



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

DIPLOMARBEIT

Measurable selection in Optimal Transport and Skorokhod Embedding

Institut für
Finanz- und Versicherungsmathematik
Technische Universität Wien

Betreuer
Univ. Prof. Mathias BEIGLBÖCK

Autor
Daniel TONEIAN

Contents

1	Introducion	5
2	Definitions and Properties of Descriptive Set Theorey	7
3	Optimal Transport	11
4	Skorokhod Embedding	17
5	Barycenters	21
6	Regularity of Optimal Transport Prolems	25

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Introduction

In this paper, we are interested in problems of the following form: There is a space E with appropriate structure and a measurable function ψ on E , so that for each $\lambda \in E$ the value $\psi(\lambda)$ describes an optimization problem of some sort, which usually has more than one solution. We aim to find a measurable function ϕ on E , so that $\phi(\lambda)$ is a solution of $\psi(\lambda)$ for $\lambda \in E$. To this end, we use selection theorems.

Selection theorems are results from the field of descriptive set theory, which generally are of the following form: Given a set with a certain structure that can be interpreted as a multifunction, there is a subset with certain properties that can be interpreted as a function. Unfortunately, the resulting set usually cannot be assumed to be Borel, but at least it is analytical. This is, however, not very prohibitive, as analytical sets are still measurable for any Borel measure.

The problems we will discuss will mainly be different variations of optimal transport problems. We also show a theorem for selecting stopping-times for the Skorokhod-Embedding Problem and for finding Barycenters of measures. A number of results in this area have already been published, but have either unnecessarily restrictive assumptions [7], [11] or have a slightly different setting and prerequisite conditions that are relatively technical [6]. In this paper we try to present results that are sufficiently general so that they can be applied to many settings, and yet can be stated in terms that are easily understandable and applicable for people primarily focused on stochastics.

For two probability spaces (X, \mathcal{B}_X, μ) and (Y, \mathcal{B}_Y, ν) , a $\mathcal{B}_X \otimes \mathcal{B}_Y$ -measurable function c on

$X \times Y$ and a probability measure π on $\mathcal{B}_X \otimes \mathcal{B}_Y$ that has the marginals μ and ν , one can define the transportation cost $I(\pi)$ of π with respect to the cost-function c , namely

$$I(\pi) := \int c d\pi.$$

We interpret this as the cost that it takes to transport μ to ν according to the cost-function c . The Kantorovich Optimal Transportation Problem is concerned with finding the minimal cost of transporting μ to ν ,

$$K(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} I(\pi),$$

where $\Pi(\mu, \nu)$ denotes the set of all probability measures π so that π has the marginals μ and ν . If minimizers exist we call them optimal transport plans. The existence of minimizers is well studied and under broad assumptions guaranteed [13], [12].

Sometimes people do not want mass from μ to be split up in the process of transportation and thus are interested in the Monge Optimal Transportation problem, which is of the same kind as the Kantorovich problem, with the key difference that the infimum is taken over all measures π so that π has the correct marginals and that π can be written as $\pi = (Id \times T)\#\mu$, where T is a measurable function from X to Y and $\#$ denotes the push-forward.

The Skorochod-Embedding Problem concerns finding measurable stopping times τ for a given Brownian motion B and a probability distribution μ , so that the stopped process B_τ has the same distribution as μ . Solutions to this problem have been studied in the literature [3] and the solvability is guaranteed under rather weak assumptions.

In physics, the concept of barycenters has proven very useful. This concept of the center of mass can be generalized to finding the center of measures in an analogous fashion. These objects have been studied, for example, by [1] and [5], on whose results we build upon.

Furthermore we show that there are cost functions c , so that the minimal-transportation-value-map K defined above is not measurable. Hence, we cannot hope to find similar results as stated in this paper by merely requiring c to be measurable. This approach surprisingly leads to the discovery that any optimization problem with a measurable function to be minimized, is equivalent to a transport problem with a measurable cost function. We also show that optimizing over a lower-semicontinuous function is significantly harder than solving the transport problem for a lower-semiconinuous cost function.

Definitions and Properties of Descriptive Set Theory

Definition 2.0.1. *A separable, completely metrizable topological space is called a Polish space.*

In many fields of stochastics the concept of σ -algebras has proven very expedient, in particular for its stability under many setoperations. They leave, however, some useful properties to be desired, as for example the images of Borelsets for even basic functions, such as projections, are not guaranteed to be Borel. If collections of subsets are needed that are closed under operations of this kind, then the idea of analytic sets is useful. The proofs of most of the following theorems can be found in [9].

Definition 2.0.2. *Let X be a Polish space. A set $A \subseteq X$ is said to be analytic, if there is a Polish space Y and a continuous function $f : Y \rightarrow X$ so that $F(Y) = A$. The set of all analytic sets of a space X is denoted as $\Sigma_1^1(X)$.*

Remark 2.0.3. *If A is not empty, then the space Y can be chosen to be $\mathcal{N} = \mathbb{N}^{\mathbb{N}}$.*

Typically, there are more analytic sets than Borel sets, which can be seen in the next theorem.

