Geometry and Optimization of Relative Arbitrage

Ting-Kam Leonard Wong

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Soumik Pal, Chair

Krzysztof Burdzy

Zhen-Qing Chen

Program Authorized to Offer Degree:
Mathematics

University of Washington

Abstract

Geometry and Optimization of Relative Arbitrage

Ting-Kam Leonard Wong

Chair of the Supervisory Committee:
Professor Soumik Pal
Department of Mathematics

This thesis is devoted to the mathematics of volatility harvesting, the idea that extra portfolio growth may be created by systematic rebalancing. First developed by E. R. Fernholz in the late 90s and the early 2000s, stochastic portfolio theory provides a novel mathematical framework to analyze this phenomenon. A major result of the theory is the construction of portfolio strategies that outperform the market portfolio under realistic conditions. These portfolios are called relative arbitrage opportunities.

In this thesis we adopt a discrete time, pathwise approach which reveals deep connections with optimal transport, nonparametric statistics and information geometry. Our main object of study is functionally generated portfolio, a family of volatility harvesting investment strategies with remarkable properties.

This thesis consists of three parts. Part I gives a convex-analytic treatment of functionally generated portfolio and relates it with optimal transport theory. The optimal transport point of view provides the geometric structure required in order that a portfolio map is volatility harvesting.

Part II turns to optimization of functionally generated portfolio. We introduce an optimization problem analogous to shape-constrained maximum likelihood density estimation. The Bayesian version of this problem leads naturally to an extension of T. M. Cover's universal portfolio and large deviations.

Finally, in Part III we introduce and study the information geometry of exponentially concave functions, a deep and elegant geometric framework underlying the ideas of Part I. It extends the dually flat geometry of Bregman divergence studied by S. Amari and others, leading to plenty of problems for further study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am very grateful to my adviser Prof. Soumik Pal for his exceptional guidance and support throughout my PhD years. Indeed, many results in the following pages are joint work with him. Not only did he led me to an exciting area of research, but he also set a role model for good attitude and taste. During our first discussion he said I should write a thesis that I am proud of. I always remember this, and I am happy to say that I achieved it.

I wish to express my thanks to my committee members Prof. Krzysztof Burdzy, Prof. Zhen-Qing Chen and Prof. Matt Lorig for their detailed comments and suggestions. In particular, I appreciate the support and encouragement of Prof. Matt Lorig who gave me chances to speak at workshops and seminars and introduced me to eminent researchers and practitioners.

This thesis would not have been possible without the pioneering work of Dr. E. R. Fernholz on stochastic portfolio theory. His 2002 monograph *Stochastic Portfolio Theory* is a goldmine of ideas and research problems; it contains an optimal mix of interesting mathematics and practical applications that I admire very much. The other hero is Shun-ichi Amari, the founder of information geometry. It is satisfying to discover that this beautiful theory has a lot to do with portfolio theory and optimal transport.

My work on the mathematics of volatility harvesting began during an internship at Parametric Portfolio Associates in the summer of 2013. At that time we were trying to understand when and why a rebalanced portfolio outperform a buy-and-hold portfolio. I would like to express my gratitude to Paul Bouchey, now the CIO of Parametric, as well as David Stein, Vassilii Nemtchinov, Mahesh Pritamani, Alex Paulsen and Tim Li.

My understanding of the subject benefited from discussions with many researchers at the University of Washington and other institutions. In particular, I would like to thank

Professors Jon Wellner, Tatiana Toro, Robin Graham, Hoi-Ying Wong, Philip Yam, Jean-Pierre Fouque, Tomoyuki Ichiba, Ioannis Karatzas, Johannes Ruf, Erhan Bayraktar, Sergey Nadtochiy, Steve Shreve, Dmitry Kramkov, Kasper Larsen, Scott Robinson, Jin Ma, Jianfeng Zheng, Michael Magill, Walter Schachermayer and Christa Cuchiero for sharing their insights.

I am indebted to Mr. Danny To and Mr. Kai Wing Leung for opening my eyes to the world of knowledge when I was a secondary school student. Their teaching impacted me for a lifetime. I learned a lot from discussions with Prof. Sam Wong when I was an undergrad at the Chinese University of Hong Kong. I would also like to thank Prof. Ka-Sing Lau, my former supervisor at CUHK, for leading me to research.

At various stages I received warm supports from many friends. I would like to say thank you to all of you.

Last but not least, I would like to thank my parents Fion Leung and Raymond Wong, my sister Jaqueline Wong, and my wife Sze-Wing, for their love and support throughout the years. They make everything possible.

DEDICATION

To my parents and Sze-Wing

Chapter 1
INTRODUCTION

The mathematics of this thesis is motivated by some real world financial problems. In this chapter we explain the financial background and survey some relevant financial and mathematical literature. We also summarize the main results of the thesis.

1.1 Portfolio management

Consider investing in financial markets. The portfolio of an investor is the collection of all financial assets (including stocks, bonds, currencies, derivatives, etc.) the investor currently holds.[1] The problem of portfolio selection is to decide, based on available information and the investor's beliefs, the composition of the portfolio in order to maximize future wealth. Portfolio selection is by nature dynamic. As market conditions change the desired portfolio also changes, and the investor needs to adjust her holdings by trading in the market. We use the term portfolio management to refer to the comprehensive process of designing and maintaining the portfolio through time.

As one can imagine, managing a large portfolio is not an easy task. For various reasons many investors, both private and institutional, hire professional investment managers to manage their investments. The size of the asset management industry is enormous. According to a study by Boston Consulting Group [93], the global value of asset under management (AUM) rose to 62.4 trillion USD in 2012. Understanding the behaviors of financial markets and implementing successful investment strategies present great intellectual and technological challenges. Indeed, the growth of the investment industry goes hand in hand with the development of financial economics, mathematics, computer science and psychology, among

[1] In this thesis we focus on equity portfolios, i.e., portfolios consisting only of stocks.

other areas. A fascinating history can be found in the books [11], [12] and [88].

1.1.1 Modern portfolio theory

The modern approach to quantitative portfolio selection originated with Markowitz's seminal paper [75]. Also known as mean-variance analysis, this approach finds the optimal portfolio by maximizing the expected return of the portfolio for a given level of risk, where risk is measured by variance.

Mathematically, let $R_1, \ldots, R_n$ be the simple (arithmetic) returns of the assets (over some investment horizon) modeled as random variables. Let $\pi_1, \ldots, \pi_n$ be the proportions of capital invested in the assets. These are called portfolio weights and satisfy $\sum_{i=1}^n \pi_i = 1$. The portfolio return is $R_\pi = \sum_{i=1}^n \pi_i R_i$. Treating the first and second moments of $(R_1, \ldots, R_n)$ as parameters, the prototype mean-variance optimization problem is

$$\max_{\pi_1, \ldots, \pi_n} \underbrace{\sum_{i=1}^n \pi_i \mathbb{E} R_i}_{\mathbb{E} R_\pi}, \quad \text{subject to} \quad \underbrace{\sum_{i,j=1}^n \pi_i \pi_j \mathrm{Cov}(R_i, R_j)}_{\mathrm{Var}(R_\pi)} \leq \sigma_0^2, \tag{1.1.1}$$

where $\sigma_0^2 > 0$ is a fixed level of risk depending on the investor's preference. Needless to say, estimation of the expected returns and the covariance matrix is highly nontrivial and is the major challenge of the approach. We refer the reader to the treatise [24] and the book [78] for mathematical details as well as practical considerations.[2]

1.2 The market portfolio

Let us focus on equity portfolios. The collection of all stocks available in the market can be viewed as a portfolio, called the market portfolio. By definition, the market portfolio is the aggregate portfolio held by all investors. The most important property of the market portfolio is that it is capitalization-weighted. For each stock, let $X_i(t) > 0$ be its market

[2]The above discussion does not cover high frequency and algorithmic trading which has become significant in recent years. A modern mathematical treatment of this topic is [20].

capitalization defined as the multiple of the stock price and the number of outstanding shares. The portfolio weight of stock i in the market portfolio is then given by

$$\mu_i(t) = \frac{X_i(t)}{X_1(t) + \cdots + X_n(t)}. \tag{1.2.1}$$

We call $\mu_i(t)$ the market weight of stock i. Since the market portfolio plays an extremely important role in this thesis, in this section we provide some historical motivations.

1.2.1 Capital asset pricing model (CAPM)

From an equilibrium perspective, Sharpe [91] asked what happens if all investors are mean-variance optimizers in the sense of (1.1.1).[3] Under certain strong assumptions (including homogeneity of expectations and the presence of a risk-free asset), he showed that the market portfolio is mean-variance efficient, i.e., it has the highest expected return given its variance. Moreover, he showed that the expected return of each asset satisfies the relation

$$\mathbb{E}R_i = R_f + \beta_i \left(\mathbb{E}R_m - R_f\right), \tag{1.2.2}$$

where R_f is the risk-free return, R_m is the market return, and $\beta_i = \text{Cov}\left(R_i, R_m\right)/\text{Var}(R_m)$ is called the beta of stock i. The capital asset pricing model advocates the idea that aggregate market risk cannot be eliminated by diversification (i.e., holding multiple assets) and thus must be rewarded (i.e., yields higher expected return) in equilibrium.

The relation (1.2.2) may also be regarded as a regression equation (in this context it is called the single index model). This simplifies tremendously the structure of the covariance matrix needed in mean-variance optimization. Extensions of (1.2.2) gave rise to various asset pricing models which attempt to explain and predict the cross-sectional variations of stock returns in terms of economic and statistical factors. Such models are heavily applied

[3]The intellectual history of financial economics is convoluted. For example, the CAPM was developed independently by Treynor (1961, 1962), Sharpe (1964), Lintner (1965) and Mossin (1966). However, Sharpe's work was the one that became well-known. (In fact, as Markowitz himself noted, mean-variance analysis was developed independently by Arthur D. Roy (1952).) Since our main purpose is to explain the financial ideas relevant to our study, we do not attempt to cite all original papers. For a more accurate history of the subject to refer the reader to [88].

in portfolio theory and management (see for example [38], [6] and [24]). An important example is the BARRA risk model which is used to estimate the covariance matrix and factor exposures.

1.2.2 Efficient market hypothesis (EMH)

The capital asset pricing model suggests that the market portfolio is in some sense a desirable portfolio. This gives a theoretical justification of index funds whose purpose is to track the performance of some market index (such as S&P500). Since forecasting and stock picking are not attempted, this form of investment is known as passive management.

A further boost to passive management is given by the efficient market hypothesis which was first developed in the mid-60s and early 70s. This is a huge topic (see for example the surveys [73] and [72] and the references therein) and we only discuss it briefly.[4] The main idea is the following. Since a financial market incorporates information efficiently, it is difficult, if not impossible, for investors to obtain abnormal profits by acting on historical patterns and publicly available information. For example, if everyone anticipates that a stock will rise, the price would have risen already. Moreover, there have been empirical studies (such as [71]) which found that the majority of mutual fund managers failed to outperform the passive market portfolio consistently. These findings casted serious doubts on the effectiveness of active management. In 2013 the Nobel price in economics was awarded to Eugune Fama and Robert J. Shiller for their work on the efficient market hypothesis.

1.2.3 The market portfolio as a benchmark

For the purposes of this thesis, the main importance of the market portfolio is that it is a standard benchmark of portfolio performance. Thus, it makes sense to measure portfolio values relative to that of the market portfolio (see (2.2.2) below). Indeed, the objective of many investment funds is to outperform the corresponding market index. An industry

[4] Also see the popular and influential account [74].

standard measure of relative performance (see [24]) is the information ratio defined by

$$\text{information ratio} = \frac{\mathbb{E}[\text{active return}]}{\sqrt{\text{Var}(\text{active return})}},$$

where

$$\text{active return} = \text{portfolio return} - \text{market return}.$$

A high information ratio means consistent outperformance (and, obviously, more compensation for the portfolio manager).

Before going to the next section, let us remark that the efficient market hypothesis is no longer the dominant paradigm in financial economics (not to mention the investment industry). To say the least, there is now a huge literature on market anomalies, i.e., empirical facts that seem to contradict the efficient market hypothesis. Moreover, investors are not as rational as the classical theory assumes. A discussion can be found in [24, Section 2.4].

1.3 Volatility harvesting

From the perspective of mean-variance analysis, investors are risk adverse and therefore volatility should be minimized. On the the hand, volatility harvesting shows that volatility can sometimes be a source of portfolio profit. Indeed, the development of volatility harvesting is closer in spirit to optimal gambling and information theory, and is in some sense orthogonal to mainstream financial economics.[5]

1.3.1 Examples

We begin with a classic example (see [69, Example 15.2] and [33]). Consider two assets whose prices fluctuate as follows (see Figure 1.1 (left)). Asset 1 earns −50% return for all odd periods and 100% return for all even periods. On the other hand, asset 2 is a risk-free asset whose return is always 0%.

If one buys and holds any of the two assets, clearly no long term growth will be created. Nevertheless, if the investor rebalances the portfolio so that equal amount of capital is

[5] See [85] for an interesting popular science account of information theory applied to gambling and finance.

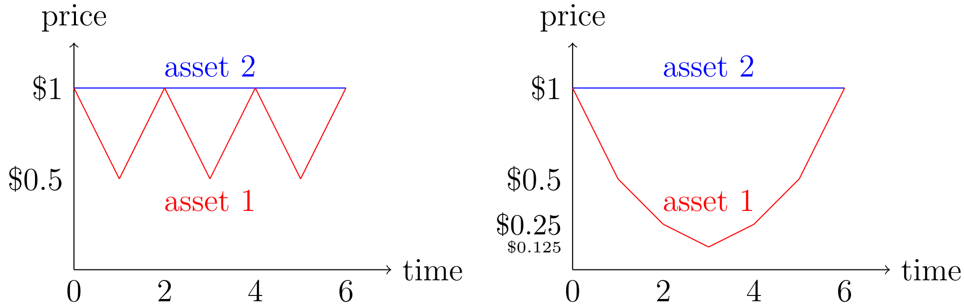

Figure 1.1: Illustration of volatility harvesting. Asset 2 is cash and asset 1 either goes up by a factor of 2 or goes down by a factor of 0.5. Six periods are shown in the figures. Left: The price pattern is $- + - + -+$. Right: The price pattern is $- - - + ++$.

invested in the two assets at the beginning of each period, the resulting equal-weighted portfolio outperforms any buy-and-hold portfolio exponentially in time.

To see why, note that the return of the equal-weighted portfolio in the first period is

$$\frac{1}{2} \times (-50\%) + \frac{1}{2} \times 0\% = -25\%,$$

and the return for the second period is

$$\frac{1}{2} \times 100\% + \frac{1}{2} \times 0\% = 50\%.$$

Over two periods the value of the portfolio grows by a factor of

$$(1 - 0.25) \times (1 + 0.5) = 0.75 \times 1.5 = 1.125.$$

Since this is strictly larger than 1, compounding gives exponential outperformance. The extra profit apparently comes from the volatility of asset 1 which is 'harvested' by buying low and selling high via maintaining the portfolio weights at $\left(\frac{1}{2}, \frac{1}{2}\right)$.

Let us give a more realistic example, taken from [13], of the same phenomenon. In Figure 1.2 (left) we plot the monthly stock prices of Starbucks and Apple from 1994 to 2012 (normalized so that they begin at \$1). Consider two portfolios: (i) the buy-and-hold portfolio which initially invests \$0.5 in each of the stocks and (ii) the equal-weighted

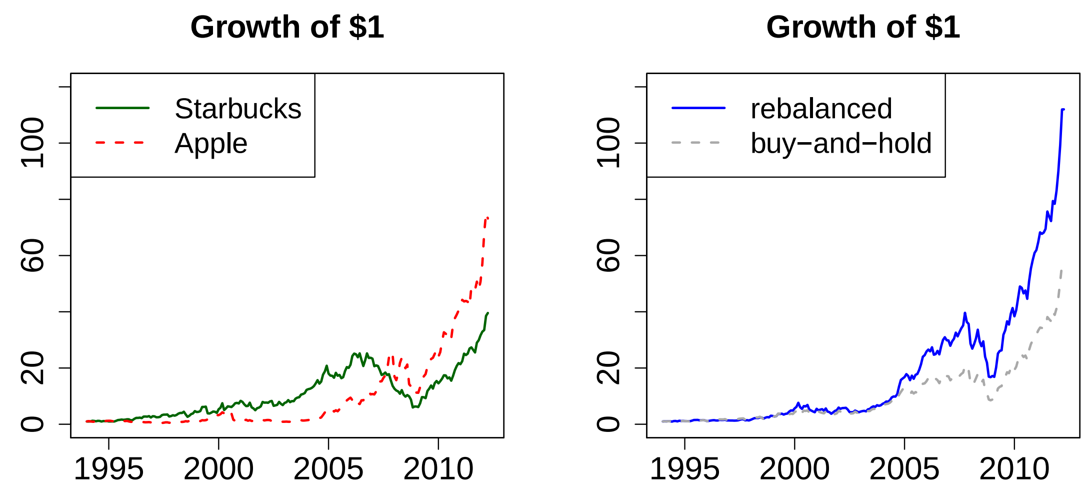

Figure 1.2: Apple-Starbucks example.

portfolio rebalanced monthly. Figure 1.2 (right) plots the time series of the values of the two portfolios. In this period, the equal-weighted portfolio outperforms significantly the buy-and-hold portfolio and the two constituent stocks.

Further empirical examples can be found, for example, in [13], [84] and [14]. Note that a buy-and-hold portfolio is capitalization-weighted: if $X_i(t)$ are the (normalized) prices of the underlying assets, the value of a buy-and-hold portfolio has the form

$$Z(t) = c_1 X_1(t) + \cdots + c_n X_n(t).$$

Thus c_i is proportional to the number of shares of stock i held in the portfolio (which is constant for a buy-and-hold portfolio). The corresponding portfolio weights are

$$\pi_i(t) = \frac{c_i X_i(t)}{c_1 X_1(t) + \cdots + c_n X_n(t)}, \tag{1.3.1}$$

which have the same form as the market weights (1.2.1). Indeed, the market portfolio (and more generally, a market index such as S&P500) can be viewed as an approximate buy-and-hold portfolio.[6] This gives rise to the following fundamental question:

[6]While a market index represents the overall performance of the market, it does not contain all the

When and why does a rebalanced portfolio outperform a buy-and-hold portfolio?

1.3.2 Rebalanced portfolio in gambling and information theory

Rebalanced portfolios also arise in the study of optimal gambling in information theory (see [63], [17], [1], [27] and [70]). Suppose $X_1(t), \ldots, X_n(t)$ are the stock prices in discrete time. At each time t the investor chooses a vector $\pi(t)$ of non-negative portfolio weights summing to 1. The value of her portfolio satisfies

$$Z(t+1) = Z(t) \sum_{i=1}^{n} \pi_i(t) \frac{X_i(t+1)}{X_i(t)}.$$

Suppose $X(t) = (X_1(t), \ldots, X_n(t))$ is a stochastic process, and let $\{\mathcal{F}_t\}$ be the natural filtration. Under mild conditions, it can be shown that the portfolio

$$\pi(t) = \arg\max_{\pi(t)} \mathbb{E}[\log Z(t+1) | \mathcal{F}_t] \tag{1.3.2}$$

beats any other portfolio asymptotically. The criterion (1.3.2) is known as the Kelly criterion. In particular, if the gross returns $\frac{X_i(t+1)}{X_i(t)}$ are i.i.d., the Kelly portfolio (also called the log-optimal portfolio) $\pi(t) \equiv \pi$ is constant over time and is thus a rebalanced portfolio. Computation of the Kelly portfolio assumes that the distribution of the price process is known. Learning the optimal portfolio when the underlying distribution is unknown (or even when there are no distributional assumptions) leads to the concept of universal portfolio [28].

Since the Kelly criterion does not involve the investor's preference, this idea is not welcomed by economists who are accustomed to utility maximization. See [90] for Samuelson's famous critique of log-optimal investing.

stocks in the market. It usually contains only the largest stocks in the market, and its members may change during periodic reconstitutions. Moreover, market capitalizations may change due to corporate actions such as IPO, split and merger. These complications cause the market index to deviate from a pure buy-and-hold portfolio.

1.3.3 Explaining the rebalancing premium

The extra profit that a rebalanced portfolio earns over a buy-and-hold portfolio is sometimes called the 'rebalancing premium'. It is both interesting and important to understand the source of this 'premium'. From an asset pricing perspective, [84] analyzed the factor exposures of the equal-weighted portfolio in US equity market. A disadvantage of this approach is that the analysis is subject to statistical errors and depends on the sample used.

It was observed by many researchers that the following consequence of Jensen's inequality plays an important role in rebalancing. If R_i is the simple return of stock i, we have

$$\log\left(1+\sum_{i=1}^{n}\pi_i R_i\right) \geq \sum_{i=1}^{n}\pi_i \log(1+R_i). \tag{1.3.3}$$

In words, we have

log return of portfolio ≥ weighted average log return of underlying assets.

The difference between the two quantities in (1.3.3) has been studied under many names. It is called the diversification return in [50], [37], [102] and [22], the excess growth rate in [42] and [41], the rebalancing premium in [13], the volatility return in [51], among others.

In spite of the large volume of literature on the theory and practice of rebalancing, there are some confusions among academics and practitioners. This, we believe, is due to the fact that in most theoretical analyses of rebalancing very specific assumptions (such as geometric Brownian motion) are imposed on the dynamics of asset prices, giving a false impression that volatility harvesting only works in those situations. An important case in point is the (wrong) assertion that rebalancing is profitable only when the underlying price changes are negatively correlated like in the first example in Section 1.3.1 (see [22]). There is thus a strong need to study the precise conditions under which rebalancing or volatility harvesting beats a capitalization-weighted portfolio.

1.4 Stochastic portfolio theory

While mainstream portfolio theory depends heavily on utility maximization, distributional assumptions and asset pricing models, stochastic portfolio theory offers a novel mathematical framework for portfolio management. This mathematical framework was first developed by E. R. Fernholz. Some ideas are implicit in his early paper [42] with Shay, and his work eventually led to the 2002 monograph [41]. More recent results are surveyed in [48] (also see [98]).

A major idea of stochastic portfolio theory is to investigate properties of portfolios that are independent of distributional assumptions on stock returns. Moreover, under certain realistic structural conditions on market behaviors, some rebalancing portfolio can be shown to outperform the market portfolio. In Fernholz's formulation, the vector process $X(t) = (X_1(t), \ldots, X_n(t))$ of market capitalizations is modeled as a general Itô process in continuous time. A (self-financing) portfolio is given by a vector-valued progressively measurable process $\{\pi(t) = (\pi_1(t), \ldots, \pi_n(t))\}$. Let $Z_\pi(t)$ and $Z_\mu(t)$ be the growths of \$1 of the portfolio π and the market portfolio μ respectively. We are interested in the ratio

$$V_\pi(t) = \frac{Z_\pi(t)}{Z_\mu(t)} \tag{1.4.1}$$

called the relative value of the portfolio π.

Given a portfolio process π, it can be shown by Itô calculus that $\log V_\pi(t)$ satisfies

$$d \log V_\pi(t) = \sum_{i=1}^{n} \pi_i(t) d \log \mu_i(t) + \gamma_\pi^*(t) dt, \tag{1.4.2}$$

where

$$\gamma_\pi^*(t) = \frac{1}{2} \sum_{i,j=1}^{n} \pi_i(t) \left(\delta_{ij} - \pi_j(t)\right) \frac{d\langle \log \mu_i, \log \mu_j \rangle_t}{dt} \tag{1.4.3}$$

is the excess growth rate in continuous time (see (1.3.3)).

