Geometry of relative arbitrage

Master's thesis in Mathematics
submitted to
Westfälische Wilhelms-Universität Münster

by

Jasper Schmidt

2022

Contents

Chapter 1

Introduction

A crentral question when optimizing portfolios is how to allocate funds in a way that maximizes return and minimizes risk. Mainstream mathematical frameworks for market models and portfolio optimization operate under various, sometimes problematic, assumptions. They are usually not descriptive, meaning that they are not grounded in empirical observation. The problems such models have is well known, they are however still widely in use[1]. As an examples one can consider the Capital Asset Pricing Model (CAPM) or Modern Portfolio Theory (MPT) that rely on several unjustified assumptions regarding the behaviour of market participants such as rationality and risk-adversity. All major pricing and portfolio models share these types of assumption, in particular the no-arbitrage principle is fundamental.

In an attempt to build models that don't rely on these kinds of assumptions Fernholz in 1995 lay the foundation for Stochastic Portfolio Theory (SPT). The proponents of this theory place SPT closer to the natural sciences and stress the descriptive nature of the theory[2]. In this approach the market is modelled by stochastic differential equations and prices driven by Brownian motion.

Take some self financing portfolio. If we assume that we invests equal amounts into the portfolio and into the market, the portfolio is called a *relative arbitrage* if, after some potentially large time T, it will be at least as valuable as the market investment and with probability greater 0 even more valuable. In other words, a relative arbitrage has the risk free chance to beat the market. Fernholz was able to show that under the minimalistic assumptions of SPT such arbitrages indeed exist assuming sufficient market diversity and volatility. This is of course a violation of the fundamental no-arbitrage principle of many common models.

In this thesis, that closely follows [PW15], we will also take the SPT viewpoint,

[1] `https://www.aeaweb.org/articles?id=10.1257/0895330042162430` (last accessed: 26.06.2022)

[2] `https://www.math.columbia.edu/~ik/FernKarSPT.pdf` (last accessed: 26.06.2022)

but we will further reduce the assumptions imposed on the market. In particular we choose a discrete setup where the market weights are not stochastic processes but deterministic sequences. This way the results are independent of any probabilistic assumptions.

We will first explore SPT in the deterministic setup and proof results similar to those in continuous time. This includes defining the corresponding concept to relative arbitrages in discrete time, called pseudo arbitrages. We will characterize them in two ways. First by so called functionally generated portfolios and second by optimal transport problems.

We will show that all pseudo arbitrages are functionally generated and that they satisfy a certain decomposition formula. By using this formula we will be able too see how the volatility in the market drives the portfolio value.

The second characterization that relates pseudo arbitrages to solutions of certain optimal transport problems allows an investor to construct pseudo arbitrages as long as one can solve the transport problem.

The theorems and results in chapters 2-4 are, if not otherwise noted, from [PW15].

Setup

Consider a market of $n \in \mathbb{N}$ assets. By $X_1(t), \ldots, X_n(t)$ we denote the market capitalization of each asset at time t. We assume that the value of each X_i is always greater than zero. We define the *market weight* of an asset as its value relative to the market. So the i-th market weight is given by

$$\mu_i(t) = \frac{X_i(t)}{\sum_{j=1}^n X_j(t)}. \tag{1.1}$$

All such market weights lie in the *n-dimensional unit simplex* defined by

$$\Delta^{(n)} = \left\{ p = (p_1, \ldots, p_n) \in \mathbb{R}^n \;\middle|\; p_i > 0 \ \forall i, \ \sum_{i=1}^n p_i = 1 \right\}. \tag{1.2}$$

The objects we are mainly interested in are *portfolio functions* or just *portfolios.* They are functions of the form

$$\pi : \Delta^{(n)} \to \overline{\Delta^{(n)}}, \quad \mu(t) \mapsto \pi(\mu(t))$$

that take the current market weights as an input and give out the corresponding portfolio weights. All portfolios we will consider are self-financing, meaning that we only rebalance the portfolio (and not put money in or take money out). Furthermore, portfolios take no short positions which is formalized by defining π to take no negative values. Note that, other than the market weight, a portfolio can have a zero component i.e. invest nothing into single assets.

In practice an investor might regularly look up the current market weights plug these into the portfolio function and then rebalance his portfolio accordingly.

An important portfolio is the *market (portfolio)* defined by $\pi(\mu) = \mu$. As mentioned in the introduction, our aim is to "beat" this market portfolio. To measure this we define the relative value of a portfolio π (with respect to the market portfolio) as

$$V(t) = \frac{\text{value at time t of 1 \$ invested in the portfolio}}{\text{value at time t of 1 \$ invested in the market}}.$$

Which we can also write recursively as

$$V(0) = 1,\ V(t+1) = V(t) \sum_{i=1}^{n} \pi_i(\mu(t)) \frac{\mu_i(t+1)}{\mu_i(t)}. \tag{1.3}$$

Intuitively we can see why this gives the correct value. For each asset we multiply the change in the stock price $\mu_i(t+1)/\mu_i(t)$ with the current amount invested into this particular asset $\pi_i(\mu(t))$. Summing these increases for the individual asset gives the change in value for the whole portfolio over this specific period. By multiplying with the previous total gain we get the total change of value since time 0.

This intuition can also be formalized as it was done in [PW16, Lemma 2.1]. It is clear that "beating the market" translates to $V(t) > 1$.

We now define *pseudo-arbitrages* that translate Fernholz's relative arbitrages into our setup. A portfolio π is a *pseudo-arbitrage* on a subset $K \subset \Delta^{(n)}$ if the following holds.

(i) For all sequences $\{\mu_i\}_{i=1}^{\infty}$ of market weights in K, there exists a constant $C(K, \pi) \geq 0$ with $\log V(t) \geq -C$ for all t.

(ii) For some sequence the relative value satisfies $V(t) \to \infty$ as $t \to \infty$.

Note how the first condition ensures that the portfolio cannot have unlimited loss, and how the second condition states the possibility for unbounded gain. The second condition is the deterministic equivalent to saying that with probability greater 0 we outperform the market.

It is also apparent that such a portfolio is not by itself guaranteed to outperform the market. Necessary are of course sufficient diversity and volatility to which we will come back later.

We will have to deal a lot with concave functions. We review some important facts from convex analysis. A useful reference is [Roc72].
A function $\Phi : \Delta^{(n)} \to \mathbb{R}$ is said to be *concave* if for any choice of $p, q \in \Delta^{(n)}$

$$\Phi(\alpha p + (1-\alpha)q) \geq \alpha\Phi(p) + (1-\alpha)\Phi(q)$$

holds for all $\alpha \in [0,1]$.

A tangent vector ξ that satisfies

$$\Phi(p) + \langle \xi, q - p \rangle \geq \Phi(q) \tag{1.4}$$

for all $q \in \Delta^{(n)}$ is said to belong to the *superdifferential* of Φ at p. We also write $\xi \in \partial\Phi(p)$. We can also view this as a multivalued map $\partial\Phi : \Delta^{(n)} \to T\Delta^{(n)}$. The superdifferential for a concave function on $\Delta^{(n)}$ is always a non-empty, convex and compact set. If Φ is differentiable at p we have $\partial\Phi(p) = \nabla\Phi(p)$.

If ξ is a supergradient of $\Phi(p)$, the vector $(\xi, -1)$ is the normal to a hyperplane that supports the graph of Φ at $(p, \Phi(p))$. All these properties can be found in [Roc72, Section V].

For a tangent vector v we can also define the directional derivative as

$$D_v\Phi(p) = \lim_{h \downarrow 0} \frac{\Phi(p + hv) + \Phi(p)}{h} \tag{1.5}$$

as long as the limit exists, which it does for every concave function at every point([Roc72, Theorem 23.1]). Furthermore, if Φ is differentiable at p we have

$$D_v\Phi(p) = \langle v, q - p \rangle.$$

Finally, we also need some geometric properties of the simplex, in particular that its tangent space $T\Delta^{(n)}$ is given by all vectors v with $\sum_i v_i = 0$. This and a few other useful geometry facts can be found in the appendix.

Notation. If not otherwise noted $\mu(t)$ is always a market weight sequence and $\pi(\mu(t))$ a portfolio function. For two vectors v, w we define the fraction v/w and the product vw component wise (if it is possible). Furthermore, $\langle \cdot, \cdot \rangle$ and $||\cdot||$ are the standard euclidean scalar product and norm. With $e(i)$ we denote the i-th corner of the simplex, which coincides with the i-th unit vector. We sometimes write $\vec{1}$ for $(1, \ldots, 1) \in \mathbb{R}^n$ to make clear whether we deal with a scalar or a vector. We may write $A \cdot v$ for the matrix vector product between the matrix A and the vector v. If it is unambiguous we can omit the argument of a functions for example $\pi(\mu(t)) = \pi(\mu) = \pi$.

Chapter 2

The MCM property

The MCM Property is the connecting element between funtionally generated portfolios, pseudo-arbitrage and optimal transport. We will discuss some important basic properties, and then use the relationship to functionally-generated portfolios to characterize these further.

2.1 Basic properties

Definition 1. (Multiplicative Cyclical Monotonicity)
Take a sequence $\{\mu(t)\}_{t=0}^{m+1} \subset \Delta^{(n)}$ with $\mu(0) = \mu(m+1)$. Such a sequence is called a *cycle* in $\Delta^{(n)}$.
Let π be a portfolio. We say that π is *Multiplicative Cyclical Monotone*, or short *MCM*, if for any m and any cycle of length $m+1$ the relative value $V(m+1) \geq 1$.
We say the MCM property holds on some set $K \subset \Delta^{(n)}$, if for any cycle $\{\mu(t)\}_{t=0}^{m+1} \subset K$ we have $V(m+1) \geq 1$.
Furthermore, if a portfolio satisfies $V(m+1) \geq 1$ for all cycles with $||\mu(t+1) - \mu(t)|| < \delta$ for all t, we say the portfolio satisfies the δ-MCM property.

Lemma 1. *We can rewrite the relative value of the portfolio π along some cycle $\{\mu(t)\}_{t=0}^{m+1}$ as*

$$V(m+1) = \prod_{t=0}^{m} \left(1 + \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right). \tag{2.1}$$

Therefore, if the portfolio is MCM (2.1) holds with ≥ 1.

Proof. We use the recursive definition for $V(m+1)$.

$$V(m+1) = V(m) \left\langle \pi(\mu(m)), \frac{\mu(m+1)}{\mu(m)} \right\rangle = \ldots = \prod_{t=0}^{m} \left\langle \pi(\mu(t), \frac{\mu(t+1)}{\mu(t)} \right\rangle.$$

We can add a zero on the right hand side by noting that $1 = \langle \pi(\mu), \vec{1} \rangle$, which holds true because $\pi \in \Delta^{(n)}$. We get

$$\begin{aligned} V(m+1) &= \prod_{t=0}^{m} \left(\left\langle \pi(\mu(t), \frac{\mu(t+1)}{\mu(t)} \right\rangle + 1 - \langle \pi(\mu(t)), \vec{1}\ \rangle \right) \\ &= \prod_{t=0}^{m} \left(1 + \left\langle \pi(\mu(t)), \frac{\mu(t+1)}{\mu(t)} - \vec{1} \right\rangle \right) \\ &= \prod_{t=0}^{m} \left(1 + \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right). \end{aligned}$$

For the last step above, remember that the ratio π/μ is defined component wise. $\square$

The MCM property is an artificial concept, market sequences seldom follow a cycle. As a theoretical property it is nonetheless highly important.

Our aim is the characterization of pseudo-arbitrages. For a portfolio to be a pseudo-arbitrage it cannot, for any market sequence, have unbounded loss. This is exactly what the MCM property guarantees. As the next lemma shows, if a portfolio fails the MCM property it cannot, even if the market is volatile and diverse, be a pseudo-arbitrage.

Lemma 2. *Let π be a portfolio that fails the MCM property. Then there exists a sequence $\{\mu(t)\}_{t=0}^{\infty} \subset \Delta^{(n)}$, such that the corresponding relative value $V(t) \to 0$ for $t \to \infty$. In particular, π is not a pseudo-arbitrage.*

Proof. By assumption there exists some cycle $\{\tilde{\mu}(t)\}_{t=0}^{t=m+1}$ such that $\eta := V_{\tilde{\mu}}(m+1) < 1$ (with $V_{\tilde{\mu}}$ we denote the value function induced by the cycle $\tilde{\mu}$). The idea here is to construct a sequence that repeats the above cycle over and over again. Intuitively, the value of a portfolio along such a sequence must go to zero because of the multiplicative structure of the value function.

Define the sequence $\{\mu(t)\}_{t=0}^{t=\infty}$ by $\mu(t) = \tilde{\mu}(k)$ for $t = k \pmod{m+1}$. We first verify that $V_\mu(k(m+1)) = \eta^k \to 0$ for $k \to \infty$.

$$V_\mu(k(m+1)) = \prod_{t=0}^{k(m+1)-1} \left(1 + \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right)$$

Note that the product has exactly $k(m+1)$ factors, so we can regroup the factors according to the cyclical pattern. For the factors

$$v_t := \left(1 + \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right)$$

it is immediately clear that $v_{m+1} = v_0$, thus the factors inherit the cyclicity of μ. Then

$$V(k(m+1)) = \prod_{t=0}^{k(m+1)-1} v_t = \prod_{i=0}^{k-1}\prod_{j=0}^{m} v_{j+i(1+m)} = \prod_{i=0}^{k-1}\prod_{j=0}^{m} v_j = \prod_{i=0}^{k-1} \eta = \eta^k$$

where we used the above mentioned cyclicality $v_{i+j(m+1)} = v_i$.
Now we show that $V(M+1)$ goes to zero for $M \to \infty$. To achieve this, fix a large M, choose $k \in \mathbb{N}$ and $k' \in \{1, \dots, m\}$ such that $M = k(m+1) + k'$. (We can assume that $M > m+1$). Then

$$\begin{aligned} V(M+1) &= \prod_{t=0}^{M} \left(1 + \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle\right) \\ &= \prod_{t=0}^{k(m+1)-1} \left(1 + \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle\right) \\ &\qquad \prod_{t=k(m+1)}^{k(m+1)+k'} \left(1 + \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle\right). \end{aligned}$$

The left product is η^k as seen above. The right product can be estimated upwards by some constant $C > 0$. To see this we only need to justify that $1/\mu$ is bounded, because positivity follows from Lemma 1 and we only consider a finite number of $\mu(t)$ (at maximum m). By assumption $\mu \in \Delta^{(n)}$ so all components are strictly positive, accordingly $1/\mu_i$ is finite for all $\mu \in \{\tilde{\mu}(t)\}_{t=0}^{t=m+1}$. Furthermore, we choose C as an upper bound for all k' to make sure that we can always estimate the right product upwards by C regardless of the position in the "phase".
We get that for any large M (choosing the correct k)

$$V(M+1) \leq \eta^k C.$$

By construction $M \to \infty$ also implies that k goes to infinity, which proves the claim. $\square$

Our next proposition relates portfolios that satisfy the MCM property to certain concave functions, these portfolios will later be called *functionally generated.*

Proposition 1. *Let π be a portfolio function. Then the following holds*

(i) π satisfies the MCM property if and only if there exists a concave function $\Phi : \Delta^{(n)} \to (0, \infty)$ such that

$$1 + \left\langle \frac{\pi(p)}{p}, q - p \right\rangle \geq \frac{\Phi(q)}{\Phi(p)} \qquad \forall p, q \in \Delta^{(n)} \tag{2.2}$$

(ii) For any $\delta > 0$, if π satisfies the δ-MCM property, then π also satisfies the regular MCM property.

Proof. (i) First the converse, so suppose that a concave and positive Φ on $\Delta^{(n)}$ exists, such that (2.2) holds. We need to show that for any cycle $\{\mu(t)\}_{t=0}^{t=m+1}$ the MCM property holds.

Start by using the assumption

$$V(m+1) = \prod_{t=0}^{m} \left(1 + \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right) \geq \prod_{t=0}^{m} \frac{\Phi(\mu(t+1)}{\Phi(\mu(t))}.$$

The term on the right simplifies to 1, because $\mu(t)$ is a cycle. This shows the MCM property for π.

For the other direction assume π to satisfy the MCM property. Let $\mu(0) \in \Delta^{(n)}$ and define Φ by

$$\Phi(p) = \inf_{\substack{m \geq 0, \{\mu(t)\}_{t=0}^{m+1}, \\ \mu(m+1) = p}} \left[\prod_{t=0}^{m} \left(1 + \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right) \right]. \tag{2.3}$$

Here we take all market sequences that end in p, and then set $\Phi(p)$ as the value of the portfolio that has the lowest returns under all of these market evolutions. Note that $\mu(t)$ is not necessarily a cycle here.

Remember that the pointwise infimum over affine functions is concave. We can rewrite (2.3) as

$$\Phi(p) = \inf \left\{ V(m) \left(1 + \left\langle \frac{\pi(\mu(m))}{\mu(m)}, p - \mu(m) \right\rangle \right) \,\middle|\, m \geq 0, \{\mu(t)\}_{t=0}^{m+1} \right\}.$$

From this we can directly see that we have an infimum over affine functions in p, therefore Φ is indeed concave. From the definition is is clear that $V(t) \geq 0$, so we immediately get that Φ is non-negative.

Furthermore, by evaluating Φ at $\mu(0)$ we only allow cycles in the infimum, so the MCM property applies and gives us the inequality

$$\Phi(\mu(0)) = \inf_{\substack{m\geq 0, \{\mu(t)\}_{t=0}^{m+1}, \\ \mu(m+1)=\mu(0)}} \left[\prod_{t=0}^{m} \left(1 + \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right) \right] \geq 1. \tag{2.4}$$

The cycle $\{\mu(0)\}$ of only one point is included in the index set of the infimum. For this particular cycle we attain the value 1. So the infimum is at most 1 at this point and with the above inequality the value of $\Phi(\mu(0))$ is exactly 1.
We are left to show that $\Phi(p) > 0$ for all $p \in \Delta^{(n)}$ and that (2.2) holds for this Φ. A positive concave function on the simplex is either zero everywhere or strictly positive, this be shown from the definition (see apendix[??]). Above we already showed that $\Phi(\mu(0)) > 0$, therefore Φ must be strictly positive on the simplex.

To show (2.2), fix some $p, q \in \Delta^{(n)}$ and choose some $\alpha \in \mathbb{R}$ with $\Phi(p) < \alpha$. Looking at (2.3) there exists an $m \geq 0$ and a sequence $\{\mu(t)\}_{t=0}^{m+1}$ with $\mu(m+1) = p$ from the index set of the infimum, such that

$$\prod_{t=0}^{m} \left(1 + \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right) < \alpha. \tag{2.5}$$

Now set $\mu(m+2) = q$, then for all $m \geq 0$ and sequences $\{\mu(t)\}_{t=0}^{m+2}$ with $\mu(m+2) = q$ we can estimate the infimum upwards by choosing one particular sequence, therefore

$$\Phi(q) \leq \prod_{t=0}^{m+1} \left(1 + \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right). \tag{2.6}$$

For a sequence with $\mu(m+1) = p$ and $\mu(m+2) = q$, we can combine the above equations. Pulling out the last factor in (2.6) and estimating the remaining product with (2.5) gives us

$$\begin{aligned} \Phi(q) &< \left(1 + \left\langle \frac{\pi(\mu(m+1))}{\mu(m+1)}, \mu(m+2) - \mu(m+1) \right\rangle \right) \alpha \\ &= \left(1 + \left\langle \frac{\pi(p)}{p}, q - p \right\rangle \right) \alpha. \end{aligned}$$

We can conclude by letting $\alpha \downarrow \Phi(p)$, which gives us the desired equation (2.2).

(ii) We now restrict us to δ-MCM cycles. Meaning that the MCM property only has to hold for cycles in which two successive points are at most δ apart. We repeat the second part of the proof of (i) with δ-MCM. We only used the MCM property to show that Φ is positive with (2.4). To make this work in the δ-MCM setting, we need to add another condition to the infimum. Define

$$\Phi(p) = \inf_{\substack{m\geq 0, \{\mu(t)\}_{t=0}^{m+1}, \\ \mu(m+1)=p, \\ |\mu(t+1)-\mu(t)|<\delta}} \left[\prod_{t=0}^{m} \left(1 + \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right) \right]$$

With this we can just repeat the argument. In the last step where we showed (2.2) we needed to estimated the infimum by choosing a particular sequence of the infimum's index set. If we do this here we, have to account for the additional condition we introduced. Otherwise the prove is identical and leaves us with

$$1 + \left\langle \frac{\pi(p)}{p}, q - p \right\rangle \geq \frac{\Phi(q)}{\Phi(p)} \tag{2.7}$$

for all $p, q \in \Delta^{(n)}$ with $||p - q|| < \delta$.

We multiply with $\Phi(p)$ and decompose $\Phi(p)\pi(p)/p$ into two components, one tangential to $\Delta^{(n)}$ denoted by $(\cdot)_\perp$ and a normal component denote by $(\cdot)_\|$. We say that for all $v \in \Delta^{(n)}$ $v = v_\perp + v_\|$, with $v_\| \in T\Delta^{(n)}$.

In particular we get from (2.7) that for $||p - q|| < \delta$

$$\Phi(p) + \left\langle \left(\frac{\pi(p)}{p} \Phi(p) \right)_\perp, q - p \right\rangle + \left\langle \left(\frac{\pi(p)}{p} \Phi(p) \right)_\|, q - p \right\rangle \geq \Phi(q). \tag{2.8}$$

Note that $\sum_{i=1}^{n} (q_i - p_i) = 0$, therefore $p - q$ is in $T\Delta^{(n)}$. The first scalar product above is therefore zero, because the normal component and $q - p$ are orthogonal to each other by construction. So (2.8) is precisely the definition of the superdifferential (remember that $(\cdot)_\|$ is in $T\Delta^{(n)}$ by construction), therefore $(\frac{\pi(p)}{p}\Phi(p))_\|$ is an element of $\partial\Phi(p)|_{B_\delta(p)}$. In other words, the parallel component is the superdifferential of the convex function Φ restricted to a convex neighbourhood V of p.

