TECHNISCHE UNIVERSITÄT WIEN

DIPLOMARBEIT

Dichtheit von bikausalen Monge Kopplungen

ausgeführt zur Erlangung des akademischen Grades

Diplom-Ingenieur

unter der Anleitung von

Univ.-Prof. Dipl.-Ing. Dr. techn. Mathias Beiglböck

und

Dipl.-Ing. Gudmund Pammer, PhD

durch

Stefan Schrott, BSc

Matrikelnummer: 01607388

Wien, am 31.08.2021

Unterschrift Verfasser

Unterschrift Betreuer

TECHNISCHE
UNIVERSITÄT
WIEN

DIPLOMA THESIS

Denseness of Bicausal Monge Couplings

supervised by

Univ.-Prof. Dipl.-Ing. Dr. techn. Mathias Beiglböck

and

Dipl.-Ing. Gudmund Pammer, PhD

written by

Stefan Schrott, BSc
Student ID: 01607388

Vienna, August 31, 2021

Contents

Abstract

Consider a pair (X, Y) of random variables that have both a continuous law. It is well known that there is a sequence of bijections $(F_n)_n$ such that $F_n(X)$ is distributed like Y and the pairs $(X, F_n(X))$ converge to (X, Y) in distribution. The aim of this thesis is to prove an analogous statement for stochastic processes with finitely many time steps.

We consider processes $X = (X_1, \ldots, X_N)$ and $Y = (Y_1, \ldots, Y_N)$, which are compatible in the following sense: For all t the random variable $(Y_1, \ldots, Y_t)$ is conditionally independent of X given $(X_1, \ldots, X_t)$, and conversely $(X_1, \ldots, X_t)$ is independent of Y given $(Y_1, \ldots, Y_t)$, as well.

A mapping F from the path space of X to the path space of Y is called adapted if the t-th component of $F(x_1, \ldots, x_N)$ only depends on $x_1, \ldots, x_t$. A bijection F is called biadapted if both F and F^{-1} are adapted.

The aim of this thesis is to show that (under suitable regularity assumptions) there are biadapted mappings F_n from the path sapce of X to the path space of Y s.t. $F_n(X)$ is distributed like Y and $(X, F_n(X))$ converges to (X, Y) in distribution.

The joint distribution of processes X and Y that satisfy the compatibility assumption mentioned above are exactly the bicausal couplings. Therefore, the claim is equivalent to the fact that bicausal Monge couplings are dense in the set of bicausal couplings with fixed marginals w.r.t. weak convergence of probability measures, i.e. testing against continuous bounded functions.

Kurzfassung

Sei (X, Y) ein Paar von Zufallsvariablen mit stetiger Verteilung. Es ist bekannt, dass eine Folge von Bijektionen $(F_n)_n$ existiert, sodass $F_n(X)$ wie Y verteilt ist, und die Tupel $(X, F_n(X))$ in Verteilung gegen (X, Y) konvergieren. Das Ziel dieser Arbeit ist es, eine analoge Aussage für stochastische Prozesse mit endlich vielen Zeitschritten zu beweisen.

Dazu betrachten wir Prozesse $X = (X_1, \dots, X_N)$ und $Y = (Y_1, \dots, Y_N)$, die im folgenden Sinne kompatibel sind: Für jedes t ist $(Y_1, \dots, Y_t)$ unabhängig von X gegeben $(X_1, \dots, X_t)$, und auch umgekehrt: $(X_1, \dots, X_t)$ ist unabhängig von Y gegeben $(Y_1, \dots, Y_t)$.

Eine Abbildung F vom Pfadraum von X in den Pfadraum von Y heißt adaptiert, falls die t-te Komponente von $F(x_1, \dots, X_N)$ lediglich von $x_1, \dots, x_t$ abhängt. Eine Abbildung F heißt biadaptiert, falls F bijektiv ist und F und F^{-1} beide adaptiert sind.

Das Zeil der Arbeit ist es zu zeigen, dass es (unter gewissen Regularitätsbedingungen) biadaptierte Abbildungen F_n vom Pfadraum von X in den Pfadraum von Y gibt, sodass $F_n(X)$ wie Y verteilt ist und $(X, F_n(X))$ in Verteilung gegen (X, Y) konvergiert.

Die gemeinsamen Verteilungen von Prozessen X und Y mit obiger Kompatibilitätseigenschaft sind genau die bikausalen Kopplungen von X und Y. Die obige Aussage ist also äquivalent dazu, dass bei festgehaltenen Marginalien die bikausalen Monge-Kopplungen dicht in den bikausalen Kopplungen liegen, und zwar bezüglich schwacher Konvergenz durch Testen gegen stetige beschränkte Funktionen.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Diplomarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Wien, am 31.08.2021

Stefan Schrott

Acknowledgments

I would like to thank Mathias Beiglböck for introducing me into optimal transport and for suggesting the topic of this thesis. Moreover, I want to thank my supervisors Mathias Beiglböck and Gudmund Pammer for their continuous and helpful advice throughout the process of writing this thesis.

I would also like to thank Martin Goldstern for fruitful discussions that helped me to improve this thesis.

Moreover, I also want to say thank you to my family, friends and Leo for their constant support.

Stefan Schrott

The work on this thesis was supported by the Austrian Science Fund (FWF), Grant Y782.

Introduction

The first chapter of this theses covers the static case. In Section 1.1 basic definitions and results from optimal transport are recalled very briefly, for detailed introduction to optimal transport the reader is referred to [9] and [10]. Afterwards, we prove the following result (see Theorem 1.20), which was already established in [4, Proposition A.3]:

If μ is a continuous probability on the Polish space X and ν a continuous probability on the Polish space Y, then the set of couplings between μ and ν, which are supported on the graph of a bijection are, is dense in the set of couplings between μ and ν w.r.t. weak convergence, i.e. testing against continuous bonded functions.

We give a new proof of this result, which has the advantage that it can be extended to prove the time-depended version in Chapter 2. This proof crucially relies on the representation of a coupling π between μ and ν as a coupling $\widehat{\pi}$ between $\mu \otimes \lambda$ and $\nu \otimes \lambda$, which is supported on the graph of a bijection $T : X \to Y$ (Theorem 1.15).

In the second chapter we prove a time dependent version of this result. We consider the laws of stochastic processes with values in a Polish space (say $\mathbb{R}$) and N time steps, i.e. probability measures on $\mathbb{R}^N$. The goal is to prove that (under certain regularity assumptions on the marginals) any coupling π between probability measures μ and ν on $\mathbb{R}^N$, that respects the time structure, can be approximated by couplings between μ and ν that are supported by the graph of a bijection that respects the time structure.

We have to clarify what "respecting the time structure" means. For mappings this is the notion of adaptedness: A mapping $T : \mathbb{R}^N \to \mathbb{R}^N$ is adapted if, for any $t \leq N$, there are mappings $T_t : \mathbb{R}^t \to \mathbb{R}$ such that $T(x_1, \dots, x_N) = (T_1(x_1), \dots, T_N(x_1, \dots, x_N))$. So, one just needs to know $(x_1, \dots, x_t)$ in order to calculate the first t components of $T(x_1, \dots, x_N)$.

For couplings causality is the right notion of "respecting the time structure". That is basically a relaxation of adaptedness: A coupling π is causal if one only needs to know $(x_1, \dots, x_t)$ in order to calculate $\pi^{x_1,\dots,x_N}(B)$ for sets $B \subseteq \mathbb{R}^N \times \mathbb{R}^N$ that only depend on the first t coordinates in y-direction (i.e. that are measureable w.r.t. the σ-algebra that is generated by the mapping $\mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^t : (x_1, \dots, x_N, y_1, \dots, y_N) \mapsto (y_1, \dots y_t)$).

A bijection T is called biadapted if T and T^{-1} are both adapted and a coupling π is bicausal if it is causal by itself and causal if we exchange the x- and y-coordinates. See Section 2.1 and 2.2 for more details.

Now we have explained the terms to state the main Theorem 2.26 more precisely: Let μ and ν be probability measures on $\mathbb{R}^N$ (satisfying some regularity conditions) and let π be a bicausal coupling between μ and ν. Then there is a sequence of biadapted mappings $(T_n)_n$ that push μ to ν such that the couplings $\mu \circ (\mathrm{id}, T_n)^{-1}$ weakly converge to π.

In order to prove the results of this thesis rigorously without getting lost in technical details, some measurability aspects were postponed to the Appendix.

Notation

Polish spaces and standard Borel spaces will be denoted with capital letters, such as X or Y. Collections of subsets of them (e.g. topologies or σ-algebras) will be denoted by calligraphic letters such as $\mathcal{B}$ or $\mathcal{T}$. The power set of X will be denoted by 2^X.

We will always equip the spaces X and Y with the Borel σ-algebra generated by their Polish topology. $\mathcal{P}(X)$ denotes the set of Borel probability measures on X. Probability measures are denoted with small Greek letters such as μ, ν and π. λ always denotes the Lebesgue measure on $[0,1]$. For a measure μ on (some subset of) $\mathbb{R}$ we denote its distribution function as F_μ, i.e. $F_\mu(t) := \mu((-\infty, t])$ and its inverse distribution function (or quantile function) as F_μ^{-1}.

The measurability of mapping is always to be understood w.r.t. the Borel σ-algebra. Given a measurable mapping $f : X \to Y$ and measure $\mu \in \mathcal{P}(X)$, we denote the pushforward of μ under f as $f_*\mu$, i.e. $f_*\mu(A) := \mu(f^{-1}(A))$ for all $A \subseteq Y$ Borel. For a further mapping $g : Y \to Z$ we define $g_* f_* \mu := g_*(f_*\mu) = (g \circ f)_*\mu$ to avoid unnecessary brackets.

We equip $\mathcal{P}(X)$ with the weak convergence by testing against continuous bounded functions, i.e.

$$\mu_n \rightharpoonup \mu :\iff \forall f : X \to \mathbb{R} \text{ continuous and bounded: } \int f d\mu_n \to \int f d\mu.$$

A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a set, $\mathcal{F}$ is a σ-algebra on Ω and $\mathbb{P} \in \mathcal{P}(\Omega)$. A random variable is a measurable function from Ω to a Polish space. We will denote random variables with sans serif letters to avoid conflicting notations, e.g. X and Y.

The law of a random variable X, denoted by $\mathcal{L}(\mathsf{X})$, is the probability measure $\mathsf{X}_*\mathbb{P}$. We write $\mathsf{X} \sim \mu$ for $\mathcal{L}(\mathsf{X}) = \mu$ and $\mathsf{X} \sim \mathsf{Y}$ for $\mathcal{L}(\mathsf{X}) = \mathcal{L}(\mathsf{Y})$.

A kernel from Z to X is a function $\pi : Z \to \mathcal{P}(X)$. We denote the probability measure $\pi(z)$ as π^z. We introduce a similar notation for functions: Given a function $F : Z \times X \to Y$ (which can also be seen as a function $F : Z \to Y^X$) and $z \in Z$ we define the function $F^z : X \to Y$ as $F^z(x) = F(z, x)$.

For spaces X and Y we introduce the function $e : X \times Y \to Y \times X : (x, y) \mapsto (y, x)$, i.e. e exchanges the order of X and Y.

For more details the reader may consult the appendix.

Chapter 1

The static case

1.1 Introduction to optimal transport

This section gives a very brief introduction to the optimal transport problem, for detailed introduction to optimal transport the reader is referred to [9] and [10].

The aim of optimal transport is to transport a given distribution μ on a Polish space X to another given distribution ν on a Polish space Y in the cheapest way with respect to a given cost function. The most natural way to clarify what is meant by "transport μ to ν" is considering mappings $T : X \to Y$ that push μ to ν. This leads to the Monge transport problem

$$\inf \left\{ \int c(x, T(x)) d\mu(x) : T : X \to Y \text{ s.t. } T_*\mu = \nu \right\},$$

where $c : X \times Y \to \mathbb{R}$ is a given cost function.

It turned out that a relaxed version of this problem is more accessible to analytic techniques. Here the transport mappings $T : X \to Y$ s.t. $T_*\mu = \nu$ are replaced by couplings (also called transport plans):

Definition 1.1. Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. The set of couplings between μ and ν is defined as

$$\mathrm{Cpl}(\mu, \nu) := \{\pi \in \mathcal{P}(X \times Y) : \mathrm{pr}_{X*}\pi = \mu, \mathrm{pr}_{Y*}\pi = \nu\}.$$

This leads to the socalled Kantrovich transport problem:

Definition 1.2. Let $c : X \times Y \to \mathbb{R}$ a cost function. Then Kantorovich problem is

$$\inf \left\{ \int c(x, y)\pi(dx, dy) : \pi \in \mathrm{Cpl}(\mu, \nu) \right\}.$$

A first and crucial oberservation is that transport mappings are special cases of couplings: In fact, if $T_*\mu = \nu$, then $\pi := (\mathrm{id}, T)_*\mu \in \mathrm{Cpl}(\mu, \nu)$ and $\int c(x, y) d\pi(x, y) = \int c(x, T(x)) d\mu(x)$, so the Kantrovich problem is indeed are relaxation of the Monge problem. We introduce a notation for those couplings that are "induced" by a map:

Definition 1.3. Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. The set of Monge couplings between μ and ν is defined as

$$\mathrm{Cpl}_0(\mu,\nu) := \{(id,T)_*\mu : T : X \to Y \text{ measureable}, T_*\mu = \nu\}.$$

Remark 1.4. One can easily see that $\pi \in \mathrm{Cpl}(\mu,\nu)$ is a Monge coupling if and only if its regular disintegration $\pi(dx,dy) = \mu(dx)\pi^x(dy)$ has the property that π^x is μ-a.s. a Dirac measure. So, one can interpret the relaxation from mappings to couplings as introducing some randomization: Instead of deterministically prescribing that a point x is transported to some point $T(x)$, we prescribe a probability distribution of its target, namely π^x.

We want to prove the existence of minimizes for the Kanatrovich problem because this gives a good intuition for the topic and the lemmas we need for this proof will be needed in other parts of the thesis as well.

We first observe that the set of couplings with fixed marginals is compact w.r.t. weak convergence of probability measures:

Proposition 1.5. $\mathrm{Cpl}(\mu,\nu)$ *is compact w.r.t. weak convergence of probability measures.*

Proof. Since the mappings $\pi \mapsto \mathrm{pr}_{X*}\pi$ and $\pi \mapsto \mathrm{pr}_{Y*}\pi$ are both continuous by Lemma A.44, the set $\mathrm{Cpl}(\mu,\nu)$ is closed.

We show tightness to conclude compactness with Prokhorov's Theorem A.21. For $\varepsilon > 0$ there exist compact set $K_X \subseteq X$ and $K_Y \subseteq Y$ s.t. $\mu(K_X^c) < \varepsilon/2$ and $\nu(K_Y^c) < \varepsilon/2$. Then for any $\pi \in \mathrm{Cpl}(\mu,\nu)$:

$$\pi((K_X \times K_Y)^c) \leq \pi(K_X^c \times Y) + \pi(X \times K_Y^c) = \mu(K_X^c) + \nu(K_Y^c) < \varepsilon. \qquad \square$$

For a continuous bounded function $f : X \to \mathbb{R}$ the mapping $\mu \mapsto \int f d\mu$ is per definition continuous w.r.t. weak convergence. The following lemma generalizes this fact a bit:

Lemma 1.6. *Let* $f : X \to \mathbb{R}$ *be lower semi continuous and bounded from below. Then*

$$\mathcal{P}(X) \to \mathbb{R} : \mu \mapsto \int f d\mu$$

is lower semi continuous.

Proof. We first show that there a continuous bounded functions f_k s.t. $f = \sup_k f_k$. To that end, let d be a compatible metric for X. Then the functions

$$f_k(x) := \Big[\inf_{y \in X}(f(y) + kd(x,y))\Big] \wedge k$$

are k-Lipschitz and bounded (from above by k and from below by the lower bound of f). Obviously, $f_k(x) \leq f(x)$, so it suffices to show $\sup_k f_k(x) \geq f(x)$.

For $k > f(x)$ pick $y_k \in X$ s.t.

$$f(y_k) + kd(x,y_k) \leq f_k(x) + 1/k.$$

Then

$$d(x,y_k) \leq \frac{1}{k}\Big[f_k(x) + \frac{1}{k} - f(y_k)\Big] \leq \frac{1}{k}\Big[f(x) + 1 + |\inf_{y\in X} f(y)|\Big] \to 0,$$

so $y_k \to x$. Since f is l.s.c. and by the definition of y_k we have

$$f(x) \leq \liminf_k f(y_k) \leq \liminf_k f_k(x) + 1/k \leq \sup_k f_k(x).$$

Hence, for all $k \in \mathbb{N}$:

$$\int f_k d\mu = \liminf_n \int f_k d\mu_n \leq \liminf_n \int f d\mu_n$$

and by using dominated convergence in the limit $k \to \infty$

$$\int f d\mu = \lim_k \int f_k d\mu \leq \liminf_n \int f d\mu_n. \qquad \square$$

These two lemmas are already enough to prove the existence of minimizers:

Theorem 1.7. *Let $c : X \times Y \to \mathbb{R}$ be lower semi-continuous and bounded from below. If*

$$\inf_{\pi \in \mathrm{Cpl}(\mu,\nu)} \int c d\pi < \infty,$$

there exists a minimizer in the Kantorovich problem.

Proof. For $n \in \mathbb{N}$ let $\pi_n \in \mathrm{Cpl}(\mu, \nu)$ s.t.

$$\int c d\pi_n \leq \inf_{\pi \in \mathrm{Cpl}(\mu,\nu)} \int c d\pi + \frac{1}{n}.$$

By Proposition 1.5 there exists a subsequence $(\pi_{n_k})_k$ converging weakly to some $\pi \in \mathrm{Cpl}(\mu, \nu)$. By Lemma 1.6 we have

$$\int c d\pi \leq \liminf_n \int c d\pi_n \leq \liminf_n \inf_{\pi \in \mathrm{Cpl}(\mu,\nu)} \int c d\pi + \frac{1}{n} = \inf_{\pi \in \mathrm{Cpl}(\mu,\nu)} \int c d\pi. \qquad \square$$

To close this section, we want to mention how the optimal transport problem can be used to define a metric on the set of probability measures. This metric is called Wasserstein distance.

Definition 1.8. Let X be a Polish space and d be a compatible metric. For $p \in [1, \infty)$ let $\mathcal{P}_p(X)$ be the set of all $\mu \in \mathcal{P}(X)$ s.t. $\int d(x, x_0)^p < \infty$ for some (and therefore any) $x_0 \in X$. For $\mu, \nu \in \mathcal{P}_p(X)$ define

$$\mathcal{W}_p(\mu, \nu) := \inf \left\{ \int d(x, y)^p \pi(dx, dy) : \pi \in \mathrm{Cpl}(\mu, \nu) \right\}^{1/p}$$

It is a very interesting and useful fact that the Wasserstein distance metrizes the weak convergence of probability measures:

Theorem 1.9 ([9, Theorem 7.12], [10, Theorem 6.18])**.** *Let X be a Polish space, d be a compatible metric and $p \in [1, \infty)$. Then for any sequence $(\mu_n)_n$ in $\mathcal{P}_p(X)$ and $\mu \in \mathcal{P}_p(X)$ the following are equivalent:*

(i) $\mathcal{W}_p(\mu_n, \mu) \to 0$

(ii) $\mu_n \rightharpoonup \nu$ *and* $\int d(x_0, x)^p d\mu_n(x) \to \int d(x_0, x)^p d\mu$ *for some (and therefore any)* $x_0 \in X$.

Moreover, $\mathcal{W}_p$ is complete and therefore a compatible metric for the Polish space $\mathcal{P}_p(X)$.

In particular, if d is a bounded metric, then $\mathcal{P}_p(X) = \mathcal{P}(X)$ and $\mathcal{W}_p$ metricizes the weak convergence of probability measures.

1.2 Representation of couplings between μ and ν as bijective Monge couplings between $\mu \otimes \lambda$ and $\nu \otimes \lambda$

In some sense, Monge couplings are simpler than general couplings. In this section we will consider Monge couplings, which are inducted by bijections.