Theorem 2.0.4. *Let X be an uncountable Polish space. Then $\mathcal{B}(X) \subsetneq \Sigma_1^1(X)$ holds.*

Definition 2.0.5. *Let X be a Polish space and $(P_s)_{s \in \mathbb{N}^{<\mathbb{N}}}$ a family of subsets called Suslin-scheme, which are indexed by $\mathbb{N}^{<\mathbb{N}}$, where $\mathbb{N}^{<\mathbb{N}}$ denotes the set of all finite sequences of \mathbb{N} . The Soulin Operator \mathcal{A} applied to such a family yields the set*

$$\mathcal{A}_s P_s = \bigcup_{x \in \mathcal{N}} \bigcap_n P_{x|n}.$$

Let Γ be any set of subsets of X , then $\mathcal{A}\Gamma$ is defined as the set of all $\mathcal{A}_s P_s$ for which $P_s \subset X$ is in Γ .

There are many equivalent definitions for analytic sets, the most important of which are:

Theorem 2.0.6. *Let X be a Polish space and $A \subseteq X$, then the following are equivalent:*

1. A is analytic.
2. There is a Polish space Y and a Borelset $B \subseteq X \times Y$ for which $A = \text{proj}_X(B)$.
3. There is a closed set $F \subseteq X \times \mathcal{N}$ so that $A = \text{proj}_X(F)$.
4. There is a Suslin-Scheme of closed sets so that $A = \mathcal{A}_s F_s$.

Analytic sets are stable under many operations.

Proposition 2.0.7. *Let X and Y be Polish spaces, $f : X \rightarrow Y$ Borel, $A_n \subseteq X$ as well as $A \subseteq X$ and $B \subseteq Y$ analytic then $\cup_{n \in \mathbb{N}} A_n, \cap_{n \in \mathbb{N}} A_n, f(A)$ and $f^{-1}(B)$ are, too.*

Analytic sets also have particular structural properties.

Theorem 2.0.8. *(Isomorphisms Theorem) Let X, Y be Polish spaces. Then there exists a Borel isomorphism between X and Y exactly if the cardinality of both spaces is the same. In particular, any two uncountable Polish spaces are Borel isomorphic.*

Definition 2.0.9. *Let X be a Polish space and a $A \subseteq X$. A set A is co-analytic if its complement is analytic. The set of co-analytic sets is denoted by Π_1^1 . Sets that are both analytic and co-analytic are called bi-analytic.*

Analytic sets have a close relationship to Borelsets.

Theorem 2.0.10. *Let X be a Polish space. Then the bi-analytic sets are exactly the Borelsets.*

Theorem 2.0.11. *Let X, Y be Polish spaces, and $f : X \rightarrow Y$. Then the following are equivalent:*

1. f is Borel.
2. $\text{graph}(f)$ is Borel.
3. $\text{graph}(f)$ is analytic.

In particular, Borel bijections are always also Borel isomorphisms.

Definition 2.0.12. Let X, Y be two sets and $P \subseteq X \times Y$, then a uniformisation of P is a subset $P^* \subseteq P$ so that for all $x \in X$, $\exists y P(x, y) \iff \exists ! y P^*(x, y)$.

Theorem 2.0.13. (Jankov, von Neumann uniformisation theorem) Let X, Y be Polish spaces and $P \subseteq X \times Y$ in Σ_1^1 . Then there exists a uniformisation P^* of P which is $\sigma(\Sigma_1^1)$ -measurable, that is, $P^{*-1}(B)$ is in the σ -algebra generated by the analytic sets on X for any Borel set $B \subseteq Y$.

Definition 2.0.14. Let X be a Hausdorff space. A capacity on X is a map $\gamma : \mathcal{P}(X) \rightarrow [0, \infty]$ so that:

1. $A \subseteq B \implies \gamma(A) \leq \gamma(B)$;
2. $A_0 \subseteq A_1 \subseteq \dots \implies (\gamma(A_n) \rightarrow \gamma(\bigcup_n A_n))$;
3. For any compact $K \subseteq X$ the condition $\gamma(K) < \infty$ holds; if furthermore $\gamma(K) < r$ for some $r \in \mathbb{R}$, then there is an open subset U which satisfies $U \supseteq K, \gamma(U) < r$.

Theorem 2.0.15. Let X be a Polish space and μ a finite Borel measure on X . Then the outer measure μ^* induced by μ is a capacity.

Definition 2.0.16. Let γ be a capacity on a Hausdorff space X . $A \subseteq X$ is called γ -capacitable if $\gamma(A) = \sup \{\gamma(K) : K \text{ compact}, K \subseteq A\}$. A is called universally capacitable if it is capacitable for any capacity γ .

Proposition 2.0.17. Let X be a Polish space and μ a finite Borel measure on X . If $\gamma = \mu^*$, then A is γ -capacitable exactly if it is μ -measurable.

Theorem 2.0.18. (Choquet Capacibility Theorem) Let X be a Polish space. Any analytic subset of X is universally capacitable.

Optimal Transport

Here we prove theorems about optimal transport problems. In this section, $\mathcal{P}(X)$ denotes the set of all Borel probability measures on a space X .

Lemma 3.0.1. *Let X_1, X_2, \dots, X_n be Polish spaces. The i -th marginal map $\pi_{\#}^i : \mathcal{P}(X_1 \times X_2 \times \dots \times X_n) \mapsto \mathcal{P}(X_i)$ is continuous for all i .*

Proof. Assume that μ_n converges weakly to $\mu \in \mathcal{P}(X_1 \times X_2 \times \dots \times X_n)$. The function $f \circ \pi_i : X_1 \times X_2 \times \dots \times X_n \mapsto \mathbb{R}$ is continuous and bounded for any bounded and continuous $f : X_i \mapsto \mathbb{R}$ and hence $(\pi_{\#}^i \mu_n)(f) = \mu_n(f \circ \pi_i) \mapsto \mu(f \circ \pi_i) = (\pi_{\#}^i \mu)(f)$. \square

Theorem 3.0.2. *Let X, Y be Polish spaces, $\mu \in \mathcal{P}(X \times Y)$, then there exists a measurable function $\phi : (\pi, x) \in (\mathcal{P}(X \times Y) \times X) \mapsto (\pi_x) \in \mathcal{P}(Y)$ where π_x is a disintegration of π with respect to $x \in X$.*

Proof. See [4, Proposition 7.27].

