University of Amsterdam

MSc Mathematics

Master Thesis

Stability and computation of martingale optimal transport

Author:
Hidde Fokkema

Supervisors:
dr. S.G. Cox (KdvI)
prof. dr. P.J.C. Spreij (KdvI)

Examination date:
June 30, 2021

Korteweg-de Vries Institute for
Mathematics

Abstract

In the classical theory of optimal transport, it is well known that the solution is stable with respect to weakly converging sequences. However, the issue of stability of the martingale optimal transport problem has long been unresolved. Knowing that the martingale optimal transport problem is stable is quite pressing, as it would make finding efficient numerical methods easier. Moreover, martingale optimal transport is widely used in the field of robust finance, for which stability is essential. Only recently, a proof of stability in the one dimensional case was independently given by Backhoff-Veraguas and Pammer [5], and Wiesel [55].

In this thesis we will investigate the former proof in great detail. The proof utilises many new methods and techniques to establish optimality. Most notably, in the classical theory of optimal transport it is possible to establish stability through the notion of 'c-cyclical monotonicity'. It was needed to generalise c-cyclical monotonicity and to consider the optimal transport problem not on the space $\mathbb{R} \times \mathbb{R}$, but on $\mathbb{R} \times \mathcal{P}_r(\mathbb{R})$ for $r \in [1, \infty)$. This leads to a 'weak' version of the martingale optimal transport problem for which we can define a new type of monotonicity, called martingale C-monotonicity. We will show that martingale C-monotonicity is stable with respect to converging sequences in the Wasserstein topology. This stability allows us then to derive stability of the martingale optimal transport problem. We will introduce all the new techniques, explain their motivation, and prove the results relevant to the proof of stability.

Finally, we will introduce a computational method for the martingale optimal transport problem developed by Guo and Obłój and discuss two explicit examples to showcase its effectiveness [28].

Title: Stability and computation of martingale optimal transport
Author: Hidde Fokkema, hidde.fokkema@student.uva.nl, 11010673
Supervisors: dr. S.G. Cox (KdvI), prof. dr. P.J.C. Spreij (Kdvi)
Second Examiner: dr. A. Khedher (KdvI)
Examination date: June 30, 2021

Korteweg-de Vries Institute for Mathematics
University of Amsterdam
Science Park 105-107, 1098 XG Amsterdam
http://kdvi.uva.nl

Acknowledgements

First and foremost I would like to thank my supervisors dr. Sonja Cox and dr. Peter Spreij. Not being able to meet in person because of the Covid pandemic was very unfortunate. However, they were able to make every meeting still fun and productive. So, meeting online did not feel like it held me back in the end and I have enjoyed the whole process because of them. Sonja and Peter also thoroughly proofread many draft versions of my thesis and gave me many tips to improve my writing, which is greatly appreciated. I also want to thank them for taking on the subject of stability of the martingale optimal transport problem. A subject we all were unfamiliar with and which proved to be a much more technical subject than we had suspected.

I also would like to give a spacial thanks to Julio Backhoff-Veraguas and Gudmund Pammer. They were very quick to respond whenever I had questions about their proof and also gave detailed answers. Their help greatly improved the arguments provided in this thesis.

Contents

Notation

Example(s)	Description
$\mathcal{X}, \mathcal{Y}$	Calligraphic capital letters will denote sets. In most cases, these sets will be Polish spaces.
Γ_x	Whenever $\Gamma \subseteq \mathcal{X} \times \mathcal{Y}$, the x in the subscript will indicate the fibre, i.e. $\Gamma_x = \{y \in \mathcal{Y} \mid (x, y) \in \Gamma\}$.
$B_R(x)$	Whenever $B_R(x) \subseteq \mathcal{X}$ and $\mathcal{X}$ is a metric space with metric d, then $B_R(x) = \{y \in \mathcal{X} \vert d(x, y) < R\}$, the open ball around x with radius R.
$C(\mathcal{X})$	The collection of all continuous functions from $\mathcal{X}$ to $\mathbb{R}$.
$C_b(\mathcal{X})$	The collection of all continuous and bounded functions from X to $\mathbb{R}$.
$L^\infty(\mathcal{X})$	The collection of all bounded measurable functions from $\mathcal{X}$ to $\mathbb{R}$.
$L^p(\mathcal{X})$	The collection of measurable functions, $f : \mathcal{X} \to \mathbb{R}$, with $\int_{\mathcal{X}} \Vert f(x) \Vert^p \mu(\mathrm{d}x) < \infty$ for $p \in [1, \infty)$.
$B(\mathcal{X})$	The collection of all bounded functions from $\mathcal{X}$ to $\mathbb{R}$.
$\sigma(\mathcal{I})$	If $\mathcal{I}$ is a set of sets, then this denotes the smallest sigma algebra containing $\mathcal{I}$. If $\mathcal{I}$ is a collection of functions, then it is the smallest sigma algebra making those functions measurable.
$\sigma(\mathcal{I})_b$	If $\mathcal{I}$ is a set of sets, then this denotes the set of bounded functions that are measurable with respect to the sigma algebra generated by $\mathcal{I}$. Similarly, if $\mathcal{I}$ is a collection of functions, then it is the set of bounded measurable functions with respect to the sigma algebra generated by those functions.
$\mathcal{B}(\mathcal{X})$	The Borel σ-algebra of $\mathcal{X}$.
$\mathcal{P}(\mathcal{X})$	The space of probability measures on $\mathcal{X}$.
$\mathcal{P}_r(\mathcal{X})$	The space of probability measures on $\mathcal{X}$ equipped with the Wasserstein-r topology. See Definition 1.5.3 for more information.
$\mathcal{W}_r(\mu, \nu)$	The Wasserstein-r distance between two measures $\mu, \nu \in \mathcal{P}_r(\mathcal{X})$. See Definition 1.5.1 for more information.
proj_i	Let $\mathcal{X} \times \ldots \times \mathcal{X}$ be some n-dimensional space. The function proj_i is the projection onto the i'th component.
$\Pi(\mu, \nu)$	The collection of all couplings between μ, ν, i.e. all probability measures $\mathbb{P} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with $\mathbb{P} \circ \mathrm{proj}_1^{-1} = \mu$ and $\mathbb{P} \circ \mathrm{proj}_2^{-1} = \nu$. See Definition 0.0.3 for more information.
$\mathcal{M}(\mu, \nu)$	The set of all martingale couplings between μ and ν. See Definition 2.1.1 for more information.

> The abundance of areas where optimal transport is relevant only reflects the fact that optimal transport is a simple, meaningful, natural and therefore universal concept.
>
> — Cedric Villani

Introduction

The subject of optimal transport has a long and rich history dating back to the mid Eighteenth century. Since then it has been touched by many great mathematicians and advances in the field were commonly followed by many prizes and accolades. A subject that once started by the simple question, "How can I move building material from its origin to the construction site as efficiently as possible", has since grown out to a theory that has touched many different areas of science. Some of those areas are geometry [23], fluid dynamics [45] and statistics [38]. More recently, machine learning [2, 3], computer vision [43] and finance [20, 29, 10] can be added to that list.

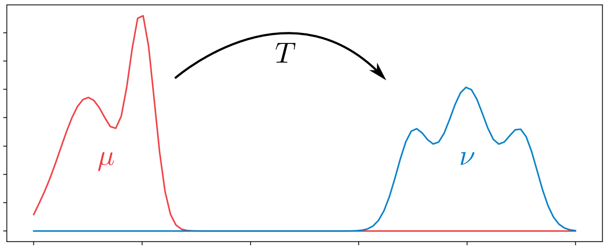


Figure 0.1.: Monge formulation of the OT problem where the measures μ and ν consist of densities.

The first notion of optimal transport was by Gaspard Monge in 1781 his paper, *Mémoire sur la théorie des déblais et des remblais*, loosely translated to *On the theory of extracting material and construction* [37]. Monge looked at the problem of extracting a certain amount of soil from a mine or quarry and transporting it to the construction site. The only thing that was known, was the amount of soil needed, the original form, and the resulting form. What needed to be determined is the mapping from the original form to its transported form. Mathematically, we can formalise this in the following way. The starting shape can be represented as a measure μ on a space called $\mathcal{X}$ and the resulting form can also be represented by another measure ν on another space called $\mathcal{Y}$. For each point $x \in \mathcal{X}$ that is moved to another point $y \in \mathcal{Y}$ we assign a cost $c(x, y)$. The problem then is to find a map $T : \mathcal{X} \to \mathcal{Y}$ that minimises

$$\int_{\mathcal{X}} c(x, T(x))\mu(\mathrm{d}x). \tag{0.1}$$

However, we cannot consider any map T. The map needs to send μ precisely to ν. This is formalised by saying that the *push-forward* measure $\mu \circ T^{-1}$ is equal to ν.

Definition 0.0.1 [Push-forward]. Let $\mathcal{X}, \mathcal{Y}$ be a Polish spaces, with probability measure $\mu \in \mathcal{P}(\mathcal{X})$ and a Borel measurable map $T : \mathcal{X} \to \mathcal{Y}$. We denote by $\mu \circ T^{-1}$ the *push-forward measure* of μ by T. It is defined as

$$\mu \circ T^{-1}(A) = \mu(T^{-1}(A)) \quad \forall A \in \mathcal{B}(\mathcal{X}).$$

Monge made many theoretical advances in the case where $\mathcal{X} = \mathcal{Y} = \mathbb{R}^3$ and the cost function was of the form $c(x, y) = |x - y|$. However, a general solution to the problem was not found. The reason being that the formulation Monge considered was very strict and difficult. In particular, the problem is non-linear in T and the set $\{T : \mathcal{X} \to \mathcal{Y} \,|\, \mu \circ T^{-1} = \nu\}$ is not well behaved. It is rarely compact for example. Furthermore, minimizing maps may not even exist!

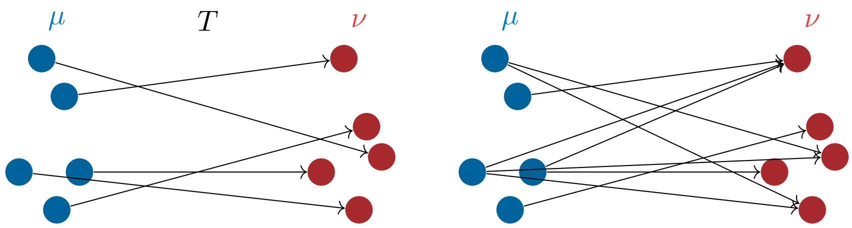


Figure 0.2.: Left: the Monge formulation of the optimal transport problem, Right: the Kantorovich formulation of the optimal transport problem. Note that in the Kantorovich case, points in the support of μ can be send to multiple target points. The measures μ and ν are discrete measures in both cases.

It took almost 200 years to make further significant progress when the Russian mathematician Leonid Vitaliyevich Kantorovich formulated a weaker version of the optimal transport problem in 1942 in his paper *On the transfer of mass* [33]. Instead of looking at maps that push μ on to ν, he considered product measures whose first marginal is equal to μ and their second marginal is equal to ν.

Definition 0.0.2 [Marginals]**.** Let $\mathcal{X}, \mathcal{Y}$ be Polish spaces, $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. Then, the *marginal* of π on $\mathcal{X}$ is defined as the measure given by $\pi \circ \mathrm{proj}_1^{-1}$, where $\mathrm{proj}_1 : \mathcal{X} \times \mathcal{Y} : (x, y) \mapsto x$, is the projection onto the first component. Similarly, we define the marginal of π on $\mathcal{Y}$, as $\pi \circ \mathrm{proj}_2^{-1}$.

Definition 0.0.3 [Couplings & transport plans]**.** Let $\mathcal{X}, \mathcal{Y}$ be two Polish spaces with probability measures $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$. We denote by $\Pi(\mu, \nu)$ the space of measures which have μ and ν as their marginals,

$$\Pi(\mu, \nu) := \{\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \,|\, \pi \circ \mathrm{proj}_1^{-1} = \mu,\ \pi \circ \mathrm{proj}_2^{-1} = \nu\}.$$

A measure $\pi \in \Pi\,(\mu, \nu)$ is called a *coupling* between μ and ν. In the literature π is also referred to as a *transport plan*.

Essentially, Kantorovich allowed each point in $\mathcal{X}$ to be split up and sent to different points in $\mathcal{Y}$, while in Monge's case a point in $\mathcal{X}$ can only be sent to one unique point in $\mathcal{Y}$. The resulting problem Kantorovich tried to solve is to find the value or the minimizing measure of

$$\inf_{\pi \in \Pi(\mu,\nu)} \int\limits_{\mathcal{X} \times \mathcal{Y}} c(x, y) \pi(\mathrm{d}x, \mathrm{d}y).$$

Kantorovich realised that one could turn this minimization problem into another maximization problem, called the *dual problem*. He needed quite technical results from functional analysis and convex analysis to show that this dual problem exists and is in fact equivalent to the original problem. This dual problem allowed him to solve the problem and devise numerical approximation schemes.

It took some time for this theory to reach the general public. The USSR was involved in an armed conflict with Germany at that time and Kantorovich work proved to be very helpful to the authorities. Kantorovich was also not unaware of the usefulness of his theoretical work and made sure that his work was only published in Russian in a limited capacity [17]. The way his theory was utilized was by finding an optimal assignment of food and supplies from storage units to the war front. It was of course in the USSR's best interest that this application was not found by their opponents. In the end, his results were published, which even resulted in him winning a Nobel prize for Economics together with the Dutch-American mathematician and economist Tjalling Koopmans.

After Kantorovich a whole sequence of influential mathematicians worked on this problem in different settings. There are too many to name them all in this introduction, see for example the two books written by Villani for a more comprehensive overview [53, 54], but we do present some of them to give an impression of the breadth of the different topics: Wilfrid Gangbo and Robert McCann studied the geometry of optimal transport [26], Yann Brenier was able to connect optimal transport with fluid dynamics [15] and Svetlozar Rachev and Ludger Rüschendorf used the theory of optimal transport to prove several limit theorems for random processes [42].

The application of optimal transport that is considered in this thesis first appeared at the start of the 2010's. Multiple articles appeared in the field of *robust finance* using the theory of optimal transport to derive bounds for prices of exotic options [25, 1, 10, 20]. The idea of robust finance is to find bounds for the prices of these financial products without imposing a model and was pioneered by Hobson in 1998 [30].

Consider the following setting: we have two probability measures μ and ν. The measure μ indicates the distribution of the price of a stock at time 0, which we will indicate with S_0. Likewise, the measure ν represents the distribution of the price of the stock some time T in the future, indicated by S_T. An exotic option is then a measurable function g of these two prices. In the non-robust setting, one would construct a model for the price process of $(S_t)_{t=0}^T$ that has μ and ν as their initial and terminal distribution. This process would then allow us to find a coupling π between μ and ν. From this measure we can calculate the price

$$P = \mathbb{E}_\pi[g(S_0, S_T)].$$

We want this price to be fair, i.e. someone that buys the option will on average not gain or lose any money on it. This means that the process $(S_t)_{t=0}^T$ should actually be a martingale. Thus, we impose the restriction on π that it should be a martingale measure.

Returning to the robust setting, we do not want to impose an explicit model for the process $(S_t)_{t=0}^T$, because this can lead to biases and prices that do not reflect the correct value of the process. The financial crisis of 2008 is an example where these kinds of models were not able to correctly calculate the prices of financial products. So, what is the next best thing we can do? Well, we can find bounds for the price, by looking at the worst and best case scenarios. This means that we look at all martingale measures π that are compatible with μ and ν. Call this set $\mathcal{M}(\mu, \nu)$ and consider

$$P_{\text{high}} = \sup_{\pi \in \mathcal{M}(\mu,\nu)} \mathbb{E}_\pi[g(S_0, S_T)], \quad P_{\text{low}} = \inf_{\pi \in \mathcal{M}(\mu,\nu)} \mathbb{E}_\pi[g(S_0, S_T)].$$

This looks very similar to the Kantorovich formulation of the optimal transport problem. The only difference is that we only look at martingale couplings instead of all couplings. Since the introduction of this *martingale optimal transport* problem it has become an interesting problem on its own [11, 21] and researchers started working on computational methods to efficiently calculate the optimizing measure [28], which we will call π^*. However, there was no guarantee that the mappings, $(\mu, \nu) \mapsto \pi^*$, $(\mu, \nu) \mapsto P_{\text{high}}$, and $(\mu, \nu) \mapsto P_{\text{low}}$ were continuous. This is a problem when deriving computational methods, because it complicates the question of convergence in numerical methods. However, two independent proofs by Gudmund Pammer and Julio Backhoff-Veraguas [5], and Johannes Wiesel [55], appeared in 2019 that showed that these mappings are indeed continuous in the case of $\mathcal{X} = \mathcal{Y} = \mathbb{R}$!

The objective of this thesis is to discuss the proof given by Julio Backhoff-Veraguas and Gudmund Pammer in detail. The methods used to show this result were all developed in the not too distant past and are relatively unknown. In this thesis we collected all the relevant methods needed for the proof. For the reader this should result in proof that is almost entirely self contained and where all big leaps are worked out carefully for the less experienced reader in this subject.

In Chapter 1 we will briefly go over the classical theory of optimal transport as formulated by Kantorovich. We will show that a minimizer always exists and in particular that finding the minimizer is a continuous operation under suitable conditions. Additionally, we will introduce the Wasserstein space, which is a space of measures, whose topology resembles the weak topology and is intimately related to optimal transport. In Chapter 2 we will introduce the martingale optimal transport problem formally. Likewise, we will show that a minimizer always exists under appropriate conditions. Afterwards, we will introduce the concepts of monotonicity and optimal transport in the space $\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$, that will be vital for the stability proof. Then, the stability results and their proofs are provided. Chapter 3 will expand on the monotonicity of the martingale optimal transport problem and we will discuss it in a more general form. Likewise, Chapter 4 will expand on topological features of the space of measures on $\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$. There, we will motivate the usefulness of the optimal transport problem on this new space and introduce the concept of so called adapted weak topologies. Finally, in Chapter 5 we will discuss a computational method for the martingale optimal transport problem and discuss some examples.

1. Optimal transport

This chapter serves as a crash course in the theory of optimal transport. We will introduce the problem formally and show several important results related to existence and stability of the solution of the problem. The proofs of these results will mostly be referenced, as they are all established results in the literature.

The topic of stability is closely related to a concept called *c-cyclical monotonicity*. The idea of monotonicity will prove vital when we discuss the martingale case of this problem in Chapter 2.

We will finish this crash course with the introduction of the Wasserstein space. This is a space of probability measures with a metric defined through an optimal transport problem. This space has many applications and is widely used in the theory of random processes and statistics. In the chapters to come, we will be discussing various convergence and topological results in the Wasserstein space. So, a good understanding of them is required.

1.1. Setting

In the sections to come, we will consider specific types of spaces. These are called Polish spaces and they are defined as follows.

Definition 1.1.1 [Polish space]**.** A *Polish* space is a separable completely metrizable topological space.

Lemma 1.1.2 [N-product of Polish space]**.** *If $\mathcal{X}$ is a Polish space, then the N-product space*

$$\mathcal{X}^N := \mathcal{X} \times \ldots \times \mathcal{X}$$

is also a Polish space.

Proof. The space $\mathcal{X}$ can be endowed with a compatible metric d. We can define a metric on $\mathcal{X}^N$

$$d^N(x,y) = \sum_{i=1}^{N} d(x_i, y_i), \quad x = (x_1, \ldots, x_N), y = (y_1, \ldots, y_N) \in \mathcal{X}.$$

It can easily be verified that this is indeed a metric and makes the space $\mathcal{X}^N$ complete. □

We will be working with probability measures defined on $\mathcal{X}$. The set of probability measures will be denoted by $\mathcal{P}(\mathcal{X})$. Unless stated otherwise, we will always consider probability measures defined on the Borel σ-algebra of a Polish space $\mathcal{X}$.

Definition 1.1.3 [Probability measures]**.** Let $\mathcal{X}$ be a polish space with the Borel σ-algebra $\mathcal{B}(\mathcal{X})$, we denote by $\mathcal{P}(\mathcal{X})$ all Borel probability measures on $\mathcal{X}$.

A specific type of measure that will return many times is the Dirac measure.

Definition 1.1.4 [Dirac measure]**.** Let $\mathcal{X}$ be a Polish space with Borel σ-algebra $\mathcal{B}(\mathcal{X})$, for $a \in \mathcal{X}$, we denote by δ_a the *Dirac measure* at the point a, which is defined by

$$\delta_a(B) = \begin{cases} 0 & a \notin B \\ 1 & a \in B \end{cases} \quad B \in \mathcal{B}(\mathcal{X}).$$

In particular, for $f \in C_b(\mathcal{X})$ this means that

$$\int\limits_{\mathcal{X}} f(x)\delta_a(\mathrm{d}x) = f(a).$$

1.2. Duality & Existence

We now have enough definitions at our disposal to introduce the optimal transport problem formally.

Definition 1.2.1 [Optimal transport problem]**.** Let $\mathcal{X}, \mathcal{Y}$ be two Polish spaces with probability measures $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ measurable with respect to the product σ-algebra on $\mathcal{X} \times \mathcal{Y}$. The function c is also commonly referred to as the *cost function*. The *optimal transport* problem is then defined by either finding the value or the minimizing coupling of the following quantity:

$$V(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int\limits_{\mathcal{X}\times\mathcal{Y}} c(x, y)\, \pi(\mathrm{d}x, \mathrm{d}y). \tag{OT}$$

One of the most important features of the optimal transport problem is that it allows a dual formulation. This dual formulation is of both theoretical and practicable interest. However, we will not deal with this notion often in this thesis and it is added for completeness of the theory only. The interesting result for the dual problem is that in most cases the optimal value for the dual problem agrees with the optimal value for the OT problem.

Definition 1.2.2 [Dual formulation OT problem]**.** Let $\mathcal{X}, \mathcal{Y}$ be two Polish spaces and $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ two probability measures and define

$$\Phi_c(\mu, \nu) := \left\{(\psi, \varphi) \in L^1(\mathcal{X}, \mu) \times L^1(\mathcal{Y}, \nu) \,|\, \varphi(y) - \psi(x) \leq c(x, y)\right\}.$$

The dual formulation for the OT problem in Definition 1.2.1 is

$$\sup_{(\psi,\varphi)\in\Phi_c(\mu,\nu)} \int\limits_{\mathcal{Y}} \varphi(y)\; \nu(\mathrm{d}y) - \int\limits_{\mathcal{X}} \psi(x)\, \mu(\mathrm{d}x).$$

A natural question to ask is: "Does there exist a minimizing coupling for the OT problem?" Luckily, this is the case. To see why this is the case, we first remark that the set of couplings $\Pi(\mu, \nu)$, as defined in Definition 0.0.3, is in fact compact.

Lemma 1.2.3 [Compactness of couplings]**.** *Let $\mathcal{X}, \mathcal{Y}$ be two Polish spaces with probability measures $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$. Then, the set $\Pi(\mu, \nu)$ is compact in the weak topology.*

Proof. See the beginning of the proof of Theorem 4.1 in [53]. □

In the proof of compactness of $\Pi(\mu,\nu)$ the author actually shows that $\Pi(\mu,\nu)$ is sequentially compact. This is enough to establish compactness, because the weak topology for probability measures is metrizable, using for example the *Lévy-Prokhorov metric*. The following theorem establishes that result.

Theorem 1.2.4 [Weak topology is metrizable]**.** *Let $(\mathcal{X}, d)$ be a metric Polish space and define for any set $A \subseteq \mathcal{X}$, the ϵ-neighbourhood,*

$$A^\epsilon := \{y \in S : d(x,y) < \epsilon \text{ for some } x \in A\}.$$

Furthermore, for two probability measures $\mu, \nu \in \mathcal{P}(\mathcal{X})$ we define the Lévy-Prokhorov metric *as*

$$\rho(\mu,\nu) := \inf\{\epsilon > 0 \,|\, \mu(A) \leq \nu(A^\epsilon) + \epsilon \text{ for all Borel sets } A\}.$$

Then, the metric ρ metrizes the weak topology on the set of probability measures.

Proof. See the proof of Theorem 11.3.3 on page 395 of [22]. □

Next, we show that sending a measure to the integrated cost value is at least lower semicontinuous in the weak topology.

Lemma 1.2.5 [Lower semicontinuity of the cost functional]**.** *Let $\mathcal{X}, \mathcal{Y}$ be two Polish spaces, and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ a lower semicontinuous functions with respect to the product topology. Let $h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ be an upper semicontinuous function with respect to the product topology, such that $c(x,y) \geq h(x,y)$. Let $\{\pi_k\}_{k\in\mathbb{N}}$ be a sequence of probability measures on $\mathcal{X} \times \mathcal{Y}$, converging weakly to some $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ in such a way that $h \in L^1(\pi_k) \cap L^1(\pi)$ for all $k \in \mathbb{N}$ and*

$$\lim_{k\to\infty} \int\limits_{\mathcal{X}\times\mathcal{Y}} h(x,y)\pi_k(\mathrm{d}x, \mathrm{d}y) = \int\limits_{\mathcal{X}\times\mathcal{Y}} h(x,y)\pi(\mathrm{d}x, \mathrm{d}y).$$

Then

$$\int\limits_{\mathcal{X}\times\mathcal{Y}} c(x,y)\pi(\mathrm{d}x, \mathrm{d}y) \leq \liminf_{k\to\infty} \int\limits_{\mathcal{X}\times\mathcal{Y}} c(x,y)\pi_k(\mathrm{d}x, \mathrm{d}y).$$

In particular, if c is non-negative, then $\pi \mapsto \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\pi(dx,dy)$ is lower semicontinuous on $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ in the weak topology.

Now that we know that the cost functional is lower semicontinuous and its domain is compact, it should be intuitive that a minimizer exists. The following result solidifies this belief.

Theorem 1.2.6 [Existence of an optimal coupling]**.** *Let $\mathcal{X}, \mathcal{Y}$ be two Polish spaces with two probability measures $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$; let $a : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ and $b : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ be two upper semicontinuous functions such that $a \in L^1(\mu), b \in L^1(\nu)$. Let $c : \mathcal{X} \times \mathcal{Y} \to [0,\infty]$ be a lower semicontinuous function, such that $c(x,y) \geq a(x) + b(y)$ for all $x \in \mathcal{X}, y \in \mathcal{Y}$. Then there exists a coupling $\pi^* \in \Pi(\mu,\nu)$ which attains the minimal value $V(\mu,\nu)$ of* (OT).

Proof. Let $\{\pi_k\}_{k\in\mathbb{N}}$ be a sequence of couplings between μ and ν such that

$$\lim_{k\to\infty} \int\limits_{\mathcal{X}\times\mathcal{Y}} c(x,y)\pi_k(\mathrm{d}x, \mathrm{d}y) = \inf_{\pi\in\Pi(\mu,\nu)} \int\limits_{\mathcal{X}\times\mathcal{Y}} c(x,y)\pi(\mathrm{d}x, \mathrm{d}y).$$

By compactness of $\Pi(\mu, \nu)$ in the weak topology , we can find a $\pi^* \in \Pi(\mu, \nu)$ such that $\pi_{k_n} \to \pi^*$ weakly for some subsequence $\{k_n\}_{k \in \mathbb{N}}$. We can use lower semicontinuity of the cost functional to show that π^* is actually the minimizing coupling. From Lemma 1.2.5 we can see that

$$\int\limits_{\mathcal{X}\times\mathcal{Y}} c(x,y)\pi^*(\mathrm{d}x, \mathrm{d}y) \leq \liminf_{n\to\infty} \int\limits_{\mathcal{X}\times\mathcal{Y}} c(x,y)\pi_{k_n}(\mathrm{d}x, \mathrm{d}y) = \inf_{\pi\in\Pi(\mu,\nu)} \int\limits_{\mathcal{X}\times\mathcal{Y}} c(x,y)\pi(\mathrm{d}x, \mathrm{d}y).$$

So, we see that π^* is minimizing. We are allowed to use Lemma 1.2.5, because $h(x, y) = a(x) + b(y)$ lies in $L^1(\pi)$ and $L^1(\pi_k)$. Furthermore, we have that $c(x, y) \geq a(x) + b(x) = h(x, y)$ by assumption. Finally, we have

$$\lim_{k\to\infty} \int\limits_{\mathcal{X}\times\mathcal{Y}} h(x,y)\pi_k(\mathrm{d}x, \mathrm{d}y) = \lim_{k\to\infty} \int\limits_{\mathcal{X}} a(x)\mu(\mathrm{d}x) + \int\limits_{\mathcal{Y}} b(y)\nu(\mathrm{d}y) = \int\limits_{\mathcal{X}\times\mathcal{Y}} h(x,y)\pi(\mathrm{d}x, \mathrm{d}y).$$

□

1.3. c-cyclical monotonicity

We are now in the position to introduce c-cyclical monotonicity, a definition that was alluded to multiple times before.

Definition 1.3.1 [Cyclical monotonicity]**.** Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ be a function. A subset $\Gamma \subseteq \mathcal{X} \times \mathcal{Y}$ is said to be *c-cyclically monotone* if, for any $N \in \mathbb{N}$, and any family $(x_1, y_1), \ldots, (x_N, y_N)$ of points in Γ, the inequality

$$\sum_{i=1}^{N} c(x_i, y_i) \leq \sum_{i=1}^{N} c(x_i, y_{i+1})$$

holds, with the convention that $y_{N+1} = y_1$. A transport plan is said to be *c-cyclically monotone* if it is concentrated on a c-cyclically monotone set.

The intuition behind this definition is the following and, unsurprisingly, finds it origin in optimal transport. Assume that we have found some minimizing coupling called π and we look at its support. Further assume that π is a discrete measure with uniform weights. We can pick N points $(x_1, y_1), \ldots,$
(x_N, y_N) from it support. If we now cyclically permute the y component of these points, as in the definition of c-cyclical monotonicity, we get the points $(x_1, y_2), \ldots, (x_N, y_1)$. If we replace the original points with these new points in the support of π, then we get a new coupling $\widetilde{\pi}$. As this is not the optimizing measure π it should be obvious that the new points cannot perform better with respect to the cost function, which is precisely what happens in the definition of c-cyclical monotonicity. If one treats this example formally, it can be shown that c-cyclical monotonicity is a necessary condition for optimality in the OT problem. What is maybe even more surprising, is that it is even sufficient for optimality!

Theorem 1.3.2. *Let $\mathcal{X}, \mathcal{Y}$ be two Polish spaces and let $c : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ be a lower semicontinuous function such that*

$$c(x, y) \geq a(x) + b(y) \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}$$

for real-valued upper semicontinuous functions $a \in L^1(\mu)$ and $b \in L^1(\nu)$. If the optimal cost $V(\mu, \nu)$ is finite, then there exists a measurable c-cyclically monotone set $\Gamma \subseteq \mathcal{X} \times \mathcal{Y}$ such that for any $\pi \in \Pi(\mu, \nu)$ the following statements are equivalent:

(a) π *is optimal;*
(b) π *is* c*-cyclically monotone;*
(c) π *is concentrated on* Γ.

If we further assume that c *is continuous, then we can ensure that* $\Gamma = \operatorname{supp}(\pi)$

Proof. See the proof of Theorem 5.11 on page 57 in [53]. □

1.4. Stability of the optimizer

One might think that c-cyclical monotonicity is nothing more than a funny consequence of optimality. However, it is precisely the reason why the solution of the OT problem is stable! It turns out that the notion of c-cyclical monotonicity is itself stable, which allows us to show stability of the optimizer.

We can now state the stability result for the OT problem. It shows that the value $V(\mu,\nu)$ and its optimizer, if a unique optimizer exists, acts continuously with respect to (μ,ν). This is precisely the type of result that was missing for the martingale optimal transport problem, until recently. In Chapter 2 we will introduce this important new result, which establishes stability for the martingale case of the OT problem.

Theorem 1.4.1 [Stability of Optimal transport]**.** *Let* $\mathcal{X}$ *and* $\mathcal{Y}$ *be Polish spaces, and let* $c : \mathcal{X} \times \mathcal{Y} \to [0,\infty)$ *be a continuous cost function. Let* $\{c_k\}_{k\in\mathbb{N}}$ *be a sequence of continuous cost functions converging uniformly to* c *on* $\mathcal{X} \times \mathcal{Y}$*. Let* $\{\mu_k\}_{k\in\mathbb{N}}$ *and* $\{\nu_k\}_{k\in\mathbb{N}}$ *be sequences of probability measures on* $\mathcal{X}$ *and* $\mathcal{Y}$ *respectively. Assume that* $\{\mu_k\}_{k\in\mathbb{N}}$ *and* $\{\nu_k\}_{k\in\mathbb{N}}$ *converge to* μ *and* ν*, respectively, in the weak topology. For each* $k \in \mathbb{N}$*, let* π_k *be an optimal coupling between* μ_k *and* ν_k*. If*

$$\forall k \in \mathbb{N}, \qquad \int_{\mathcal{X}\times\mathcal{Y}} c_k(x,y)\pi_k(\mathrm{d}x,\mathrm{d}y) < \infty,$$

then, taking subsequences and relabelling if necessary, π_k *converges weakly to some* c*-cyclically monotone coupling* $\pi \in \Pi(\mu,\nu)$*. If moreover,*

$$\liminf_{k\in\mathbb{N}} \int_{\mathcal{X}\times\mathcal{Y}} c_k(x,y)\pi_k(\mathrm{d}x,\mathrm{d}y) < \infty,$$

then the optimal transport cost $V(\mu,\nu)$ *is finite, and* π *is an optimal coupling.*

Proof. See the proof of Theorem 5.20 on page 77 in [53]. □

1.5. Wasserstein space

The final section of this chapter will serve as an introduction to the Wasserstein space. This space of probability measures is a natural consequence of the optimal transport problem with the compatible metric d as its cost function. The topology that is generated is almost similar to the weak topology, but is slightly stronger. However, it allows for a metrizable topology that is even separable. First, we need to introduce the metric on this space.

Definition 1.5.1 [Wasserstein distance]**.** Let $(\mathcal{X},d)$ be a Polish metric space, and let $r \in [1,\infty)$. For any two probability measures μ,ν on $\mathcal{X}$, the *Wasserstein distance of order* r (also denoted as the Wasserstein-r distance or $\mathcal{W}_r$-distance) between μ and ν is defined by the formula

$$\mathcal{W}_r(\mu,\nu) = \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d(x,y)^r \, \pi(\mathrm{d}x,\mathrm{d}y) \right)^{\frac{1}{r}}.$$

Proposition 1.5.2 [Wasserstein distance]. *Let $(\mathcal{X}, d)$ be a Polish metric space, $r \in [1, \infty)$. Then, the Wasserstein distance of order r is a metric.*

Proof. See the proof after Definition 6.1 on page 94 in [53]. □

Definition 1.5.3 [Wasserstein space]. Let $(\mathcal{X}, d)$ be a Polish metric space, $r \in [1, \infty)$ and $x_0 \in \mathcal{X}$, the *Wasserstein space of order r* (also denoted the Wasserstein-r space) is defined as

$$\mathcal{P}_r(\mathcal{X}) := \left\{ \mu \in \mathcal{P}(\mathcal{X}) \,\middle|\, \int_{\mathcal{X}} d(x_0, x)^r \, \mu(\mathrm{d}x) < \infty \right\}.$$

Proposition 1.5.4. *Let $(\mathcal{X}, d)$ be a Polish metric space, $r \in [1, \infty)$, The space $\mathcal{P}_r(\mathcal{X})$ does not depend on the choice of the point x_0 and $\mathcal{W}_r$ defines a finite metric on $\mathcal{P}_r(\mathcal{X})$.*

Proof. See the proof after Definition 6.4 on page 95 in [53]. □

We can argue that $\mathcal{W}_r$ is a natural metric to consider on the space $\mathcal{P}_r(\mathcal{X})$. Namely, consider two points $x, y \in \mathcal{X}$ with their Dirac measures δ_x and δ_y. The only coupling between those measures is the product measure, which results in the following Wasserstein distance

$$\mathcal{W}_r(\delta_x, \delta_y) = \left(\int_{\mathcal{X} \times \mathcal{X}} d(x', y')^r \delta_x(\mathrm{d}x') \delta_y(\mathrm{d}y') \right)^{\frac{1}{r}} = d(x, y).$$

We can see from this example, that $\mathcal{W}_r$ allows us to isometrically embed $\mathcal{X}$ into $\mathcal{P}_r(\mathcal{X})$ through the map $x \mapsto \delta_x$.

The topology that is created using this metric can be tied to the weak topology of probability measures in the following sense.

Theorem 1.5.5. *Let $(\mathcal{X}, d)$ be a Polish metric space, $r \in [1, \infty)$, $\mu \in \mathcal{P}_r(\mathcal{X})$ and a sequence $\{\mu_k\}_{k \in \mathbb{N}}$ with $\mu_k \in \mathcal{P}_r(\mathcal{X})$ for all $k \in \mathbb{N}$, then the following statements are equivalent*

(i) $\mathcal{W}_r(\mu_k, \mu) \to 0$;

(ii) $\mu_k \to \mu$ *weakly, and* $\displaystyle\int_{\mathcal{X}} d(x_0, x)^r \, \mu_k(\mathrm{d}x) \to \int_{\mathcal{X}} d(x_0, x)^r \, \mu(\mathrm{d}x)$;

(iii) $\mu_k \to \mu$ *weakly, and* $\displaystyle\limsup_{k \to \infty} \int_{\mathcal{X}} d(x_0, x)^r \, \mu_k(\mathrm{d}x) \leq \int_{\mathcal{X}} d(x_0, x)^r \, \mu(\mathrm{d}x)$;

(iv) $\mu_k \to \mu$ *weakly, and* $\displaystyle\lim_{R \to \infty} \limsup_{k \to \infty} \int_{\mathcal{X} \setminus B_R(x_0)} d(x_0, x)^r \, \mu_k(\mathrm{d}x) = 0$;

(v) For all continuous functions φ for which there exists a $C \in \mathbb{R}$ such that $|\varphi(x)| \leq C(1 + d(x_0, x)^r)$, one has

$$\int_{\mathcal{X}} \varphi(x) \mu_k(\mathrm{d}x) \to \int_{\mathcal{X}} \varphi(x) \mu(\mathrm{d}x).$$

Proof. See the proof of Theorem 6.9 on page 101 in [53] □

In some cases we will look at product measures, whose components converge in the $\mathcal{W}_r$-topology. The following lemma tells us that the product measure also converges in that case.

Lemma 1.5.6. *Let $(\mathcal{X}, d)$ be a Polish metric space and $\{p_k\}_{k\in\mathbb{N}}$ with $p_k, p \in \mathcal{P}_r(\mathcal{X})$ and $p_k \to p$ in the $\mathcal{W}_r$-topology. Consider the N-product measure $p_k^{\otimes N}$ with $N \in \mathbb{N}$. Then it is also true that $p_k^{\otimes N} \to p^{\otimes N}$ in the $\mathcal{W}_r^N$-topology. In particular, $p_k^{\otimes N} \to p^{\otimes N}$ weakly as well.*

Proof. By Lemma 1.1.2, we already know that $\mathcal{X}^N$ is a Polish space with metric d^N. We see, where for convenience we have set $x = (x_1, \dots, x_N)$ and $y = (y_1, \dots, y_N)$,

$$\begin{aligned}
\mathcal{W}_r^N(p_k^{\otimes N}, p^{\otimes N}) &= \inf_{\gamma\in\Pi(p_k^{\otimes N}, p^{\otimes N})} \int\limits_{\mathcal{X}^N\times\mathcal{X}^N} d^N(x,y)\gamma(\mathrm{d}x, \mathrm{d}y) \\
&\leq \sum_{i=1}^N \inf_{\pi\in\Pi(p_k,p)} \int\limits_{\mathcal{X}\times\mathcal{X}} d(x_i,y_i)\pi(\mathrm{d}x_i, \mathrm{d}y_i) \\
&= N\mathcal{W}_r(p_k, p) \to 0.
\end{aligned}$$

□

We conclude this section with the fact that the Wasserstein space is itself a Polish space, whenever the underlying domain is Polish.

Theorem 1.5.7 [Topology of the Wasserstein space]**.** *Let $(\mathcal{X}, d)$ be a Polish metric space, $r \in [1, \infty)$, then the Wasserstein space $(\mathcal{P}_r(\mathcal{X}), \mathcal{W}_r)$ is also a Polish metric space. Moreover, any probability measure $p \in \mathcal{P}_r(\mathcal{X})$ can be approximated in $\mathcal{W}_r$-topology by a sequence of probability measures $\{p_k\}_{k\in\mathbb{N}}$ with $p_k \in \mathcal{P}_r(\mathcal{X})$ and $|\mathrm{supp}(p_k)| < \infty$ for all $k \in \mathbb{N}$.*

Proof. See the proof of Theorem 6.18 on page 105 in [53]. □

2. Martingale Optimal Transport

In the previous chapter, the classical OT problem was introduced. Moreover, we illustrated how optimality of a coupling π can be read of a set Γ with $\pi(\Gamma) = 1$, by checking if Γ is c-cyclical monotone. Whenever c is continuous, it is even only necessary to check the support of π by Theorem 1.3.2. In this chapter we will introduce a stochastically restricted version of the OT problem, called *martingale optimal transport*. The martingale version of the optimal transport problem finds its origin in the pricing of exotic options in a robust, or model-independent setting [10, 20, 29]. Unfortunately, c-cyclically monotonicity will no longer be enough to check optimality for the martingale optimal transport problem. So, we will need a new definition of monotonicity. Even though the martingale version of the OT problem has been known and used for quite some time, a stability result in the same vein as Theorem 1.4.1 was not known to be true until recently. After the introduction of the martingale optimal transport problem, we will state this new stability result for the case $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, which was given and proved by Backhoff-Veraguas and Pammer in [5]. Once the problem statement is clear, we will introduce the necessary tools to prove this result. Although much of theory can be done on a high level of generality by considering general Polish spaces $\mathcal{X}$ and $\mathcal{Y}$, we will assume that $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ for some $d \in \mathbb{N}$. We do not lose much generality, as most of the proofs can be copied almost verbatim in the general case, but it allows us to discuss the results in a more familiar setting.

