



universität
wien

MASTERARBEIT

Titel der Masterarbeit

An Introduction to Stochastic Portfolio Theory via Optimal Transport

verfasst von

Dominik Imgrüth, BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, August 2015

Studienkennzahl lt. Studienblatt: A 066 821

Studienrichtung lt. Studienblatt: Masterstudium Mathematik

Betreuer: Assoz. Prof. Dr. Mathias Beiglböck

CONTENTS

1	INTRODUCTION	7
1.1	A Word on Notation.....	7
2	OPTIMAL TRANSPORT	9
2.1	Aspects of Probability Theory	9
2.1.1	Framework and Notation.....	9
2.1.2	Weak Convergence and Theorem of Prohorov.....	10
2.2	Convex Analysis	11
2.2.1	Derivatives of Convex Functions	11
2.2.2	Convex Conjugates.....	12
2.2.3	Convexity and Cyclical Monotonicity: Rockafellar's Theorem....	13
2.2.4	Application: Concave Functions.....	16
2.3	Optimal Transport.....	17
2.3.1	Kantorovich's Transport Problem	17
2.3.2	Kantorovich Duality	21
2.3.3	Monge's Optimal Transport Problem.....	21
2.3.4	Knot-Smith Criterion & Brenier's Theorem.....	22
3	OPTIMALITY OF PSEUDO-ARBITRAGES	25
3.1	Pseudo-Arbitrage and Multiplicative Cyclical Monotonicity.....	25
3.1.1	Financial Framework	25
3.1.2	Performance of Portfolios	26
3.1.3	Multiplicative Cyclical Monotonicity	28
3.2	An Optimal Transport Characterisation for Pseudo-Arbitrage	31
3.2.1	Set-up of the Optimal Transport	31
3.2.2	The Characterisation Theorem	32
4	CONNECTIONS AND RÉSUMÉ.....	37
	BIBLIOGRAPHY.....	39
	SUPPLEMENTS.....	41
	Deutsche Zusammenfassung.....	41
	Lebenslauf	42

ABSTRACT

The theory of Optimal Transport can be used in many fields of mathematics. In this work it will be used to characterise pseudo arbitrages, following a paper of Soumik Pal and Ting-Kam Leonard Wong [PW14]. This characterisation can be found in Theorems 54 and 58. Therefore some theory of Convex Analysis and Optimal Transport will be covered, including Rockafellar's Theorem 17 about cyclical monotone sets. A multiplicative form of cyclical monotonicity will be introduced, for which we present Rockafellar-type Theorems 51 and 56. Furthermore we provide examples for pseudo-arbitrages, see Section 3.1.

INTRODUCTION

This master thesis is about an application of Optimal Transport in financial mathematics. To cover these topics we start in Chapter 2 to develop the theory of Optimal Transport, without going into great detail of this non-trivial theory. Instead, we try to understand the mechanics of it by using a special case, namely the case of a *quadratic cost function*.

Chapter 3 carries on with an introduction to a discrete time model of fixed finite stocks. The aim is to find portfolios that perform well compared to the marked, so-called *pseudo-arbitrages*. Finally we find a characterisation for these pseudo-arbitrages via Optimal Transport, using the results of Chapter 2.

A short comparison of the results of Chapters 2 & 3 will follow in the brief Chapter 4.

Before starting properly, it should be said, that the aim of this work is to give a first insight in a very huge topic, namely *Stochastic Portfolio Theory*, and to be 'easy to read' at the same time.

1.1 A WORD ON NOTATION

Two different notations are used for the euclidean inner product which are both well-established, namely, $\langle x, y \rangle$ and $x \cdot y$ for any two vectors $x, y \in \mathbb{R}^n$, both of course equal $\sum_{i=1}^n x_i y_i$. The simple reason why both notations will be applied is the following. $\langle \cdot, \cdot \rangle$ will be used, when the reader should be reminded that this is a scalar product, and $\cdot$ will be used to point out the similarity to the usual multiplication in $\mathbb{R}$, for instance in the well-known formula $(x - y) \cdot (x - y) = x \cdot x - 2x \cdot y + y \cdot y$. Furthermore $|\cdot|$ denotes the euclidean norm, such that the last equation can be rewritten as $|x - y|^2 = |x|^2 - 2x \cdot y + |y|^2$.

One of the major examples of measures is of course the *Lebesgue-measure*. It will be denoted by $\text{Leb} : \mathcal{B} \rightarrow [0, \infty]$, where $\mathcal{B}$ is the Borel- σ -algebra of $\mathbb{R}^d$. The dimension $d \in \mathbb{N}$ will always be clear from context. More on measures can be found in Section 2.1.2.

More on notational issues will be covered through out the text. Most of the notation is *standard notation*, so that readers who are familiar to mathematical texts will not face any troubles.

OPTIMAL TRANSPORT

This chapter tries to capture all relevant materials and knowledge for a better understanding of Chapter 3, where parts of the underlying paper [PW14] are discussed. Therefore, the theory of Optimal Transport will be developed. Before starting to discuss Optimal Transport we need some concepts of probability theory and convex analysis. This will be provided in the following two Sections 2.1 and 2.2.

A very good source for more on Optimal Transport is [Vil03], whereas [Vil09] includes more advanced topics.

2.1 ASPECTS OF PROBABILITY THEORY

With the goal to formulate the Theorem of Prohorov, we start by recalling some definitions of probability theory.

2.1.1 Framework and Notation

We begin elemental by defining a wide used framework for probability theory. By this we mean that our underlying spaces should be of the following form.

Definition 1. A complete separable metrizable space is called **Polish space**.

It can be presumed that Polish spaces are rich enough for studying probability theory, after reading the following line, found in [Kle13, Page 251]: “Praktisch sind alle Räume, die in der Wahrscheinlichkeitstheorie bedeutsam sind, polnische Räume.”, which can be translated as: In practice, every relevant space in probability theory is a Polish space.

Completeness is inherited by closeness, so statements like Lemma 2 below are easy to prove.

Lemma 2. *If X is a Polish metric space¹ and $Y \subseteq X$ is closed, then Y is also a Polish metric space.*

Here is a more meaningful theorem, which is of course harder to verify than Lemma 2, cf. [Kec95, Theorem 3.11]:

¹ A Polish metric space is a complete separable metric space.

Theorem 3. *If X is a Polish space, then any G_δ -subset of X is Polish, i.e. $Y \subseteq X$ is Polish if $\exists (O_i)_{i \in \mathbb{N}}$ open subsets of X such that $Y = \bigcap_{i \in \mathbb{N}} O_i$. Furthermore, if Y is a Polish subspace of a Polish space X , then Y is a G_δ -subset of X .*

Examples of Polish (metric) spaces are $(\mathbb{R}^n, d)$ where d denotes the usual (euclidean) distance and $(\mathcal{C}([0, 1]), \|\cdot\|_\infty)$ where $\|f\|_\infty := \sup_{x \in [0, 1]} f(x)$ is the usual *sup* norm and $\mathcal{C}$ denotes the space of continuous functions:

$$\mathcal{C}(E) := \mathcal{C}^0(E) := \{\varphi : E \rightarrow \mathbb{R} : \varphi \text{ continuous}\}$$

as a special case of $\mathcal{C}^k$, which is the space of k -times continuously differentiable functions. Furthermore the set of continuous bounded functions will be denoted by $\mathcal{C}_b$. Sometimes continuity of functions can be slightly weakened in the following way:

Definition 4. A function $f : X \rightarrow [-\infty, +\infty]$ is called **lower semi-continuous (lsc)** at $x \in X$ if $f(x) \leq \liminf_{y \rightarrow x} f(y)$. The function is called **upper semi-continuous (usc)** at $x \in X$ if $f(x) \geq \limsup_{y \rightarrow x} f(y)$.

2.1.2 Weak Convergence of Probability Measures and Theorem of Prohorov

Let $(\Omega, \mathcal{F})$ be some measure space. Then we may define the set of all probability measures on $(\Omega, \mathcal{F})$ by $\mathcal{P}(\Omega, \mathcal{F})$. If it is clear from context which σ -algebra is taken, we write $\mathcal{P}(\Omega) := \mathcal{P}(\mathcal{F}, \Omega)$. For instance, if E is some Polish space, then $\mathcal{P}(E)$ is the set of all *Borel* probability measures, i.e. the corresponding σ -algebra is generated by the open sets of E .

A subset $\Pi \subseteq \mathcal{P}(\Omega, \mathcal{F})$ is called **relatively weak compact** if for every sequence $(\pi_i)_{i \geq 1}$ in Π there exists a subsequence $(\pi_{i_k})_{k \geq 1}$ and a probability measure μ such that $\pi_{i_k} \xrightarrow{w} \mu$ as $k \rightarrow \infty$, i.e.

$$\forall f \in \mathcal{C}_b(\Omega) : \int f d\pi_{i_k} \xrightarrow{k \rightarrow \infty} \int f d\mu,$$

or equivalently in probability notation

$$\forall f \in \mathcal{C}_b(\Omega) : \mathbb{E}_{\pi_{i_k}} f \xrightarrow{k \rightarrow \infty} \mathbb{E}_\mu f.$$

Note that μ does not have to lie in Π .

For many purposes this convergence is appropriate, so this asks for a characterisation, which can be found in the famous next theorem.

Theorem 5 (Prohorov (1956)). *Let (E, d) be a metric space and consider a family of probability measures $\mathcal{F} \subseteq \mathcal{P}(E)$.*

- (i) *If $\mathcal{F}$ is tight, then $\mathcal{F}$ is relatively weak compact.*
- (ii) *If E is Polish and $\mathcal{F}$ is relatively weak compact, then $\mathcal{F}$ is tight.*

For completeness we recall the definition of *tightness*:

Definition 6. A probability measure μ defined on some measure space $(\Omega, \mathcal{F})$ is called **tight** if

$$\forall \varepsilon > 0 \exists \text{ compact } K \subset E : \mu(K) > 1 - \varepsilon. \quad (1)$$

More generally, a family Π of probability measures is said to be **tight** if (1) holds $\forall \mu \in \Pi$.

A detailed proof of Prohorov's Theorem 5 is given e.g. in [Kle13, Chapter 13]. For more details on this topic, the reader is politely referred to [Bil68].

2.2 CONVEX ANALYSIS

Most of the content of this section is taken from [Roc70], which is a standard reference in convex analysis. We attend to convex and concave functions and the differentiability of them, which will result in defining *sub-* and *super-gradients* and discuss some properties of them. Another aspect will be Cyclical Monotonicity which is of importance when discussing *optimality* of a transference plan, cf. Theorem 37.

Definition 7. A function $f : S \subseteq \mathbb{R}^n \rightarrow (-\infty, +\infty]$ is called **convex** if for each two points $x \neq y \in S$ and every $\lambda \in [0, 1]$ the following inequality holds

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \quad (2)$$

The function f is called **strict convex** if (2) holds with $<$ for every $\lambda \in (0, 1)$. Next, a convex function f is called **proper** if $f \not\equiv +\infty$.

2.2.1 Derivatives of Convex Functions

A function $f : (a, b) \subset \mathbb{R} \rightarrow \mathbb{R}$ is convex iff its second derivative is non-negative on the interval (a, b) . The proposition below contains the analogous statement for functions defined on $\mathbb{R}^n$:

Proposition 8. Let $C \subseteq \mathbb{R}^n$ and $f \in \mathcal{C}^2(C, \mathbb{R})$. Then f is convex iff its Hessian $H_x := (h_{ij}(x))_{1 \leq i, j \leq n}$ with

$$h_{ij}(x) := \frac{\partial^2 f}{\partial \xi_i \partial \xi_j}(\xi_1, \dots, \xi_n)$$

is positive semi-definite for all $x \in C$.

Next consider convex functions $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$. Recall, that the directional derivative in direction $v \in \mathbb{R}^n$ at a point $x \in \mathbb{R}^n$ where $f(x) < \infty$ is defined as follows:

$$D_v f(x) := \lim_{\lambda \searrow 0} \frac{f(x + \lambda v) - f(x)}{\lambda}. \quad (3)$$

From [Roc70, Theorem 23.1] we know that the directional derivative of a convex function exists at every point where the function is finite.

Whenever a convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is differentiable at $x \in \mathbb{R}^n$

$$\forall z \in \mathbb{R}^n \quad f(z) \geq f(x) + \langle \nabla f(x), z - x \rangle \quad (4)$$

is fulfilled and has the geometric meaning, that the whole graph of f lies above its tangent hyperplane at x . Also recall the following formula, where the connection of the gradient and the directional derivative is shown nicely:

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be differentiable at $x \in \mathbb{R}^n$, then we have for any $v \in \mathbb{R}^n$

$$D_v f(x) = \langle \nabla f(x), v \rangle.$$

We now want to generalize the gradient of a convex function in the sense of (4) and use this inequality as a definition:

Definition 9. Let $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ be a proper convex function and $x \in \mathbb{R}^n$. x^* is called a **sub-gradient of f at x** if

$$\forall z \in \mathbb{R}^n \quad f(z) \geq f(x) + \langle x^*, z - x \rangle, \quad (5)$$

and the set $\partial f(x)$ of all sub-gradients x^* of f at x is called **sub-differential of f at x** . The resulting multi-valued mapping $\partial f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $x \mapsto \partial f(x)$ is called **sub-differential of f** .