Theorem 3.0.3. *Let (E, Σ, m) be a σ -finite measurable space, X and Y Polish spaces, $c : X \times Y \mapsto [0, \infty]$ lower semi-continuous and consider a measurable function $\psi : \lambda \in E \mapsto (\mu_\lambda, \nu_\lambda) \in \mathcal{P}(X) \times \mathcal{P}(Y)$.*

If moreover for m -a.a. λ there exists a Monge solution to the optimal transport problem between μ_λ and ν_λ , then there exists a function $(\lambda, x) \mapsto T(\lambda, x)$ which is measurable with respect to $\Sigma \otimes \mathcal{B}(X)$ and such that

$$T(\lambda, x)$$

is an optimal Monge transport for m -a.a. λ .

Proof. Since c is assumed to be nonnegative and lower semi-continuous, the function $F : \pi \mapsto \int c d\pi$ is also lower semi-continuous [13, Lemma 4.3]. Consider the function

$$G : (\mu, \nu) \mapsto \inf_{\pi \in \Pi(\mu, \nu)} \int c d\pi,$$

which is clearly lower semi-continuous as well, since for any $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)$ and any sequence (μ_n, ν_n) that converges weakly to (μ, ν) , we can find a $\pi_n \in \Pi(\mu_n, \nu_n)$ that satisfies

$$\int c d\pi_n < \inf_{\pi \in \Pi(\mu_n, \nu_n)} \int c d\pi + \varepsilon.$$

$(\pi_n)_{n \in \mathbb{N}}$ has a subsequence that converges to a $\hat{\pi} \in \Pi(\mu, \nu)$ [13, Lemma 4.4]. Now we find

$$\inf_{\pi \in \Pi(\mu, \nu)} \int c d\pi < \int c d\hat{\pi} < \liminf_n \inf_{\pi \in \Pi(\mu_n, \nu_n)} \int c d\pi + \varepsilon$$

which proves that G is lower semi-continuous.

The function $H : (\mu, \nu, \sigma) \mapsto (\mu, \nu, \int c d\sigma)$ is Borel, since any lower semi-continuous function is Borel as well. This means that the set $A := \{(\mu, \nu, \pi) : \pi \text{ is an optimal transport plan with respect to } (\mu, \nu)\} = h^{-1}(\text{graph}(G))$ is also Borel and hence also the nonempty set $B := \{(\mu, \nu, \pi) : \exists T : X \rightarrow Y, \text{ so that } T \text{ is measurable and } \nu = T\#\mu, \pi = (x \mapsto (x, T(x)))\#\mu, \pi \text{ is an optimal transport plan with respect to } (\mu, \nu)\}$, which is the intersection of A and the set of triples $C := \{(\mu, \nu, \pi) : \exists T : X \rightarrow Y, \text{ so that } T \text{ is measurable and } \nu = T\#\mu\}$. C is Borel because the set of Monge-maps is Borel [2, Prop. 5.5] and, by Lemma 3.0.1 so is the function that maps a measure to its marginals. Now we can simply express the required function as a combination of simpler functions:

$$(\lambda, x) \xrightarrow{f} (\mu, \nu, x) \xrightarrow{g} (\pi, x) \xrightarrow{h} \pi_x \xrightarrow{i} \int y d\pi_x(y),$$

where f is defined by $f(\lambda, x) := (\psi(\lambda), x)$, g is obtained from B by the Jankov, von Neuman Uniformization theorem, since the resulting set can be interpreted as the graph of a total function and h is a disintegration as guaranteed by the previous theorem. Since all these simpler functions are at least analytic and analytic sets are universally capacitable by 2.0.18, we can modify it on a set with vanishing m -measure to obtain a $\Sigma \otimes \mathcal{B}(X)$ measurable function T so that m -a.a. $T(\lambda, x)$ is optimal. \square

If we assume a slightly stronger set of axioms than ZFC, we can drop our requirements for c as long as we know that solutions exist. Namely, we need to assume Σ_1^1 -determinacy. For the ramifications of this assumption, see [9].

Theorem 3.0.4. (Σ_1^1 – determinacy) *Let X and Y be Polish spaces and let (E, Σ, m) be a σ -finite measurable space and consider a measurable function $\lambda \in E \mapsto (\mu_\lambda, \nu_\lambda) \in \mathcal{P}(X) \times \mathcal{P}(Y)$. If for m -a.a. λ the optimal transport problem between μ and ν has a solution, then there exists a function $\lambda \mapsto \pi_\lambda$ which is measurable with respect to Σ and such that π_λ is an optimal transport plan with respect to the marginals $(\mu_\lambda, \nu_\lambda)$ for m -a.a. λ .*

If moreover for m -a.a. λ there exists a Monge solution to the optimal transport problem between μ_λ and ν_λ , then there exists a function $(\lambda, x) \mapsto T(\lambda, x)$ which is measurable with respect to $\Sigma \otimes \mathcal{B}(\mathbb{R}^d)$ and such that

$$T(\lambda, x)$$

is an optimal Monge transport for m -a.a. λ .