This is almost the desired statement. In fact, the property that a vector is a superdifferential is a local property. This can be seen in [Roc72, Theorem 23.2], which states that

$$\partial\Phi(p) = \{\xi \in T\Delta^{(n)} \colon D_\nu \Phi(p) =< \xi, \nu >, \forall \nu \in T\Delta^{(n)}\}.$$

The directional derivative is defined locally, so for a convex neighbourhood V of p we have $\partial(\Phi|_V)(p) = \partial\Phi(p)$. Equation (2.7) holds without restrictions on p and q over the whole simplex (i.e. its the same as (2.2)). The result now follows from part (i). □

Corollary 1. *Let π satisfy the MCM property on some set $K \subset \Delta^{(n)}$. Then there exists a positive concave function Φ on $\Delta^{(n)}$ such that (2.2) holds for all $p, q \in K$.*

Proof. This is easily obtained by making small modifications to the proof of Proposition 1, (i). We need to change the definition of Φ by restricting the infimum to sequences $\{\mu(t)\}_{t=0}^{m}$ that lie in K. Only $\mu(m+1)$ can still be chosen in $\Delta^{(n)}$ to make sure that Φ is still defined on the whole simplex. Then Φ is, as before, positive and concave on the simplex.

To show equation (2.2) in the final step, just let p, q be in K and repeat the proof. □

2.2 Functionally generated portfolios

In the first section we explored some important basic properties. We also saw that for a portfolio to satisfy the MCM property it is equivalent that there exist a positive concave function that satisfies (2.2). We stress this property with the following definition.

Definition 2. (Functionally generated portfolios)
We call a portfolio function π to be *functionally generated*, if there exists a positive concave function Φ on the simplex such that (2.2) holds. We also say Φ is a *generating function* of π.

Lemma 3. *Let Φ be a positive and concave function, then for all $p \in \Delta^{(n)}$*

$$\partial \log \Phi(p) = \frac{1}{\Phi(p)} \partial \Phi(p) := \left\{ \frac{1}{\Phi(p)} \xi, \xi \in \partial \Phi(p) \right\} \tag{2.9}$$

Proof. First note that $\log \Phi(p)$ is indeed a concave function (see remark to B1 in the appendix). Let $p \in \Delta^{(n)}$ be a point where Φ is differentiable. By the chain rule we have

$$\partial \log \Phi(p) = \nabla \log \Phi(p) = \frac{1}{\Phi(p)} \nabla \Phi(p) = \frac{1}{\Phi(p)} \partial \Phi(p).$$

The more interesting case is when Φ is not differentiable. For a finite concave function Φ we will show that

$$\partial \Phi(p) = \mathrm{cl}(\mathrm{conv}(S(p))), \tag{2.10}$$

where $S(p)$ is the set of all limits of sequences $\nabla(p_1), \nabla(p_2), \ldots$ with $p_i \to p$. Then, for a point p at which Φ is not differentiable we can construct a sequence $p_i \to p$ such that Φ is differentiable at p$_i$. We get

$$\partial \log \Phi(p) \ni v' \xleftarrow{i\to\infty} \partial \log \Phi(p_i) = \frac{1}{\Phi(p_i)} \partial \Phi(p_i) \xrightarrow{i\to\infty} v \in \frac{1}{\Phi(p)} \partial \Phi(p).$$

The limits exist, because the graphs of $\partial\Phi$ and $\partial \log \Phi$ are closed [Roc72, Theorem 24.4], and have to coincide. We get $v = v' \in S(p)$. By choosing different sequences p_i we can get the whole set $S(p)$. Now, take $w \in \partial \log \Phi(p)$. We distinguish two cases.

In the first case, there exist some $w_1, w_2 \in S(p)$ such that w is the convex combination of w_1 and w_2. Then we also have $w_1, w_2 \in \frac{1}{\Phi(p)} \partial\Phi(p)$ so the convex convex combination w is also included in $\frac{1}{\Phi(p)}\partial\Phi(p)$ by the convexity of superdifferentials.

In the second case, $w \in \mathrm{cl}(\mathrm{conv}(S(p)) \setminus \mathrm{conv}(S(p))$ then there exists a sequence $w_i \to w$, where $w_i \in \mathrm{conv}(S(p))$. For all w_i case one applies so we again have the inclusion into both superdifferentials. Superdifferentials are closed, so the limit w also lies in both.

The reverse set inlcusion $\frac{1}{\Phi(p)}\partial\Phi(p) \subset \partial \log \Phi(p)$ can be done in the same way. We are left to justify (2.10). The inclusion $\partial\Phi(p) \subset \mathrm{cl}(\mathrm{conv}(S(p))$ holds because of the above mentioned fact that the graph of $\partial\Phi$ is closed. For the reverse look at [Roc72, Theorem 25.6]. This theorem is the generalisation of (2.10). It states that $\partial\Phi(p) = \mathrm{cl}(\mathrm{conv}(S(p)) + K(p)$, where $K(p)$ is the normal cone to the domain of Φ at p. Looking into the proof of the reverse inclusion we see that the author has shown that the convex hull of the superdifferential consists of its extreme points and extreme directions, the latter is what is captured by $K(p)$. In our case, however, the concave function is bounded and therefore the superdifferential is bounded as well. In turn there are not extreme directions in the set and we can disregard $K(p)$. $\square$

We will often have to deal with the superdifferential of generating functions. In the following Proposition we will prove the interesting claim there is a one to one correspondence between the superdifferential of a positive concave function and a functionally generated portolfio of which the generator is said concave function.

Proposition 2. *Let Φ be a positive concave function on $\Delta^{(n)}$.*

(i) Assume the portfolio π to be functionally generated by Φ. Then, for all $p \in \Delta^{(n)}$, the tangent vector $v = (v_1, \dots, v_n)$ with

$$v_i = \frac{\pi_i}{p_i} - \frac{1}{n} \sum_{j=1}^{n} \frac{\pi_j}{p_j} \tag{2.11}$$

is in $\partial \log \Phi(p)$.

(ii) For any $v \in \partial \log \Phi(p)$, the vector $\pi = (\pi_1, \dots, \pi_n)$ with

$$\frac{\pi_i}{p_i} = v_i + 1 - \sum_{j=1}^{n} p_j v_j \tag{2.12}$$

is in $\overline{\Delta^{(n)}}$.

These two operations that map the portfolio vector onto the supergradient and vice versa, are inverse to each other.

Proof. (i) We first note that v is actually a tangent vector, namely

$$\sum_{i=1}^{n} v_i = \sum_{i=1}^{n} \frac{\pi_i}{p_i} - \sum_{j=1}^{n} \frac{\pi_j}{p_j} = 0.$$

$\pi(p)/p$ itself is not a tangent vector so we need this summand to project onto the tangent space. Note that Φ generates π by assumption, so equation (2.2) holds.

From the definition of v we have for all $\xi \in T\Delta^{(n)}$

$$\left\langle \frac{\pi(p)}{p} - v, \xi \right\rangle = \frac{1}{n} \sum_{j=1}^{n} \frac{\pi_j(p)}{p} \overbrace{\sum_{i=1}^{n} \xi_i}^{=0} = 0$$

meaning that $\frac{\pi(p)}{p} - v$ is perpendicular to $T\Delta^{(n)}$. As $q - p$ is also a tangent vector, we can set $\xi = q - p$. Then

$$\left\langle \frac{\pi(p)}{p}, q - p \right\rangle = \langle v, q - p \rangle + \left\langle \frac{\pi(p)}{p} - v, q - p \right\rangle = \langle v, q - p \rangle. \tag{2.13}$$

The part that is perpendicular to $T\Delta^{(n)}$ is zero in the scalar product, which leads to the above simplification.

We apply (2.13) to the definition of functionally generated portfolios (2.2) so we can replace $\pi(p)/p$ by v in the scalar product. After multiplying with $\Phi(p)$ we arrive at

$$\Phi(p) + \langle v\Phi(p), q-p\rangle \geq \Phi(q).$$

Which is precisely the definition of $v\Phi(p) \in \partial\Phi(p)$, as long as $v\Phi(p)$ is a tangent vector. With Lemma 3 we get

$$v\Phi(p) \in \partial\Phi(p) \iff v = v\Phi(p)\frac{1}{\Phi(p)} \in \partial \log \Phi(p).$$

It is left to verify that $v\Phi(p) \in T\Delta^{(n)}$, but this is immediately clear, because Φ is real-valued and v is a tangent vector.

(ii) Need to check that $\pi \in \Delta^{(n)}$. First, the sum over the components of π is indeed 1.

$$\sum_{i=1}^n \pi_i = \sum_{i=1}^n \left(v_i p_i + p_i - p_i \sum_{j=1}^n p_j v_j \right) = \sum_{i=1}^n v_i p_i + 1 - \left(\sum_{j=1}^n p_j v_j \right) \left(\sum_{i=1}^n p_i \right) = 1.$$

Next we check that $\pi_i \geq 0$. To do so, fix $p \in \Delta^{(n)}$ and $0 < t < 1$. We define a second point q by $q - p = t(e(i) - p)$. It is easy to verify that $\sum_i q_i = 1$ and $q_i > 0$ (because $t < 1$), so q is in the simplex as well.

By assumption $v \in \partial \log \Phi(p)$, so from Lemma 3 we have that $\Phi(p)v \in \partial\Phi(p)$. The definition of the superdifferential then states that $\Phi(q) - \Phi(p) \leq \langle \Phi(p)v, q-p\rangle$ holds for all $q \in \Delta^{(n)}$. We can use this to get

$$\begin{aligned} -\Phi(p) &\leq \Phi(p + t(e(i) - p)) - \Phi(p) \\ &\leq \langle \Phi(p)v, t(e(i) - p)\rangle \\ &= t\Phi(p)\langle v, e(i) - p\rangle, \end{aligned}$$

where the first simple estimate uses $\Phi > 0$. Expanding the inner product and dividing by $\Phi(p)$ yields

$$-1 \leq t\left(v_i - \langle p, v\rangle\right).$$

Now let t converges to 1 from below, this leads to

$$0 \leq v_i + 1 - \langle p, v\rangle = \pi_i.$$

which proves part (ii).

Finally we show that the two mappings defined in (i) and (ii) are inverse functions of each other.

Fix a $p \in \Delta^{(n)}$, for notational clearness we define

$$\phi : \overline{\Delta^{(n)}} \to \partial \log \Phi(p),\ \pi(p) \mapsto v = \left(\frac{\pi_i(p)}{p_i} - \frac{1}{n} \sum_{j=1}^{n} \frac{\pi_j(p)}{p_j} \right)_{1 \leq i \leq n} \tag{2.14}$$

and

$$\psi : \partial \log \Phi(p) \to \overline{\Delta^{(n)}},\ v \mapsto \pi(p) = \left(p_i v_i + p_i - p_i \sum_{j=1}^{n} p_j v_j \right)_{1 \leq i \leq n} \tag{2.15}$$

First, let π be functionally generated. Then, for $\phi(\pi) = v$

$$(\psi \circ \phi)(\pi(p))_i = p_i v_i + p_i - p_i \sum_{j=1}^{n} p_j v_j. \tag{2.16}$$

Note, that

$$p_i \sum_{j=1}^{n} p_j v_j = p_i \underbrace{\sum_{h=1}^{n} \pi_h}_{=1} - \frac{p_i}{n} \underbrace{\sum_{j=1}^{n} p_j}_{=1} \sum_{k=1}^{n} \frac{\pi_k}{p_k} = p_i - \frac{p_i}{n} \sum_{k=1}^{n} \frac{\pi_k}{p_k}$$

and

$$p_i v_i = \pi_i - \frac{p_i}{n} \sum_{j=1}^{n} \frac{\pi_j}{p_j}.$$

Using these (2.16) becomes

$$(\psi \circ \phi)(\pi)_i = \pi_i(p).$$

Conversely, let $v \in \partial \log \Phi(p)$ and set $\psi(v) = \pi$, we can then use equation (2.14)

$$\begin{aligned}
(\phi \circ \psi)(v)_i &= \frac{\pi_i(p)}{p_i} - \frac{1}{n} \sum_{j=1}^{n} \frac{\pi_j(p)}{p_j} \\
&= v_i + 1 - \sum_{k=1}^{n} p_k v_k - \frac{1}{n} \sum_{j=1}^{n} \left(v_j + 1 - \sum_{l=1}^{n} p_l v_l \right) \\
&= v_i - \sum_{k=1}^{n} p_k v_k - \frac{1}{n} \underbrace{\sum_{j=1}^{n} v_j}_{=0} + \sum_{l=1}^{n} p_l v_l.
\end{aligned}$$

Where everything cancels except v_i. □

This proposition allows us to equivalently characterize a portfolio by its generator, or more precisely the superdifferential of the generator.

Proposition 3. *Let Φ be a concave and positive function that generates the portfolio π. Then the following holds*

(i) *Φ is unique up to a multiplicative constant.*

(ii) *For all $p \in \Delta^{(n)}$ and $i \in \{1, \ldots, n\}$ we have*

$$1 + D_{e(i)-p} \log \Phi(p) \leq \frac{\pi_i}{p_i} \leq 1 - D_{p-e(i)} \log \Phi(p). \tag{2.17}$$

If Φ is differentiable, then π is given by

$$\pi_i = p_i \left(1 + D_{e(i)-p} \log \Phi(p)\right). \tag{2.18}$$

(iii) *If π is continuous, then Φ is continuously differentiable.*
More generally, if $\pi \in C^k$ then $\Phi \in C^{k+1}$

(iv) *If $\Phi : \Delta^{(n)} \to (0, \infty)$ is concave and differentiable, and we define π by (2.18). Then, π is generated by Φ.*

Proof. (i) Assume π has two generating functions Φ_1 and Φ_2. Pick two arbitrary, non-identical points p and q in the simplex and denote the line from p to q by l. We can now restrict the generating functions to this line and get two one-dimensional concave functions $\Phi_1|_l$ and $\Phi_1|_l$.

Proposition 2 describes the connection between the log-supergradient of the generating function and of the portfolio, the representation was independent of the generator. For some $r \in \Delta^{(n)}$, the function that maps a vector $v \in \partial \log \Phi(r)$ to a portfolio vector and its inverse are independent of the generating function. As an easy notation write $\pi \mapsto v$ and $v \mapsto \pi$ for the functions described in (2.11) and (2.12) respectively.

Take any vector $v \in \partial \log \Phi_1(r)$ and map it onto π by (2.12). This π is independent of any generating function. If we apply (2.11) we again get v because the functions are inverse to each other, but note that Proposition 2 (i) states that if Φ is a generating function $\pi \mapsto v$ lies in its log-supergradient. By assumption Φ_2 is a generating function so consequently $v \in \partial \log \Phi_2(r)$. Interchanging the indices and repeating this idea implies

$$\partial \log \Phi_1(r) = \partial \log \Phi_2(r) \tag{2.19}$$

for all r in the simplex.

Both generating functions are concave so they are differentiable almost everywhere. The set of points where both functions are differentiable is also a set of full measure, simply by the fact that the union of sets of non-differentiable points stays countable.

Returning back to the restriction of the functions to the line l. At a differentiable point the supergradients coincide with each other and with the derivative. Therefore

$$\log \Phi_1'|_l = \log \Phi_2'|_l \tag{2.20}$$

almost everywhere. We now want to apply [Roc72, Corollary 24.2.1], the theorem requires the functions to be 1-dimensional which is the reason we restricted to the line. The theorem states that[1]

$$\begin{aligned}\log \Phi_1|_l(q) - \log \Phi_1|_l(p) &= \int_p^q \log \Phi_1'|_l(t)dt \\ &= \int_p^q \log \Phi_2'|_l(t)dt = \log \Phi_2|_l(q) - \log \Phi_2|_l(p).\end{aligned}$$

Most importantly we can replace the integrand because of (2.20) above that holds almost surely. This leads to

$$\frac{\Phi_2(p)}{\Phi_1(p)} = \frac{\Phi_2(q)}{\Phi_1(q)}.$$

We chose p, q arbitrary, so the ratio Φ_1/Φ_2 is a constant over the simplex.

(ii) Choose some $h \in \mathbb{R} \setminus \{0\}$ such that $p + h(e(i) - p) \in \Delta^{(n)}$. The sum over the components of the vector is always 1, so the only problematic part here is that the components also need to be positive, but this can always be achieved by choosing h small enough. Set $q = p + h(e(i) - p)$, then by the definition of functionally generated portfolios we get

$$1 + \left\langle \frac{\pi(p)}{p}, h(e(i) - p) \right\rangle \geq \frac{\Phi(q)}{\Phi(p)} = \frac{\Phi(p + h(e(i) - p))}{\Phi(p)}. \tag{2.21}$$

We can expand the inner product on the left hand side

$$\begin{aligned}\left\langle \frac{\pi(p)}{p}, h(e(i) - p) \right\rangle &= \sum_{j=1}^n \frac{\pi_j(p)}{p_j} h(e(i)_j - p_j) \\ &= \frac{\pi_i(p)}{p_i} h - h \sum_{j=1}^n \pi_j(p) \\ &= h \left(\frac{\pi_i(p)}{p_i} - 1 \right).\end{aligned}$$

[1] The notation here hides the fact that there is some underlying parametrization of the line l. In the sense of $[0, 1] \ni t \mapsto \Phi_1|_l(\gamma(t))$ where γ parametrizes the line. This is not the crucial point here and the same issue will be discusses in depth in Theorem 1 so we will omit this detail for now.

Apply this to (2.21), take the logarithm on both sides and divide by $h > 0$ to obtain

$$\frac{\log\left(1 + h\left(\frac{\pi_i}{p_i} - 1\right)\right)}{h} \geq \frac{\log \Phi\left(p + h(e(i) - p)\right) - \log \Phi(p)}{h}. \tag{2.22}$$

We are interested in the limit $h \to 0$. First we analyse the limit from above. Note that from the Taylor expansion we have

$$\log(1 + x) = x - \frac{x^2}{2} + R_3(x)$$

where R_3 is an $O(x^3)$-term. Using this approximation on the numerator the left hand side becomes

$$\frac{\log\left(1 + h\left(\frac{\pi_i}{p_i} - 1\right)\right)}{h} = \frac{h\left(\frac{\pi_i}{p_i} - 1\right)}{h} - \frac{h^2\left(\frac{\pi_i}{p_i} - 1\right)^2}{2h} + \frac{R^3(h)}{h} \xrightarrow{h\downarrow 0} \frac{\pi_i}{p_i} - 1.$$

The right hand side in (2.22) is precisely the definition of the directional derivative

$$\lim_{h \downarrow 0} \frac{\log\left(\Phi\left(p + h(e(i) - p)\right)\right) - \log \Phi(p)}{h} = D_{e(i)-p} \log \Phi(p).$$

This shows the first inequality.

To show the second inequality we look at the limit from below. To do so first divide by $h < 0$ in the step before (2.22). Note that the inequality reverses, so we have

$$\frac{\log\left(1 + h\left(\frac{\pi_i}{p_i} - 1\right)\right)}{h} \leq \frac{\log \Phi\left(p + h(e(i) - p)\right) - \log \Phi(p)}{h}. \tag{2.23}$$

Now take the limit $h \uparrow 0$. Notice that the convergence on the left-hand side is unchanged. We define $\bar{h} = -h$. Then the right hand side becomes

$$-\lim_{\bar{h} \downarrow 0} \frac{\log \Phi\left(p + \bar{h}(p - e(i))\right) - \log \Phi(p)}{\bar{h}} = -D_{p-e(i)} \log \Phi(p).$$

Which gives us the desired inequality.

If Φ is differentiable, for every tangent vector v the equality

$$D_v \log \Phi(p) = \langle v, \nabla \log \Phi(p) \rangle = -\langle -v, \nabla \log \Phi(p) \rangle = -D_{-v} \log \Phi(p)$$

holds. In turn (2.17) becomes

$$1 + D_{e(i)-p} \log \Phi(p) = \frac{\pi_i}{p_i}$$

which was the last statement in (ii).

(iii) Let π be continuous. Then

$$p \mapsto \frac{\pi(p)}{p} - \frac{1}{n}\sum_{j=1}^{n}\frac{\pi_j(p)}{p_j} \tag{2.24}$$

is continuous and in $\partial \log \Phi(p)$ by Proposition 2. We use [Rai88, Theorem 4] which says that the Fréchet differentiability of $\log \Phi$ holds if and only if there exists a selection of the sub-differential map which is norm to norm continuous. Note that in $\mathbb{R}^n$ all norms are equivalent, so the continuity condition is satisfied by the selection (2.24). Furthermore one can show that regular differentiability follows almost immediately from the definition of Fréchet differentiability (see apendix). Putting this together we get that $\log \Phi$ is differentiable.

We can use [Roc72, Corollary, 25.5.1] which states that finite convex functions (such as our Φ) on an open convex set (here $\Delta^{(n)}$) that are differentiable are actually continuous differentiable. This shows the claim for $k = 0$.

Now assume that $\pi \in C^k$. We know that Φ is at least of regularity C^1. We use the coordinates

$$\varphi : \Delta^{(n)} \to U \subset \mathbb{R}^{n-1}, p = (p_1, \ldots, p_{n-1}, p_n) \mapsto (p_1, \ldots, p_{n-1})$$

which just project a point from simplex into the $\mathbb{R}^{n-1}$ plain. To ensure that φ can be inverted, we define U such that $U = \varphi\left(\Delta^{(n)}\right)$. The inverse is given by

$$\varphi^{-1}(p_1, \cdots, p_{n-1}) = \left(p_1, \cdots, p_{n-1}, 1 - \sum_{i=1}^{n-1} p_i\right).$$

The definition for C^{k+1} differentiability on a manifold is that

$$\frac{\partial}{\partial p_i}\left(\log \Phi \circ \varphi^{-1}\right)(p_1, \ldots, p_{n-1}) \tag{2.25}$$

needs to be C_k for all i. We assume the equality

$$\frac{\partial}{\partial p_i}(\log \Phi \circ \varphi^{-1})(p_1, \ldots, p_{n-1}) = D_{e(i)-e(k)} \log \Phi(p) \tag{2.26}$$

which we will show further below. We can then use the differentiability of Φ to get

$$\begin{aligned}
D_{e(i)-e(n)} \log \Phi(p) &= \langle e(i) - e(n), \nabla \log \Phi(p)\rangle \\
&= \langle e(i) - p, \nabla \log \Phi(p)\rangle - \langle e(n) - p, \nabla \log \Phi(p)\rangle \\
&= D_{e(i)-p} \log \Phi(p) - D_{e(n)-p} \log \Phi(p) \\
&\overset{(ii)}{=} \frac{\pi_i}{p_i} - \frac{\pi_n}{p_n}.
\end{aligned}$$

This shows the regularity of (2.25) is C_k.
The equality (2.26) is left to show. To distinguish vectors in $\mathbb{R}^n$ and $\mathbb{R}^{n-1}$ we denote vectors in $\mathbb{R}^{n-1}$ by a bar over the variable. For example $\bar{p} := (p_1, \dots, p_{n-1}) = \varphi(p) \in \mathbb{R}^{n-1}$. Similarly for the unit vectors, $\overline{e(i)} \in \mathbb{R}^{n-1}$ but $e(i) \in \mathbb{R}^n$. Accordingly is $\overline{e(i)}$ only defined for $i = 1, \cdots, n-1$.
We use the definition of partial derivatives on the left hand side of (2.26)

$$\frac{\partial}{\partial p_i}(\log \Phi \circ \varphi^{-1})(\bar{p}) = \lim_{h \to 0} \frac{\log \Phi(\varphi^{-1}(\bar{p} + h\overline{e(i)})) - \log \Phi(\varphi^{-1}(\bar{p}))}{h}. \tag{2.27}$$

The right hand side in (2.26) is just a directional derivative so with the definition we have

$$\lim_{h \downarrow 0} \frac{\log \Phi(p + h(e(i) - e(n)) - \log \Phi(p)}{h}. \tag{2.28}$$

Here the limit is equal to the limit from above (and below), because Φ is continuously differentiable at least once. When we compare (2.27) and (2.28) we see it is left to show that

$$\varphi^{-1}\left(\bar{p} + h\overline{e(i)}\right) = p + h\left(e(i) - e(n)\right). \tag{2.29}$$

This is a straight forward calculation

$$\begin{aligned}\varphi^{-1}\left(\bar{p} + h\overline{e(i)}\right) &= \left(\overbrace{\bar{p} + h\overline{e(i)}}^{\in\mathbb{R}^{n-1}}, \overbrace{1 - \sum_{j=1}^{n-1}\left(\bar{p}_j + h\overline{e(i)}_j\right)}^{\in\mathbb{R}}\right) \\ &= \overbrace{\left(\bar{p}, 1 - \sum_{j=1}^{n-1}\bar{p}_j\right)}^{=p} + h\left(\overline{e(i)}, -\sum_{j=1}^{n-1}\overline{e(i)}\right).\end{aligned}$$

Which leaves us with

$$\varphi^{-1}\left(\bar{p} + h\overline{e(i)}\right) = p + h\left(\overline{e(i)}, -1\right) = p + h(e(i) - e(n)).$$

from which the claim (2.26) follows.