There's a sometimes useful characterisation using the directional derivative:

Proposition 10 ([Roc70, Theorem 23.2]). *Let $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ be a proper convex function and let $x \in \mathbb{R}^n$ such that $f(x) < \infty$. Then $x^* \in \partial f(x)$ iff $D_v f(x) \geq \langle x^*, v \rangle \forall v \in \mathbb{R}^n$.*

If f is differentiable it is clear from Definition 9 and (4), that $\nabla f(x) \in \partial f(x)$ for every $x \in \mathbb{R}^n$. Even more can be said:

Proposition 11 ([Roc70, Theorem 25.1]). *Let $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ be a proper convex function and $x \in \mathbb{R}^n$ such that $f(x)$ is finite. If f is differentiable at x , then the set $\partial f(x)$ contains only one element: $\partial f(x) = \{\nabla f(x)\}$. Conversely, if $\#(\partial f(x)) = 1$ (i.e. f has a unique sub-gradient at x), then f is differentiable at x .*

2.2.2 Convex Conjugates

Definition 12. Let $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ be a proper function. Then its **convex conjugate function f^*** is defined by

$$f^*(y) := \sup_{x \in \mathbb{R}^n} (x \cdot y - f(x)) \quad \forall y \in \mathbb{R}^n. \quad (6)$$

We show that f^* is convex indeed:

For any two points $y_1, y_2 \in \mathbb{R}^n$ and $\lambda \in [0, 1]$ arbitrary

$$\begin{aligned} f^*(\lambda y_1 + (1 - \lambda)y_2) &= \sup_{x \in \mathbb{R}^n} (\lambda y_1 \cdot x + (1 - \lambda)y_2 \cdot x - 1 \cdot f(x)) \\ &= \sup_{x \in \mathbb{R}^n} (\lambda y_1 \cdot x + (1 - \lambda)y_2 \cdot x - \lambda f(x) - (1 - \lambda) \cdot f(x)) \\ &\leq \lambda \sup_{x \in \mathbb{R}^n} (x \cdot y_1 - f(x)) + (1 - \lambda) \sup_{x \in \mathbb{R}^n} (x \cdot y_2 - f(x)) \\ &= \lambda f^*(y_1) + (1 - \lambda)f^*(y_2). \end{aligned}$$

From (6) **Fenchel's inequality**

$$\forall x, y \in \mathbb{R}^n : f^*(y) + f(x) \geq x \cdot y \quad (7)$$

follows immediately. If we have equality in (7), then this is a characterisation of sub-differentials in the following sense:

Proposition 13. *Let $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ be a proper (lsc) convex function and $y \in \mathbb{R}^n$. Then $y \in \partial f(x)$ iff $\forall x \in \mathbb{R}^n : f^*(y) + f(x) = x \cdot y$. And, in the same fashion, $x \in \partial f^*(y)$ iff $\forall y \in \mathbb{R}^n : f^*(y) + f(x) = x \cdot y$.*

This can be proven in a straight forward way, see e.g. the proof of [Vil03, Proposition 2.4].

Remark 14. As a side-remark, the theory of conjugacy is a theory about *the best* inequality of the type

$$x \cdot y \leq f(x) + g(y) \quad \forall x, y \in \mathbb{R}^n,$$

where $f, g : \mathbb{R}^n \rightarrow (-\infty, \infty]$. For a better understanding what '*the best*' means, consider a set $W := \{(f, g) : f, g : \mathbb{R}^n \rightarrow (-\infty, \infty] \text{ such that } x \cdot y \leq f(x) + g(y) \forall x, y \in \mathbb{R}^n\}$. Then a pair of functions $(f, g) \in W$ yields the **best** inequality, if for any pair $(f', g') \in W$ with $f' \leq f$ and $g' \leq g$ it follows that $f' = f$ and $g' = g$.

Now, clearly:

$$\begin{aligned} (f, g) \in W &\Leftrightarrow g(y) \geq \sup_{x \in \mathbb{R}^n} (x \cdot y - f(x)) = f^*(y) \quad \forall y \in \mathbb{R}^n \\ &\Leftrightarrow f(x) \geq \sup_{y \in \mathbb{R}^n} (x \cdot y - g(y)) = g^*(x) \quad \forall x \in \mathbb{R}^n \end{aligned}$$

and so the best pairs in W are precisely those pairs which satisfy $g = f^*$ and $f = g^*$.

By this remark, it should not be surprising that one uses convex conjugates to find optimality in some context. Cf. also Theorem 32, where a minimizing (i.e. optimal) pair of functions is of the form (f, f^*) .

2.2.3 Convexity and Cyclical Monotonicity: Rockafellar's Theorem

Definition 15 ([Vil03]). A subset $\Gamma \subseteq \mathbb{R}^n \times \mathbb{R}^n$ is called **cyclical monotone (CM)** if $\forall (x_1, y_1), \dots, (x_m, y_m) \in \Gamma$

$$\sum_{i=1}^m \langle y_i, x_{i+1} - x_i \rangle \leq 0,$$

with the convention, that $x_{m+1} := x_1$.

For functions f we also use the notation: f is CM, which means that the set $\text{Graph}(f)$ is CM.

Remark 16. In the 1-dimensional case this is just another notation for *non-decreasing*: A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is CM iff it is non-decreasing.

The essence of the following proof is the *Rearrangement Inequality* [HLP56, Theorem 368]: for every choice of real numbers $a_1 \leq \dots \leq a_n$ and $b_1 \leq \dots \leq b_n$ and for every permutation $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ it follows that

$$\sum_{i=1}^n a_i b_{n-i+1} \leq \sum_{i=1}^n a_i b_{\sigma(i)} \leq \sum_{i=1}^n a_i b_i.$$

Proof. (SKETCH) Let f be CM. Take any $x_1, x_2 \in \mathbb{R}$, then for the cycle $x_1 \rightarrow x_2 \rightarrow x_1$ Definition 15 implies that $f(x_1)(x_2 - x_1) + f(x_2)(x_1 - x_2) \leq 0$. This is, if $x_1 < x_2$, then $f(x_1) \leq f(x_2)$. Since points $x_1, x_2 \in \mathbb{R}$ were arbitrary, this shows that f is non-decreasing.

On the other hand, let f be non-decreasing and take any cycle $(x_i)_{i=1}^m$ in $\mathbb{R}$. We assume here for simplicity, that this cycle is ordered, i.e. $x_i \leq x_{i+1}$ for all $1 \leq i \leq m-1$. It follows

$$\begin{aligned} \sum_{i=1}^m f(x_i)(x_{i+1} - x_i) &= \sum_{i=1}^{m-1} f(x_i)(x_{i+1} - x_i) + f(x_m)(x_1 - x_m) \\ &\leq f(x_{m-1}) \sum_{i=1}^{m-1} (x_{i+1} - x_i) - f(x_m)(x_m - x_1) \\ &= f(x_{m-1})(x_m - x_1) - f(x_m)(x_m - x_1) \leq 0, \end{aligned}$$

where we used twice that f is non-decreasing. Now consider the case of an unordered cycle $(x_i)_{i=1}^m$ in $\mathbb{R}$ and assume wlog that there is just one point, where the cycle is unordered, i.e. $\exists! j$ such that $x_{j-1} \leq x_{j+1} \leq x_j \leq x_{j+2}$. Then first of all take a look at the following relationship, which will be used in a moment:

$$\begin{aligned} B &:= f(x_{j-1})(x_j - x_{j-1}) + f(x_j)(x_{j+1} - x_j) + f(x_{j+1})(x_{j+2} - x_{j+1}) \\ &= f(x_{j-1})(x_j - x_{j+1} + x_{j+1} - x_{j-1}) + f(x_j)(x_{j+1} - x_j) \\ &\quad + f(x_{j+1})(x_{j+2} - x_j + x_j - x_{j+1}) \\ &\leq f(x_{j-1})(x_{j+1} - x_{j-1}) + f(x_{j+1})(x_{j+2} - x_j) + f(x_j)(x_j - x_{j+1}) =: A. \end{aligned}$$

This inequality holds because $f(x_{j-1})(x_j - x_{j+1}) - f(x_j)(x_j - x_{j+1}) \leq 0$ and $f(x_{j+1}) \leq f(x_j)$ (both because f is non-decreasing). Now consider the primal equation

$$\begin{aligned} \sum_{i=1}^m f(x_i)(x_{i+1} - x_i) &= \left(\sum_{i \notin \{j-1, j, j+1\}} f(x_i)(x_{i+1} - x_i) + A \right) + B - A \\ &\leq \sum_{i \notin \{j-1, j, j+1\}} f(x_i)(x_{i+1} - x_i) + A \leq 0, \end{aligned}$$

where the first inequality follows from the previous considerations and the second inequality follows because the remaining expression is a *sum over a ordered cycle*, which was our simple case.

If there is more than one unordered element, proceed inductively in the same fashion. $\square$

Definition 15 is very useful for characterising sub-differentials of convex functions, and in the sequel (as we will see later) for identifying optimal transference plans. Here is the connection between CM and convex functions:

Theorem 17 (Rockafellar's Theorem, [Vil03, Theorem 2.27]). $\emptyset \neq \Gamma \subseteq \mathbb{R}^n \times \mathbb{R}^n$ is CM iff $\exists$ proper lsc convex function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ such that $\Gamma \subseteq \text{Graph}(\partial f)$. Moreover maximal CM subsets are exactly sub-differentials of proper lsc convex functions.

Remark 18. A set Γ is called **maximal CM** if there is no proper superset of Γ that is CM.

For a better understanding, the following corollary of Theorem 17 is proved here:

Corollary 19. Let $f \in C^1(\mathbb{R})$. f is convex iff $\text{Graph}(f')$ is CM.

This means, together with Remark 16, this corollary says in particular that a function is convex iff its derivative is non-decreasing.

Proof. Let f be convex. Then

$$\forall x, z \in \mathbb{R} : f(z) - f(x) \geq f'(x)(z - x), \quad (8)$$

cf. (4). Now choose points $(x_1, y_1), \dots, (x_m, y_m) \in \text{Graph}(f')$, i.e. choose arbitrary $x_i \in \mathbb{R}$, then $y_i = f'(x_i)$ for all i . Now from (8) we obtain

$$\begin{aligned} f(x_2) - f(x_1) &\geq f'(x_1)(x_2 - x_1) \\ &\vdots \\ f(x_m) - f(x_{m-1}) &\geq f'(x_{m-1})(x_m - x_{m-1}) \\ f(x_1) - f(x_m) &\geq f'(x_m)(x_1 - x_m). \end{aligned}$$

Hence, by summing over these equations,

$$0 = \sum_{i=1}^m (f(x_{i+1}) - f(x_i)) \geq \sum_{i=1}^m f'(x_i)(x_{i+1} - x_i),$$

using the convention $x_{m+1} = x_1$. The points $(x_1, y_1), \dots, (x_m, y_m) \in \text{Graph}(f')$ were arbitrary, so $\text{Graph}(f')$ is CM.

Now let $\text{Graph}(f')$ be CM. Like in the proof of Remark 16, we have by Definition 15 ($m = 2$) for any $x_1, x_2 \in \mathbb{R}$

$$f'(x_1)(x_2 - x_1) + f'(x_2)(x_1 - x_2) \leq 0,$$

therefore, if $x_1 < x_2$, then $f'(x_1) \leq f'(x_2)$ follows, i.e. f' is a non-decreasing function. Hence f has to be convex. $\square$

2.2.4 Application: Concave Functions

Definition 20. A function $f : X \rightarrow [-\infty, +\infty)$ is called **concave** if $-f$ is convex.

In the same fashion strict concavity is defined. Also: a concave function is called **proper** if $f \not\equiv -\infty$.

Analogue to Definition 9 one defines the *super*-gradient for concave functions:

Definition 21. Let $f : \mathbb{R}^n \rightarrow [-\infty, +\infty)$ be a proper concave function and $x \in \mathbb{R}^n$. x^* is called a **super-gradient of f at x** if

$$\forall z \in \mathbb{R}^n \quad f(z) \leq f(x) + \langle x^*, z - x \rangle, \quad (9)$$

and the set $\partial f(x)$ of all super-gradients x^* of f at x is called **super-differential of f at x** . The resulting multi-valued mapping $\partial f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $x \mapsto \partial f(x)$ is called **super-differential of f** .

Remark 22. A note on notation: using the ∂ -symbol for sub-gradients as well as for super-gradients may be a slight abuse of notation. Nevertheless it will always be clear, since the considered functions will always be specified to be concave or convex. Similarly: statements like Proposition 11 stay true for the concave version using 'super-gradient' instead of 'sub-gradient'.