Proof. The set $A := \{(\mu, \nu, \pi) : \pi \text{ is an optimal transport plan with respect to } (\mu, \nu)\}$ is co-analytic, since we can write its complement as the union of the open set $\{(\mu, \nu, \pi) : \pi \text{ does not have the marginals } (\mu, \nu)\}$ and the analytic set $\{(\mu, \nu, \pi) : \pi \text{ has marginals } (\mu, \nu), \text{ but is not optimal}\} = \text{proj}_{1,2,3}\{(\mu, \nu, \pi, \sigma) : \pi \text{ and } \sigma \text{ have marginals } (\mu, \nu) \text{ and } \int c d\sigma < \int c d\pi\}$. By [9, 36.20] we see that A is universally measurable and we can proceed analogously to the proof of the previous theorem. \square

Remark 3.0.5. *From these two results we see that by assuming some regularity of the cost function c the requirement that μ does not give any mass to sets with Hausdorff dimension smaller or equal to $d - 1$ can be dropped in [7, Theorem 1.1].*

The following theorem, its Lemmas and its proof are based on [11] but in a slightly different setting, while the proof is virtually the same.

Lemma 3.0.6. *The marginaliser map $\Pi : \mathcal{P}(X_1 \times X_2 \times \cdots \times X_n) \mapsto \mathcal{P}(X_1) \times \mathcal{P}(X_2), \dots, \mathcal{P}(X_n)$ defined by*

$$\Pi(\mu) = (\pi_{\#}^1 \mu, \dots, \pi_{\#}^n \mu)$$

maps closed sets into Borel sets for any Polish spaces X_1, X_2, \dots, X_n .

Proof. Π is continuous by 3.0.1, and thus Borel as well. The spaces $\mathcal{P}(X_1 \times X_2 \times \cdots \times X_n)$ and $\mathcal{P}(X_1) \times \mathcal{P}(X_2), \dots, \mathcal{P}(X_n)$ are Polish. The preimage of Π for any singleton is compact. This means all requirements for [11, Lemma 3.8] are met and give the desired result. \square

Lemma 3.0.7. *Let (E, Σ) be a measurable space, X and Y Polish spaces and consider a measurable function $\psi : \lambda \in E \mapsto (\mu_\lambda, \nu_\lambda) \in \mathcal{P}(X) \times \mathcal{P}(Y)$. The map $F(\lambda) := \Pi(\psi(\lambda))$ is measurable as a set-valued mapping from E to $\mathcal{P}(X \times Y)$.*

Proof. We write $F(\lambda)$, that is the set of couplings of μ_λ and ν_λ , as the preimage of the previously defined marginaliser,

$$F(\lambda) = \Pi^{-1}(\{(\mu_\lambda, \nu_\lambda)\}).$$

$F(\lambda) \cap A \neq \emptyset$ is equivalent to $\mu \in F(\lambda)$ for some $\mu \in A$, that is, $\Pi(\mu) = (\mu_\lambda, \nu_\lambda)$ for some $\mu \in A$. Therefore, the set-valued inverse of F may be written as

$$F^{-1}(A) = \{\lambda \in E : F(\lambda) \cap A \neq \emptyset\} = \{\lambda \in E : (\mu_\lambda, \nu_\lambda) \in \Pi(A)\}.$$

By the previous Lemma $\Pi(A)$ is a Borel set in $X \times Y$ whenever $A \subset \mathcal{P}(X \times Y)$ is closed. Since the map $\lambda \mapsto (\mu_\lambda, \nu_\lambda)$ is assumed to be measurable, we may conclude that $F^{-1}(A)$ is a measurable subset of E for any closed $A \subset \mathcal{P}(X \times Y)$.

Theorem 3.0.8. *Let (E, Σ) be a measurable space, X and Y Polish spaces, $c : X \times Y \mapsto [0, \infty]$ lower semi-continuous and consider a measurable function $\psi : \lambda \in E \mapsto (\mu_\lambda, \nu_\lambda) \in \mathcal{P}(X) \times \mathcal{P}(Y)$. Then there exists a function $\lambda \mapsto \pi_\lambda$ which is measurable with respect to Σ and such that π_λ is an optimal transport plan with respect to the marginals $(\mu_\lambda, \nu_\lambda)$.*

Proof. Let $\Gamma(\lambda) = \Pi(\mu_\lambda, \nu_\lambda)$ for $\lambda \in E$. Due to the previous Lemma this map is measurable as a set-valued map and takes compact values. Define for each $q \in \mathbb{Q}$ the level sets $L_q := \{\mu \in \mathcal{P}(X \times Y) : \mu(c) \leq q\}$, which are closed since the map $\mu \mapsto \mu(c)$ is lower semi-continuous [13, Lemma 4.3]. The set-valued mapping

$$\hat{L}_q(\lambda) := \begin{cases} L_q \cap \Gamma(\lambda), & \text{if } L_q \cap \Gamma(\lambda) \neq \emptyset \\ \Gamma(\lambda), & \text{otherwise,} \end{cases} \quad (3.1)$$

has compact values. To see that it is also measurable as a set-valued map, consider for a closed set F

$$\begin{aligned} \hat{L}_q^{-1}(F) &= \{\lambda \in E : \hat{L}_q(\lambda) \cap F \neq \emptyset\} \\ &= \{\lambda \in E : \Gamma(\lambda) \cap L_q(\lambda) \cap F \neq \emptyset\} \cup \{\lambda : \Gamma(\lambda) \cap F \neq \emptyset, \Gamma(\lambda) \cap L_q = \emptyset\} \\ &= \Gamma^{-1}(F \cap L_q) \cup (\Gamma^{-1}(F) \setminus \Gamma^{-1}(L_q)), \end{aligned} \quad (3.2)$$

which is measurable as $\lambda \mapsto \Gamma(\lambda)$ is measurable. It is clear that

$$\Gamma_{opt}(\lambda) := \{\mu \in \Gamma(\lambda) : \mu(c) \text{ is minimal}\} = \bigcap_{q \in \mathbb{Q}} \hat{L}_q(\lambda).$$