The equation (1.4.2) is important for several reasons. First, unlike in mean-variance analysis, (relative) portfolio value is measured in logarithmic scale which is additive in time. Second, the dynamics of the relative value is expressed solely in terms of $\pi(t)$ and the

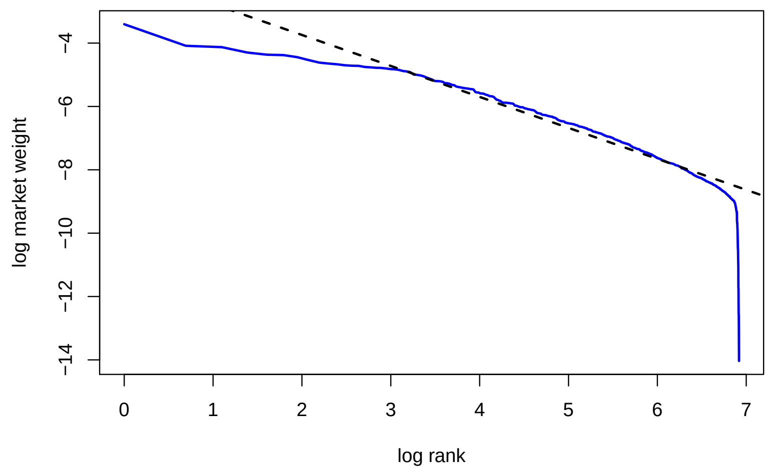

Figure 1.3: Capital distribution curve of the Russel 1000 index in June 2015 (this data set is taken from [79]). The dotted line is the Pareto approximation fitted by ordinary least squares. The slope of the line is -0.95.

market weights $\mu(t)$. This forces us to consider the properties of market weights instead of the absolute prices. Mathematically, the state space of the market becomes the open unit simplex

$$\Delta_n := \left\{ p = (p_1, \ldots, p_n) \in \mathbb{R}^n : p_i > 0, \sum_{i=1}^n p_i = 1 \right\}. \tag{1.4.4}$$

The capital distribution of the market is defined by rearranging the market weights from largest to smallest:

$$\mu_{(1)}(t) \geq \mu_{(2)}(t) \geq \cdots \geq \mu_{(n)}(t)$$

An important fact (see [41, Chapter 4]) is that the capital distribution of a large market is stable over time and is approximately Pareto distributed (see Figure 1.3 for an example). On the mathematical side, this leads to a large literature on Atlas and ranked-based models of interacting diffusions; see [8, 23, 57, 46, 56].

For a general portfolio process $\{\pi(t)\}$, integration of (1.4.2) contains a stochastic integral which is difficult to analyze. A remarkable discovery of Fernholz is that for a special class of portfolio processes which are functions of $\mu(t)$, (1.4.2) can be integrated free of stochastic integrals. For these portfolios – called functionally generated portfolios – $\log V_\pi(t)$ depends solely on $\mu(0)$, $\mu(t)$ and a finite variation process related to time-aggregated market volatility. Using this almost sure pathwise decomposition formula, Fernholz was able to formulate conditions under which the portfolio outperforms the market portfolio with probability 1 for all sufficiently long horizons, i.e., there exists a constant t_0 such that

$$\mathbb{P}\left(Z_\pi(t) > Z_\mu(t)\right) = 1, \quad \text{for all } t \geq t_0. \tag{1.4.5}$$

Such a portfolio is called a relative arbitrage opportunity with respect to the market portfolio. The existence of relative arbitrage opportunities implies that the underlying market model does not admit any equivalent martingale measure. See for example [80, 39, 40, 9, 89, 61] for the analysis of strict local martingales and optimal arbitrages that arise in this context.

The conditions Fernholz used are (i) diversity and (ii) sufficient volatility. By definition, the market is said to be diverse if there exists $\delta > 0$ such that

$$\sup_{t \geq 0} \max_{1 \leq i \leq n} \mu_i(t) \leq 1 - \delta \tag{1.4.6}$$

almost surely. That is, no stock is ever allowed to dominate most of the market. A version of sufficient volatility is non-degeneracy, i.e., the matrix $\sigma_{ij}(t) = \frac{d\langle \log \mu_i, \log \mu_j \rangle_t}{dt}$ satisfies a uniform elliptic condition. Under these conditions relative arbitrage opportunities can be constructed using functionally generated portfolios (see [41, 49, 47, 7]). A remarkable advantage of these portfolios is that their portfolio weights are deterministic functions of the current market weights, i.e., $\pi(t) = F(\mu(t))$ for certain maps F. In other words, implementation of these portfolios does not require dynamic estimation of parameter and optimization.

1.5 Outline of the thesis

Fernholz's results give a partial answer to the question about rebalancing: for a rebalanced portfolio to outperform the market portfolio, a sufficient condition is that the market is diverse and sufficiently volatile (in suitable senses). While Fernholz (and most authors in stochastic portfolio theory) works in continuous time, actual trading takes place at discrete time points. Also, despite its theoretical significance, the original definition of functionally generated portfolio (see [41, Theorem 3.1.5]) is somewhat obscure, and it was not immediate why this is a natural family of portfolios to consider.[7] (See [60] for a recent study which interprets portfolio generating functions as Lyapunov functions.)

To address these issues we adopt a discrete time, pathwise approach which not only clarifies the arguments but also reveals deep mathematical connections with optimal transport (Part I), nonparametric statistics (Part II) and information geometry (Part III). To illustrate why a discrete time approach may be better, let us consider again the first example in Section 1.3.1. Suppose the returns of asset 1 are reshuffled over time as in Figure 1.1 (right). Now except the fourth period the price changes are positively correlated. The growth of the equal-weighted portfolio over the six periods remains unchanged because we can rearrange the factors:

$$0.75 \times 0.75 \times 0.75 \times 1.5 \times 1.5 \times 1.5 = (0.75 \times 1.5)^3.$$

Thus, the key driver of the long term growth of the rebalancing portfolio is the number of times the growth factor 0.75×1.5 can be matched.[8] Note that matching of two opposing moves happening at different times is not captured by continuous time stochastic calculus. More importantly, the discrete time approach allows us to focus on path properties of market relevant to volatility harvesting without any stochastic modeling assumptions. Indeed, the basic definitions to be given in Chapter 2 involve no probability at all.

[7] In a talk given in the 2015 conference 'Stochastic Portfolio Theory and related topics' at Columbia University, Fernholz said that the concept of functionally generated portfolio was discovered after numerous computations.

[8] This also shows that negative correlation is not needed for rebalancing to be profitable, at least in the long run.

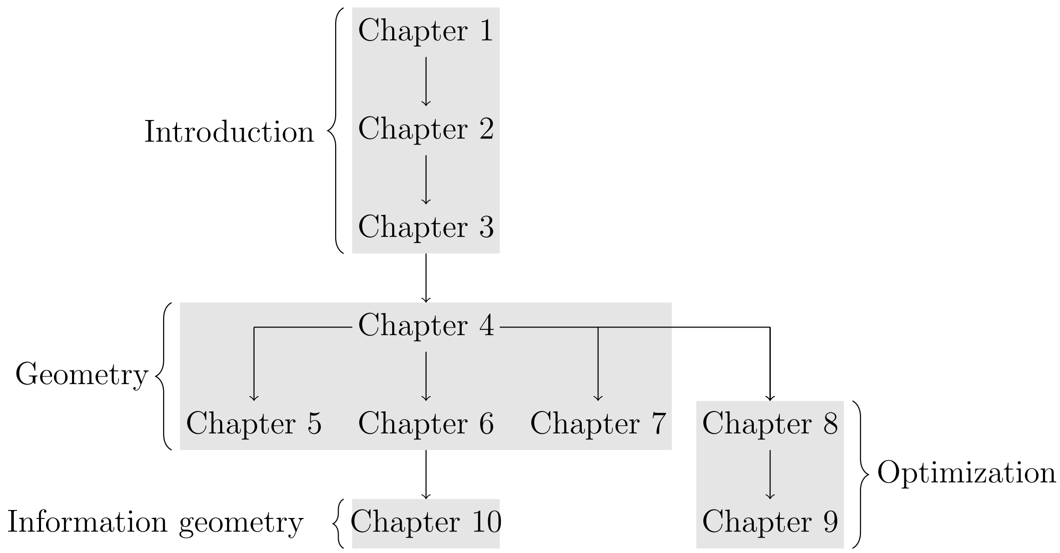

Figure 1.4: Interdependence of the chapters.

1.5.1 Outline

Now we give a more detailed description of the content of the thesis which is mainly based on the results in [81, 83, 103, 104, 14, 82]. The papers [81, 83, 82] are joint work with Soumik Pal, and the paper [14] is joint work with Paul Bouchey and Vassillii Nemtchinov. See Figure 1.4 for the interdependence of the chapters.

In Chapter 2 we introduce the mathematical set up which will be adopted throughout the thesis. As noted above we work under a discrete time model. Under some simplifying assumptions, the relative value (see (1.4.1)) of a portfolio π relative to the market portfolio (with weights given by (1.2.1)) satisfies $V_\pi(0) = 1$ and the relation

$$\frac{V_\pi(t+1)}{V_\pi(t)} = \sum_{i=1}^{n} \pi_i(t) \frac{\mu_i(t+1)}{\mu_i(t)}. \tag{1.5.1}$$

Here $\{\mu(t)\}_{t=0}^{\infty}$ is a more or less arbitrary sequence taking values in the simplex Δ_n. For the most part we focus on portfolios whose weights are deterministic functions of the current market weights. This gives rise to the concept of portfolio map, i.e., a function mapping Δ_n into $\overline{\Delta}_n$, the closed unit simplex.

In Chapter 3 we illustrate the pathwise approach by analyzing constant-weighted portfolios, i.e., rebalanced portfolios whose weights are constant over time. In particular, we formulate conditions under which a constant-weighted portfolio outperforms the market portfolio in the long run. Constant-weighted portfolios are basic examples of functionally generated portfolio. We also study the performance of constant-weighted portfolios as a function of the portfolio weights.

The rest of the thesis is divided into three parts.

Part I: Geometry

In Part I we give a convex-analytic treatment of functionally generated portfolio in relation to optimal transport theory.

In Chapter 4 we define a functionally generated portfolio as a portfolio map $\pi : \Delta_n \rightarrow \overline{\Delta}_n$ associated with a concave function $\Phi : \Delta_n \rightarrow (0, \infty)$, called the generating function of π (Definition 4.1.1). Its logarithm $\varphi = \log \Phi$ is said to be exponentially concave on Δ_n. The analysis of exponentially concave functions plays an important role in our study.

Geometrically, the portfolio weights of a functionally generated portfolio are given in terms of the supergradients of the log generating function $\varphi = \log \Phi$. For such a portfolio, we can decompose its relative performance in the form (Proposition 4.1.3)

$$\log \frac{V_\pi(t+1)}{V_\pi(t)} = \varphi(\mu(t+1)) - \varphi(\mu(t)) + T\left(\mu(t+1) \mid \mu(t)\right), \tag{1.5.2}$$

where

$$T\left(q \mid p\right) := \log\left(\sum_{i=1}^n \pi_i(p) \frac{q_i}{p_i}\right) - \left(\varphi(q) - \varphi(p)\right) \geq 0 \tag{1.5.3}$$

is called the L-divergence of the portfolio π (the L stands for logarithmic). It is a nonnegative functional on $\Delta_n \times \Delta_n$ measuring the market volatility harvested by the portfolio. It is a generalization of excess growth rate defined in (1.3.3).

Functionally generated portfolio can also be characterized by the following property: for any discrete cycle $\{\mu(t)\}_{t=0}^{m+1}$ in Δ_n satisfying $\mu(m+1) = \mu(0)$, we have

$$V_\pi(m+1) = \prod_{t=0}^{m} \left(\sum_{i=1}^n \pi_i(\mu(t)) \frac{\mu_i(t+1)}{\mu_i(t)}\right) \geq 1. \tag{1.5.4}$$

The property (1.5.4) is called multiplicative cyclical monotonicity (MCM) (Definition 4.2.2). The MCM property captures geometrically the idea of volatility harvesting.

Chapter 5 gives a financial justification to the concept of functionally generated portfolio. Using the notion of pseudo-arbitrage (Definition 5.1.1), we show that functionally generated portfolios are, in a sense, the only portfolio maps capable of generating relative arbitrage opportunities if the market is only assumed to be diverse (in a generalized sense) and sufficiently volatile.

Chapter 6 is a key chapter. Using the MCM property, we show that functionally generated portfolios can be interpreted as the optimal transport maps of a remarkable optimal transport problem. To formulate this transport problem, we use the exponential coordinate system of Δ_n viewed as a smooth manifold: for $p \in \Delta_n$, let

$$\theta_i = \log \frac{p_i}{p_n}, \quad i = 1, \ldots, n-1.$$

We call $\theta = (\theta_1, \ldots, \theta_{n-1}) \in \mathbb{R}^{n-1}$ the exponential coordinates of p. Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}^{n-1}$ and consider the Kantorovich optimal transport problem with the cost function

$$c(\theta, \phi) = \psi(\theta - \phi), \quad \theta \in \mathcal{X}, \quad \phi \in \mathcal{Y}, \tag{1.5.5}$$

where

$$\psi(x) = \log \left(1 + \sum_{i=1}^{n-1} e_i^x\right).$$

We prove that the optimal coupling is deterministic, and is given by

$$\phi_i = \theta_i - \log \frac{\pi_i(\theta)}{\pi_n(\theta)}, \quad i = 1, \ldots, n-1,$$

where π is a functionally generated portfolio viewed as a function of the exponential coordinates. The main idea is to show that the MCM property is equivalent to c-cyclical monotonicity (Theorem 6.3.1). This result shows that the following objects are essentially equivalent: (i) functionally generated portfolio, (ii) exponentially concave function, (iii) c-concave function, (iv) the optimal transport map. In Part III we will study these objects using the tools of information geometry.

Given two functionally generated portfolios, it is natural to ask if one is 'more volatility harvesting' than the other. The L-divergence (1.5.3) reflects the concavity of the generating function and provides a natural partial ordering among functionally generated portfolios:

$$\tau \succeq \pi \Leftrightarrow T_{\tau}\left(q \mid p\right) \geq T_{\pi}\left(q \mid p\right). \tag{1.5.6}$$

In Chapter 7 we study the maximal elements of the partial order defined by (1.5.6). Let $e(1) = (1, 0, \ldots, 0)$ and $\overline{e} = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$ be a corner and the barycenter of Δ_n respectively. Restricting to the class of continuously differentiable portfolio maps, we show that if π is generated by a C^2 symmetric concave function Φ and

$$\int_0^1 \frac{1}{\Phi(t\overline{e} + (1-t)e(1))^2} dt = \infty, \tag{1.5.7}$$

then π is maximal (Theorem 7.1.7). In other words, no portfolio maps can beat π in all diverse and sufficiently volatile market. Some portfolios that satisfy (1.5.7) are the equal-weighted portfolio and the entropy-weighted portfolio.

Part II: Optimization

In Part I we showed that functionally generated portfolios are volatility harvesting. Given historical or simulated data, it is natural to find an optimal portfolio, where optimality is defined in terms of expected or asymptotic growth rate. A major difficulty is that the set $\mathcal{FG}$ of functionally generated portfolios is infinite dimensional.[9]

Our approach is motivated by a seemingly unrelated problem in nonparametric statistics. A density f_0 on $\mathbb{R}^d$ is said to be log-concave if $\log f_0$ is concave. For example, Gaussian densities are log-concave. Suppose $X_1, \ldots, X_N$ are random samples from an unknown log-concave density f_0. The nonparametric maximum likelihood estimate of f_0 is defined by

$$\widehat{f} = \underset{f \text{ log-concave}}{\arg\max} \sum_{i=1}^{N} \log f(X_i). \tag{1.5.8}$$

[9]Nevertheless, a blessing fact is that $\mathcal{FG}$ is convex.

It can be shown that (1.5.8) has a unique solution which satisfies nice statistical properties (see for example [31]). Since $\log f$ is concave, the optimization problem (1.5.8) is said to be shape-constrained.

In Chapter 8 we introduce an analogous problem for functionally generated portfolio. Let $\mathbb{P}$ be a Borel probability measure on $\Delta_n \times \Delta_n$ representing historical or simulated data. Consider the optimization problem

$$\widehat{\pi} = \operatorname*{arg\,max}_{\pi \in \mathcal{FG}} \int_{\Delta_n \times \Delta_n} \log \left(\sum_{i=1}^n \pi_i(p) \frac{q_i}{p_i} \right) d\mathbb{P}, \tag{1.5.9}$$

where $\mathcal{FG}$ is the family of functionally generated portfolios. The solution $\widehat{\pi}$ can be interpreted as the maximum likelihood estimate of the optimal portfolio (note that $\mathbb{P}$ can be regarded as an approximation of the true process). In Chapter 8 we establish existence and uniqueness results (Theorems 8.2.2 and 8.3.1) which are analogous to those for (1.5.8). Note that (1.5.9) is conceptually more difficult than (1.5.8) since a functionally generated portfolio is given by the supergradients of an exponentially concave function. We show further that the 'estimator' $\widehat{\pi}$ is consistent: under certain regularity conditions, if $\mathbb{P}^{(N)} \rightarrow \mathbb{P}$ weakly, and $\widehat{\pi}^{(N)}$ and $\widehat{\pi}$ are the estimators for $\mathbb{P}^{(N)}$ and $\mathbb{P}$, then $\widehat{\pi}^{(N)} \rightarrow \widehat{\pi}$ almost everywhere on Δ_n.

In Chapter 9 we consider a Bayesian version of the problem (1.5.9). Endow the space $\mathcal{FG}$ with the topology of uniform convergence. Let ν_0 be a Borel probability measure on $\mathcal{FG}$, interpreted as the prior distribution. Given the path of the market weights up to time t, we can define the posterior distribution ν_t by

$$\nu_t(B) = \frac{1}{\widehat{V}(t)} \int_B V_\pi(t) d\nu_0(t), \quad B \subset \mathcal{FG} \text{ Borel}, \tag{1.5.10}$$

where

$$\widehat{V}(t) = \int_{\mathcal{FG}} V_\pi(t) d\nu_0(t) \tag{1.5.11}$$

is the normalizing factor.

The posterior distribution ν_t can be interpreted in another way. Imagine a hypothetical market whose assets are the elements of $\mathcal{FG}$. That is, consider a market of portfolios. Let ν_0 be the initial distribution of wealth in the market. Then (1.5.10) is the wealth distribution at time t analogous to the market weight (1.2.1), and (1.5.11) is the total (relative) value of the abstract market. While the capital distribution of a a large equity market is typically stable, the wealth distribution of a family like $\mathcal{FG}$ is likely to become more and more concentrated. In Chapter 9 we quantify this by a large deviations principle (LDP). Furthermore, if one invest according to the posterior mean

$$\widehat{\pi}(t) = \int_{\mathcal{FG}} \pi(\mu(t)) d\nu_t(\pi),$$

the relative value of the portfolio is equal to (1.5.11). Under suitable conditions, we show (Theorem 9.1.3) that this portfolio has the universality property

$$\lim_{t \to \infty} \frac{1}{t} \log \frac{\widehat{V}(t)}{\max_{\pi \in \mathcal{FG}} V_{\pi}(t)} = 0.$$

This result generalizes Cover's univeral portfolio [28] to the family of functionally generated portfolios.

Part III: Information geometry

Finally, in Chapter 10 we show that information geometry provides the appropriate framework to study the geometric ideas in Part I (see Figure 1.5). For a smooth exponentially concave function φ on Δ_n, the L-divergence (1.5.3) can be written in the form

$$T\left(q \mid p\right) = \log\left(1 + \nabla\varphi(p) \cdot (q - p)\right) - \left(\varphi(q) - \varphi(p)\right), \quad p, q \in \Delta_n.$$

We show that the L-divergence induces a Riemannian metric g on Δ_n as well as a pair (∇, ∇^*) of affine connections which are dual to each other. This geometric structure has deep connections with the optimal transport problem studied in Chapter 6. To give an example of our results we state the following generalized Pythagorean theorem (Theorem 10.1.1): for $p, q, r \in \Delta_n$, the equality

$$T\left(q \mid p\right) + T\left(r \mid q\right) = T\left(r \mid p\right)$$

holds if and only if the ∇^*-geodesic joining p and q is g-orthogonal to the ∇-geodesic joining q and r. This extends the information geometry of Bregman divergence to L-divergence.

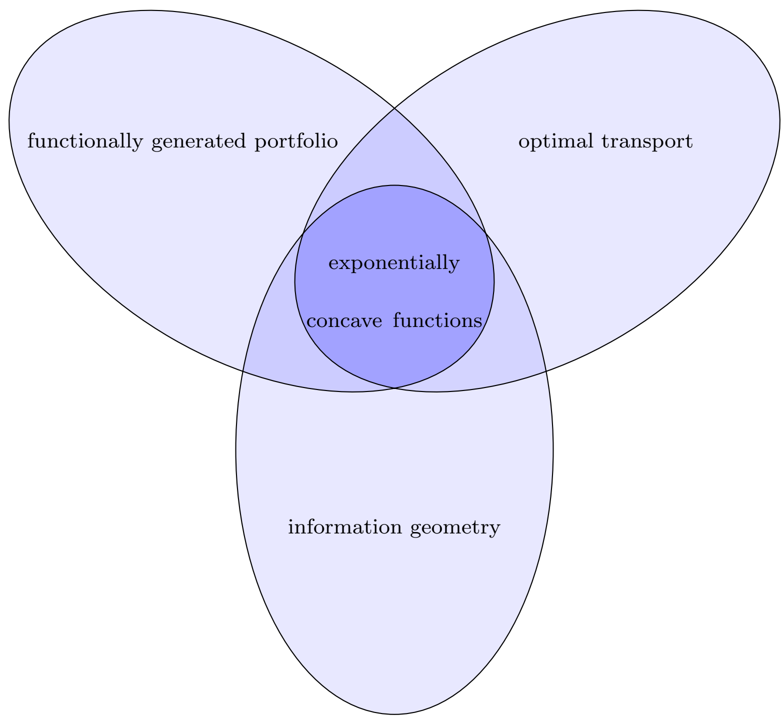

Figure 1.5: Main theme of Part III.

Chapter 2

A DISCRETE TIME MARKET MODEL

In this section we introduce the mathematical set up which will be adopted throughout this thesis. In our framework, the dynamics of asset prices are given exogenously as sequences in the unit simplex satisfying certain structural conditions. In this sense, we are taking a model-free, pathwise approach. Nevertheless, our model contains significant and important simplifications of the way real equity markets operate. These idealizations will be made explicit as we go along.

2.1 Stocks and market weights

Consider an equity market with $n \geq 2$ stocks. We assume that the stocks are infinitely divisible. Without loss of generality, we suppose that each stock has a single share outstanding. Accordingly, the price of the stock is equal to its market capitalization. Time is taken to be discrete. For $i \in \{1, \ldots, n\}$ and $t \geq 0$, we let $X_i(t) > 0$ be the market capitalization of stock i at time t. For convenience, we use dollar ($) as the unit of money. Note that $X_1(t) + \cdots + X_n(t)$ is the total capitalization of the market at time t.

Definition 2.1.1 (Market weight)**.** The market weight of stock i at time t is defined by

$$\mu_i(t) = \frac{X_i(t)}{X_1(t) + \cdots + X_n(t)}. \tag{2.1.1}$$

We let $\mu(t)$ be the vector $(\mu_1(t), \ldots, \mu_n(t))$.