(iv) Define π by (2.18), that is

$$\pi_i = p_i\left(1 + D_{e(i)-p}\log \Phi(p)\right).$$

In part (ii) we already assumed that π is generated by Φ, we now need to show that even without that assumption π lies in the simplex. First fix a $p \in \Delta^{(n)}$. We claim that

$$\pi_i \geq 0 \iff D_{e(i)-p}\Phi(p) \geq -\Phi(p) \tag{2.30}$$

From the definition it is clear that $\pi_i \geq 0$ if and only if $1 + D_{e(i)-p} \log \Phi(p) \geq 0$. Using the positivity of Φ this is equivalent to

$$\Phi(p) D_{e(i)-p} \log \Phi(p) \geq -\Phi(p).$$

We can use the differentiability and calculate

$$\begin{aligned} -\Phi(p) &\leq \Phi(p) D_{e(i)-p} \log \Phi(p) \\ &= \Phi(p)\langle e(i) - p, \nabla \log \Phi(p)\rangle \\ &= \langle e(i) - p, \nabla \Phi(p)\rangle \\ &= D_{e(i)-p} \Phi(p) \end{aligned}$$

which shows (2.30).

We can now use (2.30) to show that all components of the portfolio are non-negative. In particular for some h with $0 < h < 1$ we have

$$\Phi(p + h(e(i) - p)) = \Phi((1-h)p + he(i)) \geq (1-h)\Phi(p) + h\Phi(e(i))$$

by the concavity of Φ which leads to

$$\Phi(p + h(e(i) - p)) - \Phi(p) \geq h(\Phi(e(i)) - \Phi(p)).$$

We now can use this to estimate the directional derivative in a useful way.

$$D_{e(i)-p}\Phi(p) = \lim_{h \downarrow 0} \frac{\Phi(p + h(e(i) - p)) - \Phi(p)}{h} \geq \Phi(e(i)) - \Phi(p) \geq -\Phi(p).$$

Most importantly, with (2.30) this shows that $\pi_i \geq 0$ for all i.

The second condition necessary to show that π is in the simplex is that $\langle \pi, \vec{1}\rangle = 1$. Remember that $\sum_i p_i = 1$ then the only piece missing is

$$\begin{aligned} \sum_{i=1}^n p_i D_{e(i)-p} \log \Phi(p) &= \sum_{i=1}^n p_i \langle e(i) - p, \nabla \log \Phi(p)\rangle \\ &= \sum_{i=1}^n p_i \sum_{j=1}^n (e(i)_j - p_j) \nabla \log \Phi(p)_j \\ &= \sum_{j=1}^n \nabla \log \Phi(p)_j \sum_{i=1}^n p_i (e(i)_j - p_j) \\ &= \sum_{j=1}^n \nabla \log \Phi(p)_j \left[\sum_{i=1}^n p_i e(i)_j - p_j \sum_{i=1}^n p_i\right] \\ &= \sum_{j=1}^n \nabla \log \Phi(p)_j \left[\sum_{i=1}^n p_i \delta_{ij} - p_j\right] = 0. \end{aligned}$$

We have shown that for a positive and concave Φ the function π, defined by (2.18), is indeed a portfolio.

To conclude rewrite the portfolio using the differentiability of Φ as

$$\frac{\pi_i}{p_i} = 1 + \langle e(i) - p, \nabla \log \Phi(p) \rangle = 1 + v_i - \langle p, v \rangle$$

for $v \in \partial \log \Phi(p)$. This is precisely equation (2.12). We showed above that π is a portfolio so we can apply Proposition 2 which shows that Φ functionally generates π. $\square$

2.3 The L-divergence

In this section we will investigate the mechanism that connects functionally generated portfolios and peseudo-arbitrages in detail. To formulate this we need to define an object that can capture the volatility of the market, this will be the L-divergence.

Definition 3. (L-divergence)
Let π be a portfolio function that is generated by some positive and concave Φ. Then we can define the *L-divergence* $T : \Delta^{(n)} \times \Delta^{(n)} \to \mathbb{R}$ for the pair (Φ, π) as follows

$$T(q|p) = \log\left(1 + \left\langle \frac{\pi(p)}{p}, q - p \right\rangle\right) - (\log \Phi(q) - \log \Phi(p)) \tag{2.31}$$

for any $p, q \in \Delta^{(n)}$.

While not a metric in general, because it is not symmetric, we can still regard the L-divergence as a device to capture some notion of distance. This distance like properties depend on the type of underlying function Φ as the next lemma shows.

Lemma 4. (Properties of the L-divergence) *Let T be the L-divergence for a pair (Φ, π) as defined above, where Φ is a positive concave function that generates π. Then, for two non-identical points $p, q \in \Delta^{(n)}$*

(i) $T(q|q) = 0$,

(ii) $T(q|p) \geq 0$,

(iii) $T(q|p) > 0$, *if and only if Φ is strictly concave on the line from p to q.*

Proof. The first statement is clear. Statement (ii) follows immediately from the definition of functionally generated portfolios (2.2).

To show (iii), first note that T is strictly positive if and only if

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

holds for p and q, while greater or equal always holds. We now prove that

$$1 + \left\langle \frac{\pi(p)}{p}, q - p \right\rangle = \frac{\Phi(q)}{\Phi(p)} \tag{2.32}$$

holds, if and only if Φ is affine on the line from p to q. This suffices to proof the claim.

First, let Φ be affine. We can then write $\Phi(q) = a + \langle b, q - p\rangle$ for some $a \in \mathbb{R}$ and $b \in \mathbb{R}^n$. More precisely $a = \Phi(p)$ and because Φ is assumed to be affine and therefore differentiable $b = \nabla\Phi(p) = \partial\Phi(p)$.
That means

$$\frac{\Phi(q)}{\Phi(p)} = 1 + \left\langle \frac{b}{a}, q - p \right\rangle. \tag{2.33}$$

Furthermore, $\frac{b}{a} \in \frac{1}{\Phi(p)}\partial\Phi(p) = \partial \log \Phi(p)$, which we can use in Proposition 2 (let $v = \frac{b}{a}$), to obtain

$$\frac{\pi_i}{p_i} = \frac{b_i}{a} + 1 - \left\langle p, \frac{b}{a} \right\rangle. \tag{2.34}$$

Plugging $\frac{b}{a}$ from equation (2.34) into (2.33) we finally reach

$$\frac{\Phi(q)}{\Phi(p)} = 1 + \left\langle \frac{\pi}{p}, q - p \right\rangle - \overbrace{\left\langle \vec{1}, q - p \right\rangle}^{=0} + \left\langle p, \frac{b}{a} \right\rangle \overbrace{\sum_i (q_i - p_i)}^{=0}.$$

So we get (2.32), as desired.

Now, to prove the converse, let Φ be positive and concave, but also assume that (2.32) holds. Apply Proposition 2 on π/p in (2.32), then

$$\frac{\Phi(q)}{\Phi(p)} = 1 + \sum_i \frac{\pi_i}{p_i}(q_i - p_i) = 1 + \sum_i \left(v_i(q_i - p_i) - \frac{1}{n}\sum_j \frac{\pi_j}{p_j}(q_i - p_i) \right)$$

for some tangent vector $v \in \partial \log \Phi(p)$. The term with the j-indexed sum is zero. We can simplify to

$$\frac{\Phi(q)}{\Phi(p)} = 1 + \langle v, q - p\rangle.$$

We can write $v = \frac{1}{\Phi(p)}w$ for some $w \in \partial\Phi(p)$, so this becomes

$$\Phi(q) = \Phi(p) + \langle w, q - p\rangle$$

Or in other words, Φ is affine on the line from p to q, which we wanted to show. $\square$

The following Lemma is central to understanding the relationship of functionally generated portfolios and pseudo-arbitrages.

Lemma 5. *Let π be generated by some positive and concave Φ. Denote by T the L-divergence corresponding to the pair (Φ, π). Then the relative value function $V(t)$ satisfies*

$$\log V(t) = \log \frac{\Phi(\mu(t))}{\Phi(\mu(0))} + A(t), \tag{2.35}$$

where $A(t) = \sum_{k=0}^{t-1} T(\mu(k+1)|\mu(k))$ is non-decreasing. Furthermore, the generating function Φ is affine if, and only if, $A(t) \equiv 0$ for all market sequences.

Proof. Since we showed in Lemma 4, that the L-divergence in non-negative and only 0, if Φ is affine, the properties of A follow immediately. From the product structure of $V(t)$ (see for example Lemma 1) we have

$$\frac{V(t+1)}{V(t)} = 1 + \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle.$$

The right hand side appears in the definition of T, which we can rewrite as

$$\begin{aligned} T(\mu(t+1)|\mu(t)) &= \log\left(1 + \left\langle \frac{\pi(p)}{p}, q - p \right\rangle\right) - \log \frac{\Phi(\mu(t+1))}{\Phi(\mu(t))} \\ &= \log \frac{V(t+1)}{V(t)} - \log \frac{\Phi(\mu(t+1))}{\Phi(\mu(t))}. \end{aligned}$$

Next take the sum to get

$$A(t) = \sum_{k=0}^{t-1} T(\mu(t+1)|\mu(t)) = \sum_{k=0}^{t-1} \log \frac{V(t+1)}{V(t)} - \sum_{k=0}^{t-1} \log \frac{\Phi(\mu(t+1))}{\Phi(\mu(t))}.$$

We can directly see, that both sums are telescopic, and simplify to

$$A(t) = \log \frac{V(t)}{V(0)} - \log \frac{\Phi(\mu(t))}{\Phi(\mu(0))},$$

which gives us the desired equation after noting that by definition $V(0) = 1$. $\square$

From this decomposition formula we can get a good idea of the mechanism that relates functionally generated portfolios to pseudo arbitrages. As previously mentioned the L-divergence can be seen as a kind of non-symmetric metric, in this sense it captures the "distance" between two consecutive market positions. The relative value is determined by two parts, the important one is the drift-term $A(t)$ that captures these "distances", which is basically equivalent to the volatility of $\mu(t)$. This part will drive up the relative value for *sufficiently volatile* market sequences. The first summand in the decomposition is some fluctuation based on the current market weight. We can see that under the condition of sufficient volatility we get unbounded increases in the $A(t)$ term and therefore in the relative value.

Additionally, we only required π to be generated by some positive and concave Φ for the decomposition to hold. While it is also clear that this does not work for affine generators, these are the only requirement for value growth (after some time). This shows the important role functionally generated portfolios play.

2.4 Local properties of MCM

The theorem presented in this section expands on the idea behind Lemma 2. Remember that we showed that if the MCM property fails we can find a cycle along which the relative value goes to zero, meaning the portfolio cannot be a pseudo arbitrage. This theorem will show that such a sequence can even be found very close to a single point. In this sense satisfying (or failing) MCM is a local property.

Figures 2.2 and 2.3 are taken from [PW15, Theorem 8] and [PW14, Theorem 8], respectively.

Theorem 1. *Let π be a portfolio function. If π fails the MCM property the following holds:*

(i) For any $\delta > 0$ we can find a sequence of market weights with jump sizes smaller than δ such that the relative value $V(t)$ of the portfolio goes to zero along that sequence. In particular π cannot be a pseudo arbitrage over any set that contains the sequence.

(ii) For any $\delta > 0$ we can find a $p \in \Delta^{(n)}$ such that the sequence from (i) lies entirely in a δ ball around p.

Proof. (i) In Lemma 2 we have already seen that such a sequence exists with arbitrary jump sizes. Furthermore, we have shown that if a portfolio fails the MCM property it also fails the δ-MCM property (contraposition of Proposition 1 (ii)). That means we have at least one cycle $\{\mu_t\}_{t=0}^m$ with jump sized $|\mu_i - \mu_{i+1}| < \delta$ for all i that fails the MCM property. By using this cycle in Lemma 2 we can construct the desired sequence.

To proof (ii) we will show the following claim, which is the contraposition of the statement.
Claim. Assume that there exists a $\delta > 0$, such that for any $p \in \Delta^{(n)}$ the MCM property holds over $B_\delta(p) \subset \Delta^{(n)}$, then the MCM property holds on the whole simplex.

We quickly review line integrals which are crucial to the proof. For two points p_1, p_2 in the simplex let $\gamma : [0,1] \to \Delta^{(n)}$ be a piecewise linear curve with $\gamma(0) = p_1$ and $\gamma(1) = p_2$. We define the line integral for some function f along the curve γ as

$$I_\gamma(f) := \int_\gamma f(x)dx = \int_0^1 \langle f(\gamma(t)), \gamma'(t)\rangle dt. \tag{2.36}$$

Except for the orientation, the line integral does not depend on the parametrization of the curve.

Throughout we assume that for some $\delta > 0$ the above described local MCM property holds. We will only consider piecewise linear curves in the proof. For ease of notation we write $[p_1, p_2]$ for the curve in $\mathbb{R}^n$ connecting the two points p_1, p_2 by a straight line.

Let $\omega(\mu) := \frac{\pi(\mu)}{\mu}$. The proof consists of two parts:
(a) Fix some $p \in \Delta^{(n)}$, for $p_1, p_2 \in B_\delta(p)$ we have

$$I_{[p_1,p_2]}(\omega) = \log \tilde{\Phi}(p_2) - \log \tilde{\Phi}(p_1) \tag{2.37}$$

where $\tilde{\Phi}$ is the generating function[2] of π over $B_\delta(p)$. In particular, for any closed curve γ in $B_\delta(p)$ it holds that

$$I_\gamma(\omega) = 0. \tag{2.38}$$

(b) For all $p, q \in \Delta^{(n)}$ and all piecewise linear loops γ_1, γ_2 in $\Delta^{(n)}$ that go from p to q. We get

$$I_{\gamma_1}(\omega) = I_{\gamma_2}(\omega). \tag{2.39}$$

Assuming (a) and (b) hold we can define Φ on the whole simplex. Fix some $p_0 \in \Delta^{(n)}$ and set

$$\log \Phi(p) := \int_\gamma \omega(\mu)d\mu \tag{2.40}$$

[2]One has to be careful with this notation as it hides the fact that for each $p \in \Delta^{(n)}$ there is a (potentially) different $\tilde{\Phi}_p$ that satisfy the MCM property **only** in their respective δ-ball. In the proof we will only deal with individual generators so it will always be clear what the corresponding δ-ball is.

for some piecewise linear curve γ from p_0 to p. This is well defined by (b), because the integral doesn't change when choosing different curves γ.

Furthermore, take some r in the simplex, then for all $q \in B_\delta(r)$

$$\log \Phi(q) - \log \tilde{\Phi}(q) = \log \Phi(r) - \log \tilde{\Phi}(r).$$

To see this, rewrite the difference by applying (a) to $I_{[r,q]}(\omega)$. Thus,

$$\log \Phi(q) - \log \tilde{\Phi}(q) = \log \Phi(q) - \log \tilde{\Phi}(r) - I_{[r,q]}(\omega).$$

Let γ be some curve from p_0 to q and use the definition for $\log \Phi(q)$. The two integrals can be combined to get

$$\begin{aligned}
\log \Phi(q) - I_{[r,q]}(\omega) &= I_\gamma(\omega) - I_{[r,q]}(\omega) \\
&\overset{\text{(b)}}{=} I_{[p_0,q]}(\omega) - I_{[r,q]}(\omega) \\
&= I_{[p_0,q]}(\omega) + I_{[q,r]}(\omega) = I_{\gamma'}(\omega)
\end{aligned}$$

Where γ' is some curve from p_0 to r. By (b) the integral $I_{\gamma'}$ is independent of the path and therefore of q. The difference of $\log \Phi$ and $\log \tilde{\Phi}$ is then

$$\log \Phi(q) - \log \tilde{\Phi}(q) = I_{\gamma'}(\omega) - \log \tilde{\Phi}(r) = \log \Phi(r) - \log \tilde{\Phi}(r).$$

Therefore Φ is locally over $B_\delta(r)$, up to constant, equal to $\tilde{\Phi}$, meaning that Φ is at least locally concave. At any point r in the simplex, we have an open convex neighbourhood over which our function Φ is concave. We can use [Hör07, Theorem 2.1.25] to show that Φ is indeed concave on $\Delta^{(n)}$.

In the last step we show that π is MCM with regard to the generating function Φ. As we saw, $\log \Phi(p) - \log \tilde{\Phi}(p)$ is constant for all $p \in B_\delta(r)$ in the simplex. It follows that for every $p \in \Delta^{(n)}$

$$\partial \log \Phi(p) = \partial \log \tilde{\Phi}(p).$$

Intuitively this is immediately clear, because the supergradients of the function define supporting hyperplanes that should be invariant under translations. We can also see this from the definition for $\xi \in \partial \log \Phi(p)$. We have

$$\forall q \in B_\delta(p): \quad \langle \xi, q - p \rangle \geq \log \Phi(q) - \log \Phi(p) = \log \tilde{\Phi}(q) - \log \tilde{\Phi}(p).$$

That means that $\xi \in \partial \log \Phi|_{B_\delta(p)}(p)$ is equivalent to $\xi \in \partial \log \tilde{\Phi}|_{B_\delta(p)}(p)$, where $\tilde{\Phi}$ is the local generator over $B_delta(p)$. As discussed in Proposition 1 (ii), the supergradient is a local property so we can disregard the restriction to $B_\delta(p)$. In Proposition 2 we showed that if the log-superdifferential of two functions are identical they generate the same portfolio. So Φ generates π.

Assuming (a) and (b) hold we able to show that a global generator exists.

Now part (a) and (b) need to be shown.
(a) The aim is to show that the line integral with regard to ω over all closed piecewise linear curves in B_δ is 0. By $\mu(t)$ we explicitly parametrize the curve $l = [p_1, p_2]$ as the convex combination

$$\mu : [0,1] \to \Delta^{(n)}, \quad t \mapsto (1-t)p_1 + tp_2.$$

There are two ways how we can look at the function $\log \Phi(\mu(t))$. First, as a function on the real numbers $t \mapsto \log \Phi(\mu(t))$, or as a function on the simplex $\mu(t) \mapsto \log \Phi(\mu(\cdot))|_t$. Both views are, of course, closely related and usually a distinction would be irrelevant, but in this case it is important to recognize the following fact:

$$D_1 \log \Phi(\mu(\cdot))|_t = D_{p_2-p_1} \log \Phi(\cdot)|_{\mu(t)} \tag{2.41}$$

Again, note the difference in formulation. On the left hand side, the direction vector is just $1 \in \mathbb{R}$ (this is simply a one sided derivative), on the right hand side we have a "typical" directional derivative with directional vector $p_2 - p_1 \in \mathbb{R}^n$.

This relationship follows immediately from the definition of directional derivatives and the fact that $\mu(t) + (p_2 - p_1)h = \mu(t+h)$. The statement hinges on the fact that we are just considering a slice of the simplex along the line l and is not true in general.

At this point note that $\log \Phi(\mu(t))$ is indeed a concave function in t (see appendix). So we can apply [Roc72, Theorem 23.4] which relates the directional derivative to the supergradient. The statement is

$$D_{p_2-p_1} \log \Phi(\cdot)|_{\mu(t)} = \inf_{v \in \partial \log \Phi(\mu(t))} \langle v, p_2 - p_1 \rangle.$$

The superdifferential is closed and bounded so we can find a minimizer v^* to which we can apply Proposition 2 (ii) to get

$$D_{p_2-p_1} \log \Phi(\cdot)|_{\mu(t)} = \left\langle \frac{\pi(\mu(t))}{\mu(t)}, p_2 - p_1 \right\rangle.$$

The normalizing factors that come up in Proposition 2 (ii) are irrelevant here because their scalar product with $p_2 - p_1$ is 0.

Here we already see why our initial step, to use a different directional derivative, paid off. For the directional derivative on $\mathbb{R}$ we do not have the structure from Proposition 2 (ii) available that allowed us to connect Φ to the portfolio π in a meaningful way (the simplex on $\mathbb{R}$ is just a point). So far we have shown that

$$D_1 \log \Phi(\mu(\cdot))|_t = \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu'(t) \right\rangle.$$

Note that $\mu'(t) = p_2 - p_1$, which simply follows from our definition of μ.

At this point, apply [Roc72, Theorem 24.1] which states that for any concave function on $\mathbb{R}$ its one sided derivative (i.e. D_1 or $-D_{-1}$) are non-increasing.

Finally, the statement follows from [Roc72, Corollary 24.2.1]. This corollary is similar to the fundamental theorem of calculus but the concavity of the function allows for an application in situation where only the one-sided derivative is guaranteed. The necessary preconditions, which are satisfied in this setting, are that the concave function needs to be finite on an open interval in $\mathbb{R}$. The corollary then states, that

$$\log \Phi(\mu(1)) - \log \Phi(\mu(0)) = \int_0^1 D_1 \log \Phi(\mu(t))dt$$

From which we can conclude that

$$\log \Phi(\mu(1)) - \log \Phi(\mu(0)) = \int_0^1 \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu'(t) \right\rangle dt = I_l(\omega).$$

We can extend the result to all piecewise linear curves in $B_\delta(p)$. Just decompose the curve into its linear parts and apply the above formula. This proofs claim (a).