Since $\Gamma_{opt}(\lambda)$ is a countable intersection of compact-valued mappings $\lambda \mapsto \hat{L}_q(\lambda)$, and because $\mathcal{P}(X \times Y)$ is a complete separable metric space, it follows [8, Theorem 3.1 and 4.1] that

it is measurable as a set-valued mapping. Because $\Gamma_{opt}(\lambda)$ is non-empty, compact-valued and measurable, we may apply the measurable selection theorem [10] to obtain the desired result. \square

Remark 3.0.9. *This result can easily be generalized. The first part of the theorem holds true in multi marginal optimal transport problems, with the same proof. We can also add countably many linear constraints in the form $\int f d\pi = 0$ for suitable f as long as we can guarantee that there still are solutions to the optimal transport problem under these constraints.*

Skorokhod Embedding

We can also apply a similar technique to get measurable selections for the Skorokhod-Embedding-Problem. First we recall the setting we are in and state some important definitions related to the problem. For a more complete discussion of the spaces involved and their properties see [3].

Let $\Omega = (\Omega, \mathcal{G}, (\mathcal{G}_t)_{t \geq 0}, \mathbb{P})$ be a stochastic basis, rich enough to support a Brownian motion \mathcal{B} and a uniformly distributed \mathcal{G}_0 -random variable, independent of \mathcal{B} . Let $\gamma : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \mapsto \mathbb{R}$ be a cost function, where $C_0(\mathbb{R}_+)$ denotes the set of continuous functions on \mathbb{R}_+ starting in 0.

For a measure μ on \mathbb{R} with mean 0 and finite second momentum, the Skorokhod Embedding Problem concerns the question of finding a stopping time τ , so that the stopped Brownian Motion \mathcal{B}_τ is distributed according to μ and $\mathbb{E}[\tau] < \infty$. Many of such stopping times can be constructed, and many canonical constructions will be minimizers of certain functions γ as defined above, namely

$$P_\gamma(\mu) = \inf \left\{ \int \gamma(\mathcal{B}, \tau) d\mathbb{P} : \tau \text{ is a stopping time} \right\}.$$

Here, we want to select minimizers for a given γ while μ changes in a measurable fashion.

We will also consider the space $C_0(\mathbb{R}_+)$ equipped with the topology of uniform convergence on compact sets. Elements of this space will be denoted by ω . The canonical process on this space will be denoted by $(B)_{t \geq 0}$, so, $B_t(\omega) = \omega_t$. Furthermore, let this space be equipped with the Wiener Measure \mathbb{W} . Let \mathcal{F}^0 be the natural filtration of $C_0(\mathbb{R}_+)$.

We want to show the following theorem:

Theorem 4.0.1. *Assume that $\gamma : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \mapsto \mathbb{R}$ is optional and lower semi-continuous and bounded from below in the sense that for some constants $a, b, c \in \mathbb{R}_+$*

$$-(a + bs + c \max_{r \leq s} \mathcal{B}_r^2) \leq \gamma(\mathcal{B}_t, s)$$

holds on $C_0(\mathbb{R}_+) \times \mathbb{R}_+$. Let (E, Σ, m) be a σ -finite measurable space and consider a measurable function $\psi : \lambda \in E \mapsto \mu_\lambda \in \mathcal{W}_2(\mathbb{R}) \cap \{\mu \in \mathcal{P}(\mathbb{R}) : \int x d\mu(x) = 0\}$. Then, there exists a function $\lambda \mapsto \tau_\lambda$ such that it is measurable with respect to the σ -algebra generated by convergence in probability and $m(d\lambda)$ -almost everywhere τ_λ is a minimizer of the functional

$$P_\gamma(\mu) = \inf \left\{ \int \gamma(\mathcal{B}, \tau) d\mathbb{P} : \tau \text{ is a stopping time} \right\}. \quad (4.1)$$

To facilitate this, we will shift to a different space.

Let S be the set of stopped Paths

$$S = \{(f, s) : f : [0, s] \text{ is continuous, } f(0) = 0\},$$

endowed with the polish topology generated by the metric

$$d_s((f, s), (g, t)) := \max(t - s, \sup_{0 \leq u \leq s} |f(u) - g(u)|, \sup_{s \leq u \leq t} |g(u) - f(s)|),$$

for $(f, s), (g, t) \in S, s \leq t$. This space is related to $C_0(\mathbb{R}_+) \times \mathbb{R}_+$ by the function

$$r : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \mapsto S, r(\omega, t) = (\omega|_{[0, t]}, t),$$

which is a continuous open mapping.

Furthermore, we need to define randomized stopping times, which are measures that are supposed to be interpreted as stopping times that do not stop at a single time, but rather have a distribution associated with them. For the following definition, let $\mathcal{P}^{\leq 1}(X)$ denote the set of all positive Borel measures on a space X , with a mass less or equal 1.