Definition 1.10. For $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ we define the set of Monge couplings supported on the graph of a bijection:

$$\mathrm{Cpl}_{00}(\mu, \nu) := \{(id, T)_*\mu : T : X \to Y \text{ bimeasureable}, T_*\mu = \nu\}.$$

A relation $R \subseteq X \times Y$ is the graph of a bijection from X to Y if and only if R is the graph of a mapping from X to Y and the inverse relation $R^{-1} \subseteq Y \times X$ the graph of a mapping from Y to X. The same is true for couplings:

Lemma 1.11. *$\pi \in \mathrm{Cpl}_{00}(\mu, \nu)$ if and only if $\pi \in \mathrm{Cpl}_0(\mu, \nu)$ and $e_*\pi \in \mathrm{Cpl}_0(\nu, \mu)$.*

Proof. Assume that $\pi \in \mathrm{Cpl}_0(\mu, \nu)$ and $e_*\pi \in \mathrm{Cpl}_0(\nu, \mu)$, i.e. there exist Borel mappings $S : X \to Y$ and $T : Y \to X$ such that $\pi = (\mathrm{id}, S)_*\mu$ and $e_*\pi = (\mathrm{id}, T)_*\nu$. By Theorem A.13 the sets $\mathrm{graph}(S)$ and $\mathrm{graph}^{-1}(T) := \{(T(y), y) : y \in Y\}$ are both Borel. Clearly, $\pi(\mathrm{graph}(S)) = 1$ and $\pi(\mathrm{graph}^{-1}(T)) = 1$. Hence, $R := \mathrm{graph}(S) \cap \mathrm{graph}^{-1}(T)$ is Borel and $\pi(R) = 1$. Moreover, R is the graph of a bijection between $\mathrm{pr}_X(R)$ and $\mathrm{pr}_Y(R)$. It is easy to see that $\mathrm{pr}_X(R) = (\mathrm{id}, S)^{-1}(\mathrm{graph}^{-1}(T))$ and $\mathrm{pr}_Y(R) = (\mathrm{id}, T)^{-1}(\mathrm{graph}^{-1}(S))$, so $\mathrm{pr}_X(R)$ and $\mathrm{pr}_Y(R)$ are both Borel. Hence, $R : \mathrm{pr}_X(R) \to \mathrm{pr}_Y(R)$ is a Borel isomorphism by Theorem A.13. By Theorem A.14 there exists a Borel isomorphism $G : X \setminus \mathrm{pr}_X(R) \to Y \setminus \mathrm{pr}_Y(R)$. Then the mapping

$$F : X \to Y : x \mapsto \begin{cases} R(x) & x \in \mathrm{pr}_X(R) \\ G(x) & x \notin \mathrm{pr}_X(R) \end{cases}$$

is a Borel ismorphism satisfying $\pi = (\mathrm{id}, F)_*\mu$. □

The main goal of Chapter 1 is to show that those couplings are dense in $\mathrm{Cpl}(\mu, \nu)$ if the marginals μ and ν are both continuous. A crucial step in the proof of this will be a representation of a coupling $\pi \in \mathrm{Cpl}(\mu, \nu)$ by a coupling $\widehat{\pi} \in \mathrm{Cpl}_{00}(\mu \otimes \lambda, \nu \otimes \lambda)$. By "representation" we mean that we can recover π from $\widehat{\pi}$ by projecting from $X \times [0, 1] \times Y \times [0, 1]$ onto $X \times Y$. Precisely, the aim of this section is to prove the following statement:

Theorem 1.12. *Let X and Y be standard Borel spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and let $\pi \in \mathrm{Cpl}(\mu, \nu)$. Then there exists some measurable bijection $T : X \times [0,1] \to Y \times [0,1]$ satisfying the following properties:*

(i) $T_*(\mu \otimes \lambda) = \nu \otimes \lambda$

(ii) *If $pr_{XY} : X \times [0,1] \times Y \times [0,1] \to X \times Y$ denotes the projection and $\widehat{\pi} := (id, T)(\mu \otimes \lambda)$, it holds $pr_{XY}(\widehat{\pi}) = \pi$.*

Remark 1.13 (see also [5, Lemma 3.22])**.** Since the proof of this is quite technical, we consider an easier problem, namely to represent a coupling $\pi \in \mathrm{Cpl}(\mu, \nu)$ by a Monge coupling $\widehat{\pi} \in \mathrm{Cpl}(\mu \otimes \lambda, \nu)$ in a way s.t. we can recover π from $\widehat{\pi}$ when projecting from $X \times [0,1] \times Y$ onto $X \times Y$.

In Remark 1.4 we have argued that general couplings are in some sense a randomized version of Monge couplings. So, the idea is to add an additional coordinate, which is distributed as λ and realizes that randomization. By Theorem A.14 we can assume that $X = Y = [0,1]$. Define the mapping $T(x,u) := F^{-1}_{\pi^x}(u)$. Clearly, $T(x,\cdot)_*\lambda = \pi^x$. If $f : X \times Y \to \mathbb{R}$ is a measureable function, we see that

$$\int f(x,y) d(\mathrm{id}, T)_*(\mu \otimes \lambda) = \int f(x, F^{-1}_{\pi^x}(u)) d\lambda(u) d\mu(x) = \int f(x,y) d\pi^x(y) d\mu(x) = \int f(x,y) d\pi(x,y),$$

so $\mathrm{pr}_{XY*} T_*(\mu \otimes \lambda) = \pi$ and hence $T_*(\mu \otimes \lambda) \in \mathrm{Cpl}(\mu \otimes \lambda, \nu)$.

If $\pi = (\mathrm{id}, T)_*\mu \in \mathrm{Cpl}(\mu, \nu)$ is a Monge coupling, the conditional probabilities w.r.t. the first coordinate are Dirac measures (i.e. $\pi^x = \delta_{T(x)}$), however the conditional probabilities w.r.t. the second coordinate do not need to be Dirac, unless the Monge mapping is injective. So, in some sense a Monge coupling can still contain randomness (given some $y \in Y$ one can in general not determine "from which x the mass in y came").

Hence, it is reasonable that a representation with Monge couplings supported on the graph of a bijection will in general need more randomization. The following easy example shows that this is in fact true:

Expample 1.14. Let $X = Y = [0,1]$, $\mu = \lambda$, $\nu = \delta_0$ and $\pi = \lambda \otimes \delta_0$. Assume that there is a bijection $T : X \times [0,1] \to Y$ s.t. $\widehat{\pi} = (\mathrm{id}, T)_*\lambda^2 \in \mathrm{Cpl}(\mu \otimes \lambda, \nu) = \mathrm{Cpl}(\lambda^2, \delta_0)$. Then, on the one hand $T^{-1}(\{0\})$ has to contain exactly one element, on the other hand $\lambda^2(T^{-1}(\{0\})) = 1$, which is a contradiction.

The idea for the proof of the representation of $\pi \in \mathrm{Cpl}(\mu, \nu)$ as a coupling $\widehat{\pi} \in \mathrm{Cpl}_{00}(\mu \otimes \lambda, \nu \otimes \lambda)$ is to consider the following mapping:

$$T : X \times [0,1] \to Y \times [0,1] : (x,u) \mapsto (y,v), \text{ where } y = F^{-1}_{\pi^x}(u), v = F_{\pi^y}(x).$$

As in Remark 1.13, we calculate y given some x and the randomization u as $y = F^{-1}_{\pi^x}(u)$. For the definition of v we observe that it is the randomization variable that belongs to the Y-component and that we already have a prescribed value for y. So, we have to ask: "given y and knowing that the result of our calculation is x, what is the suitable value for the randomization?" Therefore,

v should satisfy $x = F_{\pi^y}^{-1}(v)$, so $v = F_{\pi^y}(x)$. Moreover, it is easy to check that the mapping[1]

$$S : Y \times [0,1] \to X \times [0,1] : (y,v) \mapsto (x,u), \text{ where } x = F_{\pi^y}^{-1}(v), u = F_{\pi^x}(y)$$

is the inverse of T, in particular T is a bijection.

This fact (as well as the last step in the motivation of the definition of T) crucially depend on the fact that the mappings F_{π^x} and F_{π^y} are all bijective. This is only possible for couplings π, whose conditional probabilities π^x and π^y are all non-atomic. Since this is no reasonable assumption on couplings (it in fails many important cases), we will have to use a somewhat more sophisticated construction to overcome this issue.

Moreover, we will prove a more general, parameterized version of this theorem in order to avoid measureability issues in Chapter 2. This proof relies on Corollary A.41, which is proven in the appendix.

Theorem 1.15. *Let X, Y, Z be standard Borel spaces and π a kernel from Z to $X \times Y$. Let μ denote the kernel from Z to X defined by $\mu^z := {\mathrm{pr}_X}_* \pi^z$ and denote ν be the kernel from Z to Y defined by $\nu^z := {\mathrm{pr}_Y}_* \pi^z$. (i.e. $\pi^z \in \mathrm{Cpl}(\mu^z, \nu^z)$ for all $z \in Z$.)*

Then there exists a Borel measurable mapping $T : Z \times X \times [0,1] \to Y \times [0,1]$ s.t. for all $z \in Z$ the mappings $T^z : X \times [0,1] \to Y \times [0,1] : (x,u) \mapsto T(z,x,u)$ are Borel isomorphisms satisfying

(i) ${T^z}_*(\mu^z \otimes \lambda) = \nu^z \otimes \lambda$

(ii) ${\mathrm{pr}_{XY}}_*(id, T^z)_*(\mu^z \otimes \lambda) = \pi^z$.

Proof. By Corollary A.41 there exits a measureable mapping

$$G : (Z \times X) \times Y \times [0,1] \to [0,1]^2$$

s.t. for all $(z,x) \in Z \times X$ the mapping $G^{z,x} := G(z,x,\cdot) : Y \times [0,1] \to [0,1]^2$ is a Borel isomorphism satisfying $G_*^{z,x}(\pi^{z,x} \otimes \lambda) = \lambda^2$.

Again by Corollary A.41, there exits a measureable mapping

$$H : (Z \times Y) \times X \times [0,1] \to [0,1]^2$$

s.t. for all $(z,y) \in Z \times Y$ the mapping $H^{z,y} := H(z,y,\cdot) : X \times [0,1] \to [0,1]^2$ is a Borel isomorphism satisfying $H_*^{z,y}(\pi^{z,y} \otimes \lambda) = \lambda^2$.

Consider the mapping $S : Z \times X \times [0,1]^3 \to Y \times [0,1]^3$ defined by $S(z, x_1, x_2, u_1, u_2) = (y_1, y_2, v_1, v_2)$, where

$$(y_1, y_2) = (G^{z,x_1})^{-1}(u_1, u_2) \qquad (v_1, v_2) = H^{z,y_1}(x_1, x_2).$$

Clearly, S is Borel. For $z \in Z$ we denote $S^z := S(z, \cdot) : X \times [0,1]^3 \to Y \times [0,1]^3$.

Our aim is to show that for all $z \in Z$ the mapping S^z is a Borel isomorphism satisfying

[1] One has to be careful when reading the definition of S: Basically, S is the same mapping as T but for the coupling $e_*\pi \in \mathrm{Cpl}(\nu, \mu)$, where $e(x,y) := (y,x)$. Hence, the $F_{\pi^y}^{-1}$ in definition of S are the quantile functions of π conditioned on some $y \in Y$, i.e. we do not have just changed the names of the variables x and y when defining an inverse function, in fact we disintegrate w.r.t. to another coordinate as in the definition of T.

(i) $S^z{}_*(\mu^z \otimes \lambda^3) = \nu^z \otimes \lambda^3$

(ii) $\text{pr}_{XY*}(id, S^z)_*(\mu^z \otimes \lambda^3) = \pi^z$.

In order to prove the injectivity of S^z, let $(x_1, x_2, u_1, u_2) \neq (\bar{x}_1, \bar{x}_2, \bar{u}_1, \bar{u}_2)$ be given. We have to show that $(y_1, y_2, v_1, v_2) := S^z(x_1, x_2, u_1, u_2)$ and $(\bar{y}_1, \bar{y}_2, \bar{v}_1, \bar{v}_2) := S^z(\bar{x}_1, \bar{x}_2, \bar{u}_1, \bar{u}_2)$ are different. In the case $(y_1, y_2) \neq (\bar{y}_1, \bar{y}_2)$ there is nothing to prove, so we may assume $(y_1, y_2) = (\bar{y}_1, \bar{y}_2)$. We distinguish two cases:

Case 1: $x_1 = \bar{x}_1$. This implies $(G^{z,x_1})^{-1} = (G^{z,\bar{x}_1})^{-1}$ and by the injectivity of this mapping we get $(u_1, u_2) = (\bar{u}_1, \bar{u}_2)$. Since $(x_1, x_2, u_1, u_2) \neq (\bar{x}_1, \bar{x}_2, \bar{u}_1, \bar{u}_2)$ this implies $x_2 \neq \bar{x}_2$ and by the injectivity of H^{z,y_1} this implies $(v_1, v_2) \neq (\bar{v}_1, \bar{v}_2)$.

Case 2: $x_1 \neq \bar{x}_1$. Then by the injectivity of H^{z,y_1} again $(v_1, v_2) \neq (\bar{v}_1, \bar{v}_2)$.

For proving the surjecitvity, let (y_1, y_2, v_1, v_2) be given. By the surjectivity of H^{z,y_1} there are (x_1, x_2) such that $H^{z,y_1}(x_1, x_2) = (v_1, v_2)$. Now, by the surjectivity of $(G^{z,x_1})^{-1}$, there exists (u_1, u_2) such that $(G^{z,x_1})^{-1}(u_1, u_2) = (y_1, y_2)$.

We have shown that S^z is a Borel measurable bijection and by Theorem A.13 it is a Borel isomorphism.

Property (i). Let $f : Y \times [0, 1]^3$ be measurable function. Then it holds

$$\begin{aligned}
&\int f(y_1, y_2, v_1, v_2) dS^z_*(\mu^z \otimes \lambda^3)(y_1, y_2, v_1, v_2) = \\
&\quad = \int f((G^{z,x_1})^{-1}(u_1, u_2), H(z, \text{pr}_Y((G^{z,x_1})^{-1}(u_1, u_2)), x_1, x_2)) d(\mu^z \otimes \lambda^3)(x_1, x_2, u_1, u_2) \\
&\quad = \int f(y_1, y_2, H^{z,y_1}(x_1, x_2)) \underbrace{d(G^{z,x_1})^{-1}_* \lambda^2(y_1, y_2)}_{= d(\pi^{z,x_1} \otimes \lambda)(y_1, y_2)} d(\mu^z \otimes \lambda)(x_1, x_2) \\
&\quad = \int f(y_1, y_2, H^{z,y_1}(x_1, x_2)) \underbrace{d\pi^{z,x_1}(y_1) d\mu^z(x_1)}_{\substack{= d\pi^z(x_1, y_1) \\ = d\pi^{z,y_1}(x_1) d\nu^z(y_1)}} d\lambda^2(x_2, y_2) \\
&\quad = \int f(y_1, y_2, H^{z,y_1}(x_1, x_2)) d\pi^{z,y_1}(x_1) d\lambda(x_2) d\nu^z(y_1) d\lambda(y_2) \\
&\quad = \int f(y_1, y_2, v_1, v_2) \underbrace{dH^{z,y_1}_*(\pi^{z,y_1} \otimes \lambda)(v_1, v_2)}_{= d\lambda^2(v_1, v_2)} d\nu^z(y_1) d\lambda(y_2) \\
&\quad = \int f(y_1, y_2, v_1, v_2) d(\nu^z \otimes \lambda^3)(y_1, y_2, v_1, v_2),
\end{aligned}$$

which yields $S^z(\mu^z \otimes \lambda^3) = \nu^z \otimes \lambda^3$.

Property (ii). Note that $(\text{pr}_X \circ (G^{z,x_1})^{-1})_* \lambda^2 = \pi^{z,x_1}$ and that

$$\text{pr}_{XY} \circ (\text{id}_{[0,1]^4}, S^z) : X \times [0, 1]^3 \to X \times Y : (x_1, x_2, u_1, u_2) \mapsto (x_1, \text{pr}_Y((G^{z,x_1})^{-1}(u_1, u_2))).$$

Hence, for any measurable function $f : X \times Y \to \mathbb{R}$ we have

$$\int f(x_1, y_1) d\mathrm{pr}_{XY*}(\mathrm{id}_{[0,1]^4}, S^z)_*(\mu^z \otimes \lambda^3)(x_1, y_1) =$$

$$= \int f(x_1, \mathrm{pr}_Y((G^{z,x_1})^{-1}(u_1, u_2))) d\mu^z(x_1) d\lambda^3(x_2, u_1, u_2)$$

$$= \int f(x_1, y_1) \underbrace{d(G^{z,x_1})_*^{-1}\lambda^2(u_1, u_2)}_{= d\pi^{z,x_1}(y_1) d\lambda(y_2)} d\mu^z(x_1)$$

$$= \int f(x_1, y_1) d\pi^{z,x_1}(y_1) d\mu^z(x_1)$$

$$= \int f(x_1, y_1) d\pi^z(x_1, y_1),$$

which shows that $\mathrm{pr}_{XY*}(\mathrm{id}_{X\times[0,1]^3}, S^z)_*(\mu^z \otimes \lambda^3) = \pi^z$.

By Theorem A.19 there exits a Borel isomorphism $h : [0, 1] \to [0, 1]^3$ satisfying $h_*\lambda = \lambda^3$. Define[2]

$$T := (\mathrm{id}_Y \times h^{-1}) \circ S \circ (\mathrm{id}_Z \times \mathrm{id}_X \times h) : Z \times X \times [0, 1] \to Y \times [0, 1]$$

Clearly, T is measurable as composition. For $z \in Z$ it holds $T^z = (\mathrm{id}_Y \times h^{-1}) \circ S^z \circ (\mathrm{id}_X \times h)$, so T^z is a Borel isomorphism as composition of Borel isomorphisms.

Moreover, it holds

$$\begin{aligned} T^z_*(\mu^z \otimes \lambda) &= (\mathrm{id}_Y \times h^{-1})_* S^z_* (\mathrm{id}_X \times h)_*(\mu^z \otimes \lambda) = (\mathrm{id}_Y \times h^{-1})_* S^z_*(\mu \otimes \lambda^3) \\ &= (\mathrm{id}_Y \times h^{-1})_*(\nu^z \otimes \lambda^3) = \nu^z \otimes \lambda. \end{aligned}$$

Denote $\widehat{\pi} := (\mathrm{id}, T)_*(\mu \otimes \lambda)$. In order to check that $\widehat{\pi^z} = (\mathrm{id}, T^z)_*(\mu^z \otimes \lambda)$ satisfies $\mathrm{pr}_{XY*}\widehat{\pi^z} = \pi^z$, observe that

$$(\mathrm{id}_{X\times[0,1]}, T) = ((\mathrm{id}_X \times h^{-1}), (\mathrm{id}_Y \times h^{-1})) \circ (\mathrm{id}_{X\times[0,1]^3}, S) \circ (\mathrm{id}_X \times h)$$

and therefore

$$\mathrm{pr}_{XY} \circ (\mathrm{id}_{X\times[0,1]}, T) = \mathrm{pr}_{XY} \circ (\mathrm{id}_{X\times[0,1]^3}, S) \circ (\mathrm{id}_X \times h),$$

where pr_{XY} denotes on the left hand side the projection $X \times [0, 1] \times Y \times [0, 1] \to X \times Y$ and on the right hand side the projection $X \times [0, 1]^3 \times Y \times [0, 1]^3 \to X \times Y$.

Using this, we see that

$$\begin{aligned} \mathrm{pr}_{XY*}\widehat{\pi^z} &= \mathrm{pr}_{XY*}(\mathrm{id}_{X\times[0,1]}, T^z)_*(\mu^z \otimes \lambda) = \mathrm{pr}_{XY*}(\mathrm{id}_{X\times[0,1]^3}, S^z)_*(\mathrm{id}_X \times h)_*(\mu^z \otimes \lambda) \\ &= \mathrm{pr}_{XY*}(\mathrm{id}_{X\times[0,1]^3}, S^z)_*(\mu^z \otimes \lambda^3) = \pi^z, \end{aligned}$$

which shows that T has all the desired properties. □

We close this section with an example that shows that this representation is not unique:

Expample 1.16. Let $X = Y = [0, 1]$, $\mu = \nu = \lambda$ and $\pi = \lambda^2$. Then $T(x, u) := (u, x)$ and $S(u, x) := (u, 1 - x)$ are both representation of π. It is easy to check that $T_*(\mu \otimes \lambda) = T_*\lambda^2 = \lambda^2$ and $\mathrm{pr}_{XY} \circ (\mathrm{id}, T)(x, u) = (x, u)$, so $\mathrm{pr}_{XY*}(\mathrm{id}, T)_*(\mu \otimes \lambda) = \mathrm{pr}_{XY*}(\mathrm{id}, T)_*\lambda^2 = \lambda^2 = \pi$. Since the mapping $x \mapsto 1 - x$ pushes λ to λ, one can easily see that S is a suitable representation as well.

[2]To clarify the notation: For $f : A \to B$ and $g : A \to C$ we define $(f, g) : A \to B \times C : a \mapsto (f(a), g(a))$. For $f : A \to B$ and $g : C \to D$ we define $f \times g : A \times C \to B \times D : (a, c) \mapsto (f(a), g(c))$.

1.3 Denseness of couplings supported by the graph of a bijection

The aim of this section is to prove that $\mathrm{Cpl}_{00}(\mu,\nu)$ is dense in $\mathrm{Cpl}(\mu,\nu)$ if μ and ν are continuous measures. For that purpose, we need to approximate a given coupling $\pi \in \mathrm{Cpl}(\mu,\nu)$ by a sequence $(\pi_n)_n$ in $\mathrm{Cpl}_{00}(\mu,\nu)$.

To that end, we use the representation of π as $\widehat{\pi} = (\mathrm{id}, T)_*(\mu \otimes \lambda) \in \mathrm{Cpl}_{00}(\mu \otimes \lambda, \nu \otimes \lambda)$ from the previous section. Then we choose sequences of partitions (with mesh converging to zero) of the spaces X and Y (Theorem A.45) and bijections between X and $X \times [0,1]$ (and respectively between Y and $Y \times [0,1]$), which are compatible with these partitions (Proposition 1.19). In the proof of Theorem 1.20 we show that the concatenation of these compatible bijections and T is a suitable approximating sequence.

For a partition $\mathcal{M}$ of a metric space X we define its mesh as $||\mathcal{M}|| := \sup_{M\in\mathcal{M}} \mathrm{diam}(M)$. The following proposition that is proven in the appendix ensures the existence of a suitable sequence of partitions.

Proposition 1.17. *Let X be a Polish space and d be a compatible metric. Then there exists a sequence $(\mathcal{M}_n)_{n\in\mathbb{N}}$ of partitions of X that consist of at most countably many Borel subsets of X satisfying $\lim_{n\to\infty} ||\mathcal{M}_n|| = 0$.*

Proof. See Theorem A.45 and Remark A.46. □

The following condition for weak convergence is convenient for proving the convergence of the approximating sequence that we construct in the proof of Theorem 1.20.