2.1. The problem

As was stated in the introduction of this thesis and this chapter, martingale optimal transport finds its origin in the theory of robust finance. In that field, the goal is to find prices, or at least bounds on prices, that do not depend on an underlying model, such as the Black-Scholes or Hull-White model.

Recall the example given in the introduction of this thesis. We had two stock prices S_0 and S_T, whose laws were given by the measures μ and ν respectively. We also had a measurable function g representing some financial derivative. The next step was to look at all martingale measures π that are compatible with μ and ν. Formally, this means that π must be a martingale coupling between μ and ν.

Definition 2.1.1 [Martingale coupling]**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $X, Y : \Omega \to \mathbb{R}^d$ random variables and $\mu, \nu \in \mathcal{P}\left(\mathbb{R}^d\right)$. A coupling $\pi \in \Pi\left(\mu, \nu\right) \subseteq \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is a *martingale coupling* if

$$\mathbb{E}[Y \mid X] = X \quad \text{for } (X, Y) \sim \pi.$$

The set of all martingale couplings is denoted $\mathcal{M}\left(\mu, \nu\right)$.

For our setting it is beneficial to look at an equivalent definition of a martingale coupling, as it allows us to easily do calculations.

Theorem 2.1.2 [Equivalent definition martingale coupling]**.** *Let $\mu, \nu \in \mathcal{P}\left(\mathbb{R}^d\right)$ and $\pi \in \Pi(\mu, \nu)$, then π is a martingale coupling if and only if there exists a regular disintegration $\{\pi_x\}_{x \in \mathbb{R}^d}$ of π, as* *defined in Definition A.1.12, such that*

$$x = \int_{\mathbb{R}^d} y\, \pi_x(\mathrm{d}y).$$

Proof. "$\Rightarrow$" This is a direct consequence of Theorem A.1.14 and Theorem A.1.13. Let $\pi \in \mathcal{M}(\mu,\nu)$ and $(X,Y) \sim \pi$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then the latter theorem gives us a kernel such that $\pi_x(\cdot) = \mathbb{P}(Y \in \cdot \,|\, X = x)$. Using the former theorem with $f(X,Y) = Y$, we get

$$x = \mathbb{E}[Y \,|\, X = x] = \int_{\mathbb{R}^d} y\, \pi_x(\mathrm{d}y).$$

"$\Leftarrow$" Conversely, if we have a coupling between μ and ν with this disintegration property, then applying Theorem A.1.14 gives us that π is a martingale coupling immediately. □

Contrary to the normal set of couplings $\Pi(\mu,\nu)$ it could be that $\mathcal{M}(\mu,\nu) = \varnothing$. A classic result known as Strassen's theorem gives a necessary and sufficient condition to check whether the set $\mathcal{M}(\mu,\nu)$ is actually non-empty. For this we need the following notion.

Definition 2.1.3 [convex order]. Let $\mu, \nu \in \mathcal{P}\left(\mathbb{R}^d\right)$. We say that μ *is smaller than* ν *in the convex order*, denoted by $\mu \preceq \nu$, if

$$\int_{\mathbb{R}^d} \varphi(x)\, \mu(\mathrm{d}x) \leq \int_{\mathbb{R}^d} \varphi(x)\, \nu(\mathrm{d}x),$$

for all convex functions $\varphi : \mathbb{R}^d \to \mathbb{R}$.

Strassen's theorem now says martingale couplings exist, whenever μ and ν are in convex order.

Theorem 2.1.4 [Strassen's theorem]. *Let $\mu, \nu \in \mathcal{P}\left(\mathbb{R}^d\right)$. Then, the set $\mathcal{M}\left(\mu,\nu\right) \neq \varnothing$ if and only if both μ and ν have finite means and $\mu \preceq \nu$*

Proof. See proof of Theorem 8, page 434, of [47]. □

We return to our example one last time. We were left with the set of martingale couplings between μ and ν. The final step in robust finance would be to find a bound on the financial derivative $g(S_0, S_T)$. This is done by calculating

$$P_{\text{low}} = \inf_{\pi \in \mathcal{M}(\mu,\nu)} \mathbb{E}_\pi[g(S_0, S_T)] \quad P_{\text{high}} = \sup_{\pi \in \mathcal{M}(\mu,\nu)} \mathbb{E}_\pi[g(S_0, S_T)]. \tag{2.1}$$

The price of the financial derivative should then be between P_{low} and P_{high}. However, to achieve these bounds, we did not need to impose any kind of model! The calculations done in (2.1) is precisely the martingale optimal transport problem.

Definition 2.1.5 [Martingale optimal transport]. Let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a measurable cost function and $\mu \in \mathcal{P}\left(\mathbb{R}^d\right), \nu \in \mathcal{P}(\mathbb{R}^d)$ such that $\mu \preceq \nu$. The martingale optimal transport problem is defined by finding the value or the minimizing coupling, if it exists, of

$$V_M\left(\mu,\nu\right) := \inf_{\pi \in \mathcal{M}(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y)\, \pi\left(\mathrm{d}x, \mathrm{d}y\right). \tag{MOT}$$

2.2. Existence of MOT optimizer

In the case of the normal OT problem, we were guaranteed to find an optimizing coupling by Theorem 1.2.6. This was in essence a consequence of $\Pi(\mu,\nu)$ being compact and the cost functional

$$\pi \mapsto \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)\pi(\mathrm{d}x,\mathrm{d}y)$$

being lower semicontinuous, whenever c is also lower semicontinuous. Luckily, we can also always find an optimizer for the MOT problem. To see this, we only need to show that $\mathcal{M}(\mu,\nu)$ is compact in $\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$.

Lemma 2.2.1 [Compactness of $\mathcal{M}(\mu,\nu)$]. *Let $r\in[1,\infty)$, $\mu,\nu\in\mathcal{P}_r(\mathbb{R}^d)$ and $\mu\preceq\nu$, then $\mathcal{M}(\mu,\nu)$ is compact in both the weak topology and the $\mathcal{W}_r$-topology.*

Proof. We will show compactness in the weak topology first. It is clear that $\mathcal{M}(\mu,\nu)\subseteq\Pi(\mu,\nu)$. We already know that $\Pi(\mu,\nu)$ is weakly compact. So, we only need to show that $\mathcal{M}(\mu,\nu)$ is closed. We will look ahead a bit in this thesis and use the characterisation of $\mathcal{M}(\mu,\nu)$ given by Lemma 2.4.1. We only need to check sequentially compactness, because the weak topology is metrizable by Theorem 1.2.4.

Let $\{\pi_k\}_{k\in\mathbb{N}}$ be a sequence in $\mathcal{M}(\mu,\nu)$ that converges weakly to some π in $\mathcal{P}(\mathcal{X}\times\mathcal{Y})$. In fact, we have convergence in the $\mathcal{W}_r$-topology, which can be shown as follows. As we already have weak convergence, we only need to check that the rth moment also converges. This is quickly verified,

$$\begin{aligned}\lim_{k\to\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d} d((0,0),(x,y))^r\pi_k(\mathrm{d}x,\mathrm{d}y) &= \lim_{k\to\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d} |x|^r+|y|^r\pi_k(\mathrm{d}x,\mathrm{d}y)\\ &= \lim_{k\to\infty}\int_{\mathbb{R}^d}|x|^r\mu(\mathrm{d}x,\mathrm{d}y)+\int_{\mathbb{R}^d}|y|^r\nu(\mathrm{d}x)\\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d}|x|^r+|y|^r\pi(\mathrm{d}x,\mathrm{d}y).\end{aligned}$$

Here, we used that all the marginals of $\{\pi_k\}_{k\in\mathbb{N}}$ are the same. In particular, convergence in $\mathcal{W}_1$ means that we can test the convergence against functions $g\in C(\mathcal{X}\times\mathcal{Y})$ such that there exists a $C\in\mathbb{R}$, for which

$$|g(x,y)|\le C(1+d(x,y)^r). \tag{2.2}$$

The functions $f(x)(x-y)$ with $f\in C_b(\mathbb{R}^d)$ satisfy the requirement given in (2.2). So, we see

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} f(x)(x-y)\pi(\mathrm{d}x,\mathrm{d}y)=\lim_{k\to\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d} f(x)(x-y)\pi_k(\mathrm{d}x,\mathrm{d}y)=\lim_{k\to\infty}0=0.$$

This shows that $\pi\in\mathcal{M}(\mu,\nu)$, which means that $\mathcal{M}(\mu,\nu)$ is closed in the weak topology. Hence, it is also compact in the weak topology.

To establish compactness in the $\mathcal{W}_r$-topology, we remark that the above argument also shows that $\mathcal{M}(\mu,\nu)$ is closed in the $\mathcal{W}_r$-topology. Furthermore, we can also argue that $\Pi(\mu,\nu)$ is compact in the $\mathcal{W}_r$-topology. Take any subsequence $\{\pi_k\}_{k\in\mathbb{N}}$ with $\pi_k\in\Pi(\mu,\nu)$. By weak compactness we can find a weakly converging subsequence, which we will also call $\{\pi_k\}_{k\in\mathbb{N}}$ after relabelling. This sequence converges to some $\pi \in \Pi(\mu,\nu)$ in the weak topology. By a similar argument as above it can be argued that this convergence also holds in the $\mathcal{W}_r$-topology, which shows compactness. Thus, $\mathcal{M}(\mu,\nu)$ is a closed subset of a compact set in the $\mathcal{W}_r$-topology, which shows compactness of $\mathcal{M}(\mu,\nu)$ itself. □

Just as in the classical case, we can use compactness to show existence of a martingale optimizer.

Theorem 2.2.2 [Existence of a martingale optimizer]. *Let $c : \mathbb{R}^d \times \mathbb{R}^d \to [0,\infty)$ be a lower semicontinuous function and let $\mu,\nu \in \mathcal{P}_1(\mathbb{R}^d)$ such that $\mu \preceq \nu$. Then there exists a $\pi^* \in \mathcal{M}(\mu,\nu)$ such that*

$$\int\limits_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)\pi^*(\mathrm{d}x,\mathrm{d}y) = \inf_{\pi\in\mathcal{M}(\mu,\nu)} \int\limits_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)\pi(\mathrm{d}x,\mathrm{d}y).$$

Proof. This can be shown by repeating the proof of Theorem 1.2.6 and replace $\Pi(\mu,\nu)$ with $\mathcal{M}(\mu,\nu)$ □

2.3. Need for larger space

It turns out that it is not possible to find an immediate analogue to a c-cyclical monotone set for the MOT problem. Indeed, Juillet showed in [31] that the optimality of a specific MOT problem cannot be verified by looking at the support of candidate optimal couplings, even when the cost function is continuous. This is in sharp contrast with Theorem 1.3.2 in the classical case. Juillet considered $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ with cost function $c(x,y) = (1+\tanh(-x))\sqrt{y^2+1}$ and showed that there exists a Borel set Γ for the optimizer of this MOT problem such that

- $\pi(\Gamma) = 1$,
- for every $(x,y^-),(x,y^+),(x',y') \in \Gamma$ satisfying $x < x'$ and $y^- < y^+$, the element $y' \notin (y^-,y^+)$. This is called the *left-monotone property*.

The existence of such a set Γ for a coupling π is both necessary and sufficient for optimality in this case. We will now give an example of a MOT problem, whose minimizing coupling is not supported on a set that enjoys the left-monotone property, i.e. $\mathrm{supp}(\pi) \neq \Gamma$.

Let $\mu = \frac{1}{2}\lambda|_{[-1,1]}$, $\nu = \frac{1}{4}(\delta_{-1}+2\delta_0+\delta_1)$ and $c(x,y) = (1+\tanh(-x))\sqrt{y^2+1}$. Juillet proved that the mass contained in $[-1,0]$ gets mapped to $\{-1,0\}$ by the optimizing coupling and $[0,1]$ is mapped to $\{0,1\}$.This means that every point in $[-1,0]$ gets its mass split and one part is sent to $\{-1\}$ and the other to $\{0\}$. So, if we take the points $-\epsilon$, ϵ and $1-\epsilon$, then their masses are sent to $\{-1,0\}$, $\{0,1\}$ and $\{0,1\}$, respectively. Letting $\epsilon \to 0$, we see that the points $(0,-1)$, $(0,1)$ and

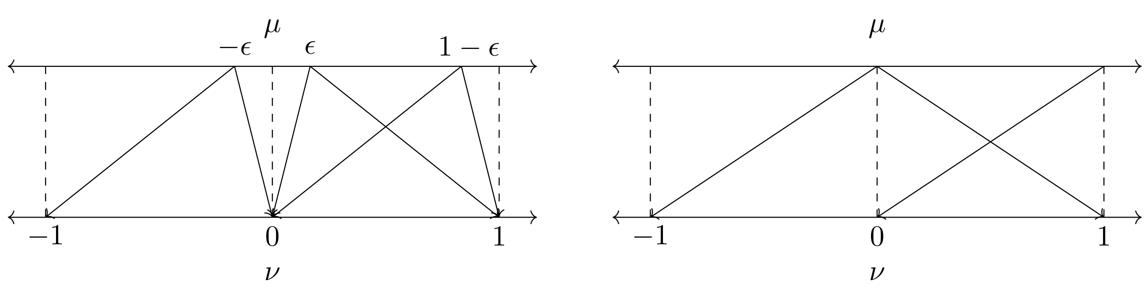


Figure 2.1.: Violation of the left-monotone property, Example 2.11 in [31].

$(1, 0)$ are all part of the support of the minimizing coupling π^*. However, these points violate the left-monotone property. See Figure 2.1 for a visual representation.

So, we cannot merely base the optimality of an optimizer on the support of a coupling in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. However, Backhoff-Veraguas and Pammer argued in [5] that in the lifted space $\mathcal{P}(\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d))$, we are able to define a suitable set from which we can read optimality.

In the sections to come we will need a topology on the space $\mathcal{P}(\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d))$ to discuss various convergence results. We will introduce that topology now.

- Let d be a metric on $\mathbb{R}^d$.
- Let $r \in [1, \infty)$, the space $\mathcal{P}_r(\mathbb{R}^d)$ is equipped with the Wasserstein-r metric $\mathcal{W}_r$, as introduced in Definition 1.5.1.
- On the space $\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d)$ we can now define the metric

$$|\widetilde{d}((x, \mu), (y, \nu))|^r := d(x, y)^r + \mathcal{W}_r(\mu, \nu)^r, \quad (x, \mu), (y, \nu) \in \mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d).$$

- Finally, we can define $\mathcal{P}_r(\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d))$ properly with its corresponding metric. Let $(x_0, p_0) \in \mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d)$. The elements in this set are those measures $P \in \mathcal{P}(\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d))$, for which

$$\int\limits_{\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d)} \widetilde{d}((x, p), (x_0, p_0))^r \, P(\mathrm{d}x, \mathrm{d}p) < \infty$$

  holds. This definition does not depend on (x_0, p_0), just as in Definition 1.5.3. For measures $P, Q \in \mathcal{P}_r(\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d))$, the metric on $\mathcal{P}_r(\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d))$ is now given by

$$|D_r(P, Q)|^r := \inf_{\gamma \in \Pi(P,Q)} \int\limits_{\left(\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d)\right)^2} \widetilde{d}((x, p), (y, q))^r \gamma\left(\mathrm{d}x, \mathrm{d}p, \mathrm{d}y, \mathrm{d}q\right).$$

In some cases we will not look at the full space $\mathcal{P}_r(\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d))$, but only at $\mathcal{P}_r(\mathcal{P}_r(\mathbb{R}^d))$, for which the Wasserstein metric is defined analogously.

To lift the MOT problem to this larger space we use an embedding named J with left inverse $\hat{I}$. We will introduce the map and display some of its important properties. Some of the results in this section will be proven in Chapter 4, as they require additional knowledge of the topological features of the space $\mathcal{P}_r(\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d))$. In Chapter 4 we will also give more background information and motivation as to why this map is an interesting object in its own right.

Definition 2.3.1 [J embedding]. Let $\mu, \nu \in \mathcal{P}\left(\mathbb{R}^d\right)$ and $\pi \in \Pi(\mu, \nu)$, then J is defined as

$$\begin{aligned} J : \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) &\to \mathcal{P}(\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)), \\ \pi(\mathrm{d}x, \mathrm{d}y) &\mapsto \pi \circ \mathrm{proj}_1^{-1}(\mathrm{d}x)\delta_{\pi_x}(\mathrm{d}p) = \mu(\mathrm{d}x)\delta_{\pi_x}(\mathrm{d}p) \end{aligned} \tag{2.3}$$

with left inverse

$$\hat{I} : \mathcal{P}(\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)) \to \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), P \mapsto \int\limits_{\mathcal{P}(\mathbb{R}^d)} p(\mathrm{d}y)P(\mathrm{d}x, \mathrm{d}p). \tag{2.4}$$

That $\hat{I}$ is a left-inverse can be seen by a direct computation,

$$\begin{aligned} \hat{I}(J(\pi))(\mathrm{d}x, \mathrm{d}y) &= \int\limits_{\mathcal{P}(\mathbb{R}^d)} p(\mathrm{d}y)J(\pi)(\mathrm{d}x, \mathrm{d}p) = \int\limits_{\mathcal{P}(\mathbb{R}^d)} p(\mathrm{d}y)\delta_{\pi_x}(\mathrm{d}p)\pi \circ \mathrm{proj}_1^{-1}(\mathrm{d}x) \\ &= \pi_x(\mathrm{d}y)\pi \circ \mathrm{proj}_1^{-1}(\mathrm{d}x) = \pi(\mathrm{d}x, \mathrm{d}y). \end{aligned}$$

We will also need a version of $\hat{I}$ which is only defined on the second component, which is called the intensity map I,

$$I : \mathcal{P}(\mathcal{P}(\mathbb{R}^d)) \to \mathcal{P}(\mathbb{R}^d), Q \mapsto \int\limits_{\mathcal{P}(\mathbb{R}^d)} p(\mathrm{dy})\, Q(\mathrm{d}p). \tag{2.5}$$

The map J is in general not a nice map to work with. For example, it is rarely continuous. However, it does enjoy some useful properties. The most important one, for us, being that it maps relative compact sets to relative compact sets in the $\mathcal{W}_r$-topology. The maps $\hat{I}$ and I are better behaved, as they are continuous with respect to the $\mathcal{W}_r$-topology.

Theorem 2.3.2. *Let $r \in [1, \infty)$ and define the maps $J, \hat{I}$ and I as in* (2.3)*,* (2.4) *and* (2.5)*. The following statements are true:*

(i) A set $\Pi \subseteq \mathcal{P}_r(\mathbb{R}^d \times \mathbb{R}^d)$ is relatively compact in the $\mathcal{W}_r$-topology if and only if $J(\Pi) \subseteq \mathcal{P}_r(\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d))$ is relatively compact in the $\mathcal{W}_r$-topology.

(ii) A set $\Lambda \subseteq \mathcal{P}_r(\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d)$ is relatively compact in the $\mathcal{W}_r$-topology if and only if $\hat{I}(\Lambda) \subseteq \mathcal{P}_r(\mathbb{R}^d \times \mathbb{R}^d)$ is relatively compact in the $\mathcal{W}_r$-topology.

(iii) The maps $\hat{I}$ and I are continuous in both the weak and $\mathcal{W}_r$-topology.

Proof. First, we prove continuity of $\hat{I}$ and I. We will only show continuity for I in full detail, as the proof for $\hat{I}$ is completely analogue. Note that for a measure $P \in \mathcal{P}_r(\mathcal{P}_r(\mathbb{R}^d))$, the mapping I is fully determined by

$$I(P)(f) = \int\limits_{\mathcal{P}(\mathbb{R}^d)} p(f) P(\mathrm{d}p) \quad \forall f \in C_b(\mathbb{R}^d).$$

Now, let $\{P_k\}_{k \in \mathbb{N}}$ be a sequence in $\mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ that converges weakly to P. In particular, this means that we can test the convergence against functions $g \in C_b(\mathcal{P}(\mathbb{R}^d))$. We can use the specific functions g_f given by

$$g_f(p) = \int\limits_{\mathbb{R}^d} f(y) p(\mathrm{d}y) = p(f), \quad f \in C_b(\mathbb{R}^d),$$

which is a bounded continuous mapping, as it is bounded and linear. Integrating these functions and taking limits gives us

$$\lim_{k \to \infty} I(P_k)(f) = \lim_{k \to \infty} \int\limits_{\mathcal{P}(\mathbb{R}^d)} p(f) P_k(\mathrm{d}p) = \lim_{k \to \infty} \int\limits_{\mathcal{P}(\mathbb{R}^d)} g_f(p) P_k(\mathrm{d}p) = \int\limits_{\mathcal{P}(Y)} g_f(p) P(\mathrm{d}p) = I(P)(f).$$

So, we see that I is a continuous mapping in the weak topology. This proof can be repeated with functions $f \in C(\mathbb{R}^d)$ such that $|f(y)| < C(1 + d(y_0, y)^r)$ for some $C \in \mathbb{R}$ to establish continuity in the $\mathcal{W}_r$-topology.

The proofs of items (i) and (ii) require considerable preparatory work, as we need a characterisation of relative compactness in the $\mathcal{W}_r$-topology of $\mathcal{P}_r(\mathcal{P}_r(\mathbb{R}^d))$, which will make it easier to check if a set is relatively compact. The full proof can be found on page 72. □

With the maps J and $\hat{I}$ in mind we will look at the set of measures on $\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ that have their image under $\hat{I}$ in $\Pi(\mu, \nu)$ or $\mathcal{M}(\mu, \nu)$,

$$\Lambda(\mu, \nu) = \{P \in \mathcal{P}(\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)) \,|\, \hat{I}(P) \in \Pi(\mu, \nu)\} \tag{2.6}$$

$$\Lambda_M(\mu, \nu) = \{P \in \mathcal{P}(\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)) \,|\, \hat{I}(P) \in \mathcal{M}(\mu, \nu)\}. \tag{2.7}$$

Definition 2.3.3 [Weak (martingale) couplings]. The measures in $\Lambda(\mu,\nu)$ will be called *weak couplings* of μ and ν and the measures in $\Lambda_M(\mu,\nu)$ will be called *weak martingale couplings* between μ and ν.

Note that for $P \in \Lambda(\mu,\nu)$ we have

$$\hat{I}(P) = \int_{\mathcal{P}(\mathbb{R}^d)} p(\mathrm{d}y)P(\mathrm{d}x,\mathrm{d}p) = \int_{\mathcal{P}(\mathbb{R}^d)} p(\mathrm{d}y)P_x(\mathrm{d}p)\mu(\mathrm{d}x) = \mu(\mathrm{d}x)I(P_x)(\mathrm{d}y).$$

Another useful property that we will need is $\Lambda(\mu,\nu)$ and $\Lambda_M(\mu,\nu)$ are both compact. Just as was the case for $\Pi(\mu,\nu)$ and $\mathcal{M}(\mu,\nu)$.

Lemma 2.3.4. *Let $r \in [1,\infty)$ and $\mu,\nu \in \mathcal{P}_r(\mathbb{R}^d)$. The sets $\Lambda(\mu,\nu)$ and $\Lambda_M(\mu,\nu)$ as defined in* (2.6) *and* (2.7)*, respectively, are compact in the $\mathcal{W}_r$-topology.*

Proof. First, we show that $\Lambda(\mu,\nu)$ is compact. Note that $\Lambda(\mu,\nu) = \hat{I}^{-1}(\Pi(\mu,\nu))$. The map $\hat{I}$ is continuous by property (iii) of Lemma 2.3.2 and $\Pi(\mu,\nu)$ was seen to be compact in the $\mathcal{W}_r$-topology by the argument at the end of the proof of Lemma 2.2.1, hence closed. So, the set $\Lambda(\mu,\nu)$ is closed. An application of property (ii) of Theorem 2.3.2 gives is that $\Lambda(\mu,\nu)$ is relatively compact, because $\hat{I}(\Lambda(\mu,\nu)) = \Pi(\mu,\nu)$. From $\Lambda(\mu,\nu)$ being closed and relatively compact we conclude that $\Lambda(\mu,\nu)$ is compact.

Continuing to $\Lambda_M(\mu,\nu)$, we can similarly see that $\Lambda_M(\mu,\nu)$ is closed, because of $\Lambda_M(\mu,\nu) = \hat{I}^{-1}(\mathcal{M}(\nu,\mu))$ and $\mathcal{M}(\mu,\nu)$ being closed by virtue of Lemma 2.2.1. The set $\Lambda_M(\mu,\nu)$ is now seen to be a closed subset of the compact set $\Lambda(\mu,\nu)$, meaning that $\Lambda_M(\mu,\nu)$ is also compact. $\square$

It is possible to generalise the OT problem and the MOT problem by considering cost functionals on the larger space, $C : \mathbb{R}^d \times \mathcal{P}\left(\mathbb{R}^d\right) \to [0,\infty)$. There are two natural ways of using these cost functionals in the martingale case.

Definition 2.3.5 [Martingale optimal weak transport]. Let $\mu,\nu \in \mathcal{P}(\mathbb{R}^d)$ and $C : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to [0,\infty)$ measurable, the two generalised versions of the martingale optimal transport are given by

$$\inf_{\pi \in \mathcal{M}(\mu,\nu)} \int_{\mathbb{R}^d} C(x,\pi_x)\mu(\mathrm{d}x), \tag{MOWT}$$

$$\inf_{P \in \Lambda_M(\mu,\nu)} \int_{\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)} C(x,p)P(\mathrm{d}x,\mathrm{d}p). \tag{MOWT'}$$

Note that these classes of problems encompass the MOT problem by using the specific cost functional

$$C(x,p) = \int_{\mathbb{R}^d} y p(\mathrm{d}y).$$

2.4. Martingale optimal transport as a linear constraint problem

Just as was the case for the classical OT problem, we want to find a useful monotonicity principle that can identify optimal martingale couplings. Luckily for us, Beiglböck and Griessler introduced a general monotonicity principle for 'optimal transport-like' problems in [14]. However, In our quest to find a useful monotonicity principle we need to cast the MOT problem into the same form Beiglöck and Griessler considered. In Chapter 3 we will give a brief overview of their result and illustrate their proofs. The way we need to phrase the MOT problem, as given in Definition 2.1.1, is as an optimization problem under a set of linear constraints. The following lemma ensures that this restatement of the problem is equivalent to the MOT problem. Any proof that is missing from this section can be found in Chapter 3.

Lemma 2.4.1 [Martingale constraints]. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ and let $\mathcal{F}_{\mu,\nu}$ be the set of functions given by*

$$\begin{aligned}\mathcal{F}_{\mu,\nu} &:= \left\{ f \in C_b(\mathbb{R}^d \times \mathbb{R}^d) \mid f(x,y) = g(x) - \int\limits_{\mathbb{R}^d} g(x)\mu(\mathrm{d}x),\ g \in C_b(\mathbb{R}^d) \right\} \\ &\cup \left\{ f \in C_b(\mathbb{R}^d \times \mathbb{R}^d) \mid f(x,y) = g(y) - \int\limits_{\mathbb{R}^d} g(y)\nu(\mathrm{d}y),\ g \in C_b(\mathbb{R}^d) \right\} = \mathcal{F}_\mu \cup \mathcal{F}_\nu.\end{aligned} \tag{2.8}$$

Let $\mathcal{F}_M$ be a set of functions given by

$$\begin{aligned}\mathcal{F}_M &:= \left\{ f \in C_b(\mathbb{R}^d \times \mathbb{R}^d) \mid f(x,y) = g(x)(y-x),\ g \in C_b(\mathbb{R}^d) \right\} \cup \mathcal{F}_{\mu,\nu} \\ &= \mathcal{F}_m \cup \mathcal{F}_{\mu,\nu}.\end{aligned} \tag{2.9}$$

Then, a measure $\pi \in \mathcal{P}\left(\mathbb{R}^d \times \mathbb{R}^d\right)$ is a martingale coupling *if and only if*

$$\int\limits_{\mathbb{R}^d \times \mathbb{R}^d} f(x,y)\pi\left(\mathrm{d}x, \mathrm{d}y\right) = 0 \quad \forall f \in \mathcal{F}_M.$$

Proof. Intuitively, the set $\mathcal{F}_{\mu,\nu}$ ensures that π is actually a coupling between μ and ν. The set $\mathcal{F}_m$ in turn ensures that π possesses the martingale property.

"$\Rightarrow$" If π is a martingale coupling between μ and ν, then there exists a disintegration $\{\pi_x\}_{x \in \mathbb{R}^d}$ such that

$$x = \int\limits_{\mathbb{R}^d} y\, \pi_x\left(\mathrm{d}y\right) \quad \mu \text{ a.s.}$$

Filling this in with $f \in \mathcal{F}_m$ gives us

$$\begin{aligned}\int\limits_{\mathbb{R}^d \times \mathbb{R}^d} f(x,y)\, \pi(\mathrm{d}x, \mathrm{d}y) &= \int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} g(x)(x-y)\, \pi_x(\mathrm{d}y)\mu(\mathrm{d}x) \\ &= \int\limits_{\mathbb{R}^d} g(x) \left(x - \int\limits_{\mathbb{R}^d} y\, \pi_x(\mathrm{d}y) \right) \mu(\mathrm{d}x) = 0.\end{aligned}$$

Also, for $f \in \mathcal{F}_\mu$ we immediately see

$$\begin{aligned}\int\limits_{\mathbb{R}^d \times \mathbb{R}^d} f(x,y)\pi(\mathrm{d}x, \mathrm{d}y) &= \int\limits_{\mathbb{R}^d \times \mathbb{R}^d} g(x)\pi_x(\mathrm{d}y)\mu(\mathrm{d}x) - \int\limits_{\mathbb{R}^d} g(x)\mu(\mathrm{d}x) \\ &= \int\limits_{\mathbb{R}^d} g(x)\mu(\mathrm{d}x) - \int\limits_{\mathbb{R}^d} g(x)\mu(\mathrm{d}x) = 0.\end{aligned}$$

Analogously, we can show that the integral with respect to any $f \in \mathcal{F}_\nu$ is equal to zero.

"$\Leftarrow$" For this direction we will be using the monotone class theorem, see Theorem A.1.1. It needs to be shown that probability measures in $\Pi_{\mathcal{F}_\mu}$ have μ as their first marginal, the measures in $\Pi_{\mathcal{F}_\nu}$ have ν as their marginal, and $\Pi_{\mathcal{F}_m}$ contains martingale measures. All three arguments follow the same pattern. So, we will only show the proof for the martingale property in full detail. The other two proofs can then be inferred by following the same steps.

Let $\mathcal{H}$ be the following set:

$$\mathcal{H} = \left\{ h(x) \in B(\mathbb{R}^d) \middle| \int\limits_{\mathbb{R}^d \times \mathbb{R}^d} h(x)(x-y)\, \pi(\mathrm{d}x, \mathrm{d}y) = 0 \right\}.$$

This is clearly a vector subspace of $B(\mathbb{R}^d)$. Now set $\mathcal{K} = C_b(\mathbb{R}^d)$. By assumption we have $\mathcal{K} \subseteq \mathcal{H}$. Furthermore, the constant function $g(x) = 1$ is trivially continuous and bounded. So, $1 \in \mathcal{K} \subseteq \mathcal{H}$. As the product of continuous and bounded functions is again continuous and bounded, we find that $\mathcal{K}$ is closed under multiplication. For a bounded and monotonically increasing sequence $\{h_n\}_{n\in\mathbb{N}}$ with $h_n \in \mathcal{H}$, $h_n \geq 0$, and limit h, we see that

$$\begin{aligned}
\int\limits_{\mathbb{R}^d \times \mathbb{R}^d} h(x)(x-y)\, \pi(\mathrm{d}x, \mathrm{d}y) &= \int\limits_{\mathbb{R}^d} h(x) \left(x - \int\limits_{\mathbb{R}^d} y\, \pi_x(\mathrm{d}y) \right) \pi \circ \mathrm{proj}_1^{-1}(\mathrm{d}x) \\
&= \int\limits_{\mathbb{R}^d} \lim_{n\to\infty} h_n(x) \left(x - \int\limits_{\mathbb{R}^d} y\, \pi_x(\mathrm{d}y) \right) \pi \circ \mathrm{proj}_1^{-1}(\mathrm{d}x) \\
&= \lim_{n\to\infty} \int\limits_{\mathbb{R}^d} h_n(x) \left(x - \int\limits_{\mathbb{R}^d} y\, \pi_x(\mathrm{d}y) \right) \pi \circ \mathrm{proj}_1^{-1}(\mathrm{d}x) = 0.
\end{aligned}$$

We were able to pull the limit outside the integral in the third equality because of the monotone convergence theorem, see Theorem A.1.2. The monotone class theorem then tells us that $\sigma(\mathcal{K})_b \subseteq \mathcal{H}$. By Theorem A.1.7 we see that $\sigma(C_b(\mathbb{R}^d)) = \mathcal{B}(\mathbb{R}^d)$. Thus, $\mathcal{H}$ contains all bounded Borel measurable functions on $\mathbb{R}^d$. Define the function $h(x) = x - \int_{\mathbb{R}^d} y \pi_x(\mathrm{d}y)$ and look at the indicator function of $\{h(x) > 0\}$, which is measurable by the measurability of $\pi_x(\mathrm{d}y)$. It follows that $1_{\{h(x)>0\}} \in \mathcal{H}$. From this we see

$$\int\limits_{\mathbb{R}^d} 1_{\{h(x)>0\}} \left(x - \int\limits_{\mathbb{R}^d} y\, \pi_x(\mathrm{d}y) \right) \pi \circ \mathrm{proj}_1^{-1}(\mathrm{d}x) = 0.$$

For $\{h(x) < 0\}$ we can get a similar result. Combining these two gives us that,

$$x = \int\limits_{\mathbb{R}^d} y\, \pi_x(\mathrm{d}y) \quad \pi \circ \mathrm{proj}_1^{-1}\text{-a.s.},$$

which means that the measures in $\Pi_{\mathcal{F}_m}$ are martingale measures. To see that $\Pi_{\mathcal{F}_\mu}$ and $\Pi_{\mathcal{F}_\nu}$ contain the measures with μ and ν marginals. We can repeat the argument using the sets

$$\mathcal{H} = \left\{ h(x) \in B(\mathbb{R}^d) \middle| \int\limits_{\mathbb{R}^d \times \mathbb{R}^d} h(x)\pi(\mathrm{d}x, \mathrm{d}y) - \int\limits_{\mathbb{R}^d} h(x)\mu(\mathrm{d}x) = 0 \right\},$$

and $\mathcal{K}$ defined as all functions in $\mathcal{K}$ that are bounded continuous functions. This will show that $\Pi_{\mathcal{F}_\mu}$ contain all measures that have μ as their first marginal. If we replace μ with ν in $\mathcal{H}$ and $\mathcal{K}$, then we get the sets we need to show that $\Pi_{\mathcal{F}_\nu}$ contains the measures with ν as their second marginal.

To conclude, we now have that if $\pi \in \Pi_{\mathcal{F}_\mu} \cap \Pi_{\mathcal{F}_\nu} \cap \Pi_{\mathcal{F}_m}$ then it is a martingale coupling between μ and ν. $\square$

Using the set of functions $\mathcal{F}_M$, we can formulate a monotonicity definition in Definition 2.4.2 that will turn out to be useful to identify optimality . Indeed, we will see in Theorems 2.4.4 and 2.4.5 that this definition of monotonicity fully characterises an optimal martingale coupling in the 1-dimensional case.

Definition 2.4.2 [$(c, \mathcal{F}_M)$-competitors and monotonicity].
Let $c : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty)$ be measurable.

(1) A measure $\alpha' \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is called an $\mathcal{F}_M$-*competitor* of $\alpha \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ if and only if $\alpha(f) = \alpha'(f)$ for all $f \in \mathcal{F}_M$.
(2) We call $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$ $(c, \mathcal{F}_M)$-*monotone* if and only if for any probability measure $\alpha \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$, that is finitely supported on Γ, and any $\mathcal{F}_M$-competitor α' of α, we have $\alpha(c) \leq \alpha'(c)$.
(3) A martingale coupling π is called $(c, \mathcal{F}_M)$-*monotone*, if it is supported on a $(c, \mathcal{F}_M)$-monotone set.

One might wonder what it means for a measure α' to be an $\mathcal{F}_M$-competitor for some other measure α. The following proposition tells us that α and α' need to have the same marginals and the mean of the regular disintegration with respect to the first component need to be the same.

Proposition 2.4.3. *Let $\alpha, \alpha \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$, both finitely supported, such that α' is an $\mathcal{F}_M$-competitor of α, then*

$$\alpha \circ \mathrm{proj}_i^{-1} = \alpha' \circ \mathrm{proj}_i^{-1} \quad \text{for } i = 1, 2$$
$$\int_{\mathbb{R}^d} y\alpha_x(\mathrm{d}y) = \int_{\mathbb{R}^d} y\alpha'_x(\mathrm{d}y) \quad \text{for } (\alpha \circ \mathrm{proj}_1^{-1}) - \textit{almost all } x \in \mathbb{R}^d.$$

Proof. We can use similar arguments as in the proof of Lemma 2.4.1 to show what $\mathcal{F}_M$-competitor measures look like and prove Proposition 2.4.3.

Let us first look at the functions from $\mathcal{F}_\mu$. We see that for all $g \in C_b(\mathbb{R}^d)$

$$\begin{aligned}
\int_{\mathbb{R}^d\times\mathbb{R}^d} g(x)\alpha(\mathrm{d}x, \mathrm{d}y) - \int_{\mathbb{R}^d\times\mathbb{R}^d} g(x)\mu(\mathrm{d}x) = \int_{\mathbb{R}^d\times\mathbb{R}^d} g(x)\alpha'(\mathrm{d}x, \mathrm{d}y) - \int_{\mathbb{R}^d\times\mathbb{R}^d} g(x)\mu(\mathrm{d}x) \quad &\implies \\
\int_{\mathbb{R}^d\times\mathbb{R}^d} g(x)\alpha(\mathrm{d}x, \mathrm{d}y) = \int_{\mathbb{R}^d\times\mathbb{R}^d} g(x)\alpha'(\mathrm{d}x, \mathrm{d}y) \quad &\iff \\
\int_{\mathbb{R}^d} g(x)\alpha \circ \mathrm{proj}_1^{-1}(\mathrm{d}x) = \int_{\mathbb{R}^d} g(x)\alpha' \circ \mathrm{proj}_1^{-1} \circ (\mathrm{d}x) &
\end{aligned}$$
.

Using a similar monotone class argument as was used in the proof of Lemma 2.4.1, it follows that α and α' have the same first marginal. Analogously, we can show that the second marginals of α

and α' must be equal. Finally, we can use the functions in $\mathcal{F}_m$ to see that for all $g \in C_b(\mathbb{R}^d)$

$$\int\limits_{\mathbb{R}^d\times\mathbb{R}^d} g(x)(x-y)\alpha(\mathrm{d}x,\mathrm{d}y) = \int\limits_{\mathbb{R}^d\times\mathbb{R}^d} g(x)(x-y)\alpha'(\mathrm{d}x,\mathrm{d}y) \qquad \Longleftrightarrow$$

$$\int\limits_{\mathbb{R}^d\times\mathbb{R}^d} g(x)(x-y)\alpha_x(\mathrm{d}y)\alpha\circ\mathrm{proj}_1^{-1}(\mathrm{d}x) = \int\limits_{\mathbb{R}^d\times\mathbb{R}^d} g(x)(x-y)\alpha_x(\mathrm{d}y)\alpha'\circ\mathrm{proj}_1^{-1}(\mathrm{d}x) \qquad \Longrightarrow$$

$$\int\limits_{\mathbb{R}^d} g(x)y\alpha_x(\mathrm{d}y)\alpha\circ\mathrm{proj}_1^{-1}(\mathrm{d}x) = \int\limits_{\mathbb{R}^d} g(x)y\alpha_x(\mathrm{d}y)\alpha'\circ\mathrm{proj}_1^{-1}(\mathrm{d}x).$$

Using the monotone class theorem again it is possible to show that this actually means that

$$\int\limits_{\mathbb{R}^d} y\alpha_x(\mathrm{d}y) = \int\limits_{\mathbb{R}^d} y\alpha'_x(\mathrm{d}y) \quad \text{for } (\alpha\circ\mathrm{proj}_1^{-1}) - \text{almost all } x \in \mathbb{R}^d.$$

So, we see that an $\mathcal{F}_M$-competitor of α is a measure that has the same marginals and which regular disintegrations with respect to x have the same means. □

A consequence of the result proved by Beiglböck and Griessler in [14] is that if a coupling is optimal for the MOT problem, then it is automatically $(c, \mathcal{F}_M)$-monotone. We will look at this result in more detail in Chapter 3.

Theorem 2.4.4 [Monotonicity of MOT]. *Let $c : \mathbb{R}\times\mathbb{R} \to [0,\infty)$ be a measurable function, $\mu,\nu \in \mathcal{P}_r(\mathbb{R}^d)$, and assume that π^* is such that*

$$\int\limits_{\mathbb{R}\times\mathbb{R}} c(x,y)\pi^*(\mathrm{d}x,\mathrm{d}y) = \inf_{\pi\in\mathcal{M}(\mu,\nu)} \int\limits_{\mathbb{R}\times\mathbb{R}} c(x,y)\pi(\mathrm{d}x,\mathrm{d}y).$$

Then π^* is $(c,\mathcal{F}_M)$-monotone.

Proof. To prove this theorem, it is better to work in a more general framework. However, as that framework is not the topic of this chapter, we will delay it to Chapter 3. We will see that Theorem 2.4.4 is a direct consequence of Theorem 3.1.1. □

A key insight by Beiglböck and Juillet, and Griessler in [12, 27] is that for $d = 1$, $(c, \mathcal{F}_M)$-monotonicity is actually a sufficient property to verify optimality. This was one of the crucial results needed to establish stability for the MOT problem.