Definition 4.0.2. *We call a measure $\xi \in \mathcal{P}^{\leq 1}(C_0(\mathbb{R}_+) \times \mathbb{R}_+)$ a randomized stopping time, if $\text{proj}_{C_0(\mathbb{R}_+)}(\xi) \leq \mathbb{W}$ and the process $A_\omega^\xi(t) = \xi_\omega([0, t])$ is optional, where $(\xi_\omega)_{\omega \in C_0(\mathbb{R}_+)}$ is a disintegratoin of ξ in the first coordinate. We call the set of all randomized stopping times RST.*

Let $(\bar{C}_0(\mathbb{R}_+), \bar{\mathcal{F}}, (\bar{\mathcal{F}}_t)_{t \geq 0}, \bar{\mathbb{W}})$, where $\bar{C}_0(\mathbb{R}_+) = (C_0(\mathbb{R}_+) \times [0, 1], \bar{\mathbb{W}}(A_1 \times A_2) = \mathbb{W}(A_1)\mathcal{L}(A_2)$ (where \mathcal{L} denotes Lebesgue measure) and $\bar{\mathcal{F}}_0$ is the completion of $\mathcal{F}^0 \otimes \mathbb{B}([0, 1])$. We will

write $\bar{B} = (\bar{B}_t)_{t \geq 0}$ for the process given by $\bar{B}_t(\omega, u) = \omega_t$. Then, given a randomized stopping time ξ , the random time

$$\rho(\omega, u) := \inf\{t \geq 0 : \xi_\omega([0, t]) \geq u\}$$

defines an $\bar{\mathcal{F}}$ -stopping time, called the representation of ξ . Let T denote the projection

$$T : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \mapsto \mathbb{R}_+.$$

Definition 4.0.3. For a measure μ on \mathbb{R} with mean 0 and finite second moment, we denote by $\text{RST}(\mu)$ the set of ξ with representation ρ for which $B_\xi = \mu$, i.e. $\bar{B}_\rho \sim \mu$ and $\xi(T) = \bar{\mathbb{E}}[\rho] = \int x^2 \mu(dx)$.

Proof of 4.0.1. First, by [3, Lemma 3.2], we see that we can define γ^S on a S so that $\gamma = \gamma^S \circ r$. γ^S obviously satisfies the same lower bound and is lower semi-continuous.

Furthermore, by [3, Lemma 3.11] we see that we can work on $C_0(\mathbb{R}_+) \times \mathbb{R}$ now and consider randomized stopping times instead of regular stopping times. Observe that by the construction of the stopping time associated with a randomized stopping time in [3, Lemma 3.11] and [3, Lemma 3.15] we see that we can associate randomized stopping times on $C_0(\mathbb{R}_+) \times \mathbb{R}$ with stopping times on Ω in a measurable way.

We want to show that $P_\gamma^S(\mu)$ is lsc with respect to μ . To that end, consider a sequence $\mu_n \in \mathcal{W}^2$ with mean 0 that converges to a $\mu \in \mathcal{W}^2$ with mean 0. For each μ_n take $\xi_n \in \text{RST}(\mu_n)$ so that $\int \gamma^S \circ r(\omega, t) \xi_n(d\omega, dt) \leq P_\gamma^S(\mu) + \varepsilon$ for $\varepsilon \geq 0$. Since $\xi_n \in \text{RST}(\mu)$, we have

$$\bar{\mathbb{E}}[\rho_n] = \mathbb{E}_{\xi_n} T = \int x^2 d\mu_n(x) \rightarrow \int x^2 d\mu(x), \quad (4.2)$$

and thus $\sup_n \mathbb{E}_{\xi_n} T$ is bounded and hence $T(\xi_n)$ is tight. But the projection on the first coordinate is tight for all randomized stopping times and therefore $(\xi_n)_{n \in \mathbb{N}}$ is tight and $(\xi_n)_{n \in \mathbb{N}}$ converges weakly to a ξ . ξ is a randomized stopping time because the set of all randomized stopping times is closed [3, Corollary 3.10]. The calculation in 4.2 shows that $\xi(T) = \bar{\mathbb{E}}[\rho] = \int x^2 \mu(dx)$, and $B_\xi = \mu$ is obvious. Thus, $\xi \in \text{RST}(\mu)$. Now

$$P_\gamma^S(\mu) \leq \int \gamma^S \circ r(\omega, t) \xi(d\omega, dt) \leq \liminf_{n \rightarrow \infty} \int \gamma^S \circ r(\omega, t) \xi_n(d\omega, dt) \leq P_\gamma^S(\mu_n) + \varepsilon$$

holds since $\int \gamma^S \circ r(\omega, t) \xi(d\omega, dt)$ is lower semi-continuous [3, Theorem 4.1] and we see that $P_\gamma^S(\mu)$ is indeed lower semi-continuous.

We can now once more proceed just like in the proof of 3.0.3. □

Barycenters

In physics, the notion of barycenters proved very useful. Recall that the barycenter of a n objects r_1, r_2, \dots, r_n with the masses m_1, m_2, \dots, m_n is defined as

$$r_B = \sum_{i=1}^n \lambda_i r_i,$$

where $\lambda_i = \frac{m_i}{\sum_{i=1}^n m_i}$ or, equivalently, minimizers of

$$x \mapsto \sum_{i=1}^n \lambda_i (x - x_i)^2$$

and for continuous mass distributions,

$$r_B = \frac{1}{\int \rho(x) dx} \int x \rho(x) dx$$

where $\rho(x)$ denotes the density at the point x .