Lemma 1.18. *Let X be a Polish space and d a compatible metric. For each $n \in \mathbb{N}$ let $\mathcal{M}_n$ be a partition of X consisting of at most countably many Borel sets such that $\lim_{n\to\infty} ||\mathcal{M}_n|| = 0$. Let $\mu_n, \mu \in \mathcal{P}(X)$ be satisfying $\mu_n(M) = \mu(M)$ for all $M \in \mathcal{M}_n$. Then $\mu_n \to \mu$ weakly.*

Proof. For $n \in \mathbb{N}$ consider the coupling

$$\pi_n := \sum_{M\in\mathcal{M}_n} \frac{1}{\mu(M)} \mu_n|_M \otimes \mu|_M.$$

If $||\cdot||$ denotes the total variation norm, we see that

$$\sum_{M\in\mathcal{M}_n} \left\| \frac{\mu_n|_M \otimes \mu|_M}{\mu(M)} \right\| = \sum_{M\in\mathcal{M}_n} \mu_n(M) = \mu_n(X) = 1,$$

so the sum in the definition of π_n converges absolutely w.r.t. the totalvariation norm, hence the sum converges w.r.t. weak convergence as well. Since the pushforward w.r.t. a continous function is continuous w.r.t. weak convergence it holds

$$\mathrm{pr}_{X*}\pi_n = \sum_{M\in\mathcal{M}_n} \mathrm{pr}_{X*} \frac{\mu_n|_M \otimes \mu|_M}{\mu(M)} = \sum_{M\in\mathcal{M}_n} \mu_n|_M = \mu_n$$

and since $\mu_n(M) = \mu(M)$ for all $M \in \mathcal{M}_n$ it holds $\mathrm{pr}_{Y*}\pi_n = \mu$ as well. By replacing d by $\widehat{d}(x,y) := \max\{d(x,y),1\}$ we can assume that d is bounded. For estimating the Wasserstein distance of μ_n and μ using the coupling π_n, note that $\pi_n(M \times M) = \mu(M)$ for all $M \in \mathcal{M}_n$. It holds

$$\begin{aligned}\mathcal{W}_1(\mu_n,\mu) &\leq \int d\,d\pi_n \leq \sum_{M\in\mathcal{M}_n} \int d\,d\pi_n \leq \sum_{M\in\mathcal{M}_n} \mathrm{diam}(M)\pi_n(M\times M)\\ &\leq \sup_{M\in\mathcal{M}_n} \mathrm{diam}(M) \sum_{M\in\mathcal{M}_n} \mu(M) = \sup_{M\in\mathcal{M}_n} \mathrm{diam}(M) \to 0,\end{aligned}$$

which implies $\mu_n \rightharpoonup \mu$ by Theorem 1.9. □

The following proposition is (up to a few technicalities) a consequence of the isomorphism theorem for measures (Theorem A.19), which states that for any two continuous probabilities measures, there exists a bijection that pushes the first measures to the second. A detailed proof can be found in the appendix.

Proposition 1.19. *Let X be a Polish space, $\mathcal{M}$ be an at most countable partition of X consisting of Borel sets and $\mu \in \mathcal{P}(X)$ be continuous. Then there exists a Borel isomorphism $\Phi^{\mathcal{M}}_{\mu} : X \to X \times [0,1]$ such that for all $M \in \mathcal{M}$ it holds $(\Phi^{\mathcal{M}}_{\mu})_*(\mu|_M) = (\mu|_M) \otimes \lambda$.*

Proof. See appendix. □

Now we are ready to prove the main theorem of this section:

Theorem 1.20. *Let X, Y be standard Borel spaces, let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ be both continuous. Then couplings supported by the graph of a bijective function are dense in $\mathrm{Cpl}(\mu,\nu)$, i.e.*

$$\mathrm{Cpl}(\mu,\nu) = \overline{\mathrm{Cpl}_{00}(\mu,\nu)}.$$

Proof. By Proposition 1.5 the set $\mathrm{Cpl}(\mu,\nu)$ is closed, so it suffices to show that any $\pi \in \mathrm{Cpl}(\mu,\nu)$ can be approximated by a sequence $\pi_n \in \mathrm{Cpl}_{00}(\mu,\nu)$ w.r.t. weak convergence.

According to Theorem 1.12 there exists a coupling $\widehat{\pi} = (id,T)_*(\mu\otimes\lambda) \in \mathrm{Cpl}_{00}(\mu\otimes\lambda,\nu\otimes\lambda)$ such that

(i) $T_*(\mu\otimes\lambda) = \nu\otimes\lambda$,

(ii) $\mathrm{pr}_{XY*}\widehat{\pi} = \pi$.

Let $(\mathcal{A}_n)_{n\in\mathbb{N}}$ and $(\mathcal{B}_n)_{n\in\mathbb{N}}$ be sequences of partitions of X and Y consisting of countably many Borel sets and satisfying $\lim_{n\to\infty} ||\mathcal{A}_n|| = 0$ and $\lim_{n\to\infty} ||\mathcal{B}_n|| = 0$. According to Proposition 1.19, for any $n \in \mathbb{N}$ there exist bijections $\Phi_n : X \to X \times [0,1]$ and $\Psi_n : Y \to Y \times [0,1]$ such that

(iii) $\Phi_{n*}(\mu|_A) = (\mu|_A) \otimes \lambda$ for all $A \in \mathcal{A}_n$

(iv) $\Psi_{n*}(\nu|_B) = (\nu|_B) \otimes \lambda$ for all $B \in \mathcal{B}_n$

For $n \in \mathbb{N}$ define the mapping

$$T_n := \Psi_n^{-1} \circ T \circ \Phi_n : X \to Y.$$

It is easy to see that T_n is bijective and satisfies $T_{n*}\mu = \nu$. We need to check that $\pi_n := (\mathrm{id}, T_n)_*\mu \rightharpoonup \pi$. Note that $\mathcal{A}_n \otimes \mathcal{B}_n := \{A \times B : A \in \mathcal{A}_n, B \in \mathcal{B}_n\}$ are partitions of $X \times Y$ consisting of countably many Borel sets satisfying $\lim_{n\to\infty} ||\mathcal{A}_n \otimes \mathcal{B}_n|| = 0$. Hence, by Lemma 1.18 it suffices to show for all $n \in \mathbb{N}$:

$$\forall A \in \mathcal{A}_n \, \forall B \in \mathcal{B}_n : \, \pi_n(A \times B) = \pi(A \times B).$$

This is a consequence of the properties (i) to (iv) of the mappings Φ_n, Ψ_n and T:

$$\begin{aligned}
\pi_n(A \times B) &= \mu(A \cap T_n^{-1}(B)) = \mu|_A((\Phi_n^{-1} \circ T^{-1} \circ \Psi_n)(B)) \overset{(iii)}{=} (\mu|_A \otimes \lambda)(T^{-1}(\Psi_n(B))) \\
&= (\mu \otimes \lambda)(A \times [0,1] \cap T^{-1}(\Psi_n(B))) \overset{(i)}{=} (\nu \otimes \lambda)(T(A \times [0,1]) \cap \Psi_n(B)) \\
&= (\nu \otimes \lambda)(\Psi_n(\Psi_n^{-1}(T(A \times [0,1])) \cap B)) = \Psi_{n*}(\nu \otimes \lambda)(\Psi_n^{-1}(T(A \times [0,1])) \cap B) \\
&= \nu|_B(\Psi_n^{-1}(T(A \times [0,1]))) \overset{(iv)}{=} (\nu|_B \otimes \lambda)(T(A \times [0,1])) \\
&= (\nu \otimes \lambda)(T(A \times [0,1]) \cap B \times [0,1]) = (\mathrm{id}, T^{-1})_*(\nu \otimes \lambda)(A \times [0,1] \times B \times [0,1]) \\
&= \widehat{\pi}(A \times [0,1] \times B \times [0,1]) \overset{(ii)}{=} \pi(A \times B)
\end{aligned}$$

□

1.4 Discussion of the result

Remark 1.21. It is well known that $\mathrm{Cpl}_0(\mu,\nu)$ is dense in $\mathrm{Cpl}(\mu,\nu)$ if μ is continuous. We can prove this result with little effort using the tools that we have developed so far. The proof can be carried out exactly as the proof of Theorem 1.20 with one important exception: The existence of the bijections $\Psi_n : Y \to Y \times [0,1]$ that push ν to $\nu \otimes \lambda$ (and are compatible with the given partition) fails. However, if we just replace Ψ_n^{-1} by $\mathrm{pr}_Y : Y \times [0,1] \to Y$ in the definition of the mappings T_n, the only property of T_n that we loose is its injectivity. Hence, we have constructed a sequence of mappings T_n that push μ to ν s.t. $(\mathrm{id}, T_n)_*\mu$ converges to the given coupling π.

An immediate consequence of Theorem 1.20 is that the optimal transport problem yields the same value if we restrict ourselves to Monge couplings or even to Monge couplings, which are supported by the graph of a bijection:

Corollary 1.22. *Let $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ be continuous and $c : X \times Y \to \mathbb{R}$ be continuous and bounded. Then*

$$\inf_{\pi \in \mathrm{Cpl}(\mu,\nu)} \int c(x,y)\pi(dx,dy) = \inf_{\pi \in \mathrm{Cpl}_0(\mu,\nu)} \int c(x,y)\pi(dx,dy) = \inf_{\pi \in \mathrm{Cpl}_{00}(\mu,\nu)} \int c(x,y)\pi(dx,dy).$$

Proof. We can assume that $\inf_{\pi \in \mathrm{Cpl}(\mu,\nu)} \int c(x,y)\pi(dx,dy) < \infty$ (otherwise the claim is trivial) and, clearly, it suffices to prove

$$\inf_{\pi \in \mathrm{Cpl}_{00}(\mu,\nu)} \int c(x,y)\pi(dx,dy) \leq \inf_{\pi \in \mathrm{Cpl}(\mu,\nu)} \int c(x,y)\pi(dx,dy).$$

By Theorem 1.7 there exits $\pi \in \mathrm{Cpl}(\mu, \nu)$ s.t.

$$\int c d\pi = \inf_{\pi \in \mathrm{Cpl}(\mu,\nu)} \int c(x,y)\pi(dx,dy).$$

By Theorem 1.20 there exists a sequence $\pi_n \in \mathrm{Cpl}_{00}(\mu, \nu)$ s.t. $\pi_n \rightharpoonup \pi$ and by the definition of weak convergence we have

$$\lim_n \int c d\pi_n = \int c d\pi = \inf_{\pi \in \mathrm{Cpl}(\mu,\nu)} \int c(x,y)\pi(dx,dy). \qquad \square$$

Remark 1.23. Theorem 1.7 states that the infimum in $\inf\{\int c d\pi : \pi \in \mathrm{Cpl}(\mu,\nu)\}$ is attained. We want to discuss under which assumptions the infima in $\inf\{\int c d\pi : \pi \in \mathrm{Cpl}_0(\mu,\nu)\}$ and $\inf\{\int c d\pi : \pi \in \mathrm{Cpl}_{00}(\mu,\nu)\}$ are attained.

We first discuss an example, where the infimum is not attained that was given in [8, Section 1.4]. Let $X = Y = [-1,1] \times [0,1]$ and for $t \in [-1,1]$ denote $f_t : [0,1] \to [-1,1] \times [0,1] : y \mapsto (t,y)$. Consider the continuous measures $\mu := f_{0*}\lambda$ and $\nu := \frac{1}{2}(f_{-1*}\lambda + f_{1*}\lambda)$ and the cost function $c(x,y) = |x-y|$.

Since $\mathrm{dist}(\mathrm{supp}(\mu), \mathrm{supp}(\nu)) = 1$, every transportplan has at least cost 1, so $\inf\{\int c d\pi : \pi \in \mathrm{Cpl}(\mu,\nu)\} \geq 1$. The coupling $d\pi(x,y) = d\lambda(x) d\pi^x(y)$, where $\pi^x = \frac{1}{2}(\delta_{(0,1)} + \delta_{(0,-1)})$ shows that $\inf\{\int c d\pi : \pi \in \mathrm{Cpl}(\mu,\nu)\} = 1$ and Corollary 1.22 implies $\inf\{\int c d\pi : \pi \in \mathrm{Cpl}_0(\mu,\nu)\} = 1$.

Assume that $\pi = (\mathrm{id}, T)_*\mu$ is a Monge coupling with cost 1, i.e. $\int |x - T(x)| d\mu = 1$. Due to $\mathrm{dist}(\mathrm{supp}(\mu), \mathrm{supp}(\nu)) = 1$ it holds $|x - T(x)| \geq 1$, hence $|x - T(x)| = 1$ for μ-almost all x, which implies that $T(0,y) \in \{(-1,y),(1,y)\}$ for λ-almost all y. Clearly, the sets $M_\pm := \{(0,y) : T(0,y) = (\pm 1, y)\}$ are disjoint, their union has full measures and it holds $T_*\mu(\{-1\} \times M_- \cup \{1\} \times M_+) = 1$. However, since $\nu(\{-1\} \times A) = \nu(\{1\} \times A)$ for all $A \subseteq [0,1]$, it holds $\nu(\{-1\} \times M_- \cup \{1\} \times M_+) = \frac{1}{2}$, which is a contradiction to $T_*\mu = \nu$.

Since $\mathrm{Cpl}_{00}(\mu,\nu) \subseteq \mathrm{Cpl}_0(\mu,\nu)$ and $\inf\{\int c d\pi : \pi \in \mathrm{Cpl}_{00}(\mu,\nu)\} = \inf\{\int c d\pi : \pi \in \mathrm{Cpl}_0(\mu,\nu)\}$ by Corollary 1.22, this example implies that in general the infimum in $\inf\{\int c d\pi : \pi \in \mathrm{Cpl}_{00}(\mu,\nu)\}$ is not attained, as well.

However, Brenier's Theorem (see e.g. [9, Theorem 2.12]) states that for $X = \mathbb{R}^n$ and μ absolutely continuous w.r.t. Lebesgue and $c(x,y) = |x-y|^2$ there exists a unique minimizer for $\inf\{\int c d\pi : \pi \in \mathrm{Cpl}(\mu,\nu)\}$ and that this minimizer is a Monge coupling.

This has the following consequence in the case that μ and ν are both absolutely continuous: Denote $e(x,y) := (y,x)$. Since the cost is symmetric, π is the minimizer in the transport problem from μ to ν if and only if $e_*\pi$ is the minimizer in the transport problem from ν to μ. Hence, for the minimizer π of $\inf\{\int c d\pi : \pi \in \mathrm{Cpl}(\mu,\nu)\}$ both π and $e_*\pi$ are Monge couplings, i.e. $\pi \in \mathrm{Cpl}_{00}(\mu,\nu)$. Therefore, the infimum in $\inf\{\int c d\pi : \pi \in \mathrm{Cpl}_{00}(\mu,\nu)\}$ is attained in this case.

In particular, one can restrict to couplings supported on the graph of bijections, when calculating the Wasserstein distance of two probability measures:

Corollary 1.24. *Let (X,d) a Polish space and $p \in [1,\infty)$. Then*

$$\mathcal{W}_p(\mu,\nu) = \inf\left\{\int d(x,y)^p : \pi \in \mathrm{Cpl}_{00}(\mu,\nu)\right\}^{1/p}$$

Proof. This is a immediate consequence of Corollary 1.22. □

In the remaining part of this section, we show that the assumption that μ and ν are continuous is (except for a trivial case) necessary. To avoid tedious case distinctions, we assume for the rest of this section that X and Y are both uncountable (and have therefore cardinality continuum, cf. Theorem A.6). First, we give a necessary and sufficient condition for the existence of at least one coupling that is supported by the of graph of a bijection.

Proposition 1.25. *The following are equivalent for $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$:*

(i) There exists a Borel isomorphism $f : X \to Y$ s.t. $f_*\mu = \nu$.

(ii) There exists injective sequences $(x_n)_{n\in\mathbb{N}}$ in X and $(y_n)_{n\in\mathbb{N}}$ in Y s.t. $\mu(\{x_n\}) = \nu(\{y_n\})$ for all $n \in \mathbb{N}$ and for all $x \in X' := X \setminus \{x_n : n \in \mathbb{N}\}$ and all $y \in Y' := Y \setminus \{y_n : n \in \mathbb{N}\}$ it holds $\mu(\{x\}) = \nu(\{y\}) = 0$.

Proof. (i) $\implies$ (ii): Since μ has at most countably many atoms there exists the desired sequence $(x_n)_{n\in\mathbb{N}}$. Since f is bijective, the sequence defined by $y_n := f(x_n)$ is again injective and has the property $\nu(\{y_n\}) = \mu(f^{-1}(\{y_n\})) = \mu(\{x_n\})$. If $y \in Y$', it holds $f^{-1}(y) \in X'$ and therefore $\nu(\{y\}) = \mu(f^{-1}(\{y\})) = 0$.

(ii) $\implies$ (i): Clearly, $\mu|_{X'}$ and $\nu|_{Y'}$ are continuous measures on the standard Borel spaces X' and Y'. By Theorem A.19 there exits a Borel isomorpism $g : X' \to Y'$ s.t. $g_*\mu|_{X'} = \nu|_{Y'}$. It is easy to see that

$$f : X \to Y : x \mapsto \begin{cases} g(x) & x \in X' \\ y_n & x = x_n \end{cases}$$

has the desired properties. □

Using this, we achieve a necessary and sufficient condition on the marginals for the denseness of couplings supported by the graph a bijection.

Theorem 1.26. *Let X, Y be Polish spaces and $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$. Then $\mathrm{Cpl}_{00}(\mu,\nu)$ is dense in $\mathrm{Cpl}(\mu,\nu)$ if and only if one of the followings statements is true:*

(i) μ and ν are both continuous measures

(ii) μ and ν are both Dirac measures.

Proof. Assume that there are μ and ν, which do not satisfy (i) or (ii), s.t. $\mathrm{Cpl}_{00}(\mu,\nu)$ is dense in $\mathrm{Cpl}(\mu,\nu)$. In particular, there exists a Borel isomorphism pushing μ to ν. Hence, by Proposition 1.25 there exist sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ s.t. $\mu(\{x_n\}) = \nu(\{y_n\})$ for all $n \in \mathbb{N}$ and for all $x \in X \setminus \{x_n : n \in \mathbb{N}\}$ and all $y \in Y \setminus \{y_n : n \in \mathbb{N}\}$ it holds $\mu(\{x\}) = \nu(\{y\}) = 0$. We assume wlog that $\mu(\{x_n\})$ is decreasing and set $\alpha := \mu(\{x_1\})$. Since μ is not continuous, we have $\alpha > 0$ and therefore $\mu(\{x_n : n \in \mathbb{N}\}) < \infty$ implies that $k := \max\{n : \mu(\{x_n\}) = \alpha\}$ is finite. We distinguish two cases:

Case 1: $k\alpha = 1$. Then $\mu = \frac{1}{k}\sum_{i=1}^k \delta_{x_i}$ and $\nu = \frac{1}{k}\sum_{i=1}^k \delta_{y_i}$, where $k \geq 2$ because μ and ν are assumed not to be Dirac measures. Then $\mathrm{Cpl}_{00}(\mu,\nu)$ is finite and therefore closed, but it does not contain $\mu \otimes \nu$.

Case 2: $k\alpha < 1$. The set $A := \{(x_i, y_j) : i, j \in \{1, \dots, k\}\}$ is finite and therefore closed. Any $\pi \in \mathrm{Cpl}_{00}(\mu, \nu)$ satisfies $\pi(A) = k\alpha$ because any bijection pushing μ to ν is in particular a bijection of the atoms of μ with mass α to the atoms of ν with mass α. Since mass cannot escape from closed sets in weak limits, any coupling π in the weak closure of $\mathrm{Cpl}_{00}(\mu, \nu)$ satisfies $\pi(A) \geq k\alpha$. However, $(\mu \otimes \nu)(A) = (k\alpha)^2 < k\alpha$. □

To close this section we give an explicit example of a sequence of couplings supported on graphs of bijections that approximate a given coupling.

Expample 1.27. Let $X = Y = [0, 1]$, $\mu = \nu = \lambda$ and $\pi = \lambda^2$. Following the construction in the proof of Theorem 1.20 we need a bijection $T : [0, 1]^2 \to [0, 1]^2$ such that $T_*(\mu \otimes \lambda) = \nu \otimes \lambda$ and $\mathrm{pr}_{XY*}(\mathrm{id}, T)_*(\mu \otimes \lambda) = \pi$. A suitable choice is $T(x, u) := (u, x)$, because $T_*(\mu \otimes \lambda) = T_*\lambda^2 = \lambda^2$ and $\mathrm{pr}_{XY} \circ (\mathrm{id}, T)(x, u) = (x, u)$, so $\mathrm{pr}_{XY*}(\mathrm{id}, T)_*(\mu \otimes \lambda) = \mathrm{pr}_{XY*}(\mathrm{id}, T)_*\lambda^2 = \lambda^2 = \pi$.

Moreover, we need a sequence $(\mathcal{M}_n)_n$ of partitions of $[0, 1]$ (having the properties stated in Proposition 1.17) and bijections $\Phi_n : [0, 1] \to [0, 1]^2$ that $\Phi_{n*}(\lambda|_M) = \lambda|_M \otimes \lambda$ for all $M \in \mathcal{M}_n$. We choose the partitions

$$\mathcal{M}_n := \{[0, 2^{-n}], \dots, (k \cdot 2^{-n}, (k + 1) \cdot 2^{-n}], \dots, ((2^n - 1) \cdot 2^{-n}, 1]\}.$$

We will define the bijections Φ_n by using dyadic expansions, which are uniquely defined up to λ-nullsets, which we will ignore in this example. Hence $0, x_1x_2x_3\dots$ denote the number $x = \sum_{i=1}^{\infty} x_i 2^{-i}$, where x_i will always be elements of $\{0, 1\}$.

Clearly, the mapping

$$\Phi_n : \begin{cases} [0, 1] \to [0, 1]^2 \\ 0, x_1x_2x_3\dots \mapsto (0, x_1 \dots x_n x_{2n+1} \dots x_{3n} x_{4n+1} \dots ; 0, x_{n+1} \dots x_{2n} x_{3n+1} \dots x_{4n} x_{5n+1} \dots) \end{cases}$$

is a bijection with inverse

$$\Phi_n^{-1} : \begin{cases} [0, 1]^2 \to [0, 1] \\ (0, x_1x_2x_3\dots ; 0, y_1y_2y_3\dots) \mapsto 0, x_1 \dots x_n y_1 \dots y_n x_{n+1} \dots x_{2n} y_{n+1} \dots y_{2n} x_{3n} \dots \end{cases}$$

and it is straight forward to check that Φ_n pushes $\lambda|_{[k2^{-n}, (k+1)2^{-n}]}$ to $\lambda|_{[k2^{-n}, (k+1)2^{-n}]} \otimes \lambda$ for all $k < 2^n$.