For later use, the next short proposition is stated. As a special case it proves that, given a concave function Φ , $\log \Phi$ is also concave. The proof is very elementary.

Proposition 23. Let $g : V \subseteq \mathbb{R} \rightarrow \mathbb{R}$ be a non-decreasing, concave function and $f : U \subseteq \mathbb{R}^n \rightarrow V$ be a concave function, where U and V are convex. Then the composition $g \circ f : U \rightarrow \mathbb{R}$ is concave.

Proof. Let $\lambda \in [0, 1]$ arbitrary but fixed and let $u_i \in U$ for $i = 1, 2$. f is concave and g is non-decreasing, hence

$$g(f(\lambda u_1 + (1 - \lambda)u_2)) \geq g(\lambda f(u_1) + (1 - \lambda)f(u_2)).$$

Now by concavity of g for $v_i \in V$, $i = 1, 2$

$$g(\lambda v_1 + (1 - \lambda)v_2) \geq \lambda g(v_1) + (1 - \lambda)g(v_2)$$

and for $v_i := f(u_i) \in V$, $i = 1, 2$ the statement follows. $\square$

Proposition 24 (Chain-Rule for Super-Differentials). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ be concave functions such that the composition $g \circ f$ is again concave. Then the following relationship between super-differentials is true for every $p \in \mathbb{R}^n$:

$$\partial(g \circ f)[p] \supseteq \partial g[f(p)]\partial f[p].^2 \quad (10)$$

² We use $[\cdot]$ -symbols here to point out explicitly, at which point the corresponding super-differentials are evaluated.

Proof. Let $p \in \mathbb{R}^n$ be arbitrary and take $\xi \in \partial g[f(p)]\partial f[p]$, i.e. there are $\alpha \in \partial g[f(p)]$ and $\beta \in \partial f[p]$ such that $\xi = \alpha \cdot \beta$.

Now let $z \in \mathbb{R}^n$ be arbitrary but fixed. By using the super-gradient-inequality (9) for $f(z) \in \mathbb{R}$ we have

$$g(f(z)) \leq g(f(p)) + \alpha \cdot (f(z) - f(p)),$$

because $\alpha \in \partial g[f(p)]$. Using now $\beta \in \partial f[p]$, we infer for $z \in \mathbb{R}^n$ (again by (9))

$$f(z) \leq f(p) + \langle \beta, z - p \rangle.$$

Stacking the last two inequalities together one obtains

$$g(f(z)) \leq g(f(p)) + \langle \alpha \cdot \beta, z - p \rangle = g(f(p)) + \langle \xi, z - p \rangle.$$

But z was arbitrary so by (9), $\xi \in \partial(g \circ f)[p]$ follows. $\square$

As immediate consequence of Proposition 24 together with Proposition 23 and Proposition 11 we infer the following set-inclusion, which will be of importance in Chapter 3. Let $f : \mathbb{R}^n \rightarrow [0, \infty)$ be concave, then

$$\forall p \in \mathbb{R}^n \quad \partial \log f(p) \supseteq \frac{1}{f(p)} \partial f(p), \quad (11)$$

which has the following meaning: *If ξ is a super-gradient of $\frac{1}{f} \partial f$ at a point p , then ξ is also a super-gradient of $\log f$ at p .*

2.3 OPTIMAL TRANSPORT

Now we are ready to start with Optimal Transport. As a first step we introduce a formulation of Optimal Mass Transport. This will be called *Kantorovich's Transport Problem*.

2.3.1 Kantorovich's Transport Problem

Consider two probability spaces (X, μ) and (Y, ν) and a measurable function $c : X \times Y \rightarrow [0, \infty]$ which will be called **cost function**. Then *Kantorovich's Transport Problem* is the aim to find a minimizer for the functional

$$\pi \rightarrow I_c[\pi] := \int_{X \times Y} c(x, y) d\pi(x, y) \quad (12)$$

among all $\pi \in \Pi(\mu, \nu)$ where

$$\begin{aligned} \Pi(\mu, \nu) := \{ \pi \in \mathcal{P}(X \times Y) : \pi(A \times Y) = \mu(A) \ \forall A \text{ mb } \subseteq X, \\ \pi(X \times B) = \nu(B) \ \forall B \text{ mb } \subseteq Y \} \end{aligned}$$

is the set of probability measures on $X \times Y$ with marginals μ and ν , respectively. Here and henceforth 'mb' is used as abbreviation for *measurable*.

Definition 25. $\Pi(\mu, \nu)$ is called the set of all **transference plans between** (X, μ) and (Y, ν) . If $I_c[\pi_*] = \min_{\pi \in \Pi(\mu, \nu)} I_c[\pi]$ then π_* will be called **optimal transference plan** (w.r.t. to the cost function c).

Remark 26. If U and V are random variables in some spaces X and Y , one can define a transference plan $\tilde{\pi}$ via marginals $\text{law}(U)$ and $\text{law}(V)$, respectively. $\tilde{\pi}$ then is the joint law of (U, V) . In this case we call $\tilde{\pi}$ **coupling of U and V** .

In this setting, the minimizing problem (12) can be turned into minimizing

$$(U, V) \rightarrow \mathbb{E}[c(U, V)]$$

over all pairs (U, V) of random variables U in X , and V in Y , such that $\mu = \text{law}(U)$ and $\nu = \text{law}(V)$.

Lemma 27. *The set $\Pi(\mu, \nu)$ is non-empty and convex.*

Proof. The product measure $\delta : X \times Y \rightarrow [0, 1]$, $\delta(A \times B) = \mu(A)\nu(B)$ for A and B mb subsets of X and Y , respectively, is always contained in $\Pi(\mu, \nu)$, therefore it's non-empty.

For convexity, take two measures $\pi_1, \pi_2 \in \Pi(\mu, \nu)$, take $\lambda \in [0, 1]$ arbitrary and consider the new measure

$$\eta := \lambda\pi_1 + (1 - \lambda)\pi_2.$$

Then η belongs to $\Pi(\mu, \nu)$:

- (i) η is clearly a probability measure on $X \times Y$,
- (ii) for $A \subseteq X$ mb,

$$\begin{aligned} \eta(A \times Y) &= \lambda\pi_1(A \times Y) + (1 - \lambda)\pi_2(A \times Y) \\ &= \lambda\mu(A) + (1 - \lambda)\mu(A) \\ &= \mu(A), \end{aligned}$$

- (iii) for $B \subseteq Y$ mb analogue: $\eta(X \times B) = \nu(B)$.

□

Lemma 28. *Let (X, μ) and (Y, ν) be two probability spaces. Then the following are equivalent*

- (i) $\pi \in \Pi(\mu, \nu)$
- (ii) π is a measure on $X \times Y$ such that $\forall \text{mb } (\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$:

$$\int_{X \times Y} (\varphi(x) + \psi(y)) d\pi(x, y) = \int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y). \quad (13)$$

Furthermore, if X and Y are Polish, properties (i) and (ii) are equivalent to

- (iii) π is a measure on $X \times Y$ such that $\forall (\varphi, \psi) \in \mathcal{C}_b(X) \times \mathcal{C}_b(Y)$ (13) holds.

Proof. We show the equivalence of (i) and (ii).

First suppose (ii). Let $A \subseteq X$ be a mb subset and define $\varphi := \mathbb{1}_A$. Then of course

$\varphi \in L^1(d\mu)$ because $\int_X \varphi d\mu = \mu(A) \leq 1$ since μ is a probability measure. Hence one can write

$$\pi(A \times Y) = \int_{X \times Y} \mathbb{1}_{A \times Y}(x, y) d\pi(x, y) = \int_{X \times Y} \mathbb{1}_A(x) \mathbb{1}_Y(y) d\pi(x, y).$$

Using $\mathbb{1}_Y(y) \equiv 1$ and finally (13), one arrives at

$$\pi(A \times Y) = \int_{X \times Y} \mathbb{1}_A(x) d\pi(x, y) = \int_X \mathbb{1}_A(x) d\mu(x) = \mu(A).$$

By an analogous statement $\pi(X \times b) = \nu(b)$ follows and hence (i) is shown.

Now let $\varphi \in L^1(d\mu)$ and $\psi \in L^1(d\nu)$ be arbitrary. Choose monotone increasing sequences of simple functions $(\varphi_k)_{k \geq 1}$ and $(\psi_k)_{k \geq 1}$ such that $\varphi_k \nearrow \varphi$ and $\psi_k \nearrow \psi$ point-wise as $k \rightarrow \infty$. For simple functions (i) $\Rightarrow$ (ii) is almost trivial:

First, for functions $\varphi_k := \mathbb{1}_{A_k}$ and $\psi_k := \mathbb{1}_{B_k}$ where A_k and B_k are mb subsets of X and Y , respectively, one has

$$\begin{aligned} \int_{X \times Y} (\mathbb{1}_{A_k}(x) + \mathbb{1}_{B_k}(y)) d\pi(x, y) &= \int_{X \times Y} \mathbb{1}_{A_k}(x) d\pi(x, y) + \int_{X \times Y} \mathbb{1}_{B_k}(y) d\pi(x, y) \\ &= \pi(A_k \times Y) + \pi(X \times B_k) \\ &= \int_X \mathbb{1}_{A_k}(x) d\mu(x) + \int_Y \mathbb{1}_{B_k}(y) d\nu(y) \end{aligned}$$

where the last equality follows because we assumed (i), i.e. $\pi \in \Pi(\mu, \nu)$, and therefore (ii) is shown in the case of indicator functions. Statement (ii) follows for simple functions $\varphi_k := \sum_{i=1}^n \mathbb{1}_{A_k}$ and $\psi_k := \sum_{j=1}^m \mathbb{1}_{B_k}$ easily. Using the *Monotone Convergence Theorem*, (ii) follows. $\square$

The first question is, if one always (under some conditions, which will be specified later) can find a minimizer π_* of $I(\pi)$. First consider the case of a *quadratic cost function* $c : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, \infty)$ with $c(x, y) = \frac{|x-y|^2}{2}$, i.e. we consider the Kantorovic minimization problem

$$\inf_{\pi \in \Pi(\mu, \nu)} I[\pi] := \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{|x-y|^2}{2} d\pi(x, y), \quad (14)$$

and give a more general result afterwards.

Proposition 29 (Existence of an optimal transference plan). *Let μ, ν be two Borel probability measures on $\mathbb{R}^n$ with finite second order moments. Then there is an optimal π_* for the minimization problem described in (14).*

Proof. First we note that $I[\pi]$ is finite for every $\pi \in \Pi(\mu, \nu)$:

$$\begin{aligned} I[\pi] &= \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{|x-y|^2}{2} d\pi(x, y) \\ &\leq \int_{\mathbb{R}^n \times \mathbb{R}^n} |x|^2 d\pi(x, y) + \int_{\mathbb{R}^n \times \mathbb{R}^n} |y|^2 d\pi(x, y) \\ &= \int_{\mathbb{R}^n} |x|^2 d\mu(x) + \int_{\mathbb{R}^n} |y|^2 d\nu(y) < \infty \end{aligned}$$

since μ, ν have finite second order moments.

We want to use Prohorov's Theorem 5 and therefore we show, that $\Pi(\mu, \nu)$ is tight. For this aim let $\varepsilon > 0$ be arbitrary and take compact (thus mb) $K \subseteq X$ and $L \subseteq Y$ such that $\mu(X \setminus K) \leq \frac{\varepsilon}{2}$ and $\nu(Y \setminus L) \leq \frac{\varepsilon}{2}$. Such sets K, L exist because every probability measures on $\mathbb{R}^n$ is tight (cf. for example the more general result [Bil68, Theorem 1.4]: *If S is separable and complete, then each probability measure on $(S, \cdot)$ is tight*). This leads to

$$\begin{aligned} \pi((X \times Y) \setminus (K \times L)) &\leq \pi(X \times (Y \setminus L)) + \pi((X \setminus K) \times Y) \\ &= \nu(Y \setminus L) + \mu(X \setminus K) < \varepsilon \end{aligned}$$

for $\pi \in \Pi(\mu, \nu)$ arbitrary, and therefore $\Pi(\mu, \nu)$ is tight and by Prohorov relatively weak compact. Next, $\Pi(\mu, \nu)$ is weakly closed: take a sequence of probability measures $(\pi_k)_{k \geq 1}$ in $\Pi(\mu, \nu)$ with $\pi_k \xrightarrow{w} \pi$ as $k \rightarrow \infty$. By Lemma 28 for k arbitrary

$$\begin{aligned} \pi_k \in \Pi(\mu, \nu) &\Leftrightarrow \forall (\varphi, \psi) \in \mathcal{C}_b(X) \times \mathcal{C}_b(Y) : \\ \int_{X \times Y} (\varphi(x) + \psi(y)) d\pi_k(x, y) &= \int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y). \end{aligned}$$

But then also $\forall (\varphi, \psi) \in \mathcal{C}_b(X) \times \mathcal{C}_b(Y)$

$$\begin{aligned} \int_{X \times Y} (\varphi(x) + \psi(y)) d\pi(x, y) &= \lim_{k \rightarrow \infty} \int_{X \times Y} (\varphi(x) + \psi(y)) d\pi_k(x, y) \\ &= \int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y). \end{aligned}$$

and this shows $\pi \in \Pi(\mu, \nu)$, hence weak closeness of $\Pi(\mu, \nu)$.