Analogous to the discrete case, we can define for measures $(\nu_i)_{i \in \{1, \dots, n\}} \in \mathcal{W}_2(\mathbb{R}^d)$ and weights $(\lambda_i)_{i \in \{1, \dots, n\}} \in \mathbb{R}, \lambda \geq 0, \sum_{i=1}^n \lambda_i = 1$ barycenters, that is, minimizers in $\mathcal{W}(\mathbb{R}^d)$ of the functional $J_{\nu_1, \dots, \nu_n, \lambda_1, \dots, \lambda_n}(\nu) = \sum_{i=1}^n \lambda_i \mathcal{W}_2^2(\nu_i, \nu)$, where $\mathcal{W}_2^2(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu, \nu)} \int |x - y|^2 d\gamma(x, y)$ denotes the squared Wasserstein metric and $\mathcal{W}_2(\mathbb{R}^d)$ the Wasserstein space of all probability measures μ for which $\int d(x_0, x)^2 \mu(dx)$ is finite. These objects have been studied for example by [1].

Theorem 5.0.1. *Let (E, Σ, m) be a σ -finite measurable space and consider a measurable function $\psi : \lambda \in E \mapsto (\nu_1, \dots, \nu_n, \lambda_1, \dots, \lambda_n) \in (\mathcal{W}_2(\mathbb{R}^d))^n \times (\mathbb{R})^n$. Then there exists*

a measurable function $\lambda \mapsto \nu_\lambda$ such that $m(d\lambda)$ -almost everywhere ν_λ is a barycenter of $(\nu_1, \dots, \nu_n, \lambda_1, \dots, \lambda_n)$.

Proof. The functional $J_{\nu_1, \dots, \nu_n, \lambda_1, \dots, \lambda_n}(\nu)$ is continuous in all its $2n+1$ arguments with respect to the topology of the Wasserstein space since the Wasserstein distance is a compatible metric for the topology. Let $(\mu_i)_{i \in \mathbb{N}} \in \mathcal{W}_2(\mathbb{R}^d)$ be a countable dense collection of measures of the separable space $\mathcal{W}_2(\mathbb{R}^d)$. Due to the measures being dense we can rewrite the function

$$f : (\nu_1, \dots, \nu_n, \lambda_1, \dots, \lambda_n) \rightarrow \inf_{\nu \in \mathcal{W}_2(S)} J_{\nu_1, \dots, \nu_n, \lambda_1, \dots, \lambda_n}(\nu)$$

as

$$(\nu_1, \dots, \nu_n, \lambda_1, \dots, \lambda_n) \rightarrow \inf_{i \in \mathbb{N}} J_{\nu_1, \dots, \nu_n, \lambda_1, \dots, \lambda_n}(\mu_i).$$

For any input of f there exists a minimizer, as has been shown in [1, Proposition 2.3]. f is Borel as it is the infimum of a countable number of continuous functions. Now the same chain of arguments as in 3.0.3 can be used to get the desired result. \square

We can also generalize barycenters with continuous mass distribution to measures, as was done by [5]. Let $\Gamma \in \mathcal{P}(\mathcal{P}(\mathbb{R}^n))$, then a barycenter to the distribution Γ is a minimizer of the functional

$$\mu \mapsto \int \mathcal{W}_2^2(\mu, \nu) d\Gamma(\nu).$$

Let $M^+(\Omega)$ denote the set of all Borel probability measures that are concentrated on a compact set $\Omega \subset \mathbb{R}^n$.

With the same argument as above, utilizing the separability of $\mathcal{P}(\mathbb{R}^n)$ we can show the following:

Theorem 5.0.2. *Let (E, Σ, m) be a σ -finite measurable space and consider a measurable function $\psi : \lambda \in E \mapsto \Gamma_\lambda \in \mathcal{P}(M^+(\Omega))$.*

There exists a function $\lambda \mapsto \mu_\lambda$ which is measurable with respect to Σ and such that μ_λ is a barycenter for m -a.a. Γ_λ .

Proof. To be able to use the same argument as above all we have to do is to check that solutions actually exist for any Γ_λ . Let ν^n be a minimizing sequence. Since they all are concentrated on a compact set, by Prokhorov's theorem, there is a subsequence that weakly converges to a measure ν , which, again, is concentrated on Ω . Now ν is a minimizer, since by Fatou's lemma

$$\int \frac{1}{2} \mathcal{W}_2^2(\mu_\theta, \nu) d\Gamma_\lambda(\theta) \leq \int \liminf_{n \rightarrow \infty} \frac{1}{2} \mathcal{W}_2^2(\mu_\theta, \nu^n) d\Gamma_\lambda(\theta) \leq \liminf_{n \rightarrow \infty} \int \frac{1}{2} \mathcal{W}_2^2(\mu_\theta, \nu^n) d\Gamma_\lambda(\theta).$$

Hence, ν indeed is a minimizer and the desired barycenter. \square

Regularity of Optimal Transport Problems

Theorem 6.0.1. *There exist measurable cost functions c on Polish spaces for which the associated transport problem does not depend measurably on the marginal distributions.*

Proof. Let $X := P \times [0, 1)$, where P is some uncountable Polish space – we may take $P = [0, 1)$.