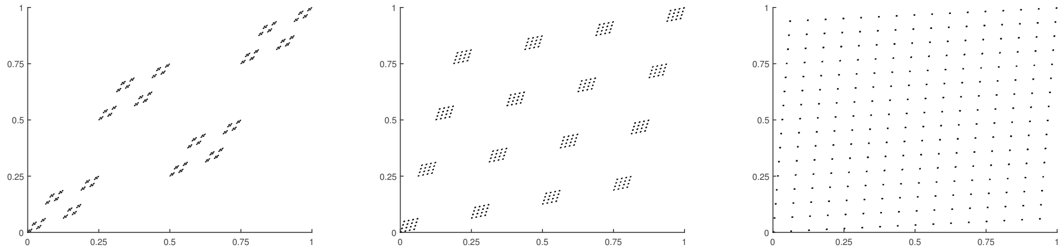
In the proof of Theorem 1.20 we see that the mappings

$$T_n := \Phi_n^{-1} \circ T \circ \Phi_n$$

have the desired properties. It is easy to calculate this composition explicitly:

$$T_n : \begin{cases} [0, 1] \to [0, 1] \\ 0, x_1x_2x_3 \mapsto 0, x_{n+1} \dots x_{2n} x_1 \dots x_n x_{3n+1} \dots x_{4n} x_{2n+1} \dots x_{3n} x_{5n+1} \dots \end{cases}$$

In other words, the mapping T_n acts on some $x \in [0, 1]$ as follows: For all $k > 0$ swap the $(2k - 1)$-th and the $(2k)$-th digits of x in its expansion w.r.t. the base 2^n.



(a) $n = 1$ (b) $n = 2$ (c) $n = 4$

Figure 1.1: Plots of the functions T_n.

1.5 Results in probabilistic notation

In the section we state the important results from this chapter in probabilistic notation. We start this section with elementary observations:

$\mathrm{Cpl}(\mu, \nu)$ is set of joint laws of random variables $\mathsf{X} \sim \mu$ and $\mathsf{Y} \sim \nu$: Given $\pi \in \mathrm{Cpl}(\mu, \nu)$, the random variables pr_X and pr_Y on the probability space $(X \times Y, \mathcal{B}_X \otimes \mathcal{B}_Y, \pi)$ have the desired properties. Conversely, given some Borel probabilty space $(\Omega, \mathcal{F}, \mathbb{P})$, an X-valued random variable $\mathsf{X} \sim \mu$ and a Y-valued random variable $\mathsf{Y} \sim \nu$ it holds, $\mathcal{L}(\mathsf{X}, \mathsf{Y}) = (\mathsf{X}, \mathsf{Y})_* \mathbb{P} \in \mathrm{Cpl}(\mu, \nu)$.

The Monge couplings correspond to pairs of random variables of the form $(\mathsf{X}, F(\mathsf{X}))$.

The representation of couplings from Section 1.2 written in probabilistic notation looks as follows:

Corollary 1.28. *Let X be an X-valued and Y be an Y-valued random variable. Then there exists a bimeasurable mapping $F : X \times [0, 1] \to Y \times [0, 1]$ such that for any uniform random variable U independent of X and any uniform random variable V independent of Y it holds:*

(i) $F(\mathsf{X}, \mathsf{U}) \sim (\mathsf{Y}, \mathsf{V})$

(ii) $(\mathsf{X}, F_1(\mathsf{X}, \mathsf{U})) \sim (\mathsf{X}, \mathsf{Y})$,

where F_1 denotes the first comment of F, i.e. $\mathrm{pr}_Y \circ F$.

Proof. This is a consequence of Theorem 1.12. □

The version of this representation for kernels (Theorem 1.15) corresponds to a version of the previous corollary with conditional probabilities:

Corollary 1.29. *Let X be a X-valued, Y be a Y-valued and Z be a Z-valued random variable. Then there exists a measurable bijection $F : Z \times X \times [0, 1] \to Y \times [0, 1]$ such that for any uniform random variable U conditionally independent of X given Z and any uniform random variable V conditionally independent of Y given Z it holds a.s.:*

(i) $\mathbb{P}(F(\mathsf{Z}, \mathsf{X}, \mathsf{U}) \in B | \mathsf{Z}) = \mathbb{P}((\mathsf{Y}, \mathsf{V}) \in B | \mathsf{Z}) \qquad \forall B \subseteq Y \times [0, 1]$ *Borel*

(ii) $\mathbb{P}((\mathsf{X}, F_1(\mathsf{Z},\mathsf{X},\mathsf{U})) \in B|\mathsf{Z}) = \mathbb{P}((\mathsf{X},\mathsf{Y}) \in B|\mathsf{Z}) \qquad \forall B \subseteq X \times Y$ *Borel*

Proof. By Theorem A.27 there exists a kernel π from Z to $X \times Y$ such that $\mathcal{L}(\mathsf{Z})$-a.s. it holds

$$\mathbb{P}((\mathsf{X},\mathsf{Y}) \in A|\mathsf{Z} = z) = \pi^z(A) \qquad A \subseteq X \times Y \text{ Borel},$$

i.e. $\pi^z = \mathcal{L}((\mathsf{X},\mathsf{Y})|\mathsf{Z} = z)$. Denote $\mu^z := \mathrm{pr}_{X*}\pi^x = \mathcal{L}(\mathsf{X}|\mathsf{Z} = z)$ and $\nu^z := \mathrm{pr}_{Y*}\pi^x = \mathcal{L}(\mathsf{Y}|\mathsf{Z} = z)$. By Theorem 1.15 there exists a Borel measurable mapping $F : Z \times X \times [0,1] \to Y \times [0,1]$ s.t. for all $z \in Z$ the mappings $F(z,\cdot) : X \times [0,1] \to Y \times [0,1] : (x,u) \mapsto F(z,x,u)$ are Borel isomorphisms satisfying

(a) $F(z,\cdot)_*(\mu^z \otimes \lambda) = \nu^z \otimes \lambda$

(b) $\mathrm{pr}_{XY*}(id, F(z,\cdot))_*(\mu^z \otimes \lambda) = \pi^z$.

It is easy to see, that (a) implies (i) and (b) implies (ii). □

Convergence of random variables in distribution is equivalent to weak convergence of their laws. Therefore, the denseness result from Section 1.3 corresponds to the following result for random variables:

Corollary 1.30. *Let* X *be an* X*-valued and* Y *be an* Y*-valued random variable. Then there exist measurable bijections* $F_n : X \to Y$ *such that* $F(\mathsf{X}) \sim \mathsf{Y}$ *and* $(\mathsf{X}, F_n(\mathsf{X}))$ *converges to* (X,Y) *in distribution.*

Proof. This is an immediate consequence of Theorem 1.20. □

Chapter 2

The time dependent case

The aim of this chapter is to extend the results to finitely many time steps. For clearly, we could just apply our results from Chapter 1 to measures on the path space, but we aim a construction that respects the arrow of time: Given measures μ and ν on path spaces of processes with finitely many times steps (say $\mathbb{R}^N$), we want to approximate a given coupling with Monge couplings supported on the graph of bijections $F : \mathbb{R}^N \to \mathbb{R}^N$, which satisfy the following property: For $t \leq N$, the t-th component of $F(x_1, \ldots, x_N)$ does only depend on $(x_1, \ldots, x_t)$ and the t-th component of $F^{-1}(y_1, \ldots, y_N)$ does only depend on $(y_1, \ldots, y_t)$. Such mappings are called biadapted, see Definitions 2.5 and 2.14 for details.

The main result of this thesis is that (under certain regularity assumptions on the marginals) such an approximation is possible if and only if the coupling is a so-called bicausal coupling, see Theorem 2.26.

In the first two sections of this chapter, we explain what causal and bicausal couplings are. Afterwards, we extend our techniques from Chapter 1 to the time dependent case, which enables to prove our main Theorem 2.26 in an analogue way as we have proven the denseness of couplings supported on the graph of bijections in Theorem 1.20.

Before we start, we need to introduce some notation for this chapter: $N \in \mathbb{N}$ will always be the number of time steps that we consider. $X_1 \ldots, X_N$, $Y_1, \ldots, Y_N$ are always Polish spaces.

$\prod_{i=1}^N X_i$ will be the path space of the first process, whose law will be denoted by μ and $\prod_{i=1}^N Y_i$ will be the path space of the second process, whose law will be denoted by ν.

For $1 \leq s < t \leq N$ we introduce the abbreviation $X_{s:t} := \prod_{i=s}^t X_i$. We use the same abbreviation for elements of $X_{s:t}$, i.e. $(x_s, x_{s+1}, \ldots, x_t) =: x_{s:t}$, and for subsets, i.e. $A_s \times A_{s+1} \times \cdots \times A_t =: A_{s:t}$ for $A_i \subseteq X_i$. We use X as a shorthand for $X_{1:N}$. For the Y-component we use analogous notations.

For $t \leq N$ define $\mathcal{F}_t^X$ as the σ-algebra on $X \times Y$ generated by the projections $X \times Y \to X_{1:t}$: $(x, y) \mapsto x_{1:t}$ (and $\mathcal{F}_t^Y$ respectively).

Let $\mu \in \mathcal{P}(X)$. Then by Theorem A.24 we can decompose μ as $\mu(dx) = \mu_1(dx_1)\mu^{x_1}(dx_{2:N})$, where $\mu_1 \in \mathcal{P}(X_1)$ and $x_1 \mapsto \mu^{x_1}$ is a kernel from X_1 to $X_{2:N}$. Iterating this yields

$$\mu(dx) = \mu_1(dx_1)\mu^{x_1}(dx_2)\cdots\mu^{x_{1:N-1}}(dx_N),$$

i.e. $\mu_1 \in \mathcal{P}(X_1)$ and for all $t < N$ there are kernels $x_{1:t} \mapsto \mu^{x_{1:t}}$ from $X_{1:t}$ to X_{t+1}.

2.1 Causal couplings

We start with a quick motivation for causal transport. For $\varepsilon \geq 0$ consider the measures $\mu_\varepsilon := \frac{1}{2}(\delta_{(\varepsilon,1)} + \delta_{(-\varepsilon,-1)}) \in \mathcal{P}(\mathbb{R}^2)$. It is easy to see that $\mathcal{W}_1(\mu_\varepsilon, \mu_0) = \varepsilon$, so μ_ε converges to μ_0 w.r.t. Wasserstein.

However, if we consider μ_ε as the law of a real-valued stochastic process with two timesteps, the natures of μ_0 and μ_ε for $\varepsilon > 0$ are totally different. For the case $\varepsilon = 0$ the corresponding process is a martingale that starts at 0 and is $+1$ or -1 in the second step, whereas for $\varepsilon > 0$ the corresponding process is deterministic. Hence, Wasserstein convergence of the laws, is a "bad" notion of convergence for laws of stochastic processes.

Now the idea is to restrict in the definition of the Wasserstein distance to couplings that "respect the time structure" in order to get a metric that induces a stronger topology that reflects properties of stochastic processes better.

The following definition clarifies what is meant by coupling that "respect the time structure".

Definition 2.1. Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. A coupling $\pi \in \mathrm{Cpl}(\mu, \nu)$ is called causal if for any $t < N$ and $B \in \mathcal{F}_t^Y$ the mapping $X \ni x \mapsto \pi^x(B)$ is $\mathcal{F}_t^X$-measurable. We denote the set of causal couplings between μ and ν as $\mathrm{Cpl}_c(\mu, \nu)$.

This condition is obviously equivalent to saying that for any $t < N$ the mapping $X \to \mathcal{P}(Y_{1:t}) : x \mapsto \mathrm{pr}_{Y_{1:t}*}\pi$ is measurable. The following proposition gives equivalent conditions for causality:

Proposition 2.2 ([2, Proposition 2.3])**.** *For $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and $\pi \in \mathcal{P}(X \times Y)$ the following are equivalent:*

(i) $\pi \in \mathrm{Cpl}_c(\mu, \nu)$

(ii) $\pi \in \mathrm{Cpl}(\mu, \nu)$ *and for all* $t < N$: $\mathcal{F}_N^X$ *is conditionally independent of* $\mathcal{F}_t^Y$ *given* $\mathcal{F}_t^X$

(iii) When decomposing

$$\pi(dx, dy) = \pi_1(dx_1, dy_1)\pi^{x_1,y_1}(dx_2, dy_2)\cdots\pi^{x_{1:N-1},y_{1:N-1}}(dx_N, dy_N)$$

it holds

(a) $\pi_1 \in \mathrm{Cpl}(\mathrm{pr}_{1*}\mu, \mathrm{pr}_{1*}\nu)$

(b) for all $t < N$ *and* π*-almost all* $x_{1:t}, y_{1:t}$: $\mathrm{pr}_{X*}\pi^{x_{1:t},y_{1:t}} = \mu^{x_{1:t}}$

(c) for all $t < N$ *and* ν*-almost all* $y_{1:t}$: $\pi^{y_{1:t}}(dy_{t+1}) = \nu^{y_{1:t}}(dy_{t+1})$.

(iv) $\pi \in \mathrm{Cpl}(\mu, \nu)$ *and for all* $t \leq N$, $h_t \in C_b(Y_{1:t})$ *and* $g \in C_b(X)$ *it holds*

$$\int h_t(y_{1:t})\Big[g(x_{1:N}) - \int g(x_{1:t}, \bar{x}_{t+1:N})\mu^{x_{1:t}}(d\bar{x}_{t+1:N})\Big]d\pi = 0.$$

Proof. (i) $\implies$ (ii): Let $\pi \in \mathrm{Cpl}_c(\mu,\nu)$ and fix $t < N$. By Lemma A.43 this implies $\pi^{x_{1:N}}(B) = \pi^{x_{1:t}}(B)$ a.s. for all $B \in \mathcal{F}_t^Y$. Since $\mathcal{F}_t^X \subseteq \mathcal{F}_N^X$ this implies for all $B \in \mathcal{F}_t^Y$ that[1] $\pi^{\mathcal{F}_t^X,\mathcal{F}_N^X}(B) = \pi^{\mathcal{F}_N^X}(B) = \pi^{\mathcal{F}_t^X}(B)$. Now Proposition A.30 yields the conditional independence of $\mathcal{F}_N^X$ and $\mathcal{F}_t^Y$ given $\mathcal{F}_t^X$.

(ii) $\implies$ (i): It is easy to see that one can read the proof of (i) $\implies$ (ii) in the converse direction as well.

(ii) $\implies$ (iii): Fix $t < N$. It is clear that $\pi \in \mathrm{Cpl}(\mu,\nu)$ implies (a) and (c). Since $\mathcal{F}_t^Y$ is conditionally independent of $\mathcal{F}_t^Y$ given $\mathcal{F}_t^X$ Proposition A.30 (and Remark A.31) yields $\pi^{\mathcal{F}_t^X,\mathcal{F}_t^Y}(B) = \pi^{\mathcal{F}_t^X}(B)$ for all $B \in \mathcal{F}_N^X$. This implies $\mathrm{pr}_{X*}(\pi^{x_{1:t},y_{1:t}}) = \mathrm{pr}_{X*}(\pi^{x_{1:t}}) = \mu^{x_{1:t}}$ a.s.

(iii) $\implies$ (ii): Clearly, (a) and (b) imply $\mathrm{pr}_{X*}\pi = \mu$, whereas (a) and (c) imply $\mathrm{pr}_{Y*}\pi = \nu$. For $t < N$ condition (b) implies $\pi^{\mathcal{F}_t^X,\mathcal{F}_t^Y}(B) = \pi^{\mathcal{F}_t^X}(B)$ for all $B \in \mathcal{F}_N^X$, which implies the conditional independence of $\mathcal{F}_N^X$ and $\mathcal{F}_t^Y$ given $\mathcal{F}_t^X$ by Proposition A.30.

(i) $\iff$ (iv)[2]: Let $\pi \in \mathrm{Cpl}(\mu,\nu)$. Clearly $\pi \in \mathrm{Cpl}_c(\mu,\nu)$ if and only if for all $t < N$ the mapping $x \mapsto \mathrm{pr}_{Y_{1:t}*}\pi : X \to \mathcal{P}(Y_{1:t})$ is $\mathcal{F}_t^X$-measurable, which is by Proposition A.17 equivalent to the $\mathcal{F}_t^X$-measurability of the functions $\phi_h(x_{1:N}) := \int h_t(y_{1:t})d\pi^{x_{1:N}}$, where $h_t \in C_b(Y_{1:t})$.

Fix some $h_t \in C_b(Y_{1:t})$. Observe that $\phi_h = \mathbb{E}_\pi^{\mathcal{F}_N^X}[h_t]$. Therefore, ϕ_h is $\mathcal{F}_t^X$-measurable if and only if

$$(\mathrm{id} - \mathbb{E}_\pi^{\mathcal{F}_t^X})\mathbb{E}_\pi^{\mathcal{F}_N^X} h_t = 0.$$

Since this is clearly $\mathcal{F}_N^X$-measurable and by the density of $C_b(X)$ in $L^2(X,\mathcal{F}_N^X,\mu)$ this is equivalent to

$$\forall g \in C_b(X) : \mathbb{E}_\mu[g(\mathrm{id} - \mathbb{E}_\pi^{\mathcal{F}_t^X})\mathbb{E}_\pi^{\mathcal{F}_N^X} h_t] = 0.$$

Since $g(\mathrm{id} - \mathbb{E}_\pi^{\mathcal{F}_t^X})\mathbb{E}_\pi^{\mathcal{F}_N^X} h_t$ is F_N^X-measurable and $\mu = \mathrm{pr}_{X*}\pi$ we can replace the $\mathbb{E}_\mu$ by $\mathbb{E}_\pi$. Since $\mathcal{F}_t^X$ and $\mathcal{F}_N^X$ are orthogonal projections, $(\mathrm{id} - \mathbb{E}_\pi^{\mathcal{F}_t^X})\mathbb{E}_\pi^{\mathcal{F}_N^X}$ is selfadjoint, so the latter is equivalent to

$$\forall g \in C_b(X) : \mathbb{E}_\pi[((\mathrm{id} - \mathbb{E}_\pi^{\mathcal{F}_t^X})\mathbb{E}_\pi^{\mathcal{F}_N^X} g)h_t] = 0.$$

Since g is $\mathcal{F}_N^X$-measurable, $g = \mathbb{E}_\pi^{\mathcal{F}_N^X} g$, so the latter expression is exactly the integral in condition (iv). ☐

We give examples of causal couplings. First we return to the measure μ_ε from the motivating example and consider the optimal coupling between μ_ε and μ_0 for the Euclidean cost.

Expample 2.3. For $\varepsilon > 0$ consider the measures

$$\mu = \frac{1}{2}\left(\delta_{(\varepsilon,1)} + \delta_{(-\varepsilon,-1)}\right) \qquad \nu = \frac{1}{2}\left(\delta_{(0,1)} + \delta_{(0,-1)}\right) \qquad \pi = \frac{1}{2}\left(\delta_{(\varepsilon,1,0,1)} + \delta_{(-\varepsilon,-1,0,-1)}\right).$$

[1] the following notations mean conditional probabilities of the measure π w.r.t. the σ-fields denoted in the upper indizies

[2] This part of the proof uses many properties of condtional expectations stated in Theorem A.26, without explicitly referring to them.

Clearly, $\pi \in \mathrm{Cpl}(\mu, \nu)$. If we decompose the measures

$$\mu_1 = \frac{1}{2}\left(\delta_\varepsilon + \delta_{-\varepsilon}\right) \qquad \mu^{\pm\varepsilon} = \delta_{\pm 1}$$
$$\nu_1 = \delta_0 \qquad \nu^0 = \frac{1}{2}\left(\delta_1 + \delta_{-1}\right)$$
$$\pi_1 = \frac{1}{2}\left(\delta_{(\varepsilon,0)} + \delta_{(-\varepsilon,0)}\right) \qquad \pi^{\pm\varepsilon,0} = \delta_{(\pm 1,\pm 1)},$$

we see that $\pi \in \mathrm{Cpl}_c(\mu, \nu)$. However, if we exchange the roles of μ and ν, we obtain a coupling that is not causal: $e_*\pi \in \mathrm{Cpl}(\nu, \mu) \setminus \mathrm{Cpl}_c(\nu, \mu)$.

Expample 2.4. Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ be arbitrary. We show that the product measure $\pi := \mu \otimes \nu$ is causal between μ and ν.

To keep the notation simple, we just prove the case $N = 2$. We consider the decompositions

$$\mu(dx) = \mu_1(dx_1)\mu^{x_1}(dx_2), \quad \nu(dy) = \nu_1(dy_1)\mu^{y_1}(dy_2), \quad \pi(dx, dy) = \pi_1(dx_1, dy_1)\pi^{x_1,y_1}(dx_2, dy_2).$$

It is clear that $\pi_1 = \mu_1 \otimes \nu_1$. Therefore, by Proposition 2.2 it suffices to show

$$\pi^{x_1,\dots,x_t,y_1,\dots,y_t} = \mu^{x_1,\dots,x_t} \otimes \nu^{y_1,\dots,y_t} \qquad \text{a.s.}$$

For all $f : X \times Y \to \mathbb{R}$ bounded and measurable it holds

$$\begin{aligned}\int f d\pi &= \int f(x_1, x_2, y_1, y_2) d\mu(x_1, x_2) d\nu(y_1, y_2)\\ &= \int f(x_1, x_2, y_1, y_2) d\mu^{x_1}(x_2) d\mu_1(x_1) d\nu^{y_1}(y_2) d\nu_1(y_1)\\ &= \int f(x_1, x_2, y_1, y_2) \underbrace{d\mu^{x_1}(x_2) d\nu^{y_1}(y_2)}_{= d\mu^{x_1}\otimes\nu^{y_1}(x_2,y_2)} \underbrace{d\mu_1(x_1) d\nu_1(y_1)}_{= d\pi_1(x_1,y_1)},\end{aligned}$$

which implies $d\pi^{x_1 y_1} = d\mu^{x_1} \otimes d\nu^{y_1}$ a.s. by the uniqueness of the disintegration.

We have introduced a notion of couplings that respect the time structure. For mappings it is much simpler to find an appropriate definition:

Definition 2.5. A map $T : X \to Y$ is called adaped if for all $t \leq N$ there exits measureable functions $T_t : X_{1:t} \to Y_t$ such that

$$T(x_{1:N}) = (T_1(x), T_2(x_{1:2}), \dots, T_N(x_{1:N})).$$

The term adapted is due to the fact that T is adapted if and only if T_t is $\mathcal{F}^X_t$-measurable for all $t \leq N$. The following lemma states that these two notions of "respecting time struce" fit together, namely a Monge coupling is causal if and only if it is supported by an adapted mapping.