(b) We want to show that the line integral with regard to ω over any piecewise linear curve only depends on its start and end point. We take two curves γ_1 and γ_2, both with start point p and end point q, to check if the line integrals over these two curves coincide.

It is sufficient to show the claim for $\gamma_2(t) = (1-t)p + tq$, the straight line from p to q. To see why this suffices let $\tilde{\gamma_i}$ be two arbitrary piecewise linear curves. Assume we already showed the statement with the above γ_2, then we could show

$$I_{\tilde{\gamma}_1}(\omega) = I_{\gamma_2}(\omega) = I_{\tilde{\gamma}_2}(\omega)$$

which gives us the full generality.

To reduce the problem further, we can also assume γ_1 to just be a triangle curve, i.e. stitching two lines together. For two lines $[p, r]$ and $[r, q]$ we define the triangle curve as

$$[p, r] \cup [r, q] := \begin{cases} (1-2t)p + 2tr & 0 \leq t \leq \frac{1}{2}, \\ (2-2t)r + (2t-1)q & \frac{1}{2} \leq t \leq 1. \end{cases} \tag{2.42}$$

This can be seen by an induction. Assume the statement holds for γ_1, a curve with $n \geq 3$ corners. Our task will be to show this assumption later on.

Now take some curve γ_1 with $n + 1$ curves. We denote the corners by $\{p = r_1, r_2, \ldots, r_{n+1} = q\}$. Let $\tilde{\gamma_1}$ be the same curve as γ_1 except that it stops at the

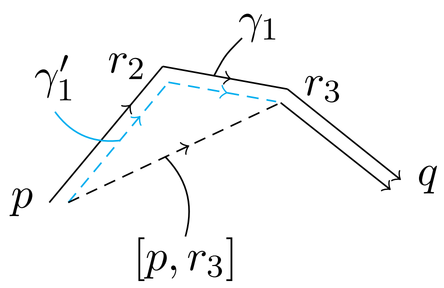

Figure 2.1: Decomposition in the induction for $n = 3$.

r_n corner. So $\tilde{\gamma}_1$ goes from p to r_n and only has n corners. By the induction assumption

$$I_{\tilde{\gamma}_1}(\omega) = I_{[p,r_n]}(\omega)$$

and

$$I_{[p,r_n]\cup[r_n,q]}(\omega) = I_{[r_n,q]}(\omega)$$

hold. The large curve γ_1 was decomposed into one curve with n and one curve with 3 corners (see Figure 2.1).

Putting everything together yields

$$I_{\gamma_1}(\omega) = I_{\tilde{\gamma}_1}(\omega) + I_{[r_n,q]}(\omega) = I_{[p,r_n]}(\omega) + I_{[r_n,q]}(\omega) = I_{[p,q]}(\omega).$$

That means that if we can show the statement for three corners, i.e. a triangle curve, we prove the induction assumption and the statement holds for curves with more corners as well.

Altogether, to show (b) it suffices to show that

$$I_{\gamma_1}(\omega) = I_{\gamma_2}(\omega)$$

holds for any p and q, where γ_1 is any triangle curve from p to q and $\gamma_2(t) = (1-t)p + tq$ is the line from p to q.

We distinguish two cases.

Case 1. The curves γ_1 and γ_2 are close to each other. Meaning that $\sup_{0\leq t\leq 1} ||\gamma_1(t) - \gamma_2(t)|| < \frac{\delta}{2}$.

Basically the only tool we have for this is part (a). Conceptually, we want to generalize part (a) to arbitrary large loops. To do so, we will decompose the large loops into smaller ones that fit into a δ-ball and then apply (a).

We choose points on the line γ_2 that are equidistant apart by less than $\delta/2$. By $0 = u_0 < u_1 < \cdots < u_m = 1$, with each $u_i \in (0,1)$ we denote the index points. We connect each $\gamma_2(u_i)$ to $\gamma_1(u_i)$ by a straight line (see Figure 2.2). The special situation at $u_0 = 0$ and $u_m = 1$ is included in this set up, just note that the "lines" $[\gamma_2(u_0), \gamma_1(u_0)]$ and $[\gamma_2(u_m), \gamma_1(u_m)]$ are actually points. This way we get closed loops for each i which we call $\beta_i := [\gamma_2(u_i), \gamma_1(u_i)] \cup [\gamma_1(u_i), \gamma_1(u_{i+1})] \cup [\gamma_1(u_{i+1}), \gamma_2(u_{i+1})] \cup [\gamma_2(u_{i+1}), \gamma_2(u_1)]$. Here we extended to notation for triangle curves to piecewise linear curves with more corners. As mentioned above, the beginning and end points are included in this with the only difference that these loops only have three instead of four corners. We claim that $\beta_i \in B_\delta(\gamma_2(u_i))$ and therefore part (a) applies. We need to check that all corners of β_i are at most δ

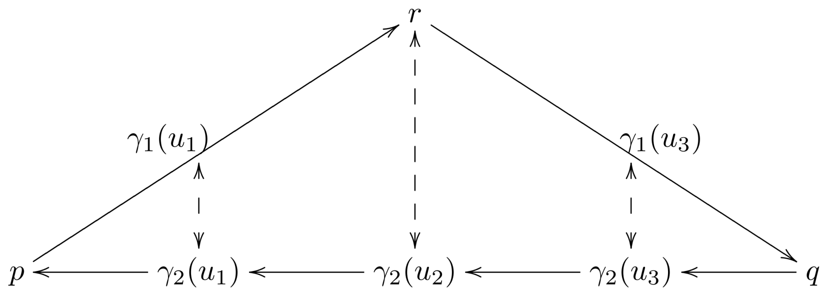

Figure 2.2: Decomposition of a large loop.

apart from $\gamma_2(u_i)$. By construction and assumption it is clear that

$$||\gamma_2(u_i) - \gamma_2(u_{i+1})|| < \frac{\delta}{2}, \quad \text{and} \quad ||\gamma_2(u_i) - \gamma_1(u_i)|| < \frac{\delta}{2}.$$

We can use the triangle inequality to get

$$||\gamma_2(u_i) - \gamma_1(u_{i+1})|| \leq ||\gamma_2(u_i) - \gamma_2(u_{i+1})|| + ||\gamma_2(u_{i+1}) - \gamma_1(u_{i+1})|| < \delta.$$

So β_i is indeed inside a δ-ball. By (a) we get $I_{\beta_i}(\omega) = 0$.

To conclude, we put the decomposed loops back together. Intuitively this will leave the original cycle of γ_1 and γ_2 because the straight lines in-between cancel out. This intuition is indeed correct.

$$\begin{aligned}
0 &= \sum_{i=0}^{m} I_{\beta_i} \\
&= \overbrace{\sum_i I_{[\gamma_1(u_i), \gamma_1(u_{i+1})]}}^{=I_{\gamma_1}} + \overbrace{\sum_j I_{[\gamma_2(u_{j+1}), \gamma_2(u_1)]}}^{=-I_{\gamma_2}} + \sum_k \left(I_{[\gamma_2(u_k), \gamma_1(u_k)]} + I_{[\gamma_1(u_{k+1}), \gamma_2(u_{k+1})]} \right) \\
&= I_{\gamma_1} - I_{\gamma_2} + I_{[p,p]} + I_{[q,q]} \\
&= I_{\gamma_1} - I_{\gamma_2}.
\end{aligned}$$

This proves the first case.

Case 2. The second case deals with curves that are farther apart than $\delta/2$. We want to reduce this case back to Case 1.

We first consider the ball $B_{\delta/4}(r)$, this ball intersects the γ_1 curve at two points, which we denote by r_1 and r_2. We define the point r_3 in the following way. The three points p, q and r define a plane. If we think of $B_{\delta/4}(r)$ as a ball in $\mathbb{R}^n$ (one dimension higher than the plane the simplex lies in) the intersection of the balls boundary $\partial B_{\delta/4}(r)$ with the plane is a circle. There is a unique line through the point r that cuts the $[p, q]$ line orthogonally. We call the intersection of this line with the above mentioned circle r_3 (see Figure 2.3).

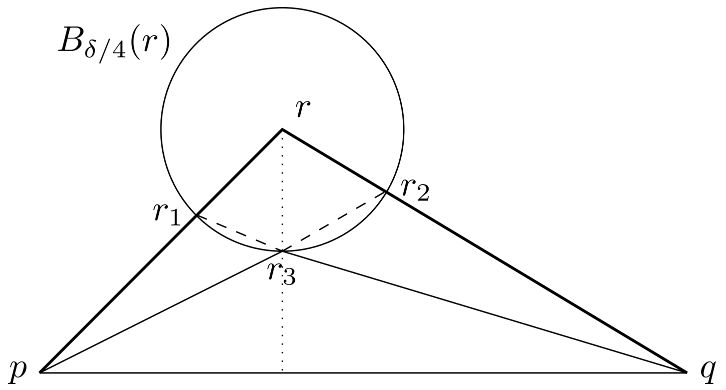

Figure 2.3: Reduction to a smaller triangle. Fat line = γ_1.

As mentioned, we want to reduce this to case 1, more concretely we want a triangle curve where the corner r is close to the line $[p, q]$. In the figure we can see how this will work out. If we can show that the line integral along γ_1 (i.e. along the point r) is identical to the line integral along $[p, r_3] \cup [r_3, q]$, we have an iterative method to replace the large triangle by a smaller one.

The loop along the lines $[r_1, r], [r, r_2], [r_2, r_3]$ and $[r_3, r_1]$, which we will call β_1, is clearly inside $B_\delta(r)$. By Case 1 the line integral over ω along this loop is zero.

We have thus reduced the line integral along γ_1 into the integral along β_1 and the triangular curves $[p, r_1] \cup [r_1, r_3] \cup [r_3, p] =: \beta_2$ and $[r_2, q] \cup [q, r_3] \cup [r_3, r_2] =: \beta_3$. Written in symbols: $I_{\gamma_1} = I_{\beta_1} + I_{\beta_2} + I_{\beta_3} = I_{\beta_2} + I_{\beta_3}$.

First we show that the integral over $[p, r_1] \cup [r_1, r_3]$ is equal to the integral over $[p, r_3]$, which brings us one step closer to reduce the large triangle to the smaller p, r_3, q one. We will show that the two curves are closer than $\delta/2$ apart, hence Case 1 leads to the desired outcome. To do so we estimate the distance from r_1 to the line $[p, r_3]$ from above by the distance $||r_1 - r_3||$, but these points are on the boundary of $B_{\delta/4}(r)$ so their distance is at most $\delta/2$. We can apply Case 1 (after re-indexing the curves to $[0,1]$). The same way we can show that the integral along $[r_3, r_2] \cup [r_2, q]$ is equal to the integral along $[r_3, q]$.

We have seen that

$$I_{[p,r]\cup[r,q]}(\omega) = I_{[p,r_3]\cup[r_3,q]}(\omega).$$

However, the distance between the curve along r_3 and the curve $[p,q]$ is $\delta/4$ less than the distance between the large triangle along p and the line $[p,q]$. This means that after applying this scheme finitely many times we reach a situation where we can apply Case 1.

To summarize, this second step was possible because the distance from the original triangle to the p, r_1, q triangle was small enough that all the "compensation" curves β_1, β_2 and β_3 were covered by Claim 1.

This shows the second claim which was the last thing to show. □

2.5 Characterization of pseudo-arbitrage

So far we have only dealt with functionally generated portfolios. In the decomposition formula we could see how pseudo-arbitrages could relate to functionally generated portfolios, in this theorem we will make these findings more manifest and firmly connect functionally generated portfolios to pseudo-arbitrages.

Theorem 2. *Let π be a portfolio function on the open and convex set $K \subset \Delta^{(n)}$. Then, π is a pseudo-arbitrage if and only if there exists a concave function $\Phi : \Delta^{(n)} \to (0, \infty)$ that satisfies the following properties*

(i) Φ restricted to K is not an affine function,

(ii) There exists an $\epsilon > 0$ such that $\inf_{p \in K} \Phi(p) \geq \epsilon$,

(iii) For all $p \in K$ the vector

$$\left(\frac{\pi_i}{p_i} - \frac{1}{n} \sum_{j=1}^{n} \frac{\pi_j}{p_j} \right)_{i=1}^{n}$$

is in the supergradient of $\log \Phi$ at the point p.

Remember that condition (iii) is by Proposition 2 equivalent to Φ generating π.

Proof. Assume that (i), (ii) and (iii) hold. We use the decomposition formula (2.35). Note that a positive concave function on a bounded set is bounded from above. This follows almost immediately from the definition. By (ii), for any sequence $\{\mu(t)\}_{t=1}^{\infty} \subset K$ we have

$$0 < C_1 \leq \frac{\epsilon}{\sup_{x \in K} \Phi(x)} \leq \frac{\Phi(\mu(t))}{\Phi(\mu(0))} \leq \frac{\sup_{x \in K} \Phi(x)}{\epsilon} \leq C_2.$$

Accordingly, we get that $\log \frac{\Phi(\mu(t))}{\Phi(\mu(0))}$ is bounded for all market sequence. From our previous deliberations we know, because Φ is not affine on K, the L-divergence is non-zero at every point. In particular we can construct a sequence, such that the relative value goes to infinity. To do so, one possibility is to choose any $\mu_1, \mu_2 \in K$ with $T(\mu_1|\mu_2) = d_1 > 0$ and $T(\mu_2|\mu_1) = d_2 > 0$ (meaning μ_i that are not identical). Then define $\{\tilde{\mu}(t)\}_{t=0}^{\infty} = \{\mu_1, \mu_2, \mu_1, \mu_2, \dots\}$.

The drift term satisfies $\lim_{t \to \infty} A(t) = \sum_{t=1}^{\infty}(d_1 + d_2) = \infty$, so the relative value goes to infinity for some market sequence by the decomposition formula. Furthermore, we have seen above that $\log \frac{\Phi(\mu(t))}{\Phi(\mu(0))}$ is bounded from below by some constant $C(K, \pi) = \log(C_1)$. Therefore the log-relative value is bounded independent of the choice of market sequence by $\log V(t) \geq \log(C_1)$. We have shown that π is indeed a pseudo-arbitrage.

For the converse assume that π is a pseudo-arbitrage. A necessary condition for this is that π is functionally generated, otherwise we could construct a contradicting sequence of market weights (see Lemma 2). From this we can apply Proposition 2 to obtain (iii).

To show (i), use the representation formula again. Φ is bounded on the simplex, if $A(t)$ were bounded as well, the pseudo arbitrage property would be impossible. Remember that the L-divergence is zero if and only if the underlying Φ is affine on the line between the arguments. In turn it is impossible for Φ to be affine on K, because otherwise the unboundedness of $\log V(t)$ would be impossible.

We are left to show point (ii), i.e. that there exists a positive constant that bounds Φ from below in K. We extend Φ to $\overline{\Delta^{(n)}}$ and denote the extension by Φ as well. If we assume that the infimum actually attains 0, there exists a sequence in K that converges to some $q \in \overline{\Delta^{(n)}}$ such that $\Phi(q) = 0$. We want to show that this assumption leads to the value process converging to zero, which would be a contradiction.

Choose some point $p \in K$ and a sequence $\{\lambda(t)\}_{t=0}^{\infty}$ that is monotonically increasing in $[0, 1)$. We then define the convex combination of p and q in the following way.

$$\mu(t) := (1 - \lambda(t))p + \lambda(t)q.$$

This is of course a sequence $\{\mu(t)\}_{t=0}^{\infty}$ depending on λ and indirectly on t. During this proof we will formulate a few requirements on $\lambda(t)$ that we have to additionally impose on the sequence. The first requirement is that we choose the points $\lambda(t)$

in a way that $\log \Phi(\mu(t))$ is differentiable for all t. This is possible because Φ is differentiable almost everywhere. Along $\mu(t)$ the gradient and superdifferential coincide. Together with Proposition 2 we get

$$\left\langle \frac{\pi(p)}{p}, q-p \right\rangle = \langle \partial \log \Phi(p), q-p \rangle = D_{q-p} \log \Phi(p).$$

By inserting two consecutive elements of the sequence $\mu(t)$ and noting that $\mu(t+1) - \mu(t) = q - p(\lambda(t+1) - \lambda(t))$ we have

$$\left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle = D_{q-p} \log \Phi(\mu(t))(\lambda(t+1) - \lambda(t)).$$

We estimate the left hand side from below using the inequality $\log(1+x) \leq x$ for $x > -1$. The condition for $x = \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle$ is satisfied because by (iii) Φ generates π so we can use the proof of Lemma 4 (iii) where we showed that $1 + \langle \frac{\pi(p)}{p}, q-p \rangle > \frac{\Phi(q)}{\Phi(p)} \geq 0$ for all p and q.

After summing over all t we arrive at

$$\sum_{t=0}^{\infty} \log \left(1 + \left\langle \frac{\pi(p)}{p}, \mu(t+1) - \mu(t) \right\rangle \right) \leq \sum_{t=0}^{\infty} D_{q-p} \log \Phi(\mu(t))(\lambda(t+1) - \lambda(t)). \tag{2.43}$$

The left hand side can already be recognized as $\log V(t)$, the log relative value function for the sequence $\mu(t)$. If we can show that this $\lim_{t\to\infty} \log V(t) = -\infty$ we have found a contradiction because by definition a pseudo arbitrage can not have a log-relative-value function that is unbounded from below. For this we will use the previous theorem. The idea is to estimate the right hand side of (2.43) by the line integral

$$\int_{[p,q]} \frac{\pi(\xi)}{\xi} d\xi = \log \Phi(q) - \log \Phi(p) = -\infty.$$

The evaluation of the integral is a direct consequence of part (a) and (b) in the proof of Theorem 1. To make the notation a little bit easier we define

$$f(\lambda(t)) := D_{q-p} \log \Phi(\mu(t)).$$

We also use the same idea as in the proof of Theorem 1 in conjunction with the fact that $t \mapsto \lambda(t)$ is monotonically increasing, to show that $t \mapsto f(\lambda(t))$ is monotone decreasing. Remember that we used the convex combination structure of μ to relate the directional derivative to a one-sided derivative and then applied the one-dimensional convex analysis result [Roc72, Theorem 24.1] to obtain the monotonicity.

We can view the sequence $\{\lambda(t)\}_{t=1}^{N}$ for some $N \in \mathbb{N}$ as a partition of the interval $[0, \lambda(N)]$. We can then connect the integral over f with the aforementioned line integral in the following way.

$$\begin{aligned}\int_0^{\lambda(N)} f(x)dx &= \int_0^{\lambda(N)} D_{q-p} \log \Phi((1-x)p + xq)dx \\ &= \int_0^{\lambda(N)} \langle q-p, \nabla \log \Phi((1-x)p + xq)\rangle dx.\end{aligned}$$

Although we have no special regularity assumptions on Φ, the second equation is unproblematic because Φ is differentiable almost everywhere. Taking the limit $N \to \infty$ yields

$$\int_0^1 f(x)dx = \int_0^1 \langle q-p, \nabla \log \Phi((1-x)p + xq)\rangle dx.$$

With the usual technique involving Proposition 2 we can rewrite the gradient expression as

$$\int_0^1 f(x)dx = \int_0^1 \left\langle q-p, \frac{\pi((1-x)p+xq)}{(1-x)p+xq} \right\rangle dx.$$

Let $\gamma(t) = (1-t)p + tq$ be the straight line connecting p and q, then the above equation becomes

$$\int_0^1 f(x)dx = \int_0^1 \left\langle \gamma'(x), \frac{\pi(\gamma(x))}{\gamma(x)} \right\rangle dx = \int_{[p,q]} \frac{\pi(\xi)}{\xi} d\xi = -\infty. \tag{2.44}$$

In the final step we estimate the integral on the left hand side, by the right hand side in (2.43). Again consider the integral over $f(x)$ in the interval $[0, \lambda(N)]$. The points $\{\lambda(t)\}_{t=0}^{N}$ partition the interval. We can estimate the integral from below by summing over the width of the intervall times the minimal value of the function in the respective interval. This is closely related to the construction of the Riemann integral. We have

$$\begin{aligned}\int_0^{\lambda(N)} f(x)dx &\geq \sum_{t=0}^{N} \inf_{\lambda(t)\leq x\leq\lambda(t+1)} f(x)(\lambda(t+1) - \lambda(t)) \\ &= \sum_{t=0}^{N} f(\lambda(t+1))(\lambda(t+1) - \lambda(t))\end{aligned}$$

where the last equality follows from the monotonicity of f.

We now state another condition on the sequence $\{\lambda(t)\}_t$. We require that

$$\lambda(t+1) - \lambda(t) \geq \lambda(t+2) - \lambda(t-1), \qquad \text{for all } t.$$

In other words, we want the distance between two consecutive points to become smaller and smaller. This is always possible even when considering the other conditions on the sequence.

Using this fact we can rewrite the integral estimate as

$$\begin{aligned}\int_0^{\lambda(N)} f(x)dx &\geq \sum_{t=0}^{N} f(\lambda(t+1))(\lambda(t+2) - \lambda(t+1)) \\ &= \sum_{t=1}^{N+1} f(\lambda(t))(\lambda(t+1) - \lambda(t)).\end{aligned}$$

Taking the limit $N \to \infty$ leaves us with

$$\int_0^1 f(x)dx \geq \sum_{t=1}^{\infty} f(\lambda(t))(\lambda(t+1) - \lambda(t)).$$

This together with (2.43) and (2.44) is the contradiction that the relative value of a pseudo arbitrage can be unbounded from below if we assume 0 to be a limit point of the generator Φ. $\square$

2.6 Differentiable portfolios

In Proposition 3 we have seen that the regularity of π is related to the regularity of the generator Φ. We will now look at MCM portfolios that are at least C^1, by the aforementioned proposition the generator of such a portfolio has to be in C^2.

Lemma 6. *Let $T(q|p)$ be the L-divergence we defined before, then for $p \in \Delta^{(n)}$ and $v \in T\Delta^{(n)}$*

$$T(p + tv|p) = H(p)(tv, tv) + o(t^2). \tag{2.45}$$

We define H as

$$\frac{-1}{2\Phi(p)} \frac{d^2}{dt^2} \Phi(p + tv)\Big|_{t=0} =: H(p)(v, v).$$

and call it the drift quadratic form of (π, Φ).

The proof can be found in Appendix Section C.

We define two objects that will appear later on.