Now $\Pi(\mu, \nu)$ is relative weak compact and weakly closed. This implies $\Pi(\mu, \nu)$ is weakly compact, hence I has a minimizer:

Indeed, take a *minimizing sequence* $(\pi_k)_{k \geq 1}$ in $\Pi(\mu, \nu)$, i.e. $I(\pi_k) \xrightarrow{k \rightarrow \infty} \inf_{\pi \in \Pi(\mu, \nu)} I(\pi)$. This sequence admits a accumulation point $\pi_* \in \Pi(\mu, \nu)$. Now take a non-decreasing sequence of continuous bounded functions $(c_l)_{l \geq 1}$ such that $c_l \nearrow c$.

$$\begin{aligned} \int c(x, y) d\pi_*(x, y) &= \int \lim_{l \rightarrow \infty} c_l(x, y) d\pi_*(x, y) \\ &= \lim_{l \rightarrow \infty} \int c_l(x, y) d\pi_*(x, y) \quad (\text{monotone convergence}) \\ &\leq \lim_{l \rightarrow \infty} \limsup_{k \rightarrow \infty} \int c_l(x, y) d\pi_k(x, y) \quad (\text{accumulation point}) \\ &\leq \limsup_{k \rightarrow \infty} \int c(x, y) d\pi_k(x, y) \quad (c_l \leq c) \\ &= \inf_{\pi \in \Pi(\mu, \nu)} I(\pi) \quad (\text{minimizing sequence}), \end{aligned}$$

thus π_* is a minimizer of I . □

This statement can be generalized in the following setting:

Theorem 30 ([Vil09, Theorem 4.1]). *Let (X, μ) and (Y, ν) be two Polish probability spaces and let $a : X \rightarrow [-\infty, +\infty)$ and $b : Y \rightarrow [-\infty, +\infty)$ be two usc functions that are integrable: $a \in L^1(\mathrm{d}\mu)$ and $b \in L^1(\mathrm{d}\nu)$. Let $c : X \times Y \rightarrow (-\infty, \infty]$ a lsc cost function such that $\forall x \in X, y \in Y : c(x, y) \geq a(x) + b(y)$. Then*

$$\exists \pi_* \in \Pi(\mu, \nu) : \inf_{\pi \in \Pi(\mu, \nu)} I[\pi] = I[\pi_*].$$

2.3.2 Kantorovich Duality

Define for a mb pair of functions $(\varphi, \psi) \in L^1(\mathrm{d}\mu) \times L^1(\mathrm{d}\nu)$

$$J(\varphi, \psi) := \int_X \varphi \mathrm{d}\mu + \int_Y \psi \mathrm{d}\nu.$$

Note by Lemma 28, $\pi \in \Pi(\mu, \nu)$ is equivalent to

$$J(\varphi, \psi) = \int_{X \times Y} (\varphi(x) + \psi(y)) \mathrm{d}\pi(x, y).$$

Theorem 31 (Kantorovich Duality, [Vil03, Theorem 1.3]). *Let (X, μ) and (Y, ν) be two Polish (Borel) probability spaces and let $c : X \times Y \rightarrow [0, \infty]$ a lsc cost function. Define*

$$\Phi_c := \{(\varphi, \psi) \in L^1(\mathrm{d}\mu) \times L^1(\mathrm{d}\nu) : (\varphi, \psi) \text{ mb} \wedge \varphi(x) + \psi(y) \leq c(x, y) \forall (x, y) \in X \times Y\}.$$

Then we have the dual problem

$$\inf_{\pi \in \Pi(\mu, \nu)} I[\pi] = \sup_{(\varphi, \psi) \in \Phi_c} J(\varphi, \psi). \quad (15)$$

Furthermore, the set Φ_c can be restricted to the set of continuous bounded functions:

$$\sup_{(\varphi, \psi) \in \Phi_c} J(\varphi, \psi) = \sup_{(\varphi, \psi) \in \Phi_c \cap (\mathcal{C}_b \times \mathcal{C}_b)} J(\varphi, \psi). \quad (16)$$

2.3.3 Monge's Optimal Transport Problem

As a special case of the *Kantorovich Optimal Transport Problem* we introduce Monge's version. To transport mass from X to Y Monge used a deterministic (mb) map $T : X \rightarrow Y$. This is a more intuitive way since this means that *no mass be split* when transported from X to Y . This leads to the following minimizing problem which is called *Monge's Optimal Transport Problem*. Minimize

$$I_c[T] := \int_X c(x, T(x)) \mathrm{d}\mu(x) \quad (17)$$

over all mb maps $T : X \rightarrow Y$ such that $T\# \mu = \nu$.

For a heuristic interpretation see e.g. [Vil03, Introduction].

³ ' $\forall (x, y) \in X \times Y$ ' is a small abuse of notation, since φ and ψ are only defined on equivalence classes. We mean the following: $\exists E_X \subseteq X, E_Y \subseteq Y$ with $\mu(X \setminus E_X) = 0$ and $\nu(Y \setminus E_Y) = 0$ such that $\varphi(x) + \psi(y) \leq c(x, y)$ holds for every pair $(x, y) \in E_X \times E_Y$.

2.3.4 Knot-Smith Criterion & Brenier's Theorem

Theorem 32 (Knot-Smith Optimality Criterion for Quadratic Cost, [Vil03, Theorem 2.12 (i)]). *Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ with finite second order moments and consider the transport problem (14). Then $\pi \in \Pi(\mu, \nu)$ is optimal iff*

$$\exists \text{ convex lsc } \varphi : \text{Supp}(\pi) \subseteq \text{Graph}(\partial\varphi).$$

In this case the pair (φ, φ^) has to be a minimizer in the problem*

$$\inf \left\{ \int_{\mathbb{R}} \varphi \, d\mu + \int_{\mathbb{R}} \psi \, d\nu : (\varphi, \psi) \in \tilde{\Phi} \right\}, \quad (18)$$

where

$$\begin{aligned} \tilde{\Phi} := \{(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu) : (\varphi, \psi) \text{ mb} \wedge \\ x \cdot y \leq \varphi(x) + \psi(y) \, \forall (x, y) \in X \times Y\}.^4 \end{aligned}$$

Remark 33. (i) Note that

$$\begin{aligned} \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : \pi(B_\delta(x, y)) > 0 \, \forall \delta > 0\} &= \text{Supp}(\pi) \\ &\subseteq \text{Graph}(\partial\varphi) = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : y \in \partial\varphi(x)\} \end{aligned}$$

is equivalent to ask for $\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n$:

$$\pi(B_\delta(x, y)) > 0 \, \forall \delta > 0 \Rightarrow y \in \partial\varphi(x)$$

Therefore Theorem 32 says: $\pi \in \Pi(\mu, \nu)$ is optimal iff

$$\exists \text{ convex lsc } \varphi : \text{for } d\pi\text{-almost all } (x, y) : y \in \partial\varphi(x).$$

- (ii) One may wonder about the *new* minimization problem (18), which looks similar to the introduced maximization problem in the Kantorovich Duality (15). In fact we use the specific form of the cost function, to see that the resulting (18) and (15) yield the same optimal transference plan:

At first define $M_2 := \int_{\mathbb{R}^n} \frac{|x|^2}{2} d\mu(x) + \int_{\mathbb{R}^n} \frac{|y|^2}{2} d\nu(y) < \infty$ as the half of the sum of the second order moments of μ and ν which is finite by assumption. Consider

$$\begin{aligned} \sup_{(\varphi, \psi) \in \Phi_c} J(\varphi, \psi) &= \\ &= \sup_{\Phi_c} \left[\int_{\mathbb{R}^n} \left(\varphi(x) - \frac{|x|^2}{2} \right) d\mu(x) + \int_{\mathbb{R}^n} \left(\psi(y) - \frac{|y|^2}{2} \right) d\nu(y) + M_2 \right] \\ &= M_2 - \inf_{\Phi_c} \left[\int_{\mathbb{R}^n} \left(\frac{|x|^2}{2} - \varphi(x) \right) d\mu(x) + \int_{\mathbb{R}^n} \left(\frac{|y|^2}{2} - \psi(y) \right) d\nu(y) \right]. \end{aligned}$$

⁴ See ³.

Now, set $\tilde{\varphi}(x) := \frac{|x|^2}{2} - \varphi(x)$ and analogously $\tilde{\psi}(y) := \frac{|y|^2}{2} - \psi(y)$. Note, since the second order moment of μ exist, $\tilde{\varphi} \in L^1(d\mu)$ iff $\varphi \in L^1(d\mu)$. Also: $\tilde{\varphi}$ is mb iff φ is mb, since $\frac{|\cdot|^2}{2}$ is continuous and thus mb. The same holds of course for $\tilde{\psi}$. Using now the special form of c we deduce $(\varphi, \psi) \in \Phi_c$ iff $\frac{|x-y|^2}{2} \geq \frac{|x|^2}{2} - \tilde{\varphi}(x) + \frac{|y|^2}{2} - \tilde{\psi}(y)$ for $d\mu$ -almost all $x \in \mathbb{R}^n$ and dv -almost all $y \in \mathbb{R}^n$ (cf. Definition of Φ_c in Theorem 31). By $|x-y|^2 = |x|^2 - 2x \cdot y + |y|^2$ one obtains equality of the last infimum with (18).

This means: The optimal transference plan obtained by (18) is the same as that obtained in the Kantorovich Duality, although the values of (18) and (15) differ.

Theorem 34 (Brenier's Theorem [Vilo3, Theorem 2.32]). *Let μ and ν two probability measures on $\mathbb{R}^n$ such that $\mu \ll \text{Leb}$. Then $\exists$ mb $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that*

$$T\#\mu = \nu \text{ and } T = \nabla \varphi$$

for some convex function $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$. The map T is unique up to $d\mu$ -null sets.

For a more advanced result, we define a more general form of cyclical monotonicity:

Definition 35 ([Vilo9]). Let X, Y be arbitrary sets and let $c : X \times Y \rightarrow (-\infty, +\infty]$ be a function. A subset $\Gamma \subseteq X \times Y$ is called **c-cyclically monotone** (**c-CM**) if $\forall n \in \mathbb{N}$ and arbitrary points $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$ the following inequality is fulfilled:

$$\sum_{i=1}^n c(x_i, y_i) \leq \sum_{i=1}^n c(x_i, y_{i+1}) \text{ where } y_{n+1} := y_1. \quad (19)$$

We also call a transference plan c-CM, if it is concentrated on a c-CM set.

In the simple case of the quadratic cost function $c(x, y) = \frac{|x-y|^2}{2}$, this definition coincides with the ordinary cyclical monotonicity:

Lemma 36. *For $c : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, \infty)$, $c(x, y) = \frac{|x-y|^2}{2}$ a set $\Gamma \subseteq \mathbb{R}^n \times \mathbb{R}^n$ is CM iff it is c-CM.*

Proof. Let $\Gamma \subseteq \mathbb{R}^n \times \mathbb{R}^n$ be a c-CM set and take n arbitrary points $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$. Define $y_{n+1} := y_1$, then by Definition 35, we have

$$\sum_{i=1}^n |x_i - y_i|^2 \leq \sum_{i=1}^n |x_i - y_{i+1}|^2.$$

By expanding this, one gets the equivalent

$$-2 \sum_{i=1}^n x_i \cdot y_i + \sum_{i=1}^n |y_i|^2 \leq -2 \sum_{i=1}^n x_i \cdot y_{i+1} + \sum_{i=1}^n |y_{i+1}|^2$$

and finally by using $y_{n+1} = y_1$

$$\sum_{i=1}^n x_i \cdot (y_{i+1} - y_i) \leq 0.$$

Since points $(x_i, y_i) \in \Gamma$ were chosen arbitrarily, Γ is CM.