Lemma 2.6.
$$\mathrm{Cpl}_c(\mu, \nu) \cap \mathrm{Cpl}_0(\mu, \nu) = \{(id, T)_*\mu : T \text{ adapted }, T_*\mu = \nu\}$$

Proof. Let $d\pi = d\delta_{T(x)} d\mu(x)$, where $T_*\mu = \nu$, be causal and fix $t \leq N$. Then for all $B \in \mathcal{F}^Y_t$, the mapping

$$\phi_B : x \mapsto \delta_{T(x)}(B)$$

is $\mathcal{F}_t^X$ measurable. Since $T^{-1}(B) = \phi_B^{-1}(\{1\})$, this implies that T is $\mathcal{F}_t^X$-$\mathcal{F}_t^Y$-measurable. Hence, $\mathrm{pr}_{Y_{1:t}} \circ T$ is $\mathcal{F}_t^X$-measurable. By Lemma A.43 this implies that there is a measurable function $\widetilde{T}_t : X_{1:t} \to Y_{1:t}$ s.t. $\mathrm{pr}_{Y_{1:t}} \circ T = \widetilde{T}_t \circ \mathrm{pr}_{X_{1:t}}$. Set $T_t := \mathrm{pr}_{Y_t} \circ \widetilde{T}_t$.

Conversely, let $d\pi = d\delta_{T(x)} d\mu(x)$, where T is adapted and satisfies $T_*\mu = \nu$. For $t < N$ and $B \in \mathcal{F}_t^Y$ we need to show that ϕ_B is $\mathcal{F}_t^X$-measurable. Since ϕ_B only takes the values 0 and 1, it suffices to show that $\phi_B^{-1}(\{1\}) \in \mathcal{F}_t^X$. Indeed, $B = \mathrm{pr}_{Y_{1:t}}^{-1}(B')$ for some measurable $B' \subseteq Y_{1:t}$ and $\phi_B^{-1}(\{1\}) = T^{-1}(B) = (\mathrm{pr}_{Y_{1:t}} \circ T)^{-1}(B') \in \mathcal{F}_t^X$, since $\mathrm{pr}_{Y_{1:t}} \circ T$ is $\mathcal{F}_t^X$-measurable. □

The following proposition is the key argument for the existence of minimizes in the socalled causal transport problem.

Proposition 2.7 ([3, Theorem 3.1])**.** $\mathrm{Cpl}_c(\mu, \nu)$ *is weakly compact.*

Proof. Denote $\mathcal{T}_X$ the topology on X and $\mathcal{T}_Y$ the topology on Y. The mappings

$$\begin{gathered} X_1 \ni x_1 \mapsto \mu^{x_1} \in \mathcal{P}(X_{2:N}) \\ X_{1:2} \ni x_{1:2} \mapsto \mu^{x_{1:2}} \in \mathcal{P}(X_{3:N}) \\ \vdots \\ X_{1:N-1} \ni (x_{1:N-1}) \mapsto \mu^{x_{1:N-1}} \in \mathcal{P}(X_N) \end{gathered}$$

are all Borel. By Theorem A.9 there exists a Polish topology $\mathcal{T}_X' \supseteq \mathcal{T}_X$ s.t. these mappings are continuous from $(X, \mathcal{T}_X')$ to $\mathcal{P}(X_{t:N})$ and $(X, \mathcal{T}_X')$ has the same Borel sets as $(X, \mathcal{T}_X)$.

Denote $\mathcal{V}_1$ the topology on $\mathcal{P}(X \times Y)$ generated by testing against $\mathcal{T}_X \times \mathcal{T}_Y$-continuous and bounded functions and $\mathcal{V}_2$ the topology on $\mathcal{P}(X \times Y)$ generated by testing against $\mathcal{T}_X' \times \mathcal{T}_Y$-continuous and bounded functions. Since the functions appearing in Proposition 2.2 (iii) are $\mathcal{V}_2$-continuous, $\mathrm{Cpl}_c(\mu, \nu)$ is $\mathcal{V}_2$-closed. Since $\mathrm{Cpl}_c(\mu, \nu)$ is tight, it is $\mathcal{V}_2$-compact by Prokhorov's Theorem A.21 and therefore in particular $\mathcal{V}_1$-compact. □

As for the ordinary optimal transport problem, we can prove the existence of minimizes for the so-called causal transport problem:

Theorem 2.8. *Let $c : X \times Y \to \mathbb{R}$ be lower semi-continuous and bounded from below. If*

$$\inf_{\pi \in \mathrm{Cpl}_c(\mu,\nu)} \int c\, d\pi < \infty,$$

this infimum is attained

Proof. The proof is completely analogously to the proof of Theorem 1.7, replacing Proposition 2.7 with Proposition 1.5. □

2.2 Bicausal couplings

Loosely speaking, a bicausal coupling is a coupling that is causal in both directions:

Definition 2.9. A causal coupling $\pi \in \mathrm{Cpl}_c(\mu, \nu)$ is called bicausal if $e_*\pi \in \mathrm{Cpl}_c(\nu, \mu)$.

Expample 2.10. Clearly, the calculation in Example 2.4 shows that the product coupling is bicausal. The coupling considered in Example 2.3 is not bicausal.

We will derive a few facts about bicausal couplings from the corresponding facts about causal couplings. First, we derive a necessary and sufficient criterion for bicausality which will be crucial later on.

Proposition 2.11 ([2, Proposition 5.1])**.** *Let $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ and $\pi \in \mathcal{P}(X \times Y)$. Then the following are equivalent:*

(i) $\pi \in \mathrm{Cpl}_{bc}(\mu, \nu)$

(ii) When decomposing

$$\pi(dx, dy) = \pi_1(dx_1, dy_1)\pi^{x_1,y_1}(dx_2, dy_2) \cdots \pi^{x_{1:N-1},y_{1:N-1}}(dx_N, dy_N)$$

it holds

(a) $\pi_1 \in \mathrm{Cpl}(\mathrm{pr}_{1*}\mu, \mathrm{pr}_{1*}\nu)$

(b) for all $t < N$ and π-almost all $(x_{1:t}, y_{1:t})$: $\pi^{x_{1:t},y_{1:t}} \in \mathrm{Cpl}(\mu^{x_{1:t}}, \nu^{y_{1:t}})$.

Proof. (i) $\implies$ (ii): This is an immediate consequence of (i) $\implies$ (iii) (a) and (b) in Proposition 2.2 applied to π and $e_*\pi$.

(ii) $\implies$ (i): Clearly, (ii) implies that π and $e_*\pi$ both satisfy the conditions stated in Proposition 2.2 (iii), which implies that $\pi \in \mathrm{Cpl}_c(\mu, \nu)$ and $e_*\pi \in \mathrm{Cpl}_c(\nu, \mu)$. □

As for causal couplings the set of bicausal couplings is compact, which implies the existence of minimizers to the bicausal transport problem.

Corollary 2.12. $\mathrm{Cpl}_{bc}(\mu, \nu)$ *is weakly compact.*

Proof. This is an immediate consequence of Proposition 2.7 (and Lemma A.44). □

Theorem 2.13. *Let $c : X \times Y \to \mathbb{R}$ be lower semi-continuous and bounded from below. If*

$$\inf_{\pi \in \mathrm{Cpl}_c(\mu,\nu)} \int c d\pi < \infty,$$

this infimum is attained.

Proof. The proof is completely analogously to the proof of Theorem 1.7, replacing Corollary 2.12 with Proposition 1.5. □

The analogue of bicausility for mappings is the following:

Definition 2.14. A map $T : X \to Y$ is called biadapted if it is bijective and T, T^{-1} are both adapted.

Corollary 2.15.

$$\mathrm{Cpl}_{bc}(\mu,\nu) \cap \mathrm{Cpl}_{00}(\mu,\nu) = \{(id, T)_*\mu : T \text{ biadapted }, T_*\mu = \nu\}$$

Proof. This is an immediate consequence of Lemma 2.6. □

We close this section with inductive characterizations of bicausality and biadaptedness, which will be helpful in the next sections.

Lemma 2.16. *Let $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ and $\pi \in \mathrm{Cpl}(\mu,\nu)$. Then $\pi \in \mathrm{Cpl}_{bc}(\mu,\nu)$ if and only if, when decomposing π as $d\pi(x,y) = d\pi_1(x_1,y_1)d\pi^{x_1,y_1}(x_{2:N},y_{2:N})$, one has $\pi^{x_1,y_1} \in \mathrm{Cpl}_{bc}(\mu^{x_1},\nu^{y_1})$ for π_1-almost all (x_1,y_1).*

Proof. This is an immediate consequence of Proposition 2.11. □

Lemma 2.17. *Let $T_1 : X_1 \to Y_1$ be a bijection and let $S : X_1 \times X_{2:N} \to Y_{2:N}$ be a measurable mapping such that for all $x_1 \in X_1$ the mapping $S^{x_1} : X_{2:N} \to Y_{2:N} : x_{2:N} \mapsto T(x_1, x_{2:N})$ is biadapted.*

Then the mapping

$$T : X_{1:N} \to Y_{1:N} : x_{1:N} \mapsto (T_1(x_1), S^{x_1}(x_{2:N}))$$

is biadapted.

Proof. It is clear that T is adapted and it is easy to check that it is injective and surjective. Therefore, there exits $F := T^{-1}$ and it suffices to show that F is adapted. To that end, we denote

$$F(y_{1:N}) = (F_1(y_{1:N}), F_{2:N}(y_{1:N}))$$

For each $y_{1:N}$ it holds

$$y_{1:N} = T(F(y_{1:N})) = (T_1(F_1(y_{1:N})), S^{F_1(y_{1:N})}(F_{2:N}(y_{1:N}))). \tag{2.1}$$

This implies $y_1 = T_1(F_1(y_{1:N}))$ and since T_1 is bijective this yields $T_1^{-1}(y_1) = F_1(y_{1:N})$; in particular, F_1 depends only on y_1. Moreover, (2.1) implies $y_{2:N} = S^{F_1(y_1)}(F_{2:N}(y_{1:N}))$. Since $S^{F_1(y_1)}$ is bijective, this implies

$$(S^{F_1(y_1)})^{-1}(y_{2:N}) = F_{2:N}(y_{1:N})$$

and since $S^{F_1(y_1)}$ is assumed to by biadapted, there exists $F_t : Y_{1:t} \to X_t$ measureable s.t.

$$F_{2:N}(y_{1:N}) = (F_2(y_{1:2}), \dots, F_N(y_{1:N})).$$ □

At the end of this section we briefly explain how to define the socalled adapted Wasserstein distance.

Definition 2.18. Let $X_1, \dots, X_N$ be Polish spaces with compatible metrics $d_1, \dots, d_N$. Then the adapted Wasserstein distance of two measures μ and $\nu \in \mathcal{P}(X)$ is defined as

$$\mathcal{AW}(\mu,\nu) := \inf \left\{ \int \sum_{i=1}^{N} d_i(x_i, y_i) d\pi(x,y) : \pi \in \mathrm{Cpl}_{bc}(\mu,\nu) \right\}.$$

We show that the measures μ_ε mentioned in the motivating example at the beginning of Section 2.1 do not converge to μ_0 w.r.t. the adapted Wasserstein distance.

Expample 2.19. For $\varepsilon > 0$ consider the measures

$$\mu = \frac{1}{2}\left(\delta_{(\varepsilon,1)} + \delta_{(-\varepsilon,-1)}\right) \qquad \nu = \frac{1}{2}\left(\delta_{(0,1)} + \delta_{(0,-1)}\right).$$

We show that $\mu \otimes \nu$ is the only bicausal couplings between μ and ν. Let $\pi \in \mathrm{Cpl}_{bc}(\mu,\nu)$ and decompose it as $d\pi(x,y) = d\pi_1(x_1,y_2) d\pi^{x_1,y_1}(x_2,y_2)$. By Proposition 2.2 it holds

$$\pi_1 \in \mathrm{Cpl}(\mu_1,\nu_1) = \mathrm{Cpl}\left(\frac{1}{2}\left(\delta_\varepsilon + \delta_{-\varepsilon}\right), \delta_0\right) = \left\{\frac{1}{2}\left(\delta_{(\varepsilon,0)} + \delta_{(-\varepsilon,0)}\right)\right\}$$
$$\pi^{\pm\varepsilon,0} \in \mathrm{Cpl}(\mu^{x_1},\nu^{y_1}) = \mathrm{Cpl}\left(\delta_{\pm 1}, \frac{1}{2}\left(\delta_1 + \delta_{-1}\right)\right) = \left\{\frac{1}{2}\left(\delta_{(\pm 1,1)} + \delta_{(\pm 1,-1)}\right)\right\}$$

Hence, $\pi = \frac{1}{4}\left(\delta_{(\varepsilon,1,0,1)} + \delta_{(\varepsilon,1,0,-1)} + \delta_{(-\varepsilon,-1,0,1)} + \delta_{(-\varepsilon,-1,0,-1)}\right) = \mu \otimes \nu$, which shows that

$$\mathcal{AW}(\mu,\nu) = \int \Big(|x_1 - y_1| + |x_2 - y_2|\Big) d\pi = \frac{1}{4}\left(\varepsilon + \sqrt{4+\varepsilon^2} + \sqrt{4+\varepsilon^2} + \varepsilon\right) \geq 1,$$

hence μ_ε does not converge top μ_0 w.r.t. the adapted Wasserstein distance.

2.3 Time dependent version of the representation of couplings

The aim is of this chapter is to prove a time dependent version of the representation of a coupling $\pi \in \mathrm{Cpl}(\mu,\nu)$ as a bijective Monge coupling $\widehat{\pi} \in \mathrm{Cpl}_{00}(\mu \otimes \lambda, \nu \otimes \lambda)$. More concretely, we want to represent a bicausal coupling π by a coupling $\widehat{\pi}$, which is supported on the graph of a biadapted mapping.

To this end, we will have to add an additional coordinate for randomization in each time step in the X- and Y-component. To keep notations short, we introduce the following abbreviations:

$$\widehat{X}_t := X_t \times [0,1], \qquad \widehat{X}_{1:t} := \prod_{i=1}^{t} \widehat{X}_i, \qquad \widehat{X} := \widehat{X}_{1:N}. \tag{2.2}$$

We will not always be careful about the ordering of the spaces X_t and $[0,1]$ in the definition of $\widehat{X}_{1:t}$ as product of those spaces. Instead of this, we agree to use consistent letters to name elements of those spaces unambiguously: Elements of $\widehat{X}_t$ are always called (x_t, u_t), where $x_t \in X_t$ and $u_t \in [0,1]$.

Therefore, $((x_1,u_1),\dots,(x_t,u_t))$ denotes the same element of $\widehat{X}_{1:t}$ as $(x_1,\dots,x_t,u_1,\dots,u_t)$ does, and the latter is often abbreviated as $(x_{1:t},u_{1:t})$. When evaluating functions $f:\widehat{X}_{1:t}\to\mathbb{R}$, we use the same convention, i.e. $f(x_{1:t},u_{1:t}):=f(x_1,\dots,x_t,u_1,\dots,u_t):=f((x_1,u_1),\dots,(x_t,u_t))$.

pr_X will always denote the projection $\widehat{X}_{1:t}\to X_{1:t}:(x_{1:t},u_{1:t})\mapsto x_{1:t}$.

If $\mu\in\mathcal{P}(X)$, we define $\widehat{\mu}\in\mathcal{P}(\widehat{X})$ as $\mu\otimes\lambda^N$, when $\widehat{X}$ is regarded as $(\prod_{t=1}^N X_t)\times[0,1]^N$, i.e. for $f:\widehat{X}\to\mathbb{R}$ we define

$$\int f d\widehat{\mu}:=\int f(x_{1:N},u_{1:N})d\mu(x_{1:N})d\lambda(u_{1:N}).$$

We use the same convention for the Y-component, where elements of $\widehat{Y}_t$ are called (y_t,v_t) with $y_t\in Y_t$ and $v_t\in[0,1]$, etc.

The terms (bi)adapted and (bi)causal are always meant to be understood as the ordering of the spaces in (2.2) suggests: In each time-step we consider the spaces $\widehat{X}_t=X_t\times[0,1]$ and $\widehat{Y}_t=Y_t\times[0,1]$. Loosely speaking, *in* (not before or after) each time step we "add" one unit interval in the X-component and one unit interval in the Y-component. Explicitly, a mapping $T:\widehat{X}\to\widehat{Y}$ is adapted if for all $t\le N$ there exists mappings $T_t:\widehat{X}_{1:t}\to\widehat{Y}_t$ s.t. $T(x_{1:N},u_{1_N})=(T_1(x_1,u_1),\dots,T_N(x_{1:N},u_{1:N}))$.

Using this notation, we can formulate the main theorem of this section:

Theorem 2.20. *Let $\mu\in\mathcal{P}(X),\nu\in\mathcal{P}(Y)$ and $\pi\in\mathrm{Cpl}_{bc}(\mu,\nu)$ be given. Then there exists a biadapted mapping $T:\widehat{X}\to\widehat{Y}$ satisfying*

(i) $T_*\widehat{\mu}=\widehat{\nu}$*, or equivalently:* $\widehat{\pi}:=(id,T)_*\widehat{\mu}\in\mathrm{Cpl}_{bc}(\widehat{\mu},\widehat{\nu})$

(ii) $\mathrm{pr}_{X\times Y*}\widehat{\pi}=\pi$.

We will prove this theorem by induction on the number of time steps. In order to avoid measurability issues in the induction step, we proof a slightly more general version of Theorem 2.20.

Theorem 2.21. *Let Z be a further Polish space and let kernels μ from Z to X, ν from Z to Y and π from Z to $X\times Y$ be given. Assume that $\pi^z\in\mathrm{Cpl}_{bc}(\mu^z,\nu^z)$ for all $z\in Z$.*

Then there exists a measurable mapping $T:Z\times\widehat{X}\to\widehat{Y}$ such that for all $z\in Z$ the mappings $T^z:\widehat{X}\to\widehat{Y}:(x_{1:N},u_{1:N})\mapsto T(z,x_{1:N},u_{1:N})$ are Borel isomorphisms satisfying

(i) $T^z{}_*\widehat{\mu^z}=\widehat{\nu^z}$*, or equivalently:* $\widehat{\pi^z}:=(id,T^z)_*\widehat{\mu^z}\in\mathrm{Cpl}_{bc}(\widehat{\mu^z},\widehat{\nu^z})$

(ii) $\mathrm{pr}_{X\times Y*}\widehat{\pi^z}=\pi^z$.

Proof. For one timestep (i.e. $X=X_1$, $Y=Y_1$) bicausality is a trivial condition and biadapted is equivalent to bijective. Therefore, Theorem 1.15 is exactly the claim for one timestep.

Assume that we have already proven Theorem 2.21 for $N-1$ time steps. Let μ be a kernel from Z to $X_{1:N}$, ν be a kernel from Z to $Y_{1:N}$ and π be a kernel from Z to $X_{1:N}\times Y_{1:N}$ satisfying $\pi^z\in\mathrm{Cpl}_{bc}(\mu^z,\nu^z)$ for all $z\in Z$.

For each $z \in Z$ we can decompose μ^z, ν^z and π^z as

$$d\mu^z = d\mu^{x_1,z} d\mu_1^z \qquad d\nu^z = d\nu^{y_1,z} d_1\nu^z \qquad d\pi^z = d\pi^{x_1,y_1,z} d\pi_1^z.$$

and by Lemma 2.16 it holds $\pi^{x_1,y_1,z} \in \mathrm{Cpl}_{bc}(\mu^{x_1,z}, \nu^{y_1,z})$.

By the induction hypothesis, there exists a mapping

$$S : (X_1 \times Y_1 \times Z) \times \widehat{X}_{2:N} \to \widehat{Y}_{2:N}$$

such that for all $(x_1, y_1, z) \in X_1 \times Y_1 \times Z$ the mapping $S^{x_1,y_1,z} : X_{2:N} \to Y_{2_N} : (x_{2:N}, y_{2:N}) \mapsto S(x_1, y_1, z, x_{2:N}, y_{2:N})$ is a Borel isomorphism satisfying[3]

- $S^{x_1,y_1,z}{}_*\widehat{\mu^{x_1,z}} = \widehat{\nu^{y_1,z}}$
- $\mathrm{pr}_{XY*}\widehat{\pi^{x_1,y_1,z}} = \pi^{x_1,y_1,z}$, where $\widehat{\pi^{x_1,y_1,z}} := (\mathrm{id}, S^{x_1,y_1,z})_*\widehat{\mu^{x_1,z}}$

Moreover, by Theorem 1.15 there exists a measurable mapping

$$T_1 : Z \times \widehat{X}_1 \to \widehat{Y}_1$$

such that for all $z \in Z$, the mapping $T_1^z : \widehat{X}_1 \to \widehat{Y}_1 : (x_1, u_1) \mapsto T_1(z, x_1, u_1)$ is a Borel isomorphism satisfying $T_1^z{}_*\widehat{\mu_1^z} = \widehat{\nu_1^z}$ and $\mathrm{pr}_{XY*}(\mathrm{id}, T_1^z)_*\widehat{\mu_1^z} = \pi_1^z$.

We define the mapping

$$T : Z \times \widehat{X} \to \widehat{Y} : (z, x_{1:N}, u_{1:N}) \mapsto (T_1^z(x_1, u_1), S^{x_1, \mathrm{pr}_X(T_1^z(x_1,u_1))}(x_{2:N}, u_{2:N}))$$

We have to check that T has the desired properties:

Clearly, T is measurable as composition. Fix $z \in Z$. The mapping T^z is biadapted by Lemma 2.17.