Definition 5. (Excess growth)
For a portfolio $\pi \in \overline{\Delta^{(n)}}$ we define the *excess growth quadratic form* Γ_π by

$$\Gamma_\pi(v, v) = \frac{1}{2} \sum_{i,j=1}^{n} \frac{\pi_i(\delta_{ij} - \pi_j)}{p_i p_j} v_i v_j. \tag{2.46}$$

For $p \in \Delta^{(n)}$ and $v \in T\Delta^{(n)}$.

Definition 4. (Fisher information metric)
We define a metric for tangent vectors $u, v \in T\Delta^{(n)}$ by

$$\langle\langle u, v\rangle\rangle_p = \frac{1}{2} \sum_{i=1}^{n} \frac{1}{p_i} u_i v_i \tag{2.47}$$

where $p \in \Delta^{(n)}$. We call this metric the *Fisher information metric.*

The two definitions are related to each other in the following way. Assume we have the market portfolio, $\pi(p) = p$, then $\Gamma_\pi(u, v) = \langle\langle u, v\rangle\rangle_p$. This can be seen by a simple calculation.

We can extend the concept of a directional derivative to functions that take values in the simplex. Formally this is done by looking at differential maps which are pushforwards of tangent maps. As shown in the appendix (Lemma A2), in this setting we can associate the tangent map dF_v with the linear map $v \mapsto JF \cdot v$, where JF is the Jacobian matrix of F. A tangent vector v (at point p) gets mapped to a tangent vector dF_v (at point $F(p)$). The matrix vector product formulation is sufficient for our needs and already offers a good intuition, we will thus take this viewpoint in the next proposition.

Proposition 4. *Let π be a portfolio generated by a positive and concave C^2-function. Then the following holds*

(i) $\omega(\mu) := \frac{\pi}{\mu}$ *satisfies*

$$\langle v, J\omega(p) \cdot v\rangle \leq -\langle \omega(p), v\rangle^2 \tag{2.48}$$

for all $\mu \in \Delta^{(n)}$ and $v \in T\Delta^{(n)}$.

(ii) for all $p \in \Delta^{(n)}$ and $v \in T\Delta^{(n)}$ we have

$$\Gamma_\pi(p)(v, v) - \langle\langle J\pi(p) \cdot v, v\rangle\rangle_p \geq 0 \tag{2.49}$$

where Γ_π is the excess growth quadratic form of π and $\langle\langle \cdot, \cdot\rangle\rangle_p$ is the Fisher information metric we defined above.

Proof. (i) Let $\epsilon > 0$. We define a cycle $\{\mu(t)\}_{t=0}^2$ by $\mu(0) = p$, $\mu(1) = p + \epsilon v$ and $\mu(2) = \mu(0)$. By assumption π is a generated portfolio on the simplex so we can apply the MCM property to this cycle. Note that $\mu(2)-\mu(1) = \mu(0)-\mu(1) = -\epsilon v$. We obtain

$$\left(1 + \left\langle \frac{\pi(p)}{p}, \epsilon v \right\rangle\right)\left(1 + \left\langle \frac{\pi(p+\epsilon v)}{p+\epsilon v}, -\epsilon v \right\rangle\right) \geq 1.$$

Expanding the product yields

$$\epsilon\langle \omega(p) - \omega(p+\epsilon v), v\rangle - \epsilon^2\langle \omega(p), v\rangle\langle \omega(p+\epsilon v), v\rangle \geq 0. \tag{2.50}$$

Now divide both sides by ϵ^2 and let ϵ go to zero. We calculate the limit term by term. The first term has the limit

$$-\left\langle \frac{\omega(p+\epsilon v) - \omega(p)}{\epsilon}, v \right\rangle \overset{\epsilon\to 0}{\longrightarrow} -\langle \{D_v\omega_i(p)\}_{i=1}^n, v\rangle = \langle J\omega(p)\cdot v, v\rangle.$$

This follows directly from the definition of the directional derivative and the fact that

$$\big(D_v\omega_i(p)\big)_{i=1}^n = \begin{bmatrix} \langle \nabla\omega_1(p), v\rangle \\ \langle \nabla\omega_2(p), v\rangle \\ \cdots \\ \langle \nabla\omega_n(p), v\rangle \end{bmatrix} = \begin{bmatrix} \nabla^T\omega_1(p) \\ \nabla^T\omega_2(p) \\ \cdots \\ \nabla^T\omega_n(p) \end{bmatrix} \cdot v = J\omega(p)\cdot v. \tag{2.51}$$

Note that the third and fourth term are a matrix vector product.
The second term in (2.50) has the limit

$$-\langle \omega(p), v\rangle\langle \omega(p+\epsilon v), v\rangle \overset{\epsilon\to 0}{\longrightarrow} -\langle \omega(p), v\rangle^2$$

if $\omega(p)$ is continuous. But this follows immediately from the regularity of π which we in this section assumed to be C^1. Putting everything back into (2.50) proves the claim.

(ii) We have $\Phi : \Delta^{(n)} \to (0,\infty)$ convex and in C^2. Then

$$\frac{1}{2}\langle v, J\omega\cdot v\rangle = \frac{1}{2}\sum_{i,j=1}^n v_j v_i \partial_i\omega_j = \frac{1}{2}\sum_{i,j=1}^n v_j v_i \frac{\partial}{\partial\mu_i}\frac{\pi_j(\mu)}{\mu_j}.$$

We apply the quotient rule and expand, thus

$$\begin{aligned}\frac{1}{2}\langle v, J\omega\cdot v\rangle =& \frac{1}{2}\sum_{i,j=1}^n v_j v_i \left(\frac{\partial\pi_j(\mu)}{\partial\mu_i}\mu_j - \overbrace{\frac{\partial\mu_j}{\partial\mu_i}}^{=\delta_{ij}}\mu_j\pi_j(\mu)\right)\frac{1}{\mu_j^2} \\ =& \frac{1}{2}\sum_{i,j=1}^n v_j v_i \frac{1}{\mu_j}\frac{\partial\pi_j(\mu)}{\partial\mu_i} - \sum_{i=1}^n \frac{1}{2}\frac{\pi_i(\mu)}{\mu_i^2}v_i^2.\end{aligned}$$

The first summand can be written in compact form as

$$\frac{1}{2}\sum_{i,j=1}^{n} v_j v_i \frac{1}{\mu_j}\frac{\partial \pi_j(\mu)}{\partial \mu_i} = \frac{1}{2}\sum_{j=1}^{n} v_j \frac{1}{\mu_j}\sum_{i=1}^{n} v_i \frac{\partial}{\partial \mu_i}\pi_j(\mu) = \frac{1}{2}\sum_{j=1}^{n} v_j \frac{1}{\mu_j}(J\pi \cdot v)_j$$

which we can rewrite with Fischer information metric as

$$\langle\langle v, J\pi \cdot v\rangle\rangle_\mu.$$

Putting everything together we get

$$\frac{1}{2}\langle v, J\omega \cdot v\rangle = \langle\langle v, J\pi \cdot v\rangle\rangle_\mu - \frac{1}{2}\sum_{i=1}^{n} \frac{\pi_i(\mu)}{\mu_i^2} v_i^2.$$

We use this fact for

$$\frac{1}{2}\langle v, J\omega \cdot v\rangle + \frac{1}{2}\langle \omega, v\rangle^2 = \langle\langle v, J\pi \cdot v\rangle\rangle_\mu - \frac{1}{2}\sum_{i=1}^{n} \frac{\pi_i(\mu)}{\mu_i^2} v_i^2 + \frac{1}{2}\langle \omega, v\rangle^2. \tag{2.52}$$

If we can show that

$$-\frac{1}{2}\sum_{i=1}^{n} \frac{\pi_i(\mu)}{\mu_i^2} v_i^2 + \frac{1}{2}\langle \omega, v\rangle^2 = \Gamma_\pi(\mu)(v, v)$$

we can estimate the left side of (2.52) by 0 according to part (i) and would be finished. For this last step we have

$$\begin{aligned}
-\frac{1}{2}\sum_{i=1}^{n} \frac{\pi_i(\mu)}{\mu_i^2} v_i^2 + \frac{1}{2}\langle \omega, v\rangle^2 &= -\frac{1}{2}\sum_{i=1}^{n} \frac{\pi_i}{\mu_i^2} v_i^2 + \frac{1}{2}\sum_{i,j}^{n} \frac{\pi_i \pi_j}{\mu_i \mu_j} v_i v_j \\
&= -\frac{1}{2}\sum_{i,j}^{n} \left(\frac{\pi_i}{\mu_i \mu_j} v_i v_j \delta_{ij} - \frac{\pi_i \pi_j}{\mu_i \mu_j} v_i v_j \right) \\
&= -\frac{1}{2}\sum_{i,j}^{n} \frac{\pi_i}{\mu_i \mu_j} v_i v_j (\delta_{ij} - \pi_i) \\
&= \Gamma_\pi(\mu)(v, v).
\end{aligned}$$

Putting everything together into (2.52) and doing the aforementioned estimate leads to the claim. □

Remark 1. At this point we can formulate another characterisation of C^1 MCM portfolios. A continuously differentiable portfolio is MCM if and only if equation

(2.48) holds and all line integrals along closed piecewise linear curves over $\omega = \frac{\pi}{p}$ are zero. We give a sketch of the proof.

The first direction is clear immediately because for an MCM portfolio π Proposition 4 (2.48) holds and by Theorem 1 ω is conservative for piecewise linear curves.

For the reverse, assume (2.48) holds and ω is conservative for piecewise linear curves. We define $\log \Phi$ as in Theorem 1 (2.40). We assumed the conservative property for piecewise linear curves so the definition is unambiguous. Take some points $p, p', q \in \Delta^{(n)}$, then

$$D_{q-p} \log \Phi(p') = \lim_{h \downarrow 0} \frac{\log \Phi(p' + h(q-p)) - \log \Phi(p')}{h}.$$

By inserting the definition of $\log \Phi$ we can see that the integrals partly cancel each other and therefore

$$D_{q-p} \log \Phi(p') = \lim_{h \downarrow 0} \frac{1}{h} \int_{\gamma} \omega = \lim_{h \downarrow 0} \frac{1}{h} \int_0^1 \langle \omega(\gamma(t)), q-p \rangle \; dt$$

where $\gamma(t)$ is a the convex combination curve from p' to $p' + h(q-p)$. All functions are continuous (even differentiable) on the bounded simplex so we can pull the limit inside the integral to obtain

$$D_{q-p} \log \Phi(p') = \langle \omega(p'), q-p \rangle. \tag{2.53}$$

We use this to first show equation (2.18). Let $q = e(i)$ and $p' = p$ then

$$D_{e(i)-p} \log \Phi(p) = \langle \omega(p), e(i) - p \rangle = \frac{\pi_i(p)}{p_i} - 1.$$

We want to apply Propositioin 3 (iv), which states that a concave (and differentiable) function Φ that satisfies (2.18) generates π. The concavity is left to show.

If we can show that $\log \Phi$ is concave on every line connecting p, q for any $p, q \in \Delta^{(n)}$ then $\log \Phi$ is concave. To reduce to one dimension define

$$f : t \mapsto \log \Phi(\gamma(t)), \text{ where } \gamma(t) = (1-t)p + tq.$$

Concavity of f for any all p, q implies the concavity of $\log \Phi$. We can use the same idea as in Theorem 1 (see (2.41)) to connect the directional derivative to the regular derivative, namely

$$f'(t) = D_1 \log \Phi(\gamma(\cdot))|_t = D_{q-p} \log \Phi(\cdot)|_{\gamma(t)} = \langle \omega(\gamma(t)), q-p \rangle$$

by choosing $p' = \gamma(t)$ in (2.53). The second derivative is then, by the same method,

$$\begin{aligned}
f''(t) = D_1\langle \omega(\gamma(t)), q-p\rangle &= D_{q-p}\langle \omega(\cdot), q-p\rangle|_{\gamma(t)} \\
&= \left\langle \left(D_{q-p}\omega_i(\gamma(t))\right)_{i=1}^{n}, q-p\right\rangle \\
&\overset{(2.51)}{=} \langle J\omega \cdot (q-p), q-p\rangle \\
&\overset{(2.48)}{\leq} -\langle \omega(\gamma(t)), q-p\rangle^2 \\
&\leq 0.
\end{aligned}$$

Finally, using the definition of concavity it is easy to show that when $\log \Phi$ is concave this extends to Φ as well. By applying Proposition 3 (iv) we get the result.

Chapter 3

Optimal transport

We quickly review some central concepts in optimal transport. A good reference is [Vil08]. Specify two (polish) spaces X, Y and a cost function $c : X \times Y \to \mathbb{R} \cup \{\infty\}$. For two probability measures ρ and τ on X and Y, respectively, we define the probability *coupling* q as a probability measure on the product space whose marginals are ρ and τ. That means that

$$q(A \times Y) = \rho(A) \quad \text{and} \quad q(X \times B) = \tau(B)$$

for some sets A and B in the sigma-algebras of X and Y, respectively. By $\Pi(\rho, \tau)$ we denote the set of all couplings between ρ and τ.

The *value of the transport problem* is given by

$$\inf_{q \in \Pi(\rho,\tau)} \int_{X \times Y} c(x, y)\ q(dx, dy) = \inf_{q \in \Pi(\rho,\tau)} \mathbb{E}_q[c]. \tag{3.1}$$

Every coupling q that minimizes the transport value is called *optimal*.

We call a subset $G \subset X \times Y$ *c-cyclical monotone* (c-cm) if for any $n \in \mathbb{N}$ and any choice of points $(x_1, y_1), \ldots, (x_n, y_n)$ in G, the inequality

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

holds. The idea here is that the pairs (x_i, y_i) are optimal in the sense that permutating the pairs would increase the summed cost. One can show that it is equivalent to permutate the x_i on the right hand side instead of the y_i.

Lastly we cite a central theorem from optimal transport theory we will need.

Proposition 5. (Kantorovich duality) [Vil08, Thm 5.10]
Let the cost function $c : X \times Y \to \mathbb{R}$ be lower semi-continuous and bounded below. Further assume that the value (3.1) of the transport problem is bounded, then the coupling R is optimal, if and only if, it is concentrated on a c-cm set.

3.1 Optimal transport on the simplex

We will first consider a transport problem on the simplex. For this section specify two spaces $X = \overline{\Delta^{(n)}}$ and $Y = [-\infty, \infty)^n$, with the cost function

$$c(p, h) = \log\left(\sum_{i=1}^{n} e^{h_i} p_i\right). \tag{3.2}$$

In the following proposition we will associate the MCM property to c-cm via the above definition of π.

Proposition 6 .
Let $h : \Delta^{(n)} \to [-\infty, \infty)^n \setminus \{-\infty, \dots, -\infty\}$. The support of $(x, h(x))$ is c-cyclical monotone if, and only if, the portfolio π defined by

$$\frac{\pi_i(p)}{p_i} = \frac{e^{h_i(p)}}{\sum_{j=1}^{n} e^{h_j(p)} p_j}, \quad i = 1, \dots, n \tag{3.3}$$

is MCM.

Proof. Note that π is indeed a portfolio. It can easily be checked that $\sum_i \pi_i = 1$ and all components are non-negative.

Let $\{\mu(t)\}_{t=0}^{m+1}$ be a cycle and π the portfolio defined by (3.3). We use the definition for the relative value $V(t)$. For better readability we write $h = h(\mu(t))$. We have

$$\frac{V(t+1)}{V(t)} = \sum_{i=1}^{n} \pi_i(\mu(t)) \frac{\mu_i(t+1)}{\mu_i(t)} = \sum_{i=1}^{n} \frac{e^{h_i} \mu_i(t+1)}{\sum_{j=1}^{n} e^{h_j} \mu_j(t)} = \frac{\big\langle e^h, \mu(t+1)\big\rangle}{\langle e^h, \mu(t)\rangle}.$$

We can now take the product over all t on both sides, note that

$$\prod_{t=0}^{m} \frac{V(t+1)}{V(t)} = \frac{V(m+1)}{V(0)} = V(m+1).$$

Taking product over the indices up to $m+1$ and then the logarithm gives us

$$\log V(m+1) = \log \prod_{t=0}^{m} \frac{\big\langle e^h, \mu(t+1)\big\rangle}{\langle e^h, \mu(t)\rangle} = \sum_{t=0}^{m} \Big(\log\big\langle e^h, \mu(t+1)\big\rangle - \log\big\langle e^h, \mu(t)\big\rangle\Big).$$

On the right hand side we can see that the cost function appeared. Remember that h is a function dependant on $\mu(t)$, so we have

$$\log V(m+1) = \sum_{t=0}^{m} c\big(\mu(t+1), h(\mu(t)\big) - \sum_{t=0}^{m} c\big(\mu(t), h(\mu(t)\big). \tag{3.4}$$

If we now assume π to be MCM, we have $\log V(m+1) \geq 0$ and get

$$\sum_{t=0}^{m} c\big(\mu(t+1), h(\mu(t))\big) \geq \sum_{t=0}^{m} c\big(\mu(t), h(\mu(t))\big). \tag{3.5}$$

We did not put any restrictions on the choice of $\mu(t)$ (except it being a cycle, i.e. $\mu(0) = \mu(m+1)$ which is required for both the MCM and ccm definition), so we indeed have an equivalent definition for h being c-cm.

The converse also follows immediately, start with (3.5) but (3.4) still holds. So for all cycles $\mu(t)$ the right hand side of (3.4) is positive giving us $V(m+1) \geq 1$. □

We can now show the following theorem, which connects the solution to a transport problem with pseudo arbitrages. To be more precise, part (i) states a method that allows us to formulate a pseudo-arbitrage from the solution of the above transport problem. And part (ii) achieves the reverse result namely, for any pseudo-arbitrage we can find a corresponding solution to the transport problem.

As one one might already guess, the mechanism that connects the functionally generated portfolio world to the optimal transport world is the above proposition relating MCM and c-cm.

Theorem 3. *Let $R \in \Pi(\rho, \nu)$ be the solution to the optimal transport problem (3.1) with cost (3.2).*

(i) *Take a subset $K \subset \Delta^{(n)}$ and a function $F : K \to [-\infty, \infty)^n \backslash \{-\infty, \cdots, -\infty\}$ be such that the graph $(p, F(p))$ belongs to the support of R for all $p \in K$ (i.e. F is a local transport map solution). Define the portfolio π by*

$$\frac{\pi_i(\mu)}{\mu_i} = \frac{e^{h_i(\mu)}}{\sum_{j=1}^{n} e^{h_j(\mu)} \mu_j} \tag{3.6}$$

where we set $h = F(\mu)$. Then there exists a concave $\Phi : \Delta^{(n)} \to (0, \infty)$ and part (iii) of Theorem 2 holds. In particular, π is a pseudo arbitrage whenever (i) and (ii) in Theorem 2 hold.

(ii) *Assume π is a pseudo arbitrage over $K \subset \Delta^{(n)}$ and additionally let the set $\{\log(\pi(p)/p), p \in K\}$ be bounded from below. We can define a transport map $T : \Delta^{(n)} \to (-\infty, \infty)^n, \quad p \mapsto \log(\frac{\pi_i(p)}{p_i})$. In line with the first part, we define $T(p) =: h$. Fix a probability measure P on K. And define $Q := T_*(P)$ (i.e. as the the pushforward of P under T). So for $X \sim P$ we have $T(X) \sim Q$. Then, $(\mu, T(\mu)) \in \Pi(P, Q)$ and solves the optimal transport problem (3.1) with cost (3.2).*

Proof. (i) We are only interested in what happens on the subset K. The structure of the proof is as follows.

$$\begin{aligned} R \text{ is optimal coupling} &\overset{\text{Claim}}{\Longrightarrow} \text{support}(R) \text{ is c-cm} \\ &\overset{\text{Prop. 6}}{\Longrightarrow} \pi \text{ is MCM} \\ &\overset{\text{Def.}}{\Longrightarrow} \pi \text{ functionally generated by } \Phi. \end{aligned}$$

The first implication is left to show. The graph of F is a subset of the support of R so the second implication follows from Proposition 6.
The last implication is, of course, just the definition of functionally generated portfolios which is equivalent to part (iii) in Theorem 2 via Proposition 2.

On first sight the claim looks like the well-known from optimal transport result Proposition 5 but in our case we do not have the necessary boundedness from below in the cost function. In our particular situation the statement can be proven nonetheless. The idea is to localize the transport problem such that the cost function is bounded and we can apply the standard result. Subdivide the space into sets $Z_m \subset Z := \Delta^{(n)} \times [-\infty, \infty)^n \setminus \{-\infty, \cdots, -\infty\}$ by

$$Z_m = \left\{ (\mu, h) \in Z : \min_{1 \le i \le n} \mu_i \ge \frac{1}{m}, \quad \max_{1 \le i \le n} h_i \ge -m, \right\}, \quad m \ge 1.$$

We can immediately see, that Z_m are an increasing (meaning $Z_m \subset Z_{m+1}$) set sequence, with $\lim_{m\to\infty} Z_m = Z$. We now restrict R to Z_m and after normalizing call the resulting probability measure R_m. These R_m are well definied as long as $R(Z_m) > 0$. From the continuity of measures, we have

$$\sum_{i=1}^{k} R(Z_i) \ge R\left(\bigcup_{i=1}^{k} Z_i \right) \overset{k\to\infty}{\longrightarrow} R(Z) = 1.$$

We can therefore find some k large enough such that $R(Z_m) > 0$ for all $m \ge k$. We neglect all Z_m with an index smaller than k, thus the restriction on Z_m is well defined. Let P_m and Q_m be the marginals of R_m. On the restricted sets we define the transport problem

$$\inf_{q \in \Pi(P_m, Q_m)} \mathbb{E}_q[c]. \tag{3.7}$$

The optimal couplings for each restricted transport problem, denoted by R'_m, are supported on Z_m, where the cost function is bounded from below. Note that because R is optimal on Z and $\text{supp}(R_m) \subset \text{supp}(R)$. The coupling R_m inherits its optimality from R. This is due to the fact that optimality is equivalent to c-cm of the support, i.e. a geometric property of the space. In particular, all couplings that are supported on a subset of the support of an optimal coupling are optimal as well. This means we can set $R'_m = R_m$ because R_m is certainly a solution of (3.7). Furthermore, the cost function is bounded on each Z_m and R_m is an optimizer of the problem (3.7), thus by Proposition 5 $\text{supp}(R) \cap Z_m$ is c-cm.

We finally show that $\text{supp}(R)$ is c-cm as well. Choose an arbitrary $l \in \mathbb{N}$ and a sequence $\{x(k), y(k)\}_{k=1}^{l}$ in the support of R. If we can choose an m large enough such that the whole sequence is contained in Z_m the c-cm property follows.