The market weight vector $\mu(t)$ is a probability vector with positive components. Thus, it is an element of the open unit simplex defined by (1.4.4). We let $\overline{\Delta}_n$ be the closed unit simplex which is the closure of Δ_n in $\mathbb{R}^n$. The evolution of the market can be visualized as

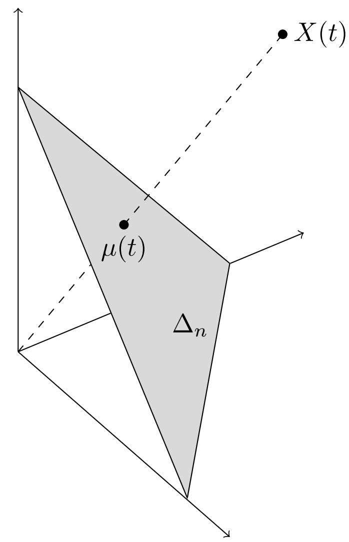

Figure 2.1: The market weight vector $\mu(t)$ is the projection of the price vector $X(t)$ onto the simplex Δ_n.

a (discrete) path in Δ_n (see Figure 2.1). Implicitly, we assume that the number of stocks is constant over time, and the firms do not go bankrupt ($X_i(t) > 0$ for all t).

For simplicity, we assume that the stocks do not pay dividends and there are no corporate actions such as public offerings. Thus all changes in market capitalizations are due to price changes. Explicitly, suppose that $R_i(t)$ is the simple return of stock i over the time period $[t, t+1]$, i.e.,

$$X_i(t+1) = X_i(t)\,(1 + R_i(t))\,.$$

From (2.1.1), the market weights can be updated by the formula

$$\mu_i(t+1) = \frac{\mu_i(t)\,(1 + R_i(t))}{\mu_1(t)\,(1 + R_1(t)) + \cdots + \mu_n(t)\,(1 + R_n(t))}. \tag{2.1.2}$$

For our purposes, the market weights contain all relevant information (see Lemma 2.2.1). Thus the basic object of our model is a sequence of market weight vectors.

Definition 2.1.2 (Market path)**.** A market path is defined by a sequence $\{\mu(t)\}_{t=0}^{\infty}$ with values in the open unit simplex Δ_n.

In Definition 2.1.2, the sequence of market weights is completely arbitrary. In particular, we do not impose the usual assumption that $\{\mu(t)\}_{t=0}^{\infty}$ is a stochastic process. Of course, to obtain meaningful results the market cannot behave arbitrarily. Here is one of the first structural conditions explored in stochastic portfolio theory.

Definition 2.1.3 (Diversity)**.** The market $\{\mu(t)\}_{t=0}^{\infty}$ is said to be diverse if there exists $\delta > 0$ such that $\max_{1 \leq i \leq n} \mu_i(t) \leq 1 - \delta$ for all t.

More generally, we may consider the condition $\mu(t) \in K$ for all t where K is a suitable subset of Δ_n. Other conditions, including sufficient volatility, will be introduced later.

2.2 Portfolio and relative value

A portfolio vector is an element of $\overline{\Delta}_n$. Its components represents the proportions of capital invested in each of the stocks. By a portfolio we mean a sequence $\pi = \{\pi(t)\}_{t=0}^{\infty}$ of portfolio weight vectors that are chosen sequentially in time. Given a portfolio π, we can define a self-financing portfolio whose distribution of capital at time t is $\pi(t)$. Note that by considering only elements of $\overline{\Delta}_n$, the portfolio is fully invested in the stock market and does not hold short positions.

By normalization, we suppose that all portfolios begin with \$1 at time 0. We also assume that there are no transaction costs. Recall that $R_i(t)$ is the simple return of stock i over the time interval $[t, t+1]$. If $Z_\pi(t)$ denotes the value of the self-financing portfolio π at time t, then $Z_\pi(0) = 1$ by definition and, by linearity of simple return, we have

$$Z_\pi(t+1) = Z_\pi(t)\left(1 + \sum_{i=1}^{n} \pi_i(t) R_i(t)\right). \tag{2.2.1}$$

The most important portfolio is the market portfolio whose portfolio vector at time t is $\mu(t)$ (see Section 1.2). With a slight abuse of notation, we will denote the market portfolio by μ. Using (2.1.2) and (2.2.1), it is easy to verify that

$$Z_\mu(t) = \frac{1}{X_1(0) + \cdots + X_n(0)}\left(X_1(t) + \cdots + X_n(t)\right).$$

Consider the ratio

$$V_\pi(t) = \frac{Z_\pi(t)}{Z_\mu(t)} \tag{2.2.2}$$

between the value of π and the value of μ. We call it the relative value of the portfolio π. To the best of our knowledge this definition is due to Fernholz (see [41, Chapter 1]).

Lemma 2.2.1 (Relative value)**.** *The relative value defined by* (2.2.2) *satisfies* $V_\pi(0) = 1$ *and*

$$V_\pi(t+1) = V_\pi(t) \sum_{i=1}^n \pi_i(t) \frac{\mu_i(t+1)}{\mu_i(t)}. \tag{2.2.3}$$

Proof. Since $Z_\pi(0) = Z_\mu(0) = 1$, it is clear that $V_\pi(0) = 1$. By (2.2.1), we have

$$\frac{Z_\pi(t+1)}{Z_\pi(t)} = \sum_{i=1}^n \pi_i(t)\left(1 + R_i(t)\right)$$

and

$$\frac{Z_\mu(t+1)}{Z_\mu(t)} = \sum_{i=1}^n \mu_i(t)\left(1 + R_i(t)\right).$$

By (2.1.2), we have

$$\begin{aligned} \frac{V_\pi(t+1)}{V_\pi(t)} &= \frac{Z_\pi(t+1)/Z_\pi(t)}{Z_\mu(t+1)/Z_\mu(t)} \\ &= \sum_{i=1}^n \pi_i(t) \frac{1 + R_i(t)}{\sum_{j=1}^n \mu_j(t)(1 + R_j(t))} \\ &= \sum_{i=1}^n \pi_i(t) \frac{\mu_i(t+1)}{\mu_i(t)}. \end{aligned}$$

□

By Lemma 2.2.1, the dynamics of the relative value depends only on the market weights, and the stock returns $R_i(t)$ enter indirectly. Since we regard the market weight vector as the primary object, we may define the relative value directly by (2.2.3). If we denote by $a \cdot b$ the Euclidean inner product and $\frac{p}{q} = \left(\frac{p_1}{q_1}, \ldots, \frac{p_n}{q_n}\right)$ the vector of componentwise ratios, we may write (2.2.3) in the form

$$\frac{V_\pi(t+1)}{V_\pi(t)} = \pi(t) \cdot \frac{\mu(t+1)}{\mu(t)}.$$

An important example is the family of constant-weighted portfolios. Besides being the simplest non-trivial portfolios, they provide a great source of intuition. In Chapter 3 we provide an in-depth analysis of constant-weighted portfolios.

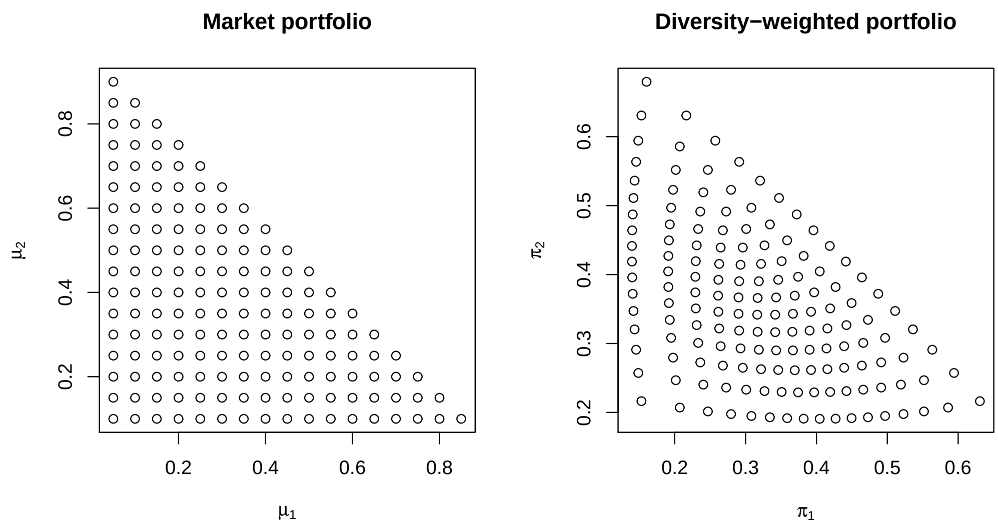

Figure 2.2: Visualization of the diversity-weighted portfolio for $\lambda = 0.5$ and $n = 3$. Each dot on the right is the image a dot on the left.

Definition 2.2.2 (Constant-weighted portfolio)**.** A portfolio π is said to be constant-weighted if $\pi(t)$ is constant over time. Abusing notation, we denote the common value by $\pi \in \overline{\Delta}_n$.

A constant-weighted portfolio chooses the same portfolio vector for all states of the market. More generally, we may let $\pi(t)$ depend deterministically on $\mu(t)$. This leads to the concept of portfolio map.

Definition 2.2.3 (Portfolio map)**.** A portfolio map is a function $F : \Delta_n \rightarrow \overline{\Delta}_n$. Given a portfolio map F, we can define a portfolio π by letting $\pi(t) = F(\mu(t))$ for all t. Abusing notation, we will also use π to denote the portfolio map itself.

As an example of portfolio map, we mention the diversity-weighted portfolio defined by

$$\pi_i(t) = \frac{\mu_i^\lambda(t)}{\sum_{j=1}^n \mu_j^\lambda(t)}, \quad i = 1, \ldots, n,$$

where $\lambda \in [0, 1]$ is a fixed parameter (see Figure 2.2). Relative to the market portfolio, the diversity-weighted portfolio overweights the small stocks and underweights the large stocks.

Note that when $\lambda = 1$ it reduces to the market portfolio, and when $\lambda = 0$ it becomes the equal-weighted portfolio $\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$. For practical applications of the diversity-weighted portfolio we refer the reader to [45] and [41, Chapter 7]. Both constant-weighted portfolios and the diversity-weighted portfolio are examples of functionally generated portfolio (see Chapter 4). Note that (see Section 4.4) it is sometimes more convenient to consider portfolio maps defined on subsets of Δ_n.

Remark 2.2.4 (Rebalancing). Most portfolios, including the constant-weighted portfolios (where π has at least two strictly positive components), require trading to maintain the desired portfolio weights. More precisely, suppose the portfolio vector is $\pi(t)$ at time t. By the consideration leading to (2.1.2), just before trading happens at time $t+1$, the proportion of capital in stock i is

$$\widetilde{\pi}_i(t+1) = \frac{\pi_i(t)(1 + R_i(t))}{\sum_{j=1}^n \pi_j(t)(1 + R_j(t))}. \tag{2.2.4}$$

The *implied weights* $\widetilde{\pi}_i(t+1)$ are sometimes called the *drifted weights* by portfolio managers. In the context of our model, rebalancing is the trading which moves the portfolio weights from $\widetilde{\pi}(t+1)$ to the new weights $\pi(t+1)$ (instead of moving from $\pi(t)$ to $\pi(t+1)$). With this terminology, a buy-and-hold portfolio – such as the market portfolio μ – is a portfolio π satisfying $\widetilde{\pi}(t+1) = \pi(t+1)$ for all t. It can be checked that π is a buy-and-hold portfolio if and only if there exist non-negative constants $c_1, \ldots, c_n$, not all zero, such that

$$\pi_i(t) = \frac{c_i \mu_i(t)}{c_1 \mu_1(t) + \cdots + c_n \mu_n(t)}$$

for all i and t (see (1.3.1)). In particular, buy-and-hold portfolios are given by portfolio maps. Buy-and-hold portfolios require no trading and thus are not subject to transaction costs.

2.3 Discussion of assumptions

Let us comment on the assumptions of the market model, several of which have already been highlighted. For our purpose, the most important equation is (2.2.3) which defines the relative value of a portfolio with respect to the market portfolio. In practice, the investment universe is constantly changing as new stocks are issued and firms may go bankrupt. Also, market capitalizations may change due to corporate actions. A capitalization-weighted market index like S&P500 is in general not a true buy-and-hold portfolio. Nevertheless, our model helps clarify the challenges of outperforming a capitalization-weighted index.

The absence of transaction costs presents a more significant problem. In fact, to the best of our knowledge there has not been any systematic study of transaction costs in the context of stochastic portfolio theory (see [41, Section 6.3] for a short discussion). For mathematical simplicity transaction cost has been ignored in this thesis.[1]

We also want to point out that in (2.2.3), it is assumed implicitly that the investor is a price taker who has no influence on stock prices. Also, in our discrete time framework, all trades are performed instantly at the beginning of each time period. Of course, this is not the case in practice. The study of market impacts and the interactions among investors fall under the field of market microstructure theory (see for example [52] and [20] for an introduction). Its relationship with stochastic portfolio theory is essentially unexplored.

2.4 Continuous time model

As mentioned in Section 1.4, stochastic portfolio theory was first developed in a continuous time setting. We choose to work in discrete time because of its simplicity and the complete absence of stochastic modeling assumptions. Nevertheless, it is sometimes more convenient to consider a continuous time model (an example is the problem of short term relative arbitrage; see [7] and [79]). For completeness, we discuss briefly the continuous time set up and refer the reader to [41, Chapter 1] and [48, Chapter 1] for more details.

In continuous time, our primary assumption is that stock prices are continuous in time. The value of a portfolio – trading continuously in time – will then be defined by an integral. A mathematically convenient and fairly general approach is to assume that the market weight

[1] Taxation and exchange rates are also absent in the present model.

$\{\mu(t)\}_{t\geq 0}$ is a continuous semimartingale defined on a given filtered probability space. In this context, a portfolio (allowing short selling) is a uniformly bounded progressively measurable process $\{\pi(t)\}_{t\geq 0}$ with values in the hyperplane $\{p = (p_1, \ldots, p_n) \in \mathbb{R}^n : p_1 + \cdots + p_n = 1\}$. The relative value process is defined as the solution of the stochastic differential equation

$$\frac{dV_\pi(t)}{V_\pi(t)} = \sum_{i=1}^{n} \pi_i(t) \frac{d\mu_i(t)}{\mu_i(t)}, \quad V_\pi(0) = 1, \tag{2.4.1}$$

which is the continuous time analogue of (2.2.3).[2] Now (1.4.2) can be derived by a routine application of Itô's lemma.

[2]Under these assumptions there exists a positive solution to (2.4.1), and this is why short selling is allowed in this context. In discrete time short selling may lead to negative portfolio value.

Chapter 3
CONSTANT-WEIGHTED PORTFOLIOS

As a prelude to what follows, in this chapter we give a treatment of constant-weighted portfolios based partly on joint work with Soumik Pal [81]. It will be shown that the results can be extended to a much larger family of portfolios called functionally generated portfolios.

3.1 A pathwise decomposition formula

Let $\pi \in \overline{\Delta}_n$ be a portfolio vector and consider the corresponding constant-weighted portfolio (see Definition 2.2.2). The portfolio will also be denoted by π. Given a market path $\{\mu(t)\}_{t=0}^{\infty} \subset \Delta_n$, we are interested in conditions under which the portfolio π outperforms the market portfolio μ. Following the treatment of [81, Section 3], we will decompose $\log V_\pi(t)$ into a sum of two terms: excess growth rate and relative entropy (Proposition 3.1.6). After discussing the theoretical implications of the decomposition in Section 3.2, we will give some empirical examples in Section 3.3.

3.1.1 Excess growth rate

We begin with the defining equation (2.2.3):

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

Taking logarithm on both sides and using the notation $\Delta A(t) := A(t+1) - A(t)$, we have

$$\begin{aligned} \Delta \log V_\pi(t) &= \log \left(\sum_{i=1}^n \pi_i \frac{\mu_i(t+1)}{\mu_i(t)} \right) \\ &= \sum_{i=1}^n \pi_i \log \frac{\mu_i(t+1)}{\mu_i(t)} + \left(\log \left(\sum_{i=1}^n \pi_i \frac{\mu_i(t+1)}{\mu_i(t)} \right) - \sum_{i=1}^n \pi_i \log \frac{\mu_i(t+1)}{\mu_i(t)} \right). \end{aligned} \tag{3.1.1}$$

We will first analyze the last term in parentheses.

Definition 3.1.1 (Excess growth rate)**.** Let $\pi = \{\pi(t)\}_{t=0}^{\infty}$ be a portfolio. The excess growth rate of π over the period $[t, t+1]$ is defined by

$$\gamma_{\pi}^{*}(t) = \log\left(\sum_{i=1}^{n} \pi_i(t) \frac{\mu_i(t+1)}{\mu_i(t)}\right) - \sum_{i=1}^{n} \pi_i(t) \log \frac{\mu_i(t+1)}{\mu_i(t)}. \tag{3.1.2}$$

The cumulative excess growth rate is defined by $\Gamma_{\pi}^{*}(t) = \sum_{s=0}^{t-1} \gamma_{\pi}^{*}(s)$.

Lemma 3.1.2. *For any portfolio π, the excess growth rate $\gamma_{\pi}^{*}(t)$ is non-negative. In particular, the cumulative excess growth rate $\Gamma_{\pi}^{*}(\cdot)$ is non-decreasing.*

Proof. This is a consequence of Jensen's inequality. Explicitly, consider a random variable $Y(t)$ which takes value $\frac{\mu_i(t+1)}{\mu_i(t)}$ with probability $\pi_i(t)$. It is easy to see that

$$\gamma_{\pi}^{*}(t) = \log \mathbb{E}_{\pi(t)} Y(t) - \mathbb{E}_{\pi(t)} \log Y(t),$$

where $\mathbb{E}_{\pi}$ is the expectation under the probability $\pi(t)$. Since log is a strictly concave function, $\gamma_{\pi}^{*}(t)$ non-negative by Jensen's inequality. Note that $\gamma_{\pi}^{*}(t)$ is strictly positive unless $Y(t)$ is almost surely constant under the probability $\pi(t)$. □

Remark 3.1.3. In terms of the stock returns $R_i(t)$, it can be shown (see [81, Lemma 3.2]) that

$$\gamma_{\pi}^{*}(t) = \log\left(1 + \sum_{i=1}^{n} \pi_i(t) R_i(t)\right) - \sum_{i=1}^{n} \pi_i(t) \log\left(1 + R_i\right).$$

In words, $\gamma_{\pi}^{*}(t)$ is the difference between the portfolio logarithmic return and the weighted average logarithmic return of the stocks (see (1.3.3). This is why $\gamma_{\pi}^{*}(t)$ is called the excess growth rate. See [81] for this numéraire invariance property as well as the chain rule.

Remark 3.1.4. By Taylor approximation, we have

$$\gamma_{\pi}^{*}(t) \approx \frac{1}{2} \sum_{i,j=1}^{n} \pi_i(t) \left(\delta_{ij} - \pi_j(t)\right) \Delta \log \mu_i(t) \Delta \log \mu_j(t) \tag{3.1.3}$$

when $\mu(t+1)$ is close to $\mu(t)$. In the continuous time limit, this becomes the excess growth rate in (1.4.3). Since we work in discrete time, for simplicity we will drop the word 'discrete'.

3.1.2 Relative entropy

Now we may write (3.1.1) in the form

$$\Delta \log V_\pi(t) = \sum_{i=1}^{n} \pi_i \log \frac{\mu_i(t+1)}{\mu_i(t)} + \gamma_\pi^*(t). \tag{3.1.4}$$

To interpret the first term on the right hand side of (3.1.4), we introduce the concept of relative entropy in information theory [27].

Definition 3.1.5 (Relative entropy)**.** For $p \in \overline{\Delta}_n$ and $q \in \Delta_n$, the relative entropy $H(p \mid q)$ is defined by

$$H(p \mid q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i},$$

with the convention $0 \log 0 = 0$.

Now we observe that

$$\begin{aligned} \sum_{i=1}^{n} \pi_i \log \frac{\mu_i(t+1)}{\mu_i(t)} &= \sum_{i=1}^{n} \pi_i \log \frac{\mu_i(t+1)}{\pi_i} - \sum_{i=1}^{n} \pi_i \log \frac{\pi_i}{\mu_i(t)} \\ &= -H(\pi \mid \mu(t+1)) + H(\pi \mid \mu(t)) \\ &= -\Delta H(\pi \mid \mu(t)). \end{aligned} \tag{3.1.5}$$

Combining (3.1.4) and (3.1.5), we obtain the decomposition

$$\Delta \log V_\pi(t) = -\Delta H(\pi \mid \mu(t)) + \gamma_\pi^*(t). \tag{3.1.6}$$

Let us summarize the above derivation in the following proposition.

Proposition 3.1.6. *Let $\pi \in \overline{\Delta}_n$ be a constant-weighted portfolio. Given a market path $\{\mu(t)\}_{t=0}^{\infty} \subset \Delta_n$, the relative value of the portfolio π satisfies the decomposition formula*

$$\log V_\pi(t) = -\left(H(\pi \mid \mu(t)) - H(\pi \mid \mu(0))\right) + \Gamma_\pi^*(t). \tag{3.1.7}$$

Proof. Sum (3.1.6) over time and recall that $\Gamma_\pi^*(\cdot)$ is the time aggregate of $\gamma_\pi^*(\cdot)$. □

In Chapter 4 we will show that a decomposition formula analogous to (3.1.7) can be derived for any functionally generated portfolio (Proposition 4.1.3). We call the general decomposition formula Fernholz's decomposition. In (3.1.7), the relative entropy term corresponds to the log generating function (Definition 4.1.1), and the excess growth rate is a special case of the L-divergence (Definition 4.1.2).

3.2 Relative arbitrage

A pathwise decomposition like (3.1.7) can be used in two ways. First, it provides a natural method of performance attribution. The relative entropy term measures how the relative value is affected by the position of the market weight vector in the unit simplex. This term is positive when $H(\pi \mid \mu(t)) < H(\pi \mid \mu(0))$, i.e., the market weight vector becomes closer to the portfolio weight. The cumulative excess growth rate is a measure of market volatility captured by the rebalancing portfolio and is always non-decreasing. The relative value of the portfolio is determined by the interplay between these two quantities. Some empirical examples will be given in Section 3.3.

Second, and more importantly, the decomposition allows us to formulate pathwise conditions under which the portfolio outperforms the market portfolio. In particular, if the market behaves in such a way that (i) the relative entropy distance $H(\pi \mid \mu(t))$ is bounded above and (ii) the cumulative excess growth rate $\Gamma_\pi^*(t)$ tends to infinity, the constant-weighted portfolio will eventually outperform the market portfolio. Here is a simple version of the existence of relative arbitrage under suitable conditions.

Proposition 3.2.1 (Relative arbitrage). *Let $\pi \in \overline{\Delta}_n$ be a constant-weighted portfolio. Let $\{\mu(t)\}_{t=0}^{\infty}$ be a market path satisfying the following conditions:*

(i) (Generalized diversity) There exists a compact set $K \subset \Delta_n$ such that $\mu(t) \subset \Delta_n$.

(ii) (Sufficient volatility) There exists $c > 0$ such that

$$\Gamma_\pi^*(t) \geq ct \tag{3.2.1}$$

for all $t \geq 0$.

Then, there exists $\epsilon \geq 0$*, depending on* K *only, such that the relative value* $V_\pi(t)$ *satisfies*

$$\log V_\pi(t) \geq -\epsilon + ct.$$

In particular, we have $V_\pi(t) > 1$ *for all* $t \geq \epsilon/c$.