In order to check that $T_*\widehat{\mu^z} = \widehat{\nu^z}$, we consider an arbitrary measurable function $f : \widehat{Y} \to \mathbb{R}$. We achieve by using the properties $S^{x_1,y_1,z}{}_*\widehat{\mu^{x_1,z}} = \widehat{\nu^{y_1,z}}$ and $T_1^z{}_*\widehat{\mu_1^z} = \widehat{\nu_1^z}$

$$\begin{aligned}
&\int f(y_{1:N}, v_{1:N}) dT_*^z\widehat{\mu^z}(y_{1:N}, v_{1:N}) = \\
&\quad = \int f(T_1^z(x_1,u_1), S^{x_1,\mathrm{pr}_X(T_1^z(x_1,u_1))}(x_{2:N}, u_{2:N})) d\widehat{\mu^{x_1,z}}(x_{2:N}, u_{2:N}) d\widehat{\mu_1^z}(x_1, u_1) \\
&\quad = \int f(T_1^z(x_1,u_1), y_{2:N}, v_{2:N}) \underbrace{dS^{x_1,\mathrm{pr}_X(T_1^z(x_1,u_1))}{}_*\widehat{\mu^{x_1,z}}(y_{2:N}, v_{2:N})}_{=d\widehat{\nu^{\mathrm{pr}_X \circ T_1^z(x_1,u_1),z}}(y_{2:N},v_{2:N})} d\widehat{\mu_1^z}(x_1, u_1) \\
&\quad = \int f(y_{1:N}, v_{1:N}) d\widehat{\nu^{y_1,z}}(y_{2:N}, v_{2:N}) dT_1^z{}_*\widehat{\mu_1^z}(y_1, v_1) \\
&\quad = \int f(y_{1:N}, v_{1:N}) d\widehat{\nu^z}(y_{1:N}, v_{1:N}),
\end{aligned}$$

[3] As the notation given at the beginning of this section suggests $\widehat{\mu^{x_1,z}}$ is defined via

$$\int f d\widehat{\mu^{x_1,z}} := \int f(x_{2:N}, u_{2:N}) d\mu^{x_1,z}(x_{2:N}) d\lambda^{N-1}(u_{2:N})$$

i.e. $T^z_* \widehat{\mu^z} = \widehat{\nu}^z$.

It remains to show that $\hat{\pi} := (\mathrm{id}, T)_* \widehat{\mu}$ satisfies $\mathrm{pr}_{XY*}\widehat{\pi} = \pi$. Using this and the properties $\mathrm{pr}_{XY*}\widehat{\pi^{x_1,y_1,z}} = \pi^{x_1,y_1,z}$ and $\mathrm{pr}_{XY*}(\mathrm{id}, T^z_1)_*\widehat{\mu^z_1} = \pi^z_1$ we obtain for any measurable function $f : X \times Y \to \mathbb{R}$:

$$
\begin{aligned}
\int f(x_{1:N}, y_{1:N}) d\mathrm{pr}_{XY*}\widehat{\pi^z}(x_{1:N}, y_{1:N}) = \\
&= \int f(\mathrm{pr}_{XY} \circ (\mathrm{id}, T^z)(x_{1:N}, u_{1:N})) d\widehat{\mu^{x_1,z}}(x_{2:N}, u_{2:N}) d\widehat{\mu^z_1}(x_1, u_1) \\
&= \int f(\mathrm{pr}_{XY}((\mathrm{id}, T^z_1)(x_1, u_1)), \mathrm{pr}_{XY}((\mathrm{id}, S^{x_1, \mathrm{pr}_X \circ T^z_1(x_1,u_1), z})(x_{2:N}, u_{2:N})) \\
&\qquad\qquad d\widehat{\mu^{x_1,z}}(x_{2:N}, u_{2:N}) d\widehat{\mu^z_1}(x_1, u_1) \\
&= \int f(\mathrm{pr}_{XY}((\mathrm{id}, T_1)(x_1, u_1)), x_{2:N}, y_{2:N}) d\pi^{x_1,y_1,z}(x_{2:N}, y_{2:N}) d\bar{\mu}(x_1) d\lambda(u_1) \\
&= \int f(x_{1:N}, y_{1:N}) d\pi^{x_1,y_1,z}(x_{2:N}, y_{2:N}) d\pi^z_1(x_1, y_1) \\
&= \int f(x_{1:N}, y_{1:N}) d\pi(x_{1:N}, y_{1:N}),
\end{aligned}
$$

which yields the desired result $\mathrm{pr}_{XY*}\widehat{\pi^z} = \pi^z$. □

2.4 Denseness of biadapted mappings in the set of bicausal couplings

Analogously to Section 1.3 we use the representation of a bicausal coupling $\pi \in \mathrm{Cpl}(\mu, \nu)$ as coupling $\widehat{\pi} = (\mathrm{id}, T)_*(\mu \otimes \lambda) \in \mathrm{Cpl}_{00}(\mu \otimes \lambda, \nu \otimes \lambda)$ to prove the denseness of couplings supported on the graph of biadapted mappings among bicausal couplings.

First, we state the regularity assumption on the marginals, which is essential for our proof:

Assumption 2.22. Let $\mu \in \mathcal{P}(X_{1:N})$. We say μ satisfies Assumption 2.22 if μ has a disintegration

$$d\mu(x_{1:N}) = d\mu_1(x_1) d\mu^{x_1}(x_2) \cdots d\mu^{x_{1:N-1}}(x_N)$$

such that μ_1 is continuous and for all $t < N$ and $x_{1:t} \in X_{1:t}$ the measure $\mu^{x_{1:t}}$ is continuous.

Remark 2.23. Let $X_t = \mathbb{R}$ for $t \in \{1, \dots, N\}$. If $\mu \in \mathcal{P}(X_{1:N}) = \mathcal{P}(\mathbb{R}^N)$ is absolutely continuous w.r.t. Lebesgue measure, it satisfies Assumption 2.22.

The part of the proof of the main theorem, where this assumption is crucial, is the following proposition:

Proposition 2.24. *For $t \in \{1, \dots, N\}$ let $\mathcal{M}_t$ be an at most countable partition of X_t consisting of Borel sets. Let $\mu \in \mathcal{P}(X)$ be satisfying Assumption 2.22.*

Then there exists a biadapted mapping $\Phi^{\mathcal{M}}_\mu : X \to \widehat{X}$ s.t. for all $(M_1, \dots, M_N) \in \mathcal{M}_1 \times \cdots \times \mathcal{M}_N$ one has ${\Phi^{\mathcal{M}}_\mu}_*(\mu|_{M_{1:N}}) = \mu|_{M_{1:N}} \otimes \lambda$.

We want to prove this propostion by induction on the number of time steps. In order to avoid measurability issues in the induction step, we have to prove a slightly stronger claim:

Proposition 2.25. *Let Z be a further Polish space and for $t \in \{1, \dots, N\}$ let $\mathcal{M}_t$ be an at most countable partition of X_t consisting of Borel sets. Let μ be a kernel from Z to X s.t. μ^z satisfies Assumption 2.22 for all $z \in Z$.*

Then there exists a measurable mapping $\Phi_\mu^{\mathcal{M}} : Z \times X \to \widehat{X}$ s.t. for all $z \in Z$ the mapping $\Phi_\mu^{\mathcal{M},z} := \Phi_\mu^{\mathcal{M}}(z, \cdot) : X \to \widehat{X}$ is a biadapted mapping satisfying the following property: For all $(M_1, \dots, M_N) \in \mathcal{M}_1 \times \dots \times \mathcal{M}_N$ it holds ${\Phi_\mu^{\mathcal{M},z}}_*(\mu^z|_{M_{1:N}}) = \widehat{\mu^z|_{M_{1:N}}}$.

Proof. We prove this Proposition by induction on N:

The claim is true for $N = 1$ by proposition A.42. Assume that the claim is true for $N - 1$. Let $\mu \in \mathcal{P}(X_{1:N})$ satisfying Assumption 2.22 and decompose it as

$$d\mu(x_{1:N}) = d\mu_1(x_1) d\mu^{x_1}(x_{2:N})$$

By the induction hypothesis there exists a measurable mapping

$$\Psi : (Z \times X_1) \times X_{2:N} \to \widehat{X}_{2:N}$$

s.t. for all $(z, x_1) \in Z \times X_1$ the mapping $\Psi^{z,x_1} := \Psi(z, x_1, \cdot) : X_{2:N} \to \widehat{X}_{2:N}$ is a biadapted mapping satisfying ${\Psi^{z,x_1}}_*(\mu^{z,x_1}|_{M_{2:N}}) = \widehat{\mu^{z,x_1}|_{M_{2:N}}}$.

By Proposition A.42 there exists a measurable mapping

$$\Phi_1 : Z \times X_1 \to \widehat{X_1}$$

s.t. for all $z \in Z$ the mapping $\Phi_1^z := \Phi_1(z, \cdot) : X_1 \to \widehat{X_1}$ is a Borel isomorphism satisfying ${\Phi_1^z}_*(\mu_1^z|_{M_1}) = \widehat{\mu_1^z|_{M_1}}$ for all $M_1 \in \mathcal{M}_1$.

Consider the mapping

$$\Phi_\mu^{\mathcal{M}} : Z \times X \to \widehat{X} : (z, x_{1:N}) \mapsto (\Phi_1^z(x_1), \Psi^{z,x_1}(x_{2:N})).$$

Clearly, $\Phi_\mu^{\mathcal{M}}$ is measurable as concatenation and for all $z \in Z$ the map $\Phi_\mu^{\mathcal{M},z}$ is biadapted by Lemma 2.17. For all $(M_1, \dots, M_N) \in \mathcal{M}_1 \times \dots \times \mathcal{M}_N$ and $f : \widehat{M_{1:N}} \to \mathbb{R}$ measurable it holds

$$\begin{aligned}
&\int f(x_{1:N}, u_{1:N}) d{\Phi_\mu^{\mathcal{M},z}}_*(\mu^z|_{M_{1:N}}) = \\
&= \int_{M_1} \int_{M_{2:N}} f(\Phi_1^z(x_1), \Psi^{z,x_1}(x_{2:N})) d\mu^{z,x_1}(x_{2:N}) d\mu_1^z(x_1) \\
&= \int_{M_1} \int_{M_{2:N}} f(\Phi_1^z(x_1), x_{2:N}, u_{2:N}) \underbrace{d{\Psi^{z,x_1}}_*(\mu^{z,x_1}|_{M_{2:N}})(x_{2:N}, u_{2:N})}_{= d\widehat{\mu^{z,x_1}|_{M_{2:N}}}(x_{2:N}, u_{2:N})} d\mu_1^z(x_1) \\
&= \int_{M_1} \int_{M_{2:N}} f(x_1, u_1, x_{2:N}, u_{2:N}) d\widehat{\mu^{z,x_1}|_{M_{2:N}}}(x_{2:N}, u_{2:N}) d{\Phi_1^z}_*(\mu_1^z|_{M_1})(x_1, u_1) \\
&= \int f(x_{1:N}, u_{1:N}) d\widehat{\mu^z|_{M_{1:N}}},
\end{aligned}$$

which yields ${\Phi_\mu^{\mathcal{M},z}}_*(\mu^z|_{M_{1:N}}) = \widehat{\mu^z|_{M_{1:N}}}$. □

Now we have provided all tools to prove the main theorem of this thesis. The structure of the upcoming proof is completely analogue to the proof of Theorem 1.20.

Theorem 2.26. *Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ both satisfy Assumption 2.22. Then the set of couplings between μ and ν supported by the graph of a biadapted mapping are dense among the bicausal couplings between μ and ν w.r.t. weak convergence, i.e.*

$$\mathrm{Cpl}_{bc}(\mu,\nu) = \overline{\{(id,T)_*\mu : T : X \to Y \textit{ biadapted, satisfying } T_*\mu = \nu\}}.$$

Proof. By Corollary 2.12 the set $\mathrm{Cpl}_{bc}(\mu,\nu)$ is closed, so it suffices to show that any $\pi \in \mathrm{Cpl}_{bc}(\mu,\nu)$ can be approximated by a sequence $\pi_n \in \mathrm{Cpl}_{00}(\mu,\nu) \cap \mathrm{Cpl}_{bc}(\mu,\nu)$ w.r.t. weak convergence.

According to Theorem 2.20 there exists a coupling $\widehat{\pi} = (id,T)_*\widehat{\mu} \in \mathrm{Cpl}_{00}(\widehat{\mu},\widehat{\nu})$ such that

(i) $T_*\widehat{\mu} = \widehat{\nu}$

(ii) $\mathrm{pr}_{XY*}\widehat{\pi} = \pi$.

For $t \in \{1,\dots,N\}$ let $(\mathcal{A}_n^t)_{n\in\mathbb{N}}$ and $(\mathcal{B}_n^t)_{n\in\mathbb{N}}$ be sequences of partitions of X_t and Y_t consisting of countably many Borel sets and satisfying $\lim_{n\to\infty}||\mathcal{A}_n^t|| = 0$ and $\lim_{n\to\infty}||\mathcal{B}_n^t|| = 0$. According to Proposition 2.24, for all $n \in \mathbb{N}$ there exist biadapted mappings $\Phi_n : X \to \widehat{X}$ and $\Psi_n : Y \to \widehat{Y}$ such that

(iii) $\Phi_{n*}(\mu|_{A_{1:N}}) = \widehat{\mu|_{A_{1:N}}}$ for all $A_t \in \mathcal{A}_n^t$

(iv) $\Psi_{n*}(\mu|_{B_{1:N}}) = \widehat{\nu|_{B_{1:N}}}$ for all $B_t \in \mathcal{B}_n^t$.

For $n \in \mathbb{N}$ define the mapping

$$T_n := \Psi_n^{-1} \circ T \circ \Phi_n : X \to Y.$$

T_n is biadapted as composition of biadapted mappings and satisfies $T_{n*}\mu = \nu$. We need to check that $\pi^n := (\mathrm{id},T^n)_*\mu \rightharpoonup \pi$. Note that $\mathcal{M}_n := \{A_{1:N} \times B_{1:N} : A_t \in \mathcal{A}_n^t, B_t \in \mathcal{B}_n^t\}$ are partitions of $X \times Y$ consisting of countably many Borel sets satisfying $\lim_{n\to\infty}||\mathcal{M}_n|| = 0$. Hence, by Lemma 1.18 it suffices to show for all $n \in \mathbb{N}$:

$$\pi_n(A_{1:N} \times B_{1:N}) = \pi(A_{1:N} \times B_{1:N}) \qquad \text{for all } A_t \in \mathcal{A}_n^t, B_t \in \mathcal{B}_n^t$$

This is a consequence of the properties (i) to (iv) of the mappings Φ_n, Ψ_n and T, as the following calculation, where were denote $A := A_{1:N}$ and $B := B_{1:N}$ shows:

$$\begin{aligned}
\pi_n(A \times B) &= \mu(A \cap T_n^{-1}(B)) = \mu|_A((\Phi_n^{-1} \circ T^{-1} \circ \Psi_n)(B)) \overset{(iii)}{=} (\mu|_A \otimes \lambda)(T^{-1}(\Psi_n(B))) \\
&= (\mu \otimes \lambda)(A \times [0,1] \cap T^{-1}(\Psi_n(B))) \overset{(i)}{=} (\nu \otimes \lambda)(T(A \times [0,1]) \cap \Psi_n(B)) \\
&= (\nu \otimes \lambda)(\Psi_n(\Psi_n^{-1}(T(A \times [0,1])) \cap B)) = \Psi_{n*}(\nu \otimes \lambda)(\Psi_n^{-1}(T(A \times [0,1])) \cap B) \\
&= \nu|_B(\Psi_n^{-1}(T(A \times [0,1]))) \overset{(iv)}{=} (\nu|_B \otimes \lambda)(T(A \times [0,1])) \\
&= (\nu \otimes \lambda)(T(A \times [0,1]) \cap B \times [0,1]) = (\mathrm{id},T^{-1})_*(\nu \otimes \lambda)(A \times [0,1] \times B \times [0,1]) \\
&= \widehat{\pi}(A \times [0,1] \times B \times [0,1]) \overset{(ii)}{=} \pi(A \times B)
\end{aligned}$$

□

2.5 Discussion

In [3] a version of the result for causal couplings was shown, i.e. causal couplings supported on the graph of adapted mappings are dense in the set of causal couplings with fixed marginals. It was sufficient to require continuity assumptions for the first timestep of the x-marginal.

Since bicauslity is causality in both directions, it is clear that we will need the same assumption for the y-marginal as well. However, the assumptions in Theorem 2.26 are much stronger than that.

We give an example that these stronger assumptions than continuity of the x- and y- marginals in the first timestep are necessary.

Expample 2.27. Let $N = 2$ and $X_1 = X_2 = Y_1 = Y_2 = [0,1]$. Consider the measures $\mu = \lambda \otimes \delta_0 \in \mathcal{P}(X_1 \times X_2)$ and $\nu = \lambda^2 \in \mathcal{P}(Y_1 \times Y_2)$.

Since μ and ν are both continuous measures, couplings between μ and ν supported on the graphs of bijections are dense in $\mathrm{Cpl}(\mu, \nu)$, see Theorem 1.20. We show that there are no bicausal couplings between μ and ν that are supported on the graph a bijection.

Assume that there exists a $\pi \in \mathrm{Cpl}_{bc}(\mu, \nu) \cap \mathrm{Cpl}_{00}(\mu, \nu)$ and decompose it as $d\pi(x_1, x_2, y_1, y_2) = d\pi_1(x_1, y_2) d\pi^{x_1,y_1}(x_2, y_2)$. By Lemma 2.16 it holds $\pi^{x_1,y_1} \in \mathrm{Cpl}_{00}(\mu^{x_1}, \nu^{y_1})$ for π_1-almost all (x_1, y_1). However, for all x_1, y_1 it holds $\mathrm{Cpl}_{00}(\mu^{x_1}, \nu^{y_1}) = \mathrm{Cpl}_{00}(\delta_0, \lambda) = \emptyset$. Therefore, such a π cannot exist and $\mathrm{Cpl}_{bc}(\mu, \nu) \cap \mathrm{Cpl}_{00}(\mu, \nu)$ is empty.

However, there is a sequence $(T_n)_n$ of adapted mappings pushing μ to ν such that $(\mathrm{id}, T_n)_* \mu \rightharpoonup \pi$: By Theorem 1.20 there is a sequence $(F_n)_n$ such that $F_{n*}\lambda = \lambda^2$ and $(\mathrm{id}, F_n)_* \lambda \rightharpoonup \lambda^3$. Then $T_n(x_1, x_2) := F_n(x_1)$ has the desired properties.

There are also cases, where $\mathrm{Cpl}_{bc}(\mu, \nu) \cap \mathrm{Cpl}_{00}(\mu, \nu)$ is not empty, but the assertion of Theorem 2.26 is still wrong:

Expample 2.28. Let $N = 2$ and $X_1 = X_2 = Y_1 = Y_2 = [0,1]$. Consider the measure $d\mu(x_1, x_2) = d\mu_1(x_1) d\mu^{x_1}(x_2)$, where $\mu_1 := \lambda$ and $\mu^{x_1} := (1 - x_1)\delta_0 + x_1 \delta_1$. Set $\nu := \mu$.

We claim that $\mathrm{Cpl}_{bc}(\mu, \nu) \cap \mathrm{Cpl}_{00}(\mu, \nu) = \{(\mathrm{id}, \mathrm{id})_* \mu\}$. Clearly, $(\mathrm{id}, \mathrm{id})_* \mu \in \mathrm{Cpl}_{bc}(\mu, \nu) \cap \mathrm{Cpl}_{00}(\mu, \nu)$. Consider an arbitrary $\pi \in \mathrm{Cpl}_{bc}(\mu, \nu) \cap \mathrm{Cpl}_{00}(\mu, \nu)$. By Lemma 2.16 it holds $\pi^{x_1,y_1} \in \mathrm{Cpl}_{00}(\mu^{x_1}, \nu^{y_1})$ for π_1-almost all (x_1, y_1). Note that $\pi_1(\{(1/2, 1/2)\}) \leq \pi_1(\{1/2\} \times [0,1]) = \lambda(\{1/2\}) = 0$ for all $\pi_1 \in \mathrm{Cpl}(\mu_1, \nu_1) = \mathrm{Cpl}(\lambda, \lambda)$, so we can neglect the point $(1/2, 1/2)$ from now on. It holds

$$\mathrm{Cpl}_{00}(\mu^{x_1}, \nu^{y_1}) = \mathrm{Cpl}_{00}((1-x_1)\delta_0 + x_1\delta_1, (1-y_1)\delta_0 + y_1\delta_1) = \begin{cases} \{(1 - x_1)\delta_{(0,0)} + x_1\delta_{(1,1)}\} & x_1 = y_1 \\ \emptyset & x_1 \neq y_1 \end{cases}$$

This implies $\mathrm{supp}(\pi_1) \subseteq \Delta := \{(x_1, x_1) : x_1 \in [0,1]\}$. By using that $(A \times B) \cap \Delta = \{(x_1, x_1) : x_1 \in A \cap B\} = ((A \cap B) \times [0,1]) \cap \Delta$ this yields

$$\pi_1(A \times B) = \pi_1((A \cap B) \times [0,1]) = \lambda(A \cap B) = \lambda((\mathrm{id}, \mathrm{id})^{-1}(A \times B)) = (\mathrm{id}, \mathrm{id})_*(A \times B),$$

so $\pi_1 = (\mathrm{id}, \mathrm{id})_* \lambda$. Hence,

$$d\pi(x_1, x_2, y_1, y_2) = d(\mathrm{id}, \mathrm{id})_* \lambda(x_1, y_1) d((1-x_1)\delta_{(0,0)} + x_1\delta_{(1,1)})(x_2, y_2) = d(\mathrm{id}, \mathrm{id})_* \mu(x_1, x_2, y_1, y_2).$$

So, $\mathrm{Cpl}_{bc}(\mu,\nu) \cap \mathrm{Cpl}_{00}(\mu,\nu) = \{(\mathrm{id},\mathrm{id})_*\mu\}$ and therefore also its closure is just $\{(\mathrm{id},\mathrm{id})_*\mu\}$, which does not contain the bicausal coupling $\mu \otimes \nu$.

To close this section, we state the consequences of Theorem 2.26 for the bicausal transport problem and for the adapted Wasserstein distance.