The x component poses no problem because the simplex is open and the Z_m converge to the full space so we can always choose a large m to encompass the sequence. The y component also poses no problem. If $(y(k))_k \subset (-\infty, \infty)^n$ this certainly works because the set is open. If $y(k)$ attains $-\infty$ in some or multiple components (but not in all) we still can find an m because

$$\Delta^{(n)}|_{Z_m} \times \underbrace{\big([-\infty, \infty) \times \cdots \times [-m, \infty) \times \cdots \times [-\infty, \infty)\big)}_{\text{for all permutations of this carthesian product}} \subset Z_m.$$

The c-cm property of the support of R follows and therefore the graph of F is c-cm as well. As mentioned above this was the only thing to show.

(ii) The structure of the proof is,

$$\begin{aligned} \pi \text{ is pseudo arbitrage on } K &\overset{\text{Lemm. 2}}{\Longrightarrow} \pi \text{ is MCM on } K \\ &\overset{\text{Prop. 6}}{\Longrightarrow} \text{support } (\mu, T(\mu)) \text{ is ccm.} \\ &\overset{\text{Claim}}{\Longrightarrow} \text{the coupling } (\mu, T(\mu)) \text{ is an optimizer.} \end{aligned}$$

The first implication follows from Lemma 2. The second implication is again Proposition 6. Only the last implication is left to show.

If we can show that the cost function is bounded, the last implication follows from the standard optimal transport result Proposition 5, that says that for a continuous and bounded cost function the optimality of the coupling is equivalent to the support being c-cm.

By assumption $h = T(p)$ is bounded from below. The measure Q, by which we denoted the distribution of h when p is distributed according to P, is therefore supported on $[-m, \infty)^n$ for some finite m. The points p are in the (open) simplex and have therefore no zero components. So the cost function is continuous and bounded

$$c(p, T(p)) \geq \log\left(\sum_{i=1}^{n} e^{-m} p_i\right) \geq -C.$$

This leaves us with the desired result. □

We have now firmly linked the pseudo arbitrage property with a particular transport problem. It is useful to consider the financial interpretation of the above portfolio π. In the theorem we defined π as

$$\pi_i = \frac{p_i e^{h_i}}{\sum_j p_j e^{h_j}}.$$

The denominator is just a normalizing factor. The nominator is more interesting. The underlying market weights are denoted by p. We see that the i-th portfolio component gets weighted according to h. If h_i is just 0 the i-th weight is identical to the marketweight. This is also the case for any constant choice of h. So h reflects the relative weighting or relative deviation of the portfolio from the market weights. For a large component h_i the portfolio weight gets overweighted relative to the market weight and similarly underweighted for small h_i.

3.2 Optimal transport in exponential coordinates

The previous theorem was already an important step in the right direction, however the transport problem was formulated on $X = \Delta^{(n)}$, this makes it difficult to find solutions. We will apply a different coordinate system and formulate the transport problem in terms of these new coordinates. That way a related transport problem can be formulated on $\mathbb{R}^{n-1}$.

The (global) coordinate system we will use is defined by

$$\iota : \Delta^{(n)} \to \mathbb{R}^{n-1}, \quad p \mapsto \left(\log \frac{\mu_1}{\mu_n}, \dots, \frac{\mu_{n-1}}{\mu_n} \right). \tag{3.8}$$

Its inversion is given by

$$\iota^{-1} : \mathbb{R}^{n-1} \to \Delta^{(n)}, \quad \theta \mapsto \left(e^{\theta_1 - \psi(\theta)}, \dots, e^{\theta_{n-1} - \psi(\theta)}, e^{-\psi(\theta)} \right), \tag{3.9}$$

where

$$\psi(\theta) := \log \left(1 + \sum_{i=1}^{n-1} e^{\theta_i} \right). \tag{3.10}$$

It is clear that this is indeed a coordinate system/chart because the maps are inverse to each other, smooth and defined on the whole simplex.

As we did before, we need to connect the MCM property and c-cyclical monotonicity. We will use a different cost function

$$c' : \mathbb{R}^{n-1} \times \mathbb{R}^{n-1} \to \mathbb{R}, \quad (\theta, \phi) \mapsto \psi(\theta - \phi), \tag{3.11}$$

where the ψ comes from the above definition of exponential coordinates.

We now relate portfolio functions to functions on $\mathbb{R}^{n-1}$. Let π be a portfolio function on the simplex, then we set

$$\iota(\pi(\mu)) = \iota(\mu) - \phi(\iota(\mu)).$$

We see that the coordinates of a portfolio $\iota(\pi(\mu))$ are a shift away from the market weight coordinates $\iota(\mu)$. This shift is represented by ϕ, which we can obtain from

$$\phi_i(\iota(\mu)) = \iota(\mu)_i - \iota(\pi(\mu))_i = \log\left(\frac{\mu_i}{\mu_n}\right) - \log\left(\frac{\pi(\mu)_i}{\pi(\mu)_n}\right) \tag{3.12}$$

for $i \in \{1, 2, \cdots, n-1\}$.

This is not simply a projection onto the functions on $\mathbb{R}^{n-1}$ but rather a one to one mapping. This becomes clear after taking a function $\phi(\theta)$ on $\mathbb{R}^{n-1}$ and defining a portfolio on the simplex by

$$\pi(\iota^{-1}(\theta)) = \iota^{-1}(\theta - \phi(\theta)). \tag{3.13}$$

These operations are inverse to each other which can be checked by a simple calculation. We can now associate a function on $\mathbb{R}^{n-1}$ to each portfolio.

This association proves to be a crucial one in relating the MCM property to c-cm in our new transport setup.

Proposition 7. *Let c' be the cost function (3.11). Let π be a portfolio function and let ϕ be the corresponding function on $\mathbb{R}^{n-1}$ as defined in (3.12) above. Then π satisfies the MCM property if and only if the graph of ϕ is c'-cm.*

Proof. We express the relative value function $V(t)$ in terms of the exponential coordinates. For an arbitrary sequence of market weights $\{\mu(t)\}_{t=0}^{\infty}$ we have

$$\begin{aligned}\frac{V(t+1)}{V(t)} &= \sum_{i=1}^{n} \pi_i(\mu(t))\frac{\mu_i(t+1)}{\mu_i(t)} \\ &= \pi_n(t)\frac{\mu_n(t+1)}{\mu_n(t)} + \sum_{i=1}^{n-1} \pi_i(\mu(t))\frac{\mu_i(t+1)}{\mu_i(t)}.\end{aligned} \tag{3.14}$$

We split the sum into these two parts because the coordinate function ι treats the first $n-1$ components different from the last one. We write $\theta(t) = \iota(\mu(t))$ as the coordinate for μ and express π in coordinates as well. For all $i \in \{1, \ldots, n-1\}$ we have

$$\pi_i(\mu) = \pi_i(\iota^{-1}(\theta)) = \iota_i^{-1}(\theta - \phi(\theta)) = e^{\theta_i - \phi_i(\theta)} e^{-\psi(\theta - \phi(\theta))},$$

where we used the definition for ϕ and ι. Furthermore, for $i = 1, \ldots, n-1$

$$\frac{\mu_i(t+1)}{\mu_i(t)} = \frac{e^{\theta_i(t+1))}}{e^{\theta_i(t)}}\frac{e^{-\psi(\theta(t+1))}}{e^{-\psi(\theta(t))}}.$$

And for $i = n$

$$\pi_n(t)\frac{\mu_n(t+1)}{\mu_n(t)} = \frac{e^{-\psi(\theta(t+1))}}{e^{-\psi(\theta(t))}}e^{-\psi(\theta(t)-\phi(\theta(t)))}.$$

Putting everything back together into (3.14) we obtain

$$\begin{aligned}
\frac{V(t+1)}{V(t)} &= \frac{e^{-\psi(\theta(t+1))}}{e^{-\psi(\theta(t))}}e^{-\psi[\theta(t)-\phi(\theta(t))]} + \sum_{i=1}^{n-1} e^{-\phi_i(\theta(t))}e^{-\psi[\theta(t)-\phi(\theta(t))]}e^{\theta_i(t+1)}\frac{e^{-\psi(\theta(t+1))}}{e^{-\psi(\theta(t))}} \\
&= \frac{e^{-\psi(\theta(t+1))}}{e^{-\psi(\theta(t))}}e^{-\psi[\theta(t)-\phi(\theta(t))]}\left(1+\sum_{i=1}^{n-1} e^{\theta_i(t+1))-\phi_i(\theta(t))}\right) \\
&= e^{\psi(\theta(t))-\psi(\theta(t+1))-\psi[\theta(t)-\phi(\theta(t))]}e^{\psi[\theta(t+1)-\phi(\theta(t)]}
\end{aligned}$$

Taking the logarithm on both sides and summing over $s = 0, \dots, t-1$ we get

$$\begin{aligned}
&\sum_{s=0}^{t-1} \log V(s+1) - \log V(s) \\
&= \sum_{r=0}^{t-1} \psi(\theta(r)) - \psi(\theta(r+1)) + \sum_{s=0}^{t-1} -\psi[\theta(s)-\phi(\theta(s))] + \psi[\theta(s+1)-\phi(\theta(s))]
\end{aligned}$$

which, after simplifying the telescopic terms, becomes

$$\log V(t) = \psi(\theta(0)) - \psi(\theta(t)) + \sum_{s=0}^{t-1} \psi[\theta(s+1)-\phi(\theta(s))] - \psi[\theta(s)-\phi(\theta(s))]. \quad (3.15)$$

To conclude, fist assume that π satisfies the MCM property. Take an arbitrary cycle $\{\mu(t)\}_{t=0}^{m+1}$ (i.e. $\mu(m+1) = \mu(0)$, which also implies that $\theta(m+1) = \theta(0)$). Now look at time $t = m+1$ in (3.15). By Assumption $\log V(m+1) \geq 0$, so we can rewrite the equation as

$$\sum_{t=0}^{m} \psi[\theta(s+1)-\phi(\theta(s)] \geq \sum_{t=0}^{m} \psi[\theta(s)-\phi(\theta(s)] \quad (3.16)$$

which is the c'-cm condition, namely

$$\sum_{t=0}^{m} c'(\theta(s+1), \phi(\theta(s)) \geq \sum_{t=0}^{m} c'(\theta(s), \phi(\theta(s)). \quad (3.17)$$

The converse is similar to the previous proof. Let the graph of ϕ be c'-cm. We take an arbitrary sequence $\{\theta(t)\}_{t=0}^{m}$ with the convention that $\theta(t+1) = \theta(0)$ (which is just a cycle as we know it). From the c'-cm assumption, we now get (3.17) which implies that in (3.15) $\log V(m+1) \geq 0$. □

Remark 2: This proposition allows us to formulate a transport problem on $\mathbb{R}^{n-1}$ and construct an MCM portfolio out of the transport solution. Assume that ϕ is a transport map that solves the transport problem with cost c' above. Then $(\theta, \phi(\theta))$ (i.e. the graph of ϕ) is c'-cm and by the above proposition the portfolio $\pi(\mu) = \iota^{-1}(\iota(\mu) - \phi(\iota(\mu))$ is MCM.

By formulating the problem on $\mathbb{R}^{n-1}$, we can use known optimal transport results to construct MCM portfolios without investigating a potentially novel transport problem. This will be done in the application section.

3.3 Relative entropy as cost function

We return to the simplex to investigate another transport problem. We define

$$\tilde{c} : \Delta^{(n)} \times \overline{\Delta^{(n)}} \to (-\infty, 0], \quad (p, q) \mapsto -H(q|p) = -\sum_{i=1}^{n} q_i \log \frac{q_i}{p_i} \tag{3.18}$$

where $H(q|p)$ is called the *relative entropy*. Unlike before, we now interpret p as the market weight and q as the portfolio weight itself, not as deviation from the market weight.

Proposition 8. *Let A be a subset of $\Delta^{(n)} \times \overline{\Delta^{(n)}}$ that is $\tilde{c}$-cm. Then for any cycle $\{p(t), q(t)\}_{t=0}^{m+1}$ in A, the portfolio $q(t)$ is MCM with regard to the market weights $p(t)$. The converse does not hold.*

Proof. Let $\{p(t), q(t)\}_{t=0}^{m+1}$ be a cycle in A, by the $\tilde{c}$-cm assumption we have

$$\sum_{t=0}^{m} \sum_{i=1}^{n} q_i(t) \log \frac{q_i(t)}{p_i(t)} \geq \sum_{t=0}^{m} \sum_{i=1}^{n} q_i(t) \log \frac{q_i(t)}{p_i(t+1)}.$$

Simplifying this equation yields

$$\sum_{t=0}^{m} \sum_{i=1}^{n} q_i(t) \log \frac{p_i(t+1)}{p_i(t)} \geq 0.$$

We now apply Jensen's inequality (with the convex function $\log(x)$) that says, for $\sum_i q_i = 1$ we can estimate

$$\sum_{i=1}^{n} q_i \log(x_i) \leq \log\left(\sum_{i=1}^{n} q_i x_i\right). \tag{3.19}$$

The condition on q_i is satisfied because $q \in \overline{\Delta^{(n)}}$. Choosing $x_i = \frac{p_i(t+1)}{p_i(t)}$ we get

$$0 \leq \sum_{t=0}^{m} \sum_{i=1}^{n} q_i(t) \log \frac{p_i(t+1)}{p_i(t)} \leq \sum_{t=0}^{m} \log \left(\sum_{i=1}^{n} q_i(t) \frac{p_i(t+1)}{p_i(t)} \right).$$

We then apply the exponential function to both sides, thus

$$1 \leq \prod_{t=0}^{m} \left(\sum_{i=1}^{n} q_i(t) \frac{p_i(t+1)}{p_i(t)} \right) = \prod_{t=0}^{m} \frac{V(t+1)}{V(t)} = V(t+1)$$

which is precisely the MCM property for the portfolio sequence $q(t)$.

Lastly, we formulate a counterexample to the converse statement. Define π as the identity portfolio, $\pi(\mu) = \mu$ for all μ. By definition for the relative value we get

$$V(t+1) = V(t) \sum_{i=1}^{n} \mu_i(t) \frac{\mu_i(t+1)}{\mu_i(t)} = V(t), \quad \forall t.$$

In particular, the portfolio satisfies the MCM property because at all times the relative value is greater or equal than 1.

Note also that $\tilde{c}(p, \pi(p)) = 0$. That means that this particular portfolio maximises the transport cost due to the fact that the maximal value of H is zero. However, from the definition it is clear that $\tilde{c}(p, q) = 0$ if and only if its arguments satisfy $p = q$, so we can choose some sequence $\{p_i\}_{i=1}^{k}$ (that is not constant) and any non-trivial permutation σ on the index set to get

$$0 = \sum_{i=1}^{k} \tilde{c}(p_i, \pi(p_i)) > \sum_{i=1}^{k} \tilde{c}(p_i, \pi(p_{\sigma(i)}))$$

which disproves the $\tilde{c}$-cm property for the identity portfolio. In particular, MCM is not sufficient for $\tilde{c}$-cm in this setup. □

In the previous section we were always able to show two things. From optimal couplings we got functionally generated portfolios and from functionally generated portfolios we got optimal couplings. The last proposition only gave us one direction.

If we have an optimal coupling for the transport problem, its support is $\tilde{c}$-cm and by Proposition 8 q is functionally generated.

We will work out some other properties of the optimal coupling in the next proposition. For this we need the following preparatory lemma.

Lemma 7. *Let φ be a proper convex function on $(0,\infty)^n$. With $\log(\cdot)$ we signify the componentwise logarithm. Note that then $-\log \Delta^{(n)} \subset (0,\infty)^n$. Assume that for every point $p \in \Delta^{(n)}$ there exists a $\pi \in \overline{\Delta^{(n)}}$ such that $\pi \in \partial\varphi(-\log p)$, a so called **sub**gradient*[1]*. Written as equation*

$$\varphi(-\log p) + \langle \pi, \log p - \log q\rangle \leq \varphi(-\log q), \quad \forall q \in \Delta^{(n)}. \tag{3.20}$$

Then, $\Phi(p) = \exp(-\varphi(-\log p))$ is concave on $\Delta^{(n)}$ and π interpreted as portfolio function is generated by Φ.

The sequence π is strictly speaking not a function of p. As described above, for each p we choose **one** π that is then in the subdifferential. We will treat this association as a mapping $p \mapsto \pi$ and accordingly write $\pi(p)$ to denote the point $\pi \in \overline{\Delta^{(n)}}$ that was associated to p.

Proof. Take some $p, q \in \Delta^{(n)}$. We use the assumption (3.20)

$$\varphi(-\log p) - \varphi(-\log q) \leq \langle \pi, \log q - \log p\rangle = \sum_{i=1}^{n} \pi_i \log\left(\frac{q_i}{p_i}\right).$$

We can use Jensen's inequality (see (3.19)) which allows us to estimate the sum upwards by exchanging sum and $\log(\cdot)$. The necessary condition, which holds true, is $\sum_i \pi_i = 1$. We estimate upwards to get

$$\begin{aligned}
\varphi(-\log p) - \varphi(-\log q) &\leq \log\left(\sum_{i=1}^{n} \pi_i \frac{q_i}{p_i}\right) \\
&= \log\left(\overbrace{1 - \sum_{i=1}^{n} \pi_i}^{=0} + \sum_{i=1}^{n} \pi_i \frac{q_i}{p_i}\right) \\
&= \log\left(1 + \left\langle \frac{\pi}{p}, q - p\right\rangle\right).
\end{aligned}$$

Then take $\exp(\cdot)$ on both sides to get

$$1 + \left\langle \frac{\pi}{p}, q - p\right\rangle \geq \frac{e^{\varphi(-\log p)}}{e^{\varphi(-\log q)}} = \frac{\Phi(q)}{\Phi(p)}. \tag{3.21}$$

Note that we already defined Φ in the lemma.

[1] Note that φ is convex. Before we only defined the **super**gradient for concave function but the definition extends to convex function by considering $-\varphi$.

To show that Φ is concave we want to employ [BSS06, Thm. 3.2.6]. This theorem states that a function is concave if we can show that for each $p \in \Delta^{(n)}$ there exists a supergradient at p. We show this by constructing concrete supergradients.

We claim that the following vector is in the supergradient of Φ at p.

$$v_p := \Phi(p) \left(\frac{\pi_i(p)}{p_i} - \frac{1}{n} \sum_{j=1}^{n} \frac{\pi_j(p)}{p_j} \right)_{1 \le i \le n}$$

To show this two things are necessary. We first need to check that v_p is actually a tangent vector and then we need to show that v_p satisfies the supergradient equation for Φ at p.

Write $v_p = \Phi(p) v'_p$. In Proposition 2 we have seen that v'_p is already a tangent vector. The tangent space is a vector space and therefore the multiple of a tangent vector is still in the space. So v_p is a tangent vector.

Now note that

$$\begin{aligned} \langle v_p, q - p \rangle &= \Phi(p) \left\langle \frac{\pi(p)}{p} - \frac{1}{n} \sum_{j=1}^{n} \frac{\pi_j(p)}{p_j}, q - p \right\rangle \\ &= \Phi(p) \left\langle \frac{\pi(p)}{p}, q - p \right\rangle - \Phi(p) \frac{1}{n} \sum_{j=1}^{n} \frac{\pi_j(p)}{p_j} \langle \vec{1}, q - p \rangle \\ &= \Phi(p) \left\langle \frac{\pi(p)}{p}, q - p \right\rangle . \end{aligned}$$

Applying equation (3.21) to this gives us precisely the superdifferential equation. Namely, that

$$\Phi(p) + \langle v_p, q - p \rangle \ge \Phi(q)$$

holds for all $q \in \Delta^{(n)}$. As discussed above, the concavity now follows from [BSS06, Thm. 3.2.6].

After having established concavity of Φ and that v_p is in the superdifferential for each p, we have $v'_p \in \partial \log \Phi(p)$. By applying Proposition 2 again we see that Φ generates π. $\square$

Proposition 9. *Let $U \subset \Delta^{(n)}$ and $V \subset \overline{\Delta^{(n)}}$. Take two probability measures P and Q that are supported on U and V, respectively. Assume that*

$$\gamma := \sum_{\pi \in V, \mu \in U} H(\pi|\mu) < \infty. \tag{3.22}$$

Then the transport problem with cost function $\tilde{c}$ has finite value and the following holds.

(i) We can find a concave function $\Phi : \Delta^{(n)} \to (0, \infty)$ such that any optimal coupling to the above problem is concentrated on the set

$$\left\{ (p, \pi) \in U \times V \middle| \left(\frac{\pi_i}{p_i} - \frac{1}{n} \sum_{j=1}^{n} \frac{\pi_j}{p_j} \right)_{1 \le i \le n} \in \partial \log \Phi(p) \right\}. \tag{3.23}$$

(ii) The function $\log \Phi$ is bounded on U because for any $p, q \in U$ we have $\Phi(q)/\Phi(p) \le \exp(\gamma)$.

Proof. We first show that the transport problem has finite value. The cost function $\tilde{c}$ is non-positive so from (3.22) we know that

$$0 \ge \mathbb{E}_R[\tilde{c}(x, \pi)] \ge -\gamma.$$

So for any optimizer the transport problem is finite.