Proof. The statement follows immediately from the decomposition (3.1.7). Note that the relative entropy term can be bounded below by

$$\epsilon := \sup_{p,q \in K} \left(H\left(\pi \mid p\right) - H\left(\pi \mid q\right) \right).$$

Since the function $H\left(\pi \mid \cdot\right)$ is continuous and K is compact, we have $\epsilon < \infty$. □

In Proposition 3.2.1, the generalized diversity condition (i) allows us to bound the relative entropy term; this together with sufficient volatility guarantee long term outperformance. A more sophisticated approach is to formulate conditions under which the sum of the two terms remains positive, at least with high probability. Using a specially designed cosine portfolio (see Example 4.4.2) and the stability of capital distribution over short horizons, Pal [79] formulated conditions that lead to short term relative arbitrage in large financial markets.

3.3 Empirical examples

Now we apply the decomposition (3.1.7) to empirical data.

Example 3.3.1 (Apple-Starbucks). We first look at a two-stock example considered in Section 1.3.1. The data used is the monthly return series of Apple and Starbucks from January 1994 to January 2016. The market consists of these two stocks and prices are normalized such that the initial market weight vector is $(0.5, 0.5)$. We consider the relative value of the equal-weighted portfolio $\pi = (0.5, 0.5)$.

The decomposition (3.1.7) is plotted in Figure 3.1. In this period, the equal-weighted portfolio outperforms the market portfolio significantly. Because the market is equal-weighted

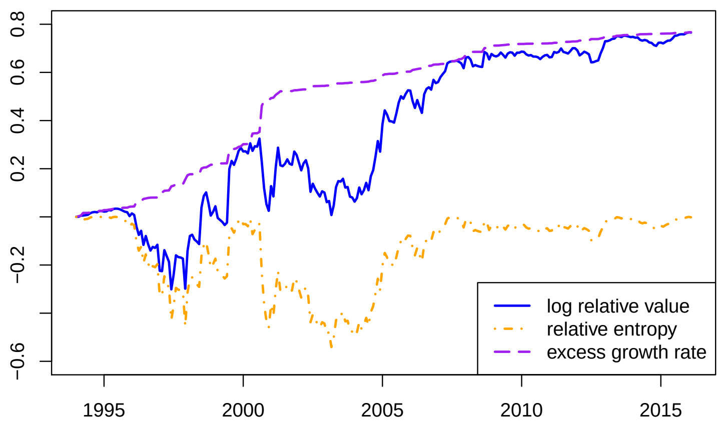

Figure 3.1: Relative value decomposition for the Apple-Starbucks example.

initially and the portfolio is equal-weighted, $q \mapsto H(\pi \mid q)$ is maximized when $q = \pi = (0.5, 0.5)$. As a result, the relative entropy term is non-positive. The relative entropy term experienced volatile fluctuations between 1997 and 2006. An important observation is that over an interval $[t_0, t_1]$ over which $H(\pi \mid \mu(t_0)) = H(\pi \mid \mu(t_1))$, the log relative return of the portfolio is positive and equals

$$\log V_\pi(t_1) - \log V_\pi(t_0) = \sum_{t=t_0}^{t_1-1} \gamma_\pi^*(t).$$

Example 3.3.2 (Emerging-market). Next we consider a more realistic example in joint work with Paul Bouchey and Vassilii Nemtchinov [14]. The dataset consists of monthly returns of S&P Global BMI country indexes of 20 emerging countries[1] from March 1997 to May 2015. Using market capitalization data in May 2015, we construct historical market weights for these countries by 'drifting the weights' back in time, based on the total returns of the country indexes. This approach excludes changes in capitalization that were due to IPOs,

[1]The countries are Argentina, Brazil, Chile, China, Colombia, Egypt, India, Indonesia, Malaysia, Mexico, Morocco, Peru, Philippines, Poland, Russia, South Africa, South Korea, Taiwan, Thailand and Turkey.

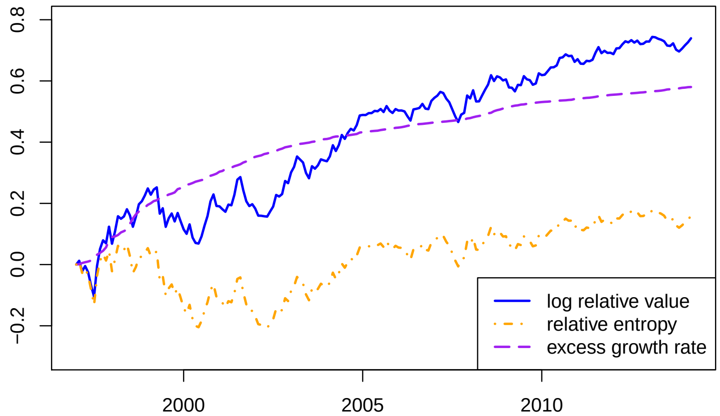

Figure 3.2: Relative value decomposition for the emerging-market example.

changes in float adjustment and stocks that left the index. By construction, all changes in market capitalization are due to returns.

In this hypothetical market of $n = 20$ assets, we again consider the relative performance of the equal-weighted portfolio. The relative value decomposition is shown in Figure 3.2. Again the equal-weighted portfolio outperforms the market portfolio. Part of this outperformance can be attributed to the relative entropy term.

3.4 The map $\pi \mapsto V_\pi(t)$

So far we have been studying the behavior of $t \mapsto V_\pi(t)$ for a fixed constant-weighted portfolio. In this section we study the behavior of $V_\pi(t)$, for t fixed, as a function of $\pi \in \overline{\Delta}_n$. This topic will be continued in Chapter 9 when we study Cover's universal portfolio.

3.4.1 A 3-stock example

Let us first look at an empirical example. Consider the monthly stock returns of Ford, Walmart and Microsoft from January 2000 to January 2016.[2] We refer to them as stocks 1, 2 and 3 in the market model. The 'market' consists of these 3 stocks and we normalize the prices so that the market weight vector is $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ at the beginning of January 2000. The market portfolio is the buy-and-hold portfolio which invests equally in each of the stocks at the beginning of January 2000.

For each constant-weighted portfolio $\pi \in \overline{\Delta}_n$, we compute the relative value $V_\pi(t)$ using (2.2.3). In Figure 3.3 we plot the map $\pi \mapsto V_\pi(t)$ at the end of January 2005 and January 2016. Since $V_\pi(0) \equiv 1$ by definition (here $t = 0$ corresponds to the beginning of January 2000), it is natural that the graph for January 2005 is close to a flat surface. In January 2016 the surface becomes more curved. The best portfolio is approximately $\pi^* = (0.17, 0.39, 0.44)$, which achieves a relative value of 1.4735, while the minimum relative value over $\overline{\Delta}_n$ is 0.4712.

We will now show that the map $\pi \mapsto V_\pi(t)$ is log-concave, i.e., $\pi \mapsto \log V_\pi(t)$ is concave, and can be approximated by a constant multiple of a Gaussian density. In fact, in continuous time the approximation becomes exact.[3]

3.4.2 Log-concavity

Given a constant-weighted portfolio $\pi \in \overline{\Delta}_n$, recall from (2.2.3) that its relative value is given by

$$\log V_\pi(t) = \sum_{s=0}^{t-1} \log \left(\sum_{i=1}^{n} \pi_i \frac{\mu_i(s+1)}{\mu_i(s)} \right). \tag{3.4.1}$$

Proposition 3.4.1 (Log-concavity). *Fix $t \geq 1$. For any market path $\{\mu(s\}_{s=0}^{t} \subset \Delta_n$, consider the relative value $V_\pi(t)$ as a function of π, where $\pi \in \overline{\Delta}_n$ ranges over all constant-weighted portfolios. Then the map $\pi \mapsto \log V_\pi(t)$ is concave in $\pi \in \overline{\Delta}_n$.*

[2] The data is obtained from Yahoo Finance. The returns are adjusted for dividends and splits.

[3] This fact has been used in [58] and [29] in the context of universal portfolio.

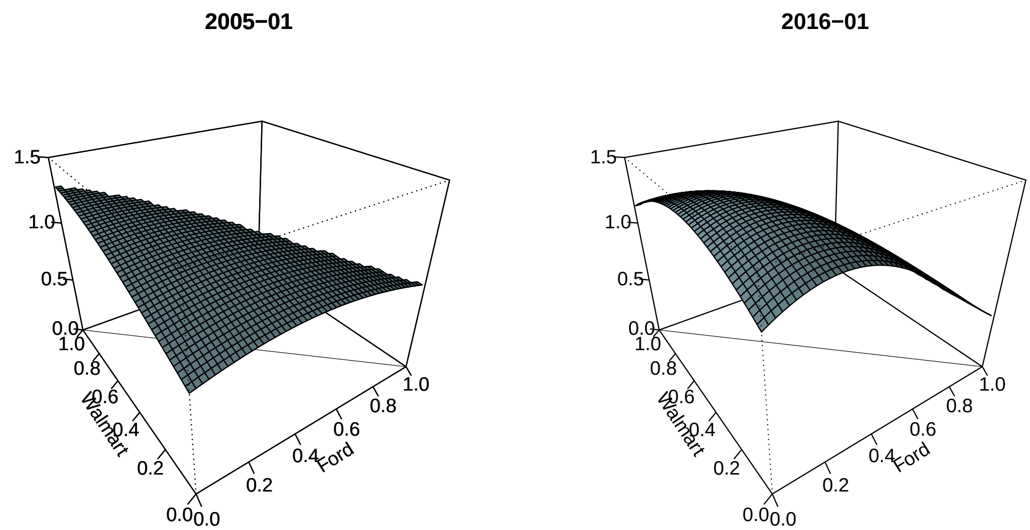

Figure 3.3: The map $\pi \mapsto V_\pi(t)$ for t = January 2005 (left) and t = January 2016 (right). Since $\pi_3 = 1 - \pi_1 - \pi_2$, the portfolio weight of stock 3 (Microsoft) is determined by those of stock 1 (Ford) and stock 2 (Walmart).

Proof. For $y \in (0, \infty)^n$ fixed, the map $\pi \in \overline{\Delta}_n \mapsto \log\left(\sum_{i=1}^n \pi_i y_i\right)$, being the composition of a linear map and an increasing concave function, is concave. Since the map $\pi \mapsto \log V_\pi(t)$ is the sum of t such maps, it is concave as well. □

3.4.3 Gaussian approximation

To obtain explicit and exact formulas we work in continuous time (see Section 2.4). Using (1.4.2), the relative value of a constant-weighted portfolio is given by

$$\log V_\pi(t) = \sum_{i=1}^n \pi_i \log \frac{\mu_i(t)}{\mu_i(0)} + \frac{1}{2} \sum_{i,j=1}^n \pi_i \left(\delta_{ij} - \pi_j\right) \langle \log \mu_i, \log \mu_j \rangle(t). \tag{3.4.2}$$

Note that (3.4.2) is valid for any $\pi \in \mathbb{R}^n$ satisfying $\pi_1 + \cdots + \pi_n = 1$. We see immediately that $\log V_\pi(t)$ is a quadratic function in π, meaning that the shape of the map $\pi \mapsto V_\pi(t)$ is Gaussian.

We are interested in the mean and covariance matrix of the density. For convenience of computation, we will adopt a coordinate system by dropping the last coordinate. Writing $Y_i(t) = \log \frac{\mu_i(t)}{\mu_n(t)} - \log \frac{\mu_i(0)}{\mu_n(0)}$ for $1 \leq i \leq n-1$ and $\Lambda_{ij}(t) = \langle Y_i, Y_j \rangle(t)$, we can write (3.4.2) in the form

$$\log V_\pi(t) = \log \frac{\mu_n(t)}{\mu_n(0)} + \sum_{i=1}^{n-1} \pi_i Y_i(t) + \frac{1}{2} \sum_{i,j=1}^{n} \pi_i \left(\delta_{ij} - \pi_j\right) \Lambda_{ij}(t).$$

Let $\theta = (\pi_1, \ldots, \pi_{n-1}) \in \mathbb{R}^{n-1}$ and write $V_\theta(t) = V_\pi(t)$. Assuming the 'relative volatility matrix' $\Lambda(t) = \left(\Lambda_{ij}(t)\right)_{1 \leq i,j \leq n-1}$ is invertible, we can complete the squares and write, in matrix notations (where all vectors are column vectors),

$$\log V_\theta(t) = \log V^*(t) - \frac{1}{2} \left(\theta - \theta^*(t)\right)' \Lambda(t) \left(\theta - \theta^*(t)\right), \quad \theta \in \mathbb{R}^{n-1}. \tag{3.4.3}$$

In (3.4.3),

$$\theta^*(t) = \Lambda^{-1}(t) \left(Y(t) + \frac{1}{2} \text{diag}(\Lambda(t)) \right)$$

and

$$V^*(t) = V_{\theta^*(t)}(t). \tag{3.4.4}$$

From (3.4.3), we see that $\theta^*(t)$ achieves the maximum relative value[4] (over the time interval $[0, t]$) over all constant-weighted portfolios. Thus

$$V^*(t) = \max_{\theta \in \mathbb{R}^{n-1}} V_\theta(t).$$

Moreover, the graph of the map $\theta \mapsto V_\theta$ is proportional to the density of the normal

$$N\left(\theta^*(t), \Lambda^{-1}(t)\right)$$

distribution. In discrete time a similar formula can be derived using second order Taylor approximation (see (3.1.3)). The graphs in Figure 3.3 are approximations of this Gaussian density restricted to the unit simplex.

Since $\Lambda(s) \leq \Lambda(t)$ whenever $s < t$, we expect that the graph of $\theta \mapsto V_\theta(t)$ becomes more and more concentrated in time.[5] This leads naturally to the topic of concentration of wealth

[4] Since $V_\theta(t) = Z_\theta(t)/Z_\mu(t)$ (see (2.2.2)), it maximizes the portfolio value as well.

[5] Note that the covariance matrix depends only on the quadratic variations. In other words, if $\Lambda(t)$ is fixed, the concentration does not depend on the value of $\mu(t)$.

in a family of portfolios. Its connection with large deviations and Cover's universal portfolio will be studied in Chapter 9.

Part I

GEOMETRY

Chapter 4
FUNCTIONALLY GENERATED PORTFOLIO

In this chapter we introduce the concept of functionally generated portfolio which is the main object of study in this thesis. Functionally generated portfolio was first introduced by Fernholz in [43] (see also [41, Chapter 3]) as a systematic method of constructing relative arbitrage opportunities under the conditions of diversity and sufficient volatility. Further generalizations are considered in [44] and [95].

In joint work with Soumik Pal [83] we showed that functionally generated portfolio has deep connections with convex analysis and optimal transport, thereby giving theoretical justifications of the concept. Moreover, we showed that functionally generated portfolios are, in a sense to be made precise, the only portfolio maps that are volatility harvesting. These results will be explained in this and the following two chapters. Throughout our development a key role will be played by multiplicative cyclical monotonicity (Definition 4.2.2).

4.1 Functionally generated portfolio

Functionally generated portfolios are a family of portfolio maps $\pi : \Delta_n \rightarrow \overline{\Delta}_n$ satisfying certain properties. There are several equivalent ways of defining functionally generated portfolio. We will first give a definition which is analogous to the approach of Chapter 3 and is easy to state.

Definition 4.1.1 (Functionally generated portfolio)**.** Let $\pi : \Delta_n \rightarrow \overline{\Delta}_n$ be a portfolio map and $\Phi : \Delta_n \rightarrow (0, \infty)$ be a concave function on Δ_n. We say that π is generated by Φ if

$$\sum_{i=1}^{n} \pi_i(p) \frac{q_i}{p_i} \geq \frac{\Phi(q)}{\Phi(p)} \tag{4.1.1}$$

for all $p, q \in \Delta_n$.[1] We call $\varphi = \log \Phi : \Delta_n \to (0, \infty)$ the log-generating function. We denote by $\mathcal{FG}$ the collection of all portfolio maps that are functionally generated.

To motivate this definition let us consider constant-weighted portfolios. Recall from (3.1.6) that the relative value of a constant-weighted portfolio $\pi \in \overline{\Delta}_n$ satisfies

$$\Delta \log V_\pi(t) = -\Delta H\left(\pi \mid \mu(t)\right) + \gamma_\pi^*(t) \geq -\Delta H\left(\pi \mid \mu(t)\right).$$

The last inequality holds because the excess growth rate $\gamma_\pi^*(t)$ is non-negative. By (2.2.3), we have

$$\Delta \log V_\pi(t) = \log \frac{V_\pi(t+1)}{V_\pi(t)} = \log \left(\sum_{i=1}^n \pi_i \frac{\mu_i(t+1)}{\mu_i(t)} \right).$$

Writing $p = \mu(t)$, $q = \mu(t+1)$ and exponentiating both sides, we have

$$\sum_{i=1}^n \pi_i \frac{q_i}{p_i} \geq \exp\left(H\left(\pi \mid p\right) - H\left(\pi \mid q\right) \right) = \frac{\Phi(q)}{\Phi(p)},$$

where

$$\Phi(p) = p_1^{\pi_1} \cdots p_n^{\pi_n} \tag{4.1.2}$$

is the geometric mean with weights $\pi_1, \ldots, \pi_n$. It follows from the definition that the constant-weighted portfolio is generated by the geometric mean (4.1.2). A way to think about functionally generated portfolio is that in the decomposition (3.1.6) the relative entropy is replaced by an arbitrary (concave) functions of the market weight vector.

The other term in the decomposition (3.1.6) is the excess growth rate (Definition 3.1.1). Its generalization in this context is called the L-divergence.

Definition 4.1.2 (L-divergence)**.** Let π be a functionally generated portfolio with log generating function φ. The L-divergence of the portfolio π is the functional $T : \Delta_n \times \Delta_n \to [0, \infty)$ defined by

$$T\left(q \mid p\right) = \log \left(\sum_{i=1}^n \pi_i(p) \frac{q_i}{p_i} \right) - \left(\varphi(q) - \varphi(p) \right), \quad p, q \in \Delta_n. \tag{4.1.3}$$

If we need to make π or φ explicit we write T_π or T_φ.

[1] It can be shown that concavity of the generating function is a consequence of the inequality (4.1.1). Fernholz's original definition (see [41, Theorem 3.1.5]) does not require concavity of Φ.

The L-divergence of π is well-defined because the log generating function φ is unique up to an additive constant (see Proposition 4.3.2(i)). When π is a constant-weighted portfolio, its L-divergence is equal to the excess growth rate. An alternative definition of L-divergence, which depends only on φ, will be given in Section 4.5.

With the above definitions, we can write down immediately a decomposition formula for a functionally generated portfolio analogous to that of a constant-weighted portfolio.[2] We call it Fernholz's decomposition to acknowledge his fundamental work.

Proposition 4.1.3 (Fernholz's decomposition)**.** *Let π be a functionally generated portfolio with log generating function φ. For any market path, the relative value of π satisfies the decomposition*

$$\log V_\pi(t) = (\varphi(\mu(t)) - \varphi(\mu(0))) + \sum_{s=0}^{t-1} T\left(\mu(t+1) \mid \mu(t)\right). \tag{4.1.4}$$

We write $A(t) = \sum_{s=0}^{t-1} T\left(\mu(t+1) \mid \mu(t)\right)$ and call it the drift process.

In particular, suppose K is a subset of Δ_n and the market path satisfies the following conditions:

(i) The generating function Φ is bounded below from zero on K.

(ii) (Generalized diversity) $\mu(t) \in K$ for all t.

(iii) (Sufficient volatility) $A(t) \uparrow \infty$ as $t \uparrow \infty$.

Then the relative value of π satisfies $\lim_{t\to\infty} V_\pi(t) = \infty$.

Proof. The decomposition follows immediately from (4.1.1) and Definition 4.1.2. The second statement can be proved using the argument of Prosposition 3.2.1. □

In Table 4.1 we give several examples of functionally generated portfolio. See [41, Chapter 3] for more examples. The formulas of the portfolio weights can be verified using Proposition

[2] Compare our approach with the proof of [41, Theorem 3.1.5].

Name	$\pi_i(p)$	$\Phi(p)$
Market	p_i	1
Buy-and-hold	$\frac{c_i p_i}{\sum_{j=1}^n c_j p_j}$	$\sum_{j=1}^n c_j p_j$
Diversity-weighted	$\frac{p_i^\lambda}{\sum_{j=1}^n p_j^\lambda}$	$\left(\sum_{j=1}^n p_j^\lambda\right)^{\frac{1}{\lambda}}$
Equal-weighted	$\frac{1}{n}$	$(p_1 p_2 \cdots p_n)^{\frac{1}{n}}$
Constant-weighted	π_i	$p_1^{\pi_1} \cdots p_n{}^{\pi_n}$
Entropy-weighted	$\frac{-p_i \log p_i}{\sum_{j=1}^n -p_j \log p_j}$	$\sum_{j=1}^n -p_j \log p_j$

Table 4.1: Examples of functionally generated portfolios

4.3.1 below. It is important to note that the market portfolio is generated by a constant function which is flat. More generally, buy-and-hold portfolios are generated by linear functions with non-negative coefficients. These functions have zero curvature which explains the lack of rebalancing. The diversity-weighted portfolio has been introduced in Section 2.2. Together with the entropy and constant-weighted portfolios, these are some of the first functionally generated portfolios studied in stochastic portfolio theory.

4.2 Multiplicative cyclical monotonicity

In this section we give an equivalent – yet more fundamental – characterization of functionally generated portfolio. The main mathematical tools are convex analytic and we begin by reviewing some basic definitions.

4.2.1 Preliminaries in convex analysis

For our purpose we will specialize our treatment to 'concave' analysis on the unit simplex Δ_n. For more background in convex analysis we refer the reader to [87] which is the standard reference of the subject. For now we will require little more than the definitions of concavity and superdifferential. Additional tools will be introduced as needed.

Let $f : \Delta_n \to \mathbb{R}$. It is said to be concave if

$$f(\lambda p + (1 - \lambda)q) \geq \lambda f(p) + (1 - \lambda)f(q) \tag{4.2.1}$$

for all $p, q \in \Delta_n$ and $\lambda \in [0, 1]$. Concave functions enjoy many nice properties. For example, if $f : \Delta \to \mathbb{R}$ is concave, it is automatically continuous. Moreover, it is Lipschitz on any compact subset of Δ_n. By Rademacher's theorem, this implies that f is almost everywhere differentiable on Δ_n.

Next we will define the superdifferential which is a generalization of derivative. By a tangent vector of Δ_n we mean a vector $v \in \mathbb{R}^n$ satisfying $\sum_{i=1}^n v_i = 0$, i.e., v is parallel to Δ_n. We denote by $T\Delta_n$ the vector space of tangent vectors of Δ_n.

Definition 4.2.1 (Superdifferential)**.** Let $f : \Delta_n \to \mathbb{R}$ be concave and $p \in \Delta_n$. A supergradient of f at p is a tangent vector $v \in T\Delta_n$ such that

$$f(p) + v \cdot (q - p) \geq f(q) \tag{4.2.2}$$

for all $q \in \Delta_n$. The superdifferential $\partial f(p)$ of f at p is the set of supergradients of f at p.

Geometrically, $\partial f(p)$ corresponds to the collection of supporting hyperplanes of the graph of f at the point $(p, f(p))$. If f is concave, then $\partial f(p)$ is a non-empty compact convex subset of $T\Delta^n$ (equipped with the usual topology).

4.2.2 Multiplicative cyclical monotonicity

Let $\pi : \Delta_n \to \overline{\Delta_n}$ be a portfolio map. Recall this means that the portfolio vector at time t is $\pi(\mu(t))$ where $\mu(t)$ is the market weight vector. We want to capture the idea that π is volatility harvesting.