Theorem 2.4.5. *Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ with $\mu \preceq \nu$. Let $c : \mathbb{R}\times\mathbb{R} \to [0,\infty)$ be measurable and $c(x,y) \le a(x) + b(y)$ for $a \in L^1(\mu)$ and $L^1(\nu)$. Then any $(c,\mathcal{F}_M)$-monotone $\pi \in \mathcal{M}(\mu,\nu)$ is optimal for the MOT problem with measures μ and ν.*

Proof. For this proof we will need a considerable amount of preliminary work and knowledge of the general theory of monotonicity for the MOT problem. So, this proof will be delegated to Chapter 3 on page 61. □

Similarly as with the martingale couplings it is possible to define the weak martingale couplings in terms of a set of continuous constraint functions analogously to Lemma 2.4.1.

Lemma 2.4.6 [Weak martingale constraints]. *Let $\mu,\nu\in\mathcal{P}_r(\mathbb{R}^d)$ and let $\widetilde{\mathcal{F}}_{\mu,\nu}$ be the set of functions given by*

$$\widetilde{\mathcal{F}}_{\mu,\nu}=\left\{f\in C_b(\mathbb{R}^d\times\mathcal{P}_r(\mathbb{R}^d))\,|\,f(x,p)=g(x)-\int\limits_{\mathbb{R}^d}g(x)\mu(\mathrm{d}x),\ g\in C_b(\mathbb{R}^d)\right\}$$
$$\cup\left\{f\in C_b(\mathbb{R}^d\times\mathcal{P}_r(\mathbb{R}^d))\,|\,f(x,p)=\int\limits_{\mathbb{R}^d}g(y)p(\mathrm{d}y)-\int\limits_{\mathbb{R}^d}g(y)\nu(\mathrm{d}y),\ g\in C_b(\mathbb{R}^d)\right\}=\widetilde{\mathcal{F}}_{\mu}\cup\widetilde{\mathcal{F}}_{\nu}.$$

Let the set $\widetilde{\mathcal{F}}_M$ of constraint functions be given by

$$\begin{aligned}\widetilde{\mathcal{F}}_M:=&\Big\{f\in C_b(\mathbb{R}^d\times\mathcal{P}_1(\mathbb{R}^d))\,\big|\,f(x,p)=h(x)g(p)\int_{\mathbb{R}^d}(x-y)p(\mathrm{d}y),\\&\quad g\in C_b(\mathcal{P}_1(\mathbb{R}^d)),h\in C_b(\mathbb{R}^d)\Big\}\cup\widetilde{\mathcal{F}}_{\mu,\nu}\\=&\,\widetilde{\mathcal{F}}_m\cup\widetilde{\mathcal{F}}_{\mu,\nu}.\end{aligned}$$

Then, a coupling $P\in\Lambda(\mu,\nu)$ is a weak martingale coupling, i.e. $P\in\Lambda_M(\mu,\nu)$, if and only if

$$\int\limits_{\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)}f(x,p)\,P(\mathrm{d}x,\mathrm{d}p)=0\quad\forall f\in\widetilde{\mathcal{F}}_M.$$

Proof. This proof follows the same idea as the proof of Lemma 2.4.1. Again, we only show the full argument for the weak martingale property of the measures.

“$\Rightarrow$" If P is a weak martingale coupling, then it should be that $\hat{I}(P)$ is a martingale coupling between μ and ν. This means actually that P gives full measure to the set $M=\{(x,p)\in\mathbb{R}\times\mathcal{P}_1(\mathbb{R}^d)\,|\,x=\int_{\mathbb{R}^d}yp(\mathrm{d}y)\}$, as

$$x=\int\limits_{\mathbb{R}^d}y\hat{I}(P)_x(\mathrm{d}y)=\int\limits_{\mathcal{P}_1(\mathbb{R}^d)}\int\limits_{\mathbb{R}^d}yp(\mathrm{d}y)P_x(\mathrm{d}p).$$

Which shows that $\int_{\mathbb{R}^d}yp(\mathrm{d}y)=x$ for P_x almost all p. The x was chosen arbitrarily. So we conclude that $\int_{\mathbb{R}^d}p(\mathrm{d}y)=x$ for P almost all (x,p), which shows that $P(M)=1$.

Now, let $f\in\widetilde{\mathcal{F}}_m$, we then see that

$$\begin{aligned}\int\limits_{\mathbb{R}^d\times\mathcal{P}_1(\mathbb{R}^d)}f(x,p)P(\mathrm{d}x,\mathrm{d}p)&=\int\limits_M h(x)g(p)\int\limits_{\mathbb{R}^d}(x-y)p(\mathrm{d}y)P(\mathrm{d}x,\mathrm{d}y)\\&=\int\limits_M h(x)g(p)\left(x-\int\limits_{\mathbb{R}^d}yp\,(\mathrm{d}y)\right)P(\mathrm{d}x,\mathrm{d}p)=0.\end{aligned}$$

In the last equality we have used that $x=\int_{\mathbb{R}^d}yp(\mathrm{d}y)$, P-almost surely. The integrals of functions in $\widetilde{\mathcal{F}}_\mu$ and $\widetilde{\mathcal{F}}_\nu$ can be calculated analogously to the equations above. The result is quickly seen to be equal to zero as well.

“$\Leftarrow$" We will use Theorem A.1.1 again. Just as was the case in the proof of Lemma 2.4.1, we will only show that $\widetilde{\mathcal{F}}_m$ ensures that $\hat{I}(P)$ is a martingale measure. The proofs that $\widetilde{\mathcal{F}}_\mu$ and $\widetilde{\mathcal{F}}_\nu$ ensure the correct marginals for P follow by similar arguments. However, we need to add a couple of arguments to prove the martingale property of $\hat{I}(P)$.

In this case we start with continuous and bounded functions that can be written as products and want to extend this to bounded measurable functions on $\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$. We start by defining the sets $\mathcal{K}$ and $\mathcal{H}$,

$$\mathcal{K} = \{f \in C_b(\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)) \,|\, f(x,p) = h(x)g(p), h(x) \in C_b(\mathbb{R}^d),\, g(p) \in C_b(\mathcal{P}(\mathbb{R}^d))\}$$

and

$$\mathcal{H} = \left\{ k \in B(\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)) : \int\limits_{\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)} k(x,p) \int\limits_{\mathbb{R}^d} (x-y)\, p(\mathrm{d}y)\, P(\mathrm{d}x, \mathrm{d}p) = 0 \right\}.$$

The set $\mathcal{K}$ is clearly closed under multiplication. The constant function is again trivially continuous and bounded. So $1 \in \mathcal{K} \subseteq \mathcal{H}$. To check the third requirement of Theorem A.1.1, let $\{k_n\}_{n\in\mathbb{N}}$ be a monotone increasing sequence in $\mathcal{H}$ with $k_n \geq 0$ converging to some function k. Using the monotone convergence theorem we can show that $k \in \mathcal{H}$.

$$\begin{aligned}\int\limits_{\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)} k(x,p) \int\limits_{\mathbb{R}^d} (x-y)p(\mathrm{d}y)P(\mathrm{d}x,\mathrm{d}p) &= \int\limits_{\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)} \lim_{n\to\infty} k_n(x,p) \int\limits_{\mathbb{R}^d} (x-y)p(\mathrm{d}y)P(\mathrm{d}x,\mathrm{d}p) \\ &= \lim_{n\to\infty} \int\limits_{\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)} k_n(x,p) \int\limits_{\mathbb{R}^d} (x-y)p(\mathrm{d}y)P(\mathrm{d}x,\mathrm{d}p) = 0.\end{aligned}$$

Now applying Theorem A.1.1 tells us that $\sigma(\mathcal{K})_b \subseteq \mathcal{H}$. It remains to argue that $\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathcal{P}_1(\mathbb{R}^d)) \subseteq \sigma(\mathcal{K})$. For this we note that $\mathcal{K}$ also contains functions of the type $f(x,p) = h(x)$, which generate $\mathcal{B}(\mathbb{R}^d)$ by Theorem A.1.7. Similarly, we also have functions of the type $f(x,p) = g(p)$. These functions in turn generate $\mathcal{B}(\mathcal{P}_1(\mathbb{R}^d))$ by Theorem A.1.7, as this is the Borel sigma algebra with respect to the Wasserstein-1 topology. This gives us that

$$\begin{aligned}\mathcal{B}(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d) &\subseteq \sigma(\mathcal{K}), \\ \mathbb{R}^d \times \mathcal{B}(\mathcal{P}_1(\mathbb{R}^d)) &\subseteq \sigma(\mathcal{K}),\end{aligned}$$

but for $\sigma(\mathcal{K})$ to be a sigma algebra it must be that the product sigma algebra $\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathcal{P}_1(\mathbb{R}^d))$ is contained in it.

We can now use that $\sigma(\mathcal{K})_b$ contains all bounded measurable functions. So, we can copy the final argument of the proof of Lemma 2.4.1 to conclude that

$$x = \int\limits_{\mathbb{R}^d} y p(\mathrm{d}y) \quad P-\text{a.s},$$

which shows that $\hat{I}(P)$ is a martingale measure. To show that $\widetilde{\mathcal{F}}_\mu$ ensures that the first marginals are equal to μ, we can use the same sets $\mathcal{H}$ as at the end of the proof of Lemma 2.4.1. To show that $\hat{I}(P)$ has ν as its second marginal we can use the sets

$$\mathcal{H} = \left\{ h \in B(\mathcal{P}_1(\mathbb{R}^d) \,\middle|\, h(p) = \int\limits_{\mathbb{R}^d} g(y)p(\mathrm{d}y) - \int\limits_{\mathbb{R}^d} g(y)\nu(\mathrm{d}y), g \in B(\mathbb{R}^d) \right\},$$

and $\mathcal{K}$ is the same as $\mathcal{H}$, but the bounded functions g are all replaced with bounded and continuous functions. $\square$

As was the case for the MOT problem, we can define a monotonicity principle for the MOWT problem that is a necessary condition. First, we state the relevant notion of monotonicity.

Definition 2.4.7 [Martingale C-monotonicity]**.** Let $C : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to [0, \infty)$ be measurable.

(1) We call $\Gamma \subseteq \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$ *martingale C-monotone* if and only if for any $N \in \mathbb{N}$, any collection $(x_1, p_1), \ldots, (x_N, p_N) \in \Gamma$, and $q_1, \ldots, q_N \in \mathcal{P}_1(\mathbb{R}^d)$ such that $\sum_{i=1}^N p_i = \sum_{i=1}^N q_i$ and $\int_{\mathbb{R}^d} y p_i(\mathrm{d}y) = \int_{\mathbb{R}^d} y q_i(\mathrm{d}y)$, we have

$$\sum_{i=1}^N C(x_i, p_i) \leq \sum_{i=1}^N C(x_i, q_i).$$

(2) A probability measure $P \in \mathcal{P}(\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d))$, which is supported on a martingale C-monotone set, is then called *martingale C-monotone*.

(3) A probability measure $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is called martingale C-monotone if $J(\pi)$ is *martingale C-monotone*. Here, the map J is the embedding defined in Definition 2.3.1.

The necessity of martingale C-monotonicity for the MOWT problem is now stated in Theorem 2.4.8.

Theorem 2.4.8. *Let $C : \mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d) \to [0, \infty)$ be measurable and $P^* \in \Lambda_M(\mu, \nu)$ optimal for (MWOT') with finite value. Then P^* is martingale C-monotone. Moreover, if C additionally satisfies for all $x \in \mathbb{R}^d$ and $Q \in \mathcal{P}_r(\mathcal{P}_r(\mathbb{R}^d))$*

$$C(x, I(Q)) \leq \int_{\mathcal{P}(\mathbb{R}^d)} C(x, p) Q(\mathrm{d}p), \tag{2.10}$$

then any optimizer π^* of (MWOT) with finite values is martingale C-monotone. Note that (2.10) *is satisfied if C is convex in second argument.*

Proof. To state this proof we will need more knowledge of the general theory of monotonicity, which is provided in Chapter 3. The proof can be found on page 54. $\square$

If $C(x, \cdot)$ is convex for all $x \in \mathbb{R}^d$, then it is possible to expand the martingale C-monotone set Γ in such a way that it includes all measures that can be written as the empirical mean of measures already in Γ.

Lemma 2.4.9. *Let $C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$ be measurable, $C(x, \cdot)$ is convex for all $x \in \mathbb{R}^d$ and $\Gamma \subseteq \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ be martingale C-monotone with*

$$\Gamma \subseteq \left\{ (x, p) \in \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \,|\, x = \int_{\mathbb{R}^d} y p(\mathrm{d}y) \right\}.$$

Then, the enlarged set

$$\widetilde{\Gamma} = \left\{ \left(x, \frac{1}{k} \sum_{i=1}^k p_i \right) \,\middle|\, (x, p_i) \in \Gamma, i = 1, \ldots, k \in \mathbb{N} \right\},$$

is also martingale C-monotone. In particular if $P \in \mathcal{P}(\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d))$ is martingale C-monotone on Γ, then it is martingale C-monotone on $\widetilde{\Gamma}$ as well.

Proof. Let $(x_1, p_1), \dots, (x_N, p_N) \in \widetilde{\Gamma}$ with competitors $(x_1, q_1), \dots, (x_N, q_N) \in \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$. We can write every p_i as

$$p_i = \frac{1}{k_i} \sum_{n=1}^{k_i} p_i^n, \; k_i \in \mathbb{N}, p_i^n \in \Gamma_{x_i}. \tag{2.11}$$

We can actually assume that $k_i = k_j =: k$ for every $i, j = 1, \dots, N$. To show that this is true, assume this is not the case and look at the least common multiple of $\{k_1, \dots, k_N\}$ and call this k. Define $l_i := \frac{k}{k_i} \in \mathbb{N}$ for each k_i. We can now duplicate each set $\{p_i^1, \dots, p_i^{k_i}\}$, l_i times, from which we see that

$$\frac{1}{k} \sum_{n=1}^{k} p_i^{\overline{n}} = \frac{1}{k_i l_i} \sum_{n=1}^{l_i k_i} p_i^{\overline{n}} = \frac{l_i}{k_i} \sum_{n=1}^{k_i} p_i^n = \frac{1}{k_i} \sum_{n=1}^{k_i} p_i^n = p_i.$$

Where we have set $\overline{n} = ((n-1) \bmod k_i) + 1$ to cycle through the l_i copies of $\{p_i^1, \dots, p_i^{k_i}\}$. So, we see that we can assume that all k_i's are equal. Otherwise, we can apply the above procedure to make them equal, without changing the resulting sum. Now using (2.11) with $k_i = k \in \mathbb{N}$ for every $i = 1, \dots, N$ gives us

$$\begin{aligned}
\sum_{i=1}^{N} C(x_i, p_i) &= \sum_{i=1}^{N} C\left(x_i, \frac{1}{k} \sum_{n=1}^{k} p_i^n\right) \\
&\leq \sum_{i=1}^{N} \frac{1}{k} \sum_{n=1}^{k} C(x_i, p_i^n) = \frac{1}{k} \sum_{i=1}^{N} \sum_{n=1}^{k} C(x_i, p_i^n).
\end{aligned}$$

Where we use the convexity in the second argument of $C(x, \cdot)$ for every $x \in \mathbb{R}^d$ to justify the inequality above. If we copy the original competitors $\{q_1, \dots, q_N\}$, k times, then this results into competitors for the set of measures $\{p_i^n \mid i = 1, \dots, N, n = 1, \dots, k\}$,

$$\sum_{i=1}^{N} \sum_{n=1}^{k} p_i^n = \sum_{i=1}^{n} k p_i = k \sum_{i=1}^{N} p_i = k \sum_{i=1}^{N} q_i = \sum_{i=1}^{N} \sum_{n=1}^{k} q_i.$$

It is also easily checked that means of q_i and p_i^n are equal as

$$\int_{\mathbb{R}^d} y p_i^n(\mathrm{d}y) = x_i = \int_{\mathbb{R}^d} y p_i(\mathrm{d}y) = \int_{\mathbb{R}^d} y q_i(\mathrm{d}y).$$

This shows that the k copies of $\{q_i, \dots, q_N\}$ are competitors of $\{p_i^n \mid i = 1, \dots, N, n = 1, \dots, k\}$. We can now use the fact that each (x_i, p_i^n) was already in the martingale monotone set Γ to find

$$\begin{aligned}
\sum_{i=1}^{N} C(x_i, p_i) &\leq \frac{1}{k} \sum_{i=1}^{N} \sum_{n=1}^{k} C(x_i, p_i^n) \\
&\leq \frac{1}{k} \sum_{i=1}^{N} \sum_{n=1}^{k} C(x_i, q_i) \\
&= \frac{1}{k} \sum_{i=1}^{N} k C(x_i, q_i) = \sum_{i=1}^{N} C(x_i, q_i).
\end{aligned}$$

If P is martingale C-monotone on Γ, then by $\Gamma \subseteq \operatorname{co}(\Gamma)$, we trivially have that $P(\widetilde{\Gamma}) = 1$ and by the first part we also have that $\widetilde{\Gamma}$ is martingale C-monotone.

$\square$

Martingale C-monotonicity is defined by looking for a Γ set in the lifted space $\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d)$. Later on, in the proofs of Theorem 2.4.11 and Lemma 2.4.12 it will be useful to consider a version of Definition 2.4.7 with regard to first component $\mathbb{R}^d$.

Proposition 2.4.10 [Martingale C-monotonicity characterisation]**.** *Let $C : \mathbb{R}^d \times \mathcal{P}\left(\mathbb{R}^d\right) :\to [0, \infty)$ be measurable. A coupling $\pi \in \mathcal{M}(\mu, \nu)$ is martingale C-monotone if and only if there exists a set $\Gamma \subseteq \mathbb{R}^d$ with $\mu(\Gamma) = 1$ such that for any finite number of points $x_1, \dots, x_N \in \Gamma$ and $q_1, \dots, q_N \in \mathcal{P}_1\left(\mathbb{R}^d\right)$ with $\sum_{i=1}^N \pi_{x_i} = \sum_{i=1}^N q_i$ and $\int_{\mathbb{R}^d} y q_i(\mathrm{d}y) = x_i$, we have*

$$\sum_{i=1}^N C(x_i, \pi_{x_i}) \leq \sum_{i=1}^N C(x_i, q_i). \tag{2.12}$$

Proof. "$\Leftarrow$" Consider a martingale C-monotone set $\tilde{\Gamma} \subseteq \mathbb{R}^d$, which we can lift to the space of $\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$ by taking

$$\Gamma = \left\{(x, p) \in \tilde{\Gamma} \times \mathcal{P}_1(\mathbb{R}^d) \,|\, p = \pi_x\right\}.$$

This set is a martingale C-monotone set for $J(\pi)$. To see this, take any finite sequence $(x_1, p_1), \dots, (x_N, p_N) \in \Gamma$. By construction it must be that $x_i = \int_{\mathbb{R}^d} y p_i(\mathrm{d}y)$. We can now check the martingale C-monotonicity by taking competitors $(x_1, q_1), \dots, (x_N, q_N) \in \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$, which gives

$$\sum_{i=1}^N C(x_i, p_i) = \sum_{i=1}^N C(x_i, \pi_{x_i}) \leq \sum_{i=1}^N C(x_i, q_i),$$

by the inequality in (2.12). For this set we also have that

$$J(\pi)(\Gamma) = \int_{\tilde{\Gamma}} \int_{\{p = \pi_x\}} \delta_{\pi_x}(\mathrm{d}p)\mu(\mathrm{d}x) = \mu(\tilde{\Gamma}) = 1.$$

"$\Rightarrow$" We start with $J(\pi)$ being martingale C-monotone as in Definition 2.4.7 with a martingale C-monotone set Γ, having the property that

$$1 = J(\pi)(\Gamma) = \int_{\Gamma} \delta_{\pi_x}(\mathrm{d}p)\mu(\mathrm{d}x) = \mu(\{x \in \mathbb{R}^d \,|\, (x, \pi_x) \in \Gamma\}).$$

The final set in μ is a projection of a Borel set and thus an analytic set by Theorem A.1.10. Additionally, we can write this set as the union of a Borel set and a null set by Theorem A.1.11. In other words, there exists a Borel measurable set $\tilde{\Gamma} \subseteq \{x \in \mathbb{R}^d | (x, \pi_x) \in \Gamma\}$ with $\mu(\tilde{\Gamma}) = 1$. We can now say that π is martingale C-monotone as in Definition 2.4.10 with the set $\tilde{\Gamma}$. $\square$

An incredibly useful result for martingale C-monotonicity is that it is actually a stable notion. This will be one of the main ingredients to show stability of the MOT problem in the next section.

Theorem 2.4.11 [Stability of martingale C-monotonicity]**.** *Let $r \in [1, \infty)$, $C, C_k \in C(\mathbb{R} \times \mathcal{P}_r(\mathbb{R}))$, $k \in \mathbb{N}$, and C_k converges uniformly to C. If $P \in \mathcal{P}_r(\mathbb{R} \times \mathcal{P}_r(\mathbb{R}))$ and $\{P^k\}_{k \in \mathbb{N}}$ is a sequence in $\mathcal{P}_r(\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d))$ such that $P^k \to P$ in the $\mathcal{W}_r$-topology, and the measure P^k is martingale C_k-monotone for all $k \in \mathbb{N}$, then P is martingale C-monotone.*

Moreover, let $\pi \in \mathcal{P}_r(\mathbb{R} \times \mathbb{R})$ and $\{\pi^k\}_{k \in \mathbb{N}}$ be a sequence in $\mathcal{P}_r(\mathbb{R} \times \mathbb{R})$, and assume that $C_k(x, \cdot)$ is convex in the second argument for all $k \in \mathbb{N}$ and $x \in \mathbb{R}$. If it is true that $\pi^k \to \pi$ in the $\mathcal{W}_r$-topology and the measure π^k is martingale C_k-monotone for all $k \in \mathbb{N}$, then π is martingale C-monotone.

Proof. This proof requires many technical arguments and some additional results. That it is why it postponed until the last section of this chapter. The proof can be found on page 43. □

We have now seen how we can rephrase the MOT and MOWT problems as linear constraint problems. This rephrasing allowed us to define monotonicity principles that are related to the MOT and MOWT problem. Using the cost functional $C(x,p) = \int_{\mathbb{R}^d} yp(\mathrm{d}y)$ it is possible to switch between $(c,\mathcal{F}_M)$-monotonicity and martingale C-monotonicity. Lemma 2.4.12 tells us that $(c,\mathcal{F}_M)$-monotonicity and martingale C-monotonicity are actually equivalent.

Lemma 2.4.12 [$(c,\mathcal{F}_M)$-monotonicity and martingale C-monotonicity equivalence]**.** *Let $r \in [1,\infty)$, $\mu,\nu \in \mathcal{P}_r(\mathbb{R})$ such that $\mu \preceq \nu$, $\pi \in \mathcal{M}(\mu,\nu)$, $b \in L^1(\nu)$, $c : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ be jointly measurable, and $c(x,\cdot)$ be upper semicontinuous and $|c(x,y)| \le \alpha(x)|b(y)|$ for all $x \in \mathbb{R}$ for some $\alpha(x) > 0$. Then, π is $(c,\mathcal{F}_M)$-monotone with $\mathcal{F}_M$ as in Lemma 2.4.1 if and only if π is martingale C-monotone with $C(x,p) = \int_{\mathbb{R}} c(x,y)\,p(\mathrm{d}y)$.*

Proof. Again, the proof of this Lemma requires a lot of work. So it is delayed to the end of this chapter. See the proof on page 46. □

2.5. Stability in $\mathbb{R}\times\mathbb{R}$

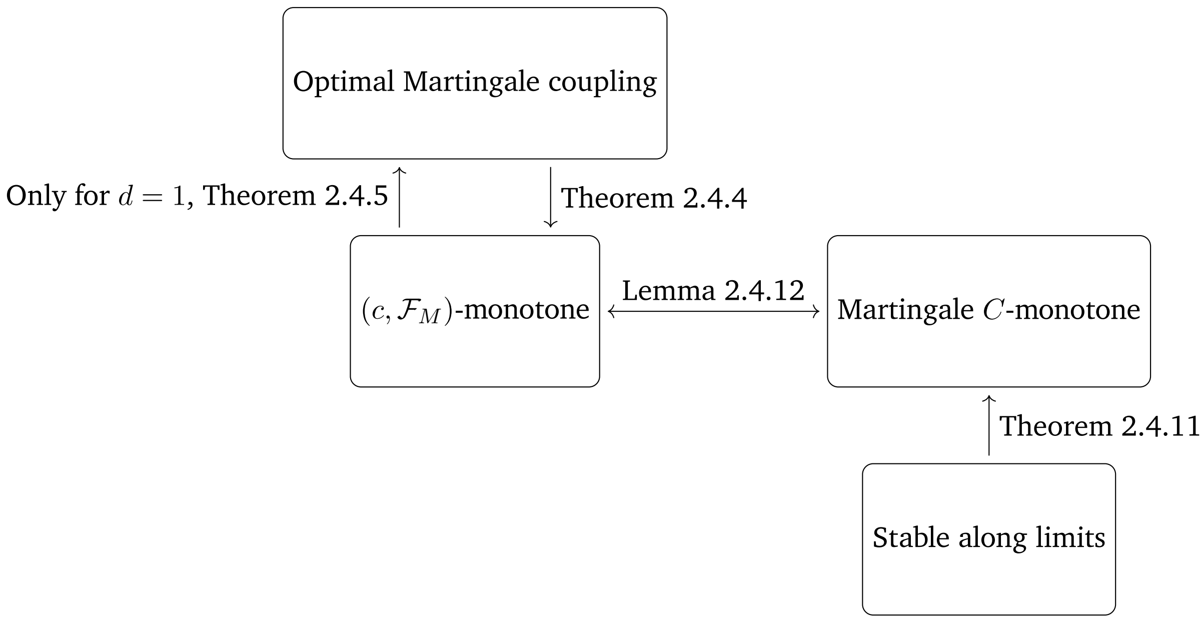


Figure 2.2.: Sketch of the proof of stability with $d = 1$

In the Chapter 1, we introduced a stability result for the OT problem with Theorem 1.4.1. That theorem informally says that finding $V(\mu,\nu)$ and the coupling that attains $V(\mu,\nu)$ are continuous operations. Until recently, it was an open question if a similar result for the MOT problem exists. By the independent works of Backhoff-Veraguas and Pammer in [5] and Wiesel in [55], we now have a positive answer to the question of stability in the martingale case. This result was proven for the special case of of the MOT problem with measures μ and ν whose support is contained in $\mathbb{R}$. Although there are still many unanswered questions for the general case, the outlook for stability in higher dimensions is less hopeful. It was shown in [18] that in the case of $\mathbb{R}^d$ with $d \ge 2$, it is possible to construct a specific sequence that behave unstably. It is not clear however, if it is possible to set certain conditions that would ensure that the MOT is stable in general.

We will examine the result of Backhoff-Veraguas and Pammer, as its statement is the closest to Theorem 1.4.1 and the techniques and tools used in the proof are closely related to the theory of the classical optimal transport problem. The main results of Backhoff-Veraguas and Pammer are given in Theorem 2.5.1 and Corollary 2.5.2.

Theorem 2.5.1 [Stability of MOT]. *Let $c, c_k : \mathbb{R} \times \mathbb{R} \to [0, \infty)$, $k \in \mathbb{N}$, be continuous cost functions such that c_k converges uniformly to c. Let $\{\mu_k\}_{k\in\mathbb{N}}$, $\{\nu_k\}_{k\in\mathbb{N}}$ be sequences in $\mathcal{P}_1(\mathbb{R})$, with $\mu_k \preceq \nu_k$, converging in $\mathcal{W}_1$ to μ and ν, respectively, where we also assume that $\mu \preceq \nu$. For each $k \in \mathbb{N}$ let $\pi_k \in \mathcal{M}(\mu_k, \nu_k)$ be an optimizer of MOT with cost c_k between the marginals μ_k and ν_k. If $c(x, y) \le a(x) + b(y)$ with $a \in L^1(\mu), b \in L^1(\nu)$, and*

$$\limsup_{k\to\infty} \int_{\mathbb{R}\times\mathbb{R}} c_k(x, y)\pi_k(\mathrm{d}x, \mathrm{d}y) < \infty,$$

then any weak accumulation point of $\{\pi_k\}_{k\in\mathbb{N}}$ is an optimizer of MOT for the cost function c. In particular if the latter has a unique optimizer π, then $\pi_k \to \pi$ weakly.

Theorem 2.5.1 shows stability of the minimizing martingale coupling for the MOT problem. As a consequence, it is also true that the value of MOT converges.

Corollary 2.5.2. *Let $c, c_k : \mathbb{R} \times \mathbb{R} \to [0, \infty)$, $k \in \mathbb{N}$, be continuous cost functions such that c_k converges uniformly to c. Let $\{\mu_k\}_{k\in\mathbb{N}}$, $\{\nu_k\}_{k\in\mathbb{N}}$ be sequences in $\mathcal{P}_r(\mathbb{R})$, with $r \in [1, \infty)$, converging in the $\mathcal{W}_r$-topology to μ and ν, respectively, and $\mu_k \preceq \nu_k$ for each $k \in \mathbb{N}$. Suppose that*

$$c(x, y) \le K(1 + |x|^r + |y|^r), \text{ for some } K > 0.$$

Then we have

$$\lim_{k\to\infty} \inf_{\pi\in\mathcal{M}(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} c_k(x, y)\, \pi_k(\mathrm{d}x, \mathrm{d}y) = \inf_{\pi\in\mathcal{M}(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} c(x, y)\, \pi(\mathrm{d}x, \mathrm{d}y).$$

We will give a sketch of the proof stability in one dimension for the MOT problem due to Backhoff-Veraguas and Pammer. What we will actually do is move from the MOT problem to martingale C-monotonicity through $(c, \mathcal{F}_M)$-monotonicity, which is possible because of Theorem 2.4.4 and Lemma 2.4.12. Martingale C-monotonicity is actually stable by virtue of Theorem 2.4.11. Once stability of martingale C-monotonicity is established, we move back to the MOT. This roadmap is visualised Figure 2.2. This last step is only possible in the one dimensional case and is due to Theorem 2.4.5.

With all the results given in the Sections 2.2 and 2.3, we can formulate the proofs of Theorem 2.5.1 and Corollary 2.5.2, the two main results of this thesis.

Proof of Theorem 2.5.1. We can follow the lines in Figure 2.2. By Theorem 2.4.4 it must be that π_k is $(c, \mathcal{F}_M)$-monotone, as each π_k is a minimizer for the MOT problem. Next, we also see that π_k is martingale C-monotone by Lemma 2.4.12. We can apply Lemma 2.4.12 because each c_k is assumed to be continuous and $\{c_k\}_{k\in\mathbb{N}}$ converges uniformly to c with $c(x, y) \le a(x) + b(y)$ for some $a \in L^1(\nu), b \in L^1(\mu)$. The uniform convergence allows us to bound each c_k with integrable functions. The martingale C-monotone property is retained along the limit because of Theorem 2.4.11. This means that π is also martingale C-monotone. Note that each C_k is defined as

$$C_k(x,p) = \int_{\mathbb{R}} c_k(x,y)p(\mathrm{d}y),$$

and C is defined analogously. The function C_k is linear in the second argument, meaning that C_k is convex in the second argument. Furthermore, we can use the uniform convergence of $\{c_k\}_{k\in\mathbb{N}}$ to show uniform convergence of $\{C_k\}_{k\in\mathbb{N}}$ to C. Let $\epsilon > 0$, then we can find a $k_0 \in \mathbb{N}$ such that $k \geq k_0$ implies that

$$\sup_{(x,y)\in\mathbb{R}\times\mathbb{R}} |c_k(x,y) - c(x,y)| < \epsilon.$$

Using this we get, for $k \geq k_0$,

$$\begin{aligned}
\sup_{(x,p)\in\mathbb{R}\times\mathcal{P}_1(\mathbb{R})} |C_k(x,p) - C(x,p)| &= \sup_{(x,p)\in\mathbb{R}\times\mathcal{P}_1(\mathbb{R})} \left| \int_{\mathbb{R}} c_k(x,y) - c(x,y)p(\mathrm{d}y) \right| \\
&\leq \sup_{p\in\mathcal{P}_1(\mathbb{R})} \int_{\mathbb{R}} \sup_{x\in\mathbb{R}} |c_k(x,y) - c(x,y)| p(\mathrm{d}y) \\
&< \sup_{p\in\mathcal{P}_1(\mathbb{R})} \int_{\mathbb{R}} \epsilon p(\mathrm{d}y) = \epsilon.
\end{aligned}$$

So, we see that $\{C_k\}_{k\in\mathbb{N}}$ converges uniformly to C. By assumption we have that π_k converges in the $\mathcal{W}_1$-topology to π. It is now possible to apply Theorem 2.4.11, which tells us that π is martingale C-monotone. We can now move back and say that π must be $(c, \mathcal{F}_M)$-monotone by a second application of Lemma 2.4.12. As a final step we can conclude that π is optimal for MOT by an application of Theorem 2.4.5. □

Proof of Corollary 2.5.2. By Theorem 2.5.1 we already have that any accumulation point of $\{\pi^k\}_{k\in\mathbb{N}}$ is an optimizer for MOT. Let d be the metric on $\mathbb{R}\times\mathbb{R}$ given by

$$d(x,y)^r = |x_1 - y_1|^r + |x_2 - y_2|^r,\ x = (x_1, x_2), y = (y_1, y_2).$$

This metric gives rise to the Wasserstein-r space $\mathcal{P}_r(\mathbb{R}\times\mathbb{R})$. It follows that

$$\begin{aligned}
\lim_{k\to\infty} \int_{\mathbb{R}\times\mathbb{R}} d(0,(x,y))^r \pi_k(\mathrm{d}x,\mathrm{d}y) &= \lim_{k\to\infty} \int_{\mathbb{R}\times\mathbb{R}} |x|^r \mu_k(\mathrm{d}x) + \int_{\mathbb{R}} |y|^r \nu_k(\mathrm{d}y) \\
&= \int_{\mathbb{R}} |x|^r \mu(\mathrm{d}x) + \int_{\mathbb{R}} |y|^r \nu(\mathrm{d}y) = \int_{\mathbb{R}x\mathbb{R}} d(0,(x,y))^r \pi(\mathrm{d}x,\mathrm{d}y).
\end{aligned}$$

By Theorem 1.5.5 we can now say that the accumulation points of $\{\pi_k\}_{k\in\mathbb{N}}$ under the $\mathcal{W}_r$-topology coincide with the accumulation points under the weak topology. We will now bound the limsup and liminf to find the limit. For the liminf we immediately have

$$\liminf_{k\to\infty} \int_{\mathbb{R}\times\mathbb{R}} c_k(x,y)\pi_k(\mathrm{d}x,\mathrm{d}y) \geq \inf_{\pi\in\mathcal{M}(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} c(x,y)\pi(\mathrm{d}x,\mathrm{d}y).$$

Let π^* be an accumulation point of $\{\pi_k\}_{k\in\mathbb{N}}$. Then, there exists some subsequence π_{k_n} such that $\pi_{k_n} \to \pi^*$ in the $\mathcal{W}_r$-topology. As $c_k \to c$ uniformly and $c(x,y) \leq K(1+|x|^r+|y|^r) = K(1+d((0,0),(x,y))^r$ for some $K \geq 0$, we can use the $\mathcal{W}_r$ convergence to show that

$$\lim_{n\to\infty} \int\limits_{\mathbb{R}\times\mathbb{R}} c_{k_n}(x,y)\pi_{k_n}(\mathrm{d}x,\mathrm{d}y) = \int\limits_{\mathbb{R}\times\mathbb{R}} c(x,y)\pi^*(\mathrm{d}x,\mathrm{d}y).$$

Namely, let $\epsilon > 0$ and find a $N_0 \in \mathbb{N}$ such that $k_n \geq N_0$ implies that

$$\sup_{(x,y)\in\mathbb{R}\times\mathbb{R}} |c_{k_n}(x,y) - c(x,y)| < \frac{\epsilon}{2}.$$

Additionally, by the $\mathcal{W}_r$-convergence we can find a $N_1 \in \mathbb{N}$ such that $k_n \geq N_1$ implies that

$$\left| \int\limits_{\mathbb{R}} c(x,y)(\pi_{k_n} - \pi^*)(\mathrm{d}x,\mathrm{d}y) \right| < \frac{\epsilon}{2}.$$

If we pick $N = \max\{N_0, N_1\}$, we see for $k_n \geq N$ that

$$\begin{aligned}
\left| \int\limits_{\mathbb{R}\times\mathbb{R}} c_{k_n}(x,y)\pi_{k_n}(\mathrm{d}x,\mathrm{d}y) - \int\limits_{\mathbb{R}\times\mathbb{R}} c(x,y)\pi^*(\mathrm{d}x,\mathrm{d}y) \right| &\leq \left| \int\limits_{\mathbb{R}\times\mathbb{R}} c_{k_n}(x,y)\pi_{k_n}(\mathrm{d}x,\mathrm{d}y) \right. \\
&\left. - \int\limits_{\mathbb{R}\times\mathbb{R}} c(x,y)\pi_{k_n}(\mathrm{d}x,\mathrm{d}y) + \int\limits_{\mathbb{R}\times\mathbb{R}} c(x,y)\pi_{k_n}(\mathrm{d}x,\mathrm{d}y) - \int\limits_{\mathbb{R}\times\mathbb{R}} c(x,y)\pi^*(\mathrm{d}x,\mathrm{d}y) \right| \\
&\leq \int\limits_{\mathbb{R}\times\mathbb{R}} |c_{k_n}(x,y) - c(x,y)|\pi_{k_n}(\mathrm{d}x,\mathrm{d}y) + \int\limits_{\mathbb{R}\times\mathbb{R}} |c(x,y)|(\pi_{k_n} - \pi^*)(\mathrm{d}x,\mathrm{d}y) \\
&< \int\limits_{\mathbb{R}\times\mathbb{R}} \frac{\epsilon}{2}\pi_{k_n}(\mathrm{d}x,\mathrm{d}y) + \frac{\epsilon}{2} = \epsilon.
\end{aligned}$$

From this we can see that

$$\limsup_{k\to\infty} \int\limits_{\mathbb{R}\times\mathbb{R}} c_k(x,y)\pi_k(\mathrm{d}x,\mathrm{d}y) \leq \int\limits_{\mathbb{R}\times\mathbb{R}} c(x,y)\pi^*(\mathrm{d}x,\mathrm{d}y) = \inf_{\pi\in\mathcal{M}(\mu,\nu)} \int\limits_{\mathbb{R}\times\mathbb{R}} c(x,y)\pi(\mathrm{d}x,\mathrm{d}y).$$

We now have both necessary bounds on the $\limsup$ and $\liminf$ to conclude

$$\lim_{k\to\infty} \inf_{\pi\in\mathcal{M}(\mu,\nu)} \int\limits_{\mathbb{R}\times\mathbb{R}} c_k(x,y)\pi_k(\mathrm{d}x,\mathrm{d}y) = \inf_{\pi\in\mathcal{M}(\mu,\nu)} \int\limits_{\mathbb{R}\times\mathbb{R}} c(x,y)\pi(\mathrm{d}x,\mathrm{d}y).$$

□

2.6. Remaining proofs

In this section we will finish up the remaining proofs of this chapter, which are the proofs of Theorem 2.4.11 and Lemma 2.4.12. We delayed these proofs, because they are quite lengthy and require some additional results.

2.6.1. Competitor sequences

To carry the martingale C-monotonicity property through to the limit in Theorem 2.4.11 we need to be able to find suitable competitors in each step of the sequence. Lemma 2.6.1 ensures that this is always possible.

Lemma 2.6.1. *Let $N \in \mathbb{N}$ and $p_i \in \mathcal{P}_r(\mathbb{R})$ with competitor $q_i \in \mathcal{P}_r(\mathbb{R})$, $i = 1, \dots N$, i.e.*

$$\sum_{i=1}^{N} p_i = \sum_{i=1}^{N} q_i, \tag{2.13}$$

$$\int_{\mathbb{R}} y p_i(\mathrm{d}y) = \int_{\mathbb{R}} y q_i(\mathrm{d}y) \quad i = 1, \dots, N. \tag{2.14}$$

Suppose there are sequences $\{p_1^k, \dots, p_N^k\}$, $k \in \mathbb{N}$, of measures in $\mathcal{P}_r(\mathbb{R})$ with $p_i^k \to p_i$ in the $\mathcal{W}_r$-topology. Then, there exist approximative sequences $\{q_1^k, \dots, q_N^k\}_{k \in \mathbb{N}}$ of competitors, i.e.

$$\sum_{i=1}^{N} p_i^k = \sum_{i=1}^{N} q_i^k, \tag{2.15}$$

$$\int_{\mathbb{R}} y p_i^k(\mathrm{d}y) = \int_{\mathbb{R}} y q_i^k(\mathrm{d}y), \tag{2.16}$$

$$q_i^k \to q_i \text{ in } \mathcal{W}_r \quad i = 1, \dots, N,$$

for all $k \in \mathbb{N}$.

Before we start the proof of Lemma 2.6.1, we need to show the following intermediate result.