Corollary 2.29. *Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ satisfy Assumption 2.22 and $c : X \times Y \to \mathbb{R}$ be continuous and bounded. Then it holds*

$$\inf\left\{\int c d\pi : \pi \in \mathrm{Cpl}_{00}(\mu,\nu)\right\} = \inf\left\{\int c(x,T(x))d\mu(x) : T \text{ biadapted s.t. } T_*\mu = \nu\right\}.$$

Proof. We can assume that $\inf_{\pi\in\mathrm{Cpl}_{bc}(\mu,\nu)} \int c(x,y)\pi(dx,dy) < \infty$ (otherwise the claim is trivial). By Theorem 2.13, in this case there exits $\pi \in \mathrm{Cpl}(\mu,\nu)$ s.t.

$$\int c d\pi = \inf_{\pi\in\mathrm{Cpl}(\mu,\nu)} \int c(x,y)\pi(dx,dy).$$

By Theorem 2.26 there exists a sequence $\pi_n \in \mathrm{Cpl}_{00}(\mu,\nu)$ s.t. $\pi_n \rightharpoonup \pi$ and by Lemma 1.6

$$\lim_n \int c d\pi_n = \int c d\pi = \inf_{\pi\in\mathrm{Cpl}(\mu,\nu)} \int c(x,y)\pi(dx,dy).$$

□

Corollary 2.30. *Let $\mu,\nu \in \mathcal{P}(X)$ satisfy Assumption 2.22. Then it holds*

$$\mathcal{AW}(\mu,\nu) = \inf\left\{\int d(x,T(x))d\mu(x) : T \text{ biadapted s.t. } T_*\mu = \nu\right\}.$$

2.6 Results in probabilistic notation

In this section we state the main results of this chapter in probabilistic notion.

Theorem 2.31. *Consider the discrete time processes $\mathsf{X} = (\mathsf{X}_1,\ldots,\mathsf{X}_N)$ and $\mathsf{Y} = (\mathsf{Y}_1,\ldots,\mathsf{Y}_N)$, where X_t takes values in X_t and Y_t that values in Y_t, which satisfy the following properties*

- *For all $t \le N$: $\mathcal{F}^Y_t$ and $\mathcal{F}^X_N$ are independent given $\mathcal{F}^X_t$*
- *For all $t \le N$: $\mathcal{F}^X_t$ and $\mathcal{F}^Y_N$ are independent given $\mathcal{F}^Y_t$*

Then there exists a biadapted mapping $F = (F_1,\ldots,F_N) : \widehat{X} \to \widehat{Y}$ satisfying

(i) If $\mathsf{U}_1,\ldots,\mathsf{U}_N$ are mutually independent uniformly distributed random variables independent of X and $\mathsf{V}_1,\ldots,\mathsf{V}_N$ are mutually independent uniformly distributed random variables independent of Y, the process Z defined as

$$\mathsf{Z}_t := F_t(\mathsf{X}_1,\mathsf{U}_1,\ldots,\mathsf{X}_t,\mathsf{U}_t)$$

satisfies

$$\mathsf{Z} \sim (\mathsf{Y}_1,\mathsf{V}_1,\mathsf{Y}_2,\mathsf{V}_2,\ldots,\mathsf{Y}_N,\mathsf{V}_N).$$

(ii) The process W*, which is defined as the projections of the* Y*-compensates of the process* Z*, satisfies* $(\mathsf{X}, \mathsf{W}) \sim (\mathsf{X}, \mathsf{Y})$*.*

Theorem 2.32. *Consider the discrete-time processes* $\mathsf{X} = (\mathsf{X}_1, \dots, \mathsf{X}_N)$ *and* $\mathsf{Y} = (\mathsf{Y}_1, \dots, \mathsf{Y}_N)$*, where* X_t *takes values in* X_t *and* Y_t *that values in* Y_t*, which satsify the following properties*

- *For all* $t \leq N$*:* $\mathcal{F}_t^Y$ *and* $\mathcal{F}_N^X$ *are independent given* $\mathcal{F}_t^X$
- *For all* $t \leq N$*:* $\mathcal{F}_t^X$ *and* $\mathcal{F}_N^Y$ *are independent given* $\mathcal{F}_t^Y$
- *For all* $t \leq N$*: the law of* X_t *given* $\mathsf{X}_1 = x_1, \dots, \mathsf{X}_{t-1} = x_{t-1}$ *is a.s. continuous*
- *For all* $t \leq N$*: the law of* Y_t *given* $\mathsf{Y}_1 = y_1, \dots, \mathsf{Y}_{t-1} = y_{t-1}$ *is a.s. continuous*

Then there exists a sequence $(F^k)_{k \in \mathbb{N}}$ *of biadapted functions* $F^k : X \to Y$ *satisfying* $F^k(\mathsf{X}) \sim \mathsf{Y}$*, such that* $(\mathsf{X}, F^k(\mathsf{X}))$ *converges to* (X, Y) *in distribution for* $k \to \infty$*.*

Proof. This is a consequence of Theorem 2.26. □

Appendix A

A.1 Preliminaries from topology and descriptive set theory

This section is a brief introduction to Polish spaces and a few aspects of descriptive set theory, for a detailed introduction on these topics the reader is refereed to [6, Chapters 1 and 2].

Definition A.1. A topological space $(X, \mathcal{T})$ is a Polish space if

- X is seperable, i.e. there is a countable dense subset of X
- X is completely metrizable, i.e. there exits a metric d, which is complete and induces the topology $\mathcal{T}$.

A metric d having these properties is called a compatible metric.

We collect some important facts about Polish spaces:

Recall that a subset of a topological space is G_δ if and only if it is the countable intersection of open sets. Closed subsets of metric spaces are G_δ.

Theorem A.2 ([6, Theorem 4.14])**.** *X is a Polish space if and only if it is homeomorphic to a G_δ-subset of the Hilbertcube, that is $[0,1]^{\mathbb{N}}$ equipped with the product topology.*

Definition A.3. Let X be a topological space. A point $x \in X$ is an isolated point if $\{x\}$ is open. Otherwise, it is a limit point of X. The space X is called dense-in-itself if all $x \in X$ are limit points of X. $A \subseteq X$ is perfect if it is closed in X and dense-in-itself w.r.t. the subspace topology.

Theorem A.4 (Cantor-Bendixson, [6, Theorem 6.4])**.** *Let X be a Polish space. Then X can be uniquely written as $X = P \cup C$, where P is a perfect set and C is at most countable and open.*

Theorem A.5 ([6, Theorem 6.2])**.** *Any non-empty perfect Polish space contains a homeomorphic copy of $\{0,1\}^{\mathbb{N}}$.*

An immediate consequence of the last three theorems is:

Theorem A.6. *Any uncountable Polish space contains a homeomorphic copy of $\{0,1\}^{\mathbb{N}}$ and has cardinality $2^{\aleph_0}$.*

Proof. Write $X = P \cup C$, where P is perfect and C is at most countable. Since X is uncountable, P is not empty, so it contains a homeomorphic copy of $\{0,1\}^{\mathbb{N}}$ and X has at least cardinality $2^{\aleph_0}$. Theorem A.2 implies that X has at most cardinality $2^{\aleph_0}$. □

A measurable space $(X, \mathcal{A})$ is a set X with a σ-algebra $\mathcal{A} \subseteq 2^X$.

If $(X, \mathcal{A}_X)$ and $(Y, \mathcal{A}_Y)$ are measurable spaces, a mapping $f : X \to Y$ is called measurable if for all $A \in \mathcal{A}_X$ we have $f^{-1}(A) \in \mathcal{A}_Y$. $(X, \mathcal{A}_X)$ and $(Y, \mathcal{A}_Y)$ are called isomorphic if there exists a bijection $f : X \to Y$ s.t. f and f^{-1} are both measureable.

If $(X, \mathcal{T})$ is topological space the Borel-σ-algebra on X is the smallest σ-algebra on X containing all open subsets of X.

Definition A.7. A measureable space $(X, \mathcal{A})$ is called standard Borel space if it is isomorphic to a Polish space Y equipped with the Borel-σ-algebra. A bijection f between two standard Borel spaces s.t. f and f^{-1} are both Borel is called Borel isomorphism.

Theorem A.8 ([6, Theorem 13.1])**.** *Let $(X, \mathcal{T})$ be a Polish space and $B \subseteq X$ be Borel. Then there is a Polish topology $\mathcal{T}_B \supseteq \mathcal{T}$ s.t. B is clopen in $\mathcal{T}_B$ and $\mathcal{B}(\mathcal{T}) = \mathcal{B}(\mathcal{T}_B)$.*

Theorem A.9 ([6, Theorem 13.11])**.** *Let $(X, \mathcal{T})$ be a Polish space, Y be a second countable space and $f : X \to Y$ be Borel measurable. Then there exists a Polish topology $\mathcal{T}_f \supseteq \mathcal{T}$ such that f is $\mathcal{T}_f$-continuous and $\mathcal{B}(\mathcal{T}_f) = \mathcal{B}(\mathcal{T})$.*

Theorem A.10. *A measureable space $(X, \mathcal{A})$ is a standard Borel space if and only if it is isomorphic to a Borel subset of a Polish space.*

Proof. Let $(X, \mathcal{A})$ be isomorphic to a Borel subset B of the Polish space $(Y, \mathcal{T})$. By Theorem A.8, there exists a Polish topology $\mathcal{T}_B \supseteq \mathcal{T}$ s.t. B is clopen in $\mathcal{T}_B$ and $\mathcal{B}(\mathcal{T}) = \mathcal{B}(\mathcal{T}_B)$. As open subset of $(Y, \mathcal{T}_B)$, the space $(B, \mathcal{T}_B|_B)$ is again Polish and has the same Borelsets a $(B, \mathcal{T}|_B)$. □

Corollary A.11. *Every uncountable standard Borel space contains a homeomorphic copy of $\{0,1\}^{\mathbb{N}}$ and has therefore cardinality $2^{\aleph_0}$.*

Theorem A.12 (Borel Cantor Bernstein Schröder)**.** *Let X, Y be standard Borel spaces, $f : X \to Y$ and $g : Y \to X$ be Borel injections. Then there are Borel sets $A \subseteq X$ and $B \subseteq Y$ s.t. $f(A) = Y \setminus B$ and $g(B) = X \setminus A$. In particular, the mapping*

$$h : X \to Y : x \mapsto \begin{cases} f(x) & x \in A \\ g^{-1}(x) & x \notin A \end{cases}$$

is a Borel isomorphism from X to Y.

Theorem A.13 ([6, Theorem 14.12])**.** *Let X, Y be standard Borel spaces and $f : X \to Y$ a mapping. Then f is Borel if and only if* $\operatorname{graph}(f)$ *is Borel.*

In particular, if $f : X \to Y$ is Borel and bijective, then f is a Borel isomphism (i.e. f^{-1} is measureable).

Theorem A.14 ([6, Theorem 15.6], Borel isomorphism Theorem)**.** *Let X, Y be standard Borel spaces. Then X and Y are Borel isomorphic if and only if they have the same cardinality.*

Corollary A.15. *Let X be a standard Borel space. Then X is Borel isomorphic to exactly one of the following standard Borel spaces:*

- $\{1,\dots,n\}$ *with the σ-algebra $2^{\{1,\dots,n\}}$ for some $n \in \mathbb{N}$*
- $\mathbb{N}$ *with the σ-algebra $2^{\mathbb{N}}$*
- $[0,1]$ *with the Borel-σ-algebra induced by the standard topology*

A.2 Probability measures on Polish spaces

Let X be topological space. A Borel measure is a measure defined on the Borel-σ-algebra of X. We denote the space of Borel probability measures with $\mathcal{P}(X)$. We equip $\mathcal{P}(X)$ with the topology of the weak convergence of probability measures, i.e. a sequence $(\mu_n)_n$ in $\mathcal{P}(X)$ converges to $\mu \in \mathcal{P}(X)$ if and only if for any $f : X \to \mathbb{R}$ continuous and bounded $\int f d\mu_n \to \int f d\mu$.

Theorem A.16. *If X is Polish, then $\mathcal{P}(X)$ is Polish.*

We give equivalent characterizations for the Borel-σ-algebra on $\mathcal{P}(X)$:

Proposition A.17. *The the Borel-σ-algebra on $\mathcal{P}(X)$ is*

- *the σ-algebra generated by the mappings $\mu \mapsto \int f d\mu$, where f varies over all bounded real-valued Borel functions*
- *the σ-algebra generated by the mappings $\mu \mapsto \int f d\mu$, where f varies over all real-valued continuous bounded functions*
- *the σ-algebra generated by the mappings $\mu \mapsto \mu(B)$, where B varies over all Borel subsets of X*

Definition A.18. Let X be a standard Borel space and $\mu \in \mathcal{P}(X)$. μ is continuous if for all $x \in X : \mu(\{x\}) = 0$. We denote the set of continuous probability measures on X as $\mathcal{P}_c(X)$.

An immediate consequence of this definition is that all countable sets are null sets for continuous measures. Therefore, Corollary A.11 yields that $\mathcal{P}_c(X) \neq \emptyset$ implies that X has cardinality contiuum.

Note that the pushforward of a continuous measure under a bijection is again a continous measure.

Theorem A.19 ([6, Theorem 17.41], Isomorphism theorem for measures)**.** *Let X be a standard Borel space and $\mu \in \mathcal{P}_c(X)$. Then there is a Borel isomorphism $f : X \to [0,1]$ s.t. $f_*\mu = \lambda$, where λ denotes the Lebesgue measure on $[0,1]$.*

This obviously implies: If X and Y are standard Borel spaces, $\mu \in \mathcal{P}_c(X)$, $\nu \in \mathcal{P}_c(Y)$, then there exists a Borel isomorphism $f : X \to Y$ s.t. $f_*\mu = \nu$.

We close this section with a crucial compactness criterion for the spaces of probability measures.

Definition A.20. $M \subseteq \mathcal{P}(X)$ is called tight if for any $\varepsilon > 0$ there exists a compact set $K \subseteq X$ s.t. for all $\mu \in M$ it holds $\mu(K^c) < \varepsilon$.

Theorem A.21 (Prokhorov)**.** *$M \subseteq \mathcal{P}(X)$ is relatively compact if and only if it is tight.*

A.3 Disintegration and kernels

In this section we briefly discussion conditioning and kernels. The most important measure theoretic definition in this context is the following:

Definition A.22. Let X and Y be standard Borel spaces. A kernel from X to Y is a Borel measurable mapping from X to $\mathcal{P}(Y)$.

Using Proposition A.17 we see that this is equivalent to the following conditions, which are often stated as the definition of a kernel:

Definition A.23 (Alternative definition for kernels)**.** Let X and Y be standard Borel spaces. A kernel from X to Y is a mapping $\pi : X \times \mathcal{B}_Y \to [0,1]$ such that

- $\forall x \in X : \pi(x, \cdot)$ is a probability measure on Y,
- $\forall B \in \mathcal{B}_Y : x \mapsto \pi(x, B)$ is Borel measurable.

The following theorem can be regarded as a generalization of Fubini's theorem to measures on a product space that are not necessarily of product structure.

Theorem A.24 ([6, Example 17.35], Measure disintegration Theorem I)**.** *Let X and Y be standard Borel spaces and $\pi \in \mathcal{P}(X \times Y)$. Then there exists a Borel measurable mapping $X \ni x \mapsto \pi^x \in \mathcal{P}(Y)$ such that*

- $\pi^x(\{x\} \times Y) = 1$ *for* $\mathrm{pr}_{X*}\pi$*-almost all* $x \in X$
- $\int f d\pi = \int \left[\int f d\pi^x\right] d\mathrm{pr}_{X*}\pi(x)$ *for all* $f : X \to \mathbb{R}$ *bounded and Borel.*

Moreover, the mapping $x \mapsto \pi^x$ is $\mathrm{pr}_{X*}\pi$-a.s. unique.

A more general case is the following:

Theorem A.25 ([6, Example 17.35], Measure disintegration Theorem II)**.** *Let X and Y be standard Borel spaces, $f : X \to Y$ be Borel. Let $\mu \in \mathcal{P}(X)$ and $\nu := f_*\mu \in \mathcal{P}(Y)$. Then there exists a Borel measurable mapping $Y \ni y \mapsto \mu^y \in \mathcal{P}(X)$ such that*

- $\mu^y(f^{-1}(\{y\})) = 1$ *for* ν*-almost all* $y \in Y$
- $\int f d\mu = \int \left[\int f d\mu^y\right] d\nu(y)$ *for all* $f : X \to \mathbb{R}$ *bounded and Borel.*

Moreover, the mapping $y \mapsto \mu^y$ is ν-a.s. unique.

The probabilistic perspective of this is conditioning, which we introduce here very briefly:

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If $\mathcal{G} \subseteq \mathcal{F}$ is a sub-σ-algebra, $L^2(\Omega, \mathcal{G}, \mathbb{P}|_{\mathcal{G}})$ is a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Therefore there exists an orthogonal projection $\mathbb{E}^{\mathcal{G}}[\,\cdot\,] := \mathbb{E}[\,\cdot\,|\mathcal{G}] : L^2(\Omega, \mathcal{F}, \mathbb{P}) \to L^2(\Omega, \mathcal{G}, \mathbb{P}|_{\mathcal{G}})$. One can extend this operator uniquely to an operator $\mathbb{E}^{\mathcal{G}} : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to L^1(\Omega, \mathcal{G}, \mathbb{P}|_{\mathcal{G}})$:

Theorem A.26 ([5, Theorem 6.1], Existence and properties of conditional expectation)**.** *Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-field. Then there exits a unique linear operator $\mathbb{E}^{\mathcal{G}} : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to L^1(\Omega, \mathcal{G}, \mathbb{P}|_{\mathcal{G}})$ such that*

(i) $\mathbb{E}[(\mathbb{E}^{\mathcal{G}}\mathsf{X})\boldsymbol{1}_A] = \mathbb{E}[\mathsf{X}\boldsymbol{1}_A]$ *for all* $\mathsf{X} \in L^1(\Omega, \mathcal{F}, \mathbb{P}), A \in \mathcal{G}$

This operator has the following properties:

(ii) $\mathsf{X} \geq 0 \implies \mathbb{E}^{\mathcal{G}}\mathsf{X} \geq 0$

(iii) $\mathbb{E}|\mathbb{E}^{\mathcal{G}}\mathsf{X}| \leq \mathbb{E}|\mathsf{X}|$

(iv) $0 \leq \mathsf{X}_n \nearrow \mathsf{X} \implies \mathbb{E}^{\mathcal{G}}\mathsf{X}_n \nearrow \mathbb{E}^{\mathcal{G}}\mathsf{X}$

(v) $\mathbb{E}^{\mathcal{G}}(\mathsf{XY}) = \mathsf{X}\mathbb{E}^{\mathcal{G}}\mathsf{Y}$ *if* X *is* $\mathcal{G}$*-measurable*

(vi) $\mathbb{E}[\mathsf{X}\mathbb{E}^{\mathcal{G}}\mathsf{Y}] = \mathbb{E}[\mathsf{Y}\mathbb{E}^{\mathcal{G}}\mathsf{X}] = \mathbb{E}[(\mathbb{E}^{\mathcal{G}}\mathsf{X})(\mathbb{E}^{\mathcal{G}}\mathsf{Y})]$

(vii) $\mathbb{E}^{\mathcal{G}}\mathbb{E}^{\mathcal{H}}\mathsf{X} = \mathbb{E}^{\mathcal{G}}\mathsf{X}$

If X, Y are random variables, $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra and A a Borel set, we define $\mathbb{E}^{\mathsf{Y}}\mathsf{X} := \mathbb{E}[\mathsf{X}|\mathsf{Y}] := \mathbb{E}[\mathsf{X}|\sigma(\mathsf{Y})]$. Moreover, we define $\mathbb{P}^{\mathcal{G}}(\mathsf{X} \in A) := \mathbb{P}[\mathsf{X} \in A|\mathcal{G}] := \mathbb{E}^{\mathcal{G}}[\mathbf{1}_A]$ and $\mathbb{P}^{\mathsf{Y}}(\mathsf{X} \in A) := \mathbb{P}(\mathsf{X} \in A|\mathsf{Y}) := \mathbb{P}(\mathsf{X} \in A|\sigma(\mathsf{Y}))$.

Theorem A.27 ([5, Theorem 6.3], conditional distribution)**.** *Let X be a standard Borel space, Y be a measurable space, X be an X-valued random variable and Y be a Y-valued random variable. Then there exits a kernel μ from Y to X satisfying*

$$\mathbb{P}(\mathsf{X} \in B|\mathsf{Y}) = \mu(\mathsf{Y}, B) \qquad B \subseteq X \text{ Borel}$$

on a set with full measure w.r.t $\mathcal{L}(\mathsf{Y})$. This kernel is $\mathcal{L}(\mathsf{Y})$-a.s. unique.

Remark A.28. Theorem A.25 and A.27 are basically the same statement, but first is in measure theoretic notation and the second is in probabilistic notation.

We briefly discuss conditional independence.

Definition A.29. Let $\mathcal{G}_1, \ldots, \mathcal{G}_n, \mathcal{H} \subseteq \mathcal{F}$ be σ-algebras. $\mathcal{G}_1, \ldots \mathcal{G}_n$ are conditionally independent given $\mathcal{H}$ if and only if

$$\mathbb{P}^{\mathcal{H}}\left(\bigcap_{i=1}^{n} G_i\right) = \prod_{i=1}^{n} \mathbb{P}^{\mathcal{H}}(G_i) \qquad \text{for all } G_i \in \mathcal{G}_i.$$

Proposition A.30 ([5, Proposition 6.8])**.** *Let $\mathcal{G}_1, \mathcal{G}_2, \mathcal{H} \subseteq \mathcal{F}$ be σ-algebras. Then $\mathcal{G}_1$ is conditionally independent of $\mathcal{G}_2$ given $\mathcal{H}$ if and only if for all $G_2 \in \mathcal{G}_2$ it holds $\mathbb{P}^{\mathcal{G}_1,\mathcal{H}}(G_2) = \mathbb{P}^{\mathcal{H}}(G_2)$.*

Remark A.31. It is interesting to observe that conditional independence is a symmetric condition (in $\mathcal{G}_1$ and $\mathcal{G}_2$), but the equivalent condition given in Proposition A.30 is not symmetric. This yields the equivalence

$$[\forall G_2 \in \mathcal{G}_2 : \mathbb{P}^{\mathcal{G}_1,\mathcal{H}}(G_2) = \mathbb{P}^{\mathcal{H}}(G_2)] \iff [\forall G_1 \in \mathcal{G}_1 : \mathbb{P}^{\mathcal{G}_2,\mathcal{H}}(G_1) = \mathbb{P}^{\mathcal{H}}(G_1)],$$

which is essentially the reason why the conditions (i) and (iii) in Proposition 2.2 are equivalent.