A simple situation is the following. Suppose there exists $m \geq 0$ such that the market path $\{\mu(t)\}_{t=0}^{\infty}$ satisfies $\mu(t + (m + 1)) = \mu(t)$ for all t. In other words, the market path is

$(m+1)$-periodic. If π is volatility harvesting, we expect that π does not suffer in this market. Over each cycle of length $m+1$, the log relative return of the portfolio π is

$$\omega := \prod_{t=0}^{m} \left(\pi(\mu(t)) \cdot \frac{\mu(t+1)}{\mu(t)} \right).$$

Iterating, the relative value at time $k(m+1)$ is $V_\pi(k(m+1)) = \omega^k$. If $\omega < 1$, the relative value converges to 0 as $t \to \infty$. In order that the relative value does not decay to 0, we require that $\omega \geq 1$ over the cycle.

The above discussion leads to the following definition.

Definition 4.2.2 (Multiplicative cyclical monotonicity (MCM))**.** By a cycle in the unit simplex we mean a finite sequence $\{\mu(t)\}_{t=0}^{m+1} \subset \Delta_n$ satisfying $\mu(m+1) = \mu(0)$. Let $\pi : \Delta_n \to \overline{\Delta}_n$ be a portfolio map. We say that π is multiplicatively cyclical monotone if

$$\prod_{t=0}^{m} \left(\pi(\mu(t)) \cdot \frac{\mu(t+1)}{\mu(t)} \right) \geq 1 \tag{4.2.3}$$

for all cycles $\{\mu(t)\}_{t=0}^{m+1}$ in Δ_n.

Note that (4.2.3) may be written in the form

$$\sum_{t=0}^{m} \log \left(1 + \frac{\pi(\mu(t))}{\mu(t)} \cdot (\mu(t+1) - \mu(t)) \right) \geq 1. \tag{4.2.4}$$

Remark 4.2.3. The MCM property is a multiplicative version of the classicical notion of cyclical monotonicity in convex analysis [87, Section 24]. For completeness and motivation let us review this concept briefly. Let ρ be a multivalued map from $\mathbb{R}^n$ to $\mathbb{R}^n$, i.e., $\rho(x)$ is a non-empty subset of $\mathbb{R}^n$ for each $x \in \mathbb{R}^n$. We say that ρ is cyclically monotone if for any cycle $\{x(t)\}_{t=0}^{m+1}$ and any sequence $\{x^*(t)\}_{t=0}^{m+1}$ such that $x^*(t) \in \rho(x(t))$ for all t, we have

$$\sum_{t=0}^{m} x^*(t) \cdot (x(t+1) - x(t)) \leq 0. \tag{4.2.5}$$

Note the similarity between (4.2.5) and (4.2.4).[3] Here is the main result concerning cyclical monotonicity.

[3]The signs are different because (4.2.5) is for convex functions, wheras (4.2.4) is for concave functions.

Theorem 4.2.4. *[87, Theorem 24.8] Let ρ be a multivalued map from $\mathbb{R}^n$ to $\mathbb{R}^n$. Then there exists a closed proper convex function f on $\mathbb{R}^n$ such that $\rho(x) \subset \partial f(x)$ if and only if ρ is cyclically monotone.*

The main result of this section is the following analogue of Theorem 4.2.4.

Theorem 4.2.5. *Let $\pi : \Delta_n \to \overline{\Delta}_n$ be a portfolio map. Then π is multiplicatively cyclical monotone if and only if it is functionally generated.*

Proof. Our proof is an adaption of the proof of [87, Theorem 24.8]. First suppose that π is generated by a concave function $\Phi : \Delta_n \to (0, \infty)$. Using (4.1.1), for any cycle $\{\mu(t)\}_{t=0}^{m+1}$ in Δ_n we have

$$\prod_{t=0}^{m} \left(\pi(\mu(t)) \cdot \frac{\mu(t+1)}{\mu(t)} \right) \geq \prod_{t=0}^{m} \frac{\Phi(\mu(t+1))}{\Phi(\mu(t))} = 1.$$

This shows that π is multiplicatively cyclical monotone.

Conversely, suppose that π is multiplicatively cyclical monotone. We will construct a function $\Phi : \Delta_n \to (0, \infty)$ such that (4.1.1) holds. Fix an arbitrary point $\mu(0) \in \Delta_n$ and define Φ on Δ_n by

$$\Phi(p) = \inf \prod_{t=0}^{m} \left(\pi(\mu(t)) \cdot \frac{\mu(t+1)}{\mu(t)} \right), \tag{4.2.6}$$

where the infimum is taken over all $m \geq 0$ and all sequences $\{\mu(t)\}_{t=0}^{m+1}$ in Δ_n with $\mu(0)$ fixed and $\mu(m+1) = p$. Financially, $\Phi(p)$ is the greatest lower bound of the relative value of π over a finite market path from $\mu(0)$ to p.

We claim that π is generated by the function Φ. First we note that Φ, being the pointwise infimum of a family of non-negative affine functions on Δ_n, non-negative and concave on Δ_n. We also observe that $\Phi(\mu(0)) = 1$ by the MCM property (put $m = 0$). It follows by concavity and non-negativity that Φ is everywhere positive on Δ_n.

It remains to establish (4.1.1). Let $p, q \in \Delta_n$ be given, and let $\alpha > \Phi(p)$. By definition of Φ, there exists $m \geq 0$ and a sequence $\{\mu(t)\}_{t=0}^{m+1}$ with $\mu(m+1) = p$ such that

$$\prod_{t=0}^{m} \left(\pi(\mu(t)) \cdot \frac{\mu(t+1)}{\mu(t)} \right) < \alpha.$$

Setting $\mu(m+2) = q$ and applying the definition (4.2.6) onces more, we have

$$\Phi(q) \leq \left(\pi(p) \cdot \frac{q}{p}\right) \alpha.$$

The proof is completed by letting $\alpha \downarrow \Phi(p)$. □

4.3 Basic properties

In this section we establish several useful properties of functionally generated portfolios. First we note that if $\Phi : \Delta_n \to (0, \infty)$ is concave, then $\varphi = \log \Phi$ is also concave on Δ_n and

$$\partial\varphi(p) = \frac{1}{\Phi(p)} \partial\Phi(p) = \left\{ \frac{1}{\Phi(p)} v : v \in \partial\Phi(p) \right\}.$$

Being the logarithm of a concave function, φ is said to be exponentially concave. For $1 \leq i \leq n$, let $e(i) = (0, \ldots, 1, \ldots, 0)$ be the vertex of Δ_n in the p_i-direction.

Proposition 4.3.1. *Let $\Phi : \Delta_n \to (0, \infty)$ be concave, and let $\varphi = \log \Phi$.*

(i) Suppose the portfolio map $\pi : \Delta_n \to \overline{\Delta}_n$ is generated by Φ. For $p \in \Delta_n$, the tangent vector $v = (v_1, \ldots, v_n)$ defined by

$$v_i = \frac{\pi_i(p)}{p_i} - \frac{1}{n} \sum_{j=1}^{n} \frac{\pi_j(p)}{p_j} \tag{4.3.1}$$

is an element of $\partial\varphi(p)$.

(ii) Conversely, if $v \in \partial\varphi(p)$ is a supergradient of φ at p, the vector $\pi = (\pi_1, \ldots, \pi_n)$ defined by

$$\frac{\pi_i}{p_i} = v_i + 1 - \sum_{j=1}^{n} p_j v_j, \quad i = 1, \ldots, n \tag{4.3.2}$$

is an element of $\overline{\Delta}_n$. In particular, any selection of $\partial\varphi$ (a map $v : \Delta_n \to T\Delta_n$ satisfying $v(p) \in \partial\varphi(p)$ for all p) defines via (4.3.2) a portfolio generated by Φ.

Finally, the operations $\pi \mapsto v$ and $v \mapsto \pi$ defined by (4.3.1) and (4.3.2) are inverses of each other.

Proof. (i) Let $p \in \Delta_n$. By (4.1.1), we have

$$1 + \frac{\pi(p)}{p} \cdot (q - p) \geq \frac{\Phi(q)}{\Phi(p)}$$

for all $q \in \Delta_n$. Note that $\frac{\pi(p)}{p}$ is not a tangent vector of Δ_n. The normalization (4.3.1) projects it to v which is a tangent vector. Since $\frac{\pi(p)}{p} - v$ is perpendicular to $T\Delta_n$, the inner product does not change if $\frac{\pi(p)}{p}$ is replaced by v. It follows that $v \in \partial\varphi(p)$.

(ii) It is easy to verify that $\sum_{i=1}^n \pi_i = 1$. To see that $\pi_i \geq 0$ for each i, consider the point $q = p + t(e(i) - p)$ where $t \in [0, 1)$. Since $\Phi(p)v \in \partial\Phi(p)$, we have

$$\begin{aligned} -\Phi(p) &\leq \Phi(p + t(e(i) - p)) - \Phi(p) \quad \text{(since } \Phi(q) > 0) \\ &\leq \Phi(p)v \cdot t(e(i) - p) \\ &= t\Phi(p)\left(v_i - \sum_{j=1}^n p_j v_j\right). \end{aligned}$$

Letting $t \uparrow 1$ and dividing both sides by $\Phi(p)$, we get the desired inequality $\pi_i \geq 0$.

That $\pi \mapsto v$ and $v \mapsto \pi$ are inverses of each other can be verified by a direct computation.

□

Proposition 4.3.2. *Let π be a portfolio map generated by a concave function Φ on Δ_n, and let $\varphi = \log \Phi$.*

(i) The generating function Φ is unique up to a positive multiplicative constant.

(ii) For $p \in \Delta_n$ and $i = 1, \ldots, n$ we have

$$1 + D_{e(i)-p}\varphi(p) \leq \frac{\pi_i(p)}{p_i} \leq 1 - D_{p-e(i)}\varphi(p).$$

Here $D_v f(p)$ is the directional derivative of f in the direction v at p. In particular, if Φ is differentiable, the portfolio map is given by the formula

$$\pi_i(p) = p_i\left(1 + D_{e(i)-p}\varphi(p)\right), \quad i = 1, \ldots, n. \tag{4.3.3}$$

(iii) If π is continuous, then Φ is continuously differentiable. More generally, if π is of class C^k, then Φ is of class C^{k+1}.

Proof. (i) Suppose π is generated by both Φ_1 and Φ_2. Let $p, q \in \Delta_n$ and consider the line segment ℓ from p to q. Consider the restrictions of $\log \Phi_i$ to ℓ, denoted by $\log \Phi_i|_\ell$. They can be parameterized as one-dimensional concave functions. In particular, they are differentiable on ℓ except at most for countably many points on ℓ. By Proposition 4.3.1, the vector $\frac{\pi(p)}{p}$ defines a supporting hyperplane of the log generating function. It follows that $\log \Phi_1$ and $\log \Phi_2$ have parallel supporting hyperplanes at all points of Δ_n. In particular, the derivatives of $\log \Phi_i|_1$ and $\log \Phi_i|_2$ agree almost everywhere on ℓ. By the fundamental theorem of calculus for concave functions (see [87, Corollary 24.2.1]), we have

$$\log \Phi_1(q) - \log \Phi_1(p) = \log \Phi_2(q) - \log \Phi_2(p).$$

Since p and q are arbitrary, Φ_2/Φ_1 is a positive constant.

(ii) By definition of π, for $h \in \mathbb{R} \setminus \{0\}$ small enough such that $p + h(e(i) - p) \in \Delta_n$, the superdifferential inequality (4.1.1) gives

$$1 + \frac{\pi(p)}{p} \cdot h(e(i) - p) \geq \frac{\Phi(p + h(e(i) - p))}{\Phi(p)}. \tag{4.3.4}$$

Note that the inner product is given by

$$\frac{\pi(p)}{p} \cdot h(e(i) - p) = h\left(\frac{\pi_i(p)}{p_i} - 1\right).$$

Taking logarithm on both sides of (4.3.4), we have

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

Dividing by h and taking the limits as $h \downarrow 0$ and $h \uparrow 0$, we obtain the desired inequalities. The next statement is proved by noting that if Φ (and hence $\varphi = \log \Phi$) is differentiable, for every tangent vector v we have $D_v\varphi(p) = -D_{-v}\varphi(p)$.

(iii) Suppose the map π is continuous. By Proposition 4.3.1, π defines a continuous a continuous selection of the superdifferential of φ. By [86, Proposition 4] φ, and hence Φ, is differentiable on Δ_n. By [87, Corollary 25.5.1], Φ is actually continuously differentiable.

It is easy to see that if π is of class C^k then Φ is of class C^{k+1}. $\square$

Proposition 4.3.3 (Convexity)**.** *The set $\mathcal{FG}$ of functionally generated portfolios is convex. Indeed, suppose $\pi^{(1)}$ and $\pi^{(2)}$ are functionally generated portfolios with generating functions $\Phi^{(1)}$ and $\Phi^{(2)}$. For λ, the portfolio map*

$$\pi = \lambda\pi^{(1)} + (1-\lambda)\pi^{(2)} \tag{4.3.5}$$

is functionally generated, and a generating function is the geometric mean

$$\Phi = \left(\Phi^{(1)}\right)^{\lambda}\left(\Phi^{(2)}\right)^{1-\lambda}.$$

Proof. We need to show that the portfolio π defined by (4.3.5) is generated by the geometric mean Φ. For any p, q, we have

$$\pi^{(i)}(p)\cdot\frac{q}{p} \geq \frac{\Phi^{(i)}(q)}{\Phi^{(i)}(p)}, \quad i = 1, 2.$$

By the AM-GM inequality, we have

$$\pi(p)\cdot\frac{q}{p} \geq \lambda\frac{\Phi^{(1}(q)}{\Phi^{(1)}(p)} + (1-\lambda)\frac{\Phi^{(2)}(q)}{\Phi^{(2)}(p)} \geq \frac{\Phi(q)}{\Phi(p)}.$$

$\square$

4.4 Functionally generated portfolios on subsets of Δ_n

Let Ω be a subset of Δ_n. Instead of requiring a portfolio map to be defined on all of Δ_n, we may consider portfolio maps defined only on Ω. The idea is that the portfolio is only used when the market weight remains in Ω. When the market weight exits Ω, one chooses another portfolio map. This provides more flexibility for investment purposes.

Note that the proof of Theorem 4.2.5 does not require π to be defined on the whole of Δ_n. Thus we have the following extension.

Corollary 4.4.1. *Let $\Omega \subset \Delta_n$ be any (non-empty) subset, and let $\pi : \Omega \to \overline{\Delta}_n$. If π satisfies the MCM property for cycles in Ω, there exists a concave function $\Phi : \Delta_n \to (0, \infty)$ such that* (4.1.1) *holds for any $p \in \Omega$ and $q \in \Delta_n$. In particular, π can be extended to Δ_n as a portfolio generated by Φ.*

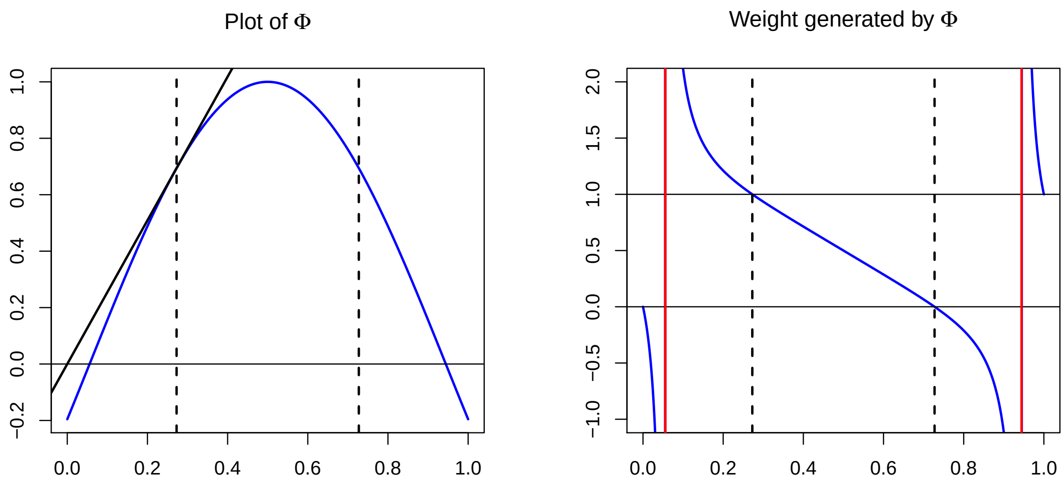

Figure 4.1: Cosine portfolio for $n = 2$ and $k = 2.5$. Left: Plot of $p_1 \mapsto \Phi(p)$. Right: Plot of $p_1 \mapsto \pi_1(p)$ using the formula $\pi_i(p) = p_i(1 + D_{e(i)-p} \log \Phi(p))$. The dotted vertical lines show the endpoints of Ω_0. The red vertical lines show the endpoints of Ω.

Proof. The proof of Theorem 4.2.5 yields a positive concave function Φ on Δ_n such that (4.1.1) holds for any $p \in \Omega$ and $q \in \Delta_n$. For $p \notin \Omega$, let $\pi(p)$ be given by (4.3.2) where v is any element of $\partial \log \Phi(p)$. Then the extended portfolio map is generated by Φ. □

By Corollary 4.4.1, an MCM portfolio map on Ω can be extended to an MCM portfolio on Δ_n. Thus, there is no loss of generality if we restrict to functionally generated portfolios on Δ_n. Nevertheless, it is sometimes more convenient to define a functionally generated portfolio locally.

Example 4.4.2 (Cosine portfolio). Let $p_0 \in \Delta_n$ be fixed and $k > 0$. Consider the function

$$\Phi(p) = \cos\left(k\|p - p_0\|\right)$$

where $\|\cdot\|$ is the Euclidean norm. For $p \in \Omega$ where

$$\Omega = \left\{p \in \Delta_n : \|p - p_0\| < \frac{\pi}{2k}\right\},$$

the function Φ is positive and concave. This function is used in [79] to construct short term relative arbitrage in large markets.

Note that Φ does not generate a $\overline{\Delta}_n$-valued portfolio map on the whole of Ω (see Figure 4.1 where $n = 2$ and $k = 2.5$). The maximal domain Ω_0 of π (for which it is $\overline{\Delta}_n$-valued and MCM) is approximately $0.2726 \leq x_1 \leq 0.7273$. Outside Ω_0, Φ can be extended to a positive concave function on Δ_2 by letting it be affine outside Ω_0 (see the black tangent line on the left). In this context, we may say that π is generated by Φ on Ω_0.

4.5 L-divergence

The L-divergence of a functionally generated portfolio has been defined in Definition 4.1.2. Letting $\varphi = \log \Phi$ be the log generating function, we have

$$T\left(q \mid p\right) = \log\left(\pi(p) \cdot \frac{q}{p}\right) - (\varphi(q) - \varphi(p))$$

for all p, q. For simplicity we restrict to differentiable generating functions in this section. Since $\frac{\pi(p)}{p}$ is essentially the gradient of φ at p (see Proposition 4.3.1), we may express $T\left(q \mid p\right)$ solely in terms of φ.

Proposition 4.5.1. *Let π be a functionally generated portfolio with a differentiable log generating function φ. Then, for any p, q we have*

$$T\left(q \mid p\right) = \log\left(1 + \nabla\varphi(p) \cdot (q - p)\right) - (\varphi(q) - \varphi(p)). \tag{4.5.1}$$

For this reason, we may call $T\left(\cdot \mid \cdot\right)$ the L-divergence of the differentiable exponentially concave function φ.

Proof. This is an immediate consequence of Proposition 4.3.1. □

The L-divergence should be distinguished from the Bregman divergence of φ defined by

$$D[q : p] = \nabla\varphi(p) \cdot (q - p) - \left(\varphi(q) - \varphi(p)\right). \tag{4.5.2}$$

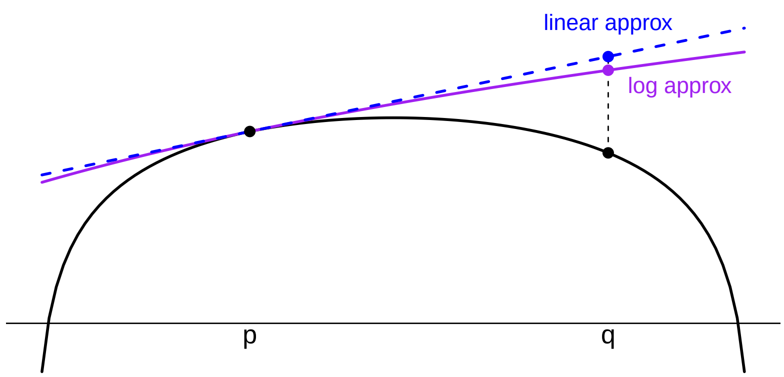

Figure 4.2: Bregman divergence and L-divergence. The Bregman divergence is defined using the linear approximation, and the L-divergence is defined using the logarithmic approximation.

Bregman divergence was introduced by Bregman in [16] and is widely applied in statistics and optimization For example, the relative entropy

$$H(q \mid p) = \sum_{i=1}^{n} q_i \log \frac{q_i}{p_i}$$

is the Bregman divergence of the Shannon entropy $\varphi(p) = -\sum_{i=1}^{n} p_i \log p_i$. While L-divergence appears to be a variant of Bregman divergence with a logarithmic correction, the logarithm makes the two objects fundamentally different. Indeed, the Bregman divergence is non-negative as soon as φ is concave. Since φ is the logarithm of a concave function, its extra concavity cannot be captured by the usual Bregman divergence (see Figure 4.2). In Chapter 10 we will show that the L-divergence induces a remarkable geometric structure on the simplex Δ_n.'

Chapter 5

PSEUDO-ARBITRAGE

In Chapter 4 we claimed that functionally generated portfolios are the only portfolio maps that are relative arbitrage opportunities under (only) the conditions of diversity and sufficient volatility. In this chapter we make this statement rigorous using the concept of pseudo-arbitrage introduced in [83].

5.1 Pseudo-arbitrage

Let π be a portfolio map. We want to capture the idea that π is a relative arbitrage with respect to the market whenever it is diverse and sufficiently volatile. Recall in Definition 2.1.3 we defined diversity as a property of market paths:

$$\sup_{t \geq 0} \max_{1 \leq i \leq n} \mu_i(t) \leq 1 - \delta. \tag{5.1.1}$$

Alternatively, (5.1.1) is equivalent to $\mu(t) \in K$ for all t where K is the set $\{p \in \Delta_n : \max_{1 \leq i \leq n} p_i \leq 1 - \delta\}$. In this chapter we consider a generalized diversity condition where K can be any subset of Δ_n. This leads us to consider market paths with values in K.

Definition 5.1.1 (Pseudo-arbitrage)**.** Let K be a subset of Δ_n. A portfolio map π is said to be a pseudo-arbitrage on K (with respect to the market portfolio) if the following properties hold.

(i) There exists a constant $C = C(K, \pi) \geq 0$ such that for all market paths $\{\mu(t)\}_{t=0}^{\infty}$ taking values in K, we have $\inf_{t \geq 0} \log V_\pi(t) \geq -C$.

(ii) There exists a market path $\{\mu(t)\}_{t=0}^{\infty} \subset K$ along which $\lim_{t \to \infty} V_\pi(t) = \infty$.

Definition 5.1.1 formalizes some necessary requirements in order that a given portfolio map is guarenteed to outperform the market under the generalized diversity condition $\mu(t) \in K$ and sufficient volatility. First, (i) requires that the portfolio is never allowed to lose more than a fixed amount. That is, the downside risk is uniformly bounded below regardless of the market movement in a fixed region. Intuitively, this is because if (i) fails, the unfavorable market movement may repeat again and again, causing the relative value to tend to zero. Along this market path the portfolio has no hope of beating the market. Property (ii) essentially says that the portfolio is not a buy-and-hold portfolio (which satisfies (i)).