Let $A \subseteq P \times [0, 1)$ be a measurable set whose projection onto P is not measurable. Such a set must exist, since any uncountable Polish space has analytic sets that are not Borel, analytic sets can be defined by such projections and $[0, 1)$ is Borel-isomorphic to \mathcal{N} by the isomorphisms theorem (2.0.8).

Define

$$c : X \times X \rightarrow \mathbb{R}$$

$$c(p, x, q, y) = \begin{cases} -1_A(p, (y - x) \bmod 1) & \text{if } p = q \\ 0 & \text{otherwise} \end{cases}.$$

We show that the value function of the optimal transport problem with c as cost function cannot depend measurably on the marginal measures, i.e. that

$$V : \mathcal{P}(X) \times \mathcal{P}(X) \rightarrow \mathbb{R}$$

$$V(\mu, \nu) := \inf_{\pi \in \text{Cpl}(\mu, \nu)} \int c \, d\pi$$

is not measurable.

More specifically, we show that $V \circ \iota$ is not measurable, where ι is the measurable map

$$\begin{aligned} \iota : P &\rightarrow \mathcal{P}(X) \times \mathcal{P}(X) \\ \iota(p) &:= \left((x \mapsto (p, x))_{\#}(\lambda), (x \mapsto (p, x))_{\#}(\lambda) \right) \end{aligned}$$

and λ is Lebesgue measure on $[0, 1)$.

$$(V \circ \iota)(p) = \inf_{\pi \in \text{Cpl}(\lambda, \lambda)} \int c(p, x, p, y) d\pi(x, y)$$

For any $z \in [0, 1)$ the measure

$$\pi^z := (x \mapsto (x, (z + x) \bmod 1))_{\#}(\lambda),$$

which puts mass only on the diagonal shifted by z , satisfies $\pi^z \in \text{Cpl}(\lambda, \lambda)$ and

$$\int c(p, x, p, y) d\pi^z(x, y) = -1_A(p, z).$$

If for $p \in P$ there is no $z \in [0, 1)$ s.t. $(p, z) \in A$, then $\int c(p, x, p, y) d\pi(x, y) = 0$ for all $\pi \in \text{Cpl}(\lambda, \lambda)$.

In summary,

$$(V \circ \iota)^{-1}(\{-1\}) = \{p : \exists z \in [0, 1) \text{ s.t. } (p, z) \in A\}$$

which by construction is not measurable. \square

Remark 6.0.2. *The idea of this proof can be further adapted to show that any optimisation problem $V'(p) = \inf_{x \in [0, 1)} G(p, x)$, where the argument is from some arbitrary space P and G is the function on $P \times [0, 1)$ to be minimized, is equivalent to an optimal transport problem, by changing the cost function in the previous proof to*

$$\begin{aligned} c : (P \times [0, 1)) \times (P \times [0, 1)) &\rightarrow \mathbb{R} \\ c(p, x, q, y) &= \begin{cases} G(p, y - x \bmod 1) & \text{if } p = q \\ \infty & \text{otherwise} \end{cases} \end{aligned}$$

Arbitrary lower semi-continuous target functions cannot, however, be reduced to optimal transport problems with lower semi-continuous cost functions. To see this, observe that $G(p, x) = -1_A$ where A is closed set on $P \times P$, for some Polish space P , whose projection onto P is not measurable (such sets exist, as can be seen in the proof of [9, 14.2], P could be chosen as \mathcal{N}). Since A is closed, G is lower semi-continuous and resulting function

$V'(p) = \inf_{x \in [0,1]} G(p, x)$ is not measurable. If there existed a cost function c that induces an equivalent optimal transport problem, then we could deduce that the associated optimal transport problem would be lower semi-continuous and hence measurable (see the proof of 3.0.3), which is a contradiction.

Bibliography

- [1] Martial Agueh and Guillaume Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.
- [2] Julio Backhoff Veraguas, Mathias Beiglböck, Manu Eder, and Alois Pichler. Fundamental properties of process distances. *arXiv preprint arXiv:1701.03955*, 2017.
- [3] Mathias Beiglböck, Alexander MG Cox, and Martin Huesmann. Optimal transport and skorokhod embedding. *Inventiones mathematicae*, 208(2):327–400, 2017.
- [4] Dimitir P Bertsekas and Steven Shreve. *Stochastic optimal control: the discrete-time case*. 2004.
- [5] Jérémie Bigot and Thierry Klein. Characterization of barycenters in the wasserstein space by averaging optimal transport maps. *ESAIM: Probability and Statistics*, 22:35–57, 2018.
- [6] Vladimir Bogachev and Il’ya Malofeev. Kantorovich problems and conditional measures depending on a parameter. *arXiv preprint arXiv:1904.03642*, 2019.
- [7] Joaquin Fontbona, Hélène Guérin, Sylvie Méléard, et al. Measurability of optimal transportation and strong coupling of martingale measures. *Electronic Communications in Probability*, 15:124–133, 2010.
- [8] Charles J Himmelberg and T Parthasarathy. Measurable relations. *Fund. math*, 87(1):53–72, 1975.
- [9] Alexander Kechris. *Classical descriptive set theory*, volume 156. Springer Science & Business Media, 2012.
- [10] Kazimierz Kuratowski and CZESŁAW Ryll-Nardzewski. A general theorem on selectors. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys*, 13(6):397–403, 1965.

- [11] Lasse Leskelä, Matti Vihola, et al. Conditional convex orders and measurable martingale couplings. *Bernoulli*, 23(4A):2784–2807, 2017.
- [12] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.
- [13] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.