Lemma 2.6.2. *Let $s \in \mathbb{R} \cup \{\infty\}$, $F_p : \mathbb{R} \to [0, 1]$ the cumulative distribution of $p \in \mathcal{P}(\mathbb{R})$, let $\{p_i\}_{i=1}^N$ and $\{q_i\}_{i=1}^N$ with $p_i, q_i \in \mathcal{P}(\mathbb{R})$ for some $N \in \mathbb{N}$ and which satisfy* (2.13) *and* (2.14)*. Define*

$$I_s^1 := \{i \in \{1, \dots, N\} \mid F_{p_i}(s) = 1\}, \quad I_s^2 := \{i \in \{1, \dots, N\} \mid F_{q_i}(s) = 1\}.$$

If

$$j \in \{1, \dots, N\} \setminus I_s^1 \implies p_j((-\infty, s)) = 0, \tag{2.17}$$

then, $I_s^2 = I_s^1$ and

$$j \in \{1, \dots, N\} \setminus I_s^2 \implies q_j((-\infty, s)) = 0. \tag{2.18}$$

Proof. By the equality of the means and (2.17), we immediately have

$$\begin{aligned} 0 &= \sum_{i \in I_s^1} \int_{\mathbb{R}} y p_i(\mathrm{d}y) - \int_{\mathbb{R}} y q_i(\mathrm{d}y) \\ &= \int_{\mathbb{R}} y (\sum_{i \in I_s^1} p_i - q_i)^+(\mathrm{d}y) - \int_{\mathbb{R}} y (\sum_{i \in I_s^1} q_i - p_i)^+(\mathrm{d}y). \end{aligned} \tag{2.19}$$

Where $(\cdot)^+$ denotes the positive part of a signed measure. Next, we can use (2.13) and (2.17) to get

$$\sum_{i \in I_s^1} q_i|_{(-\infty,s)} \leq \sum_{i=1}^{N} q_i|_{(-\infty,s)} \overset{(2.13)}{=} \sum_{i=1}^{N} p_i|_{(-\infty,s)} \overset{(2.17)}{=} \sum_{i \in I_s^1} p_i|_{(-\infty,s)}.$$

From which we see that $(\sum_{i\in I_s^1} p_i - q_i)$ has non-negative values only on $(-\infty, s)$ and is negative otherwise. It follows that $(\sum_{i\in I_s^1} p_i - q_i)^+$ is concentrated on $(-\infty, s)$. By swapping q_i and p_i and noticing that the sign of the sum then changes, we see that $(\sum_{i\in I_s^1} q_i - p_i)^+$ is concentrated on $[s, \infty)$. Together with (2.19) this shows that $(\sum_{i\in I_s^1} p_i - q_i)^+ = (\sum_{i\in I_s^1} q_i - p_i)^+ = 0$, which we will show by contradiction. Assume that this is not the case. This would mean that $(\sum_{i\in I_s^1} p_i - q_i)$ has positive mass in $(-\infty, s)$ and $(\sum_{i\in I_s^1} q_i - p_i)$ in $[s, \infty)$. In particular, we can use the disjoint supports to bound the barycenters,

$$
\begin{aligned}
\int_{\mathbb{R}} y(\sum_{i\in I_s^1} p_i - q_i)^+(\mathrm{d}y) - \int_{\mathbb{R}} y(\sum_{i\in I_s^1} q_i - p_i)^+(\mathrm{d}y)& \\
&< s(\sum_{i\in I_s^1} p_i - q_i)^+((-\infty, s)) - s(\sum_{i\in I_s^1} q_i - p_i)^+([s, \infty)) \\
&= s\left(\sum_{i\in I_s^1} p_i((-\infty, s)) - q_i((-\infty, s)) - q_i([s, \infty)\right) \\
&= s\left(|I_s^1| - \sum_{i\in I_s^1} q_i(\mathbb{R})\right) = 0,
\end{aligned}
$$

which is a contradiction with (2.19). So, we see that $(\sum_{i\in I_s^1} p_i - q_i)^+ = (\sum_{i\in I_s^1} q_i - p_i)^+ = 0$, which implies that $\sum_{i\in I_s^1} p_i = \sum_{i\in I_s^2} q_i$ and $I_s^1 \subseteq I_s^2$. To show equality of the index sets I_s^1 and I_s^2, we assume that $j \in I_s^2 \setminus I_s^1$. We have the following chain of equalities

$$
\sum_{i=1}^N q_i|_{(-\infty,s)} = \sum_{i=1}^N p_i|_{(-\infty,s)} = \sum_{i\in I_s^1} p_i|_{(-\infty,s)} = \sum_{i\in I_s^2} q_i|_{(-\infty,s)}.
$$

This already shows that we must have that $q_j((-\infty, s)) = 0$. However, $j \in I_s^2$ implies that $q_j((s, \infty)) = 0$. From these two facts together, we infer that $q_j = \delta_s$. The measure q_j is a competitor for p_j. So, their means must coincide, which shows that the mean of p_j is equal to s. We should then also have that p_j gives some mass to $(-\infty, s)$. However, this violates (2.17), as $j \notin I_s^1$. This gives us $I_s^1 = I_s^2$. Now, assume that $j \in \{1, \ldots, N\} \setminus I_s^2$, then q_j cannot have any mass on $(-\infty, s)$. By (2.13) and $I_s^1 = I_s^2$ we must have the equalities

$$
\sum_{i\in I_s^1} p_i = \sum_{i=1}^N p_i|_{(-\infty,s)} = \sum_{i=1}^N q_i|_{(-\infty,s)} = \sum_{i\in I_s^1} q_i.
$$

We conclude that $q_j(-\infty, s) > 0$ is not possible and (2.18) is satisfied. □

Proof of Lemma 2.6.1. We will break this proof down into three parts. First, we will show that that there exist approximating sequences $\{q_1^k, \ldots, q_N^k\}$ which satisfy condition (2.15). Once we have these approximative sequences, we will show that given some assumptions on the support of pairs of measures (q_i^k, q_j^k), we can correct the mean of q_i^k, using q_j^k, to make it equal to the mean of p_i^k. Finally, it will be shown that we can actually iterate this correcting technique of the measures until all the means are correct.

For all steps in the proof, We will assume that each measure is not a discrete measure supported on one point, because then it would also be clear how to construct the competitors. The reason being that we will take the convex hull of the support at some point, which we want to have non-zero Lebesque measure.

Step 1: Construct sequences $\{q_1^k, \ldots, q_N^k\}$ with $q_i^k \to q_i$ whenever $k \to \infty$ with $\sum_i q_i^k = \sum_i p_i^k$.

In this step we can explicitly construct the needed sequences. Call $R = \sum_i q_i = \sum_i p_i$. From this it is easy to see that $q_i \ll R$ for all $i = 1, \ldots, N$. By the Radon-Nykodim theorem we can find a measurable function $Z_i = \frac{\mathrm{d}q_i}{\mathrm{d}R}$, such that for every measurable A

$$q_i(A) = \int_A q_i(\mathrm{d}x) = \int_A Z_i(x)R(\mathrm{d}x) = \sum_{j=1}^N \int_A Z_i(x)q_j(\mathrm{d}x).$$

From this we can see that $0 \leq Z_i \leq 1$ as $0 \leq q_i(A) \leq 1$ and q_i will appear on the left hand side as well. Define subprobabilities $m_{i,j}$ as follows. For each measurable A

$$m_{i,j}(A) = \int_A Z_j(x)p_i(\mathrm{d}x).$$

With a slight abuse of notation we will write this as $m_{i,j} = Z_j p_i$. Note that these $m_{i,j}$ are indeed subprobabilities as they are measures by definition and

$$m_{i,j}(A) = \int_A Z_j(x)p_i(\mathrm{d}x) \leq p_i(A), \quad \text{because } 0 \leq Z_j \leq 1.$$

These subrobabilities have the following 2 properties:

$$\sum_{i=1}^N m_{i,j} = Z_j \sum_{i=1}^N p_i = Z_j R = q_j, \tag{2.20}$$

$$\sum_{j=1}^N m_{i,j} = p_i \sum_{j=1}^N Z_j = p_i \frac{\mathrm{d}(q_1 + \ldots + q_N)}{\mathrm{d}R} = p_i. \tag{2.21}$$

Next, by Theorem 1.2.6 we know that the value $\mathcal{W}_r(p_i, p_i^k)$ is attained by some coupling between p_i and p_i^k. Call this coupling $\chi^{k,i}$, with regular disintegration $\{\chi_z^{k,i}\}_{z \in \mathbb{R}}$ with respect to p_i, which exists by virtue of Theorem A.1.14. We are now ready to define the q_i^k as follows:

$$q_{i,j}^k(\mathrm{d}y) := \int_{\mathbb{R}} \chi_z^{k,i}(\mathrm{d}y)m_{i,j}(\mathrm{d}z),$$

$$q_i^k := \sum_{j=1}^N q_{i,j}^k.$$

For this latter sum we see

$$\begin{aligned}
\sum_{j=1}^N |\mathcal{W}_r(q_{i,j}^k, m_{i,j})|^r &= \sum_{j=1}^N \inf_{\pi \in \Pi(q_{i,j}^k, m_{i,j})} \int_{\mathbb{R}\times\mathbb{R}} d(z,y)^r \pi(\mathrm{d}z, \mathrm{d}y) \\
&\leq \sum_{j=1}^N \int_{\mathbb{R}\times\mathbb{R}} d(z,y)^r \chi_z^{k,i}(\mathrm{d}y)m_{i,j}(\mathrm{d}z) \\
&= \int_{\mathbb{R}\times\mathbb{R}} d(z,y)^r \chi_z^{k,i}(\mathrm{d}y)p_i(\mathrm{d}z) = \mathcal{W}_r(p_i, p_i^k) \to 0.
\end{aligned}$$

Here, the inequality follows from the definition of $q^k_{i,j}$ and the second to last equality follows from property 2.21 if the subprobabilities $m_{i,j}$. This tells us that each $q^k_{i,j}$ converges to $m_{i,j}$, which we can use to prove the convergence of q^k_j to q_j:

$$\begin{aligned}\mathcal{W}_r(q_j, q^k_j) &= \mathcal{W}_r\left(\sum_{i=1}^N m_{i,j}, \sum_{j=1}^N q^k_{i,j}\right)\\ &\leq \sum_{i=1}^N\sum_{j=1}^N \mathcal{W}_r(m_{i,j}, q^k_{i,j}) \to 0.\end{aligned}$$

It remains to check that these q^k_i indeed satisfy the sum condition

$$\begin{aligned}\sum_{i=1}^N q^k_i &= \sum_{i=1}^N\sum_{j=1}^N q^k_{i,j}\\ &= \sum_{i=1}^N\sum_{j=1}^N \int_{z\in\mathbb{R}} \chi^{k,i}_z(\mathrm{d}y) m_{i,j}(\mathrm{d}z)\\ &= \sum_{i=1}^N \int_{z\in\mathbb{R}} \chi^{k,i}_z(\mathrm{d}y) p_i(\mathrm{d}z)\\ &\overset{(2.21)}{=} \sum_{i=1}^N \int_{z\in\mathbb{R}} \chi^{k,i}(\mathrm{d}y, \mathrm{d}z) = \sum_{i=1}^N p^k_i.\end{aligned}$$

Where the second to last equality follows from $\{\chi^{k,i}_z\}_{z\in\mathbb{R}}$ being a regular disintegration with respect to p_i. The final equality is then just marginalizing the coupling.

So the $\{q^k_1, \ldots, q^k_N\}$ all converge to $\{q_1, \ldots, q_N\}$ in $\mathcal{W}_r$ and satisfy the sum condition for being feasible competitors of $\{p^k_1, \ldots, p^k_N\}$.

However, in general the means of $\{q^k_1, \ldots, q^k_N\}$ will not be correct, in the sense that they do not necessarily satisfy condition (2.16). The next two steps of the proof will be devoted to creating a procedure that allows us to correct the means iteratively, while keeping condition (2.15) in tact. To do this we first define

$$S_i := \mathrm{co}(\mathrm{supp}(q_i)), \quad i = 1, \ldots, N.$$

Where $\mathrm{co}(\cdot)$ denotes the convex hull of a set.

Step 2, if

$$\lambda(S_i \cap S_j) > 0 \text{ or } S_j \subseteq \mathrm{int}(S_i), \tag{2.22}$$

then we can ensure that the mean of q^k_i is the same as p^k_i . Consider the sequences $\{q^k_i\}_{k\in\mathbb{N}}$ and $\{q^k_j\}_{k\in\mathbb{N}}$ obtained in the previous step, which converge to q_i and q_j in $\mathcal{W}_r$ respectively.

In Figure 2.3 it is shown how the two cases of (2.22) could look like. As shown in the picture let $O^+_{i,j}, O^-_{i,j}, O^+_{j,i}$ and $O^-_{j,i}$ be open intervals around the points $\inf S_i$, $\sup S_i$, $\inf S_j$ and $\sup S_j$ respectively. These points are all actually minima and maxima, because the convex hull of a 1-dimensional closed set is also closed. This means that these points are also part of the supports of q_i and q_j. This ensures that any open interval around them has positive measure. Finally, we can pick $O^+_{i,j}, O^-_{i,j}, O^+_{j,i}$ and $O^-_{j,i}$ small enough so that

$$\begin{gathered} q_i(O^{\pm}_{i,j}) > 0,\ q_j(O^{\pm}_{j,i}) > 0,\\ \sup O^+_{i,j} < \inf O^-_{j,i},\ \sup O^+_{j,i} < \inf O^-_{i,j}.\end{gathered}$$

Remark that condition (2.22) is possible to occur and we are able to find the sets $O^+_{i,j}, O^-_{i,j}, O^+_{j,i}$ and $O^-_{j,i}$, because we assumed that all measures $\{q_1, \ldots, q_n\}$ are not Dirac measures. By the Portmanteau theorem (Theorem A.1.4), we also have that

$$\liminf_{k\to\infty} q_i^k(O^{\pm}_{i,j}) > 0, \ \liminf_{k\to\infty} q_j^k(O^{\pm}_{j,i}) > 0. \tag{2.23}$$

As both q_i and q_j have the same means as p_i and p_j by assumption we have

$$\begin{aligned} \lim_{k\to\infty} \left| \int_{\mathbb{R}} y p_i^k(\mathrm{d}y) - \int_{\mathbb{R}} y q_i^k(\mathrm{d}y) \right| = 0, \\ \lim_{k\to\infty} \left| \int_{\mathbb{R}} y p_j^k(\mathrm{d}y) - \int_{\mathbb{R}} y q_j^k(\mathrm{d}y) \right| = 0. \end{aligned} \tag{2.24}$$

We can now define the corrected measures $\tilde{q}_j^k$ and $\tilde{q}_i^k$. We will explain what the constants $\alpha_+^k \geq 0$ and $\alpha_-^k \geq 0$ are afterwards. The corrected measures are given by

$$\begin{aligned} \tilde{q}_j^k &= q_j^k + \alpha_+^k \left(\frac{q_i^k|_{O^-_{i,j}}}{q_i^k(O^-_{i,j})} - \frac{q_j^k|_{O^+_{j,i}}}{q_j^k(O^+_{j,i})} \right) + \alpha_-^k \left(\frac{q_i^k|_{O^+_{i,j}}}{q_i^k(O^+_{i,j})} - \frac{q_j^k|_{O^-_{j,i}}}{q_j^k(O^-_{j,i})} \right), \\ \tilde{q}_i^k &= q_i^k - \tilde{q}_j^k + q_j^k. \end{aligned} \tag{2.25}$$

The limit in (2.23) ensures that the fractions in (2.25) are well defined for $k \geq k_0$ for some $k_0 \in \mathbb{N}$. Additionally, by (2.24) we can find a $k_1 \geq k_0$ large enough for which we can pick $\{\alpha_+^k\}_{k\geq k_0}$ and $\{\alpha_-^k\}_{k>k_0}$ such that $\tilde{q}_i^k$ and $\tilde{q}_j^k$ are probabilities and

$$\int_{\mathbb{R}} y \tilde{q}_i^k(\mathrm{d}y) = \int_{\mathbb{R}} y p_i^k(\mathrm{d}y).$$

Here, we set $\alpha_+^k > 0$ and $\alpha_-^k = 0$ if the mean of p_j^k is larger than that of q_j^k and $\alpha_+^k = 0$ and $\alpha_-^k > 0$ otherwise. By (2.24) we also have that both $\alpha_+^k \to 0$ and $\alpha_-^k \to 0$ and we already know that $q_i^k \to q_i$. So we also have that $\tilde{q}_i^k \to q_i$ in the $\mathcal{W}_r$-topology. If we let the sequences now start from k_1 and replace q_i^k and q_j^k with $\tilde{q}_i^k$ and $\tilde{q}_j^k$, then $\{q_i^k\}_{k\in\mathbb{N}}$ will be a correct competitor sequence for p_i^k.

What we have done now is used some of the mass of q_j^k to adjust the mean of q_i^k in such a way that it is equal to the mean of p_i^k. So using the pair (q_j^k, q_i^k) we can correct one of the means whenever condition (2.22) is met. In the next step, we will show that we can iteratively repeat this method, so that we can correct all the means.

Step 3: All the means can be corrected. Let $s \in \mathbb{R} \cup \{\infty\}$ be such that either (2.17) or (2.18) holds, then we have that $I_s^1 = I_s^2 =: I_s$ and $\sum_{i\in I_s} p_i = \sum_{i\in I_s} q_i$. We now have two index sets, I_s

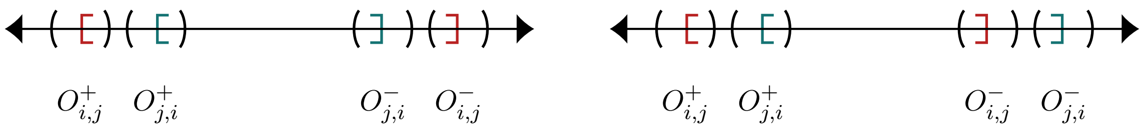


Figure 2.3.: Construction of the sets $O^+_{i,j}O^-_{i,j}, O^-_{j,i}, O^-_{j,i}$. The red brackets indicate S_i and the blue brackets indicate S_j

and $\{1,\ldots,N\}\setminus I_s$, which indicate two sets of probability measures with disjoint supports. We can thus split the problem of correcting the means of q_i^k with $i\in\{1,\ldots,N\}$ into doing this for q_i^k with $i\in I_s$ and then for $i\in\{1,\ldots,N\}\setminus I_s$. The procedures in both cases are identical. So, we can assume without loss of generality that $\{1,\ldots,N\}=I_s$ and that s is minimal.

To finish up the proof, we introduce the notion of a pivot.

Definition 2.6.3 [Pivot]**.** An index $i\in I_s$ is an *pivot*, if there exists a $k_0\in\mathbb{N}$ such that for all $n\in\{1,\ldots,N\}$ with $\sup S_n\le\inf S_i$, the means of q_n^k satisfy (2.16) for $k\ge k_0$.

If i is a pivot, then we know that all the measures that are supported to the left of S_i are already correct. Let us consider the case that i is a pivot. By step 2 of this proof, we can consider each $j\neq i$ for which (2.22) holds and whose means are not corrected yet. For these indices we can find a $k_1\in\mathbb{N}$ such that

$$k\ge k_1 \implies \int_{\mathbb{R}} y q_j^k(\mathrm{d}y)=\int_{\mathbb{R}} y p_j^k(\mathrm{d}y).$$

There are now two cases to consider:

(1) The requirement (2.22) holds for all $j\neq i$ with $\inf S_i<\sup S_j$. These are all the indices that are not correct yet, by the assumption that i is a pivot. We can thus correct all these means, by step 2 of the this proof, for k large enough. We only need to correct the mean of q_i^k itself now. However, it turns out, that this is already the case

$$\int_{\mathbb{R}} y q_i^k(\mathrm{d}y)=\sum_{n=1}^N\int_{\mathbb{R}} y p_n^k(\mathrm{d}y)-\sum_{j\neq i}\int_{\mathbb{R}} y q_j^k(\mathrm{d}y)=\int_{\mathbb{R}} y p_i^k(\mathrm{d}y).$$

Now, we are actually done and have our desired sequence of competitors

(2) There are indices $m\in\{1,\ldots,N\}$, such that S_m and S_i do not satisfy (2.22). This means that S_m is disjoint from S_i and by i being a pivot, we also know that S_m must be to the right of S_i, i.e. $\sup S_i<\inf S_m\le\sup S_m$. By minimality of s we can find another index $l\in\{1,\ldots,N\}$ such that $\inf S_l<\sup S_m<\sup S_l$. If this were not the case, then we could actually take $\tilde{s}=\inf S_m<s$ and consider the set $I_{\tilde{s}}$, for which (2.17) and (2.18) holds. This contradicts the minimality of s.
For these S_m and S_l we see that 2.22 holds and we can use $\{q_l^k\}_{k\in\mathbb{N}}$ to correct the means of $\{q_m^k\}_{k\in\mathbb{N}}$ for k large enough. It also turns out that l is a pivot. Take any $\hat{l}\in\{1,\ldots,N\}$ with $\sup S_{\hat{l}}\le\inf S_i$. We either have that $\sup S_{\hat{l}}\le\inf S_i$ as well, for which we know that the means are already correct. Otherwise, we have that $\inf S_i<\sup S_{\hat{l}}\le\inf S_l<\sup S_i$, but in this case we see that $S_{\hat{l}}$ satisfies (2.22). In the previous step the mean would thus already be corrected. Finally, note that

$$\sup S_i<\sup S_l. \tag{2.26}$$

In other words. We moved the pivot more to the right.

We can now start with i, such that $\inf S_i$ is minimal. We can iterate the above steps and move the pivot index each step at least one set to the right by (2.26). As there are only finitely many sets, this procedure is ensured to end, which means that we have found our sequences. □

2.6.2. Proof of martingale C-monotonicity

As a refresher of our memory, we repeat Theorem 2.4.11.

Theorem 2.4.11 [Stability of martingale C-monotonicity]. *Let $r \in [1,\infty)$, $C, C_k \in C(\mathbb{R}\times\mathcal{P}_r(\mathbb{R}))$, $k \in \mathbb{N}$, and C_k converges uniformly to C. If $P \in \mathcal{P}_r(\mathbb{R}\times\mathcal{P}_r(\mathbb{R}))$ and $\{P^k\}_{k\in\mathbb{N}}$ is a sequence in $\mathcal{P}_r(\mathbb{R}^d\times\mathcal{P}_r(\mathbb{R}^d))$ such that $P^k \to P$ in the $\mathcal{W}_r$-topology, and the measure P^k is martingale C_k-monotone for all $k \in \mathbb{N}$, then P is martingale C-monotone.*

Moreover, let $\pi \in \mathcal{P}_r(\mathbb{R}\times\mathbb{R})$ and $\{\pi^k\}_{k\in\mathbb{N}}$ be a sequence in $\mathcal{P}_r(\mathbb{R}\times\mathbb{R})$, and assume that $C_k(x,\cdot)$ is convex in the second argument for all $k\in\mathbb{N}$ and $x\in\mathbb{R}$. If it is true that $\pi^k \to \pi$ in the $\mathcal{W}_r$-topology and the measure π^k is martingale C_k-monotone for all $k\in\mathbb{N}$, then π is martingale C-monotone.

Proof. For the first part, it will be our goal to find a martingale C-monotone set for P from the martingale C_k-monotone sets $\tilde{\Gamma}^k$ of P_k. For all $\epsilon \geq 0$, we can define the following set

$$\tilde{\Gamma}^{\epsilon}_N = \left\{ (x_i,p_i)_{i=1}^N \in (\mathbb{R}\times\mathcal{P}_r(\mathbb{R}))^N \middle| \forall m_1,\dots,m_N \in \mathcal{P}_r(\mathbb{R}) \text{ s.t. } \sum_{i=1}^N p_i = \sum_{i=1}^N m_i \text{ and} \right.$$

$$\left. \int_{\mathbb{R}} y m_i(\mathrm{d}y) = \int_{\mathbb{R}} y p_i(\mathrm{d}y) \text{ for } i=1,\dots,N, \text{ we have } \sum_{i=1}^N C(x_i,p_i) \leq \sum_{i=1}^N C(x_i,m_i)+\epsilon \right\}.$$

This set is closed. Indeed, take any convergent sequence $\big((x_i^k,p_i^k)_{i=1}^N\big)_{k\in\mathbb{N}} \in \tilde{\Gamma}^{\epsilon}_N$ with limit $(x_i,p_i)_{i=1}^N$ and any competitor $(x_i,q_i)_{i=1}^N$ for $(x_i,p_i)_{i=1}^N$. We can use Lemma 2.6.1 to find a sequence of competitors $\big((x_i^k,q_i^k)_{i=1}^N\big)_{k\in\mathbb{N}}$ converging to $(x_i,q_i)_{i=1}^N$. The continuity of C gives us

$$\begin{aligned}\sum_{i=1}^N C(x_i,p_i) &= \lim_{k\to\infty}\sum_{i=1}^N C(x_i^k,p_i^k)\\ &\leq \lim_{k\to\infty}\sum_{i=1}^N C(x_i^k,q_i^k)+\epsilon = \sum_{i=1}^N C(x_i,q_i)+\epsilon,\end{aligned}$$

showing that $(x_i,p_i)_{i=1}^N \in \tilde{\Gamma}^{\epsilon}_N$.

This set resembles a martingale C-monotone set closely. However, it is defined on the N-product space, instead of $\mathbb{R}\times\mathcal{P}_r(\mathbb{R})$, and the C-monotonicity can fail with a difference of ϵ. Let us first deal with the latter issue. We want to show that there are no problems with letting $\epsilon\to 0$. For this consider the product measures $\{P^{k,\otimes N}\}_{k\in\mathbb{N}}$ and $P^{\otimes N}$, which also have the property that $P^{k,\otimes N} \to P^{\otimes N}$ in the $\mathcal{W}_r^N$ topology by Theorem 1.5.6.

Each P_k is C_k-monotone and C_k converges uniformly to C, meaning that for every $\epsilon > 0$ we can find a k_0 such that for $k > k_0$,

$$\sup_{(x,p)\in\mathbb{R}\times\mathcal{P}_r(\mathbb{R})} |C_k(x,p) - C(x,p)| < \frac{\epsilon}{2N}.$$

Taking any $(x_i,p_i)_{i=1}^N$ with $(x_i,p_i) \in \tilde{\Gamma}^k$ and competitors $(x_i,q_i)_{i=1}^N$and using the previous assertion, we see that $(x_i,p_i)_{i=1}^N \in \tilde{\Gamma}^{\epsilon}_N$, namely

$$\begin{aligned}\sum_{i=1}^N C(x_i,p_i) &\leq \sum_{i=1}^N C_k(x_i,p_i) + \frac{\epsilon}{2}\\ &\leq \sum_{i=1}^N C_k(x_i,q_i) + \frac{\epsilon}{2}\\ &\leq \sum_{i=1}^N C(x,q_i)+\epsilon.\end{aligned}$$

Here, we used the uniform convergence of $C_k \to C$ in the first and last inequality, and the C_k-monotonicity in the second inequality. In particular this means that $P^{k,\otimes N}(\tilde{\Gamma}_N^\epsilon) = 1$ for $k \geq k_0$ for some $k_0 \in \mathbb{N}$. If we put this together with $\tilde{\Gamma}_N^\epsilon$ being closed, Wasserstein convergence implying weak convergence by Proposition 1.5.5 and the Portmanteau theorem (Theorem A.1.4), then we find

$$1 = \limsup_{k \in \mathbb{N}} P^{k,\otimes N}(\tilde{\Gamma}_N^\epsilon) \leq P^{\otimes N}(\tilde{\Gamma}_N^\epsilon).$$

Consequently, we can let $\epsilon \downarrow 0$ and the above argument then tells us that

$$P^{\otimes N}(\tilde{\Gamma}_N^0) = \lim_{\epsilon \downarrow 0} P^{\otimes N}(\tilde{\Gamma}_N^\epsilon) = 1.$$

Continuing the proof, we set $\tilde{\Gamma}_N := \tilde{\Gamma}_N^0$.

To finish up the first part, we need to transform $\tilde{\Gamma}_N$ from the N-product space to a set $\tilde{\Gamma}$ on $\mathbb{R} \times \mathcal{P}_r(\mathbb{R})$. To do this, note that products of open sets form a basis for the product topology on $(\mathbb{R} \times \mathcal{P}_r(\mathbb{R})^N$. Also, as the space $(\mathbb{R} \times \mathcal{P}_r(\mathbb{R}))^N$ is a complete separable space by Theorem 1.5.7, we can cover any open set with a countable collection of open sets from the basis. Specifically, we can cover the complement of $\tilde{\Gamma}_N$ with countably many sets,

$$\tilde{\Gamma}_N^c = \bigcup_{k \in \mathbb{N}} \bigotimes_{i=1}^N O_{k,i} \quad O_{k,i} \text{ open in } \mathbb{R} \times \mathcal{P}_r(\mathbb{R}). \tag{2.27}$$

Now, for each of the sets $\bigotimes_{i=1}^N O_{i,k}$ we can decompose

$$0 = P^{\otimes N}\left(\bigotimes_{i=1}^N O_{k,i}\right) = \prod_{i=1}^N P(O_{k,i}), \tag{2.28}$$

as $P^{\otimes N}$ is product measure. For each $k \in \mathbb{N}$, we pick only those sets $O_{k,i}$ that have zero measure with respect to P, of which there must be at least one for each k by (2.28), and take the union. We call this set A_N,

$$A_N := \bigcup_{k \in \mathbb{N}} \bigcup_{\substack{i \in \{1,\dots,N\} \\ P(O_{k,i})=0}} O_{k,i}.$$

This is a countable union and thus also a null set. We can do this for every $N \in \mathbb{N}$ and define the closed set

$$\tilde{\Gamma} := \left(\bigcup_{N \in \mathbb{N}} A_N\right)^c.$$

By construction, we have

$$P(\tilde{\Gamma}) = 1 - P\left(\bigcup_{N \in \mathbb{N}} A_N\right) \geq 1 - \sum_{N=1}^\infty P(A_N) = 1.$$

Next, we verify that $\tilde{\Gamma}$ is a C-monotone set. Indeed, if we take a finite sequence $(x_j, p_j)_{j=1}^N$ for some $N \in \mathbb{N}$, with $(x_j, p_j) \in \tilde{\Gamma}$, then $(x_j, p_j) \notin A_N$ for all $N \in \mathbb{N}$. Recall the definition of $\tilde{\Gamma}_N^c$ given in (2.27). The set A_N is constructed in such a way that for each $k \in \mathbb{N}$ at least one $O_{i,k} \subseteq A_N$. We are now in the situation, where no (x_j, p_j) of the finite sequence is in A_N, meaning that for each $k \in \mathbb{N}$ there is some $O_{i,k}$ such that $(x_i, p_i) \notin O_{i,k}$, but then it can never be that $(x_j, p_j)_{j=1}^N$ is a part of $\tilde{\Gamma}_N^c$. So, we find that $(x_j, p_j)_{j=1}^N \in \tilde{\Gamma}_N$ and $\tilde{\Gamma}$ is martingale C-monotone as defined in Definition 2.4.7.

For the second part of the theorem, we apply the map J from Definition 2.3.1 to π^k. By Lemma 2.3.2 we know that J preserves relative compactness, meaning that $\{J(\pi^k)\}_{k\in\mathbb{N}}$ has an accumulation point in $\mathcal{P}(\mathbb{R} \times \mathcal{P}(\mathbb{R}))$, called P. The measure P actually has to be in $\Lambda_M(\mu, \nu)$, as this latter set is compact by Lemma 2.3.4 and each $J(\pi^k) \in \Lambda_K(\mu, \nu)$. Finally, using the continuity of the map I we also see that $\mu(\mathrm{d}x)I(P_x)(\mathrm{d}y) \in \mathcal{M}(\mu, \nu)$ and $\mu(\mathrm{d}x)I(P_x)(\mathrm{d}y) = \pi(\mathrm{d}x, \mathrm{d}y)$

By assumption, each $J(\pi^k)$ is martingale C_k-monotone and by the first part of this theorem, we can now find some martingale C-monotone set $\tilde{\Gamma}$ for P. We will show that this is a martingale C-monotone set for π, by showing the equivalent definition given in Proposition 2.4.10. As the set $\tilde{\Gamma}$ is closed, it will be enough to show that (x, π_x) is some limit point for a sequence in $\tilde{\Gamma}$ for all $x \in \mathbb{R}$.

Take for any $x \in \mathbb{R}$ with the property that $P_x(\tilde{\Gamma}_x) = 1$, a sequence of random variables $\{X_i^x\}_{i\in\mathbb{N}}$ i.i.d. distributed according to P_x defined on some probability space $(\Omega, \mathbb{P})$. Each X_i^x is a random measure in $\mathcal{P}_r(\mathbb{R})$. We can test these measures against functions $g \in C(\mathbb{R})$, which are bounded in the sense that

$$|g(y)| \leq C(1 + |y|^r), \quad C \in \mathbb{R}.$$

By the fact that the r'th moment exists for the measure X_i^x, we find

$$\mathbb{E}[X_i^x(g)] = \int\limits_{\mathcal{P}_r(\mathbb{R})} \int\limits_{\mathbb{R}} g(y)\, p(\mathrm{d}y)\, P_x(\mathrm{d}p) = I(P_x)(g) = \pi_x(g).$$

By the large law of numbers we thus have $\mathbb{P}$-almost surely

$$\frac{1}{n}\sum_{i=1}^{n} X_i^x(g) \to I(P_x)(g) = \pi_x(g). \tag{2.29}$$

Using a countable set of test functions it is now possible to extend this limit result from $\mathbb{R}$ to $\mathcal{P}_r(\mathbb{R})$, where the limit is taken in the $\mathcal{W}_r$ sense. For this we need to check that the following two limits hold for $\mathbb{P}$-almost all ω

(i) $\frac{1}{n}\sum_{i=1}^{n} X_i^x \to \pi_x$ weakly for $n \to \infty$,

(ii) $\frac{1}{n}\sum_{i=1}^{n} \int\limits_{\mathbb{R}} |y|^r X_i^x(\mathrm{d}y) \to \int\limits_{\mathbb{R}} |y|^r \pi_x(\mathrm{d}y)$ for $n \to \infty$.

The second requirement is easily verified as obviously $|y|^r < |1 + |y|^r|$. For the first property we use a countable set of test functions that we will use to check the convergence of the characteristic functions,

$$F = \{f(y) = e^{iuy} | u \in \mathbb{Q}\}.$$

Because the set F is countable, the $\mathbb{P}$-a.s. convergence of (2.29) will hold for this set simultaneously. All the functions in F are also appropriately bounded,

$$|f(y)| = |e^{iuy}| = 1 \leq |1 + |y|^r|.$$

Integrating these functions gives

$$\frac{1}{n}\sum_{i=1}^{n} \int\limits_{\mathbb{R}} e^{iuy} X_i^x(\mathrm{d}y) \to \int\limits_{\mathbb{R}} e^{iuy} \pi_x(\mathrm{d}y), u \in \mathbb{Q}. \tag{2.30}$$

This means that the characteristic functions of $\frac{1}{n}\sum_{i=1}^n X_i^x$ converges to the characteristic function of π_x for each $u \in \mathbb{Q}$. We want to extend this from $\mathbb{Q}$ to $\mathbb{R}$. To this end, note that for each $u \in \mathbb{R}\backslash\mathbb{Q}$ there exists a sequence $q_k \in \mathbb{Q}, k \in \mathbb{N}$ such that $\lim\limits_{k\to\infty} q_k = u$. The resulting sequence of functions $f_k(y) = e^{iq_k y}$ then converge pointwise to $f(y) = e^{iuy}$, by continuity of the characteristic function, and are uniformly bounded by 1. The convergence is even uniform, as the characteristic function is uniformly continuous. Dominated convergence allows us to pull the limit out of the integral. We then need to swap the limits, which is possible because the sum in (2.30) converges absolutely and the characteristic function is uniformly continuous by Theorem B.1.1. This gives us

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n X_i^x(f) = \lim_{n\to\infty}\lim_{k\to\infty}\frac{1}{n}\sum_{i=n}^n X_i^x(f_k) = \lim_{k\to\infty}\lim_{n\to\infty}\frac{1}{n}\sum_{i=n}^n X_i^x(f_k) = \lim_{k\to\infty}\pi_x(f_k) = \pi_x(f).$$

Convergence of the characteristic function for all $u \in \mathbb{R}$ ensures weak convergence. We conclude

$$\frac{1}{n}\sum_{i=1}^n X_i^x \to \pi_x \text{ in } \mathcal{W}_r.$$

As this happens $\mathbb{P}$-almost surely, we can take any $\omega \in \Omega$ such that the above limit holds. Taking $p_i^x = X_i^x(\omega)$, then gives us the desired sequence of measures. It could still be that $\frac{1}{n}\sum_{i=1}^n p_i^x$ is not actually part of $\tilde{\Gamma}$. However, by convexity of $C(x,\cdot)$ we can enlarge the set $\tilde{\Gamma}$ to include all convex combinations of measures, with the enlarged set still being martingale C-monotone, by Lemma 2.4.9. Finally, we can invoke the closedness of $\tilde{\Gamma}$ to conclude that $(x,\pi_x) \in \tilde{\Gamma}$ for μ almost all x, which implies that $\tilde{\Gamma}$ is a martingale C-monotone set for π by Proposition 2.4.10. $\square$

2.6.3. Proof of $(c, \mathcal{F}_M)$-monotonicity and martingale C-monotonicity equivalence

Again, we will restate Lemma 2.4.12 to remind ourselves of all the assumptions.

Lemma 2.4.12 [$(c, \mathcal{F}_M)$-monotonicity and martingale C-monotonicity equivalence]. *Let $r \in [1,\infty)$, $\mu,\nu \in \mathcal{P}_r(\mathbb{R})$ such that $\mu \preceq \nu$, $\pi \in \mathcal{M}(\mu,\nu)$, $b \in L^1(\nu)$, $c : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ be jointly measurable, and $c(x,\cdot)$ be upper semicontinuous and $|c(x,y)| \le \alpha(x)|b(y)|$ for all $x \in \mathbb{R}$ for some $\alpha(x) > 0$. Then, π is $(c,\mathcal{F}_M)$-monotone with $\mathcal{F}_M$ as in Lemma 2.4.1 if and only if π is martingale C-monotone with $C(x,p) = \int_{\mathbb{R}} c(x,y)\,p(\mathrm{d}y)$.*

Proof. In the proofs of both implications we will start with one monotone set and construct the other one from it.

"$\Rightarrow$" We have a $(c,\mathcal{F}_M)$-monotone set $\hat{\Gamma} \subseteq \mathbb{R}\times\mathbb{R}$ from which we can construct the following set

$$\Gamma = \left\{ (x,p) \in \mathbb{R}\times\mathcal{P}_1(\mathbb{R}) \,\middle|\, p(\hat{\Gamma}_x) = 1, \int_{\mathbb{R}} y\,p(\mathrm{d}y) = x, \int_{\mathbb{R}} |b(y)|\,p(\mathrm{d}y) < \infty \right\},$$

where $\hat{\Gamma}_x = \{y \in \mathbb{R} \,|\, (x,y) \in \hat{\Gamma}\}$. Note that $(x,\pi_x) \in \Gamma$ for μ-almost all $x \in \mathbb{R}$, as $\pi \in \mathcal{M}(\mu,\nu)$ and $\int_{\mathbb{R}} |b(y)|\,\pi_x(\mathrm{d}y) < \infty$ for μ-almost all x, because

$$\int_{\mathbb{R}}\int_{\mathbb{R}} |b(y)|\,\pi_x(\mathrm{d}y)\,\mu(\mathrm{d}x) = \int_{\mathbb{R}}\int_{\mathbb{R}} |b(y)|\pi(\mathrm{d}x,\mathrm{d}y) = \int_{\mathbb{R}} |b(y)|\,\nu(\mathrm{d}y) < \infty.$$

We will now show that Γ is a martingale C-monotone set. To achieve this, take any finite sequence $(x_1, p_1), \dots, (x_N, p_N) \in \Gamma$ with competitors $(x_1, q_1), \dots, (x_N, q_N) \in \mathbb{R} \times \mathcal{P}_r(\mathbb{R})$. We will construct sequences $\{p_i^k\}_{k\in\mathbb{N}}$ for each (x_i, p_i) such that p_i^k is finitely supported on Γ_{x_i}, $x_i = \int y p_i^k(\mathrm{d}y)$ and

$$\lim_{k\to\infty} \int_{\mathbb{R}} c(x_i, y) p_i^k(\mathrm{d}y) = \int_{\mathbb{R}} c(x_i, y) p_i(\mathrm{d}y), \tag{2.31}$$

$$\lim_{k\to\infty} \int_{\mathbb{R}} |b(y)|\, p_i^k(\mathrm{d}y) = \int_{\mathbb{R}} |b(y)| p_i(\mathrm{d}y). \tag{2.32}$$

For each i, let $\{X_k^i\}_{k\in\mathbb{N}}$ be an iid sequence random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which are p_i distributed. By the Glivenko-Cantelli theorem (Theorem A.1.8, we know that the empirical distribution

$$F_n(t) := \frac{1}{k} \sum_{j=1}^{k} 1_{(-\infty, X_j^i]}(t)$$

converges uniformly to the distribution of p_i, $\mathbb{P}$-almost surely. In particular, this means that $F_n(t)$ converges pointwise for all continuity points of the distribution of p_i, which shows weak convergence of the empirical distribution. Consequently, we conclude that the empirical measure

$$p_i^j(\omega) := \frac{1}{k} \sum_{j=1}^{k} \delta_{X_j^i(\omega)}$$

converges weakly to p_i. Note that $p_i^k(\Gamma_{x_i}) = 1$. To achieve convergence in the $\mathcal{W}_1$-topology, we need to check that the first moment converges. This is a consequence of the Law of Large Numbers,

$$\int_{\mathbb{R}} |y| p_i^k(\mathrm{d}y) = \frac{1}{k} \sum_{j=1}^{k} |X_j^i| \to \int_{\mathbb{R}} |y| p_i(\mathrm{d}y) = \mathbb{E}[|X_1^j|] \quad \mathbb{P}-\text{ a.s.}$$

So, we found a sequence of finitely supported measures, which converges in the $\mathcal{W}_1$-metric to p_i. Another consequence the Law of Large Numbers, is that (2.31) and (2.32) also hold. The reason being that for all $f \in L^1(p_i)$ we have convergence of the empirical means of $\{f(X_i^k)\}_{k\in\mathbb{N}}$ by the Law of Large Numbers. Specifically, this means

$$\int_{\mathbb{R}} f(y) p_i^k(\mathrm{d}y) = \frac{1}{k} \sum_{j=1}^{k} f(X_j^j) \to \int_{\mathbb{R}} f(y) p_i(\mathrm{d}) = \mathbb{E}[f(X_1^j)] \quad \mathbb{P}-\text{ a.s.}$$

By assumption we have that $c(x_i, y), b(y) \in L^1(p_i)$. So, we also get almost sure convergence of the integrals of $c(x_i, y)$ and $b(y)$ with respect to p_i^k. Now, fix any $\omega \in \Omega$ for which the above limits hold and set $p_i^k = p_i^k(\omega)$.