The other direction of the Lemma follows from exactly the same arguments. $\square$

So the next theorem is a generalized version of the *Knot-Smith criterion* (cf. Theorem 32 statement (i)) together with the Theorem of Rockafellar (cf. Theorem 17):

Theorem 37 (c-Cyclical Monotonicity Characterises Optimality of Transference Plans [Vil09, Theorem 5.10]). *Let (X, μ) and (Y, ν) be two Polish probability spaces and let $a : X \rightarrow \mathbb{R}$ and $b : Y \rightarrow \mathbb{R}$ be two usc functions that are integrable: $a \in L^1(\mathrm{d}\mu)$ and $b \in L^1(\mathrm{d}\nu)$. Let $c : X \times Y \rightarrow \mathbb{R}$ be a lsc cost function such that $\forall x \in X, y \in Y : c(x, y) \geq a(x) + b(y)$. Furthermore suppose that $\inf_{\pi \in \Pi(\mu, \nu)} I_c[\pi]$ is finite, then $\forall \pi \in \Pi(\mu, \nu)$:*

$$\pi \text{ is optimal} \Leftrightarrow \pi \text{ is } c\text{-CM}$$

OPTIMALITY OF PSEUDO-ARBITRAGES

In this chapter, we want to use the theory of Optimal Transport for a problem in a financial market. The market itself will be very simple, in particular we only consider a discrete time set-up as we will see in Section 3.1. Theorems 30 and 37 will play an important role in Section 3.2, where we finally find a characterisation of pseudo-arbitrages.

3.1 PSEUDO-ARBITRAGE AND MULTIPLICATIVE CYCLICAL MONOTONICITY

3.1.1 Financial Framework

Given a stock market with n stocks, namely $X_1(t), \dots, X_n(t)$ for some given time $t \geq 0$, the value of the whole market at time t is clearly $S(t) := \sum_{i=1}^n X_i(t)$. Define the **market weights** at time t by $\mu_i(t) := \frac{X_i(t)}{S(t)}$ ($i = 1, \dots, n$), as the relative values of the single stocks compared to the whole market. If we define the **simplex** to be

$$\Delta^n := \left\{ (p_1, \dots, p_n) \in (0, 1)^n : \sum_{i=1}^n p_i = 1 \right\}, \quad (20)$$

then the market weights belong to the closure of the simplex $\mu_i(t) \in \overline{\Delta^n}$ for every t . $\overline{\Delta^n}$ itself is a $(n - 1)$ -dimensional manifold in $\mathbb{R}^n$ and we will work with the usual topology, whenever considering $\overline{\Delta^n}$ or subsets of it. For later use, we also define the following. Choose $\varepsilon \in (0, \frac{n-1}{n})$ and define the **simplex with cut corners** by

$$\Delta_c^n(\varepsilon) := \left\{ (p_1, \dots, p_n) \in (0, 1 - \varepsilon)^n : \sum_{i=1}^n p_i = 1 \right\}. \quad (21)$$

Also, take $\kappa \in (0, \frac{1}{n})$, then the **simplex without border zone** is defined as

$$\Delta_b^n(\kappa) := \left\{ (p_1, \dots, p_n) \in (\kappa, 1)^n : \sum_{i=1}^n p_i = 1 \right\}. \quad (22)$$

Note here, that for $\kappa = \varepsilon$ we have the inclusion $\Delta_b^n(\kappa) \subset \Delta_c^n(\kappa)$.

Portfolios are the central objects and worth to be discussed shortly. Following [PW14], we define portfolios as maps:

Definition 38. A **portfolio** is a mb map $\mathbf{p} : \Delta^n \rightarrow \overline{\Delta^n}$.

The interpretation is as follows: if the current market weights are given by $\mu(t) = (\mu_1(t), \dots, \mu_n(t))$ the corresponding portfolio is $\mathfrak{p}(\mu(t)) \in \overline{\Delta^n}$. If the initial wealth is w , we hold, accordingly to this portfolio, $w \cdot \mathfrak{p}_i(\mu(t))$ of the i -th stock. $\mathfrak{p}(\mu(t)) \in \overline{\Delta^n}$ implies that these portfolios are *long-only*.

There are some major examples as the *market portfolio*, the *constant weighted portfolio* or the *diversity-weighted portfolio*. We will develop these portfolios in the next sections to identify their properties. So far, we will just give the definitions of them. Of course all of them are portfolios in the sense of Definition 38.

Example 39. The **market portfolio** is the easiest example of a portfolio, which one may think of. At time $t = 0$ it buys the same amount of each stock and holds them forever (*Buy and Hold Strategy*). Therefore it is simply the identity map on Δ^n , i.e. $\mathfrak{p} = \text{Id} : \Delta^n \rightarrow \Delta^n$, $\mathfrak{p}(p) = p$.

Next, fix a point $\eta \in \Delta^n$. The **constant-weighted portfolio** buys or sells stocks to fix the portfolio weights to the point $\eta \in \Delta^n$. As a special case consider the **equally-weighted portfolio** $\mathfrak{p}_e : \Delta^n \rightarrow \Delta^n$, $\mathfrak{p}_e \equiv \eta \equiv \left(\frac{1}{n}, \dots, \frac{1}{n}\right)$.

As third example, which is the first non-trivial mentioned here, the **diversity-weighted portfolio** is defined as $\mathfrak{p}_d : \Delta^n \rightarrow \Delta^n$, $\mathfrak{p}_d(p) = (\mathfrak{p}_1(p), \dots, \mathfrak{p}_n(p))$ with $\mathfrak{p}_i(p) := \frac{\sqrt{p_i}}{\sum_{j=1}^n \sqrt{p_j}}$. This can be seen as a *more compressed* version of the market portfolio: the number of shares of the more expensive stocks will be reduced slightly, whereas the number of shares of the less expensive stocks will be increased.

3.1.2 Performance of Portfolios

Given some stocks $X_1(t), \dots, X_n(t)$, e.g. the S&P500, we want to find a portfolio $\mathfrak{p}(t) := \mathfrak{p}(\mu(t)) = (\mathfrak{p}_1(t), \dots, \mathfrak{p}_n(t))$ (renewed e.g. every day) that *outperforms the market*. To check if our portfolio $\mathfrak{p}$ outperforms the market, we take a look at the **relative value process** (in discrete time) given by

$$\frac{V(t+1)}{V(t)} := 1 + \left\langle \frac{\mathfrak{p}(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \geq 0, \quad (23)$$

where non-negativity can be seen e.g. from formula (25). The relative value process $V(t)$ in (23) is defined as we want it to be, as we think of it as

$$V(t) = \frac{V_{\mathfrak{p}}(t)}{V_{\mu}(t)} = \frac{\text{value at time } t \text{ of \$1 invested in the portfolio } \mathfrak{p}}{\text{value at time } t \text{ of \$1 invested in the market portfolio } \mu}. \quad (24)$$

This is shown in the following short informative lemma.

Lemma 40. *The relative value process (24) fulfils (23).*

Proof. CLAIM:

$$\frac{V(t+1)}{V(t)} = \sum_{i=1}^n \mathfrak{p}_i(\mu(t)) \frac{\mu_i(t+1)}{\mu_i(t)}. \quad (25)$$

Then the assertion follows because

$$\begin{aligned} 1 + \left\langle \frac{\mathbf{p}(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle &= 1 + \left\langle \frac{\mathbf{p}(\mu(t))}{\mu(t)}, \mu(t+1) \right\rangle - \langle \mathbf{p}(\mu(t)), 1 \rangle \\ &= \sum_{i=1}^n \mathbf{p}_i(\mu(t)) \frac{\mu_i(t+1)}{\mu_i(t)}, \end{aligned}$$

using the fact $\sum_{i=1}^n \mathbf{p}_i = 1$ (follows from $\mathbf{p} \in \overline{\Delta^n}$).

So let's prove the CLAIM. First note that the ratio $\frac{X_i(t+1)}{X_i(t)}$ is the relative growth/decrease of the i 'th stock from $t \mapsto t+1$. Therefore, obviously

$$V_{\mathbf{p}}(t+1) = V_{\mathbf{p}}(t) \sum_{i=1}^n \mathbf{p}_i(\mu(t)) \frac{X_i(t+1)}{X_i(t)}$$

and

$$V_{\mu}(t+1) = V_{\mu}(t) \sum_{i=1}^n \mu_i(t) \frac{X_i(t+1)}{X_i(t)}.$$

Using this and $\mu_i(t)S(t) = X_i(t)$ for every t we get

$$\begin{aligned} V(t+1) &= \frac{V_{\mathbf{p}}(t+1)}{V_{\mu}(t+1)} = \frac{V_{\mathbf{p}}(t) \sum_{i=1}^n \mathbf{p}_i(\mu(t)) \frac{\mu_i(t+1)S(t+1)}{\mu_i(t)S(t)}}{V_{\mu}(t) \sum_{i=1}^n S(t)^{-1} X_i(t+1)} \\ &= V(t) \frac{S(t+1) \sum_{i=1}^n \mathbf{p}_i(\mu(t)) \frac{\mu_i(t+1)}{\mu_i(t)}}{\sum_{i=1}^n X_i(t+1)} \end{aligned}$$

hence the CLAIM follows using again $S(t+1) = \sum_{i=1}^n X_i(t+1)$. $\square$

A similar proof can be found in [PW13].

Remark 41. An explicit formula for the relative value process, using $V(0) = 1$ and (23) inductively, is given by:

$$V(m+1) = \prod_{t=0}^m \left(1 + \left\langle \frac{\mathbf{p}(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right), \quad m \in \mathbb{N}_0. \quad (26)$$

Next, the term '*outperforming the market*' has to be made precise. Here a version of *arbitrage* comes into play:

Definition 42 ([PW14, Definition 2]). A portfolio $\mathbf{p} : \Delta^n \rightarrow \overline{\Delta^n}$ is called a **pseudo-arbitrage (PA)** on a mb subset $K \subseteq \Delta^n$ if the following properties hold:

- (i) $\exists C = C(K, \mathbf{p}) \geq 0$ such that whenever $(\mu(t))_{t \geq 0} \subseteq K \Rightarrow \log V(t) \geq -C \forall t \geq 0$ and
- (ii) $\exists \text{seq } (\mu(t))_{t \geq 0} \subset K$ along which $\lim_{t \rightarrow \infty} \log V(t) = \infty$.

In words of Pal & Wong [PW14, Page 3]: "the downside risk is uniformly bounded below regardless of the market movements in a fixed region, and there is a possibility of unbounded gain."

Remark 43. As easy examples of subsets $K \subseteq \Delta^n$ in Definition 42, one can think of the simplex with cut corners (21) or the simplex without border zone (22). (21) represents a market where non of the stocks can be more worth than $100 \cdot (1 - \varepsilon)$ percent of the whole market. Similarly, (22) represents a market where no stock is allowed to fall below the $100 \cdot \kappa$ -percentile of the market.

3.1.3 Multiplicative Cyclical Monotonicity

Now we define some property which is similar to part (i) of Definition 42 with $C \equiv 0$ on the one hand. This property will later be used to characterise PA's. On the other hand it is a multiplicative form of *cyclical monotonicity*, recall Definition 15.

Definition 44 ([PW14, Definition 3]). A portfolio $\mathfrak{p}$ satisfies the **multiplicative cyclical monotonicity (MCM) condition** if over any cycle $(\mu(t))_{t=0}^m \subseteq \Delta^n$ (i.e. $\mu(m+1) = \mu(0)$) it holds $V(m+1) \geq 1$, i.e.

$$\prod_{t=0}^m \left(1 + \left\langle \frac{\mathfrak{p}(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right) \geq 1. \quad (27)$$

Also, $\mathfrak{p}$ satisfies the MCM condition on a convex subset $K \subseteq \Delta^n$ if (27) holds over any cycle $(\mu(t))_{t=0}^m \subseteq K$.

Hereafter we will sometimes say, that a portfolio is MCM, meaning that it fulfils the MCM condition.

Remark 45. Moreover we will call a set $\Gamma \subseteq \Delta^n \times \overline{\Delta^n}$ **MCM** if for any cycle $((p_t, p'_t))_{t=0}^m \subseteq \Gamma$ we have

$$\prod_{t=0}^m \left(1 + \left\langle \frac{p'_t}{p_t}, p_{t+1} - p_t \right\rangle \right) \geq 1.$$

This is an extension of Definition 44 for multi-valued mappings since any set $\Gamma \subseteq \Delta^n \times \overline{\Delta^n}$ can be represented as multi-valued mapping $\mathfrak{p} : \Delta^n \rightarrow \overline{\Delta^n}$ via $\Gamma = \text{Graph}(\mathfrak{p})$. This analogy can be considered throughout the remaining text.

Remark 46 (Important note on performance). Assume a portfolio $\mathfrak{p}$ *underperforms the market* in the sense that $\mathfrak{p}$ is not MCM, i.e. there is a cycle $(\mu(t))_{t=0}^{m+1}$, $\mu(m+1) := \mu(0)$, such that for some $\alpha \in (0, 1)$

$$\prod_{t=0}^m \left(1 + \left\langle \frac{\mathfrak{p}(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right) < \alpha$$

holds. Then for the cycles $\mu^{(k)}$ defined by $\mu^{(k)}(t) := \mu(t) \bmod m$ follows

$$0 \leq \prod_{t=0}^{m \cdot k} \left(1 + \left\langle \frac{\mathfrak{p}(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right) < \alpha^k \rightarrow 0 \text{ as } k \rightarrow \infty.$$

Therefore, a portfolio that is not MCM can not be a PA, since the downside risk is not bounded. This of course is equivalent to: $\text{PA} \Rightarrow \text{MCM}$. We also note here, that the converse is not true in general. This is shown in the next example.