To show (i) we want to use the special structure of the cost function to reformulate the problem in quadratic form and apply standard optimal transport results. Let R' be a solution to our problem

$$\inf_{R \in \Pi(P,Q)} \int \tilde{c}(x, \pi)\ dR(x, \pi).$$

We define a change of coordinates by $f(x) = -\log(x)$. This defines a new measure $\tilde{P}$ via the pushforward $\tilde{P} := f_*(P)$. This transforms the transport problem in the following way. Let $T(x, \pi) = (-\log x, \pi)$, then

$$\begin{aligned}
\int \tilde{c}(x, \pi)\ dR(x, \pi) &= \int \tilde{c}(\exp(\log(x)), \pi)\ dR(x, \pi) \\
&= \int \big(\tilde{c}(\exp(\cdot), \cdot) \circ T\big)(-x, \pi)\ dR(x, \pi) \\
&= \int \tilde{c}(\exp(-x), \pi)\ d\tilde{R}(x, \pi).
\end{aligned}$$

Where $\tilde{R} = T_*(R)$ is coupling between $\tilde{P}$ and Q which comes from the change of variable. Note that while Q is still supported on the closed simplex, $\tilde{P}$ is now supported on $(0,\infty)^n$. We further rewrite the cost function as

$$\begin{aligned}\tilde{c}(\exp(-x),\pi) &= -\left\langle \pi, \log\frac{\pi}{\exp(-x)}\right\rangle \\ &= -\langle \pi, \log\pi\rangle - \langle \pi, x\rangle \\ &= H(\pi) - \langle \pi, x\rangle.\end{aligned}$$

Where $H(\pi) = -\langle \pi, \log\pi\rangle$ is the so called *Shannon entropy.* The transport problem can written as

$$\begin{aligned}\inf_{R\in\Pi(P,Q)} \int \tilde{c}(x,\pi)\ dR(x,\pi) &= \inf_{\tilde{R}\in\Pi(\tilde{P},Q)} \int H(\pi) - \langle \pi, x\rangle\ d\tilde{R}(x,\pi) \\ &= \sup_{\tilde{R}\in\Pi(\tilde{P},Q)} \mathbb{E}_{\tilde{R}}[\langle \pi, x\rangle - H(\pi)]\end{aligned}$$

Furthermore, note that $\mathbb{E}_{\tilde{R}}[H(\pi)] = \mathbb{E}_Q[H(\pi)]$, because it only depends on the second marginal of R which is Q. Therefore we can write

$$\sup_{\tilde{R}\in\Pi(\tilde{P},Q)} \mathbb{E}_{\tilde{R}}[\langle \pi, x\rangle - H(\pi)] = \sup_{\tilde{R}\in\Pi(\tilde{P},Q)} \mathbb{E}_{\tilde{R}}[\langle \pi, x\rangle] - \mathbb{E}_Q[H(\pi)].$$

The second term is independent of R or $\tilde{R}$. We can leave it out without changing the optimizer (the transport value will change by this constant). We can formulate an equivalent transport problem as

$$\sup_{\tilde{R}\in\Pi(\tilde{P},Q)} \mathbb{E}_{\tilde{R}}[\langle x, \pi\rangle]. \tag{3.24}$$

With "equivalent", we mean that a solution R' to the original problem can be transformed into a solution $\tilde{R}'$ to (3.24) by following the above calculations. The same holds for a solution $\tilde{R}'$ to (3.24) that can, by inverting the above steps, be transformed to a solution of the original problem.

As mentioned above, our aim is to reformulate the problem in terms of a quadratic cost function. Note that $||\pi - x||^2 = ||\pi||^2 - 2\langle \pi, x\rangle + ||x||^2$. Using this we can again formulate an equivalent problem to (3.24) by

$$\inf_{\tilde{R}\in\Pi(\tilde{P},Q)} \mathbb{E}_{\tilde{R}}[||\pi - x||^2] = \inf_{\tilde{R}\in\Pi(\tilde{P},Q)} \mathbb{E}_{\tilde{R}}[-2\langle \pi, x\rangle] + \mathbb{E}_{\tilde{P}}[||x||^2] + \mathbb{E}_Q[||\pi||^2]$$

As before, the expectations that only depend on the marginals can be disregarded, the same goes for the constant factor 2, they only change the transport value and not the optimizer. We have seen that the problem

$$\inf_{\tilde{R}\in\Pi(\tilde{P},Q)} \mathbb{E}_{\tilde{R}}[||\pi - x||^2] \tag{3.25}$$

is equivalent to (3.24) and therefore to the original transport problem as well. By [Vil08, Theorem 4.1] there exists a solution $\tilde{R}'$ to (3.25). To such a quadratic cost function the Knott-Smith optimality criterion ([Vil03, Theorem 2.12 (i)]) is applicable. It states that a coupling $\tilde{R}'$ is optimal if and only if there exists a lower semi continuous convex function φ such that $\mathrm{Supp}(\tilde{R}') \subset \mathrm{Graph}(\partial(\varphi))$, or in other words

$$\text{for } \tilde{R}'\text{-almost all } (x, \pi), \text{ we have } \pi \in \partial\varphi(x).$$

We want to apply the previous lemma, which requires such a φ for **all** points and not just a set of measure one. We can alleviate the problem with a small trick.

First define $p := \exp(-x)$ that way $\pi \in \partial\varphi(x)$ becomes $\pi \in \partial\varphi(-\log p)$. This brings us back to the simplex because $\tilde{P}$ is supported on $(0, \infty)^n$ due to the previous change of variables.

Define a function f in the following way. Take some pair $(p, \pi) \in U \times V$
- If the pair (p, π) satisfies $\pi \in \partial\varphi(-\log p)$ we map $f(p) = \pi$.
- If the pair (p, π) does not satisfy $\pi \in \partial\varphi(-\log p)$ we simply map $f(p) = v$ for some $v \in \partial\varphi(-\log p)$.

The function by itself would not be well-defined. If there are multiple choices of $(p, \pi), (p, \pi'), \ldots$ choose a pair that satisfies the first case, if there is no such pair choose an arbitrary pair. By our previous deliberations the points in the second category are only a set of measure zero.

We associated the market weights p with the portfolio weights π that we got from the transport solution. For a zero-set of points this was not possible but because we are only interested in a statement about the support of the solution this will pose no problem in the end.

We are now in a situation where, for every $p \in U$ there exists a vector $f(p) \in \partial\varphi(-\log p)$.

Apply Lemma 7 for which we just showed the necessary conditions. We get that $\Phi(p) = \exp(-\varphi(-\log p))$ is concave on the simplex and generates π. Using Proposition 2 we get that

$$\left(\frac{\pi_i}{p_i} - \frac{1}{n} \sum_{j=1}^{n} \frac{\pi_j}{p_j} \right)_{1 \le i \le n} \in \partial\Phi(p) \tag{3.26}$$

for $\pi = f(p)$.

To summarize, in this construction Φ generates a portfolio defined by $p \mapsto f(p) = $ "$\pi(p)$". Except on a zero set this portfolio coincides with the solution to the transport problem. In particular, the support of the solution must be concentrated on the graph of f. By Proposition 2 the graph of f coincides exactly with the points that satisfy (3.26). This proofs the claim.

(ii) Take a point (p, π) from the support of the solution to the optimal transport problem R, as seen above $\pi \in \partial(-\log \varphi(p))$. By definition, for all q we have

$$\begin{aligned}\varphi(-\log q) - \varphi(-\log p) &\geq \langle \pi, -\log q + \log p \rangle \\ &= H(\pi, p) - H(\pi, q) \\ &\geq -\gamma\end{aligned}$$

and thus get the desired result

$$\frac{\Phi(p)}{\Phi(q)} = \exp(\varphi(-\log q) - \varphi(-\log p)) \leq \exp(\gamma).$$

□

Chapter 4

Application

In this chapter we will apply the theory we developed to stock data of Visa and Mastercard. We will limit ourselves to $n = 2$, so our market has only two assets. This will allow us to solve the transport problem by so called *monotone rearrangement*. We will train our model with data[1] from 01.02.2010 to 30.01.2015 and then backtest the model with data from 01.02.2015 to 19.05.2022.
This section is closely related to [PW15, Section 4]. Figures 4.1, 4.2 and 4.3 are also adapted from [PW15, Section 4].

We will solve the transport problem in exponential coordinates, the details can be found in Section 3.2. As dimension we choose $n = 2$, by (3.9) we have

$$\mu = \left(\frac{e^\theta}{1 + e^\theta}, \frac{1}{1 + e^\theta} \right)$$

where $\theta \in \mathbb{R}$ is the exponential coordinate of μ. A portfolio on $\Delta^{(2)}$ written in terms of exponential coordinates is given by

$$\pi(\mu) = \left(\frac{e^{\theta - \phi}}{1 + e^{\theta - \phi}}, \frac{1}{e^{\theta - \phi}} \right)$$

where ϕ is as in (3.12). The cost function we will use is given by

$$c(\theta, \phi) = \psi(\theta - \phi) = \log(1 + e^{\theta - \phi}),$$

as in (3.11). As discussed in Section 3, the portfolio weight is determined by the shift ϕ. Our aim is accordingly to find ϕ as a function of the exponential coordinate θ. For this we will solve the transport problem. We set up the problem

[1]The data is taken from `https://finance.yahoo.com/quote/V?p=V` and `https://finance.yahoo.com/quote/MA?p=MA`, respectively.

as in Proposition 7. For two measures P, Q that are absolutely continuous with regard to Lebesgue-measure we have

$$\inf_{R \in \Pi(P,Q)} \mathbb{E}_R[c(\theta, \phi)]. \tag{4.1}$$

We can find an explicit optimizer using monotone rearrangement. Let G and H be the cumulative distribution functions of P and Q, respectively. The *monotone transport* map is given by

$$F : \mathbb{R} \to \mathbb{R} : \quad x \mapsto \inf\{y : H(y) \geq G(x)\}.$$

One can also see this as the quantile function of H applied to the point $G(x)$. It can be shown (see [Vil03, Remark 2.19 (iv)]) that this monotone coupling $(\mathrm{id}, F)(P)$ solves the transport problem (4.1). As discussed in Remark 2 in Section 3.2 the transport map $F(\theta)$ is then the portfolio shift function $\phi(\theta)$.

In the case of $P = N(p_1, \sigma_1^2)$ and $Q = N(p_2, \sigma_2^2)$ we can show that

$$F(\theta) = m_2 + \frac{\sigma_2}{\sigma_1}(\theta - m_1).$$

Clearly, F is non-decreasing, and the pushforward map from P to Q because for $X \sim P$

$$\mathbb{P}(F(X) \leq z) = \mathbb{P}\Big(\overbrace{m_2 + \frac{\sigma_2}{\sigma_1}(X - m_1)}^{\sim Q} \leq z\Big).$$

The quantile function is uniquely determined for measures that are absolutely continuous with regard to the Lebesgue-measure. That means that our transport map the unique non-decreasing function that pushes P onto Q. Most importantly, F is indeed the rearrangement map that solves the transport problem (4.1). We can set $F = \phi$.

After finding ϕ we can always calculate the portfolio by

$$\pi(\iota^{-1}(\theta)) = \iota^{-1}(\theta - \phi(\theta)).$$

The exponential coordinate is $\theta = \log \frac{\mu_1}{\mu_2}$, then

$$\begin{aligned}
\iota(\pi(\mu)) &= \log \frac{\pi_1(\mu)}{\pi_2(\mu)} \\
&= \theta - F(\theta) = \left(1 - \frac{\sigma_2}{\sigma_1}\right) \log \frac{\mu_1}{\mu_2} + \left(\frac{\sigma_2}{\sigma_1} m_1 - m_2\right) \\
&= \alpha \log \frac{\mu_1}{\mu_2} + \log c
\end{aligned}$$

with $\alpha = 1 - \frac{\sigma_2}{\sigma_1}$ and $c = \exp\left(\frac{\sigma_2}{\sigma_1} m_1 - m_2\right)$. We pull everything back to the simplex to get

$$\iota^{-1}(\iota(\pi(\mu))) = \left(\frac{\exp\left(\alpha \log \frac{\mu_1}{\mu_2} + \log c\right)}{1 + c\left(\frac{\mu_1}{\mu_2}\right)^2}, \frac{1}{1 + \left(\frac{\mu_1}{\mu_2}\right)^\alpha} \right) = \left(\frac{c\mu_1^\alpha}{c\mu_1^\alpha + \mu_2^\alpha}, \frac{\mu_2^\alpha}{c\mu_1^\alpha + \mu_2^\alpha} \right).$$

Which means that when P and Q are normal, the portfolio is given by

$$\pi(\mu) = \left(\frac{c\mu_1^\alpha}{c\mu_1^\alpha + \mu_2^\alpha}, \frac{\mu_2^\alpha}{c\mu_1^\alpha + \mu_2^\alpha} \right).$$

We can derive some intuition about the portfolio by investigating this representation. The mean of Q influences how much the portfolio over- or underweights the first component. Assume $m_2 \gg m_1$ then $c \ll 1$ and in turn π_1 gets less weight. This effect gets modified by the ratio of variances.

If Q has a small variance relative to P we have $\frac{\sigma_2}{\sigma_1} \ll 1$ and the above influence of m_1 becomes stronger. A small Q variance also has the effect that the portfolio stays closer to the market portfolio. We get $\sigma \approx 1$ and $c \approx 1$ for m_1 close to zero. The resulting portfolio is basically the market portfolio.

Model training[2]. We first need to fix some probability measures P and Q. The distribution of P reflects how the investor expects the market (represented by θ) to behave in the future. Higher dispersion reflects higher uncertainty. The measure Q on the other hand, reflects the investors expectation on how the assets will behave relatively to each other.

For our stock data, we took the average price for each day and normed each stock price to the first day. That means the resulting time series represents the gain (or loss) of a single dollar investment at 01.02.2010. In other words we normed the market capitalization of each asset to the this date. The initial market weight is therefore be $\mu(0) = (0.5, 0.5)$.

Choice of P. At first we need to choose the distribution P that represents the investors future believe of how the market weight θ will behave. For simplicity we fit the training data of θ to a Normal distribution, although the histogram suggest a much harder to model bimodal distribution. We also assume that the distribution will not change in the future so let $P = N(-0.098, 0.090)$ which is the results of a straight forward Normal distribution fitting of θ. As mentioned, we could also choose a higher variance if we were more uncertain about the future development of the market.

[2]The code for the calculations can be found at `https://github.com/JasperSchmidt/MasterThesis/blob/main/ThesisApplication.ipynb`

Choice of Q. The choice of Q is not so simple. As seen above it represents the distribution of ϕ which describes how the portfolio shifts away from the market weight. A very diffuse distribution Q might lead to a portfolio that is vastly different from the market. We can also interpret this as the risk affinity of the investor. The more concentrated the distribution the less can the portfolio outperform or underperform the market meaning that it is less risky. Our discussion above suggests that if ϕ takes values larger than 0 it will underweigh the first stock (given that P correctly describes the future market weights). To illustrate the influence

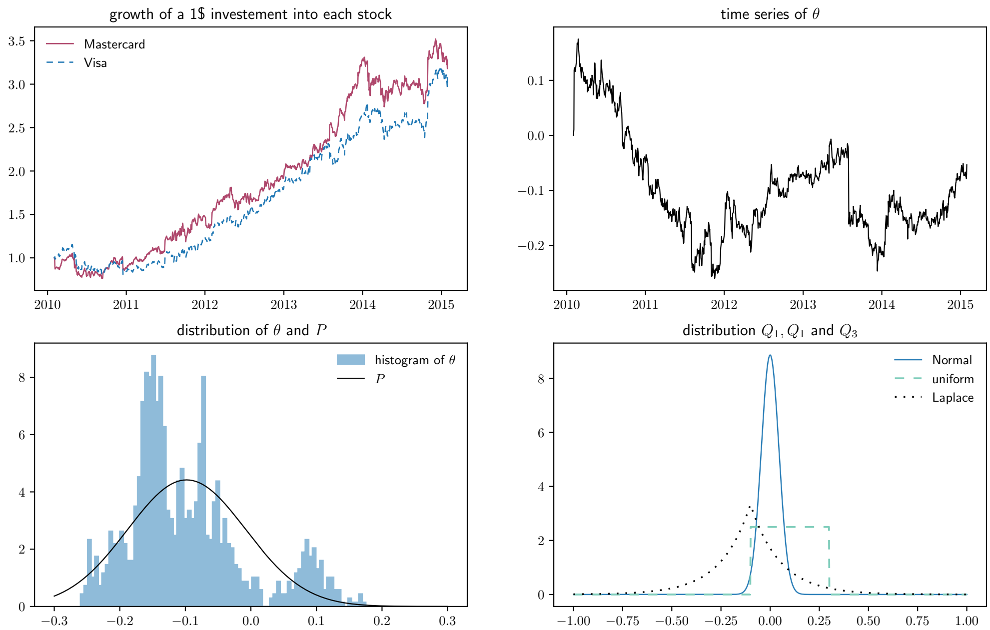

Figure 4.1: The normed stock price evolution (top left), the exponential coordinate of the market (top right), fitting of the market distribution P (bottom left) and the three choices of Q (bottom right).

of the choice of distribution we will create three models with a Normal, uniform and Laplace distribution (see Figure 4.1). We call the resulting portfolio $\pi^{(1)}$, $\pi^{(2)}$ and $\pi^{(3)}$.

The normal distribution $Q_1 = N(0, 0.045)$ has expectation zero and a small variance (less than half of P). We expect the resulting portfolio to stay close to the market weight and not severely overweight either asset. It also should not outperform the buy and hold strategy by much. This is an example for a strategy that is less risky.

The uniform distribution $Q_2 = \text{Uniform}(-0.1, 0.3)$ has bounded support and therefore limits the maximal portfolio deviation. It is still relatively concentrated.

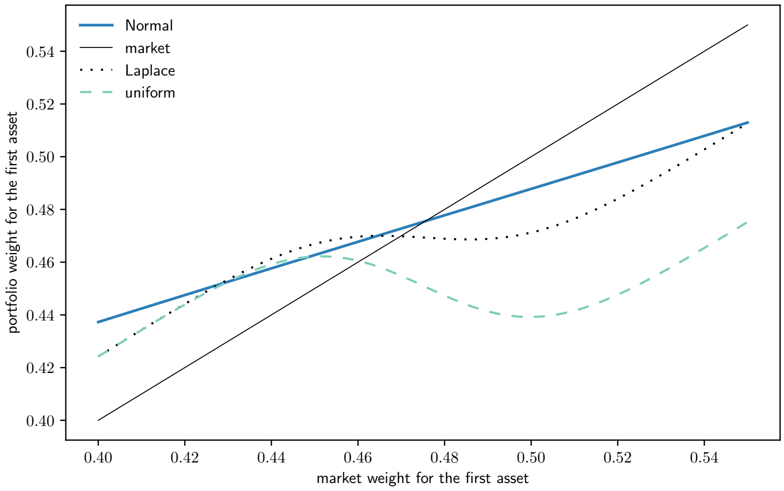

Figure 4.2: First portfolio component against first market component (Visa)

In three quarter of the cases the value will be greater than zero so we expect Q_2 to systematically underweigh the first stock.

The Laplace distribution $Q_3 = \text{Laplace}(\text{location} = -0.1, \ \text{scale} = 0.15)$ is similar to the Normal distribution but with the notable difference that its tails are much fatter, this allows the portfolio to differ more widely from the market weights. We additionally chose an expected value that is less than zero, so we expect it to overweight the first asset slightly.

We can simulate the weights for each portfolio. Portfolio $\pi^{(1)}$, that is based on the Normal distribution, has weights that are relatively close to the market weight. We could get even closer by choosing a smaller variance for Q_1. The Laplace and uniform portfolio also behave similar to what we expected. We can see that the uniform distribution strongly underweights the first asset. The Laplace portfolio, qualitatively, has some similarity to the normal portfolio but occasionally exaggerates the weighting. This is due to the higher dispersion of the distribution that allows the portfolio weights to differ more from the market.

Backtesting. We can now backtest the portfolios for the years 2015-2022 and calculate the relative value of each portfolio (See Figure 4.3). The uniform portfolio outperforms both the Normal and the Laplace portfolio by a comfortable margin. The results with with 0.5% to 1.25% for the whole period seem rather small, it should be noted however that we are looking the relative value of each portfolio meaning that these are "excess" gains on top of the regular buy and hold gain. The most important feature is that we outperformed the buy and hold strategy reliably after 2-3 years. The portfolio with the most gains ($\pi^{(2)}$) was also the one with the biggest losses in the first few years. This choice appears to have been more risky which allowed for the higher reward. On the same note, the less risky portfolio $\pi^{(1)}$ had the least returns and least losses (in the first years) while still outperforming the market.

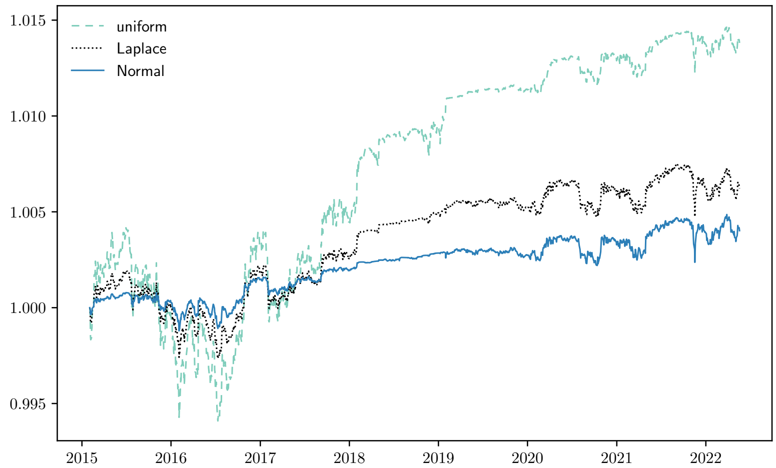

Figure 4.3: Relative value of the portfolios $\pi^{(1)}, \pi^{(2)}$ and $\pi^{(3)}$.

We also want to touch on the issue of trading costs. Usually market participants have to pay a certain fee to buy or sell assets. This seems to be especially problematic for our approach because we rely on constant rebalancing of the portfolio. To illustrate this we compare the relative values of $\pi^{(2)}$ with and without trading costs.

We change the relative value formula (1.3) to

$$V(0) = \alpha,\ V(t+1) = \alpha V(t) \sum_{i=1}^{n} \pi_i(\mu(t)) \frac{\mu_i(t+1)}{\mu_i(t)}.$$

for some $\alpha \in [0,1]$. With this we can describe the relative value if the investor has to pay $(1-\alpha)V(t)$ as trading cost per transaction.

In Figure 4.4 we see that even for of 0,001% trading cost per transaction the portfolio stops beating the market. It should be mentioned however that this heuristic could certainly be optimized.

Also note that when rebalancing the investor does not need to sell her whole portfolio. Assume the portfolio weights for the previous time were $\pi(t-1) = (0.5, 0.5)$ and the current weights are $\pi(t) = (0.55, 0.45)$, therefore only 10% of the portfolio need to actually be rebalanced.

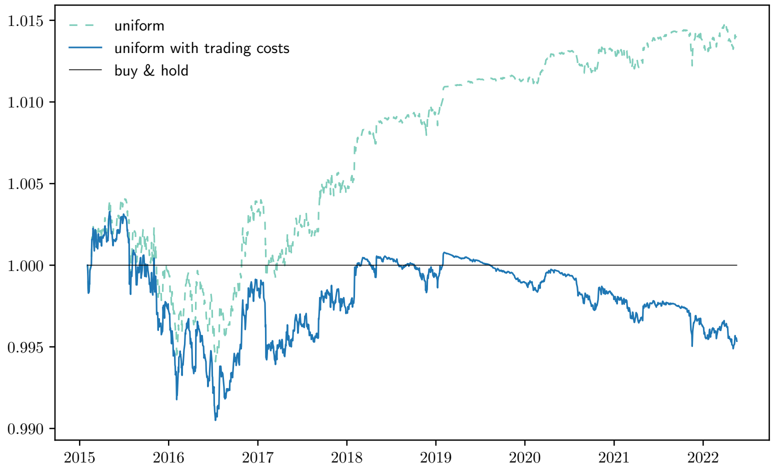

Figure 4.4: The relative value of different portfolios with and without trading costs for $\alpha = 0.99999$.