The measures could still have the wrong means, i.e. $\int_{\mathbb{R}} y p_i^k(\mathrm{d}y) \neq x_i$. We can correct this by modifying the sequence a bit. Fix $y^+, y^- \in \Gamma_{x_i}$ such that $y^+ > x_i$ and $y^- < x_i$ and set $x_i^k = \int_{\mathbb{R}} y p_i^k(\mathrm{d}y)$. As p_i^k is a finitely supported probability measure we can write it as $p_i^k = \sum_{j=1}^{n_k} \alpha_j \delta_{y_j}$. It is now possible to add one Dirac measure to this sum to adjust the mean in such a way that it will be correct. Let $\epsilon_k^\pm \in \mathbb{R}$ and set

$$\tilde{p}_i^{k,\pm} = \sum_{j=1}^{n_k} (1 - \epsilon_k^\pm) \alpha_j \delta_{y_j} + \epsilon_k^\pm \delta_{y^\pm} \implies \tag{2.33}$$

$$\int_{\mathbb{R}} y \tilde{p}_i^{k,\pm}(\mathrm{d}y) = (1 - \epsilon_k^\pm) x_i^k + \epsilon_k^\pm y^\pm. \tag{2.34}$$

We want that $\epsilon_k^\pm$ is chosen in such a way that (2.34) is equal to x_i. This is possible by setting

$$\epsilon_k^\pm = \frac{x_i^k - x_i}{x_i^k - y^\pm}.$$

If $x_i < x_i^k$ we pick $\tilde{p}_i^{k,-}$ and if $x_i > x_i^k$ we pick $\tilde{p}_i^{k,+}$ to ensure that sum in (2.33) only has positive terms. We can also bound $|x_i^k - y^\pm|$ from below by some C. Finally, as $x_i^k \to x_i$ by the $\mathcal{W}_r$-convergence, we know that there exists some k_0 such that $|x_i^k - x_i| < C$, meaning that $\epsilon_k^\pm \in (-1,1)$ for $k \geq k_0$. So, resetting the sequence from k_0 and taking the correct $\tilde{p}_i^{k,\pm}$ will give us a convergent sequence that has the same mean as p_i. In the following we redefine this sequence as $\{p_i^k\}_{k\in\mathbb{N}}$ for convenience's sake.

The reason for introducing these finitely supported measures that approximate p_i is that we can only apply the $(c, \mathcal{F}_M)$-monotonicity property if the measures are finitely supported. Using Lemma 2.6.1 we can find sequences of competitors with limit $\{q_1, \dots, q_N\}$. We can now derive the martingale C-monotonicity using Proposition 2.4.10

$$\begin{aligned}
\sum_{i=1}^N C(x_i, p_i) &= \sum_{i=1}^N \int_{\mathbb{R}} c(x_i, y)\, p_i(\mathrm{d}y) \\
&\overset{(2.31)}{=} \lim_{k\to\infty} \sum_{i=1}^N \int_{\mathbb{R}} c(x_i, y)\, p_i^k(\mathrm{d}y) \\
&\leq \limsup_{k\to\infty} \sum_{i=1}^N \int_{\mathbb{R}} c(x_i, y)\, q_i^k(\mathrm{d}y) \\
&\leq \sum_{i=1}^N \int_{\mathbb{R}} c(x_i, y)\, q_i(\mathrm{d}y) = \sum_{i=1}^N C(x_i, q_i).
\end{aligned}$$

The $(c, \mathcal{F}_M)$-monotonicity as defined in Definition 2.4.2 is used in the first inequality. For the final inequality we used Theorem A.1.4 and the fact that $c(x, \cdot)$ is upper semicontinuous and bounded from above by assumption. This, together with the fact that $(x, \pi_x) \in \Gamma$, μ-almost surely, shows that π is martingale C-monotone with Γ as its corresponding set.

"$\Leftarrow$" This direction will require a bit more work, although most of the work is moved to the proof of Lemma A.1.15. We start with a martingale C-monotone Borel measurable set $\Gamma \subseteq \mathbb{R} \times \mathcal{P}_r(\mathbb{R})$, on which π is martingale C-monotone. A Borel measurable set is in particular analytic. So, we can apply Lemma A.1.15 to get an analytic set $\hat{\Gamma} \subseteq \mathbb{R} \times \mathbb{R}$, with $\pi(\hat{\Gamma}) = 1$ and such that

(i) For any $(x, p) \in \Gamma$ we have that p is concentrated on the fibre $\hat{\Gamma}_x = \{y \in \mathcal{Y} \mid (x, y) \in \hat{\Gamma}\}$, i.e., $p(\hat{\Gamma}_x) = 1$.

(ii) For any $(x, y) \in \hat{\Gamma}$ we find $(x, p) \in \Gamma$, such that for any $\epsilon > 0$, we can select a Borel measurable set $K \subseteq \hat{\Gamma}_x$ for which
   (a) $p(K) \geq 1 - \epsilon$,
   (b) c restricted to the fibre $\{x\} \times K$ is continuous,
   (c) $y \in \mathrm{supp}(p) \cap K$ and for all continuous $f : K \to \mathbb{R}$, we have

$$\int_{B_\delta(y)\cap K} \frac{f(z)}{p(B_\delta(y) \cap K)}\, p(\mathrm{d}z) \to f(y) \quad \text{for } \delta \downarrow 0.$$

It will be shown that $\hat{\Gamma}$ is $(c, \mathcal{F}_M)$-monotone. Take any finite sequence $(x_1, y_1), \dots, (x_N, y_N) \in \hat{\Gamma}$ and a measure α only supported on these points, with $\mathcal{F}_M$-competitor β as in Definition 2.4.2.

By Proposition 2.4.3, this means that α and β have the same marginals and

$$\int_{\mathbb{R}} y\alpha_{x_i}(\mathrm{d}y) = \int_{\mathbb{R}} y\beta_{x_i}(\mathrm{d}y), \quad \text{for } i = 1, \dots, N.$$

By property (ii) of Lemma A.1.15, we can find points $(x_i, p_i) \in \Gamma$ and sets K_i for each (x_i, y_i) such that c is continuous on $\{x_i\} \times K_i$, $y_i \in K_i$ and

$$\int_{B_\delta(y_i)\cap K_i} \frac{c(x,z)}{p(B_\delta(y_i)\cap K_i)} p(\mathrm{d}z) \to c(x,y), \quad \text{for } \delta \downarrow 0. \tag{2.35}$$

To show that α is $(c, \mathcal{F}_M)$-monotone we will find a sequence that converges to α, for which $(c, \mathcal{F}_M)$-monotonicity holds as a consequence of martingale C-monotonicity. To this end, let

$$\alpha^k(\mathrm{d}x, \mathrm{d}y) = \sum_{i=1}^N \delta_{x_i}(\mathrm{d}x) \frac{\alpha(\{(x_i,y_i)\})}{p_i(B_{\frac{1}{k}}(y_i)\cap K_i)} p_i|_{B_{\frac{1}{k}}(y_i)\cap K_i}(\mathrm{d}y), \quad k \in \mathbb{N}.$$

Combining this definition with (2.35) we get

$$\begin{aligned}
\lim_{k\to\infty} \int_{\mathbb{R}\times\mathbb{R}} c(x,y)\alpha^k(\mathrm{d}x,\mathrm{d}y) &= \lim_{k\to\infty} \sum_{i=1}^N \alpha(\{(x_i,y_i)\}) \int_{B_{\frac{1}{k}}(y_i)\cap K_i} \frac{c(x_i,y)}{p(B_{\frac{1}{k}}(y_i)\cap K_i)} p(\mathrm{d}y) \\
&= \sum_{i=1}^N c(x_i,y_i)\alpha(\{(x_i,y_i)\}) \\
&= \int_{\mathbb{R}\times\mathbb{R}} c(x,y)\alpha(\mathrm{d}x,\mathrm{d}y),
\end{aligned} \tag{2.36}$$

where the first equality follows from integration over Dirac measures and the second from (2.35). We can check that $\alpha^k \to \alpha$ weakly whenever $k \to \infty$ completely analogously to the chain of equalities in (2.36), by integrating any $f \in C_b(\mathbb{R}\times\mathbb{R})$ against the measure α^k. We see that

$$\lim_{k\to\infty} \int_{\mathbb{R}\times\mathbb{R}} f(x,y)\alpha^k(\mathrm{d}x,\mathrm{d}y) = \int_{\mathbb{R}\times\mathbb{R}} f(x,y)\alpha(\mathrm{d}x,\mathrm{d}y).$$

We use here that restrictions of continuous functions are again continuous. This allows us to get a similar result as (2.35) with c replaced by f. We can then apply the same set of equalities as in (2.36) to get the above limit. The key point of introducing these α^k is that these measures depend on $(x_i, p_i) \in \Gamma$, to which we can apply the martingale C-monotonicity property. For that we need to construct *martingale C-monotone* competitors for all $\alpha^k_{x_i}$. This is possible by virtue of Lemma 2.6.1 and results into sequences of competitors $\{\beta^k_{x_i}\}_{k\in\mathbb{N}}$ for $\{\alpha^k_{x_i}\}_{k\in\mathbb{N}}$. To provide a little bit more detail,

- the set $\{\alpha_{x_1}, \dots, \alpha_{x_N}\}$ will act as the set $\{p_1, \dots, p_N\}$;
- the set $\{\alpha^k_{x_1}, \dots, \alpha^k_{x_N}\}_{k\in\mathbb{N}}$ will function as the set $\{p^k_1, \dots, p^k_N\}_{k\in\mathbb{N}}$;
- and the set $\{\beta_{x_1}, \dots, \beta_{x_N}\}$ as the set $\{q_1, \dots, q_N\}$.

Lemma 2.6.1 then gives us sequences of competitors $\{q^k_1, \dots, q^k_N\}_{k\in\mathbb{N}}$ for $\{\alpha^k_{x_1}, \dots, \alpha^k_{x_N}\}_{k\in\mathbb{N}}$ with $q^k_i \to \beta_{x_i}$ weakly as $k \to \infty$. It is now possible to construct $\mathcal{F}_M$-competitors β^k for each α^k,

$$\beta^k(\mathrm{d}x,\mathrm{d}y) = \sum_{i=1}^N \delta_{x_i}(\mathrm{d}x)\beta(\{(x_i,y_i)\})q^k_i(\mathrm{d}y).$$

These measures are actual $\mathcal{F}_M$-competitors for α^k as

$$\begin{aligned}\alpha \circ \mathrm{proj}_1^{-1}(\mathrm{d}x) &= \sum_{i=1}^{N} \delta_{x_i}(\mathrm{d}x)\alpha(\{(x_i, y_i)\}) = \alpha \circ \mathrm{proj}_1^{-1}(\mathrm{d}x) \\ &= \beta \circ \mathrm{proj}_1^{-1}(\mathrm{d}x) = \sum_{i}^{N} \delta_{x_i}(\mathrm{d}x)\beta(\{(x_i, y_i)\}) = \beta^k \circ \mathrm{proj}_1^{-1}(\mathrm{d}x), \\ \int_{\mathbb{R}} y\alpha_{x_i}^k &= \int_{\mathbb{R}} yq_i^k(\mathrm{d}y) = \int_{\mathbb{R}} y\beta_{x_i}(\mathrm{d}y).\end{aligned}$$

Here the first sequence of equalities relies heavily on the fact that α and β were already $(c, \mathcal{F}_M)$-competitors. The second line of equalities follows directly from Lemma 2.6.1. The last thing we need to check is $\beta^k \to \beta$ weakly. For $f \in C_b(\mathbb{R} \times \mathbb{R})$ we get

$$\begin{aligned}\lim_{k\to\infty} \int_{\mathbb{R}\times\mathbb{R}} f(x, y)\beta^k(\mathrm{d}x, \mathrm{d}y) &= \lim_{k\to\infty} \sum_{i=1}^{N} \beta(\{(x_i, y_i)\}) \int_{\mathbb{R}} f(x_i, y)q_i^k(\mathrm{d}y) \\ &= \sum_{i=1}^{N} \beta(\{(x_i, y_i)\}) \int_{\mathbb{R}} f(x_i, y)\beta_{x_i}(\mathrm{d}y) \\ &= \sum_{i=1}^{N} \int_{\mathbb{R}\times\mathbb{R}} f(x, y)\beta(\{(x_i, y_i)\})\beta_x(\mathrm{d}y)\delta_{x_i}(\mathrm{d}x) \\ &= \int_{\mathbb{R}\times\mathbb{R}} f(x, y)\beta_x(\mathrm{d}y)\mathrm{proj}_1(\beta)(\mathrm{d}x) = \int_{\mathbb{R}\times\mathbb{R}} f(x, y)\beta(\mathrm{d}x, \mathrm{d}y).\end{aligned}$$

Finally, we can combine all these results into

$$\begin{aligned}\int_{\mathbb{R}\times\mathbb{R}} c(x, y)\alpha(\mathrm{d}x, \mathrm{d}y) &\overset{(2.36)}{=} \lim_{k\to\infty} \int_{\mathbb{R}\times\mathbb{R}} c(x, y)\alpha^k(\mathrm{d}x, \mathrm{d}y) \\ &= \lim_{k\to\infty} \sum_{i=1}^{N} C(x_i, \alpha_{x_i}^k) \\ &\leq \limsup_{k\to\infty} \sum_{i=1}^{N} C(x_i, \beta_{x_i}^k) \\ &= \limsup_{k\to\infty} \int_{\mathbb{R}\times\mathbb{R}} c(x, y)\beta^k(\mathrm{d}x, \mathrm{d}y) \leq \int_{\mathbb{R}\times\mathbb{R}} c(x, y)\beta(\mathrm{d}x, \mathrm{d}y).\end{aligned}$$

The second equality is because we have chosen $C(x, p) = \int_{\mathbb{R}} c(x, y)p(\mathrm{d}y)$, the first inequality is due to martingale C-monotonicity. Recall that each (x_i, p_i) is in the martingale C-monotone set Γ and

$$\alpha_{x_i}^k = \frac{1}{p(B_{\frac{1}{k}}(y_i) \cap K_i)} p|_{B_{\frac{1}{k}}(y_i)\cap K_i}.$$

By construction the $\beta_{x_i}^k$ are competitors for $\alpha_{x_i}^k$. So their C values must be less than those of α_{x_i}. The final inequality follows by c being upper semicontinuous and Theorem A.1.4, as was the case as well in the proof of the reverse implication. $\square$

3. A general monotonicity principle

For the original OT problem we could define the notion of c-cyclical monotonicity, which in the case of OT was both a necessary and sufficient condition for optimality as seen in Theorem 1.3.2. As c-cyclical monotonicity carries no information about the conditional distribution of the couplings considered, we have no hope that this condition would be sufficient or even necessary for the MOT problem. In the search of stability of the MOT problem there were some independent monotonicity results in specific cases [12, 34], with the most general result given in [14]. The idea is to generalize the OT problem to a general class of problems that minimise some integral over a set of measures, where the measures have to satisfy some constraint. This constraint will be given in terms of a set of functions. We have already seen two specific examples of such problems by the characterisation of the MOT and MOWT problems given in Lemma 2.4.1 and Lemma 2.4.6 in terms of the sets $\mathcal{F}_M$ and $\widetilde{\mathcal{F}}_M$.

Definition 3.0.1 [Generalized moment problem]**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces, with Borel measurable cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. Assume that we have a set $\mathcal{F}$ of Borel-measurable functions on $\mathcal{X} \times \mathcal{Y}$ and write

$$\Pi_{\mathcal{F}} = \left\{ \pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \,\middle|\, \int f(x,y)\,\pi(\mathrm{d}x, \mathrm{d}y) = 0, \forall f \in \mathcal{F} \right\}.$$

The generalized moment problem associated with $\mathcal{F}$ is then given by either finding the value or minimizer, if it exists, of

$$\inf_{\pi \in \Pi_{\mathcal{F}}} \int_{\mathcal{X} \times \mathcal{Y}} c(x,y)\pi(\mathrm{d}x, \mathrm{d}y). \tag{GMP}$$

We can now generalize Definition 2.4.2 to encompass a more general set of functions $\mathcal{F}$, which will logically be called $(c, \mathcal{F})$-monotonicity.

Definition 3.0.2 [Competitors and $(c, \mathcal{F})$-monotonicity]**.** Let E be a Polish space and $\mathcal{F}$ a set of Borel measurable functions from E to $\mathbb{R}$.

(1) A measure $\alpha' \in \mathcal{P}(E)$ is called an *$\mathcal{F}$-competitor* of $\alpha \in \mathcal{P}(E)$ if and only if $\alpha(f) = \alpha'(f)$ for all $f \in \mathcal{F}$.

(2) A set $\Gamma \subseteq E$ is called *$(c, \mathcal{F})$-monotone* if and only if for any probability measure α, finitely support on Γ, and any $\mathcal{F}$-competitor α' we have $\alpha(c) \leq \alpha'(c)$.

(3) A measure $P \in \mathcal{P}(E)$, which is supported on a $(c, \mathcal{F})$-monotone set, is then also called $(c, \mathcal{F})$-monotone.

3.1. Necessity of $(c, \mathcal{F})$-monotonicity

Definition 3.0.2 also generalises the idea of c-cyclical monotonicity by looking at any finite random configuration of points instead of only allowing cyclic permutations. We now state the monotonicity principle.

Theorem 3.1.1 [Monotonicity principle]**.** *Let E be a Polish space, with Borel measurable cost function $c : E \to (-\infty, +\infty]$. Let $\mathcal{F}$ be a family of Borel-measurable functions on E satisfying*

(1) There exists a function $g : E \to [0, \infty)$ such that each element of $\mathcal{F}$ is bounded by some multiple of g. I.e., for each $f \in \mathcal{F}$ there is a constant $a_f \in \mathbb{R}^+$ such that $|f| \leq a_f g$.

(2) All functions in $\mathcal{F}$ are continuous, or $\mathcal{F}$ is at most countable.

Further assume that there exists a minimizer π^* to the GMP problem associated with $\mathcal{F}$, then π^* is $(c, \mathcal{F})$-monotone.

Remark 3.1.2. Note that the linear restraints given in Lemma 2.4.1 and Lemma 2.4.6 are of the setting in Theorem 3.1.1. The sets $\mathbb{R}^d \times \mathbb{R}^d$ and $\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ are both metric spaces, which are in particular Polish spaces. For the first assumption of the constraint sets we note that all function sets considered contain only bounded functions. For the set $\mathcal{F}_M$, it is possible to consider $g(x, y) = (x - y)$ and bound every function by $a_f g = \|f\|_\infty g$. For the other set $\tilde{\mathcal{F}}_M$, we can use the function $g(x, p) = \int (x - y) p(\mathrm{d}y)$. $\diamond$

For the proof of Theorem 3.1.1 we will need the following result, the proof of which can be found in [13], Proposition 2.1.

Lemma 3.1.3. *Let (E, m) be a Polish probability space, and M an analytic subset of $E^l, l \in \mathbb{N}$, then one of the following is true:*

(i) there exists m-null sets $M_1, \ldots, M_l \subseteq E$ such that $M \subseteq \cup_{i=1}^l \mathrm{proj}_i^{-1}(M_i)$, or

(ii) there is a measure η on E^l such that $\eta(M) > 0$ and $\eta \circ \mathrm{proj}_i^{-1} \leq m$ for $i = 1, \ldots, l$.

Proof of Theorem 3.1.1. We need to construct a monotone set Γ for the optimal transport plan π^*. To do this we will construct for each $l \in \mathbb{N}$ a set Γ_l with $\pi^*(\Gamma_l) = 1$ such that for any finite measure α concentrated on Γ_l with $|\mathrm{supp}(\alpha)| \leq l$ and

(a) $\alpha(E) \leq 1$,

(b) $\int_E g(x) \alpha(\mathrm{d}x) \leq l$,

(c) $|c|_{\mathrm{supp}(\alpha)}| \leq l$,

there is no $\mathcal{F}$-competitor α' with $\alpha(c) > \alpha'(c)$ on at most l points and satisfying the above requirements as well. If we can find such sets we can set $\Gamma = \cap_{l \in \mathbb{N}} \Gamma_l$, which will turn out to be the required set.

To achieve this goal, fix $l \in \mathbb{N}$ and define the following subset of the product space E^l,

$$\begin{aligned} M = \{(z_1, \ldots, z_l) \in E^l \mid \exists \text{ a measure } \alpha \text{ on } E, \text{ satisfying (a), (b), (c)}, \mathrm{supp}(\alpha) \subseteq \{z_1, \ldots, z_l\} \\ \text{such that there is an } \mathcal{F}\text{-competitor } \alpha' \text{ with } \alpha(c) > \alpha'(c), \text{ satisfying (a), (b), (c)}\}. \end{aligned}$$

We can realise M as the projection of a Borel set, which ensures that M is analytic. We need M to be analytic for Lemma 3.1.3. The set that we need to project is

$$\begin{aligned} \hat{M} = \Big\{ &(z_1, \ldots, z_l, \alpha_1, \ldots, \alpha_l, z_1', \ldots, z_l', \alpha_1' \ldots, \alpha_l') \in E^l \times \mathbb{R}_+^l \times E^l \times \mathbb{R}_+^l : \\ &\alpha = \sum_{i=1}^l \alpha_i \delta_{z_i} \text{ and } \alpha' = \sum_{i=1}^l \alpha_i' \delta_{z_i'} \text{ satisfy (a), (b), (c)}, |c|_{\mathrm{supp}(\alpha) \cup \mathrm{supp}(\alpha')}| \leq l \\ &\text{and } \alpha(c) > \alpha'(c) \Big\}. \end{aligned}$$

If we project $\hat{M}$ onto the first l coordinates it becomes M. The set $\hat{M}$ is Borel by the assumption on the set $\mathcal{F}$, meaning that M is analytic.

We can now apply Lemma 3.1.3 to the space (E, π^*) and the set M. There are now two cases. If (i) holds we can directly construct the set Γ_l by defining $N_l := \cup_{i=1}^l M_i$, and taking $\Gamma_l := E \setminus N_l$.

This set has full measure, as all the M_i are null sets for π^*. By definition, if α is supported on Γ_l with $|\text{supp}(\alpha)| \leq l$ and satisfying (a), (b) and (c), then we cannot find a $\mathcal{F}$-competitor α' that has $\alpha(c) > \alpha'(c)$. Such pairs of measures can only be defined on the compliment of Γ_l by construction.

On the other hand, we can show that if (ii) holds, a contradiction appears. We can assume that the measure η is concentrated on M and satisfies $\eta \circ \text{proj}_i^{-1} \leq \frac{1}{l}\pi^*$ by restricting and rescaling. An application of the Jankow-Von Neumann selection theorem (Theorem A.1.16) to the map $\text{proj}_M : \hat{M} \to M$ gives us a mapping

$$\Phi : M \to \mathbb{R}_+^l \times E^l : (z \mapsto (\alpha_1(z), \dots, \alpha_l(z), z_1'(z), \dots, z_l'(z), \alpha_1(z) \; \dots, \alpha_l'(z))$$

such that

$$(z, \Phi(z)) \in \hat{M}.$$

Additionally, the map Φ is measurable with respect to the σ-algebra generated by the analytic subsets of E^l. Note that every Borel set is analytic. So this σ-algebra encompasses the Borel σ-algebra. We construct the following measures

$$\alpha_z := \sum_{i=1}^{l} \alpha_i(z)\delta_{z_i},$$
$$\alpha_z' := \sum_{i=1}^{l} \alpha_i'(z)\delta_{z_i'(z)}.$$

From these measures we get kernels $z \mapsto \alpha_z$, $z \mapsto \alpha_z'$ from E^l with the σ-algebra generated by the analytic subsets to E with the Borel σ-algebra. We can define new measures from these kernels, ω, ω' on the Borel sets of E :

$$\omega(B) = \int_{E^l} \alpha_z(B)\eta(\mathrm{d}z),$$
$$\omega'(B) = \int_{E^l} \alpha_z'(B)\eta(\mathrm{d}z).$$

By construction we have for that for all $B \in \mathcal{B}$

$$\begin{aligned} \omega(B) = \int_{E^l} \alpha_z(B)\eta(\mathrm{d}z) &= \sum_{i=1}^{l} \int_{E^l} \alpha_i(z)\delta_{z_i}(B)\eta(\mathrm{d}z) \\ &\leq \sum_{i=1}^{l} \int_{E^l} \delta_{z_i}(B)\eta(\mathrm{d}z) = \sum_{i=1}^{l} \eta \circ \text{proj}_i^{-1}(B) \quad . \\ &\leq l\frac{1}{l}\pi^*(B) = \pi^*(B). \end{aligned} \tag{3.1}$$

In particular, this means that $\omega \leq \pi^*$. As the penultimate step we show that ω' is an $(c, \mathcal{F})$-competitor of ω with $\omega(c) > \omega'(c)$. For each $f \in \mathcal{F}$ we get

$$\int_E f(x)\omega'(\mathrm{d}x) = \int_E \int_{E^l} f(x)\alpha_z'(\mathrm{d}x)\eta(\mathrm{d}z) = \int_E \int_{E^l} f(x)\alpha_z(\mathrm{d}x)\eta(\mathrm{d}z) = \int_E f(x)\omega(\mathrm{d}x).$$

Where the first and last equality are justified because $\int_E g(x)\alpha_z(\mathrm{d}x) \leq l$, $\int_E g(x)\alpha'_z(\mathrm{d}x) \leq l$ for all z and f is bounded by g. Hence, the function f is integrable with respect to α_z and α'_z. The second equality is justified by the properties of the set $\hat{M}$, out of which these kernels were constructed.

Finally, we can state the contradiction. As we have that $|c| \leq l$, $(\omega + \omega')$-a.s., we find that

$$\int_E c(x)\omega'(\mathrm{d}x) = \int_E \int_{E^l} c(x)\alpha'_z(\mathrm{d}x)\eta(\mathrm{d}z) < \int_E \int_{E^l} c(x)\alpha_z(\mathrm{d}x)\eta(\mathrm{d}z) = \int_E c(x)\omega(\mathrm{d}x).$$

We can now construct a new measure $\pi' := \pi^* - (\omega - \omega')$, which is a probability measure because of (3.1), with

$$\int_E c(x)\pi'(\mathrm{d}x) < \int_E c(x)\pi^*(\mathrm{d}x)$$

and $\pi' \in \Pi_{\mathcal{F}}$ by construction. This is a contradiction with the assumption that π^* was optimal. $\square$

Now that we know that optimality implies $(c, \mathcal{F})$-monotone, it is immediately also true that an optimal martingale coupling should be $(c, \mathcal{F}_M)$-monotone, because of the characterisation given in Lemma 2.4.1. This shows that Theorem 2.4.4 is true. It remains to show that an optimal coupling for the MOWT problem is martingale C-monotone, which we will prove now. To remind ourselves of all the assumptions, we restate the theorem.

Theorem 2.4.8. *Let $C : \mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d) \to [0, \infty)$ be measurable and $P^* \in \Lambda_M(\mu, \nu)$ optimal for (MWOT') with finite value. Then P^* is martingale C-monotone. Moreover, if C additionally satisfies for all $x \in \mathbb{R}^d$ and $Q \in \mathcal{P}_r(\mathcal{P}_r(\mathbb{R}^d))$*

$$C(x, I(Q)) \leq \int_{\mathcal{P}(\mathbb{R}^d)} C(x, p)Q(\mathrm{d}p), \tag{2.10}$$

then any optimizer π^* of (MWOT) with finite values is martingale C-monotone. Note that (2.10) *is satisfied if C is convex in second argument.*

Proof. The first part of the theorem follows from Theorem 3.1.1. Note that if we set $c = C$ and $\mathcal{F} = \tilde{\mathcal{F}}_M$, then we are exactly in the right situation to apply Theorem 3.1.1. To see that we can apply Theorem 3.1.1, we need to check that $(C, \tilde{\mathcal{F}}_M)$-monotonicity implies martingale C-monotonicity. Recall that $(C, \tilde{\mathcal{F}}_M)$-monotonicity means that P^* is concentrated on a $(C, \tilde{\mathcal{F}}_M$-monotone set Γ. By definition, we have for any finitely supported probability measure α concentrated on Γ and an $\tilde{\mathcal{F}}_M$-competitor α' that $\alpha(C) \leq \alpha'(C)$.

We can construct probability measures that precisely encode the martingale C-monotonicity requirement. Indeed, let $(x_1, p_1), \ldots, (x_N, p_N) \in \Gamma$ and also consider $(x_1, q_1), \ldots, (x_N, q_N) \in \mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d)$ with

$$\sum_{i=1}^N p_i = \sum_{i=1}^N q_i, \tag{3.2}$$

$$x_i = \int_{\mathbb{R}^d} y p_i(\mathrm{d}y) = \int_{\mathbb{R}^d} y q_i(\mathrm{d}y), \quad i = 1, \ldots, N. \tag{3.3}$$

We will show that the following probability measures are $\tilde{\mathcal{F}}_M$-competitors,

$$\alpha = \frac{1}{N}\sum_{i=1}^{N}\delta_{(x_i,p_i)}$$
$$\alpha' = \frac{1}{N}\sum_{i=1}^{N}\delta_{(x_i,q_i)}.$$

The set $\tilde{\mathcal{F}}_M$ consists of the three sets $\tilde{\mathcal{F}}_m$, $\tilde{\mathcal{F}}_\mu$ and $\tilde{\mathcal{F}}_\nu$. First we check the integral equality for $f \in \tilde{\mathcal{F}}_m$,

$$\alpha(f) = \frac{1}{N}\sum_{i=1}^{N} h(x_i)g(p_i)\int_{\mathbb{R}^d}(x_i - y)p_i(\mathrm{d}y) \overset{(3.3)}{=} 0$$
$$\alpha'(f) = \frac{1}{N}\sum_{i=1}^{N} h(x_i)g(q_i)\int_{\mathbb{R}^d}(x_i - y)q_i(\mathrm{d}y) \overset{(3.3)}{=} 0.$$

For $f \in \tilde{\mathcal{F}}_\mu$ we get

$$\alpha(f) = \frac{1}{N}\sum_{i=1}^{N} f(x_i) - \int_{\mathbb{R}^d} f(x)\mu(\mathrm{d}x) = \alpha'(f).$$

Finally, $f \in \tilde{\mathcal{F}}_\nu$ gives us

$$\begin{aligned}\alpha(f) &= \frac{1}{N}\sum_{i=1}^{N}\int_{\mathbb{R}^d} g(y)p_i(\mathrm{d}y) \\ &\overset{(3.2)}{=} \frac{1}{N}\sum_{i=1}^{N}\int_{\mathbb{R}^d} g(y)q_i(\mathrm{d}y) = \alpha'(f).\end{aligned}$$

It is now clear that α and α' are $\tilde{F}_M$-competitors. Using that $\alpha(C) \leq \alpha'(C)$ we find martingale C-monotonicity

$$\sum_{i=1}^{N} C(x_i, p_i) = N\alpha(C) \leq N\alpha'(C) = \sum_{i=1}^{N} C(x_i, p_i).$$

Theorem 3.1.1 guarantees that P^* being optimal implies that P^* is $(C, \tilde{\mathcal{F}}_M)$-monotone and because of the above argument that implies martingale C-monotonicity.

For the second part of the theorem we first note that for any $\pi \in \mathcal{M}(\mu,\nu)$ we have that $J(\pi) \in \Lambda_M(\mu,\nu)$. For our purpose it is beneficial to actually write out what $J(\pi^*)$ is, giving

$$\begin{aligned}\int_{\mathbb{R}^d}\int_{\mathcal{P}(\mathbb{R}^d)} C(x,p)J(\pi^*)(\mathrm{d}p,\mathrm{d}x) &= \int_{\mathbb{R}^d}\int_{\mathcal{P}(\mathbb{R}^d)} C(x,p)\delta_{\pi_x^*}(\mathrm{d}p)\mu(\mathrm{d}x) \\ &= \int_{\mathbb{R}^d} C(x,\pi_x^*)\,\mu(\mathrm{d}x)\end{aligned}. \tag{3.4}$$

Furthermore, we have that for any $P \in \Lambda_M(\mu,\nu)$ that $I(P_x)(\mathrm{d}y)\mu(\mathrm{d}x) \in \mathcal{M}(\mu,\nu)$. We can now apply assumption (2.10) to get

$$\begin{aligned}
\int_{\mathbb{R}^d}\int_{\mathcal{P}(\mathbb{R}^d)} C(x,p)J(\pi^*)(\mathrm{d}p,\mathrm{d}x) &\overset{(3.4)}{=} \int_{\mathbb{R}^d} C(x,\pi^*_x)\mu(\mathrm{d}x) \\
&\leq \int_{\mathbb{R}^d} C(x,I(P_x))\mu(\mathrm{d}x) \\
&\leq \int_{\mathbb{R}^d}\int_{\mathcal{P}(\mathbb{R}^d)} C(x,p)P_x(\mathrm{d}p)\mu(\mathrm{d}x) \\
&= \int_{\mathbb{R}^d}\int_{\mathcal{P}(\mathbb{R}^d)} C(x,p)P(\mathrm{d}x,\mathrm{d}p).
\end{aligned}$$

Here, the first inequality follows from the assumption that π^* is optimal and the second inequality follows from assumption (2.10). □

3.2. Sufficiency in $\mathbb{R}\times\mathbb{R}$

In the previous section it was shown that the monotonicity principle formulated in Theorem 3.1.1 is necessary for GMP problems. What is maybe surprising is that it is a sufficient condition to check optimality for a coupling for the MOT problem in the one dimensional case. A version of this result with a specific cost function was originally proven by Beiglböck and Juillet in [12]. This result was later extended and the proof was simplified by Beiglböck and Griessler in [27]. This result is one of the main ingredients in our proof of stability of the MOT problem. For this proof we will need some intermediate results.

Proposition 3.2.1. *Let $\chi : I \to \mathbb{R}$ be convex, $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ with $\mu \preceq \nu$, as defined in Definition 2.1.3. Then, for any $\pi \in \mathcal{M}(\mu,\nu)$, the value*

$$F(\chi) = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \chi(y)\pi_x(\mathrm{d}y) - \chi(x)\right]\mu(\mathrm{d}x)$$

is well defined in $[0,\infty]$ and does not depend on the choice of π.

Proof. We can intuitively argue why this definition is well defined by noticing that

$$F(\chi) = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \chi(y)\pi_x(\mathrm{d}y) - \chi(x)\right]\mu(\mathrm{d}x) = \int_{\mathbb{R}} \chi(y)\nu(\mathrm{d}y) - \int_{\mathbb{R}} \chi(x)\mu(\mathrm{d}x).$$

The only issues that we have to resolve are integrability issues. This will be done by approximating χ with functions that have well behaved integrals.

Define χ_n to be the smallest convex function that is equal to χ_n on $[-n,n]$ and grows linearly on $|x|>n$. We can find this function explicitly by finding two affine functions $\ell_n^{\pm}(x)$ such that

$$\ell_n^-(-n) = \chi(-n), \qquad \ell_n^+(n) = \chi(n), \tag{3.5}$$

$$\chi(x) \geq \ell_n^-(x), \qquad \chi(x) \geq \ell_n^+(x) \quad \forall x \in \mathbb{R}. \tag{3.6}$$

Let us look at the construction of ℓ_n^+, as the construction of ℓ_n^- goes almost analogously. Note that the left derivative of χ exists, because of the convexity of χ, and set

$$L := \lim_{h \downarrow 0} \frac{\chi(n) - \chi(n-h)}{h}.$$

Define $\ell_n^+(x) = L(x-n) + \chi(n)$. It is clear that $\ell_n^+(n) = \chi(n)$ and taking the left derivative ensures that this is the smallest affine function satisfying 3.6. To construct $\ell_n^-(x)$ we repeat the same procedure, but take the right derivative instead of the left derivative. We can now define χ_n as

$$\chi_n(x) = \chi(x)1_{[-n,n]}(x) + \ell_n^-(x)1_{(-\infty,n)}(x) + \ell_n^+(x)1_{(n,\infty)}.$$

Notice that, as the left derivative and right derivative are non-decreasing and non-increasing respectively, that the slopes of $\ell_n^-(x)$ and $\ell_n^+(x)$ will also be non-decreasing and non-increasing in n. To show that $F(\chi_n) \to F(\chi)$ we define

$$\begin{aligned} f(x) &= \int_{\mathbb{R}} \chi(y)\pi_x(\mathrm{d}y) - \chi(x), \\ f_n(x) &= \int_{\mathbb{R}} \chi_n(y)\pi_x(\mathrm{d}y) - \chi_n(x). \end{aligned}$$

By Jensen's inequality we see that

$$f(x) = \int_{\mathbb{R}} \chi(y)\pi_x(\mathrm{d}y) - \chi(x) \geq \chi\left(\int_{\mathbb{R}} y\pi_x(\mathrm{d}y)\right) - \chi(x) = \chi(x) - \chi(x) = 0,$$

and completely analogously we find that $f_n(x) \geq 0$. For $n \leq m$ we have that χ_m agrees with χ_n on $[-n,n]$, and outside of $[-n,n]$ χ_m will grow faster than χ_n by definition. This means that $\chi_m - \chi_n$ is again convex. By another application of the Jensen's inequality to the function $\chi_m - \chi_n$ for $n \leq m$ we also get

$$\begin{aligned} f_m(x) - f_n(x) &= \int_{\mathbb{R}} \chi_m(y) - \chi_n(y)\pi_x(\mathrm{d}y) - \chi_m(x) + \chi_n(x) \\ &\geq \chi_m\left(\int_{\mathbb{R}} y\pi_x(\mathrm{d}y)\right) - \chi_n\left(\int_{\mathbb{R}} y\pi_x(\mathrm{d}y)\right) - \chi_m(x) + \chi_n(x) = 0. \end{aligned}$$

This shows that the functions f_n are monotonically increasing. So we can apply the monotone convergence theorem to the monotonically increasing sequence $\{f_n\}_n$ to get

$$\begin{aligned} F(\chi) &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \chi(y)\pi_x(\mathrm{d}y) - \chi(x)\right]\mu(\mathrm{d}x) \\ &= \int_{\mathbb{R}} \lim_{n\to\infty} \left[\int_{\mathbb{R}} \chi_n(y)\pi_x(\mathrm{d}y) - \chi_n(x)\right]\mu(\mathrm{d}x) \\ &= \lim_{n\to\infty} \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \chi_n(y)\pi_x(\mathrm{d}y) - \chi_n(x)\right]\mu(\mathrm{d}x) \\ &= \lim_{n\to\infty} \int_{\mathbb{R}} \chi_n(y)\nu(\mathrm{d}y) - \int_{\mathbb{R}} \chi_n(x)\mu(\mathrm{d}x). \end{aligned}$$

All the integrals on the right hand side exist, because the functions χ_n grow linearly. This shows that the left hand side is well defined. $\square$

Definition 3.2.2. Let $\varphi, \psi : I \to \mathbb{R}$ be measurable functions and $\chi : I \to \mathbb{R}$ a convex function such that both $\int_{\mathbb{R}}(\varphi(x) - \chi(x))\mu(\mathrm{d}x) < \infty$ and $\int_{\mathbb{R}} \psi(y) + \chi(y)\nu(\mathrm{d}y) < \infty$. We define the function G as

$$G(\varphi, \psi) = \int\limits_{\mathbb{R}} \varphi(x) - \chi(x)\mu(\mathrm{d}x) + \int\limits_{\mathbb{R}} \psi(y) + \chi(y)\nu(\mathrm{d}y) - F(\chi).$$

Proposition 3.2.3. *Definition 3.2.2 is well defined and does not depend on the choice of χ.*

Proof. Let $\chi, \chi' : I \to \mathbb{R}$ be both convex. If both $G_\chi(\varphi, \psi)$ and $G_{\chi'}(\varphi, \psi)$ are infinite, then the statement would obviously true. So, we can assume that $G_\chi(\varphi, \psi)$ is finite. We can now compute the difference to see that

$$G_\chi(\varphi, \psi) - G_{\chi'}(\varphi, \psi) = \int\limits_{\mathbb{R}} \chi'(x) - \chi(x)\mu(\mathrm{d}x) + \int\limits_{\mathbb{R}} \chi(y) - \chi'(y)\nu(\mathrm{d}x) + (F(\chi') - F(\chi)). \quad (3.7)$$

From the proof of Proposition 3.2.1 we saw that

$$\begin{aligned} F(\chi) &= \int\limits_{\mathbb{R}} \chi(y)\nu(\mathrm{d}y) - \int\limits_{\mathbb{R}} \chi(x)\mu(\mathrm{d}x), \\ F(\chi') &= \int\limits_{\mathbb{R}} \chi'(y)\nu(\mathrm{d}y) - \int\limits_{\mathbb{R}} \chi'(x)\mu(\mathrm{d}x). \end{aligned}$$

These expressions precisely cancel out all the other terms in (3.7), which is what we wanted. $\square$

Lemma 3.2.4. *Let $\varphi, \psi, \Delta : I \to \mathbb{R}$ be measurable functions and $\pi \in \mathcal{M}(\mu, \nu)$. If φ and ψ are as in Definition 3.2.2 and the function*

$$\xi(x, y) = \varphi(x) + \psi(y) + \Delta(x)(y - x)$$

has a well defined integral with respect to π in $[-\infty, \infty)$, then

$$G(\varphi, \psi) = \int\limits_{\mathbb{R}\times\mathbb{R}} \varphi(x) + \psi(y)\pi_x(\mathrm{d}y)\mu(\mathrm{d}x) = \int\limits_{\mathbb{R}\times\mathbb{R}} \xi(x, y)\pi(\mathrm{d}x, \mathrm{d}y).$$

Proof. The proof follows from a sequence of definitions and equalities. We start out by writing out the definition of G

$$\begin{aligned} G(\varphi, \psi) &= \int\limits_{\mathbb{R}} \varphi(x) - \chi(x)\mu(\mathrm{d}x) + \int\limits_{\mathbb{R}} \psi(y) + \chi(y)\nu(\mathrm{d}y) - F(\chi) \\ &= \int\limits_{\mathbb{R}} \varphi(x) - \chi(x)\mu(\mathrm{d}x) + \int\limits_{\mathbb{R}} \int\limits_{\mathbb{R}} \psi(y) + \chi(y)\pi_x(\mathrm{d}y)\mu(\mathrm{d}x) - F(\chi) \\ &= \int\limits_{\mathbb{R}} \left(\varphi(x) - \chi(x) + \int\limits_{\mathbb{R}} \psi(y) + \chi(y)\pi_x(\mathrm{d}y) - \int\limits_{\mathbb{R}} \chi(y')\pi_x(\mathrm{d}y') + \chi(x) \right) \mu(\mathrm{d}x) \\ &= \int\limits_{\mathbb{R}} \left(\varphi(x) + \int\limits_{\mathbb{R}} \psi(y)\pi_x(\mathrm{d}y) \right) \mu(\mathrm{d}x) \\ &= \int\limits_{\mathbb{R}} \int\limits_{\mathbb{R}} \varphi(x) + \psi(y)\pi_x(\mathrm{d}y)\mu(\mathrm{d}x). \end{aligned}$$

To finish the argument we remark that

$$\int_{\mathbb{R}\times\mathbb{R}} \Delta(x)(y-x)\pi(\mathrm{d}x,\mathrm{d}y) = \int_{\mathbb{R}} \Delta(x)\left(\int_{\mathbb{R}} y\pi_x(\mathrm{d}y) - x\right)\mu(\mathrm{d}x) = 0.$$

Combining this with what we found earlier gives us

$$G(\varphi,\psi) = \int_{\mathbb{R}}\int_{\mathbb{R}} \varphi(x)+\psi(y)+\Delta(x)(y-x)\pi_x(\mathrm{d}y)\mu(\mathrm{d}x) = \int_{\mathbb{R}}\int_{\mathbb{R}} \xi(x,y)\pi(\mathrm{d}x,\mathrm{d}y).$$

□

Lemma 3.2.5. *Let $\varphi,\psi : I \to \mathbb{R}$ and $\Delta : I \to \mathbb{R}$ be functions, and let $c_1 \in L^1(\mu)$ and $c_2 \in L^2(\nu)$ be such that the inequality*

$$\varphi(x)+\psi(y)+\Delta(x)(y-x) \leq c_1(x)+c_2(y)$$

holds on a set $A\times B$ with $\mu(A)=\nu(B)=1$. Then there is an interval $I' \subseteq I$ with $\mu(I')=\nu(I')=1$, and a convex function $\chi : I' \to \mathbb{R}$ such that $\varphi-\chi \leq c_1$ μ-a.e. and $\psi+\chi \leq c_2$ ν-a.e.