Example 47. The market portfolio is MCM: using that

$$\left\langle \frac{\mathfrak{p}(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle = \langle 1, \mu(t+1) - \mu(t) \rangle = 0$$

holds for any market weight sequence (and therefore for any cycle) $(\mu(t))_{t \geq 1}$ in Δ^n , we infer $V(m+1) = 1$ (cf. Definition 44) for m arbitrary. But the market portfolio is not a PA, because (ii) of Definition 42 is not fulfilled: by the same argument as before we have that $V(t) = 1 \ \forall t \geq 0$ and so $\lim_{t \rightarrow \infty} \log V(t) = 0$, independent of the market sequence. (This is more than obvious. Outperforming the market is impossible if our portfolio is the market itself!)

In Remark 46 and Example 47 we discussed PA's over the whole simplex Δ^n . The same arguments work also for any convex subset of Δ^n . The resulting statement is the content of the next proposition.

Proposition 48. *Let $\mathbf{p} : \Delta^n \rightarrow \overline{\Delta^n}$ be a portfolio. If $\mathbf{p}$ is a PA over a convex subset $K \subseteq \Delta^n$, then $\mathbf{p}$ is MCM on K . The converse is not true in general.*

In [PW14, Propostion 4] the following connection between concave functions and the MCM property can be found:

Proposition 49. *Let $\mathbf{p} : \Delta^n \rightarrow \overline{\Delta^n}$ be a multi-valued map (e.g. a portfolio). Then $\mathbf{p}$ is MCM iff $\exists$ concave $\Phi : \Delta^n \rightarrow (0, \infty)$ such that*

$$\forall p, q \in \Delta^n : 1 + \left\langle \frac{\mathbf{p}(p)}{p}, q - p \right\rangle \geq \frac{\Phi(q)}{\Phi(p)}. \quad (28)$$

In this case we call Φ a **generating function** of $\mathbf{p}$ and also use the term ' $\mathbf{p}$ is generated by Φ '.

Lemma 50. *If a portfolio $\mathbf{p} : \Delta^n \rightarrow \overline{\Delta^n}$ is MCM, then there exists a concave function $\Phi : \Delta^n \rightarrow (0, \infty)$ such that $\frac{\mathbf{p}(p)}{p}$ is a super-gradient of $\log \Phi$ at p for every $p \in \Delta^n$.*

Proof. Take $p \in \Delta^n$ arbitrary but fixed. By Proposition 49 there is a concave function $\Phi : \Delta^n \rightarrow (0, \infty)$ such that $\forall p \in \Delta^n : \Phi(p) + \left\langle \Phi(p) \frac{\mathbf{p}(p)}{p}, q - p \right\rangle \geq \Phi(q)$, i.e. by Definition 21, $\Phi(p) \frac{\mathbf{p}(p)}{p}$ is a super-gradient of Φ at p . By (11), $\frac{\mathbf{p}(p)}{p}$ is a super-gradient of $\log \Phi$ at p , as claimed. $\square$

Note that in the last proof we had shown, that condition (28) is equal to $\Phi(p) \frac{\mathbf{p}(p)}{p} \in \partial \Phi(p)$ for every $p \in \Delta^n$.

Above Proposition 49 is an analogue to the Theorem of Rockafellar 17 for MCM sets; consider the following version of Proposition 49.

Theorem 51 (Rockafellar-Type Theorem for MCM sets). *If a set $\Gamma \subseteq \Delta^n \times \overline{\Delta^n}$ is MCM, then $\exists$ concave $\Phi : \Delta^n \rightarrow (0, \infty)$ such that $(p, \frac{\mathbf{p}'}{p}) \in \text{Graph}(\partial(\log \Phi))$ for all $(p, p') \in \Gamma$. Conversely let $\Gamma \subseteq \Delta^n \times \overline{\Delta^n}$ be a set. If $\exists$ concave $\Phi : \Delta^n \rightarrow (0, \infty)$ is such that $(p, \frac{\mathbf{p}'}{p}) \in \text{Graph}(\partial(\log \Phi))$ for all $(p, p') \in \Gamma$ and $\partial(\log \Phi) = \frac{1}{\Phi} \partial \Phi$ (e.g. $\Phi \in \mathcal{C}^1(\Delta^n)$), then Γ is MCM.*

Proof. SKETCH. By rewriting Proposition 49 for MCM sets, we have that $\Gamma \subseteq \Delta^n \times \overline{\Delta^n}$ is MCM iff

$$\forall (p, p'), (q, q') \in \Gamma : 1 + \left\langle \frac{p'}{p}, q - p \right\rangle \geq \frac{\Phi(q)}{\Phi(p)}.$$

Like in the proof of Lemma 50, $\frac{p'}{p} \in \frac{1}{\Phi(p)} \partial \Phi(p)$. This shows the statement if $\partial(\log \Phi) = \frac{1}{\Phi} \partial \Phi$ holds. For the general case use (11), like in the proof of Lemma 50. □

We now give two examples of portfolios, which have generating functions. Once we had found such a generating function for a portfolio, Proposition 49 will imply that this portfolio is MCM. For one example we will show this implication explicitly.

Example 52. We claim that the concave function $\Phi : \Delta^n \rightarrow (0, \infty)$, $\Phi(p) = \prod_{i=1}^n \sqrt[n]{p_i}$ is a generating function for the equally-weighted portfolio $\mathbf{p}_e = (\frac{1}{n}, \dots, \frac{1}{n})$. This can be seen using the *inequality of arithmetic and geometric means*: for $p, q \in \Delta^n$

$$1 + \left\langle \frac{\mathbf{p}_e(p)}{p}, q - p \right\rangle = \frac{1}{n} \sum_{i=1}^n \left(\frac{q_i}{p_i} \right) \geq \sqrt[n]{\prod_{i=1}^n \left(\frac{q_i}{p_i} \right)} = \frac{\prod_{i=1}^n \sqrt[n]{q_i}}{\prod_{i=1}^n \sqrt[n]{p_i}} = \frac{\Phi(q)}{\Phi(p)}.$$

We now show that $\mathbf{p}_e$ is MCM, although we know this already by Proposition 49. In fact, this is how one direction of Proposition 49 is established.

We want to show $V(m+1) \geq 1$ for any market cycle $(\mu_t)_{t=0}^m$. Therefore let $(\mu_t)_{t=0}^m$ be an arbitrary cycle in Δ^n , then by using that Φ is a generating function of $\mathbf{p}_e$

$$\begin{aligned} V(m+1) &= \prod_{t=0}^m \left(1 + \left\langle \frac{\mathbf{p}_e(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right) \\ &\geq \prod_{t=0}^m \frac{\Phi(\mu(t+1))}{\Phi(\mu(t))} = 1, \end{aligned}$$

because $\mu(m+1) = \mu(0)$.

Now we show, that $\mathbf{p}_e$ is a PA on a appropriate subset of Δ^n . Take $\kappa \in (0, \frac{1}{n})$ and define $K := \Delta_b^n(\kappa)$ to be the simplex without border zone (cf. (22)). Then $\mathbf{p}_e$ is a PA on K . Define $C := \log \frac{1-\kappa}{\kappa} \geq 0$ (for $n \geq 2$), then we have for every market sequence $(\mu(t))_{t \geq 0}$ in K :

$$\begin{aligned} \log V(m) &= \log \left(\prod_{t=0}^{m-1} \frac{1}{n} \sum_{i=0}^t \frac{\mu_i(t+1)}{\mu_i(t)} \right) \geq \log \left(\prod_{t=0}^{m-1} \sqrt[n]{\prod_{i=1}^n \frac{\mu_i(t+1)}{\mu_i(t)}} \right) \\ &= \log \sqrt[n]{\prod_{i=1}^n \prod_{t=0}^{m-1} \frac{\mu_i(t+1)}{\mu_i(t)}} = \log \sqrt[n]{\prod_{i=1}^n \frac{\mu_i(m)}{\mu_i(0)}} \\ &\geq \log \sqrt[n]{\prod_{i=1}^n \frac{\kappa}{1-\kappa}} = \log \frac{\kappa}{1-\kappa} = -C. \end{aligned}$$

Therefore, (i) of Definition 42 is satisfied. To check (ii), we have to find a market sequence in K such that $\lim_{t \rightarrow \infty} \log V(t) = \infty$ along this sequence. Choose a start point $\mu(0) \in K$ such that $\mu(0) \neq (\frac{1}{n}, \dots, \frac{1}{n})$. Without loss of generality let $\mu_1(0) := \max_i \mu_i(0)$ and $\mu_2(0) := \min_i \mu_i(0)$. Then $\mu_1(0) > \mu_2(0)$. Now define $\mu_i(t) := \mu_i(0)$ for all $i \in \{3, \dots, n\}$ and all $t \geq 0$. The missing two values are defined as follows: $\mu_i(t) = \mu_i(0)$ for $t \equiv 0 \pmod{2}$ and $\mu_1(t) = \mu_2(0) \wedge \mu_2(t) = \mu_1(0)$ for $t \equiv 1 \pmod{2}$. This means that this sequence interchanges just the first two coordinates of our starting point. For this sequence we infer

$$\sum_{i=1}^n \frac{\mu_i(t+1)}{\mu_i(t)} = \frac{\mu_1(t+1)}{\mu_1(t)} + \frac{\mu_2(t+1)}{\mu_2(t)} + (n-2) = \frac{\mu_1(0)}{\mu_2(0)} + \frac{\mu_2(0)}{\mu_1(0)} + (n-2),$$

regardless if t is odd or even. Because $\frac{a}{b} + \frac{b}{a} > 2$ holds for every two positive numbers $a \neq b$, it follows that

$$\sum_{i=1}^n \frac{\mu_i(t+1)}{\mu_i(t)} > n,$$

independent of t . This means, there is an $\eta > 0$ such that $\frac{1}{n} \sum_{i=1}^n \frac{\mu_i(t+1)}{\mu_i(t)} = 1 + \eta$ and therefore we get along this subsequence

$$\log V(m) = \log \left(\prod_{t=0}^{m-1} \frac{1}{n} \sum_{i=1}^n \frac{\mu_i(t+1)}{\mu_i(t)} \right) = \log(1 + \eta)^m \xrightarrow{m \rightarrow \infty} \infty.$$

Example 53. $\Phi : \Delta^n \rightarrow (0, \infty)$, $\Phi(p) = (\prod_{i=1}^n \sqrt[n]{p_i})^2$ is a generating function of the diversity-weighted portfolio p_d :

$$1 + \left\langle \frac{p_e(p)}{p}, q - p \right\rangle = \sum_{i=1}^n \frac{p_i(p)}{p_i} q_i = \frac{1}{\sum_{j=1}^n \sqrt{p_j}} \sum_{i=1}^n \frac{q_i}{\sqrt{p_i}} \geq \frac{(\sum_{i=1}^n \sqrt{q_i})^2}{(\sum_{j=1}^n \sqrt{p_j})^2},$$

where the last inequality is the *Cauchy-Schwarz inequality*:

$$\left(\sum_{i=1}^n \sqrt{q_i} \right)^2 = \left(\sum_{i=1}^n \frac{\sqrt{q_i}}{p_i^{\frac{1}{4}}} \cdot p_i^{\frac{1}{4}} \right)^2 \leq \sum_{i=1}^n \frac{q_i}{\sqrt{p_i}} \sum_{j=1}^n \sqrt{p_j}.$$

Therefore p_d is MCM.

3.2 AN OPTIMAL TRANSPORT CHARACTERISATION THEOREM FOR PSEUDO-ARBITRAGE

3.2.1 Set-up of the Optimal Transport

We now want to use the knowledge about Optimal Transport to find portfolios which are PA. Therefore we first identify how our framework looks like.

Define the cost function as follows:

$$c : \overline{\Delta}^n \times \widetilde{\mathbb{R}}^n \rightarrow \mathbb{R}, \quad (p, h) \mapsto \log \left(\sum_{i=1}^n e^{h_i} p_i \right), \quad (29)$$

where $\widetilde{\mathbb{R}}^n := [-\infty, +\infty)^n \setminus \{(-\infty, \dots, -\infty)\}$. This cost function is clearly continuous and therefore lsc (and also usc). Furthermore, given a mb transport map $T : \Delta^n \rightarrow \widetilde{\mathbb{R}}^n$ we define the portfolio $\mathfrak{p} : \Delta^n \rightarrow \overline{\Delta}^n$ through

$$\mathfrak{p}_i(\mu) := \frac{e^{T_i(\mu)} \mu_i}{\sum_{i=1}^n e^{T_i(\mu)} \mu_i}, \quad i = 1, \dots, n. \quad (30)$$

This is a portfolio, since T is mb (and therefore is $\mathfrak{p}$) and obviously $\mathfrak{p}(\mu) \in \overline{\Delta}^n$ for every $\mu \in \Delta^n$.