A.4 An isomorphism theorem for kernels

The isomorphism theorem for measures (Theorem A.14) states that for any continuous measure on a standard Borel space X, there exists a bijection $f : X \to [0,1]$ s.t. $f_*\mu = \lambda$. The main goal of this section is to prove a parameterized version of that: If π is a kernel from X to Y then there exists a Borel isomorphism $f : X \times Y \to X \times [0,1]$ such that f lets the x-component fixed and for all $x \in X$ the mapping $f(x, \cdot)$ pushes π^x to λ.

Using the axiom of choice, one could choose for all $x \in X$ a Borel ismorphism f_x that pushes π^x to λ and define $f(x,y) := (x, f_x(y))$. However, there is no reason why this function f is measurable. Therefore, we repeat the construction given in the proof of the isomorphism theorem for measures given in [6, Theorem 17.41] in a way that is uniform for all x. A key role for this plays Theorem 2.4 from [7] because it ensures the existence of Borel isomorphisms that let the x-coordinate fixed under suitable conditions. We will first clarify what this means exactly:

For a set $A \subseteq X \times Y$ and $x \in X$ we define the x-section of A as $A_x := \{y \in Y : (x,y) \in A\}$.

Definition A.32. Let X, Y and Z be standard Borel spaces and let $B \subseteq X \times Y$ be Borel. A Borel parametrization of B is a Borel isomorphism $f : X \times Z \to B$ satisfying $f(\{x\} \times Z) = \{x\} \times B_x$ for all $x \in X$.

Loosely speaking, a Borel parametrization is a Borel isomorphism between B and an rectangular set, which acts within columns (i.e. lets the x-coordinate fixed). Of course, a necessary condition for the existence of a Borel parametrization is that all x-sections of B have the same cardinality. We are interested in the case, where X is uncountable and all x-sections of B are uncountable. The following theorem answers the question under which conditions Borel parametrizations exist.

Theorem A.33 ([7, Theorem 2.4])**.** *Let X and Y be uncountable standard Borel spaces and let $B \subseteq X \times Y$ be a Borel set with uncountable x-sections. Then the following are equivalent*

(i) B has a Borel parametrization.

(ii) There is a Borel set $M \subseteq B$ such that for all $x \in X$ the set M_x is compact and perfect.

(iii) There exits a kernel from X to Y such that for all $x \in X$ the measure μ^x is continuous and satisfies $\mu^x(B_x) = 1$.

Lemma A.34. *Let X and Y be Polish spaces, $C \subseteq X$ compact perfect and $f : X \to Y$ continuous and injective. Then $f(C)$ is compact perfect. In particular, homeomorphic copies of compact perfects sets are compact perfect.*

Proof. Clearly, $f(C)$ is compact and hence closed. Assume that $f(C)$ contains an isolated point y. Then there exits an open neighborhood U of the point y which satisfies $f(C) \cap (U \setminus \{y\}) = \emptyset$. Moreover, there exists an (unique) $x \in C$ such that $f(x) = y$. By the continuity of f there exists an open neighborhood V of the point x such that $f(V) \subseteq U$. Since f is injective this implies $f(V \setminus \{x\}) \subseteq U \setminus \{y\} \subseteq Y \setminus f(C)$ and again by the injectivity of f this implies $V \setminus \{x\} \subseteq X \setminus C$. Hence, x is an isolated point of C, which is a contradiction; so $f(C)$ does not have isolated points. □

Remark A.35. If $B \subseteq X \times Y$ is a Borel set with uncountable x-sections, Corollary A.11 implies that for all x the set B_x contains a homeomorphic copy of $\{0,1\}^{\mathbb{N}}$ and therefore a compact perfect set. Loosely speaking, the assertion (ii) in Theorem A.33 says the these perfect sets can be chosen in a uniform way.

We recall a few basic facts about the distribution function $F_\mu(t) := \mu([0,t])$ of a probability measure μ on $[0,1]$. The distribution function F_μ is increasing and satisfies $F_\mu(0) = 0, F_\mu(1) = 1$. Moreover, it is continuous if and only if μ is continuous. In this case it satisfies ${F_\mu}_* \mu = \lambda$ and it is easy to see that F_μ is strictly increasing if and only if every non-empty open interval has positive measure if and only if $\text{supp}(\mu) = [0,1]$. In this case F_μ is a continuous bijection between the compact Hausdorffspaces $[0,1]$ and $[0,1]$, hence a homeomorphism.

The following proposition implies that the function $(x,t) \mapsto F_{\pi^x}(t)$ is jointly measurable for kernels π s.t. π^x is continuous for all $x \in X$. (Of course the latter assumption is not necessary for the measurability, but we just need it in that case.)

Proposition A.36 ([1, Lemma 4.51])**.** *Let X, Y and Z be Polish spaces and $f : X \times Y \to Z$ be a function such that for all $x \in X$ the mapping $y \mapsto f(x,y)$ is continuous and for all $y \in Y$ the mapping $x \mapsto f(x,y)$ is Borel. Then the mapping $(x,y) \mapsto f(x,y)$ is Borel.*

Since, we want to construct a bijection the set, where the distribution functions of π^x are constant (hence not injective) are a problem and we will have to modify the function there. First, we observe that this set is Borel:

Lemma A.37. *Let $F : X \times [0,1] \to [0,1]$ a measurable mapping such that for all $x \in [0,1]$ the mapping $t \mapsto F(x,t)$ is monotone. Then the set*

$$M := \{(x,t) \in X \times [0,1] : \exists t' \neq t : F(x,t) = F(x,t')\}$$

is Borel and M_x is at most countable for all $x \in X$.

Proof. Denote the diagonal in $[0,1]^2$ as Δ and define for $n \in \mathbb{N}$ the sets

$$\begin{aligned} M_n^+ &= \{(x,t) \in X \times [0,(n-1)/n] : (F(x,t), F(x,t+1/n)) \in \Delta\}, \\ M_n^- &= \{(x,t) \in X \times [1/n,1] : (F(x,t), F(x,t-1/n)) \in \Delta\}. \end{aligned}$$

As preimages of Δ under the measureable mappings $(x,t) \mapsto (F(x,t), F(x,t+1/n))$ and $(x,t) \mapsto (F(x,t), F(x,t-1/n))$ these sets are obviously Borel. The montonicity of $t \mapsto F(x,t)$ implies

$$\begin{aligned} (x,t) \in M \iff \exists n \in \mathbb{N} \quad &\Big(t - 1/n \in [0,1] \ \wedge \ F(x,t) = F(x,t-1/n)\Big) \\ &\vee \Big(t + 1/n \in [0,1] \ \wedge \ F(x,t) = F(x,t+1/n)\Big), \end{aligned}$$

which shows that $M = \bigcup_{n \in \mathbb{N}} (M_n^+ \cup M_n^-)$. Hence, M is Borel.

M_x is at most countable because the preimage of a point under a monotone mapping is an interval and $[0,1]$ can contain at most countably many disjoint non-degenerate intervals. □

Proposition A.38. *Let X be a dense-in-itself Polish space and π a kernel from X to $[0,1]$ s.t. π^x is continuous for all $x \in X$. Then there is a Borel set $M \subseteq X \times [0,1]$ such that for all $x \in X$ the set M_x is compact perfect and satisfies $\pi^x(M_x) = 0$.*

Proof. Let $(U_n)_{n\in\mathbb{N}}$ be a base of the standard topology of $[0,1]$. Since the mappings $\mathcal{P}([0,1]) \to [0,1] : \mu \mapsto \mu(U_n)$ are Borel by Proposition A.17, the mapping

$$\Phi : \mathcal{P}([0,1]) \to \{0,1\}^{\mathbb{N}} : \mu \mapsto (z_n)_{n\in\mathbb{N}}, \text{ where } z_n = \begin{cases} 0 & \mu(U_n) > 0 \\ 1 & \mu(U_n) = 0 \end{cases}$$

is Borel as well. For $n \in \mathbb{N}$ define the set $A_n := \{z \in \{0,1\}^{\mathbb{N}} : z_0 = 0, \ldots, z_{n-1} = 0, z_n = 1\}$ and define A_∞ as the set, which only contains the constant zero sequence. Hence, for all $n \in \mathbb{N} \cup \{\infty\}$ the set A_n contains exactly the sequences $(z_k)_{k\in\mathbb{N}}$ that satisfy $\min\{k \in \mathbb{N} : z_k = 1\} = n$. This shows that $\{A_n : n \in \mathbb{N} \cup \{\infty\}\}$ is a partition of $\{0,1\}^{\mathbb{N}}$. It is easy to see that all these sets are closed and therefore Borel.

For $n \in \mathbb{N} \cup \{\infty\}$ define $B_n := (\Phi \circ \pi)^{-1}(A_n)$. (Recall that π is mapping $X \to \mathcal{P}([0,1])$.) Clearly, the sets B_n, $n \in \mathbb{N} \cup \{\infty\}$ are all Borel and form a partition of X. For $n \in \mathbb{N}$ and $x \in B_n$ it holds $U_n \subseteq [0,1] \setminus \operatorname{supp}(\pi^x)$. For $x \in B_\infty$ it holds $\operatorname{supp}(\pi^x) = [0,1]$.

For $n \in \mathbb{N}$ the open set U_n contains a nonempty open interval and therefore a non-degenerate closed interval C_n. Clearly, C_n is compact perfect.

The function

$$f : B_\infty \times [0,1] \to B_\infty \times [0,1] : (x,t) \mapsto (x, F_{\pi^x}(t))$$

is jointly measurable by Proposition A.36 and for all $x \in B_\infty$ the mapping $f(x,\cdot)$ is a homeomorphism between $\{x\} \times [0,1]$ and $\{x\} \times [0,1]$.

Denote $C_\infty \subseteq [0,1]$ the (usual) Cantor set. As isomorphic image of $\{0,1\}^{\mathbb{N}}$ it is clearly compact perfect. Define the set

$$M := f(B_\infty \times C_\infty) \cup \bigcup_{n\in\mathbb{N}} B_n \times C_n.$$

Clearly, M is Borel and any of its x-sections is a homeomorphic image of C_n for some $n \in \mathbb{N} \cup \{\infty\}$ and therefore compact perfect. $\square$

We are now able to state and prove the main theorem of this section.

Theorem A.39. *Let X and Y be standard Borel spaces and π a kernel from X to Y s.t. π^x is a continuous probability measure for all $x \in X$. Then there exists a measurable function*

$$G : X \times Y \to [0,1]$$

such that for all $x \in X$ the mapping $G^x := G(x,\cdot) : Y \to [0,1]$ is a Borel ismorphism satisfying $G^x_* \pi^x = \lambda$, where λ denotes the Lebesgue measure on $[0,1]$.

Proof. By the Borel isomorphism Theorem A.14 we can assume that $Y = [0,1]$. By Proposition A.36 the mapping

$$F : X \times [0,1] \to X \times [0,1] : (x,t) \mapsto (x, F_{\pi^x}(t))$$

is jointly measurable. By Lemma A.37 the set

$$N := \{(x,t) \in X \times [0,1] : \exists t' \neq t : F(x,t) = F(x,t')\}$$

is Borel and N_x is at most countable for all $x \in X$. The set $M := F^{-1}(N)$ is Borel and satisfies $\pi^x(M_x) = \pi^x([F(x,\cdot)]^{-1}(N_x)) = \lambda(N_x) = 0$ for all $x \in X$. Clearly, F is a bijection between $(X \times [0,1]) \setminus M$ and $(X \times [0,1]) \setminus N$.

By Proposition A.38 there exist Borel sets $A, B \subseteq X \times [0,1]$ such that for all $x \in X$ the sets A_x and B_x are compact perfect and it holds $\pi^x(A_x) = 0$ and $\lambda(B_x) = 0$. This implies that $\lambda([F(x,\cdot)](A_x)) = 0$ and $\pi^x([F(x,\cdot)]^{-1}(B_x)) = 0$.

Consider the sets $C := A \cup F^{-1}(B) \cup M$ and $D := F(A) \cup B \cup N$. Then it holds $\pi^x(C_x) \leq \pi^x(A_x) + \pi^x([F(x,\cdot)]^{-1}(B_x)) + \pi^x(M_x) = 0$ and $\lambda(D_x) \leq \lambda([F(x,\cdot)](A_x)) + \lambda(B_x) + \lambda(N_x) = 0$ for all $x \in X$ and F is a bijection between $(X \times [0,1]) \setminus C$ and $(X \times [0,1]) \setminus D$. Moreover, C and D both satisfy the assumptions of Theorem A.33, so there exist Borel parametrizations $f : X \times [0,1] \to C$ and $g : X \times [0,1] \to D$.

Clearly, the mapping

$$\widetilde{G} : X \times [0,1] \to X \times [0,1] : (x,t) \mapsto \begin{cases} F(x,t) & t \in [0,1] \setminus C_x \\ g(f^{-1}(x,t)) & t \in C_x \end{cases}$$

is a Borel parametrization of $X \times [0,1]$. Denote pr_2 the projection on the second component. It is easy to see that ${\mathrm{pr}_2}_* \widetilde{G}(x,\cdot)_* \pi^x = \lambda$ for all $x \in X$. Hence, the mapping $G := \mathrm{pr}_2 \circ \widetilde{G}$ has the desired properties. □

We state two corollaries that will be useful in this thesis.

Corollary A.40. *Let X, Y and Z be standard Borel spaces, μ a kernel from Z to X and ν a kernel from Z to Y s.t. μ^z and ν^z are continuous probability measures for all $z \in Z$. Then there exists a measurable function*

$$G : Z \times X \to Y$$

such that for all $z \in Z$ the mapping $G^z := G(z,\cdot) : X \to Y$ is a Borel ismorphism satisfying $G^z_* \mu^z = \nu^z$.

Proof. As we see at the end of the proof of Theorem A.39, there are Borel parametrizations $\widetilde{F} : Z \times X \to Z \times [0,1]$ and $\widetilde{H} : Z \times Y \to Z \times [0,1]$ such that ${\mathrm{pr}_2}_* \widetilde{F}(x,\cdot)_* \mu^x = \lambda$ and ${\mathrm{pr}_2}_* \widetilde{H}(x,\cdot)_* \nu^x = \lambda$ for all $z \in Z$. It is easy to see that $G := \mathrm{pr}_2 \circ \widetilde{H}^{-1} \circ \widetilde{F}$ has the desired properties. □

Corollary A.41. *Let X and Y be standard Borel spaces and π a kernel from X to Y. Then there exists a measurable function*

$$G : X \times Y \times [0,1] \to [0,1]^2$$

such that for all $x \in X$ the mapping $G^x := G(x,\cdot) : Y \times [0,1] \to [0,1]^2$ is a Borel ismorphism satisfying $G^x_*(\pi^x \otimes \lambda) = \lambda^2$, where λ^2 denotes the Lebesgue measure on $[0,1]^2$.

Proof. Note that $\pi^x \otimes \lambda$ is a continuous probability measure on $Y \times [0,1]$ for any $\pi^x \in \mathcal{P}(Y)$. Hence we can apply Theorem A.33 to standard Borel spaces X and $Y \times [0,1]$ and the kernel $(\pi^x \otimes \lambda)_{x \in X}$. □

At the end of this section we prove another technical proposition that we used a few times throughout this thesis.

Proposition A.42. *Let X and Z be Polish spaces, $\mathcal{M}$ be an at most countable partition of X consisting of Borel sets and μ be a kernel from Z to X s.t. μ^z is continuous for all $z \in Z$.*

Then there exists a measurable mapping $\Phi : Z \times X \to X \times [0,1]$ s.t. for all $z \in Z$ the mapping $\Phi^z := \Phi(z, \cdot) : X \to X \times [0,1]$ is a Borel isomorphism satisfying $\Phi^z_*(\mu^z|_M) = (\mu^z|_M) \otimes \lambda$ for all $M \in \mathcal{M}$.

Proof. Denote $\mathcal{M}' := \{M \in \mathcal{M} : M \text{ is uncountable}\}$ and $X' := \bigcup \mathcal{M}'$. For $M \in \mathcal{M}'$ define the kernel μ_M from Z to M as

$$\mu^z_M := \begin{cases} \frac{1}{\mu^z(M)} \mu^z|_M & \mu^z(M) > 0 \\ \rho_M & \mu^z(M) = 0, \end{cases}$$

where ρ_M is a fixed continuous probability measure on M (e.g. the pushforward of λ under a Borel isomorphism between $[0,1]$ and M that exists by Theorem A.14). By Theorem A.39 there exits a mapping $\Phi_M : Z \times M \to M \times [0,1]$ s.t. for all $z \in Z$ the mapping $\Phi^z_M := \Phi_M(z, \cdot) : M \to M \times [0,1]$ is a Borel isomorphism pushing μ^z_M to $\mu^z_M \otimes \lambda$.

Define the mapping

$$\widetilde{\Phi} : Z \times X' \to X' \times [0,1] : (z,x) \mapsto \Phi_M(z,x), \text{ if } M \text{ is the unique } M \in \mathcal{M}' \text{ s.t. } x \in M.$$

Clearly, $\widetilde{\Phi}$ is Borel as at most countable case-distinction of Borel maps and satisfies $\widetilde{\Phi}^z_*(\mu^z_M) = \mu^z_M \otimes \lambda$ for all $M \in \mathcal{M}'$.

Now, it remains to modify $\widetilde{\Phi}$ on nullsets to get a bijection Φ with the desired properties. By Proposition A.38 there are Borel sets $A \subseteq X$ and $B \subseteq X \times [0,1]$ such that for all $z \in Z$ the sets A_z and B_z are compact perfect and satisfy $\mu^z(A_z) = 0$ and $\mu^z \otimes \lambda(B_z) = 0$.

Define $C := A \cup \widetilde{\Phi}^{-1}(B) \cup X \setminus X'$ and $D := \widetilde{\Phi}(A) \cup B \cup Z \times (X \setminus X') \times [0,1]$. Clearly, C and D both satisfy the assumptions from Theorem A.33, so there exists a Borel isomorphism $\Psi : C \to D$ such that $\Psi(\{z\} \times C_z) = \{z\} \times D_z$ for all $z \in Z$. Denote $\mathrm{pr} : Z \times X \times [0,1] \to X \times [0,1]$ the projection. It is easy to check that the mapping

$$\Phi : Z \times X \to X \times [0,1] : (z,x) \mapsto \begin{cases} \mathrm{pr}(\widetilde{\Phi}(z,x)) & x \in X \setminus C_z \\ \mathrm{pr}(\Psi(z,x)) & x \in C_z \end{cases}$$

has the desired properties. □

A.5 Miscellaneous

This section contains a few more technical statements that we needed through out this thesis.

Lemma A.43 ([5, Lemma 1.13], Functional representation)**.** *Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be measurable spaces and Z be a standard Borel space. Let $f : X \to Z$ and $g : X \to Y$ be given functions.*

f is $\sigma(g)$-measurable if and only if there exits a measurable mapping $h : Y \to Z$ such that $f = h \circ g$.

Lemma A.44 ([5, Lemma 4.3])**.** *Let $f : X \to Y$ be continuous, then the mapping $\mathcal{P}(X) \to \mathcal{P}(Y) : \mu \mapsto f_*\mu$ is continuous w.r.t. weak convergence.*

$\mathbb{N}^{<\mathbb{N}}$ denotes the set of finite sequences of natural numbers. For $s \in \mathbb{N}^{<\mathbb{N}}$ denote $|s|$ the length of s. For $s_1, s_2 \in \mathbb{N}^{<\mathbb{N}}$ let $s_1 s_2$ be the conactenation of s_1 and s_2. Let $\emptyset$ be the empty sequence.

Theorem A.45 ([6, Theorem 13.9])**.** *Let X be a Polish space and d a compatible metric. Then there exists a collection $(A_s)_{s\in\mathbb{N}^{<\mathbb{N}}}$ of subsets of X, which has the following properties:*

(i) A_s *is Borel for all* $s \in \mathbb{N}^{<\mathbb{N}}$

(ii) $A_\emptyset = X$

(iii) $A_s = \bigcup_{i\in\mathbb{N}} A_{si}$ *for all* $s \in \mathbb{N}^{<\mathbb{N}}$

(iv) $A_{si} \cap A_{sj} = \emptyset$ *for all* $i \neq j \in \mathbb{N}$ *and* $s \in \mathbb{N}^{<\mathbb{N}}$

(v) $\mathrm{diam}(A_s) \leq 2^{-|s|}$ *for all* $s \in \mathbb{N}^{<\mathbb{N}} \setminus \{\emptyset\}$

Remark A.46. In particular, for each $n \in \mathbb{N}$ the collection $\{A_s : |s| = n\}$ is an at most countable bijection of X consisting of Borel sets with diameter at most 2^{-n}. Moreover, the partition $\{A_s : |s| = n+1\}$ is a refinement of the partition $\{A_s : |s| = n\}$.

Bibliography

[1] C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis - A Hitchhiker's Guide*. Springer Science Business Media, Berlin Heidelberg, 2007.

[2] J. Backhoff, M. Beiglböck, Y. Lin, and A. Zalashko. Causal transport in discrete time and applications. *SIAM J. Optim.*, 27(4):2528–2562, 2017.

[3] M. Beiglböck and D. Lacker. Denseness of adapted processes among causal couplings. *ArXiv e-prints*, 2020.

[4] W. Gangbo. The monge mass transfer problems and its applications. *NSF-CBMS Conference. Contemporary Mathematics, Vol*, 1999.

[5] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[6] A. S. Kechris. *Classical descriptive set theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

[7] R. D. Mauldin. Borel parametrizations. *Transactions of the American Mathematical Society*, 250:223–234, 1979.

[8] F. Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.

[9] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[10] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.