Proof. For notational convenience we define $f=\varphi-c_1$ and $g=\psi-c_2$. This results in

$$f(x)+g(y)+\Delta(x)(y-x) \leq 0. \tag{3.8}$$

We also define the lower convex envelope of $-g$ as in Definition B.1.2,

$$\chi(x) = \sup\{\alpha(x) \mid \alpha : I \to \mathbb{R} \text{ convex with } \alpha \leq -g \text{ on } B\}. \tag{3.9}$$

This is the largest convex function on I that is smaller than $-g$ on B by Proposition B.1.3. We need to check that this function is not equal to $-\infty$ somewhere on B. Note that for any $x \in A$ the function $f_x(y) = f(x)+\Delta(x)(y-x)$ is convex, as for $t \in [0,1]$

$$\begin{aligned} f_x((1-t)y_1+ty_2) &= f(x)+\Delta(x)((1-t)y_1+ty_2-x) \\ &= (1-t+t)f(x)+\Delta(x)((1-t)y_1+ty_2-(1-t)x+tx) \\ &= (1-t)(f(x)+\Delta(x)(y_1-x))+t(f(x)+\Delta(x)(y_2-x)) \\ &= (1-t)f_x(y_1)+tf_x(y_2). \end{aligned}$$

Also, by (3.8) we know that $f_x(y) \leq -g(y)$ on B. So, we have found a finite convex function that is contained in the set seen in (3.9), meaning that $\chi(x)$ will not be be equal to $-\infty$. It will also be shown that $\chi < \infty$ on $I' = \text{co}(B)$, where $\text{co}(\cdot)$ is the convex hull of some set. From our assumptions we know that $\nu(I')=1$ and by the fact that $\mu \preceq \nu$ we also have that $\mu(I')=1$. By definition of χ we also know that

$$f(x)+\Delta(x)(y-x) \leq \chi(y).$$

As this holds for all $x \in A$ and $y \in \text{co}(B)$, we can look at the case when $x=y$. We see that for all $x \in A \cap I'$ we have $f(x) \leq \chi(x)$, which gives us

$$f(x) = \varphi(x)-c_1(x) \leq \chi(x) \iff \varphi(x)-\chi(x) \leq c_1(x) \quad \mu-\text{a.e.}$$

Similarly, we can look at $y \in \text{co}(B)\cap I'$. From which we get

$$g(y) = \psi(y)-c_2(y) \leq \chi(y) \iff \psi(y)-\chi(y) \leq c_2(y) \quad \nu-\text{a.e.}$$

□

We are almost ready to start the proof of Theorem 2.4.5. For the final preparations, we need to know what it means for a set to be regular and irreducible, and how this relates to martingale couplings. Recall, that for a set $B \subseteq \mathbb{R}^d \times \mathbb{R}^d$, we denote by Γ_x the fibre of x in Γ, meaning $\Gamma_x = \{y \in \mathbb{R}^d \,|\, (x, y) \in \Gamma\}$.

Definition 3.2.6. Let I be an open interval. A set $A \subseteq \mathbb{R} \times \mathbb{R}$ is called *regular* on I if $A \subseteq I \times \overline{I}$ and for every $x \in I$ we have $A_x = \varnothing$ or $A_x = \{x\}$ or $x \in (\inf A_x, \sup A_x)$. Here, $\overline{I}$ indicates the closure of I.

A set A is called *irreducible* on I if $A \subseteq I \times \overline{I}$ and for every $y \in I$ there exist $x \in I$ and $y^-, y^+ \in A_x$ so that $y^- < y < y^+$.

Now that we know what regular and irreducible set are we can state the following proposition.

Proposition 3.2.7. *Assume that $c : \mathbb{R}^2 \to \mathbb{R}$ is continuous and let Γ be a $(c, \mathcal{F})$-monotone set that is regular and irreducible on some open interval I. Then there exist upper semicontinuous functions $\varphi : I \to [-\infty, \infty), \psi : \mathrm{co}(\mathrm{proj}_2(\Gamma)) \to [-\infty, \infty)$ and a measurable function $\Delta : I \to \mathbb{R}$ such that*

$$\varphi(x) + \psi(y) + \Delta(x)(y - x) \leq c(x, y)$$

for all $x \in I$, $y \in \mathrm{co}(\mathrm{proj}_2(\Gamma))$, with equality holding whenever $(x, y) \in \Gamma$. Here, co denotes the convex hull of a set.

Proof. See the proof of Proposition A.10 in [12]. □

In the proof of Theorem 2.4.5 we will need that the $(c, \mathcal{F}_M)$-monotone martingale coupling is concentrated on a $(c, \mathcal{F}_M)$-monotone set that is additionally irreducible and irregular, because we want to apply Proposition 3.2.7. The following three results ensure that we can assume that this is the case, without losing any generality. Before we give those results, we will need a definition of irreducibility of measures.

Definition 3.2.8. Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ such that $\mu \preceq \nu$. We say that (μ, ν) is irreducible if there exists an open interval I such that $\mu(I) = 1$ and $\nu(\overline{I}) = 1$ and

$$\int_{\mathbb{R}} |y - x| \mu(\mathrm{d}y) < \int_{\mathbb{R}} |y - x| \nu(\mathrm{d}y) \quad \forall x \in I.$$

Theorem 3.2.9 [Decomposition of (μ, ν)]. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ such that $\mu \preceq \nu$. Then there exist subprobabilities $\{\mu_k\}_{k \in J}$ and $\{\nu_k\}_{k \in J}$ with $\mu_k \preceq \nu_k$ and (μ_k, ν_k) irreducible on I_k for each $k \in J$, such that*

$$\mu = \sum_{k \in J} \mu_k, \quad \nu = \sum_{k \in J} \nu_k$$

.

The set J is an index set that is either equal to $\mathbb{N}$ or $\{1, \ldots, N\}$ for some $N \in \mathbb{N}$. Additionally, any martingale coupling $\pi \in \mathcal{M}(\mu, \nu)$ can uniquely be decomposed into subprobabilities $\{\pi_k\}_{k \in J}$ such that

$$\pi = \sum_{k \in J} \pi_k,$$

and each π_k is a martingale coupling between μ_k and ν_k.

Proof. See the proof of Theorem A.4 in [12]. □

We can reduce the MOT problem to the components found in Theorem 3.2.9.

Proposition 3.2.10. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ such that $\mu \preceq \nu$, $\pi \in \mathcal{M}(\mu, \nu)$ with decompositions $\{\mu_k\}_{k \in J}$, $\{\nu_k\}_{k \in J}$ and $\{\pi_k\}_{k \in J}$ as in Theorem 3.2.9. Let $c : \mathbb{R} \times \mathbb{R} \to \infty$ be lower semicontinuous with $a(x) \in L^1(\mu)$,$\nu \in L^1(\nu)$ such that $c(x, y) \leq a(x) + b(x)$. Then, a martingale coupling π is optimal if and only if every π_k is optimal for the MOT problem between μ_k and ν_k*

Proof. See the proof of Corollary A.4 in [12]. □

The final result we will need is that we can choose the $(c, \mathcal{F}_M)$-monotone set to be regular and irreducible, whenever μ and ν are already irreducible.

Lemma 3.2.11. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ such that (μ, ν) is irreducible on the set I. Let $c : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ be a measurable function. Let π be a $(c, \mathcal{F}_M)$-monotone probability measure concentrated on Γ. Then, there exists a Borel measurable set $\widetilde{\Gamma} \subseteq \Gamma \cap (I \times \hat{I})$ that is regular and irreducible on I such that $\pi(\widetilde{\Gamma}) = 1$. Moreover, $\widetilde{\Gamma}$ is still a $(c, \mathcal{F}_M)$-monotone set.*

Proof. See the proof of Lemma A.9 in [12]. □

We are now ready to state the proof of Theorem 2.4.5. We state the full theorem again as a reminder.

Theorem 2.4.5. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ with $\mu \preceq \nu$. Let $c : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ be measurable and $c(x, y) \leq a(x) + b(y)$ for $a \in L^1(\mu)$ and $L^1(\nu)$. Then any $(c, \mathcal{F}_M)$-monotone $\pi \in \mathcal{M}(\mu, \nu)$ is optimal for the MOT problem with measures μ and ν.*

Proof. As was stated before, we may assume that π is $(c, \mathcal{F})$-monotone on a regular and irreducible set Γ. If this is not the case, we can decompose μ, ν and π as in Theorem 3.2.9. For each component (μ_k, ν_k) and π_k we can invoke Lemma 3.2.11 to get a regular and irreducible $(c, \mathcal{F}_M)$-monotone set Γ for π_k. The proof below then ensures that π_k is optimal. Repeating this for each component and using Proposition 3.2.10 allows us to say that π is optimal.

So, let π be a $(c, \mathcal{F}_M)$-monotone measure and assume without loss of generality that Γ is regular and irreducible in addition to being a $(c, \mathcal{F}_M)$-monotone set for π. First invoking Proposition 3.2.7 we can get functions φ, ψ and Δ such that

$$\varphi(x) + \psi(y) + \Delta(x)(y - x) \leq c(x, y) \tag{3.10}$$

with equality on the $(c, \mathcal{F})$-monotone set Γ. We can now apply Lemma 3.2.5 and get a convex function that fits the requirements that are needed for Definition 3.2.2. Similarly as in Lemma 3.2.4 we define

$$\xi(x, y) = \varphi(x) + \psi(y) + \Delta(y)(y - x).$$

The function ξ is bounded by c, which in turn is bounded by two integrable functions c_1 and c_2. So ξ must also be integrable for every coupling between μ and ν. Let $\pi' \in \mathcal{M}(\mu, \nu)$, then by Lemma 3.2.4 we have both

$$G(\varphi, \psi) = \int\limits_{\mathbb{R} \times \mathbb{R}} \xi(x, y) \pi(\mathrm{d}x, \mathrm{d}y)$$

and

$$G(\varphi, \psi) = \int\limits_{\mathbb{R}\times\mathbb{R}} \xi(x, y)\pi'(\mathrm{d}x, \mathrm{d}y).$$

However, we also have

$$\int\limits_{\mathbb{R}\times\mathbb{R}} \xi(x, y)\pi(\mathrm{d}x, \mathrm{d}y) = \int\limits_{\mathbb{R}\times\mathbb{R}} c(x, y)\pi(\mathrm{d}x, \mathrm{d}y)$$

and

$$\int\limits_{\mathbb{R}\times\mathbb{R}} \xi(x, y)\pi'(\mathrm{d}x, \mathrm{d}y) \leq \int\limits_{\mathbb{R}\times\mathbb{R}} c(x, y)\pi'(\mathrm{d}x, \mathrm{d}y),$$

because π is concentrated on Γ and π' is in general not concentrated on Γ. The functions c and ξ are equal on the set Γ and ξ is bounded from above by c outside of Γ because of (3.10) and Lemma 3.2.5. Combining this all together gives us the desired result,

$$\int\limits_{\mathbb{R}\times\mathbb{R}} c(x, y)\pi(\mathrm{d}x, \mathrm{d}y) = \int\limits_{\mathbb{R}\times\mathbb{R}} \xi(x, y)\pi(\mathrm{d}x, \mathrm{d}y) = \int\limits_{\mathbb{R}\times\mathbb{R}} \xi(x, y)\pi'(\mathrm{d}x, \mathrm{d}y) \leq \int\limits_{\mathbb{R}\times\mathbb{R}} c(x, y)\pi'(\mathrm{d}x, \mathrm{d}y).$$

$\square$

4. Weak adapted topology

In Chapter 2 we introduced the spaces $\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d)$ and $\mathcal{P}_r(\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d))$, which helped us to create new meaningful monotonicity principles for the MOT and MOWT problem. These spaces were originally introduced to study so called adapted distances for the laws of processes. The problem these adapted distances try to resolve is the fact that it is possible to construct a sequence of processes that converge in the weak topology, while the limit process has a completely different causal structure compared to the processes in the sequence [6]. We will start this chapter with an example of such a process to explain what is meant by a process having different causal structures. After this clarifying example we will move onto our study of the maps $\hat{I}$ and J and prove that they preserve relative compactness.

4.1. Introduction

We will now show an example of processes that have different causal structures, but for which weak convergence holds. This example is taken from the introduction of [6]. It will also be our goal to illustrate why the map J resolves the emerging problems. To construct a sequence of processes for which the limit process has a different causal structure, define the probability measures on $\mathbb{R} \times \mathbb{R}$

$$\begin{aligned} \mathbb{P}^\epsilon &:= \frac{1}{2}\delta_{(\epsilon,2)} + \frac{1}{2}\delta_{(-\epsilon,-2)} \\ \mathbb{P} &:= \frac{1}{2}\delta_{(0,2)} + \frac{1}{2}\delta_{(0,-2)} \end{aligned}. \tag{4.1}$$

These measures describe 2-step processes taking values in $\mathbb{R}$. The interpretation of the $(x, y) \in \mathbb{R} \times \mathbb{R}$ in the subscripts of Dirac measures in (4.1) is that x indicates the position at time t_1 and y indicates the position at time t_2. See the Figure 4.1 for a visual representation. It can be shown

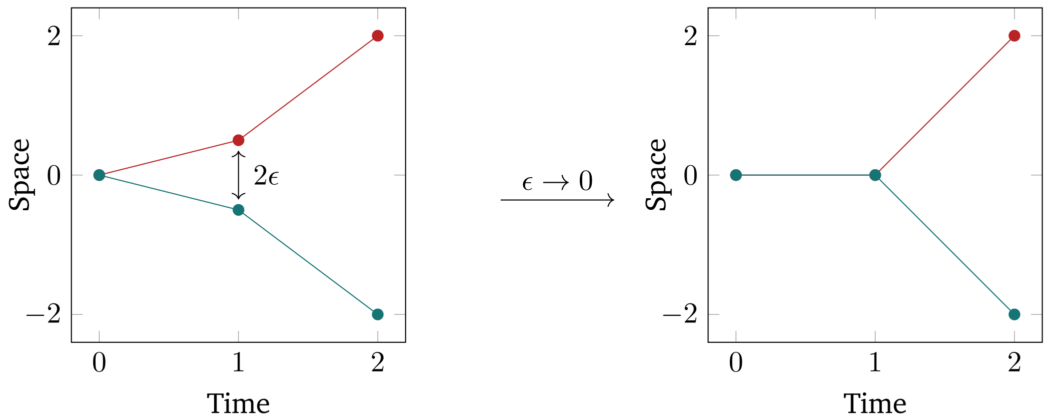


Figure 4.1.: First plot shows the process described by $\mathbb{P}^\epsilon$ and the second plot illustrates the process described by $\mathcal{P}$

that $\mathbb{P}^\epsilon \to \mathbb{P}$ in both the weak topology and the Wasserstein topology, which seems logical when looking at the pictures in Figure 4.1.

However, if we think about the causal structure of these processes, then it is not so clear why the distributions of these processes should converge to each other. The reason is the following. In the process described by $\mathbb{P}^\epsilon$ with $\epsilon > 0$, the first step is stochastic, but the second step is deterministic given the first step. In the process described by $\mathbb{P}$, the first step is deterministic, but the second step is stochastic given the first step. The fact that $\mathbb{P}^\epsilon$ does converge to $\mathbb{P}$ in the weak and Wasserstein topology has negative consequences for optimal stopping problems, utility maximization and stochastic programming [7]. Different authors have developed various techniques to circumvent these issues. All these techniques result into so called *weak adapted topologies*. We will introduce only one of these techniques, called the *adapted Wasserstein distance*. Unfortunately, there are too many other techniques and the theory behind them is too rich for a full discussion in this thesis. We refer to [7] for a more detailed overview. In that article Backhoff-Veraguas et al. proved that all these methods result in equivalent topologies.

The adapted Wasserstein distance was independently introduced by many different authors under different names. It was first introduced by Vershik in [51, 52]. Pflug and Pichler introduced an equivalent concept and showed its potential in various stochastic optimization problems [40, 41]. More recently, Laselle was able to connect this distance to a causal version of the optimal transport problem [36]. The adapted Wasserstein distance also found an application in finance in this form [6]. This is the version we will briefly introduce, but for more details we refer to its respective papers. We will need two definitions, the notion of a *bi-causal* coupling and the actual adapted Wasserstein distance.

Definition 4.1.1 [Bi-causal coupling]. Let $\mathcal{X}, \mathcal{Y}$ be Polish spaces and take $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$, with filtrations $\mathcal{F}_1^{\mathcal{X}} \subseteq \mathcal{F}_2^{\mathcal{X}}$ and $\mathcal{F}_1^{\mathcal{Y}} \subseteq \mathcal{F}_2^{\mathcal{Y}}$ for the σ-algebras of $\mathcal{X}$ and $\mathcal{Y}$ respectively. A coupling $\pi \in \Pi(\mu, \nu)$ is called *bi-causal* if the maps

$$\begin{aligned} &\mathcal{X} \to [0,1], x \mapsto \pi_x(A), \\ &\mathcal{Y} \to [0,1], y \mapsto \pi_y(B), \end{aligned}$$

are $\mathcal{F}_t^{\mathcal{X}}$- and $\mathcal{F}_t^{\mathcal{Y}}$-measurable, respectively, for all $t \in \{1,2\}$, $A \in \mathcal{F}_t^{\mathcal{X}}$ and $B \in \mathcal{F}_t^{\mathcal{Y}}$. Here, π_x and π_y indicate the regular disintegration with respect to x and y respectively. We denote this set of couplings by $\Pi_{bc}(\mu, \nu)$.

Definition 4.1.2 [Adapted Wasserstein distance]. Let $(\mathcal{X}, d)$ be a metric Polish space, $r \in [1, \infty)$ and $\mu, \nu \in \mathcal{P}_r(\mathcal{X})$, the *adapted Wasserstein distance of order* r is defined as

$$|\mathcal{AW}_r(\mu,\nu)|^r = \inf_{\pi \in \Pi_{bc}(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d(x,y)^r \pi(\mathrm{d}x, \mathrm{d}y).$$

It turns out that the embedding map J plays an important role in the theory of the adapted Wasserstein distance. Namely, for two measures $\mu, \nu \in \mathcal{P}_1(\mathcal{X} \times \mathcal{Y})$ we have the following equality by Theorem 4.4 in [50]

$$D_1(J(\mu), J(\nu)) = \mathcal{AW}_1(\mu, \nu),$$

where D_1 is the Wasserstein-1 distance on $\mathcal{P}_1(\mathcal{X} \times \mathcal{P}_1(\mathcal{Y}))$ as defined in Section 2.3. This is important, because in general the space $(\mathcal{P}_1(\mathcal{X} \times \mathcal{Y}), \mathcal{AW}_1)$ will not be complete anymore. However, we can find a completion of this space by embedding $\mathcal{P}_1(\mathcal{X} \times \mathcal{Y})$ into $\mathcal{P}_1(\mathcal{X} \times \mathcal{P}_1(\mathcal{Y}))$ through J. The completed space is then given by $(\mathcal{P}_1(\mathcal{X} \times \mathcal{P}_1(\mathcal{Y})), D_1)$. See Theorem 4.7 in [50] for more details.

Let us return to our example as described in (4.1) to illustrate that this technique resolves our problem. If we apply the map J to the measures given in (4.1) we get

$$\begin{aligned}
J(\mathbb{P}^\epsilon)(\mathrm{d}x, \mathrm{d}p) &= \mathbb{P}^\epsilon \circ \mathrm{proj}_1^{-1}(\mathrm{d}x)\delta_{\mathbb{P}^\epsilon_x}(\mathrm{d}p) \\
&= \frac{1}{2}\delta_\epsilon(\mathrm{d}x)\delta_{\mathbb{P}^\epsilon_\epsilon}(\mathrm{d}p) + \frac{1}{2}\delta_{-\epsilon}(\mathrm{d}x)\delta_{\mathbb{P}^\epsilon_{-\epsilon}}(\mathrm{d}p) \\
&= \frac{1}{2}\delta_{(\epsilon,\delta_2)}(\mathrm{d}x, \mathrm{d}p) + \frac{1}{2}\delta_{(-\epsilon,\delta_{-2})}(\mathrm{d}x, \mathrm{d}p).
\end{aligned}$$

This measure actually converges to the measure given by $\tilde{\mathbb{P}} = \frac{1}{2}\delta_{(0,\delta_2)} + \frac{1}{2}\delta_{(0,\delta_{-2})}$, while $J(\mathbb{P})$ is given by

$$\begin{aligned}
J(\mathbb{P}) &= \mathbb{P} \circ \mathrm{proj}_1^{-1}(\mathrm{d}x)\delta_{\mathbb{P}_x}(\mathrm{d}p) \\
&= \delta_0(\mathrm{d}x)\delta_{\mathbb{P}_0}(\mathrm{d}p) \\
&= \delta_0(\mathrm{d}x)\delta_{\frac{1}{2}\delta_2+\frac{1}{2}\delta_{-2}}(\mathrm{d}p) \\
&= \delta_{\left(0,\frac{1}{2}\delta_{(0,\delta_2)}+\frac{1}{2}\delta_{(0,\delta_{-2})}\right)}.
\end{aligned}$$

This illustrates that we do not have $\mathbb{P}^\epsilon \to \mathbb{P}$ when $\epsilon \to 0$ in this new topology! The new topology remembers the causal structure of the original processes and carries this structure to the limit, which is what we wanted. This concludes this short discussion on why the map J is an interesting and even a natural object to study.

4.2. Tightness & Compactness

We will now continue our study of the map J and $\hat{I}$ and prove that they both preserve relative compactness as stated in Lemma 2.3.2. We will follow the proof in [4] closely. To show that J preserves relative compactness, we will need a clear characterisation of relative compactness in the space $\mathcal{P}_r(\mathcal{P}_r(\mathbb{R}^d))$. Once that characterisation is clear, we will show that relative compactness in $\mathcal{P}_r(\mathcal{P}_r(\mathbb{R}^d)$ can be related to relative compactness in $\mathcal{P}_r(\mathbb{R}^d)$. Using that result, we can show relative compactness of J.

Recall that in the weak topology, relative compactness is equivalent to tightness by Prokhorov's theorem (Theorem A.1.6). However, when we are working in the $\mathcal{W}_r$-topology, we need to strengthen that result a bit. Lemma 4.2.1 tells us that we will need an additional uniform integrability assumption on top of tightness to establish relative compactness in the $\mathcal{W}_r$-topology.

Lemma 4.2.1. *A set $\mathcal{A} \subseteq \mathcal{P}_r(\mathbb{R}^d)$ is relatively compact in the $\mathcal{W}_r$-topology if and only if it is tight and there exists a $y' \in \mathbb{R}^d$ such that for all $\epsilon > 0$, there exists an $R > 0$, for which*

$$\sup_{\mu\in\mathcal{A}} \int_{\mathbb{R}^d\setminus B_R(y')} d(y, y')^r \mu(\mathrm{d}y) < \epsilon. \tag{4.2}$$

If the property (4.2) *holds for some y', then it holds for any $y' \in \mathbb{R}^d$ automatically.*

Proof. Let us first prove the final statement of the lemma, that (4.2) holds for all $y' \in \mathbb{R}^d$ if it is true for just one y'. Let $\epsilon > 0$ and assume that (4.2) holds for some $y' \in \mathbb{R}^d$ and $R > 0$. Pick another $y_0 \in \mathbb{R}^d$ and set $R_0 = R + d(y', y_0) + 1$. Note that the ball around y_0 with radius R_0 covers $B_R(y')$ completely. We also have that at least $R_0 > d(y', y_0) + 1$, which ensures that

$B_1(y') \subseteq B_{R_0}(y_0)$ and consequently that $1 \leq d(y,y')^r$ for $y \in \mathbb{R}^d \setminus B_{R_0}(y_0)$. Using the bound in (4.2) then gives us

$$\sup_{\mu\in\mathcal{A}} \int\limits_{\mathbb{R}^d\setminus B_{R_0}(y_0)} \mu(\mathrm{d}y) \leq \sup_{\mu\in\mathcal{A}} \int\limits_{\mathbb{R}^d\setminus B_R(y')} d(y,y')^r \mu(\mathrm{d}y) < \epsilon. \tag{4.3}$$

Using the triangle inequality we see that

$$\begin{aligned}
\sup_{\mu\in\mathcal{A}} \int\limits_{\mathbb{R}^d\setminus B_{R_0}(y_0)} d(y,y_0)^r \mu(\mathrm{d}y) &\leq \sup_{\mu\in\mathcal{A}} \int\limits_{\mathbb{R}^d\setminus B_{R_0}(y_0)} d(y,y')^r \mu(\mathrm{d}y) + \int\limits_{\mathbb{R}^d\setminus B_{R_0}(y_0)} d(y',y_0)^r \mu(\mathrm{d}y) \\
&\leq \sup_{\mu\in\mathcal{A}} \int\limits_{\mathbb{R}^d\setminus B_R(y')} d(y,y')^r \mu(\mathrm{d}y) + d(y',y_0)^r \int\limits_{\mathbb{R}^d\setminus B_R(y')} \mu(\mathrm{d}y) \\
&< \epsilon + d(y',y_0)^r \epsilon = (1 + d(y',y_0)^r)\epsilon.
\end{aligned}$$

The second inequality follows from the fact that $B_R(y') \subseteq B_{R_0}(y_0)$. The final inequality follows from (4.2) and (4.3). As ϵ is arbitrary, this shows that property (4.2) holds for any y_0 if it holds for one y'.

"$\Rightarrow$" We know that $\mathcal{W}_r$ topology is stronger than the weak topology, meaning that relative compactness in the $\mathcal{W}_r$ topology implies relative compactness in the weak topology. We can thus use Prokhorov to see that $\mathcal{A}$ is tight. Assume that (4.2) fails, meaning that there exist $y' \in \mathbb{R}^d$ and $\epsilon > 0$ such that for all $N \in \mathbb{N}$ there is a $\mu_N \in \mathcal{A}$ such that

$$\int\limits_{\mathbb{R}^d\setminus B_N(y')} d(y,y')^r \mu_N(\mathrm{d}y) \geq \epsilon, \tag{4.4}$$

which we will show implies a contradiction. We used here that property (4.2) should hold for all $y' \in \mathbb{R}^d$ by the argument at the beginning of this proof. The inequality in (4.4) implies

$$\lim_{R\to\infty} \liminf_{N\in\mathbb{N}} \int\limits_{\mathbb{R}^d\setminus B_R(y')} d(y,y')^r \mu_N(\mathrm{d}y) \geq \epsilon. \tag{4.5}$$

However, because $\mathcal{A}$ is relatively compact, we can find a convergent subsequence in $\mathcal{P}_r(\mathbb{R}^d)$ for any sequence in $\mathcal{A}$. In particular, we can do this for the sequence $\{\mu_N\}_{N\in\mathbb{N}}$ By the different characterisations of convergence in the $\mathcal{W}_r$ topology as described in Theorem 1.5.5, this immediately means that

$$\lim_{R\to\infty} \liminf_{N\in\mathbb{N}} \int\limits_{\mathbb{R}^d\setminus B_R(y')} d(y,y')^r \mu_N(\mathrm{d}y) = 0.$$

This is a clear contradiction with (4.5).

"$\Leftarrow$ Let $\mathcal{A}$ be tight and such that (4.2) holds. As $\mathcal{A}$ is tight, we know that any sequence $\{\mu_k\}_{k\in\mathbb{N}}$ has an accumulation point μ somewhere in $\mathcal{P}(\mathbb{R}^d)$ with respect to the weak topology. We will first show that actually $\mu \in \mathcal{P}_r(\mathbb{R}^d)$, i.e. that μ has an r-th moment. Then, we will show that the convergence also holds in the $\mathcal{W}_r$ topology.

By taking subsequences and renumbering the sequence, we can assume without loss of generality that $\mu_k \to \mu$ weakly for $k \to \infty$. By the monotone convergence theorem, and the Portmonteau theorem (Theorem A.1.4), we see that

$$\begin{aligned}\int\limits_{\mathbb{R}^d} d(y,y')^r \mu(\mathrm{d}y) &= \lim_{R\to\infty} \int\limits_{\mathbb{R}^d} R \wedge d(y,y')^r \mu(\mathrm{d}y) \\ &\le \lim_{R\to\infty} \liminf_{k\to\infty} \int\limits_{\mathbb{R}^d} R \wedge d(y,y')^r \mu_k(\mathrm{d}y) \le \sup_{k\in\mathbb{N}} \int\limits_{\mathbb{R}^d} d(y,y')^r \mu_k(\mathrm{d}y).\end{aligned}$$

We can now invoke (4.2) together with an $\epsilon > 0$ and see that for some $R > 0$ one has that

$$\begin{aligned}\int\limits_{\mathbb{R}^d} d(y,y')^r \mu(\mathrm{d}y) &\le \sup_{k\in\mathbb{N}} \int\limits_{B_R(y')} d(y,y')^r \mu_k(\mathrm{d}y) + \sup_{k\in\mathbb{N}} \int\limits_{\mathbb{R}^d\setminus B_R(y')} d(y,y')^r \mu_k(\mathrm{d}y) \\ &\le \sup_{k\in\mathbb{N}} \int\limits_{B_R(y')} d(y,y')^r \mu_k(\mathrm{d}y) + \epsilon < \infty.\end{aligned}$$

This shows that $\mu \in \mathcal{P}_r(\mathbb{R}^d)$. Finally, fix some, possibly other, $\epsilon > 0$. Pick $R > 0$ such that

$$\int\limits_{\mathbb{R}^d} d(y,y')^r - R^r \wedge d(y,y')^r \mu(\mathrm{d}y) < \epsilon, \tag{4.6}$$

$$\sup_{k\in\mathbb{N}} \int\limits_{\mathbb{R}^d\setminus B_R(y')} d(y,y')^r \mu_k(\mathrm{d}y) < \epsilon. \tag{4.7}$$

The bound in (4.6) can be achieved, because of the monotone convergence theorem (Theorem A.1.2). Note that the function $f_R(y) = R^r \wedge d(y,y')^r$ increases monotonically to $f(y) = d(y,y')^r$. The bound in (4.7) holds because of (4.2). The weak convergence of $\{\mu_k\}_{k\in\mathbb{N}}$ gives us

$$\lim_{k\to\infty} \int\limits_{\mathbb{R}^d} R^d \wedge d(y,y')^r \mu_k(\mathrm{d}y) = \int\limits_{\mathbb{R}^d} R^d \wedge d(y,y')^r \mu(\mathrm{d}y),$$

as $y \mapsto R^r \wedge d(y,y')^r \in C_b(\mathbb{R}^d)$. Because of this, we can find a k_0 such that for all $k \ge k_0$

$$\left| \int\limits_{\mathbb{R}^d} R^d \wedge d(y,y')^r (\mu_k - \mu)(\mathrm{d}y) \right| < \epsilon. \tag{4.8}$$

Putting these three bounded quantities together gives us for $k \geq k_0$

$$\begin{aligned}\left|\int_{\mathbb{R}^d} d(y,y')^r(\mu_k-\mu)(\mathrm{d}y)\right| &= \left|\int_{\mathbb{R}^d} R^r \wedge d(y,y')^r(\mu_k-\mu)(\mathrm{d}y)\right. \\ &\quad \left.+ \int_{\mathbb{R}^d} d(y,y')^r - R^r \wedge d(y,y')^r(\mu_k-\mu)(\mathrm{d}y)\right| \\ &\overset{(4.8)}{<} \epsilon + \left|\int_{\mathbb{R}^d} d(y,y')^r - R^r \wedge d(y,y')\mu(\mathrm{d}y)\right| \\ &\quad + \left|\int_{\mathbb{R}^d} d(y,y')^r - R^r \wedge d(y,y')^r \mu_k(\mathrm{d}y)\right| \\ &\overset{(4.6)}{<} 2\epsilon + \left|\int_{\mathbb{R}^d \setminus B_R(y')} d(y,y')^r \mu_k(\mathrm{d}y)\right| < 3\epsilon.\end{aligned}$$

As $\epsilon > 0$ was arbitrary, this shows that $\mu_k \to \mu$ in the $\mathcal{W}_r$-topology as well. $\square$

A classical result for tightness is that we can characterise tightness in $\mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ by tightness in $\mathcal{P}(\mathbb{R}^d)$ using the map I from Definition 2.3.1. This characterisation is given in Lemma 4.2.2.

Lemma 4.2.2. *A set $\mathcal{A} \subseteq \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ is tight if and only if the set of its intensities $I(\mathcal{A})$ is tight in $\mathcal{P}(\mathbb{R}^d)$, where I is defined as in* (2.5).

Proof. See page 178, Chapter 2 in [48]. $\square$

We can derive a similar result using Lemma 4.2.2 for relative compactness in the $\mathcal{W}_r$ topology in in $\mathcal{P}_r(\mathcal{P}_r(\mathbb{R}^d)$ and $\mathcal{P}(\mathbb{R}^d)$.

Lemma 4.2.3. *A set $\mathcal{A} \subseteq \mathcal{P}_r(\mathcal{P}_r(\mathbb{R}^d))$ is relatively compact if and only if the set of its intensities $I(\mathcal{A})$ is relatively compact in $\mathcal{P}_r(\mathbb{R}^d)$, where I is defined as in* (2.5).

Proof. "⇒" Using item (iii) of Theorem 2.3.2, which was already proven, we know that I is a continuous map. This immediately shows the implication from left to right, as continuous maps preserve relative compactness.

"⇐" To show the other direction, we will use the characterisation given by Lemma 4.2.1. We need to verify tightness of $\mathcal{A}$ and for some $y' \in \mathbb{R}^d$ that

$$\forall \epsilon > 0 \exists R_\epsilon > 0 : \sup_{P \in \mathcal{A}} \int_{\{p \in \mathcal{P}(\mathbb{R}^d) \,|\, W_r(p,\delta_{y'}) \geq R_\epsilon\}} \mathcal{W}_r(p,\delta_{y'})^r P(\mathrm{d}p) < \epsilon. \tag{4.9}$$

We start by showing (4.9). So let $I(\mathcal{A})$ be relatively compact in $\mathcal{P}_r(\mathbb{R}^d)$ and fix an $\epsilon > 0$. By first applying Lemma 4.2.1 to the set $I(\mathcal{A})$ we see that it is tight and that (4.2) holds. By tightness, we can find a $K > 0$ such that

$$\int_{\mathbb{R}^d} d(y,y')^r I(P)(\mathrm{d}y) \leq K,$$

for all $P \in \mathcal{A}$. Note that the product measure is the only coupling between $\delta_{y'}$ and any $p \in \mathcal{P}_r(\mathbb{R}^d)$. Using this we get for all $P \in \mathcal{A}$

$$\int\limits_{\mathcal{P}_r(\mathbb{R}^d)} \mathcal{W}_r(p, \delta_{y'})^r P(\mathrm{d}p) = \int\limits_{\mathcal{P}_r(\mathbb{R}^d)} \int\limits_{\mathbb{R}^d} d(y, y')^r p(\mathrm{d}y) P(\mathrm{d}p) = \int\limits_{\mathbb{R}^d} d(y, y')^r I(P)(\mathrm{d}y) \leq K. \tag{4.10}$$

Now we can use (4.2) to find an $R > 0$ such that

$$\int\limits_{\mathcal{P}_r(\mathbb{R}^d)} \int\limits_{\mathbb{R}^d \setminus B_R(y')} d(y, y')^r p(\mathrm{d}y) P(\mathrm{d}p) = \int\limits_{\mathbb{R}^d \setminus B_R(y')} d(y, y')^r I(P)(\mathrm{d}y) < \frac{\epsilon}{2}. \tag{4.11}$$

Set $R_\epsilon = \max\left\{\frac{2R^r K}{\epsilon}, R\right\}$ and $A_{R_\epsilon} = \{p \in \mathcal{P}_r(\mathbb{R}^d) \,|\, \mathcal{W}_r(p, \delta_{y'}) \geq R_\epsilon\}$, which gives us

$$\sup_{P \in \mathcal{A}} P(A_{R_\epsilon}) \leq \sup_{P \in \mathcal{A}} \frac{1}{R_\epsilon} \int\limits_{A_{R_\epsilon}} \mathcal{W}_r(p, \delta_{y'})^r P(\mathrm{d}p) \leq \frac{K}{R_\epsilon}.$$

The first inequality follows from Markov's inequality (Theorem A.1.3) and the second inequality follows from (4.10). Using this bound, we can find

$$\sup_{P \in \mathcal{A}} \int\limits_{A_{R_\epsilon}} \int\limits_{B_R(y')} d(y, y')^r p(\mathrm{d}y) P(\mathrm{d}p) \leq \sup_{P \in \mathcal{A}} P(A_{R_\epsilon}) R^r \leq \frac{K}{R_\epsilon} = \frac{K \epsilon R^r}{2 R^r K} = \frac{\epsilon}{2}. \tag{4.12}$$

Finally, we can combine (4.11) and (4.12) to show (4.9). This gives us

$$\begin{aligned}
\sup_{P \in \mathcal{A}} \int\limits_{A_{R_\epsilon}} \mathcal{W}_r(p, \delta_{y'})^r P(\mathrm{d}p) &\leq \sup_{P \in \mathcal{A}} \int\limits_{A_{R_\epsilon}} \int\limits_{\mathbb{R}^d} d(y, y')^r p(\mathrm{d}y) P(\mathrm{d}p) \\
\leq \sup_{P \in \mathcal{A}} \int\limits_{A_{R_\epsilon}} \int\limits_{B_{R_\epsilon}} d(y, y')^r p(\mathrm{d}y) P(\mathrm{d}p) &+ \sup_{P \in \mathcal{A}} \int\limits_{A_{R_\epsilon}} \int\limits_{\mathbb{R}^d \setminus B_{R_\epsilon}} d(y, y')^r p(\mathrm{d}y) P(\mathrm{d}p) \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}$$

Next, we need to show that $\mathcal{A}$ is tight in $\mathcal{P}(\mathcal{P}_r(\mathbb{R}^d))$. We can first invoke Lemma 4.2.2 to see that $\mathcal{A}$ is tight in $\mathcal{P}(\mathcal{P}(\mathbb{R}^d))$, meaning that for every $\epsilon > 0$ we can find a compact set $K_\epsilon \subseteq \mathcal{P}(\mathbb{R}^d)$ such that

$$\inf_{P \in \mathcal{A}} P(K_\epsilon) \geq 1 - \epsilon.$$

We will now construct a set $\widetilde{K}_\epsilon \subseteq K_\epsilon$ that is compact in $\mathcal{P}_r(\mathbb{R}^d)$ and has the property that

$$\inf_{P \in \mathcal{A}} P(\widetilde{K}_\epsilon) \geq 1 - \epsilon.$$

To construct this set, let $\{R_n\}_{n \in \mathbb{N}}$ be a sequence of numbers with $R_n > 0$ for all $n \in \mathbb{N}$ and

$$\sup_{P \in \mathcal{A}} P\left(\left\{p \in \mathcal{P}_r(\mathbb{R}^d) \,\middle|\, \int\limits_{\mathbb{R}^d \setminus B_{R_n}(y')} d_{\mathbb{R}^d}(y, y')^r p(\mathrm{d}y) \geq \frac{1}{n}\right\}\right) < \frac{\epsilon}{2^n}.$$

This is possible because of Lemma 4.2.1 and the fact that

$$P\left(\left\{p \in \mathcal{P}_r(\mathbb{R}^d) \,\middle|\, \int\limits_{\mathbb{R}^d \setminus B_{R_n}(y')} d_{\mathbb{R}^d}(y,y')^r p(\mathrm{d}y) \geq \frac{1}{n}\right\}\right) \leq n \int\limits_{\mathbb{R}^d \setminus B_{R_n}(y')} \int\limits_{\mathcal{P}_r(\mathbb{R}^d)} d(y,y') p(\mathrm{d}y) P(\mathrm{d}p)$$
$$= n \int\limits_{\mathbb{R}^d \setminus B_{R_n}(y')} d(y,y') p(\mathrm{d}y) I(P)(\mathrm{d}y).$$

The inequality follows directly from Markov's inequality as stated in A.1.3. The final expression above can be bounded uniformly in P by our assumption that $I(\mathcal{A})$ is already relatively compact. We can now define $\widetilde{K}_\epsilon$ as

$$\widetilde{K}_\epsilon := \left\{p \in K_\epsilon \,\middle|\, \int\limits_{\mathbb{R}^d \setminus B_{R_n}(y')} d(y,y')^r p(\mathrm{d}y) \leq \frac{1}{n}, \quad n \in \mathbb{N}\right\}.$$

This set is a closed set. Furthermore, it is relatively compact, because the condition in the set definition ensures precisely that Property (4.2) holds. So, $\widetilde{K}_\epsilon$ is in fact compact. We can now finish up by stating

$$\inf_{P \in \mathcal{A}} P(\widetilde{K}_\epsilon) \geq P(K_\epsilon) - \sum_{n=1}^{\infty} P\left(\left\{p \in \mathcal{P}_r(\mathbb{R}^d) \,|\, \int\limits_{\mathbb{R}^d \setminus B_{R_n}(y')} d(y,y') p(\mathrm{d}y) \geq \frac{1}{n}\right\}\right)$$
$$\geq 1 - \epsilon - \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} = 1 - 2\epsilon.$$

□

As a final preparatory step for the proof of Theorem 2.3.2, we show that mapping a measure to one of its marginals is a continuous operation. This allows us to check relative compactness of a set of measures on a product space, by looking at the marginal sets. This is a classical result when considering tightness of measures on the product space.