Next, $\overline{\Delta}^n$ and $\widetilde{\mathbb{R}}^n$ are Polish spaces: $\overline{\Delta}^n$ is a closed subset of $\mathbb{R}^n$ (which itself is Polish) and so $\overline{\Delta}^n$ is Polish. To see why $\widetilde{\mathbb{R}}^n$ is Polish, one may use Theorem 3. Define the following open sets for $n \in \mathbb{N}$

$$A_n := \left(-1 - \frac{1}{n}, 1 \right)^n \setminus \{(-1, \dots, -1)\}.$$

Then $\bigcap_{n \in \mathbb{N}} A_n = [-1, 1]^n \setminus \{(-1, \dots, -1)\}$ which is a G_δ -subset of $\mathbb{R}^n$ and therefore Polish by Theorem 3. Now $[-1, 1]^n \setminus \{(-1, \dots, -1)\} \cong \widetilde{\mathbb{R}}^n$, and thus we verified that $\widetilde{\mathbb{R}}^n$ is Polish.

3.2.2 The Characterisation Theorem

Theorem 54 (Characterisation – Part 1 [PW14, Theorem 2]). *Let $\mathbb{P}$ and $\mathbb{Q}$ be (Borel) probability measures on $\overline{\Delta}^n$ and $\widetilde{\mathbb{R}}^n$, respectively. Define c as in (29) and suppose that the problem*

$$\inf_{\pi \in \Pi(\mathbb{P}, \mathbb{Q})} \int_{\overline{\Delta}^n \times \widetilde{\mathbb{R}}^n} c(x, y) \, d\pi(x, y) \quad (31)$$

has a solution $\mathcal{R} \in \Pi(\mathbb{P}, \mathbb{Q})$. Choose a mb function $T : \overline{\Delta}^n \rightarrow \widetilde{\mathbb{R}}^n$ such that $(\mu, T(\mu)) \in \text{Supp}(\mathcal{R})$ for all $\mu \in \overline{\Delta}^n$ and define the portfolio $\mathfrak{p}$ via

$$\frac{\mathfrak{p}_i}{\mu_i} = \frac{e^{h_i}}{\sum_{i=1}^n e^{h_i} \mu_i}, \quad i = 1, \dots, n \quad (32)$$

with $h = T(\mu)$. Then there is a concave function $\Phi : \Delta^n \rightarrow [0, \infty)$ such that

$$\forall p \in \Delta^n : \frac{\mathfrak{p}(p)}{p} \text{ is a super-gradient of } \log \Phi \text{ at } p. \quad (33)$$

Furthermore, $\mathfrak{p}$ is a PA over a convex set $K \subseteq \Delta^n$ whenever the following two conditions are true

- (i) $\Phi|_K$ is not affine and
- (ii) $\inf_{p \in K} \Phi(p) > 0$.

Proof. Let $\mathcal{R} \in \Pi(\mathbb{P}, \mathbb{Q})$ be an optimal transference plan for the cost function c with finite cost, $\left| \int_{\overline{\Delta^n} \times \widetilde{\mathbb{R}}^n} c(x, y) d\mathcal{R}(x, y) \right| < \infty$. We show that $\text{Supp}(\mathcal{R})$ is c -CM. We want to use an approximation argument, therefore define for $m \in \mathbb{N}$ sets

$$L_m := \{h \in [-\infty, \infty)^n \setminus \{(-\infty, \dots, -\infty)\} : \forall i \in \{1, \dots, n\} h_i \geq -m\}.^5 \quad (34)$$

For $m \in \mathbb{N}$ we define new measures on the sets L_m : the restricted measure $\mathcal{T}_m = \mathcal{R}|_{\overline{\Delta^n} \times L_m}$ and the normalized restricted measure $\mathcal{R}_m(\cdot) = \frac{\mathcal{T}_m(\cdot)}{\mathcal{T}_m(\overline{\Delta^n} \times L_m)}$, i.e. $\mathcal{R}_m$ is again a probability measure.

We note the following properties of L_m , $\mathcal{T}_m$, $\mathcal{R}_m$ and $\mathcal{R}$:

1. The sequence $(L_m)_{m \geq 1}$ is strictly increasing with limit $\widetilde{\mathbb{R}}^n$:

$$L_m \nearrow \widetilde{\mathbb{R}}^n \text{ as } m \text{ increases.}$$

2. By this and by continuity of measures we have

$$\mathcal{R}(\overline{\Delta^n} \times L_m) \nearrow \mathcal{R}(\overline{\Delta^n} \times \widetilde{\mathbb{R}}^n) = 1 \text{ as } m \text{ increases,}$$

therefore $\mathcal{T}_m$ and $\mathcal{R}_m$ are well-defined for m large enough.

3. Clearly $\mathcal{T}_m \leq \mathcal{R}$. Hence by the *Restriction Property for Optimal Transference Plans* ([Vil09, Theorem 4.6]) the probability measure $\mathcal{R}_m$ is optimal for

$$\inf_{\pi \in \Pi(\mathbb{P}_m, \mathbb{Q}_m)} \int_{\overline{\Delta^n} \times L_m} c(x, y) d\pi(x, y), \quad (35)$$

where $\mathbb{P}_m$ and $\mathbb{Q}_m$ are the marginals of $\mathcal{R}_m$, respectively.

Now c is continuous and bounded below on $\overline{\Delta^n} \times L_m$: for $(p, h) \in \overline{\Delta^n} \times L_m$ deduce $c(p, h) = \log(\sum_{i=1}^n e^{h_i} \mu_i) \geq \log(e^{-m} \sum_{i=1}^n \mu_i) = -m$. Since $\mathcal{R}_m$ is optimal in (35) and the total cost is finite, Theorem 37 implies that $\text{Supp}(\mathcal{R}_m)$ is c -CM. But $m \in \mathbb{N}$ was arbitrary (but large enough), so $\text{Supp}(\mathcal{R})$ is c -CM.

Now take a mb function $T : X \rightarrow Y$ such that $(\mu, T(\mu)) \in \text{Supp}(\mathcal{R})$ for all μ . This is, the graph of T is c -CM, so by Lemma 55 the portfolio defined via (32) is MCM. By Lemma 50 there exists a concave function $\Phi : \Delta^n \rightarrow (0, \infty)$ such that $\frac{p(p)}{p}$ is a super-gradient of $\log \Phi$ at p for all $p \in \Delta^n$. This justifies (33).

It remains to show, that if furthermore (i) and (ii) are fulfilled for a convex set $K \subseteq \Delta^n$, $\mathfrak{p}$ is a PA over K . Therefore, take $K \subseteq \Delta^n$ arbitrary but convex and assume $\mathfrak{p}$ is not a PA over K . This means at least one of the following statements must happen (cf. Definition 42):

1. $\forall C = C(K, \mathfrak{p}) > 0 \exists (\mu(t))_{t \geq 1} \subset K \exists \tau \geq 1 : \log V(\tau) < -C$
2. $\forall \text{ sequences } (\mu(t))_{t \geq 1} \subset K : \lim_{t \rightarrow \infty} \log V(t) < \infty \text{ along } (\mu(t))_{t \geq 1}.$

⁵ Note that $L_m = [-m, \infty)^n$. By writing L_m like in (34) we remark, that this really tends to $\widetilde{\mathbb{R}}^n$ in the limit (defined by the metric induced by the Polish space $\widetilde{\mathbb{R}}^n$).

CASE 1: assume 1. holds. By (ii), $\inf_t \Phi(\mu(t)) > 0$, hence $\exists C' \geq 0 : \log \Phi(\mu(t)) > -C' \forall t \geq 0$. Define $C := C' + \log \Phi(\mu(0))$ and find τ such that $\log V(\tau) < -C$, then by the decomposition of [PW14, Lemma 7] follows

$$-C > \log V(\tau) = \log \Phi(\mu(\tau)) - \log \Phi(\mu(0)) + A(\tau) > -C$$

since A is non-negative (cf. [PW14, Lemma 7]) and this is a contradiction. So 1. can not be true. Therefore consider

CASE 2: assume 2. holds. Since Φ is a positive concave function, it is bounded above, and hence so is $\log \Phi$ and hence $\log \frac{\Phi(\cdot)}{\Phi(\mu(0))}$ is bounded above. From the last point 1. we also know, that $\log \frac{\Phi(\cdot)}{\Phi(\mu(0))}$ is bounded below *independent of choices of the market sequence* $(\mu(t))$. Since $\Phi|_K$ is not affine, we can find a market sequence $(\mu(t))_{t \geq 1} \subset K$ such that the process $A(t)$ tends to ∞ as time t increases. Collecting things together we infer

$$\infty > \lim_{t \rightarrow \infty} \log V(t) = \lim_{t \rightarrow \infty} \log \frac{\Phi(\mu(t))}{\Phi(\mu(0))} + \lim_{t \rightarrow \infty} A(t) = \infty,$$

which is again a contradiction. So therefore $\mathfrak{p}$ has to be a PA over K and the proof of Theorem Part 1 is complete. $\square$

The following lemma was used to prove Theorem 54. For a simple proof, see the proof of [PW14, Lemma 11].

Lemma 55. *Let $h : \Delta^n \rightarrow \tilde{\mathbb{R}}^n$. The graph of h is c -CM iff the portfolio $\mathfrak{p}$ defined in (32) is MCM.*

Lemma 55 is also useful to give a further version of Proposition 49 (cf. Theorem 51):

Theorem 56 (Rockafellar-Type Theorem for c -CM sets). *If a set $\mathcal{H} \subseteq \Delta^n \times \tilde{\mathbb{R}}^n$ is c -CM, then $\exists$ concave $\Phi : \Delta^n \rightarrow (0, \infty)$ such that $(p, \frac{e^h}{E(p,h)}) \in \text{Graph}(\partial(\log \Phi))$ for all $(p, h) \in \mathcal{H}$, where $E(p, h) = \sum_{j=1}^n p_j e^{h_j}$.*

Conversely let $\mathcal{H} \subseteq \Delta^n \times \tilde{\mathbb{R}}^n$ a set. If $\exists$ concave $\Phi : \Delta^n \rightarrow (0, \infty)$ such that $(p, \frac{e^h}{E(p,h)}) \in \text{Graph}(\partial(\log \Phi))$ for all $(p, h) \in \mathcal{H}$ and $\partial(\log \Phi) = \frac{1}{\Phi} \partial \Phi$, then $\mathcal{H}$ is c -CM.

Proof. Define

$$\Gamma := \left\{ (p, p') : p'_i := \frac{p_i e^{h_i}}{E(p, h)} \forall (p, h) \in \mathcal{H} \right\} \subseteq \Delta^n \times \overline{\Delta^n}.$$

Then Γ is MCM by Lemma 55. The proof is complete by using Theorem 51 and the equality $\frac{p'}{p} = \frac{e^h}{E(p, h)}$. $\square$

This Lemma was also used in the proof of Theorem 54:

Lemma 57 ([PW14, Lemma 7]). *Let $\mathfrak{p}$ be a portfolio that is generated by a positive concave function Φ and let $(\mu(t))_{t \geq 0}$ be some market weight process (i.e. a sequence). Then there is a non-decreasing process $(A(t))_{t \geq 0}$ with $A(0) = 0$ such that the relative value process $V(t)$ satisfies the decomposition*

$$\log V(t) = \log \frac{\Phi(\mu(t))}{\Phi(\mu(0))} + A(t) \quad \forall t \geq 0. \quad (36)$$

Moreover, Φ is affine iff $A \equiv 0$, independent of the market weight sequence $(\mu(t))_{t \geq 0}$.

The proof shows how $A(t)$ explicitly looks, so we give the proof:

Proof. First define $A(t) := \log V(t) - \log \frac{\Phi(\mu(t))}{\Phi(\mu(0))}$ for all $t \geq 0$. Using equation (26) and expanding $\frac{\Phi(\mu(t))}{\Phi(\mu(0))} = \prod_{k=0}^{t-1} \frac{\Phi(\mu(k+1))}{\Phi(\mu(k))}$ one can write

$$A(t) = \sum_{k=0}^{t-1} \log \left(1 + \left\langle \frac{\mathfrak{p}(\mu(k))}{\mu(k)}, \mu(k+1) - \mu(k) \right\rangle \right) - \sum_{k=0}^{t-1} \log \frac{\Phi(\mu(k+1))}{\Phi(\mu(k))}.$$

By defining

$$\mathbb{L}(q|p) := \log \left(1 + \left\langle \frac{\mathfrak{p}(p)}{p}, q - p \right\rangle \right) - \log \frac{\Phi(q)}{\Phi(p)}$$

for any $p, q \in \Delta^n$, we get

$$A(t) = \sum_{k=0}^{t-1} \mathbb{L}(\mu(k+1)|\mu(k)).$$

We claim that

(i) $\mathbb{L}(\cdot|\cdot) \geq 0$ and

(ii) $\mathbb{L}(q|p) = 0$ iff Φ is affine on the straight line joining two points $p \neq q \in \Delta^n$.