A more realistic way to look at trading costs is to look at fixed costs for each transaction. Large institutional investors and high frequency traders might be able to operate under low transaction costs for large and frequent rebalancing. A further analysis of the impact of trading costs is, however, beyond the scope of this thesis.

Chapter 5

Conclusion

The most important contribution of [PW15] to the theory of SPT is certainly that all pseudo-arbitrages have to be functionally generated. Functionally generated portfolios were already discovered in the time continuous setup but it was not clear how large the class of these portfolios were. At least for discrete time we can now say that the class is fairly large and encompasses all pseudo arbitrages.

Another important contribution was the introduction of the MCM property that as a theoretical concept was crucial to almost all of the proofs. If one recalls how we related pseudo arbitrages to functionally generated portfolios or to optimal transport it was always via the MCM property.

A third achievement that has theoretical and practical implications is the connection to optimal transport. As our empirical analysis showed, if one can solve the transport problem the construction of functionally generated portfolios and pseudo-arbitrages is relatively straight forward. This novel characterization might allow for the discovery of new results and the authors themselves have already gone into that direction (see below).

If we restrict us to differentiable portfolios we have seen that there are basically three equivalent ways to characterize a functionally generated portfolio, by the generating concave function, its log-gradient and by the transport problem. One can introduce the concept of log-concave functions. These are exactly the functions that after applying the logarithm are concave (given that the function is positive). The portfolio generators we considered can be characterized as such functions. We also saw that there is some connection to objects from information theory. For example in Proposition 4 we used the Fisher information metric to relate the tangent map of a portfolio to the excess growth, furthermore we used the relative entropy as a cost function in Proposition 9 where also the Shannon entropy appeared in the proof. This hints at a deeper connection to information theory and this was indeed explored in [PW17]. They use the L-divergence to define a new information geometry on the simplex and relate this to the the optimal transport problem from Section 3.2.

Regarding the practical application it should be mentioned that the theory as presented here has some drawbacks. First, the assumption of no trading cost. It is not clear if or how a functionally generated portfolio that incurs a fee every time it rebalances can still be a pseudo arbitrage. Our analysis in the Application section suggest that this is not the case in general. A second issue that ties back to this is how frequent rebalancing has to be done. The results presented here assume that rebalancing is done every time the underlying weights change. However, in the empirical analysis we rebalanced only once per day (without trading costs) and still outperformed the market. The questions that arise are if can we save trading costs by reducing the frequency or reduce the frequency in general. There has been some work done to answer the second question (see [PW17]). Intuitively, skipping changes in the market weight might increase the volatility for the portfolio thus leading to higher value growth[1].

Our discrete SPT approach had many advantages that allowed us to formulate the crucial MCM property and find a connection to optimal transport. A downside to this was some loss of descriptive power. In continuous time SPT the prices are dynamically modelled by SDEs which make it possible to formulate a complex market model and analyse the endogenous market behaviour. In discrete time the market is static meaning that weights are not generated by some underlying structure, they are just exogenous inputs.

In the optimal transport section we have seen that after solving the higher dimensional transport problem it would also be straight forward to construct portfolios for more than two stocks. In [Pal16] the author has shown in continuous time that one can construct functionally generated portfolios in high dimensions, that can outperform the market even in short time.

Leaving the SPT framework aside, there is also some research into model independent finance that utilizes optimal transport. The aim is to assess the impact of flawed model assumptions. When pricing or hedging an asset the investor usually models the asset by some probability measure, this entails some inherent uncertainty either by limited information or flawed models. An approach to this in the context of investment hedging can be found in [BBB20]. They introduce the so called *adapted Wasserstein distance* which is a metric for probability measures defined via an optimal transport problem. They showed that the choice of the model (i.e. the probability measure) is robust under this metric, meaning that the hedging error between two models can be bounded by the adapted Wasserstein distance.

[1] See also the talk by S. Pal on the above mentioned paper [PW17], where he connects the results to [PW15], the paper on which this thesis is based on. It can be found here: `http://www.birs.ca/events/2016/5-day-workshops/16w5134/videos/watch/201605231515-Pal.html` (last accessed: 26.06.2022)

This idea, that the "distance" between models can be captures by some metric for probability measures can be further extended to general stochastic optimization problems. In [BDO21] the authors consider the *Wasserstein distance* (a precursor to the adapted Wasserstein distance) to quantify the uncertainty of a model by looking at Wasserstein balls around the parameter measure. They discuss various applications including finance and neural networks.

APPENDIX

A: Geometry

We need some basic facts from differential geometry. This section will recall the details for the most important results that we need. Note however, that because of the simplicity of our manifold (a submanifold of $\mathbb{R}^n$ that lies in a hyperplane) the main thing to do here is to simplify the abstract results to our situation. As we cannot review all basic results in differential geometry the proofs will be sketches that show how to use the abstract results in the simple setting. Helpful references on this topic were [Lee12], [Tu10] and [Lee09].

Lemma A1.
The tangent space of the simplex is

$$T\Delta^{(n)} = \{v \in \mathbb{R}^n \text{ such that } \langle v, \vec{1}\rangle = 0\}.$$

Proof (idea). Intuitively we believe the statement because the normal to the simplex is $\vec{1}$. We now sketch how this can be seen using a general definition of the tangent space. The tangent space can be abstractly defined by *derivations*. A derivation at the point $p \in \Delta^{(n)}$ is a linear map $v : C^\infty(\Delta^{(n)}) \to \mathbb{R}$ that satisfies the product rule

$$v(fg) = f(p)vg + g(p)vf.$$

We call the set of all derivations at some point p in the manifold the tangent space, denoted by $T_p^a\Delta^{(n)}$ (the a denotes that the tangent space is defined **a**bstractly). For our application this definition is less useful. The simplex is embedded in $\mathbb{R}^n$ and there should be a corresponding tangent space in $\mathbb{R}^n$. One can show ([Lee09, Proposition 3.2]) that

$$T_p^a\Delta^{(n)} \cong T_p\Delta^{(n)}.$$

The isomorphism is given by

$$v \mapsto D_v|_p$$

where

$$D_v|_p : C^\infty(\mathbb{R}^n) \to \mathbb{R}^n, \quad f \mapsto D_v f(p) = \frac{d}{dt}\Big|_{t=0} f(p + tv).$$

This means that the directional derivative operators[2] can be used to describe all derivations.

To find the tangent space, we use another fact from the abstract theory. In particular the result below ([Lee09, Proposition 5.3]) allows us to identify the tangent space of a submanifold as a subset of the original tangent space. The vectors in the simplex have only positive components, to account for this we use $M = \mathbb{R}^n_+ = \{x_i > 0\}$ as bigger manifold. We identify the tangent space with the directional derivatives, then

$$\begin{aligned} T\Delta^{(n)}(p) &= \{w \in T^a_p M : w(f) = 0, \quad \forall f \in C^\infty(T^a_p M) \text{ with } f|_{\Delta^{(n)}} = 0\} \\ &= \{v \in \mathbb{R}^n : D_v(f) = 0, \quad \forall f \in C^\infty(\mathbb{R}^n) \text{ with } f|_{\Delta^{(n)}} = 0\}. \end{aligned}$$

Tangent spaces at all points in the simplex are identical, we therefore simply write $T\Delta^{(n)}$. Take $f = \langle \cdot, \vec{1} \rangle - 1$. The directional derivative of f is

$$D_v f(p) = \langle v, \nabla f(p) \rangle = \langle v, \vec{1} \rangle.$$

So $v \in T\Delta^{(n)}$ if $\langle v, \vec{1} \rangle = 0$. We can easily construct $n-1$ linearly independent vectors to span the tangent space. Note that the function is linear, that means that all linear combinations of base vectors already satisfy the equality themself, therefore

$$T\Delta^{(n)} = \text{span}\{v \in \mathbb{R}^n \text{ with } \langle v, \vec{1} \rangle = 0\} = \{v \in \mathbb{R}^n \text{ with } \langle v, \vec{1} \rangle = 0\},$$

which was the claim.

Lemma A2. ([Tu10, Example 8.4])
The differential map between $\mathbb{R}^n$ and $\mathbb{R}^m$ is the Jacobian matrix.

Proof (idea). Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a map of sufficient regularity. The differential map with regard to F, we denote it by dF_p, maps the tangent space at p to the tangent space at $F(p)$. That means that for some $v \in T^a_p \mathbb{R}^n$ the object $dF_p(v)$ is again an (abstract) tangent vector. In particular, $dF_p(v)$ is a derivation (see Lemma A1). We define the differential map by

$$(dF_p(v))f = v(f \circ F).$$

for $f \in C^\infty$. One important property of the differential map that we do not show, is its linearity. We can fully determine the map by investigating how it operation

[2] It should be mentioned that these directional derivatives, strictly speaking, are not the ones used throughout the thesis. Here the limit $t \to 0$ is **not** only from above. For differentiable functions the limit from above and below coincide, so this distinction plays no role here.

on base vectors. As discussed in the previous lemma we can obtain all derivations by looking at directional derivatives. In particular we can construct bases of the tangent spaces for $T_p^a\mathbb{R}^n$

$$\left\{ \frac{\partial}{\partial x_i}\bigg|_p, \quad 1 \le i \le n \right\}$$

and for $T_{F(p)}^a\mathbb{R}^m$

$$\left\{ \frac{\partial}{\partial y_i}\bigg|_{F(p)}, \quad 1 \le i \le m \right\}$$

where x_i and y_i are the coordinates for $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively.

How does dF_p act on these vectors? Apply the derivation dF_p to a base vector of $T_p^a\mathbb{R}^n$, we can express this operator as a linear combination of base vectors of $T_{F(p)}^a\mathbb{R}^m$, by the linearity. That means there exists a matrix a_j^k such that

$$dF_p\left(\frac{\partial}{\partial x_j}\bigg|_p \right) = \sum_k a_j^k \frac{\partial}{\partial y_k}\bigg|_{F(p)}.$$

This matrix a_j^k fully describes the operator.

We want to show that a_j^k is indeed the Jacobian matrix. We can plug the coordinate function $(y_1, \ldots, y_m) \mapsto y_i$ into the above functional. Then by the defining property the left side is

$$dF_p\left(\frac{\partial}{\partial x_j}\bigg|_p \right) y_i = \frac{\partial}{\partial x_j}\bigg|_p (y_i \circ F) = \frac{\partial F_i}{\partial x_j}(p).$$

And on the right side we have

$$\sum_k a_j^k \frac{\partial}{\partial y_k}\bigg|_{F(p)} y_i = \sum_k a_j^k \delta_{ik} = a_j^i.$$

We see that the linear operation with which dF_p acts on vectors is given by

$$a_j^i = \frac{\partial F_i}{\partial x_j}(p)$$

which is the definition of the Jacobian matrix.

B: Convex analysis

Lemma B1.
Let Φ be a real-valued, positive and concave function on $\mathbb{R}^n$. Let p_1, p_2 be two points. By $\mu(t)$ we denote the convex combination of p_1 and p_2. That is $\mu(t) = p_1(1-t) + p_2 t$.
Then, the function $[0,1] \to \mathbb{R},\ t \mapsto \log \Phi(\mu(t))$ is concave as well.

Proof. We show the definition of concave functions. Choose some $s, t \in [0,1]$ and $\alpha \in [0,1]$. First we look at convex combinations on the line from p_1 to p_2

$$\begin{aligned}
\mu((1-\alpha)s + \alpha t) &= p_1(1 - [(1-\alpha)s + \alpha t]) + p_2[(1-\alpha)s + \alpha t] \\
&= p_1 - p_1(1-\alpha)s - p_1\alpha t + p_2(1-\alpha)s + p_2\alpha t \\
&= (1-\alpha)(p_2 s - p_1 s) + \alpha(p_2 t - p_1 t) + p_1.
\end{aligned}$$

Plugging this into Φ we can use the fact that $\Phi(\cdot + p_1)$ is concave as well. We use this concavity to get

$$\begin{aligned}
\Phi(\mu((1-\alpha)s + \alpha t)) &\geq (1-\alpha)\Phi(s(p_2 - p_1) + p_1) + \alpha\Phi(t(p_2 - p_1) + p_1) \\
&= (1-\alpha)\Phi(\mu(s)) + \alpha\Phi(\mu(t)).
\end{aligned}$$

Here we used that $\mu(t) = t(p_2 - p_1) + p_1$. We only need to convince ourselves that the argument of μ is actually in $[0,1]$, but $s, t \in [0,1]$ holds by assumption.

To conclude use the monotonicity and concavity of log. We get

$$\log \Phi(\mu((1-\alpha)s + \alpha t)) \geq \log((1-\alpha)\Phi(\mu(s)) + \alpha\Phi(\mu(t)))$$

which again, has a convex combination of two points in the argument. We use the concavity of the logarithm to obtain

$$\log \Phi(\mu((1-\alpha)s + \alpha t)) \geq \alpha \log \Phi(\mu(t)) + (1-\alpha)\log \Phi(\mu(s)).$$

□

Remark: Note that by setting $t = 0$ and $s = 1$ we get $\mu(1-\alpha) = p_2(1-\alpha) + \alpha p_1$. We can use this, and do the same proof, to show that $p \mapsto \log \Phi(p)$ is concave as well.

Lemma B2.
Let U be a convex, open and bounded subset of $\mathbb{R}^n$ and $\Phi : U \to [0, \infty)$ a concave function. Furthermore, assume that there exists a point r with $\Phi(r) > 0$. Then

$\Phi > 0$ *everywhere on* U.

Proof. Assume there exists a $q \in U$ such that $\Phi(q) = 0$. We want to show a contradiction. We write q as a convex combination in U. So choose $t \in U$ such that there exists an $\alpha \in (0,1)$ with $q = r(1-\alpha) + t\alpha$. Because U is open and convex, such a t always exists and the convex combination is still in U. With the concavity we can find a contradiction

$$0 = \Phi(q) = \Phi(r(1-\alpha) + t\alpha) \geq (1-\alpha)\Phi(r) + \alpha\Phi(t) > 0$$

by using that $\Phi \geq 0$ and $\Phi(r) > 0$. □

C: Miscellaneous

Lemma 6. *Let* $T(q|p)$ *be the L-divergence we defined before, then for* $p \in \Delta^{(n)}$ *and* $v \in T\Delta^{(n)}$

$$T(p+tv|p) = H(p)(tv,tv) + o(t^2).$$

We define H *as*

$$\frac{-1}{2\Phi(p)} \frac{d^2}{dt^2} \Phi(p+tv)\Big|_{t=0} =: H(p)(v,v).$$

and call it the drift quadratic form of (π, Φ).

Proof. We do a taylor approximation of $t \mapsto T(p+tv|p)$ at $t = 0$. We calculate the first three polynomials. We have already seen that $T(p,p) = 0$, so the first polynomial is just zero. It is useful to recognize that by the chain rule $d/dt\Phi(p+tv) = \langle \nabla\Phi(p+tv), v\rangle$. The first derivative can be obtained by

$$\begin{aligned}\frac{d}{dt}T(p+tv|p) &= \frac{d}{dt}\left(-\log\frac{\Phi(p+tv)}{\Phi(p)} + \log\left(1 + \left\langle \frac{\pi(p)}{p}, tv\right\rangle\right)\right) \\ &= -\frac{\langle \nabla\Phi(p+tv), v\rangle}{\Phi(p+tv)} + \frac{\left\langle \frac{\pi(p)}{p}, v\right\rangle}{1 + \left\langle \frac{\pi(p)}{p}, tv\right\rangle}.\end{aligned}$$

At time $t = 0$ we get

$$\frac{d}{dt}T(p+tv|p)\Big|_{t=0} = \left\langle \frac{\pi(p)}{p}, v\right\rangle - \frac{D_v\Phi(p)}{\Phi(p)}.$$

We can simplify further. Note that by Proposition 2 there exists a $w \in \partial\log\Phi(p) = \nabla\log\Phi(p)$ such that

$$\left\langle \frac{\pi(p)}{p}, v\right\rangle = \left\langle w + \frac{1}{n}\sum_j \frac{\pi_j(p)}{p_j}, v\right\rangle = \langle \nabla\log\Phi(p), v\rangle + \overbrace{\langle v, 1\rangle}^{=0} \frac{1}{n}\sum_j \frac{\pi_j(p)}{p_j}.$$

Putting everything together leads to a satisfying simplification.

$$\frac{d}{dt}T(p+tv|p)|_{t=0} = \frac{1}{\Phi(p)}\langle \nabla\Phi(p), v\rangle - \frac{D_v\Phi(p)}{\Phi(p)} = 0.$$

The first and second taylor polynomial are therefore 0. We continue with the third,

$$\frac{d^2}{dt^2}T(p+tv|p) = \frac{d}{dt}\left(\frac{\left\langle \frac{\pi(p)}{p}, v\right\rangle}{1+\left\langle \frac{\pi(p)}{p}, tv\right\rangle} - \frac{\langle \nabla\Phi(p+tv), v\rangle}{\Phi(p+tv)}\right).$$

The first summand is

$$\begin{aligned}
\frac{d}{dt}\Big|_{t=0}\frac{\left\langle \frac{\pi(p)}{p}, v\right\rangle}{1+\left\langle \frac{\pi(p)}{p}, tv\right\rangle} &= \frac{\left\langle \frac{\pi(p)}{p}, v\right\rangle^2}{\left(1+\left\langle \frac{\pi(p)}{p}, tv\right\rangle\right)^2}\Bigg|_{t=0} \\
&= -\left\langle \frac{\pi(p)}{p}, v\right\rangle^2 \\
&= -\frac{1}{\Phi(p)^2}\langle \nabla\Phi(p), v\rangle^2
\end{aligned}$$

where the last equation uses the same simplification we applied to the second polynomial. The second summand requires some more calculations with the product rule

$$\begin{aligned}
&-\frac{d}{dt}\Big|_{t=0}\frac{\langle \nabla\Phi(p+tv), v\rangle}{\Phi(p+tv)} = -\frac{d}{dt}\Big|_{t=0}\frac{\sum \partial_i\Phi(p+tv)v_i}{\Phi(p+tv)} \\
&= -\frac{1}{\Phi(p+tv)^2}\left(\sum_i\sum_j \partial_i\partial_j\Phi(p+tv)v_iv_j\Phi(p+tv) - \langle \nabla\Phi(p+tv), v\rangle^2\right)\Bigg|_{t=0} \\
&= -\frac{1}{\Phi(p)}\sum_i\sum_j \partial_i\partial_j\Phi(p)v_iv_j + \frac{1}{\Phi(p)^2}\langle \nabla\Phi(p), v\rangle^2.
\end{aligned}$$

Putting everything back together the first summand cancels and we are left with

$$\frac{d^2}{dt^2}\Big|_{t=0}T(p+tv|p) = -\frac{1}{\Phi(p)}\sum_i\sum_j \partial_i\partial_j\Phi(p)v_iv_j = 2H(p)(v,v).$$

The taylor expansion can be written as

$$T(p+tv|p) = H(p)(v,v)t^2 + o(t^2) = H(p)(tv,tv) + o(t^2)$$

which we wanted to show. □

Lemma C1.
Let U be an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}$ a function. Assume that f is Fréchet differentiable at all $x \in U$, meaning (in this simple space) that there exists a linear function $A : \mathbb{R}^n \to \mathbb{R}$, the Fréchet derivative, with

$$\lim_{h \downarrow 0} \frac{|f(x+hv) - f(x) - A(hv)|}{||hv||} = 0$$

for all $v \in \mathbb{R}^n$ with $||v|| > 0$.
Then, f is also differentiable in U in the regular sense meaning that all partial derivatives exist and are continuous.

Proof. We show that $A(v) = D_v(x)$ for any v with $||v|| > 0$, it then follows that all partial derivatives exist and are continuous. Take any non-zero vector v, then

$$\begin{aligned} 0 &= \lim_{h \downarrow 0} \frac{|f(x+hv) - f(x) - A(hv)|}{||hv||} \\ &= \lim_{h \downarrow 0} \left| \frac{f(x+hv) - f(x)}{h||v||} - \frac{A(v)}{||v||} \right| \\ &= \frac{1}{||v||} \left| D_v(x) - A(v) \right| \end{aligned}$$

where we used the linearity in the second line. □

References

[BBB20] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, et al. *Adapted Wasserstein distances and stability in mathematical finance*, Finance Stoch 24 (2020), 601–632.

[BDO21] D. Bartl, S. Drapeau, J. Obłój, et al. *Sensitivity analysis of Wasserstein distributionally robust optimization problems*, ArXiv e-prints (2021), no. 2006.12022

[BSS06] M.S. Bazaraa, H.D. Sheralim, C. M. Shetty *Nonlinear programming: Theory and algorithms*, Wiley, 2006.

[Hör07] L. Hörmander, *Notions of convexity*, Modern Birkhäuser classics, Springer London, Limited, 1994.

[Lee09] J. M. Lee, *Introduction to Smooth Manifolds*, Graduate studies in mathematics, Springer, Vol. 107, 2009.

[Lee12] J. M. Lee, *Manifolds and Differential Geometry*, Graduate studies in mathematics, American Mathematical Society, Vol. 218, 2nd ed., 2012.

[Pal16] S. Pal, *Exponentially concave functions and high dimensional stochastic portfolio theory*, ArXiv e-prints (2016), no. 1603.01865

[PW14] S. Pal and T.-K. L. Wong, *The geometry of relative arbitrage*, ArXiv e-prints (2014), no. 1402.3720v4

[PW15] S. Pal and T.-K. L. Wong, *The geometry of relative arbitrage*, ArXiv e-prints (2015), no. 1402.3720

[PW16] S. Pal and T.-K. L. Wong, *Energy, entropy, and arbitrage*, ArXiv e-prints (2016), no. 1308.5376

[PW17] S. Pal and T.-K. L. Wong, *Exponentially concave functions and a new information geometry*, ArXiv e-prints (2017), no. 1605.05819

[Rai88] J. Rainwater, *Yet more on the differentiability of convex functions*, Proceedings of the American Mathematical Society 103 (1988), no. 3, 773-778.

[Roc72] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1972.

[Tu10] L. W. Tu, *An Introduction to Manifolds*, Springer, 2nd ed., 2010.

[Vil03] C. Villani, *Topics in optimal transport*, Graduate studies in mathematics, American Mathematical Society, 2003.

[Vil08] C. Villani, *Optimal transport, old and new*, vol. 338, Springer, 2008.