Lemma 4.2.4. *The sets $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^d)$, $\mathcal{B} \subseteq \mathcal{P}(\mathbb{R}^d)$ are tight, if and only if the set*

$$\mathcal{C} = \{\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \,|\, \pi \circ \mathrm{proj}_1^{-1} \in \mathcal{A} \text{ and } \pi \circ \mathrm{proj}_2^{-1} \in \mathcal{B}\}$$

is tight.

Proof. "⇒" Let $\epsilon > 0$ and pick $K_1 \subseteq \mathbb{R}^d$, $K_2 \subseteq \mathbb{R}^d$ such that

$$\sup_{\mu \in \mathcal{A}} \mu\left(\mathbb{R}^d \setminus K_1\right) < \frac{\epsilon}{2}$$
$$\sup_{\nu \in \mathcal{B}} \nu\left(\mathbb{R}^d \setminus K_2\right) < \frac{\epsilon}{2}.$$

This is possible, because both $\mathcal{A}$ and $\mathcal{B}$ are tight. We now see that the claim holds, as $K_1 \times K_2 \subseteq \mathbb{R}^d \times \mathbb{R}^d$ is compact and

$$\sup_{q \in \mathcal{C}} q\left(\mathbb{R}^d \times \mathbb{R}^d \setminus K_1 \times K_2\right) \leq \sup_{q \in \mathcal{C}} q\left(\left(\mathbb{R}^d \setminus K_1\right) \times \mathbb{R}^d\right) + \sup_{q \in \mathcal{C}} q\left(\mathbb{R}^d \times \left(\mathbb{R}^d \setminus K_2\right)\right)$$
$$= \sup_{\mu \in \mathcal{A}} \mu\left(\mathbb{R}^d \setminus K_1\right) + \sup_{\nu \in \mathcal{B}} \nu\left(\mathbb{R}^d \setminus K_2\right) < \epsilon.$$

"$\Leftarrow$" Let $\epsilon > 0$ and choose some compact set $K \subseteq \mathbb{R}^d \times \mathbb{R}^d$ such that

$$\sup_{\pi \in \mathcal{C}} \pi(\mathbb{R}^d \times \mathbb{R}^d \setminus K) < \epsilon.$$

Let $K_1 = \mathrm{proj}_1(K)$, then this is a compact set again, because the projection is a continuous function. We then get

$$\inf_{\mu \in \mathcal{A}} \mu(K_1) = \inf_{\pi \in \mathcal{C}} \pi \circ \mathrm{proj}_1^{-1}(K_1) = \inf_{\pi \in \mathcal{C}} \pi(K_1 \times \mathbb{R}^d) \geq \inf_{\pi \in \mathcal{C}} \pi(K) \geq 1 - \epsilon.$$

This shows tightness of $\mathcal{A}$. Similarly, we can show tightness of $\mathcal{B}$ by using proj_2. $\square$

For relative compactness in the $\mathcal{W}_r$-topology on the product space we get the following similar lemma.

Lemma 4.2.5. *Let $\pi \in \mathcal{P}_r(\mathbb{R}^d \times \mathbb{R}^d)$ and $r \in [1, \infty)$, then the mappings*

$$\begin{aligned} \mathrm{Mar}_1 &: \mathcal{P}_r(\mathbb{R}^d \times \mathbb{R}^d) \to \mathcal{P}_r(\mathbb{R}^d), \pi \mapsto \pi \circ \mathrm{proj}_1^{-1} = \mu, \\ \mathrm{Mar}_2 &: \mathcal{P}_r(\mathbb{R}^d \times \mathbb{R}^d) \to \mathcal{P}_r(\mathbb{R}^d), \pi \mapsto \pi \circ \mathrm{proj}_2^{-1} = \nu \end{aligned}$$

are continuous. Consequently, the sets $\mathcal{A} \subseteq \mathcal{P}_r(\mathbb{R}^d)$, $\mathcal{B} \subseteq \mathcal{P}(\mathbb{R}^d)$ are relatively compact if and only if the following set is relatively compact

$$\mathcal{C} = \{q \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \,|\, q \circ \mathrm{proj}_1^{-1} \in \mathcal{A} \text{ and } q \circ \mathrm{proj}_2^{-1} \in \mathcal{B}\}.$$

Proof. We show continuity for taking the first marginal only, as the proof of continuity of the second marginal is identical. Let $\{\pi_k\}_{k \in \mathbb{N}}$ be a sequence in $\mathcal{P}_r(\mathbb{R}^d \times \mathbb{R}^d)$ that converges in the $\mathcal{W}_r$-topology. In particular this means that we can test the convergence against integration of continuous functions $f(x, y)$ for which there exists a $C \in \mathbb{R}$ such that $|f(x, y)| \leq C(1 + d((x_0, y_0), (x, y))^r)$. Where d is a compatible metric on $\mathbb{R}^d \times \mathbb{R}^d$ and $(x_0, y_0) \in \mathbb{R}^d \times \mathbb{R}^d$ can be chosen arbitrarily. This class of functions in particular encompasses functions of the form $f(x, y) = g(x)$, which gives us continuity, as

$$\begin{aligned} \lim_{k \to \infty} \int_{\mathbb{R}^d} g(x) \pi_k \circ \mathrm{proj}_1^{-1}(\mathrm{d}x) &= \lim_{k \to \infty} \int_{\mathbb{R}^d \times \mathbb{R}^d} g(x) \pi_k(\mathrm{d}x, \mathrm{d}y) \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^d} g(x) \pi(\mathrm{d}x, \mathrm{d}y) = \int_{\mathbb{R}^d} g(x) \pi \circ \mathrm{proj}_1^{-1}(\mathrm{d}x). \end{aligned}$$

This shows that the first marginals of π_k also converge to the first marginal of π. We can repeat this argument but with functions of the form $f(x, y) = h(y)$ to get continuity of the second marginal.

Now that we have continuity, we can show that $\mathcal{A}$ and $\mathcal{B}$ are relatively compact if and only if $\mathcal{C}$ is relatively compact.

"$\Leftarrow$" This direction follows directly from the fact that mapping a measure to its first or second marginal is continuous, because

$$\begin{aligned} \mathcal{A} &= \mathrm{Mar}_1(\mathcal{C}), \\ \mathcal{B} &= \mathrm{Mar}_2(\mathcal{C}) \end{aligned}$$

and relative compactness is conserved by continuous mappings.

"$\Rightarrow$" The sets $\mathcal{A}$ and $\mathcal{B}$ are relatively compact in the $\mathcal{W}_r$-topology, which is stronger than the weak topology. So, $\mathcal{A}$ and $\mathcal{B}$ are also weakly relatively compact and by Prokhorov's theorem

(Theorem A.1.6) they are tight. Lemma 4.2.4 then tells us that $\mathcal{C}$ is also tight. We can now use Lemma 4.2.1 to show relative compactness. We only need to check that property (4.2) holds. We already now that this property must hold for $\mathcal{A}$ and $\mathcal{B}$. So, fix $\epsilon > 0$ and let $x' \in \mathbb{R}^d$, $y' \in \mathbb{R}^d$ and $R_1, R_2 > 0$ such that

$$\sup_{\mu \in \mathcal{A}} \int\limits_{\mathbb{R}^d \setminus B_{R_1}(x')} d(x,x')^r \mu(\mathrm{d}x) < \frac{\epsilon}{2},$$

$$\sup_{\nu \in \mathcal{B}} \int\limits_{\mathbb{R}^d \setminus B_{R_2}(x')} d(y,y')^r \nu(\mathrm{d}y) < \frac{\epsilon}{2}.$$

Take $R = \max\{R_1, R_2\}$, see that

$$\sup_{\pi \in \mathcal{C}} \int\limits_{\mathbb{R}^d \times \mathbb{R}^d \setminus B_R(x',y')} \widetilde{d}((x,y),(x',y'))^r \pi(\mathrm{d}x, \mathrm{d}y) \leq \sup_{\mu \in \mathcal{A}} \int\limits_{\mathbb{R}^d \setminus B_R(x')} d(x,x')^r \mu(\mathrm{d}x)$$
$$+ \sup_{\nu \in \mathcal{B}} \int\limits_{\mathbb{R}^d \setminus B_R(x')} d(y,y')^r \nu(\mathrm{d}y) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Here, the compatible metric $\mathbb{R}^d \times \mathbb{R}^d$ is given by $\widetilde{d}((x,y),(x',y'))^r = d(x,x')^r + d(y,y')^r$. This shows that Property 4.2.1 holds and we can conclude that $\mathcal{C}$ is relatively compact. $\square$

We now have all the ingredients to state the proof of items (i) and (ii) in Lemma 2.3.2. We want to show that the embedding map J and its left inverse $\hat{I}$ both preserve relative compactness in the $\mathcal{W}_r$-topology.

Proof of Lemma 2.3.2. Next, we show that if $\Lambda \subseteq \mathcal{P}_r(\mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d))$ is relatively compact, then $\hat{I}(\Lambda)$ also has this property. This is almost immediate. By the previous paragraph, we know that $\hat{I}$ is continuous and as continuous maps preserve relative compactness we see that $\hat{I}(\Lambda)$ is compact whenever Λ is relatively compact.

Next, we show that if $\Pi \subseteq \mathcal{P}_r(\mathbb{R}^d \times \mathbb{R}^d)$ is relatively compact, then also $J(\Pi)$. Let $\Pi \subseteq \mathcal{P}_r(\mathbb{R}^d \times \mathbb{R}^d)$ be relatively compact. Consider the following sets

$$\begin{aligned}
\Pi^1 &= \{p \in \mathcal{P}(\mathbb{R}^d) \,|\, \exists P \in \Pi \text{ such that } p = P \circ \mathrm{proj}_1^{-1}\} \\
\Pi^2 &= \{p \in \mathcal{P}(\mathbb{R}^d) \,|\, \exists P \in \Pi \text{ such that } p = P \circ \mathrm{proj}_2^{-1}\} \\
\Pi^1_J &= \{p \in \mathcal{P}(\mathbb{R}^d) \,|\, \exists P \in J(\Pi) \text{ such that } p = P \circ \mathrm{proj}_1^{-1}\} \\
\Pi^2_J &= \{p \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d)) \,|\, \exists P \in J(\Pi) \text{ such that } p = P \circ \mathrm{proj}_2^{-1}\}.
\end{aligned}$$

The first two sets are the first and second marginals, respectively. The last two sets consist of the $\mathbb{R}^d$- and $\mathcal{P}_r(\mathbb{R}^d)$-marginals of the measures in $J(\Pi)$. As the embedding J leaves the first marginal unchanged, we immediately have that $\Pi^1 = \Pi^1_J$. Now, if $p \in \Pi^2_J$, then p is the $\mathcal{P}_r(\mathbb{R}^d)$-marginal of some $J(\pi)$ with $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. The map I is just the second component of the left inverse of J. So, it must be that $I(p)$ is the second marginal of π. We see that $I(\Pi^2_J)$ is a subset of Π^2. Subsets of relatively compact subsets are again relatively compact, from which we conclude that $I(\Pi^2_J)$ is relatively compact. By Lemma 4.2.3 we know that Π^2_J is relatively compact if and only if $I(\Pi^2_J)$ is relatively compact, meaning that Π^2_J is also relatively compact. This shows that both marginal sets of $J(\Pi)$ are relatively compact, which implies that $J(\Pi)$ itself is relatively compact by Lemma 4.2.5.

If we now assume that $J(\Pi)$ is relatively compact, then it follows that $\Pi = \hat{I}(J(\Pi))$, which is relatively compact by the continuity of $\hat{I}$.

Finally, if $\hat{I}(\Lambda)$ is relatively compact, we can again argue that the set of first and second marginals are relatively compact because of Lemma 4.2.5. By Lemma 4.2.3 we now see that $\mathbb{R}^d$- and $\mathcal{P}_r(\mathbb{R}^d)$-marginals need to be relatively compact. From which we conclude that Λ must be relatively compact as well. $\square$

5. Computation

Up to this chapter, we only devoted ourselves to the theoretical aspects of the martingale optimal transport problem. However, its origins lie in an established practical application. Most notably, it is used to find robust bounds on prices of exotic options. Consequently, it is important to find numerical schemes that approximate the solution to the MOT problem well. In this chapter we will discuss the issue of actually calculating the minimizer and value of the MOT problem. Specifically, we will illustrate why the methods used in the classical OT problem do not translate well to the martingale case and we will introduce a method developed by Guo & Obłój that circumvents these issues. At the end of this chapter, we will showcase some examples to illustrate the effectiveness of the method. Throughout this chapter we will consider the version of the MOT problem that maximizes the cost functional, as this is more commonly done in practise.

5.1. Problem with OT methods

Let us first consider the OT problem for measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and consider the following setting. If μ and ν are discrete measures, then it is possible to rewrite the OT problem as *linear programming* (LP) problem [39]. This was in essence also the crucial insight of Kantorovich when he was working on the OT problem [33]. Write μ and ν as

$$\mu(\mathrm{d}x) = \sum_{i=1}^{n} \alpha_i \delta_{x_i}(\mathrm{d}x), \quad \nu(\mathrm{d}y) = \sum_{j=1}^{m} \beta_j \delta_{y_j}(\mathrm{d}y).$$

The equivalent LP problem of the OT problem with μ and ν becomes

$$\begin{aligned} \max p \in \mathbb{R}^{n\times m}_{\geq 0} \sum_{i=1}^{n}\sum_{j=1}^{m} p_{i,j} c(x_i, y_i) \text{ such that } & \sum_{j=1}^{m} p_{i,j} = \alpha_i, \quad \text{for } i = 1, \dots, n \\ & \sum_{i=1}^{n} p_{i,j} = \beta_j, \quad \text{for } j = 1, \dots, m. \end{aligned}$$

Presently, it is possible to solve these kinds of LP problems efficiently, using the simplex algorithm for example [9]. When Kantorovich was working on the OT problem, he did realise that you could extend this LP formulation in a meaningful way to non-discrete settings. However, he was not able to provide an algorithm, which was proven to work. This was possible only later, when the field of linear programming was more developed.

In more recent times, OT has found its way into many different disciplines. As a consequence, many new algorithms have been developed for different settings. See the book *Computational optimal transport* by Cuturi and Peyré for an extensive overview of such algorithms [39].

A logical step forward to finding an algorithm for the computation of the MOT problem is to once again consider an LP formulation. The good news is that this is possible, and the resulting

LP formulation looks like

$$\max p \in \mathbb{R}^{n\times m}_{\geq 0} \sum_{i=1}^{n}\sum_{j=1}^{m} p_{i,j}c(x_i,y_i) \text{ such that } \begin{aligned} &\sum_{j=1}^{m} p_{i,j} = \alpha_i, \quad \text{for } i=1,\dots,n \\ &\sum_{i=1}^{n} p_{i,j} = \beta_j, \quad \text{for } j=1,\dots,m \\ &\sum_{j=1}^{n} p_{i,j}y_j = \alpha_i x_i \quad \text{for } i=1,\dots,n. \end{aligned} \tag{5.1}$$

This formulation was first proposed by Davis et al. in [19]. For general measures one would approximate μ and ν with a sequence of discrete measures $\{(\mu_n,\nu_n)\}_{n\in\mathbb{N}}$ and solve the equivalent LP problem. The hope would then be that this approach will approximate the solution for a general MOT problem. However, there are two issues one runs into when using this approach:

(i) There are no general continuity results for the MOT problem. This means that we are not guaranteed to approximate the MOT value with a sequence of discrete measures. The stability results provided in Theorem 2.5.1 and Corollary 2.5.2 show that the MOT problem is continuous for measures defined on $\mathbb{R}$. However, for the general case there are still no known requirements to ensure continuity, if they exist at all.

(ii) Conventional methods of discretizing the measures μ and ν will in general not ensure that μ_n is less than ν_n in the convex order. The consequence of this is that the LP problem can become unsolvable. There are methods that will preserve the convex order [8], but these only work for measures defined on $\mathbb{R}$.

5.2. General method

The method given by Guo & Obłój in [28] circumvents the aforementioned issues by relaxing the martingale property and considering a sequence of Linear Programming problems, where the relaxed martingale property is gradually made stricter. The relaxation of the problem solves two problems. It ensures the restrictions in the resulting LP problem are not too strict, so that the LP problem can be solved. The second issue it solves, is that we do not have to require that the approximating μ_n and ν_n are in convex order. We will provide the full method and the convergence results in the case of $N=2$ and $d=1$ proved by Guo & Obłój. First we state what the setting is in which we will work

- The space we will work in is $\Omega = \mathbb{R}^d$.
- We will use the coordinate process $(S_k)_{k=1}^N$ for $N\in\mathbb{N}$, which is defined as

$$S_k(x_1,\dots,x_N) = x_k, \quad \forall (x_1,\dots,x_N)\in\Omega^N.$$

- The space Ω^N is endowed with the natural filtration $(\mathcal{F}_k)_{k=1}^N$ of $(S_k)_{k=1}^N$.
- The Wasserstein metric for vectors of probability measures, $\vec{\mu} = (\mu_1,\dots,\mu_k)$ and $\vec{\nu} = (\nu_1,\dots,\nu_k)$, is defined as

$$\mathcal{W}_r(\vec{\mu},\vec{\nu}) = \sum_{i=1}^{k} \mathcal{W}_r(\mu_i,\nu_i).$$

The interpretation of this setting is that we consider d stocks on N different trading dates. For a vector of probability measures $\vec{\mu} = (\mu_k)_{k=1}^N$ with $\mu\in\mathcal{P}(\Omega)$, we can define a set of couplings as

$$\Pi(\vec{\mu}) := \{\mathbb{P}\in\mathcal{P}(\Omega^N)\,|\,\mathbb{P}\circ S_k^{-1} = \mu_k, \quad \text{for } k=1,\dots,N\}.$$

Before we can introduce the relaxed version of the MOT problem we need the definition of an ϵ-approximating martingale.

Definition 5.2.1 [ϵ-approximating martingale]**.** Let $\epsilon \geq 0$, a probability measure $\mathbb{P} \in \mathcal{P}(\Omega^N)$ is said to be an *ϵ-approximating martingale measure* if for each $k = 1, \ldots, N$

$$\mathbb{E}_{\mathbb{P}}\left[|\mathbb{E}_{\mathbb{P}}[S_{k+1} \mid \mathcal{F}_k] - S_k|\right] \leq \epsilon. \tag{5.2}$$

We denote by $\mathcal{M}_\epsilon(\vec{\mu}) \subseteq \Pi(\vec{\mu})$, the set of all ϵ-approximating martingale measures that are compatible with $\vec{\mu}$. Finally, we set $\mathcal{P}_\epsilon \subseteq \mathcal{P}(\Omega)^N$ to be the collection of vectors of measures $\vec{\mu}$ such that $\mathcal{M}_\epsilon(\vec{\mu}) \neq \varnothing$.

The relaxed optimisation problem is now defined as follows.

Definition 5.2.2 [relaxed MOT problem]**.** Let $\vec{\mu} = (\mu_1, \ldots, \mu_N)$ with $\mu_k \in \mathcal{P}_1(\Omega)$ for all $k = 1, \ldots, N$, $\vec{\mu} \in \mathcal{P}_\epsilon$ and $c : \Omega^N \to \mathbb{R}$ a measurable function, the *relaxed MOT problem* is defined by finding the value or minimizing coupling of

$$V_M^\epsilon(\vec{\mu}) := \inf_{\mathbb{P} \in \mathcal{M}_\epsilon(\vec{\mu})} \mathbb{E}_{\mathbb{P}}\left[c(S_1, \ldots, S_N)\right].$$

The next theorem allows us to use the relaxed MOT to approximate the original MOT problem.

Theorem 5.2.3. *Fix a $\vec{\mu} = (\mu_1, \ldots, \mu_N)$ with $\mu_k \in \mathcal{P}_1(\Omega)$ for all $k = 1, \ldots, N$ and $\mu_k \preceq \mu_{k+1}$ for all $k = 1, \ldots, N-1$. Here, the symbol $\preceq$ indicates the convex order as described in Definition 2.1.3. Let $\{\vec{\mu}^n\}_{n \in \mathbb{N}}$ be a sequence with in $\mathcal{P}(\Omega)^N$ that converges to $\vec{\mu}$ in the $\mathcal{W}_1$ topology. Then, $\vec{\mu}^n \in \mathcal{P}_{r_n}$ with $r_n := \mathcal{W}_1(\vec{\mu}^n, \vec{\mu})$. We assume that $c : \Omega^N \to \mathbb{R}$ is a Lipschitz function. Then, the following two statements are true.*

(i) For any sequence $\{\epsilon_n\}_{n \in \mathbb{N}}$ with $\epsilon_n > 0$ such that $\epsilon_n \geq r_n$ for all $n \in \mathbb{N}$ and $\epsilon_n \to 0$, one has

$$\lim_{n \to \infty} V_M^{\epsilon_n}(\vec{\mu}^n) = V_M^0(\vec{\mu}).$$

(ii) For each $n \geq 1$, $V_M^{\epsilon_n}(\vec{\mu}^n)$ admits an optimizer $\mathbb{P}_n \in \mathcal{M}_{\epsilon_n}(\vec{\mu}^n)$, i.e. $V_M^{\epsilon_n}(\vec{\mu}^n) = \mathbb{E}_{\mathbb{P}_n}[c]$. The sequence $\{\mathbb{P}_n\}_{n \in \mathbb{N}}$ converges weakly whenever $V_M^0(\vec{\mu})$ has a unique optimizer.

Proof. See the proof of Theorem 2.2 in [28]. □

The idea is now to approximate $\vec{\mu}$ with a sequence of discrete measures, for which we can rewrite the relaxed MOT problem as an LP problem. The resulting LP problem can then be solved by any existing LP algorithm. If we approximate $\vec{\mu}$ carefully, i.e. such that the assumptions in Theorem 5.2.3 hold, then it is guaranteed that the sequence of LP problems converge to the true value of $V_M(\vec{\mu})$.

In the case of of $N = 2$ and $d = 1$ Guo & Obłój actually provide an estimation of the convergence rate of this approach.

Theorem 5.2.4. *In the same setting as in Theorem 5.2.3, let $N = 2$ and $d = 1$, which means that $\vec{\mu} = (\mu, \nu)$ and $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Additionally, assume that $\sup_{(x,y) \in \mathbb{R}^2} |\partial^2_{yy} c(x, y)| < \infty$ and ν has a finite second moment. Then, there exists a $C > 0$ such that*

$$|V_M^{\epsilon_n}(\mu^n, \nu^n) - V_M^0(\mu, \nu)| \leq C \inf_{R > 0} \lambda_n(R), \quad \forall n \geq 1,$$

where $\lambda_n : (0,\infty) \to \mathbb{R}$ is given by

$$\lambda_n(R) := (R+1)\epsilon_n + \int\limits_{(-\infty,R)\cup(R,\infty)} (|y|-R)^2 \nu(dy).$$

In particular, the convergence rate is proportional to ϵ_n if supp(ν) is bounded.

Proof. See the proof of Theorem 2.5 in [28]. □

We are now ready to state the algorithm for the MOT problem. In the following we will assume that $\vec{\mu} = (\mu_1, \dots, \mu_N)$ with $\mu_k \in \mathcal{P}_1(\Omega)$ for all $k = 1, \dots, N$ and $\mu_k \preceq \mu_{k+1}$ for all $k = 1, \dots, N-1$. The measures that we will consider are either fully discrete or admit a Lipschitz continuous density with respect to the Lebesque measure. For the approximating method we will also need the assumption that there exists $\theta > 1$ and $M_\theta < \infty$ such that

$$\int\limits_{\mathbb{R}^d} |x|^\theta \mu_N(\mathrm{d}x) \le M_\theta.$$

Because of the convex ordering this also means that $\int_{\mathbb{R}^d} |x|^\theta \mu_k(\mathrm{d}x) \le M_\theta$ for all $k = 1, \dots, N$. The approximating method follows the following steps

1. *Truncation:* First we need to bound the support of each μ_k. Let $R > 0$ and define

$$B_R := \{x = (x_1, \dots, x_d) \in \mathbb{R}^d \,|\, |x_i| \le R, \text{ for } i = 1, \dots, d\}.$$

Take R such that $\mu(B_R) > 0$. We truncate μ_k as follows

$$\mu_{n,R}(\mathrm{d}x) := 1_{B_R}(x)\mu(\mathrm{d}x) + \mu(\mathbb{R}^d \setminus B_R)\delta_0(\mathrm{d}x).$$

One can show that

$$\mathcal{W}_1(\mu_{k,R}, \mu_k) \le \frac{M_\theta}{R^{\theta-1}}.$$

In particular, this means that $\lim\limits_{R\to\infty} \mathcal{W}_1(\mu_{k,R}, \mu_k) = 0$.

2. *Discretization*. Denote by $\Omega_n \subseteq \mathbb{R}^d$ the countable subspace consisting of elements $\frac{q}{n}$ for all $q = (q_1, \dots, q_d) \in \mathbb{Z}^d$. For each $q \in \mathbb{Z}^d$, we denote

$$Z\left(\frac{q}{n}\right) := \{x = (x_1, \dots, x_d) \in \mathbb{R}^d \,|\, \lfloor n x_i \rfloor = q_i \text{ for } i = 1, \dots, d\}.$$

The discretized probability measure for μ_k, supported on Ω_n is then defined as

$$\mu_k^{(n)}\left[\left\{\frac{q}{n}\right\}\right] = \mu_k\left[(Z\left(\frac{q}{n}\right)\right].$$

For this discretizing method, one can show that

$$\mathcal{W}_1(\mu_k^{(n)}, \mu_k) \le \frac{d}{n}.$$

If we have a measure that admits a density, then $\mu_k(\mathrm{d}x) = \rho_k(x)\mathrm{d}x$ for some function ρ_k and define $\mu_{k,R}^{(n)}$ as

$$\mu_k^{(n)}\left[\left\{\frac{q}{n}\right\}\right] := \int\limits_{Z(\frac{q}{n})} \rho(x)\mathrm{d}x \quad 0 \ne \frac{q}{n} \in B_R, \tag{5.3}$$

$$\mu_k^{(n)}[\{0\}] := 1 - \sum_{\frac{q}{n} \ne 0} \mu_k^{(n)}\left[\left\{\frac{q}{n}\right\}\right].$$

The calculation of the integrals in (5.3) can be costly and inefficient. In that case we can discretize the measure as follows

$$\mu_k^{(n)}\left[\left\{\frac{q}{n}\right\}\right] := \frac{\rho(x_{\frac{q}{n}})}{n^d} \quad 0 \neq \frac{q}{n} \in B_R, \tag{5.4}$$

$$\mu_k^{(n)}\left[\{0\}\right] := 1 - \sum_{\frac{q}{n}\neq 0} \mu_k^{(n)}\left[\left\{\frac{q}{n}\right\}\right]. \tag{5.5}$$

The values $x_{\frac{q}{n}}$ are values in $Z(\frac{q}{n})$ and should be chosen such that the sum in (5.5) does not exceed 1.

3. *Optimal parameters*. We need to set all parameters in a suitable ways so that the assumptions of Theorem 5.2.3 are satsified. Let $\gamma > 1$ such that $\frac{1}{\theta} + \frac{1}{\gamma} = 1$ and set

$$R_n := \left(\frac{\theta M_\theta n}{\gamma d}\right)^{\frac{1}{\theta-1}}. \tag{5.6}$$

If we use the approximation given in (5.4) and (5.5) we need to be a bit more careful, as we are using the densities to approximate the measures. As we assume that the densities are Lipschitz continuous with Lipschitz constant L, we set the bound to

$$R_n = \left\lfloor \left(\frac{n(\theta-1)M_\theta}{L}\right)^{\frac{1}{\theta+1}} \right\rfloor. \tag{5.7}$$

It can be shown that this R_n leads to the sharpest inequality in the following bound,

$$\mathcal{W}_1(\mu_{k,R}^{(n)}, \mu_k) \leq \frac{d}{n} + \frac{M_\theta}{R^{\theta-1}}.$$

Additionally, all assumptions of Theorem 5.2.3 will be satisfied, which yields

$$\lim_{n\to\infty} V_M^{\epsilon_n}(\vec{\mu}^n) \to V_M^0(\vec{\mu}) \text{ for } \vec{\mu}^n = (\mu_{1,R_n}^{(n)}, \dots, \mu_{N,R_n}^{(n)}).$$

In the case where we are not working with densities, we can set

$$\epsilon_n = \frac{N\gamma d}{n}$$

and for the case Lipschitz densities we can use

$$\epsilon_n = N\left(\frac{1}{n} + \frac{2R^2L}{n}\right) \qquad R \in \mathbb{R} \text{ such that } \operatorname{supp}(\mu) \subseteq B_R, \tag{5.8}$$

$$\epsilon_n = N\left(\frac{1}{n} + \frac{M_\theta}{R_n^{\theta-1}} + \frac{2R_n^2L}{n} + \frac{4M_\theta}{R_n^{\theta-1}}\right) \qquad \text{unbounded support }. \tag{5.9}$$

Once we have create our discretized measures, we can rewrite the relaxed MOT problem as an LP problem. This LP problems is described as follows. The discretized measure $\vec{\mu}^n = (\mu_1^n, \dots, \mu_N^n)$ consists of discrete measures of the form

$$\mu_k^n(\mathrm{d}x) = \sum_{i_k\in I_k} \alpha_{i_k}^k \delta_{x_{i_k}^k}(\mathrm{d}x),$$

where $I_k = \{1, \ldots, n(k)\}$ labels the support of μ_n^k. Denote by $p = (p_{i_1,\ldots,i_N})_{i_1 \in I_1,\ldots,i_N \in I_N}$ the elements of $\mathbb{R}^D_{\geq 0}$ with $D := \prod_{k=1}^N n(k)$. The LP problem is then given by

$$
\begin{aligned}
&\max_{p \in \mathbb{R}^D_{\geq 0}} \sum_{i_1,\ldots,i_n} p_{i_1,\ldots,i_N} c(x^1_{i_1}, \ldots, x^N_{i_N}) \text{ such that} \\
&\sum_{i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_N} p_{i_1,\ldots,i_N} = \alpha^k_{i_k}, \quad \text{for } i_k \in I_k \text{ and } k = 1, \ldots, N, \\
&\sum_{i_1,\ldots,i_k} \left| \sum_{i_{k+1},\ldots,i_N} p_{i_1,\ldots,i_N} (x^{k+1}_{i_{k+1}} - x^k_{i_k}) \right| \leq \epsilon_n, \quad \text{for } k = 1, \ldots, N-1.
\end{aligned}
\tag{5.10}
$$

Strictly speaking, this is not an LP problem, as the third line is not a linear bound. However, we can add dummy variables $\delta^k = \left(\delta^k_{i_1,\ldots,i_k,j}\right)_{i_1 \in I_1,\ldots,i_k \in I_k, j \in J} \in \mathbb{R}^{D_k}_{\geq 0}$ with $J = \{1, \ldots, d\}$ and $D_k := d \prod_{r=1}^k n(r)$. We will then maximize over p as well as δ^k. The δ^k will be bounded by ϵ_n and we can then use the δ^k to bound the sum in the third line of (5.10). The third line in (5.10) now becomes

$$
\begin{aligned}
-\delta^k_{i_1,\ldots,i_k,j} \leq \sum_{i_{k+1},\ldots,i_N} p_{i_1,\ldots,i_N} (x^{k+1}_{i_{k+1},j} - x^k_{i_k,j}) \leq \delta^k_{i_1,\ldots,i_k,j} \text{ for } i_k \in I_k, j \in J \text{ and } k = 1, \ldots, N, \\
\sum_{i_1,\ldots,i_k,j}^{k} \delta^k_{i_1,\ldots,i_k,j} \leq \epsilon_n \text{ for } k = 1, \ldots, N-1.
\end{aligned}
$$

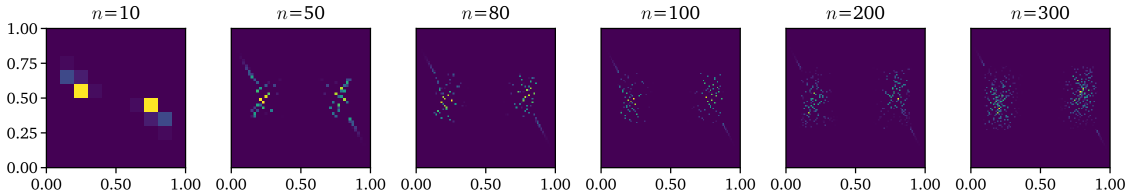
We refer to Corollary 2.4 in [28] for the details as to why this is the correct LP formulation.

5.3. Examples

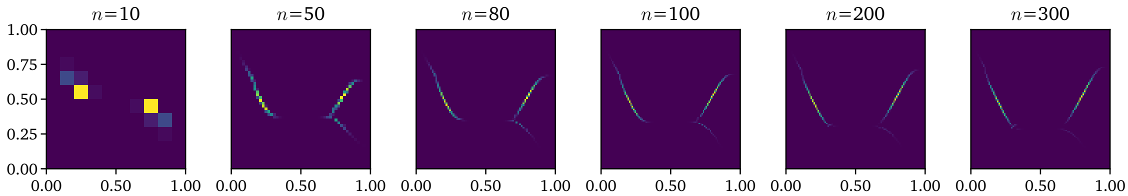
The full method can be quite daunting. We will now discuss some simpler examples, which should clarify how this computational method works in practise. The first example functions as yet another reason as to why the MOT problem is fundamentally different from the OT problem. The second example will serve as a more realistic example. We will look at the bounds of a Lookback option with log-normally distributed risky assets. In both examples, we will set $N = 2$ and $d = 1$. The LP problems that need to be solved in the following examples are solved using the *Gurobi* interface in *Python*.

5.3.1. Squared and exponential cost function

In this example we will look at the MOT problem for the cost function $c(x, y) = (x - y)^2$. This is an important cost function in the classical OT problem, as it represents the Euclidean distance. The resulting OT problem can be used to model many phenomena in physics, because that OT problem can be seen as some sort of energy functional that needs to be minimized. However,



(a) Optimal martingale couplings for $c(x, y) = (x - y)^2$.



(b) Optimal martingale couplings for $c(x, y) = \mathrm{e}^{x-y}$.

Figure 5.1.: The heatmaps indicate where the optimal martingale coupling between μ and ν is concentrated, where the y-axis represents μ and the x-axis represents ν. The number n indicates the number of discretization points. The lines are artificially made thicker for clarity.

considering this cost function in the martingale case is trivial. Let $\pi \in \mathcal{M}(\mu, \nu)$, then

$$\begin{aligned}\int\limits_{\mathbb{R}\times\mathbb{R}} (x-y)^2 \pi(\mathrm{d}x, \mathrm{d}y) &= \int\limits_{\mathbb{R}} x^2 \mu(\mathrm{d}x) - \int\limits_{\mathbb{R}\times\mathbb{R}} 2xy\pi(\mathrm{d}x, \mathrm{d}y) + \int\limits_{\mathbb{R}} y^2 \nu(\mathrm{d}y) \\ &= \int\limits_{\mathbb{R}} x^2 \mu(\mathrm{d}x) - \int\limits_{\mathbb{R}} x \int\limits_{\mathbb{R}} y \pi_x(\mathrm{d}y)\mu(\mathrm{d}x) + \int\limits_{\mathbb{R}} y^2 \nu(\mathrm{d}y) \\ &= \int\limits_{\mathbb{R}} x^2 \mu(\mathrm{d}x) - 2\int\limits_{\mathbb{R}} x^2 \mu(\mathrm{d}x) + \int\limits_{\mathbb{R}} y^2 \nu(\mathrm{d}y) \\ &= \int\limits_{\mathbb{R}} y^2 \nu(\mathrm{d}y) - \int\limits_{\mathbb{R}} x^2 \mu(\mathrm{d}x)\end{aligned}$$

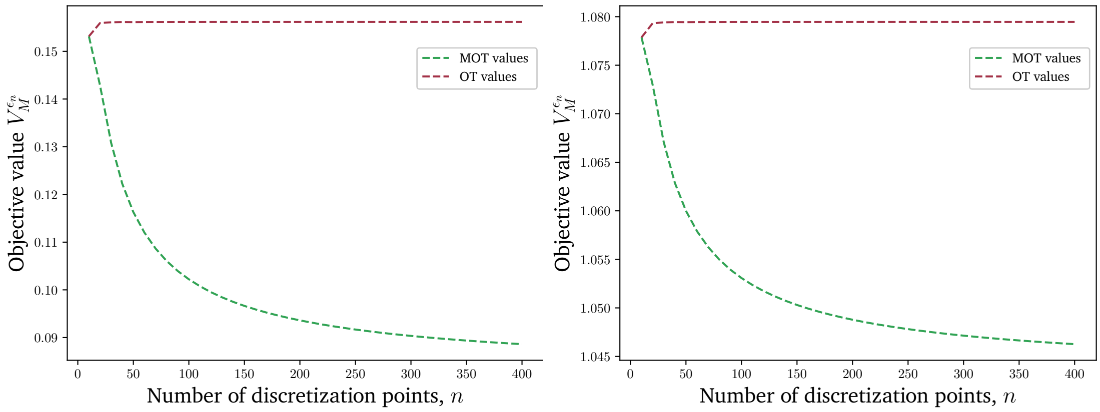
So, the integral only depends on μ and ν and not on what martingale coupling we choose. We will see this in the optimal couplings calculated with the method of Guo and Obłój, as they will appear randomly selected.

Let $\mu(\mathrm{d}x) = \rho(x)\mathrm{d}x$ and $\nu(\mathrm{d}y) = \sigma(y)\mathrm{d}y$. Let φ be the density of a standard normal distribution, the densities $\rho(x)$ and $\sigma(y)$ are given by

$$\begin{aligned}\rho(x) &= \frac{1}{C_\rho}\varphi\left(10(x - \frac{1}{2})\right)\Big|_{[0,1]}, \\ \sigma(y) &= \frac{1}{C_\sigma}\left(\frac{1}{2}\varphi\left(20(y - \frac{1}{5})\right)\Big|_{[0,1]} + \frac{1}{2}\varphi\left(20(y - \frac{4}{5})\right)\Big|_{[0,1]}\right).\end{aligned}$$

The density ρ is a truncated normal density on the domain $[0, 1]$ with mean $\frac{1}{2}$ and standard deviation $\frac{1}{10}$. The density σ is a bimodal density consisting of two truncated normal densities with means $\frac{1}{5}$ and $\frac{4}{5}$, and standard deviations $\frac{1}{20}$. Likewise, the densities are restricted to $[0, 1]$. The constants C_ρ and C_σ ensure that the densities integrate to one.

It can be verified that $\mu \preceq \nu$ and it is clear that the supports $\operatorname{supp}(\mu)$ and $\operatorname{supp}(\nu)$ are both bounded. For each n we apply the discretization step for the set $\{\frac{i}{n} \,|\, 0 \leq i < n\}$ to get the



(a) Optimal values for $c(x, y) = (x - y)^2$. (b) Optimal values for $c(x, y) = \mathrm{e}^{x-y}$.