Then $A(t)$ is non-decreasing by (i) and $A(0) = 0$ is trivially true. By (ii) Φ is affine iff $\mathbb{L}(\mu(k+1)|\mu(k)) = 0$ for any choice of market weights and for every $k \geq 0$, so $A \equiv 0$.

To finish the proof completely we owe to prove the claims: Non-negativity follows from (28) since $\mathfrak{p}$ is generated by the concave function Φ . For (ii) consider first $1 + \langle \frac{\mathfrak{p}(p)}{p}, q - p \rangle = \langle \frac{\mathfrak{p}(p)}{p}, q \rangle$. So if $\mathbb{L}(q|p) = 0$, we have $\Phi(q) = \langle \Phi(p) \frac{\mathfrak{p}(p)}{p}, q \rangle$, so therefore Φ is affine on the straight line connecting p and q . For the remaining direction, let Φ be affine on the straight line joining p and q . This means

$$\Phi(q) = \Phi(p) + \langle \nabla \Phi(p), q - p \rangle, \quad (37)$$

where $\nabla \Phi(p)$ is the unique super-gradient of Φ at p (existence and uniqueness follows, because Φ is affine and hence differentiable). On the other hand, $\Phi(p) \frac{\mathfrak{p}(p)}{p}$ is also a super-gradient of Φ at p , since $\mathfrak{p}$ is generated by Φ , cf. the proof of Lemma 50. Hence by uniqueness $\nabla \Phi(p) = \Phi(p) \frac{\mathfrak{p}(p)}{p}$, thus (37) turns into

$$\frac{\Phi(q)}{\Phi(p)} = 1 + \left\langle \frac{\mathfrak{p}(p)}{p}, q - p \right\rangle,$$

i.e. $\mathbb{L}(q|p) = 0$, as claimed. $\square$

The converse to Theorem 54 is now much easier to prove:

Theorem 58 (Characterisation – Part 2 [PW14, Theorem 2]). *Suppose a portfolio $\mathfrak{p}$ is a PA over a convex set $K \subseteq \Delta^n$ and assume $\log \frac{\mathfrak{p}(p)}{p} := \left(\log \frac{\mathfrak{p}_1(p)}{p_1}, \dots, \log \frac{\mathfrak{p}_n(p)}{p_n} \right)$ is bounded below coordinate-wise for every $p \in \Delta^n$. Define the function $T : \Delta^n \rightarrow \mathbb{R}^n$ with coordinates*

$$T_i(\mu) := \log \frac{\mathfrak{p}_i(\mu)}{\mu_i} \text{ for } i = 1, \dots, n \quad (38)$$

and consider the coupling $(\mu, T(\mu))$. Let $\mathbb{P}$ be a probability measure on K and define $\mathbb{Q} := T\#\mathbb{P}$ to be the push forward measure of $\mathbb{P}$. Then the coupling $(\mu, T(\mu))$ solves the Monge Optimal Transport Problem $\inf \mathbb{E}[c(\mu, T(\mu))]$, where $c(\mu, h) := \log \left(\sum_{i=1}^n e^{h_i} \mu_i \right)$. In particular $\mathbb{P}$ can be taken as the normalized Lebesgue measure on K .

Proof. By assumption, $\exists m \in \mathbb{N} : \log \frac{\mathfrak{p}(p)}{p} \geq -m$ and hence $\mathbb{Q}$ is supported on $L_m = [-m, \infty)^n$. On $\overline{\Delta^n} \times L_m$, c is continuous and bounded below. Now, $\mathfrak{p}$ is a PA over $K \subseteq \Delta^n$ and hence $\mathfrak{p}$ is MCM over K , by Proposition 48. Because of the special form of the cost function $c(\mu, h) = \log \left(\sum_{i=1}^n e^{h_i} \mu_i \right)$, $\text{Graph}(T|_K)$ is a c-CM set by Lemma 55. (By the definition of T , $\mathfrak{p}$ is given by $\mathfrak{p}_i(\mu) = \mu_i e^{T_i(\mu)}$ and note that $\sum_{i=1}^n \mu_i e^{T_i(\mu)} = 1$, because $\mathfrak{p} \in \overline{\Delta^n}$. Therefore Lemma 55 is indeed applicable). But now, by Theorem 37 $\text{Graph}(T|_K)$ is optimal in $\inf \mathbb{E}[c(\mu, T(\mu))]$. That one can choose $\mathbb{P}(\cdot) := \frac{\text{Leb}(\cdot)}{\text{Leb}(K)}$ is obvious, since this is clearly a probability measure on K . $\square$

CONNECTIONS AND RÉSUMÉ

As the reader may have noticed in the last section, the *Characterisation Theorems* 54 and 58 can be connected informally with some theorems of Chapter 2, in the sense that the statements are *of the same type*. We mean the following:

Consider Theorem 54, which is basically of the following form: *Suppose that $\mathcal{R} \in \Pi(\mathbb{P}, \mathbb{Q})$ is optimal, choose $(\mu, T(\mu)) \in \text{Supp}(\mathcal{R})$ and define a portfolio $\mathfrak{p}$ via (30).*

$$\Rightarrow \exists \text{ concave } \Phi : \frac{\mathfrak{p}(p)}{p} \in \partial \log \Phi(p) \quad \forall p.$$

Now recall the Knot-Smith Optimality Criterion (Theorem 32), which is of similar type:

Suppose that $\mathcal{R} \in \Pi(\mathbb{P}, \mathbb{Q})$ is optimal

$$\Rightarrow \exists \text{ convex } \varphi : (p, h) \in \partial \varphi(p) \text{ for } d\mathcal{R}\text{-almost all } (p, h).$$

Moreover, we had in Theorem 54, that

$$\begin{aligned} \exists \text{ concave } \Phi : \frac{\mathfrak{p}(p)}{p} \in \partial \log \Phi(p) \quad \forall p \text{ (plus some conditions on } \Phi) \\ \Rightarrow \mathfrak{p} \text{ is a PA and therefore MCM,} \end{aligned}$$

which is an analogue to the Theorem of Rockafellar 17, which says

$$\exists \text{ convex } f \text{ such that } \Gamma \subseteq \partial f \Rightarrow \Gamma \text{ is CM.}$$

Next we take a look at Theorem 58 which was essentially saying:

Suppose $\mathfrak{p}$ is a PA⁶ and let $\mathbb{P}$ be a probability measure. Define T by (38) and set $\mathbb{Q} := T\#\mathbb{P}$.

$$\Rightarrow T \text{ is optimal in the Monge Optimal Transport Problem.}$$

This can be interpreted as a composition of the Theorem of Rockafellar 17 and the Knot-Smith Optimality Criterion (Theorem 32): First we have the following implication

$\Gamma (= \text{Supp}(\mathcal{R}) \text{ for some transference plan } \mathcal{R} \in \Pi(\mathbb{P}, \mathbb{Q})) \text{ is CM}$

$$\Rightarrow \exists \text{ convex } f \text{ such that } \text{Supp}(\mathcal{R}) \subseteq \text{Graph}(\partial f),$$

⁶ $\mathfrak{p}$ therefore is MCM, as we had seen in Proposition 48.

by Theorem 17, and continuing,

$\Rightarrow \mathcal{R}$ is optimal,

by Theorem 32.

So beneath this stone we may call the *Characterisation Theorems* 54 and 58 **Knot-Smith & Rockafellar Type Characterisations Theorems**.

BIBLIOGRAPHY

- [Bil68] Patrick Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1968.
- [HLP56] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, Cambridge, 2nd edition, 1956.
- [Kec95] Alexander S. Kechris. *Classical Descriptive Set Theory*, volume 156 of *Graduate Texts in Mathematics*. Springer, New York, 1995.
- [Kle13] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springer, Berlin Heidelberg, 3rd edition, 2013.
- [PW13] S. Pal and T.-K. L. Wong. Energy, Entropy, and Arbitrage. arXiv:1308.5376 [q-fin.PM], August 2013.
- [PW14] S. Pal and T.-K. L. Wong. The Geometry of Relative Arbitrage. arXiv:1402.3720v4 [q-fin.PM], November 2014.
- [Roc70] Ralph Tyrrell Rockafellar. *Convex Analysis*. Princeton Mathematical Series. Princeton University Press, Princeton, New Jersey, 1970.
- [Vil03] Cédric Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics. American Mathematical Society, Providence, Rhode Island, 2003.
- [Vil09] Cédric Villani. *Optimal Transport: Old and New*, volume 338 of *A Series of Comprehensive Studies in Mathematics*. Springer, Berlin Heidelberg, 2009.

SUPPLEMENTS

DEUTSCHE ZUSAMMENFASSUNG

Die Theorie von optimalem Massetransport findet in vielen Teilgebieten der Mathematik ihre Anwendung. In der vorliegenden Arbeit wird diese Theorie verwendet, um eine Charakterisierung von sogenannten Pseudo-Arbitragen zu finden. Dadurch soll auf das Themengebiet der stochastischen Portfoliotheorie eingegangen werden. Pseudo-Arbitragen sind Portfolios, welche es erlauben, unbeschränkten Gewinn bei beschränktem Risiko zu erzielen.

Kapitel 2 beschäftigt sich eingehender mit der Theorie des optimalen Transports. Anfangs werden einige wenige Grundbegriffe der Wahrscheinlichkeitstheorie kurz beschrieben, mit dem Ziel den Satz von Prohorov zu formulieren, welcher in Abschnitt 2.3 zur Anwendung kommt. Im Abschnitt 2.2 steht der Satz von Rockafellar, der zyklisch monotone Mengen in Verbindung mit konvexen Funktionen bringt, im Fokus. Den Abschluss des zweiten Kapitels bildet ein Absatz zum Thema optimaler Transport, mit dem Ziel eine Charakterisierung optimaler Transportpläne zu gewinnen, der wiederum essentielle Bedeutung im Beweis der Charakterisierung von Pseudo-Arbitragen zukommt.

Anwendung findet die eben genannte Theorie in Kapitel 3, in dem zunächst Pseudo-Arbitragen und der dazu geeignete finanzmathematische Rahmen eingeführt werden. Einige Beispiele in Form von Portfolios veranschaulichen diese Objekte. Es wird gezeigt, dass Pseudo-Arbitragen eine Eigenschaft besitzen, welche unter dem Namen *Multiplicative Cyclical Monotonicity* (zu Deutsch: multiplikative zyklische Monotonie) definiert werden. Diese Eigenschaft kann in Verbindung mit konkaven Funktionen gebracht werden. Jene Verbindung kann somit als Analogon zum Satz von Rockafellar aus Kapitel 2 gesehen werden. Um darauf einzugehen, finden sich in Kapitel 3 zwei Sätze, die als Rockafellar-Type Theoreme bezeichnet werden.

Schließlich wird im Abschnitt 3.2 eine Charakterisierung von Pseudo-Arbitragen über ein geeignetes Transportproblem vorgestellt.

Den Abschluss der Arbeit bildet eine kurze Diskussion über die Parallelen zwischen einigen Sätzen aus den beiden Kapiteln 2 und 3.

LEBENS LAUF

DOMINIK IMGRÜTH, BSc

AUSBILDUNG

Seit 10/2013	Magisterstudium Statistik an der Universität Wien
Seit 10/2012	Masterstudium Mathematik an der Universität Wien
10/2009 – 08/2012	Bachelorstudium Mathematik an der Paris Lodron Universität Salzburg <i>-Bachelorprüfungszeugnis mit Auszeichnung bestanden</i>
09/2003 – 06/2008	HTBLA Vöcklabruck (Höhere technische Bundeslehranstalt) <i>-Ausbildungsschwerpunkt Maschinen- und Anlagentechnik</i>

ZIVILDIENTST

02/2009 – 10/2009	Pflege, Reinigungsarbeiten und administrative Tätigkeiten bei der Lebenshilfe Oberösterreich, Tießenbach 5, 4644 Scharnstein
-------------------	--

FÄHIGKEITEN UND KOMPETENZEN

Sprachkenntnisse:	Deutsch (Muttersprache) Englisch <i>-Gute Kenntnisse in Wort und Schrift</i>
IT-Kenntnisse:	-souverän im Umgang mit Microsoft Office™-Programmen (Word™, Excel™ und PowerPoint™) -Grundkenntnisse im Umgang mit der Java™-Programmierung, erworben in der Lehrveranstaltung "Einführung in die Programmierung" an der Paris Lodron Universität Salzburg -Vertrautheit mit dem Statistik-Programm R , erworben in der Lehrveranstaltung "Statistik" an der Paris Lodron Universität Salzburg sowie in den Lehrveranstaltungen „Statistisches Programmieren“ und „Computational Statistics“ an der Universität Wien
Führerschein:	Klasse B (ausgestellt im September 2007)

AUSSERUNIVERSITÄRES ENGAGEMENT

Seit 2005	Aktives Mitglied in der Freiwilligen Feuerwehr Rüstorf
-----------	--