Here are the main results of this chapter. First, we show that a pseudo-arbitrage opportunity is functionally generated.

Theorem 5.1.2. *Let K be an open convex subset of Δ_n and let $\pi : K \to \overline{\Delta}_n$ be a portfolio map on K. Then π is a pseudo-arbitrage on K if and only if the following properties hold:*

(i) π can be extended to Δ_n as a portfolio map generated by a positive concave function Φ on Δ_n.

(ii) The function Φ is not affine on K, or equivalently the L-divergence $T(q \mid p)$ is not identically zero for $p, q \in K$.

(iii) There exists $\epsilon > 0$ such that $\inf_{p \in K} \Phi(p) \geq \epsilon$.

In Theorem 4.2.5 we showed that a portfolio map is functionally generated if and only if it is multiplicatively cyclical monotone. The next result complements this and shows that failing to be MCM is a local property. Here we use $\|\cdot\|$ to denote the Euclidean norm.

Theorem 5.1.3. *Let $\pi : \Delta_n \to \overline{\Delta}_n$ be a portfolio map which does not satisfy the MCM property.*

(i) For any $\delta > 0$, there is a market path $\{\mu(t)\}_{t=0}^{\infty}$ such that $\|\mu(t+1) - \mu(t)\| < \delta$ for all t and $\lim_{t \to \infty} V_\pi(t) = 0$. Thus π cannot be a pseudo-arbitrage over any set containing this path.

(ii) For any $\delta > 0$, there exists $p \in \Delta_n$ such that the sequence in (i) can be chosen to lie entirely within the Euclidean ball centered at p with radius δ.

The proofs of Theorems 5.1.2 and 5.1.3 will be given in the next two sections.

5.2 Proof of Theorem 5.1.2

First we suppose that conditions (i), (ii) and (iii) hold. Let $\{\mu(t)\}_{t=0}^{\infty}$ be any market path in K. By (i), π is functionally generated. By Fernholz's decomposition (Proposition 4.1.3), we have

$$\log V_{\pi}(t) = \log \frac{\Phi(\mu(t))}{\Phi(\mu(0))} + \sum_{s=0}^{t-1} T\left(\mu(s+1) \mid s\right) \tag{5.2.1}$$

for all t.

Since $\mu(t) \in K$ for all t, by condition (iii) we have

$$\inf_{t \geq 0} \log \frac{\Phi(\mu(t))}{\Phi(\mu(0))} \geq \inf_{p \in K} \left(\log \Phi(p) - \log \Phi(\mu(0))\right) > -\infty.$$

Also the drift process $A(t) = \sum_{s=0}^{t-1} T\left(\mu(s+1) \mid s\right)$ is non-decreasing in t. It follows that (i) of Definition 5.1.1 holds.

By condition (ii), there exists $p, q \in K$ such that $T\left(q \mid p\right) > 0$. Now let $\{\mu(t)\}_{t=0}^{\infty}$ be any market path in K such that $(\mu(t), \mu(t+1)) = (p, q)$ for infinitely many t. Then the drift process tends to infinity as $t \to \infty$, and it follows that $\lim_{t \to \infty} V_{\pi}(t) = \infty$. Hence π is a pseudo-arbitrage over the set K.

Conversely, suppose that π is a pseudo-arbitrage on K. Then π must satisfy the MCM property on K. Thus π is generated by a concave function Φ on K.

Since π is functionally generated, we may apply Fernholz's decomposition. Consider the right hand side of (5.2.1). If the L-divergence $T\left(\cdot \mid \cdot\right)$ vanishes on $K \times K$, the drift process is identically zero. On the other hand, since a non-negative concave function on Δ_n is bounded above, the first term on the right hand side of (5.2.1) is bounded above. It follows that $\sup_{t \geq 0} V_{\pi}(t) < \infty$ for all market paths. This violates (ii) of Definition 5.1.1. Thus $T\left(\cdot \mid \cdot\right)$

does not vanish on $K \times K$. Since K is (relatively) open in Δ_n, this implies that the restriction of Φ to K is not affine.

It remains to show that Φ is bounded away from 0 on K. We proceed by contradiction. Suppose zero is a limit point of $\Phi(K)$. As a positive concave function on Δ_n, Φ can be extended continuously to $\overline{\Delta}_n$. We can thus find a point q in $\overline{K}$ such that $\Phi(q) = 0$. Fix a point $p \in K$ and let $\{\lambda(t)\}_{t=0}^{\infty}$ be a strictly increasing sequence in $[0, 1)$ converging to 1. Let $\{\mu(t)\}_{t=0}^{\infty}$ be the market path in K defined by

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

Since K is convex, we have $[p, q) \subset K$. We choose $\lambda(t)$ such that $\log \Phi$ is differentiable at $\mu(t)$ for all t (this is possible sense $\log \Phi$ is not differentiable for at most countably many points on the segment $[p, q]$). Since $\log \Phi$ is differentiable at $\mu(t)$, we have

$$\frac{\pi(\mu(t))}{\mu(t)} \cdot (\mu(t+1) - \mu(t)) = \nabla \log \Phi(\mu(t)) \cdot (\mu(t+1) - \mu(t)).$$

Using the elementary inequality $\log(1 + x) \leq x$ for $x > -1$, we get

$$\sum_{t=0}^{\infty} \log \left(\frac{\pi(\mu(t))}{\mu(t)} \cdot (\mu(t+1) - \mu(t)) \right) \leq \sum_{t=0}^{\infty} \nabla \log \Phi(\mu(t)) \cdot (\mu(t+1) - \mu(t)). \qquad (5.2.2)$$

Comparing the right hand side of (5.2.2) with the line integral

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

we may choose $\{\lambda(t)\}_{t=0}^{\infty}$ such that the right hand side of (5.2.2) is $-\infty$. Thus, along this market path the relative value $V_\pi(t)$ tends to zero as $t \to \infty$. This shows that π cannot be a pseudo-arbitrage if zero is a limit point of $\Phi(K)$. This completes the proof of Theorem 5.1.2.

5.3 *Proof of Theorem* 5.1.3

To prove (i) we need the following definition.

Definition 5.3.1 (δ-MCM)**.** Let $\delta > 0$. A portfolio map $\pi : \Delta_n \to \overline{\Delta}_n$ satisfies δ-MCM if the inequality (4.2.3) holds for all market cycles where the successive jump sizes $\|\mu(t+1) - \mu(t)\|$ are all less than δ.

Lemma 5.3.2. *Let $\pi : \Delta_n \to \overline{\Delta}_n$ be a portfolio map. Then π is MCM if and only if π is δ-MCM for all $\delta > 0$.*

Proof. It is clear that if π is MCM then π is δ-MCM for all $\delta > 0$. To prove the converse, suppose π is δ-MCM for all $\delta > 0$. The idea is to repeat the proof of Theorem 4.2.5 with the additional restriction that the jumps have sizes less than δ. Consider the function Φ defined by (4.2.6), where now the infimum is taken over all $m \geq 0$ and all choices of cycles $\{\mu(t)\}_{t=0}^{m+1}$ where $\|\mu(t+1) - \mu(t)\| < \delta$ for all t. Following the proof of Theorem 4.2.5, we see that Φ is a positive concave function on Δ_n and

$$\Phi(p) + \Phi(p)\frac{\pi(p)}{p} \cdot (q - p) \geq \Phi(p), \quad \|p - q\| < \delta. \tag{5.3.1}$$

This shows that the component of $\Phi(p)\frac{\pi(p)}{p}$ parallel to Δ_n (which is a tangent vector) is a supergradient of the restricted concave function $\Phi|_V$ at p, where V is a convex neighborhood of $p \in \Delta_n$. However, by [87, Theorem 23.2] we have

$$\partial\Phi(p) = \left\{\xi \in T\Delta_n : D_v\Phi(p) \leq \xi \cdot v, \text{ for all } v \in T\Delta_n\right\}.$$

Since the directional derivatives of Φ depends only on the values of Φ in a neighborhood of p, we observe that $\partial\Phi(P) = \partial\left(\Phi|_V\right)(p)$ for any convex neighborhood V of p. It follows that (5.3.1) holds for all $p, q \in \Delta_n$. Hence π is generated by Φ, and by Theorem 4.2.5 π is MCM. □

Now (i) of Theorem 5.1.3 is an immediate consequence of Lemma 5.3.2.

To prove (ii), we will show that given $\delta > 0$, there is a point $p \in \Delta_n$ such that the MCM property fails inside the Euclidean ball of radius δ around p. Then we may repeat the proof of (i) in this ball. This will be achieved by a method of contradiction using the following claim.

Claim. Suppose there exists $\delta > 0$ such that for any $p \in \Delta_n$, the MCM property holds over any choice of points selected within a ball of radius δ around p. Then the MCM property holds on Δ_n.

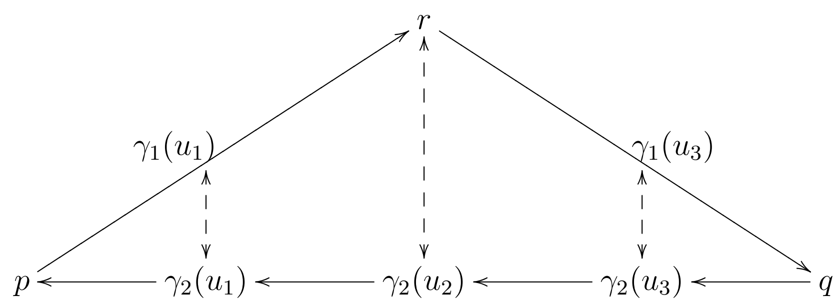

Figure 5.1: Decomposing a loop as a union of local loops. Here $r = \gamma_1(u_2)$.

To prove the above claim let us recall the notions of line integrals and conservative vector fields. Let γ be a piecewise linear curve in Δ_n indexed by a closed interval, say $[0, 1]$. The curve will be called a loop if $\gamma(0) = \gamma(1)$. The line integral of the vector field $w(\mu) := \pi(\mu)/\mu$ over any γ will be denoted by

$$I_\gamma(w) := \int_\gamma \frac{\pi(\mu)}{\mu} d\mu = \int_0^1 \frac{\pi(\mu(t))}{\mu(t)} \cdot \mu'(t) dt.$$

The line integral does not depend on parametrization, except for the orientation. By a slight abuse of notation, the line from any a to any b in Δ_n, irrespective of parametrization, will be denoted by $[a, b]$.

Let $p \in \Delta_n$ and let $B_\delta(p) = \{q \in \Delta_n : \|q - p\| < \delta\}$. Consider any loop γ whose range is contained in $B_\delta(p)$. Then we have

$$I_\gamma(w) = 0. \tag{5.3.2}$$

In other words, the vector field w is locally conservative restricted to every $B_\delta(p)$. To see (5.3.2), we use the fact that π satisfies MCM over $B_\delta(p)$. Therefore, by Theorem 4.2.5, there is a positive concave function Φ on Δ_n which generates π on $B_\delta(p)$. Consider any line $\ell = [p_1, p_2]$ contained in $B_\delta(p)$. By [87, Theorem 24.2], we have $I_\ell(\pi/\mu) = \log \Phi(p_2) - \log \Phi(p_1)$. Thus (5.3.2) holds for any piecewise linear loop in $B_\delta(p)$.

We now show that any locally bounded and conservative vector field over Δ_n must be globally conservative. While this statement is well known for smooth vector fields, we only assume that π/μ is measurable and locally bounded, and the resulting potential $\log \Phi$ is not necessarily differentiable. Since we are unable to find a reference for this result, we will give a sketch of proof and refer the reader to [83, Proof of Theorem 8] for more details.

Let $w(\mu) = \pi(\mu)/\mu$ be locally conservative in the sense of (5.3.2). Fix $p, q \in \Delta_n$ and consider two piecewise linear curves γ_1 and γ_2 from p to q. We will be done once we show

$$\int_{\gamma_1} w(\mu) d\mu = \int_{\gamma_2} w(\mu) d\mu. \tag{5.3.3}$$

Without loss of generality, we may assume that $\gamma_2(t) = (1-t)p + tq$.

In fact, we can assume that γ_1 has exactly three corners p, r, q and is a concatenation of $[p, r]$ and $[r, q]$ (we call such curves triangular). This is because once we establish (5.3.3) for such triangular curves, we can inductively eliminate corners in any other γ_1 and establish (5.3.3) in general.

For the rest of the argument we assume that γ_1 is triangular and γ_2 is $[p, q]$. Assume both γ_1 and γ_2 are indexed by $[0, 1]$.

We first suppose that $\sup_{0 \leq t \leq 1} \|\gamma_1(t) - \gamma_2(t)\| < \frac{\delta}{2}$. In this case, choose points $u_0 = 0 < u_1 < u_2, \ldots$ in $[0, 1]$ such that their images on γ_2 are a sequence of equidistant points with successive distance less than $\delta/2$. Now add lines between $\gamma_1(u_i)$ and $\gamma_2(u_i)$. Now consider each loop which is formed by the 4 oriented lines $[\gamma_2(u_{i+1}), \gamma_2(u_i)]$, $[\gamma_2(u_i), \gamma_1(u_i)]$, $[\gamma_1(u_i), \gamma_1(u_{i+1})]$, and $[\gamma_1(u_{i+1}), \gamma_2(u_{i+1})]$. See Figure 5.1.

By the triangle inequality for Euclidean distance it follows that the loop lies entirely inside $B_\delta(\gamma_2(u_i))$. Hence, by our assumption on local conservation, the integrals of w over these loops are zero. However, the sum of the integrals over all these loops is precisely the integral of w over the concatenation of lines γ_1 and $-\gamma_2$. Therefore this integral is zero, proving (5.3.3).

It can be shown by means of a simple geometric argument that any other case can be reduced to Case 1 above (see [83] for details). Now that we have shown that w is globally conservative, we can unambiguously define a function Φ on Δ_n by fixing some $p_0 \in \Delta_n$ and defining

$$\log \Phi(p) = \int_\gamma \frac{\pi}{\mu} d\mu, \quad p \in \Delta_n, \tag{5.3.4}$$

where the integral is over any piecewise linear curve from p_0 to p. Over any $B_\delta(p)$, the function Φ must coincide (up to a constant) with the concave function resulting from the local MCM property of the vector field w. Thus, Φ is locally concave on Δ_n and hence it is concave (see [55, page 58]) and generates π. This shows that π is MCM over Δ_n and this completes the proof of the theorem.

Chapter 6

OPTIMAL TRANSPORT

In this chapter we explore a remarkable connection between functionally generated portfolio and optimal transport theory. This relation is first studied in joint work with Soumik Pal in [83, 82]. We begin by reviewing some general definitions of optimal transport theory. For more background we refer the reader to [101] and [4].

6.1 Preliminaries in optimal transport

Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish (separable complete metric) spaces. By $\mathcal{P}(\mathcal{X})$ we mean the set of all Borel probability measures on $\mathcal{X}$ (same for $\mathcal{Y}$). Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a measurable function called the cost function.

Let $P \in \mathcal{P}(\mathcal{X})$ and $Q \in \mathcal{P}(\mathcal{Y})$. A coupling of the pair (P, Q) is a probability measure $R \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ whose marginals are P and Q respectively, i.e.,

$$R(A \times \mathcal{Y}) = P(A), \quad R(\mathcal{X} \times B) = Q(B)$$

for any Borel set A of $\mathcal{X}$ and any Borel set B of $\mathcal{Y}$. We denote by $\Pi(P, Q)$ the set of all couplings of (P, Q). Note that $\Pi(P, Q)$ is non-empty for any P and Q, for it contains the product measure $R = P \otimes Q$. A coupling R of (P, Q) can be represented by a random element (X, Y) whose distribution is R.

Given P and Q, the Monge-Kantorovich optimal transport problem is the problem

$$\inf_{R \in \Pi(P,Q), (X,Y) \sim R} \mathbb{E}_R[c(X, Y)]. \tag{6.1.1}$$

Here the notation means that the random element (X, Y) has distribution R. The infimum in (6.1.1) is called the value of the optimal transport problem. A coupling that attains the infimum is called an optimal coupling. If R is an optimal coupling, we say that R solves the optimal transport problem. An optimal coupling (X, Y) of (6.1.1) is said to be deterministic if there exists a measurable function $F : \mathcal{X} \to \mathcal{Y}$ such that $Y = F(X)$ almost surely.

It is well known that under mild technical conditions (see [101]), cyclical monotonicity is a necessary and sufficient solution criteria of the optimal transport problem (6.1.1).

Definition 6.1.1. Let A be a subset of $\mathcal{X} \times \mathcal{Y}$. We say that A is c-cyclical monotone if for any $m \geq 1$, any sequence $\{(x(k), y(k))\}_{k=1}^m$ in A and any permutation σ of $\{1, \ldots, k\}$, we have

$$\sum_{k=1}^{m} c(x(k), y(k)) \leq \sum_{k=1}^{m} c(x(k), y(\sigma(k))). \tag{6.1.2}$$

It can be shown that c-cyclical monotonicity is equivalent to the property that for any finite sequence $\{(x(k), y(k))\}_{k=1}^m$ in A, we have

$$\sum_{k=1}^{m} c(x(k), y(k)) \leq \sum_{k=1}^{m} c(x(k+1), y(k)) \tag{6.1.3}$$

with the convention $x(m+1) = x(1)$ and $y(m+1) = y(1)$. The proof uses the cyclical decomposition of an arbitrary permutation.

6.2 Exponential coordinate system

In this section we formulate the optimal transport problem using the exponential coordinate system. In [83] we studied two (almost equivalent) formulations of the transport problems.[1] In this thesis we focus on the formulation given in terms of the exponential coordinates. The following notations will also be used in Chapter 10.

6.2.1 Δ_n as an exponential family

Consider the open unit simplex

$$\Delta_n = \left\{ p = (p_1, \ldots, p_n) \in \mathbb{R}^n : p_i > 0, \sum_{i=1}^{n} p_i = 1 \right\}.$$

[1] Also see [83] for a related transport problem on Δ_n where the negative relative entropy is taken as the cost function.

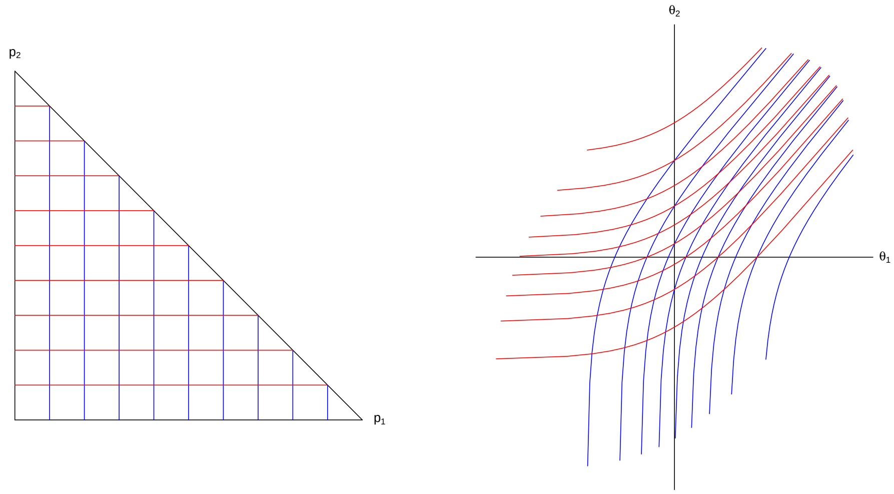

Figure 6.1: Coordinate curves of the exponential coordinate system.

We regard it as an $(n-1)$-dimensional smooth manifold. The exponential coordinate system defines a global coordinate system on Δ_n (see Figure 6.1).

Definition 6.2.1 (Exponential coordinate system)**.** The exponential coordinate

$$\theta = (\theta_1, \ldots, \theta_{n-1}) \in \mathbb{R}^{n-1}$$

of $p \in \Delta_n$ is given by

$$\theta_i = \log \frac{p_i}{p_n}, \quad i = 1, \ldots, n-1. \tag{6.2.1}$$

We denote this map by $\boldsymbol{\theta} : \Delta_n \to \mathbb{R}^{n-1}$. By convention we set $\theta_n \equiv 0$. The inverse transformation $\mathbf{p} := \boldsymbol{\theta}^{-1}$ is given by

$$p_i = \mathbf{p}_i(\theta) = e^{\theta_i - \psi(\theta)}, \quad 1 \le i \le n, \tag{6.2.2}$$

where

$$\psi(\theta) = \log\left(1 + \sum_{i=1}^{n-1} e^{\theta_i}\right) = \log\left(\sum_{i=1}^{n} e^{\theta_i}\right) \tag{6.2.3}$$

Remark 6.2.2 (Δ_n as an exponential family). Let $X = \{1, 2, \ldots, n\}$. Then the open unit simplex Δ_n can be regarded as the family of positive probability densities on X with respect to the counting measure. For $i = 1, \ldots, n-1$, let $s_i : X \to \mathbb{R}$ be the function defined by

$$s_i(x) = \begin{cases} 1 & \text{if } x = i, \\ 0 & \text{if } x \neq i. \end{cases}$$

Now we may rewrite (6.2.1) and (6.2.2) in the form

$$p_x = f(x; \theta) = \exp\left(\sum_{i=1}^{n} \theta_i s_i(x) - \psi(\theta)\right), \quad x \in X.$$

Thus we have expressed Δ_n as an exponential family of probability densities where s_i are the sufficient statistics. Moreover,

$$\psi(\theta) = \log\left(\sum_{i=1}^{n} \exp\left(\sum_{i=1}^{n-1} \theta_i s_i(x)\right)\right)$$

is the cumulant generating function of the family (see [2, Section 2.2.2]).

The exponential coordinate system is the first of several coordinate systems we will introduce on the simplex. By changing coordinate systems, any function on Δ_n can be expressed as a function on $\mathbb{R}^{n-1}$ and vice versa. Explicitly, a function φ on Δ_n can be expressed in exponential coordinates by $\theta \mapsto \varphi(\mathbf{p}(\theta))$. To simplify the notations, we simply write $\varphi(p)$ or $\varphi(\theta)$ depending on the coordinate system used. For example, if $\varphi(p) = \sum_{i=1}^{n} \pi_i \log p_i$ is the cross entropy where $\pi \in \Delta_n$, then $\varphi(\theta) = \sum_{i=1}^{n-1} \pi_i \theta_i - \psi(\theta)$.

6.2.2 Portfolio as a transport map

Now let $\pi : \Delta_n \to \Delta_n$ be a portfolio map.[2] If the current market weight is $\mu(t) = p$, the portfolio vector is $\pi(p)$. We may represent both p and $\pi(p)$ in terms of the exponential coordinate system. Let the exponential coordinates of p be $\theta = (\theta_1, \ldots, \theta_n)$. Then, there exists $\phi \in \mathbb{R}^{n-1}$ such that the exponential coordinates of $\pi(p)$ is $\theta - \phi$, i.e.,

$$\theta_i - \phi_i = \log \frac{\pi_i(p)}{\pi_n(p)}, \quad i = 1, \ldots, n-1. \tag{6.2.4}$$

[2]Note that the range of π is Δ_n but not the closed unit simplex $\overline{\Delta}_n$. We need this assumption in order to use the exponential coordinate system. In [83] we give another formulation of the transport problem which does not require this assumption. Our formulation is slightly less general but is more suitable for information geometry to be studied in Part III.