Figure 5.2.: The objective value $V_M^{\epsilon_n}(\mu^{(n)}, \nu^{(n)})$.

discretized versions $\mu^{(n)}$ and $\nu^{(n)}$ as defined in (5.4) and (5.5). It can be shown that ρ and σ are Lipschitz with $L = 12$. Putting all this together gives $\epsilon_n = \frac{2+4L}{n}$. The LP problem that needs to be solved is given by

$$\begin{aligned} \max_{p \in \mathbb{R}_{\geq 0}^{n \times n}, \delta \in \mathbb{R}_{\geq 0}^{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} p_{i,j} c(x_i, y_i) \text{ such that } \quad & \sum_{j=1}^{n} p_{i,j} = \alpha_i, \quad \text{for } i = 1, \ldots, n, \\ & \sum_{i=1}^{n} p_{i,j} = \beta_j, \quad \text{for } j = 1, \ldots, n, \\ -\delta_i \leq & \sum_{j=1}^{n} p_{i,j}(y_j - x_i) \leq \delta_i \quad \text{for } i = 1, \ldots, n, \\ & \sum_{l=1}^{n} \delta_l \leq \epsilon_n. \end{aligned} \tag{5.11}$$

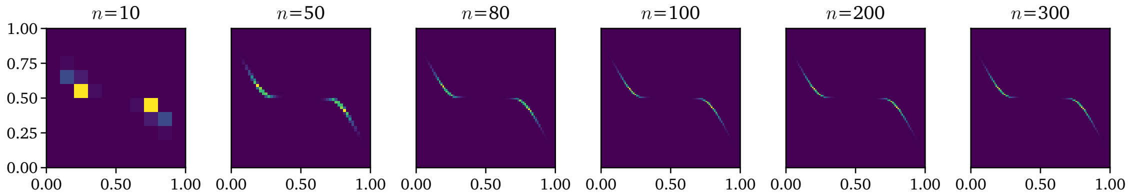
The coefficients α_i and β_j are given by

$$\alpha_i := \frac{\rho(\frac{i}{n})}{n} \quad i = 1, \ldots, n, \qquad \beta_j := \frac{\sigma(\frac{j}{n})}{n} \quad i = 1, \ldots, n,$$
$$\alpha_0 := 1 - \sum_{i=1}^{n} \alpha_i, \qquad \beta_0 := 1 - \sum_{j=1}^{n} \beta_j.$$

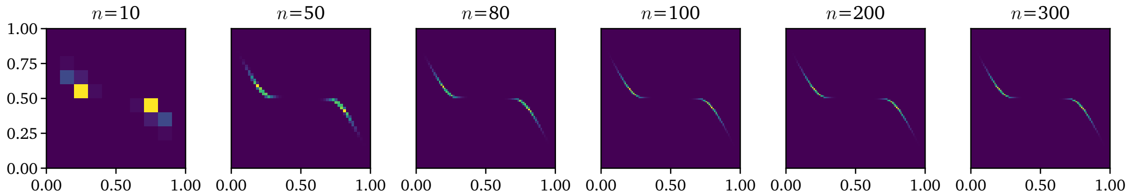
The resulting couplings found by solving the equivalent LP problem are shown in Figure 5.1a and the optimal values are shown in 5.2a. Although the couplings look random, there is still a clear descent in the optimal values, when the number of discretization steps is increased. This is most likely explained by the fact that we are are solving the LP problem of the relaxed version, for which not all couplings are optimal, as the martingale property can be violated.

We can get picture with a nicer structure, if we use the cost function $c(x, y) = \mathrm{e}^{x-y}$. All the other steps remain the same. The results are shown in Figures 5.1b and 5.2b. We can see two clear lines on which the martingale coupling is concentrated.

We can also compare the optimal martingale coupling with the optimal coupling from the classical OT problem. The LP that has to be solved is the problem shown in (5.11), with the last



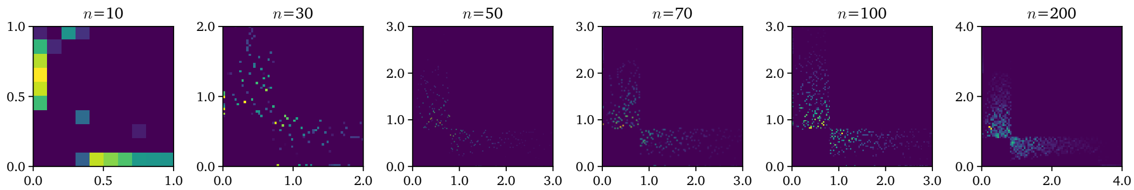
(a) Optimal couplings for $c(x, y) = (x - y)^2$.



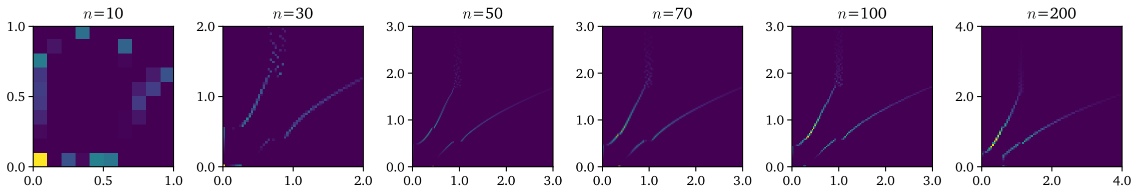
(b) Optimal couplings for $c(x, y) = \mathrm{e}^{x-y}$.

Figure 5.3.: The heatmaps indicate where the optimal coupling between μ and ν is concentrated, where the y-axis represents μ and the x-axis represents ν. The lines are artificially made thicker for clarity.

two lines removed. The optimal values are also shown in the Figures 5.2a and 5.2b. The optimal couplings are shown in Figures 5.3a and 5.3b. Now, we see that both cost functions result into two nice couplings concentrated on one curve. The optimal coupling for squared cost function also looks very similar to the exponential cost function.



(a) Optimal martingale couplings when using the theoretical bounds for ϵ_n.

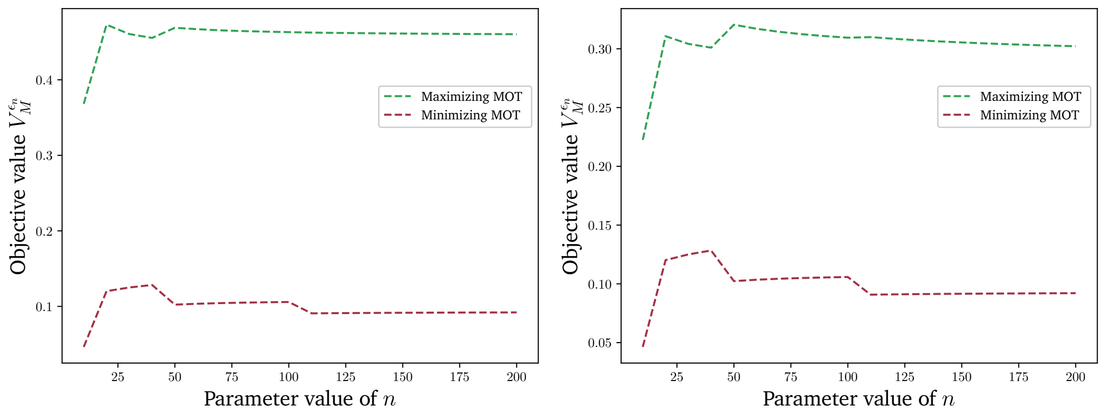


(b) Optimal martingale couplings when using the artificially decreased bounds for ϵ_n.

Figure 5.4.: Heatmaps of the optimal coupling for $c(x, y) = \max\{x, y\} - y$, where the y-axis represents μ and the x-axis represents ν. The lines are artificially made thicker for clarity.

5.3.2. Lookback option

The MOT problem was of course motivated by the robust finance setting for option pricing. With that rationale, we will look at a final example. Inspired by Example 3.8 in [28], we will consider a risky asset on two dates. The price of this asset will be denoted by S_0 and S_T. Inspired by the Black-Scholes model, we adopt a log-normal distribution for the two random variables S_0 and



(a) The optimal values using the theoretical values of ϵ_n. (b) The optimal values using artificially decreased values of ϵ_n.

Figure 5.5.: The objective value $V_M^{\epsilon_n}(\mu^{(n)}, \nu^{(n)})$ with cost function $c(x,y) = \max\{x,y\} - y$.

S_T. The densities of these log-normal distributions, $\mu(\mathrm{d}x) = \rho(x)\mathrm{d}x$ and $\nu(\mathrm{d}y) = \sigma(y)\mathrm{d}y$, are given for $x, y > 0$ by

$$\begin{aligned}
\rho(x) &= \frac{\sqrt{2}}{x\sqrt{\pi}} \mathrm{e}^{-2(\log(x)+\frac{1}{8})^2}, \\
\sigma(y) &= \frac{1}{y\sqrt{\pi}} \mathrm{e}^{-(\log(y)+\frac{1}{4})^2}.
\end{aligned}$$

Notice that the supports are now unbounded and $\mathrm{supp}(\mu) = \mathrm{supp}(\nu)$. It is well known that all moments exist for the log-normal distribution. By the convex ordering, we can bound all the moments of ρ by the moments of σ, which gives us for $\theta > 1$

$$\int_{\mathbb{R}} x^\theta \rho(x)\mathrm{d}x \leq \int_{\mathbb{R}} x^\theta \sigma(y)\mathrm{d}y = \mathrm{e}^{\frac{1}{4}\theta(\theta-1)} =: M_\theta.$$

For our application we set $\theta = 2$. We bound the domain using the optimal R_n given in (5.7) and take the discretization point $\{\frac{i}{n} : 0 \leq i < nR_n, i \in \mathbb{N}\}$. The descretized measures $\mu^{(n)}$ and $\nu^{(n)}$ are again obtained by utilising (5.4) and (5.5). Finally, we find that both ρ and σ are Lipschitz with constant $L = 12$ and set ϵ_n as described in (5.9). The LP problem that needs to be solved is given by

$$\begin{aligned}
\max_{p \in \mathbb{R}_{\geq 0}^{nR_n \times nR_n}, \delta \in \mathbb{R}_{\geq 0}^{nR_n}} \sum_{i=1}^{nR_n}\sum_{j=1}^{nR_n} p_{i,j} c(x_i, y_i) \text{ such that } & \sum_{j=1}^{nR_n} p_{i,j} = \alpha_i, \quad \text{for } i = 1, \dots, R_n, \\
& \sum_{i=1}^{nR_n} p_{i,j} = \beta_j, \quad \text{for } j = 1, \dots, nR_n, \\
-\delta_i \leq & \sum_{j=1}^{nR_n} p_{i,j}(y_j - x_i) \leq \delta_i \quad \text{for } i = 1, \dots, nR_n, \\
& \sum_{l=1}^{nR_n} \delta_l \leq \epsilon_n.
\end{aligned} \tag{5.12}$$

The coefficients α_i and β_j are given by

$$\alpha_i := \frac{\rho(\frac{i}{n})}{n} \quad i = 1, \ldots, nR_n, \qquad \beta_j := \frac{\sigma(\frac{j}{n})}{n} \quad i = 1, \ldots, nR_n,$$
$$\alpha_0 := 1 - \sum_{i=1}^{nR_n} \alpha_i, \qquad \beta_0 := 1 - \sum_{j=1}^{nR_n} \beta_j.$$

The optimal couplings are shown in Figure 5.4a with optimal values in Figure 5.5a. Unfortunately, the ϵ_n that we should take from the theory remained too large and the relaxed problem did not result in any good martingale couplings. The values of ϵ_n ranged from 10 to almost 100. If we artificially decrease the values of ϵ_n by a factor of 50, we do get nicer pictures, as seen in Figures 5.4b and 5.5b. This indicates that further research can be done to make the theoretical bounds sharper for more efficient methods. Note the increasing support of the optimal couplings by $R_n \to \infty$. In Figures 5.5a and 5.5b we also show the optimal values when we look at the minimizing MOT problem. From these values, we can get a feeling of how large the bounds can get when actually using martingale optimal transport to get bounds on prices of options.

Conclusion

The goal of this thesis was to describe the martingale optimal transport problem and discuss the recently proven stability results given by Backhoff-Veraguas and Pammer in the one dimensional case. Additionally, a numerical scheme for calculating the optimal value or optimizing martingale coupling was given and two examples were presented.

To be able to prove stability for the martingale optimal transport problem it was needed to introduce the notion of $(c, \mathcal{F}_M)$-monotonicity and martingale C-monotonicity. These definitions allowed us to find necessary properties of an optimizing martingale coupling. Moreover, we showed that $(c, \mathcal{F}_M)$-monotonicity is actually a sufficient condition for an optimizing martingale coupling in the one dimensional case. We were able to switch between $(c, \mathcal{F}_M)$-monotonicity and martingale C-monotonicity by embedding the space $\mathbb{R} \times \mathbb{R}$ into the space $\mathbb{R} \times \mathcal{P}(\mathbb{R})$, through the map J. A consequence of being able to switch between $(c, \mathcal{F}_M)$-monotonicity and martingale C-monotonicity was that the two definitions are actually equivalent. Finalle, We showed that the map J preserves the relative compactness of sets, which allowed us to carry martingale C-monotonicity to the limit of a sequence of martingale C-monotone probability measures. Combining all these result allowed us to prove the stability results of Backhoff-Veraguas and Pammer.

The techniques introduced to prove stability of the martingale optimal transport problem are all relatively new and unknown. Most notably, the introduction of a more general monotonicity principle and the map J, were introduced only recently and vital to the proof. Chapters 3 and 4 served as a short introduction to these new concepts. In Chapter 3 we showed that we could cast the martingale optimal transport into this general framework, which allowed us to derive the $(c, \mathcal{F}_M)$-monotonicity principle. Furthermore, the proof of sufficiency of $(c, \mathcal{F}_M)$-monotonicity was shown. In chapter 4 we gave a motivating example for usefulness of the map J and showed that J preserves relative compactness. Any remaining proofs were displayed in Chapter 5.

The final Chapter was dedicated to the description of the numerical method developed by Guo and Obłój. We showed how their method works, and provided their convergence results. At the end we showcased the effectiveness of the method by discussing two examples. The first example served as another argument for the fundamental difference between the classical OT problem and the MOT problem. The second example illustrated the use of MOT in the field of robust finance.

The question of stability of the martingale optimal transport problem in one dimension has now been resolved. However, the question of stability in higher dimensions is still open. In January 2021, a counter-example in dimension 2 or higher was shown to exist, casting doubts on a stability result in higher dimensions. However, it could still be the case that there are assumptions we can make to ensure stability in higher dimensions. The search of these assumption should actively be pursued and could lead to valuable insights into the behaviour of martingales in higher dimensions.

Popular summary

Financial institutions have to calculate the right prices of various financial products daily. One of the most popular variant of a financial product is an option. How to determine a correct price for such financial products is a difficult problem. In Mathematics, we can formulate this problem for specific options as follows. The pay-out of these kinds of options is determined by the value of a stock at some time $t = 0$ and the value of that same stock at some time $t = T$ in the future. The value of the stock at these times will be denote by S_0 and S_T. These two variables are random, as it is impossible to know the true value of a stock. The distributions of S_0 and S_T will be denoted by μ and ν. The amount of money you will gain from an option, is determined by applying some function g to S_0 and S_T. In general, one will only gain money from an option or you will not receive any money at all, because if the option incurs a loss, you can just not exercise the option. It remains an *option* after all!

Now that we have a mathematical formulation of an option, we want to calculate the price. But what would be a fair price? As buyer of an option you want to make money out if it, but the profit comes out of the sellers pockets. Thus, as seller you want that the price is high enough to cover those pay outs. We are thus in the situation that the buyer wants to price to be low enough, so that they expect to make a profit, while the seller wants the price to be high enough to cover those profits. The only fair price of $g(S_0, S_T)$ for both parties will be a price that ensures that making a profit is just as likely as not making a profit. For mathematicians this means that (S_0, S_T) should be a so called *martingale*. The joint distribution of (S_0, S_T) determines if (S_0, S_T) is a martingale or not. If this is the case, then we call the joint distribution a *martingale distribution*. Additionally, this joint martingale distribution should be compatible with the distributions μ and ν. We call this joint martingale distribution π and the fair price is then given by

$$P_{\text{fair}} = \mathbb{E}_\pi[g(S_0, S_T)].$$

The price is given as an expectation, because the expectation ensures that the payout $g(S_0, S_T)$ is more than the price in half of the cases, and less in the other half of the cases. A buyer will win as often as they lose as a consequence.

The difficulty lies in finding a π satisfies all requirements. Classically, this is done be building a model that can explain how S_T follows from S_O. This model will give us a process $(S_t)_{t=0}^T$, from which we can calculate π.

This method has been very successful in the past and is applied in many cases. However, there are also drawbacks to this approach. Most notably, there is no guarantee that the model that is employed, is even close to the true process governing the price of a stock. If that is the case, then the price can lie far away from the true price. This can be very dangerous for the economy. The research field of *robust finance* tries to find prices for financial products, without postulating an explicit model. One of the methods that was developed for this case is called the *martingale optimal transport* technique. One looks at *all* possible joint martingale distributions that are compatible with μ and ν, in stead of creating a model and then calculating π. We call this set $\mathcal{M}(\mu, \nu)$. We can now look at the lowest possible price and the highest possible price:

$$P_{\text{low}} = \inf_{\pi \in \mathcal{M}(\mu,\nu)} \mathbb{E}_\pi[g(S_0, S_T)], \quad P_{\text{high}} = \sup_{\pi \in \mathcal{M}(\mu,\nu)} \mathbb{E}_\pi[g(S_0, S_T)].$$

The true price is still not known, but we can put bounds on what the eventual price should be. The price can be chosen somewhere between these bounds, depending on how much risk a financial institution wants to take (or is allowed to take).

If we consider P_{low}, then we can show that under some mild assumptions, that there exists at least one distribution π that actually attains the infimum, which means that

$$P_{\text{low}} = \inf_{\pi \in \mathcal{M}(\mu,\nu)} \mathbb{E}_\pi[g(S_0, S_T)] = \mathbb{E}_{\pi^*}[g(S_0, S_T)]$$

for some $\pi^* \in \mathcal{M}(\mu, \nu)$. For a long time, it was unsure if the mappings

$$\begin{aligned} &(\mu, \nu) \mapsto \pi^*, \\ &(\mu, \nu) \mapsto P_{\text{low}}, \end{aligned} \tag{5.13}$$

were continuous. This is important especially for numerical methods, because proving convergence of numerical methods that calculate these mappings becomes a lot easier. These numerical methods are used in financial institutions. So, making sure that these algorithms actually work is of paramount importance.

Luckily, two proofs appeared in 2019 by Backhoff-Veraguas and Pammer, and Wiesel, who showed continuity of the two mappings in (5.13). The proof of Backhoff-Veraguas and Pammer will be discussed in great detail in this thesis. In Chapter 2 we will introduce the results proven by Backhoff-Veraguas and Pammer and show the necessary methods used in the proofs. We will elaborate on these techniques in Chapters 3 and 4. The final chapter will show a recently developed numerical method for the martingale optimal transport and we will calculate some examples.

A. Measure theory results

Theorem A.1.1 [Functional monotone class theorem]. *Let $(\Omega, \mathcal{F}, \mu)$ a probability space and $\mathcal{K} \subseteq B(\Omega)$, be closed under multiplication. Let $\mathcal{H}$ be a vector subspace of $B(\Omega)$ satisfying*

- *$\mathcal{K} \subseteq \mathcal{H}$.*
- *$1 \in \mathcal{H}$.*
- *$\mathcal{H}_+$ is closed under bounded increasing limits. Here, $\mathcal{H}_+$ indicates all non-negative functions in $\mathcal{H}$.*

Then, $\sigma(\mathcal{K})_b \subseteq \mathcal{H}$

Proof. See Theorem 1 in [24]. □

Theorem A.1.2 [Monotone convergence theorem]. *Let $f, f_1, f_2, \ldots$ be measurable functions on $(\Omega, \mathcal{A}, \mu)$ with $0 \leq f_n \uparrow f$. Then $\int_\Omega f_n \, d\mu \uparrow \int_\Omega f \, d\mu$*

Theorem A.1.3 [Markov's inequality]. *Let $(\Omega, \mathcal{A}, \mu)$ be a measurable space and $f : \omega \to \mathbb{R}$ a Borel measurable function. If $\epsilon > 0$, then we have*

$$\mu\left(\{x \in \mathcal{X} \mid |f(x)| > \epsilon\}\right) \leq \frac{1}{\epsilon} \int_{\mathcal{X}} |f(x)| \mu(\mathrm{d}x).$$

Proof. See the proof of Lemma 3.1, page 40, in [32]. □

Proof. See Theorem 1.19, page 11, in [32]. □

Theorem A.1.4 [Portmanteau]. *Let E be a metric space and let $\mu, \mu_1, \ldots \in \mathcal{P}(E)$. The following are equivalent:*

(i) $\mu \xrightarrow{w} \mu$;
(ii) $\int_E f \, d\mu_k \to \int_E f \, d\mu$ *for all* $f \in C_b(C)$;
(iii) $\limsup_{k \in \mathbb{N}} \int_E f \, d\mu_k \leq \int_E f \, d\mu$ *for all upper semi-continuous functions f, that are bounded from above;*
(iv) $\liminf_{k \in \mathbb{N}} \int_E f \, d\mu_k \geq \int_E f \, d\mu$ *for all lower semi-continuous functions f, that are bounded from below;*
(v) $\limsup_{k \in \mathbb{N}} \mu_k(C) \leq \mu(C)$ *for all closed C;*
(vi) $\liminf_{k \in \mathbb{N}} \mu_k(O) \geq \mu(O)$ *for all open O;*
(vii) $\lim_{k \to \infty} \mu_k(A) = \mu(A)$ *for all continuity sets A of μ.*

Proof. See Theorem 13.16, page 253, in [35]. □

Definition A.1.5 [Thightness of measures]. Let $\mathcal{X}$ be a Polish space and $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$ a set of probability measures. The set $\mathcal{A}$ is called tight if, for any $\epsilon > 0$, there is a compact subset $K_\epsilon \subseteq \mathcal{X}$ such that

$$\sup_{\mu \in \mathcal{A}} \mu(\mathcal{X} \setminus K_\epsilon) < \epsilon.$$

Theorem A.1.6 [Prokhorov]. *For any sequence of random elements $\{X_i\}_{i \in \mathbb{N}}$ in a metric space S, tightness implies relative compactness in distribution, and the two are equivalent if S is a metric Polish space.*

Proof. See the proof of Theorem 14.3, page 254, in [32]. □

Theorem A.1.7 [Borel σ algebra and C_b]**.** *Let $(\mathcal{X}, d)$ be a metric space with topology generated by open balls with respect to the metric d. We then have that $\mathcal{B}(\mathcal{X}) = \sigma(C_b(\mathcal{X}))$.*

Proof. "$\subseteq$", Take A open, as the set of open balls form a basis for this topology we can consider without loss of generality that A is of the form

$$A = B_\epsilon(x) = \{y \in \mathcal{X} \mid d(x, y) < \epsilon\} = d_x((-\infty, \epsilon))^{-1}.$$

Here we denote $d_x(y) = d(x, y)$, which is a continuous function. It is not bounded however. To circumvent this issue we can take a $K > \epsilon$ and consider the adjusted function

$$\tilde{d}_x(y) = d_x(y)1_{\{d_x(y)\leq K\}}(y) + K1_{\{d_x(y)>K\}}(y).$$

This is a continuous and bounded function, because $\tilde{d}_x(y) \leq K$, with

$$\tilde{d}_x((-\infty, \epsilon))^{-1} = d_x((-\infty, \epsilon))^{-1} = B_\epsilon(x).$$

This shows that at least the generators of $\mathcal{B}(\mathcal{X})$ are in $\sigma(C_b(\mathcal{X}))$, but then the whole σ algebra must be a part of it as well

"$\supseteq$", The generators of $\sigma(C_b(\mathcal{X}))$ are precisely the sets of the form $f^{-1}(B)$ with B open in $\mathcal{R}$ for some continuous function f. This also means that $f^{-1}(B)$ is also open, but in $\mathcal{X}$. Thus we have again that the generators of $\sigma(C_b(\mathcal{K}))$are a part of $\mathcal{B}(\mathcal{X})$. So we have that $\mathcal{B}(\mathcal{X}) \supseteq \sigma(C_b(\mathcal{X}))$. □

Theorem A.1.8 [Glivenko-Cantelli]**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and assume that $\{X_k\}_{k\in\mathbb{N}}$ are i.i.d. random variables with distribution function $F(t)$. Define the empirical distribution function as*

$$F_n(t) = \frac{1}{n}\sum_{k=1}^{n} 1_{(-\infty, X_k]}(t).$$

Then, $\lim_{n\to\infty} \sup_{t\in\mathbb{R}} |F_n(t) - F(t)| = 0$, $\mathbb{P}$ almost surely.

Proof. See the proof of Theorem 19.1 on page 266 in [49]. □

Definition A.1.9 [Analytic set]**.** Let $\mathcal{X}$ be a Polish space, a set $A \subseteq \mathcal{X}$ is called analytic if it is the image of a continuous function $f : \mathcal{Y} \to A$ for some Polish space $\mathcal{Y}$.

Theorem A.1.10. *Let $\mathcal{X}$ be a Polish space, for a set $A \subseteq \mathcal{X}$ the following are equivalent:*

1. *A is analytic.*
2. *There is a Polish space $\mathcal{Y}$ and a Borel set $B \subseteq \mathcal{X} \times \mathcal{Y}$ such that A is the projection of B.*

Proof. See Proposition 7.39, page 165, in [16]. □

Theorem A.1.11. *Let $\mathcal{X}$ be a Borel space, then every analytic subset of $\mathcal{X}$ is universally measurable, i.e. an analytic set is measurable with respect to every Borel probability measure. In particular, every analytic set can be written as the disjoint union of a Borel set and a null set for any probability measure.*

Proof. See Corollary 7.42, page 169 in [16]. □

Definition A.1.12 [Regular disintegration]. Let $(T, \mathcal{B}(T)), (S, \mathcal{B})$ be two measurable spaces with probability measures $\mu \in \mathcal{P}(T)$ and $\nu \in \mathcal{P}(S)$. If $\pi \in \Pi(\mu, \nu)$, then a *regular disintegration* with respect to μ is a collection of probability measures $\{\pi_x\}_{x \in T}$, such that for every Borel measurable function on $T \times S$ with

$$\int_{T \times S} |f(t,s)| \pi(\mathrm{d}t, \mathrm{d}s) < \infty$$

we have that

$$\int_{T \times S} f(t,s) \pi(\mathrm{d}t, \mathrm{d}s) = \int_{T \times S} f(t,s) \pi_x(\mathrm{d}s) \mu(\mathrm{d}t).$$

Theorem A.1.13 [Existence Conditional Kernel]. *Fix a Borel space $(S, \mathcal{B}(S))$ and a measurable space $(T, \mathcal{B}(T))$, and let X and Y be random elements in T and S defined on the same probability space, respectively. Then there exists a probability kernel μ from T to S satisfying $\mathbb{P}(Y \in \cdot \,|\, X) = \mu_X(\cdot)$ a.s., and μ is unique a.e. $\mathbb{P} \circ X^{-1}$.*

Proof. See Theorem 5.4, page 84 in [32]. □

Theorem A.1.14 [Existence regular disintegration]. *Let $(\Omega, \mathbb{P}, \mathcal{A})$ be a probability space and fix two measurable spaces $(T, \mathcal{B}(T)), (S, \mathcal{B}(S))$ and a sub σ algebra $\mathcal{F} \subseteq \mathcal{A}$, and a random variable Y in S such that $\mathbb{P}(Y \in \cdot \,|\, \mathcal{F}) = \nu$ for some probability measure ν. Further consider an $\mathcal{F}$-measurable random element X in T and a measurable function f on $S \times T$ with $\mathbb{E}[|f(Y,X)|] < \infty$. Then*

$$\mathbb{E}[f(Y,X) \,|\, \mathcal{F}] = \int f(s,X) \nu(\mathrm{d}s) \quad \textit{a.s.}$$

In particular, if $\mathcal{F} = \sigma(X)$ and $\mathbb{P}(Y \in \cdot \,|\, X) = \mu_X(\cdot)$ for some probability kernel μ, then

$$\mathbb{E}[f(Y,X) \,|\, X] = \int f(s,X) \mu_X(\mathrm{d}s) \quad \textit{a.s.}.$$

Note that this implies that $\mu_X(\cdot)$ is a regular disintegration.

Proof. See Theorem 5.4, page 85, in [32]. □

Lemma A.1.15 [Measurable version of Lusin's theorem]. *Let $\Gamma \subseteq \mathcal{X} \times \mathcal{P}(\mathcal{Y})$ be analytic, and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\}$ be Borel measurable. Then there exists an analytic set $\hat{\Gamma} \subseteq \mathcal{X} \times \mathcal{Y}$ with the following properties:*

- *(i) For any $(x,p) \in \Gamma$ we have that p is concentrated on the fibre $\hat{\Gamma}_x = \{y \in \mathcal{Y} \,|\, (x,y) \in \hat{\Gamma}\}$, i.e., $p(\hat{\Gamma}_x) = 1$.*
- *(ii) For any $(x,y) \in \hat{\Gamma}$ we find $(x,p) \in \Gamma$, such that for any $\epsilon > 0$, we can select a Borel measurable set $K \subseteq \hat{\Gamma}_x$ such that*
  - *(a) $p(K) \geq 1 - \epsilon$,*
  - *(b) c restricted to $\{x\} \times K$ is continuous,*
  - *(c) $y \in \operatorname{supp}(p) \cap K$ and for all continuous $f : K \to \mathbb{R}$, we have*

$$\int_{B_\delta(y) \cap K} \frac{f(z)}{p(B_\delta(y) \cap K)} p(\mathrm{d}z) \to f(y) \quad \textit{for } \delta \downarrow 0. \tag{A.1}$$

Proof. For the full proof we refer to Lemma 6.3 in [5]. The original statement of the theorem is a bit different, but from the proof we can infer that the measure

$$\frac{1}{p(B_\delta(y) \cap K)} p|_{B_\delta(y) \cap K}.$$

converges weakly to δ_y, which shows the last statement of this lemma.

□

Theorem A.1.16 [Jankow-von Neumann selection theorem]. *Let p be a surjective Borel map of a Polish space $\mathcal{X}$ onto a Polish space $\mathcal{Y}$. Then there is a function $s : \mathcal{Y} \to \mathcal{X}$, universally measurable, such that $p \circ s = id$.*

Proof. See the proof of Theorem 13 of Chapter II in [46]. □

B. Analysis results

Theorem B.1.1 [Moore-Osgood]. *Suppose* $\{a_{n,k}\}_{n,k\in\mathbb{N}^2}$ *with* $a_{n,k} \in \mathbb{R}$ *for all* $n, k \in \mathbb{N}$. *If*

$$\lim_{n\to\infty} a_{n,k} = B_k \quad \textit{uniformly in } k$$

and

$$\lim_{k\to\infty} a_{n,k} = A_n$$

then the double limit $\lim\limits_{(n,k)\to\infty} a_{n,k}$ *exists and*

$$\lim_{n\to\infty}\lim_{k\to\infty} a_{n,k} = \lim_{k\to\infty}\lim_{n\to\infty} a_{n,k}.$$

Proof. This is a consequence of Theorem 7.11 on page 149 of [44] □

Definition B.1.2 [Lower convex envelope]. Let $f : S \to \mathbb{R}$ be a lower semi-continuous function, where $S \subseteq \mathbb{R}^d$ is a non empty convex subset. The *lower convex envelope* taken over S, is the function given by

$$\breve{f}(x) = \sup\{g(x) \,|\, g \text{ is convex and } g \leq f \text{ for all } x \in S\}.$$

Proposition B.1.3. *Let* $f : S \to \mathbb{R}$ *be a lower semi-continuous function, where* $S \subseteq \mathbb{R}^d$ *is a non empty convex subset. The lower convex envelope* $\breve{f}$ *has the following properties:*

(i) $\breve{f}$ *is a convex function defined on the set* S.

(ii) $\breve{f}(x) \leq f(x)$ *for all* $x \in S$.

(iii) if h *is any other convex function such that* $h(x) \leq f(x)$ *for all* $x \in S$*, then* $h(x) \leq \breve{f}(x)$ *for all* $x \in S$.

Proof. See page 343 in [56]. □

Bibliography

[1] B. Acciaio et al. “A model-free version of the fundamental theorem of asset pricing and the super-replication theorem”. In: *Mathematical Finance* 26.2 (2016), pp. 233–251.

[2] B. Acciaio et al. “COT-GAN: Generating Sequential Data via Causal Optimal Transport”. In: *Advances in Neural Information Processing Systems 33 pre-proceedings (NeurIPS 2020)* (2020).

[3] M. Arjovsky, S. Chintala, and L. Bottou. “Wasserstein generative adversarial networks”. In: *International conference on machine learning*. PMLR. 2017, pp. 214–223.

[4] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. “Existence, duality, and cyclical monotonicity for weak transport costs”. In: *Calculus of Variations and Partial Differential Equations* 58.6 (2019), pp. 1–28.

[5] J. Backhoff-Veraguas and G. Pammer. “Stability of martingale optimal transport and weak optimal transport”. In: (2020). arXiv: `1904.04171 [math.PR]`.

[6] J. Backhoff-Veraguas et al. “Adapted Wasserstein distances and stability in mathematical finance”. In: *Finance and Stochastics* 24 (2020), pp. 601–632.

[7] J. Backhoff-Veraguas et al. “All adapted topologies are equal”. In: *Probability Theory and Related Fields* 178.3 (Dec. 2020), pp. 1125–1172. ISSN: 1432-2064. DOI: `10.1007/s00440-020-00993-8`. URL: `https://doi.org/10.1007/s00440-020-00993-8`.

[8] D. Baker. “Martingales with specified marginals”. In: *Theses, Université Pierre et Marie Curie-Paris VI* (2012).

[9] R. H. Bartels and G. H. Golub. “The simplex method of linear programming using LU decomposition”. In: *Communications of the ACM* 12.5 (1969), pp. 266–268.

[10] M. Beiglböck, P. Henry-Labordère, and F. Penkner. “Model-independent bounds for option prices—a mass transport approach”. In: *Finance and Stochastics* 17.3 (2013), pp. 477–501.

[11] M. Beiglböck, M. Nutz, N. Touzi, et al. “Complete duality for martingale optimal transport on the line”. In: *The Annals of Probability* 45.5 (2017), pp. 3038–3074.

[12] M. Beiglböck, N. Juillet, et al. “On a problem of optimal transport under marginal martingale constraints”. In: *Annals of Probability* 44.1 (2016), pp. 42–106.

[13] M. Beiglböck et al. “Optimal and better transport plans”. In: *Journal of Functional Analysis* 256.6 (2009), pp. 1907–1927.

[14] M. Beiglboeck and C. Griessler. “A land of monotone plenty”. In: *Annali Della Scuola Normale Superiore Di Pisa-classe Di Scienze* 19 (2019), pp. 109–127.

[15] J.-D. Benamou and Y. Brenier. “A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem”. In: *Numerische Mathematik* 84.3 (2000), pp. 375–393.

[16] D. P. Bertsekas and S. E. Shreve. *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific, 2007. ISBN: 1886529035.

[17] I. BOLDYREV and T. DÜPPE. “Programming the USSR: Leonid V. Kantorovich in context”. In: *The British Journal for the History of Science* 53.2 (2020), pp. 255–278. DOI: 10.1017/S0007087420000059.

[18] M. Brückerhoff and N. Juillet. “Instability of Martingale optimal transport in dimension $d \geq 2$”. In: *arXiv preprint arXiv:2101.06964* (2021).

[19] M. Davis, J. Obłój, and V. Raval. “Arbitrage bounds for prices of weighted variance swaps”. In: *Mathematical Finance* 24.4 (2014), pp. 821–854.

[20] Y. Dolinsky and H. M. Soner. “Martingale optimal transport and robust hedging in continuous time”. In: *Probab. Theory Related Fields* 160.1-2 (2014), pp. 391–427. ISSN: 0178-8051. DOI: 10.1007/s00440-013-0531-y.

[21] Y. Dolinsky and H. M. Soner. “Martingale optimal transport in the Skorokhod space”. In: *Stochastic Processes and their Applications* 125.10 (2015), pp. 3893–3931.

[22] R. M. Dudley. *Real analysis and probability*. CRC Press, 2018.

[23] A. Figalli, F. Maggi, and A. Pratelli. “A mass transportation approach to quantitative isoperimetric inequalities”. In: *Inventiones mathematicae* 182.1 (2010), pp. 167–211.

[24] P. J. Fitzsimmons. *Functional monotone class theorem*. Oct. 2011. URL: http://math.ucsd.edu/~pfitz/downloads/courses/fall11/math280a/mct.pdf.

[25] A. Galichon, P. Henry-Labordere, and N. Touzi. “A stochastic control approach to no-arbitrage bounds given marginals, with an application to Lookback”. In: (Feb. 2013). DOI: http://dx.doi.org/10.2139/ssrn.1912477. URL: https://ssrn.com/abstract=1912477.

[26] W. Gangbo and R. J. McCann. “The geometry of optimal transportation”. In: *Acta Mathematica* 177.2 (1996), pp. 113–161.

[27] C. Griessler. “An extended footnote on finitely minimal martingale measures”. In: *arXiv preprint arXiv:1606.03106* (2016).

[28] G. Guo and J. Obłój. “Computational methods for martingale optimal transport problems”. In: *Ann. Appl. Probab.* 29.6 (2019), pp. 3311–3347. ISSN: 1050-5164. DOI: 10.1214/19-AAP1481.

[29] P. Henry-Labordère. *Model-free hedging: A martingale optimal transport viewpoint*. CRC Press, 2017.

[30] D. G. Hobson. “Robust hedging of the lookback option”. In: *Finance and Stochastics* 2.4 (1998), pp. 329–347.

[31] N. Juillet et al. “Stability of the shadow projection and the left-curtain coupling”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 52.4 (2016), pp. 1823–1843.

[32] O. Kallenberg. *Foundations of modern probability*. Springer Science & Business Media, 2006.

[33] L. Kantorovich. “On the transfer of masses (originally in Russian)”. In: *Doklady Akademii Nauk* 37.2 (1942), pp. 227–229.

[34] Y.-H. Kim and B. Pass. “A general condition for Monge solutions in the multi-marginal optimal transport problem”. In: *SIAM Journal on Mathematical Analysis* 46.2 (2014), pp. 1538–1550.

[35] A. Klenke. *Probability Theory: A Comprehensive Course*. Universitext. Springer London, 2013. ISBN: 9781447153603.

[36] R. Lassalle. “Causal transport plans and their Monge–Kantorovich problems”. In: *Stochastic Analysis and Applications* 36.3 (2018), pp. 452–484.

[37] G. Monge. “Mémoire sur la théorie des déblais et des remblais”. In: *Histoire de l’Académie Royale des Sciences de Paris* (1781).

[38] V. M. Panaretos and Y. Zemel. *An invitation to statistics in Wasserstein space*. Springer Nature, 2020.

[39] G. Peyré and M. Cuturi. “Computational Optimal Transport”. In: *Foundations and Trends in Machine Learning* 11.5 - 6 (2019), pp. 355–607.

[40] G. C. Pflug and A. Pichler. “A distance for multistage stochastic optimization models”. In: *SIAM Journal on Optimization* 22.1 (2012), pp. 1–23.

[41] G. C. Pflug and A. Pichler. *Multistage stochastic optimization*. Springer, 2014.

[42] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems, Volume I: Theory, Volume II: Applications*. Springer Science & Business Media, 1998.

[43] Y. Rubner, C. Tomasi, and L. J. Guibas. “The earth mover’s distance as a metric for image retrieval”. In: *International journal of computer vision* 40.2 (2000), pp. 99–121.

[44] W. Rudin et al. *Principles of mathematical analysis*. Vol. 3. McGraw-hill New York, 1976.

[45] F. Santambrogio. “Optimal transport for applied mathematicians”. In: *Birkäuser, NY* 55.58-63 (2015), p. 94.

[46] L. Schwartz. *Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures*. Studies in mathematics. Tata Institute of Fundamental Research, 1973. ISBN: 9780195605167.

[47] V. Strassen et al. “The existence of probability measures with given marginals”. In: *Annals of Mathematical Statistics* 36.2 (1965), pp. 423–439.

[48] A.-S. Sznitman. *Topics in propagation of chaos*. Springer, 1991, pp. 165–251.

[49] A. W. Van der Vaart. *Asymptotic statistics*. Vol. 3. Cambridge university press, 2000.

[50] J. B. Veraguas et al. “Fundamental properties of process distances”. In: *Stochastic Processes and their Applications* 130.9 (2020), pp. 5575–5591.

[51] A. M. Vershik. “Descending sequences of measurable decompositions, and their applications”. In: 193.4 (1970), pp. 748–751.

[52] A. M. Vershik. “Theory of decreasing sequences of measurable partitions”. In: *Algebra i Analiz* 6.4 (1994), pp. 1–68.

[53] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008. ISBN: 9783540710509.

[54] C. Villani. *Topics in optimal transportation*. 58. American Mathematical Soc., 2003.

[55] J. Wiesel. “Continuity of the martingale optimal transport problem on the real line”. In: (2020). arXiv: `1905.04574 [math.PR]`.

[56] Y. Yajima. “Convex Envelopes in Optimization Problems”. In: *Encyclopedia of Optimization*. Ed. by C. A. Floudas and P. M. Pardalos. Boston, MA: Springer US, 2001, pp. 343–344. ISBN: 978-0-306-48332-5. DOI: `10.1007/0-306-48332-7_74`. URL: `https://doi.org/10.1007/0-306-48332-7_74`.