gamma}(dy_1).$$

The second line holds true, as

$$\int \mu_{x_1}(dx_2)\gamma_{y_1}(dx_1) = \int \mu_{x_1}(dx_2)\bar{\gamma}_{y_1}(dx_1) = \int\int \hat{x}\bar{\gamma}_{x_1,y_1}(d\hat{x})\bar{\gamma}_{y_1}(dx_1)$$
$$= \int \hat{x}\bar{\gamma}_{y_1}(d\hat{x}) = \int \hat{x}\bar{\gamma}_{y_1,\nu_{y_1}}(d\hat{x}),$$

where we used (9.1) in order to obtain the last equality. We could have also started with a coupling $\bar{\gamma} \in \Pi(\bar{\mu}, \bar{\nu})$ and chosen $\gamma \in \Pi(\mu^1, \nu^1)$ such that $\gamma_{y_1}(dx_1) = \bar{\gamma}_{y_1}(dx_1)$ and hence we get that

$$\inf_{\gamma \in \Pi_c(\mu,\nu)} \int |x_1 - y_1| + |x_2 - y_2| d\gamma$$
$$= \inf_{\gamma \in \Pi(\mu^1,\nu^1)} \int |x_1 - y_1| + W\left(\int \mu_{x_1}\gamma_{y_1}(dx_1), \nu_{y_1}\right)\gamma(dx_1, dy_1)$$
$$= \inf_{\bar{\gamma} \in \Pi(\bar{\mu},\bar{\nu})} \int\int |x_1 - y_1|\bar{\gamma}_{y_1,\hat{y}}(dx_1) + W\left(\int \hat{x}\bar{\gamma}_{y_1,\hat{y}}(d\hat{x}), \hat{y}\right)\bar{\gamma}(dy_1, d\hat{y}).$$

□

Theorem 9.2. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$. Then we have that*

$$\inf_{\gamma \in \Pi_c(\mu,\nu)} \int |x_1 - y_1| + |x_2 - y_2| d\gamma =$$
$$\sup\left\{ \int \phi(y_1, \nu_{y_1})\nu(dy_1) - \int \phi(x_1, \mu_{x_1})\mu(dx_1), \phi \in \Psi(\mathbb{R} \times \mathcal{P}_1(\mathbb{R})) \cap Lip_1 \right.$$
$$\left. \&\ \textit{conv. in sec. var.} \right\}.$$

Proof. By Lemma 9.1 we know that

$$\inf_{\gamma \in \Pi_c(\mu,\nu)} \int |x_1 - y_1| + |x_2 - y_2| d\gamma = \tag{9.2}$$
$$= \inf_{\bar{\gamma} \in \Pi(\bar{\mu},\bar{\nu})} \int\int |x_1 - y_1|\bar{\gamma}_{y_1,\hat{y}}(dx_1) + W\left(\int \hat{x}\bar{\gamma}_{y_1,\hat{y}}(d\hat{x}), \hat{y}\right)\bar{\gamma}(dy_1, d\hat{y}).$$

The RHS in (9.2) can be interpreted as a weak transport with the cost function $C : \mathbb{R} \times \mathcal{P}_1(\mathbb{R}) \times \mathcal{P}_1(\mathbb{R} \times \mathcal{P}_1(\mathbb{R})) \to \mathbb{R}$,

$$C(y_1, \hat{y}, p) = \int |x_1 - y_1| p(dx_1) + W\left(\int \hat{x}p(d\hat{x}), \hat{y}\right)$$

and therefore we can use Theorem 8.8 to get that

$$
\begin{aligned}
&\inf_{\gamma\in\Pi_c(\mu,\nu)}\int |x_1-y_1|+|x_2-y_2|d\gamma \\
&=\sup\left\{\int \phi d\bar{\nu}-\int \phi d\bar{\mu}, \phi\in\Psi(\mathbb{R}\times\mathcal{P}_1(\mathbb{R}))\cap Lip_1 \&\text{ conv. in sec. var.}\right\} \\
&=\sup\Bigg\{\int \phi(y_1,\nu_{y_1})\nu(dy_1)-\int \phi(x_1,\mu_{x_1})\mu(dx_1), \phi\in\Psi(\mathbb{R}\times\mathcal{P}_1(\mathbb{R}))\cap Lip_1 \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \&\text{ conv. in sec. var.}\Bigg\}
\end{aligned}
$$

□

Bibliography

[BBEP20] Julio Backhoff, Mathias Beiglböck, Manu Eder, and Alois Pichler. Fundamental properties of process distances. *Stochastic Processes and their Applications*, 2020.

[BBLZ17] Julio Backhoff, Mathias Beiglböck, Yiqing Lin, and Anastasiia Zalashko. Causal transport in discrete time and applications. *SIAM Journal on Optimization*, 2017.

[BBP19] Julio Backhoff, Mathias Beiglböck, and Gudmund Pammer. Existence, duality, and cyclical monotonicity for weak transport costs. *Calculus of Variations and Partial Differential Equations*, 2019.

[BS96] Dimitri P. Bertsekas and Steven E. Shreve. *Stochastic optimal control: the discrete time case*. Athena Scientific, 1996.

[Coh13] Donald L. Cohn. *Measure Theory (Second Edition)*. Springer Science & Business Media, 2013.

[Dob14] Ernst-Erich Doberkat. Measures and all that — a tutorial. *ArXiv*, 2014.

[GRST17] Nathael Gozlan, Cyril Roberto, Paul-Marie Samson, and Prasad Tetali. Kantorovich duality for general transport cost and applications. *Journal of Functional Analysis*, 2017.

[Kel84] Hans Kellerer. Duality theorems for marginal problems. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 1984.

[Vil16] Cédric Villani. *Optimal transport, old and new*. Springer Berlin Heidelberg, 2016, 2016.

[Zae15] Danila Zaev. On the monge-kantorovich problem with additional linear constraints. *Mathematical Notes*, 2015.

[Zal02] Constantin Zalinescu. *Convex analysis in general vector spaces*. World Scientific, 2002.