Rearranging, we have

$$\phi_i = \theta_i - \log \frac{\pi_i(\theta)}{\pi_n(\theta)}, \quad i = 1, \ldots, n-1, \tag{6.2.5}$$

and

$$\pi_i(\theta) = e^{\theta_i - \phi_i - \psi(\theta - \psi)}, \quad i = 1, \ldots, n-1.$$

Here we abuse notations and write $\pi(\theta)$ to mean π as a function of the exponential coordinates. In other words, ϕ is the negative shift in exponential coordinates to go from p to $\pi(p)$. By (6.2.5), each portfolio map $\pi : \Delta_n \to \Delta_n$ induces a map $F : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ that 'transport' θ to ϕ.

6.2.3 The optimal transport problem

Now we may state the optimal transport problem. Using the notations of Section 6.1, we take $\mathcal{X} = \mathcal{Y} = \mathbb{R}^{n-1}$ which is a Polish space when equipped with the usual Euclidean metric and topology. As for the cost function, we take

$$c(\theta, \phi) = \psi(\theta - \phi) = \log \left(1 + \sum_{i=1}^{n-1} e^{\theta_i - \phi_i} \right), \quad \theta \in \mathcal{X}, \quad \phi \in \mathcal{Y}. \tag{6.2.6}$$

Since the cost function is real-valued and continuous, there exists an optimal coupling under mild integrability conditions on P and Q (see [101, Theorem 4.1]).

6.3 c-cyclical monotonicity

In this section we give the connection between portfolio theory and the cost function (6.2.6). To do this, we need to express the relative value of a portfolio in terms of the exponential coordinate system. Let $\pi : \Delta_n \to \Delta_n$ be a portfolio map, and let $\{\mu(t)\}_{t=0}^{\infty}$ be a market path. Let $\theta(t) \in \mathcal{X} = \mathbb{R}^{n-1}$ be the exponential coordinate of $\mu(t)$. Then we may regard the market as a path in $\mathcal{X}$.

We write $\phi = \phi(\theta) = F(\theta)$ for the negative shift map induced by π (see (6.2.5)). For notational convenience we write $\theta_n = \phi_n = 0$. Then $\theta_i = e^{\theta_i - \psi(\theta)}$ and $\pi_i(\theta) = e^{\theta_i - \phi_i(\theta) - \psi(\theta - \phi(\theta))}$ for all $1 \leq i \leq n$.

Using (2.2.3), we have

$$
\begin{aligned}
\frac{V_\pi(t+1)}{V_\pi(t)} &= \sum_{i=1}^{n} \pi_i(\mu(t)) \frac{\mu_i(t+1)}{\mu_i(t)} \\
&= \sum_{i=1}^{n} e^{\theta_i(t)-\phi_i(\theta(t))-\psi(\theta(t)-\phi(\theta(t)))} \frac{e^{\theta_i(t+1)-\psi(\theta(t+1))}}{e^{\theta_i(t)-\psi(\theta(t))}} \\
&= e^{-\psi(\theta(t)-\phi(\theta(t)))+\psi(\theta(t))-\psi(\theta(t+1))} e^{\psi(\theta(t+1)-\phi(\theta(t)))}.
\end{aligned}
$$

Taking logarithm on both sides, we have

$$
\log \frac{V_\pi(t)}{V_\pi(t)} = \left[\psi(\theta(t)) - \psi(\theta(t+1))\right] + \left[\psi(\theta(t+1) - \phi(\theta(t))) - \psi(\theta(t) - \phi(\theta(t)))\right].
$$

Summing over time, we get

$$
\begin{aligned}
\log V_\pi(t) &= \psi(\theta(0)) - \psi(\theta(t)) + \sum_{s=0}^{t-1} \left[\psi(\theta(s+1) - \phi(s)) - \psi(\theta(s) - \phi(\theta(s)))\right] \\
&= \psi(\theta(0)) - \psi(\theta(t)) + \sum_{s=0}^{t-1} \left[c(\theta(s+1), \phi(\theta(s))) - c(\theta(s), \phi(\theta(s)))\right].
\end{aligned}
\tag{6.3.1}
$$

Now consider a discrete cycle $\{\mu(t)\}_{t=0}^{m+1}$ in Δ_n where $\mu(m+1) = \mu(m)$. Putting $t = m+1$ in (6.3.1), we have

$$
\log V_\pi(m+1) = \sum_{t=0}^{m} \left[c(\theta(t+1), \phi(\theta(t))) - c(\theta(t), \phi(\theta(t)))\right].
$$

Using (6.1.3), we summarize the results in the following theorem.

Theorem 6.3.1. *For any portfolio map $\pi : \Delta_n \to \Delta_n$ the following statements are equivalent.*

(i) There exists an exponentially concave function φ on Δ_n which generates π in the sense of Definition 4.1.1.

(ii) The portfolio map is multiplicatively cyclical monotone in the sense of Definition 4.2.2.

(iii) The graph of the map $\theta \mapsto \phi$ defined by (6.2.5) *is c-cyclical monotone.*

6.4 *Optimal transport and duality*

6.4.1 *c-concavity and duality*

Now we make use of the notion of c-concavity in optimal transport theory. The results of this section will be useful in Chapter 10. The definitions we use are standard and can be found in [4, Chapter 1]. For $f : \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$ we define its c-transform by

$$f^*(\phi) = \inf_{\theta \in \mathcal{X}} \left(c(\theta, \phi) - f(\theta) \right), \quad \phi \in \mathcal{Y}. \tag{6.4.1}$$

The c-transform of a function $g : \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\}$ is defined analogously. We say that $f : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ is c-concave if there exists $g : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ such that $f = g^*$ (similar for c-concave function on $\mathcal{Y}$). A function h (on $\mathcal{X}$ or $\mathcal{Y}$) is c-concave if and only if $h^{**} = h$.

If $f : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ is c-concave, its c-superdifferential is defined by

$$\partial^c f = \{(\theta, \phi) \in \mathcal{X} \times \mathcal{Y} : f(\theta) + f^*(\phi) = c(\theta, \phi)\}. \tag{6.4.2}$$

For $\theta \in \mathcal{X}$ we define $\partial^c f(\theta) = \{\phi \in \mathcal{Y} : (\theta, \phi) \in \partial^c f\}$. If this set is a singleton $\{\phi\}$, we call ϕ the c-supergradient of f at θ and write $\phi = \nabla^c f(\theta)$. Analogous definitions hold for a c-concave function g on $\mathcal{Y}$.

Let $f : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ be c-concave. By definition, equality holds in

$$f(\theta) + f^*(\phi) \leq c(\theta, \phi) \tag{6.4.3}$$

if and only if $(\theta, \phi) \in \partial^c f$. This is a generalized version of Fenchel's identity and will be used frequently in this section.

Lemma 6.4.1 (Exponential concavity and c-concavity)**.** *For $\varphi : \Delta_n \to \mathbb{R} \cup \{-\infty\}$ the following statements are equivalent.*

(i) φ is exponentially concave on Δ_n.

(ii) The function $f : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ defined by

$$f(\theta) = \varphi(\mathbf{p}(\theta)) + \psi(\theta)$$

is c-concave on $\mathcal{X}$.

(iii) The function $g : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ defined by

$$g(\phi) = \varphi(\mathbf{p}(-\phi)) + \psi(-\phi),$$

where $-\phi$ is the exponential coordinate, is c-concave on $\mathcal{Y}$.

Proof. We prove the implication (i) $\Rightarrow$ (ii) and the others can be proved similarly. Suppose (i) holds and consider the non-negative concave function $\Phi = e^{\varphi}$ on Δ_n. By [87, Theorem 10.3], we can extend Φ continuously up to $\overline{\Delta}_n$, the closure of Δ_n in $\mathbb{R}^n$. We further extend Φ to the affine hull H of Δ_n in $\mathbb{R}^n$ by setting $\Phi(p) = -\infty$ for $p \notin \overline{\Delta}_n$. The extended function Φ is then a closed concave function on H. By convex duality (see [87, Theorem 12.1]), there exists a family $\mathcal{C}$ of affine functions on H such that

$$\Phi(p) = \inf_{\ell \in \mathcal{C}} \ell(p), \quad p \in \Delta_n. \tag{6.4.4}$$

Since Φ is non-negative on Δ_n, each $\ell \in \mathcal{C}$ is non-negative on Δ_n. Replacing ℓ by the sequence $\ell_k = \ell + \frac{1}{k}$, we may assume without loss of generality that each $\ell \in \mathcal{C}$ is positive on Δ_n. We parameterize $\ell \in \mathcal{C}$ in the form $\ell(p) = \sum_{i=1}^n a_i p_i$ where $a_1, \ldots, a_n$ are positive constants. Now we write

$$\begin{aligned}
\log \ell(p) &= \log \left(\sum_{i=1}^n a_i p_i \right) \\
&= \log \left(1 + \sum_{i=1}^{n-1} \frac{a_i}{a_n} \frac{p_i}{p_n} \right) + \log p_n + \log a_n \\
&= c(\theta - \phi) - \psi(\theta) + \log a_n,
\end{aligned}$$

where $\phi_i := -\log \frac{a_i}{a_n}$, $i = 1, \ldots, n-1$. It follows from (6.4.4) that

$$\varphi(\theta) + \psi(\theta) = \inf_{\ell \in \mathcal{C}} \left(c(\theta - \phi) + \log a_n \right).$$

Define $h : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ by setting

$$h(\phi) = \inf \left\{ -\log a_n : \exists\ \ell(p) = \sum_{i=1}^n a_i p_i \in \mathcal{C} \text{ s.t. } \phi_i = -\log \frac{a_i}{a_n}\ \forall\ i \right\},$$

where the infimum of the empty set is $-\infty$. Thus $f = \varphi + \psi = h^*$ which shows that f is c-concave on $\mathcal{X}$. □

Now we show that the transport problem is solved by a deterministic coupling under mild conditions.

Proposition 6.4.2. *Let R be an optimal coupling for the transport problem with cost $c(\theta, \phi) = \psi(\theta - \phi)$. Suppose that P is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{n-1}$ and the optimal cost is finite. Then there exists an exponentially concave function φ on Δ_n such that*

$$(\theta, \phi) = (\theta, \nabla^c f(\theta))$$

R-almost surely, where $f = \varphi + \psi$. In particular, the optimal coupling is attained by a deterministic map.

Proof. Since R is an optimal coupling, its support is c-cyclical monotone. It follows that there exists a c-concave function f on $\mathcal{X}$ such that the support of R is a subset of $\partial^c f$, the c-superdifferential of f (see [4, Chapter 1]). Recall that $(\theta, \phi) \in \partial^c f$ if and only if $f(\theta) + f^*(\phi) = c(\theta, \phi)$. Since $-\infty + a = -\infty$ for all $a \in [-\infty, \infty)$, f and f^* are real-valued under P and Q respectively. By Lemma 6.4.1, if we define $\varphi = f - \psi$, then φ is exponentially concave. We know that an exponentially concave function is almost everywhere differentiable on its domain. Since P is assumed to be absolutely continuous, f is P-almost surely differentiable. Thus $\phi = \nabla^c f(\theta)$ P-almost surely. □

In the smooth case, the transport map $\theta \mapsto \phi$ gives another coordinate system of the simplex which will play a crucial role in Chapter 10. To prove this, we will impose the following regularity conditions on the exponentially concave function φ.

Assumption 6.4.3 (Regularity conditions)*.* We assume the following.

(i) The function φ is smooth (i.e., infinitely differentiable) on Δ_n.

(ii) The (Euclidean) Hessian of $\Phi = e^{\varphi}$ is strictly negative definite everywhere on Δ_n. In particular, Φ is strictly concave and the portfolio π generated by φ maps Δ_n into Δ_n.

The following is the c-concave version of the classical Legendre transformation [87].

Theorem 6.4.4 (c-Legendre transformation)**.** *Let φ be an exponentially concave function φ satisfying Assumption* 6.4.3*, and let π be the portfolio map generated by φ. Given φ, consider the c-concave function*

$$f(\theta) := \varphi(\theta) + \psi(\theta) \tag{6.4.5}$$

defined on $\mathcal{X} = \mathbb{R}^{n-1}$ via the exponential coordinate system.

(i) The c-supergradient of f is given by (6.2.5)*, i.e.,*

$$\nabla^c f(\theta) = \left(\theta_i - \log \frac{\pi_i(\theta)}{\pi_n(\theta)}\right)_{1 \leq i \leq n-1}, \quad \theta \in \mathcal{X}. \tag{6.4.6}$$

Moreover, the map $\nabla^c f : \mathcal{X} \to \mathcal{Y}$ is injective.

(ii) Let $\mathcal{Y}' \subset \mathcal{Y}$ be the range of $\nabla^c f$. Then the c-supergradient of f^* is given on $\mathcal{Y}'$ by

$$\nabla^c f^*(\phi) = \left(\nabla^c f\right)^{-1}(\phi).$$

In fact, the map $\nabla^c f$ is a diffeomorphism from $\mathcal{X}$ to $\mathcal{Y}'$ whose inverse is $\nabla^c f^*$. Also, the function f^* is smooth on the open set $\mathcal{Y}'$.

Proof of Theorem 6.4.4. In this proof we treat θ and ϕ as independent variables.

We prove (i) and (ii) together. We begin by observing that

$$\frac{\partial}{\partial \theta_i} f(\theta) = \pi_i(\theta), \quad 1 \leq i \leq n-1. \tag{6.4.7}$$

To see this, write $p_i = e^{\theta_i - \psi(\theta)}$. Switching coordinates and using the chain rule, we have

$$\begin{aligned} \frac{\partial}{\partial \theta_i} f(\theta) &= \sum_{j=1}^n \frac{\partial \varphi}{\partial p_j}(p)(-p_i p_j) + \frac{\partial \varphi}{\partial p_i}(p) p_i + p_i \\ &= p_i \left(1 + \frac{\partial \varphi}{\partial p_i}(p) - \sum_{j=1}^n p_j \frac{\partial \varphi}{\partial p_j}(p)\right) = \pi_i(\theta). \end{aligned}$$

Consider the c-transform of f given by

$$f^*(\phi) = \inf_{\theta \in \mathcal{X}} \left(\psi(\theta - \phi) - f(\theta)\right). \tag{6.4.8}$$

Differentiating $\psi(\theta - \phi) - f(\theta)$ and using (6.4.7), we see that $\theta \in \mathcal{X}$ attains the infimum in (6.4.8) if and only if

$$\frac{e^{\theta_i - \phi_i}}{\sum_{j=1}^n e^{\theta_j - \phi_j}} = \pi_i(\theta), \quad i = 1, \ldots, n-1.$$

Rearranging, we have

$$\phi_i = \theta_i - \log \frac{\pi_i(\theta)}{\pi_n(\theta)}, \quad i = 1, \ldots, n-1. \tag{6.4.9}$$

This proves that equality holds in

$$f(\theta) + f^*(\phi) \leq \psi(\theta - \phi) \tag{6.4.10}$$

if and only if θ and ϕ satisfies the relation (6.4.9). In particular, for $\theta \in \mathcal{X}$ the c-supergradient $\nabla^c f(\theta)$ is given by (6.4.9).

Next we we prove that the minimizer in (6.4.8), if exists, is unique. Consider instead maximization of the quantity

$$e^{\varphi(\theta) + \psi(\theta) - \psi(\theta - \phi)} = \Phi(\theta) \frac{e^{\psi(\theta)}}{e^{\psi(\theta - \phi)}}.$$

Expanding and switching to Euclidean coordinates, this equals

$$\Phi(p) \frac{\sum_{i=1}^n e^{\theta_i}}{\sum_{i=1}^n e^{\theta_i - \phi_i}} = \sup_{p \in \Delta_n} \Phi(p) \frac{1}{\sum_{i=1}^n a_i p_i}, \tag{6.4.11}$$

where $a_i = e^{-\phi_i} > 0$. Being the quotient of a strictly concave function and an affine function, the right hand side of (6.4.11) is strictly quasi-concave, i.e., its superlevel sets are strictly convex (see [15, Example 3.38]). This shows that the minimizer θ in (6.4.8) is unique if it exists.

Let $\phi \in \mathcal{Y}'$. Then there exists unique $\theta \in \mathcal{X}$ such that equality holds in (6.4.10) and $\phi = \nabla^c f(\theta)$. In particular, the c-supergradient $\partial^c f^*(\phi)$ is θ and $\nabla^c f^*(\nabla^c f(\theta)) = \theta$. This completes the proof of (i) and (ii).

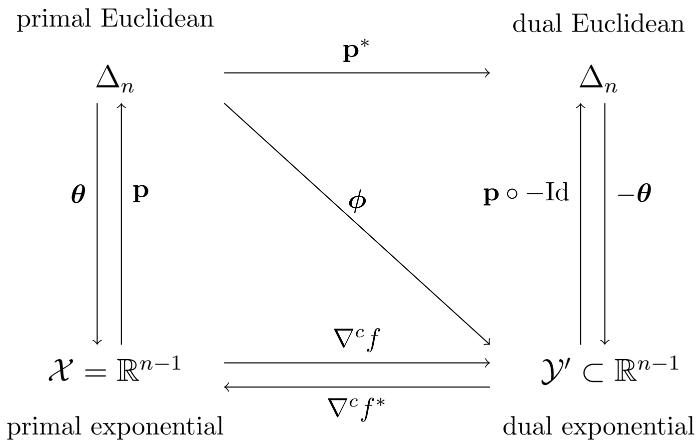

Figure 6.2: Coordinate systems on Δ_n.

It remains to prove that $\nabla^c f : \mathcal{X} \to \mathcal{Y}'$ is a diffeomorphism. Since $\nabla^c f : \mathcal{X} \to \mathcal{Y}$ is smooth and injective, by the inverse function theorem it suffices to show that the Jacobian of $\nabla^c f$ is invertible everywhere. Using the fact that the Hessian of Φ is strictly positive definite, this can be verified by a tedious calculation which we omit. Having known that the transformation $\nabla^c f^*$ is smooth, we see that

$$f^*(\phi) = c(\nabla^c f^*(\phi), \phi) - f(\nabla^c f^*(\phi))$$

is smooth on $\mathcal{Y}'$ as well. $\square$

Although $\mathcal{Y}'$ is in general a strict subset of $\mathcal{Y}$, the dual variable $\phi = \nabla^c f(\theta)$ defines a global coordinate system of the manifold Δ_n. In Chapter 10 we will use another coordinate system on Δ_n called the dual Euclidean coordinate system. Thus we have four coordinate systems on Δ_n: Euclidean, primal, dual and dual Euclidean (see Figure 6.2). In the following we will frequently switch between coordinate systems to facilitate computations. To avoid confusions let us state once for all the conventions used. We let φ and $f = \varphi + \psi$ be given.

Definition 6.4.5 (Coordinate systems)**.** For the manifold Δ_n we call the identity map

$$p = (p_1, \ldots, p_n), \quad p_i > 0, \quad \sum_{i=1}^{n} p_i = 1$$

the (primal) Euclidean coordinate system with range Δ_n. We let

$$\theta = \boldsymbol{\theta}(p) = \left(\log \frac{p_1}{p_n}, \ldots, \log \frac{p_{n-1}}{p_n}\right)$$

be the primal (exponential) coordinate system with range $\mathcal{X}$ and

$$\phi = \boldsymbol{\phi}(p) := \nabla^c f(\theta)$$

be the dual (exponential) coordinate system with range $\mathcal{Y}'$. The dual Euclidean coordinate system is defined by the composition

$$p^* = \mathbf{p}^*(p) := \mathbf{p}(-\boldsymbol{\phi}(p)).$$

From now on p, p^*, θ and ϕ always represent the same point of Δ_n. In particular, unless otherwise specified θ and ϕ are dual to each other in the sense that $\phi = \nabla^c f(\theta)$. By convention we let $\theta_n = \phi_n = 0$ for any $p \in \Delta_n$.

Notation 6.4.6 (Switching coordinate systems). We identify the spaces Δ_n, $\mathcal{X}$ and $\mathcal{Y}'$ using the coordinate systems in Definition 6.4.5. If h is a function on any one of these spaces, we write $h(p) = h(\theta) = h(\phi) = h(p^*)$ depending on the coordinate system used.

We also record a useful fact. A formula analogous to the first statement is derived in [95].

Lemma 6.4.7. *For $1 \leq i \leq n-1$, we have*

$$\frac{\partial}{\partial \theta_i} f(\theta) = \pi_i(\theta), \quad \frac{\partial}{\partial \phi_i} f^*(\phi) = -\pi_i(\phi).$$

Proof. The first statement is derived in the proof of Theorem 6.4.4. The second statement can be proved in a similar way using the chain rule. □

6.5 Empirical examples: two stocks case

In general, solving optimal transport problems (either analytically or numerically for given P and Q) is a difficult task; see for example [10] and the references therein. Designing practical algorithms for solving the transport problem with cost (6.2.6) is an interesting open problem.

In the case $n = 2$, the solution can be characterized explicitly due to the special structure of the real line and the convexity of the cost function. In this section we present the solution and give several empirical examples.

6.5.1 Monotone rearrangements

Throughout this section we assume $n = 2$. A typical point p in Δ_n is represented as

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

where $\theta \in \mathbb{R}$ is the exponential coordinate of p. A portfolio vector $\pi(p)$ with positive weights corresponding to p can be expressed as

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

for some $\phi \in \mathbb{R}$. So the exponential coordinate of $\pi(p)$ is $\theta - \phi$. We will choose ϕ as a function of θ. As ϕ increases, the portfolio underweights more and more stock 1 relative to the market weight. See Figure 6.3 for the dependence of the portfolio on ϕ at different points on the simplex. As an example, if $|\phi|$ is bounded by 0.6, the graph of the resulting portfolio will lie within the curves labeled -0.6 and 0.6.

Consider the transport problem with cost (6.2.6). Let P and Q be probability measures on $\mathbb{R}$. We assume that P is absolutely continuous with respect to the Lebesgue measure. The cost is

$$c(\theta, \phi) = \psi(\theta - \phi) = \log\left(1 + e^{\theta - \phi}\right), \quad \theta, \phi \in \mathbb{R}.$$

Here $\psi(x) = \log(1 + e^{x})$ is smooth and strictly convex. The transport problem is

$$\mathbb{E}_{R \in \Pi(P,Q)} \psi(\theta - \phi) \tag{6.5.1}$$

where (θ, ϕ) is a random element in $\mathbb{R} \times \mathbb{R}$ with distribution R.

Let G and H be the distribution functions of P and Q respectively.

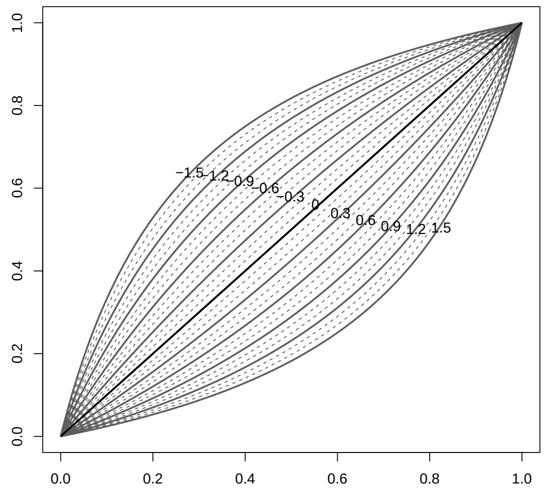

Figure 6.3: Plot of $\pi_1(p) = \frac{e^{\theta-\phi}}{1+e^{\theta-\phi}}$ as a function of $p_1 = \frac{e^{\theta}}{1+e^{\theta}}$, for different values of ϕ (labeled).

Definition 6.5.1 (Monotone rearrangement)**.** The monotone transport map from P to Q is the map $F : \mathbb{R} \to \mathbb{R}$ defined by

$$F(x) = \inf\{y : H(y) \geq G(x)\}. \tag{6.5.2}$$

In other words, F is defined by matching the quantiles of H to those of G. It is clear from (6.5.2) that F is non-decreasing. Moreover, it is easy to check that if $\theta \sim P$, then $F(\theta) \sim Q$. Thus $(\theta, F(\theta))$ is a coupling of (P, Q). In fact, F is the unique non-decreasing function (up to the null sets of P) which maps P to Q.

The following theorem is a special case of a well-known fact (see for example [59, Theorem 3.1]). Indeed, the monotone transport map remains optimal if ψ is replaced by any strictly convex function.

Theorem 6.5.2. *The coupling $(\theta, F(\theta))$ where $\theta \sim P$ and F is the monotone transport map from P to Q is the unique solution to the transport problem* (6.5.1)*.*

An explicit example is where P and Q are normal. In this case, the monotone transport map is linear and the corresponding portfolio is essentially a diversity-weighted portfolio.

Proposition 6.5.3. *Let $P = N(m_1, \sigma_1^2)$ and $Q = N(m_2, \sigma_2^2)$. Then the monotone transport map is given by*

$$F(\theta) = m_2 + \frac{\sigma_2}{\sigma_1}(\theta - m_1), \quad \theta \in \mathbb{R}.$$

Moreover, the portfolio map corresponding to the transport map F is given by

$$\pi(p) = \left(\frac{c p_1^{\alpha}}{c p_1^{\alpha} + p_2^{\alpha}}, \frac{p_2^{\alpha}}{c p_1^{\alpha} + p_2^{\alpha}} \right), \tag{6.5.3}$$

where $\alpha = 1 - \frac{\sigma_2}{\sigma_1}$ and $c = \exp\left(\frac{\sigma_2}{\sigma_1} m_1 - m_2\right)$.

Proof. Let $X \sim N(m_1, \sigma_1^2)$ and $Y \sim N(m_2, \sigma_2^2)$. The first statement follows from the fact that $(X - m_1)/\sigma_1$ and $(Y - m_2)/\sigma_2$ have the same distribution.

To show (6.5.3), note that $\theta = \log \frac{p_1}{p_2}$ and, by definition of $\pi(p)$, we have

$$\begin{aligned} \log \frac{\pi_1(p)}{\pi_2(p)} = \theta - F(\theta) &= \left(1 - \frac{\sigma_2}{\sigma_1}\right) \log \frac{p_1}{p_2} + \left(\frac{\sigma_2}{\sigma_1} m_1 - m_2\right) \\ &= \alpha \log \frac{p_1}{p_2} + \log c. \end{aligned}$$

Rearranging gives the result. □

Clearly the portfolio has the form (6.5.3) whenever the transport map is linear, so the normality assumption is not required. Nevertheless, it is instructive to see how the portfolio depends on the means and variances of P and Q. In particular, the exponent α in Proposition 6.5.3 depends on the ratio $\frac{\sigma_2}{\sigma_1}$. If $\sigma_1 = \sigma_2$, then $\alpha = 0$ and π is a constant-weighted portfolio. If $0 < \sigma_2 < \sigma_1$, then $0 < \alpha < 1$ and π is essentially the diversity-weighted portfolio. If $\sigma_2 > \sigma_1 > 0$, then α is negative and the corresponding portfolio is studied in the recent paper [99]. On the other hand, the mean m_2 of $\widetilde{\mathcal{Q}}$ represents systematic overweight/underweight of stock 1 and interacts with other parameters to determine the constant c.

We may generalize Proposition 6.5.3 as follows.

Example 6.5.4 (Product of Gaussian distributions). Let P be a product of one-dimensional Gaussian distributions:

$$P = \bigotimes_{i=1}^{n-1} N(a_i, \sigma_i^2),$$

where $a_i \in \mathbb{R}$ and $\sigma_i > 0$. Also let

$$Q = \bigotimes_{i=1}^{n-1} N(b_i, (1-\lambda)\sigma_i^2)$$

where $b_i \in \mathbb{R}$ and $0 < \lambda < 1$. Then the optimal transport map for the measures P and Q is given by the map (6.2.5), where π is the following variant of the diversity-weighted portfolio:

$$\pi_i(p) = \frac{c_i p_i^\lambda}{\sum_{j=1}^n c_j p_j^\lambda}, \quad \varphi(p) = \frac{1}{\lambda} \log \left(\sum_{j=1}^n c_j p_j^\lambda \right). \tag{6.5.4}$$

Here the coefficients c_i are chosen such that $(1-\lambda)a_i - \log \frac{c_i}{c_n} = b_i$ for all i.

6.5.2 Empirical examples

In this subsection we use a simple example to illustrate how our methodology of optimal transport might be applied in practice. Consider the monthly stock prices of Walmart (stock 1) and Microsoft (stock 2) from January 1995 to July 2015. The stock prices (normalized to be \$1 at January 1995) are plotted in Figure 6.4 (top left). The 'market' consists of the two stocks and the initial market weight is $(0.5, 0.5)$. We compute the exponential coordinate process $\theta(t) = \log \frac{\mu_1(t)}{\mu_2(t)}$ (top right). Suppose we use the first 10 years of data (120 months) as training data. Our objective is to use the training data as well as choices of $\mathcal{Q}$ to construct portfolios that will be backtested using the next 10 years of data. To do this using optimal transport, we need to specify the probability distributions P and Q on $\mathbb{R}$.

Choice of P. The measure P reflects our belief of the position of $\theta(t)$ in the future. Figure 6.4 plots the density estimate of $\theta(t)$ over the training period (bottom left). The distribution is bimodal (corresponding to the periods 1997–2000 and 2002–2004) and is mostly concentrated in the interval $[-1.2, 0]$. Suppose our belief is that the market weight will most likely remain in this region in the next decade. For simplicity, we take P to be the

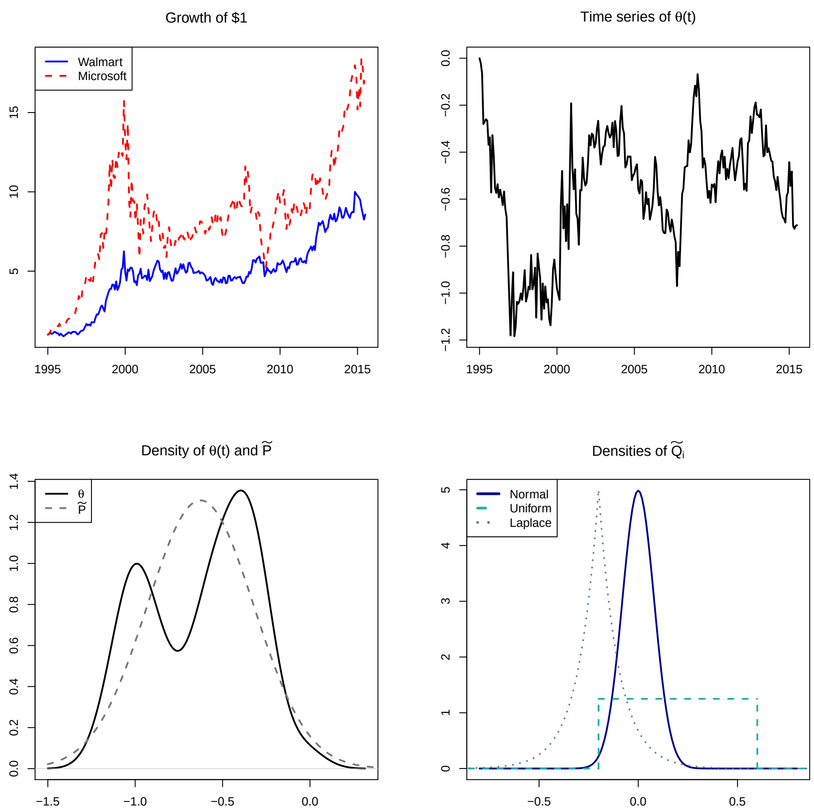

Figure 6.4: Plots of the data. Top left: Time series of the normalized stock prices. Top right: Time series of $\theta(t)$, the exponential coordinate process. Bottom left: Density estimate of $\theta(t)$ over the training period (solid curve) and density of P (dashed curve). Bottom right: Densities of our choices of Q.

normal distribution whose mean and standard deviation match those of the density estimate. Explicitly, we have

$$P = N(-0.626, 0.305).$$

A more diffuse distribution can be chosen if the investor is less certain.

Choice of Q. Recall that the portfolio has the representation

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

where ϕ is a function of θ and Q is the marginal distribution of ϕ, given that θ is distributed as P.

To illustrate the effects of different distributions we consider three distributions given as follow:

$$Q_1 = N(0, 0.08),$$
$$Q_2 = \text{Uniform}(-0.2, 0.6),$$
$$Q_3 = \text{Laplace}(\text{location} = -0.2, \text{scale} = 0.1).$$

Here we recall that the Laplace distribution with location parameter a and scale parameter b has density given by $f(x) = \frac{1}{b}\exp\left(-\frac{|x-a|}{b}\right)$. The densities of these distributions are shown in Figure 6.4 (bottom right). We denote the resulting portfolios by $\pi^{(1)}$, $\pi^{(2)}$ and $\pi^{(3)}$.

Let us give some intuitions about these distributions. Overall, the distributions we choose concentrate in the interval $[-0.6, 0.6]$. From Figure 6.3, they allow moderate deviations from the market weight but not too much (most of the time).

Note that Q_1 has mean 0 and has a rather small standard deviation (about a quarter of the standard deviation of P). This means that on average $\pi^{(1)}$ will not overweight or underweight stock 1 (Walmart) and the deviation is most of the time small. By Proposition 6.5.3, we know that $\pi^{(1)}$ is a diversity-weighted portfolio with $\alpha = 1 - \frac{0.08}{0.305} \approx 0.74$. (From (6.5.3), the portfolio is constant-weighted if $\alpha = 0$ and buy-and-hold if $\alpha = 1$.)

For Q_2, we expect that $\pi^{(2)}$ tends to underweight stock 1 (about 75% of the time provided the future empirical distribution of $\theta(t)$ is close to P). Since Q_2 has bounded support, the weight ratios of $\pi^{(2)}$ are uniformly bounded on Δ_n. However, the underweight can be significant on a certain region.

Finally, Q_3 has a Laplace distribution which has fatter tails than the normal distribution. Thus we expect that $\pi^{(3)}$ deviates more (from the market portfolio) than a diversity-weighted portfolio with matching parameters near the boundary of the simplex. Also Q_3 is chosen to have negative mean. Thus $\pi^{(3)}$ will tend to overweight stock 1. In practice, the location measure of Q should reflect the investor's belief about the relative performances of the stocks

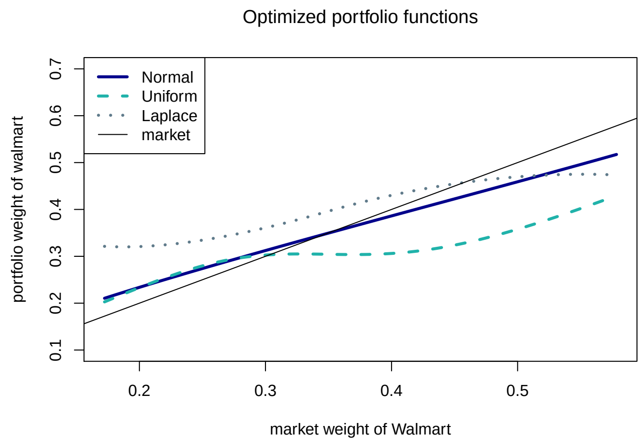

Figure 6.5: Plot of $\pi_1^{(i)}(p)$ against p_1, for $i = 1, 2, 3$. The curves are labeled using the distributions Q_i.

in the future.

Results. For each choice of Q we solve the optimal transport problem using the method of monotone rearrangement (Theorem 6.5.2). The resulting portfolio functions $\pi^{(i)}$ are plotted in Figure 6.5. The range of μ_1 shown contains more than 99.9% of the mass of P.

The features of the portfolios are consistent with our intuitions. As noted $\pi^{(1)}$ is a diversity-weighted portfolio which is quite close to the market portfolio by construction. Note that the curve intersects the market weight function around $p_1 = 0.35$. This corresponds to the median of P and is a consequence of the fact that Q_1 is symmetric about 0. Thus if P is close to reality, $\pi^{(1)}$ will overweight stock 1 half of the time and underweight stock 1 half of the time.

The portfolio $\pi^{(2)}$ consistently underweights stock 1 because Q_2 is biased towards the right, and it has the largest deviation on the range shown. Nevertheless, if we draw the curves towards the boundary points 0 and 1, the boundedness of the support of Q_2 forces $\pi^{(2)}$ to be close to the market weight near the boundary of the simplex (in the sense that the weight ratios are bounded). This is not the case for $\pi^{(1)}$ and $\pi^{(3)}$ whose distributions have

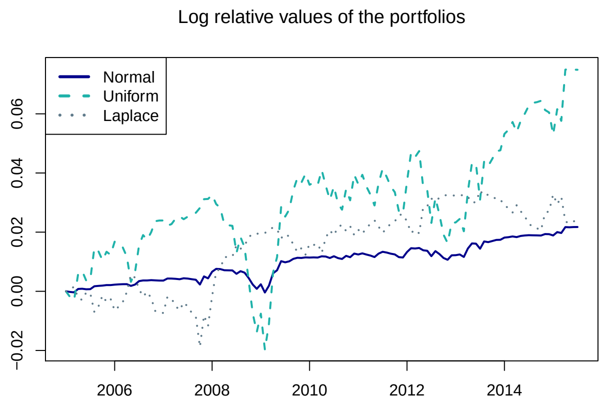

Figure 6.6: Log relative values of the portfolios $\pi^{(i)}$ in the testing period $2005 - 2015$.

unbounded supports.

As for $\pi^{(3)}$, we note that most of the curve is above the market because Q_3 has negative mean. The portfolio deviates more and more towards the boundary because the Laplace distribution has fat tails. In this case, optimal transport couples large values of $|\phi|$ with the boundary values of p which have small probability under P.

Backtesting. Finally we compute the relative values of the three portfolios with respect to the market portfolio during the testing period 2005–2015. The result is shown in Figure 6.6.

At the end of the period all three portfolios outperformed the market (by respectively 2.17%, 7.49% and 2.38%, in log scale, over the 10 year period). The amounts are not large (except perhaps for $\pi^{(2)}$), and this is mostly because the portfolios deviate only moderately from the market portfolio. While detailed analysis of the performance is beyond the scope of the paper, we note that the relative riskiness of the portfolio (with respect to the market portfolio, also called the tracking error) depends on the deviation from the market weights and hence the location and dispersion of Q. The distribution Q_2 deviates most from 0 and thus $\pi^{(2)}$ is riskier; it also has the biggest reward at the end of the period. Note that the approach of optimal transport optimizes a portfolio function over a region of Δ_n instead of picking portfolio weights period by period; it is simpler and perhaps more robust and prevents overfitting.

Chapter 7

RELATIVE CONCAVITY

Let $\pi : \Delta_n \rightarrow \overline{\Delta}_n$ be a portfolio map generated by a concave function Φ on Δ_n. In Chapter 4 we saw that its L-divergence $T(\cdot \mid \cdot)$ measures the concavity of Φ (or, rather, $\varphi = \log \Phi$) as well as the market volatility harvested by the portfolio. Using Fernholz's decomposition (Proposition 4.1.3), we can formulate conditions on $\{\mu(t)\}_{t=0}^{\infty}$ under which π is a relative arbitrage opportunity with respect to the market portfolio.

So far we have been using the market portfolilo as the benchmark. In this chapter we ask the following natural question: if we replace the market portfolio by an arbitrary functionally generated portfolio (such as the equal-weighted portfolio), can we follow the same approach and construct relative arbitrage opportunities under the conditions of diversity and sufficient volatility?

Restricting to portfolio maps that are continuously differentiable, we showed in [103] that the answer is in general no. The main idea is the following. Let τ and π be functionally generated portfolios with generating functions Ψ and Φ. If τ is 'volatility harvesting' relative to π, it must be the case that Ψ is everywhere more concave than Ψ. Thus, the concept of 'relative' volatility harvesting induces a partial ordering on functionally generated portfolios and their generating functions. We will formulate a sufficient condition for a given portfolio to be maximal with repsect to this ordering. Such portfolios cannot be beaten if only diversity and sufficient volatility are assumed. Some portfolios that are maximal in this sense are the equal-weighted portfolio and the entropy-weighted portfolio.

7.1 Maximal portfolio

In this section we define the partial ordering under which portfolio maps are compared. In Chapter 4 we used the concept of multiplicative cyclical monotonicity (MCM) to capture the idea of volatility harvesting (with respect to the market portfolio). Now we extend the definition to allow an arbitrary benchmark portfolio.

Definition 7.1.1 (Relative muliplicative cyclical monotonicity (RMCM)). Let π and τ be portfolio maps. We say that τ is multiplicatively cyclical monotone relative to π if over any discrete cycle

$$\mu(0), \mu(1), \ldots, \mu(m), \mu(m+1) = \mu(0)$$

in Δ_n we have

$$V_\tau(m+1) = \prod_{t=0}^{m} \left(\tau(\mu(t)) \cdot \frac{\mu(t+1)}{\mu(t)} \right) \geq \prod_{t=0}^{m} \left(\pi(\mu(t)) \cdot \frac{\mu(t+1)}{\mu(t)} \right) = V_\pi(m+1). \tag{7.1.1}$$

In words, τ is MCM relative to π if τ does not earn less than π over any finite cycle in Δ_n. We will show that RMCM is equivalent to the following property.

Definition 7.1.2 (Domination on compacts). Let π and τ be portfolio maps. We say that τ dominates π on compacts (written $\tau \succeq \pi$) if for any compact subset K of Δ_n, there exists a constant $C = C(\pi, \tau, K) \geq 0$ such that

$$\log \frac{V_\tau(t)}{V_\pi(t)} \geq -C, \quad t \geq 0 \tag{7.1.2}$$

for any market path $\{\mu(t)\}_{t=0}^{\infty} \subset K$.

As the next lemma shows, the partial order $\succeq$ is closely related to pseudo-arbitrage introduced in Chapter 5.

Lemma 7.1.3. *Let π and τ be portfolio maps. Suppose τ is a pseudo-arbitrage relative to π on K_j for all j, where $\{K_j\}$ is a compact exhaustion of Δ_n. Then τ dominates π on compacts.*

Proof. Condition (7.1.2) is simply condition (i) of the definition of pseudo-arbitrage. $\square$

Theorem 7.1.4. *Let π be a portfolio map generated by a concave function $\Phi : \Delta_n \to (0, \infty)$, and let τ be an arbitrary portfolio map. Then the following statements are equivalent.*

(i) τ dominates π on compacts, i.e., $\tau \succeq \pi$.

(ii) τ satisfies MCM relative to π.

(iii) τ is generated by a concave function Ψ, and the L-divergence $T_\tau(\cdot \mid \cdot)$ of (τ, Ψ) dominates $T_\pi(\cdot \mid \cdot)$ of (π, Φ) in the sense that

$$T_\tau(q \mid p) \geq T_\pi(q \mid p) \tag{7.1.3}$$

for all $p, q \in \Delta_n$.

Proof. (i) $\Rightarrow$ (ii): Suppose τ dominates π on compacts. If τ is not MCM relative to π, we can find a discrete cycle $\{\mu(t)\}_{t=0}^{m+1}$ in Δ_n over which $\eta := V_\tau(m+1)/V_\pi(m+1) < 1$. Consider the market path which goes over this cycle again and again, i.e., $\mu(t) = \mu(t + (m + 1))$ for all t. Then

$$\frac{V_\tau(k(m+1))}{V_\pi(k(m+1))} = \eta^k$$

for all $k \geq 0$ and the ratio tends to 0 as $k \to \infty$. This contradicts the hypothesis $\tau \succeq \pi$. Thus if τ dominates π on compacts then τ satisfies MCM relative to π.

(ii) $\Rightarrow$ (iii): Suppose τ satisfies MCM relative to π. Since $V_\mu(\cdot) \equiv 1$ by definition and π satisfies MCM relative to the market portfolio (by Theorem 4.2.5), τ satisfies MCM relative to the market portfolio as well. By Theorem 4.2.5 again τ has a generating function Ψ. To prove (7.1.3), let $p, q \in \Delta_n$ with $p \neq q$. Let $\{q = \mu(1), \ldots, \mu(m), \mu(m+1) = p\}$ be a partition of the line segment $[q, p]$. Setting $\mu(0) = p$, $\{\mu(t)\}_{t=0}^{m+1}$ is a cycle which starts at p, jumps to q and then returns to p along the partition. Then the RMCM inequality (7.1.1) implies that

$$\begin{aligned} &\left(1 + \frac{\tau(p)}{p} \cdot (q - p)\right) \prod_{t=1}^{m} \left(1 + \frac{\tau(\mu(t))}{\mu(t)} \cdot (\mu(t+1) - \mu(t))\right) \\ &\geq \left(1 + \frac{\pi(p)}{p} \cdot (q - p)\right) \prod_{t=1}^{m} \left(1 + \frac{\pi(\mu(t))}{\mu(t)} \cdot (\mu(t+1) - \mu(t))\right). \end{aligned} \tag{7.1.4}$$

Taking log on both sides, we have

$$\log\left(1+\frac{\tau(p)}{p}\cdot(q-p)\right)+\sum_{t=1}^{m}\log\left(1+\frac{\tau(\mu(t))}{\mu(t)}\cdot(\mu(t+1)-\mu(t))\right)$$
$$\geq \log\left(1+\frac{\pi(p)}{p}\cdot(q-p)\right)+\sum_{t=1}^{m}\log\left(1+\frac{\pi(\mu(t))}{\mu(t)}\cdot(\mu(t+1)-\mu(t))\right).$$

By the fundamental theorem of calculus for concave function and Taylor approximation, we can choose a sequence of partitions with mesh size going to zero, along which

$$\sum_{t=0}^{m}\log\left(1+\frac{\pi(\mu(t))}{\mu(t)}\cdot(\mu(t+1)-\mu(t))\right)\to\int_{\gamma}\frac{\pi}{\mu}d\mu=\log\frac{\Phi(p)}{\Phi(q)},$$
$$\sum_{t=0}^{m}\log\left(1+\frac{\tau(\mu(t))}{\mu(t)}\cdot(\mu(t+1)-\mu(t))\right)\to\int_{\gamma}\frac{\tau}{\mu}d\mu=\log\frac{\Psi(p)}{\Psi(q)},$$

where γ is the line segment from q to p. Taking the corresponding limit in (7.1.4), we obtain the desired inequality (7.1.3).

(iii) $\Rightarrow$ (i): Let $\{\mu(t)\}_{t=0}^{\infty}$ be any market path. By Proposition 4.1.3 we can write

$$\log\frac{V_\tau(t)}{V_\pi(t)}=\log\frac{\Psi(\mu(t))/\Psi(\mu(0))}{\Phi(\mu(t))/\Phi(\mu(0))}+\left(A_\tau(t)-A_\pi(t)\right),$$

where A_τ and A_π are the drift processes of τ and π respectively. By (iii), $A_\tau(t)-A_\pi(t)$ is non-decreasing in t. Since $\log\frac{\Psi(\mu(t))/\Psi(\mu(0))}{\Phi(\mu(t))/\Phi(\mu(0))}$ is bounded as long as $\mu(t)$ stays within a compact subset of Δ_n, τ dominates π on compacts. □

Having defined the partial order $\succeq$, we introduce the concept of maximal portfolio.

Definition 7.1.5 (Maximal portfolio)**.** Let $\mathcal{S}$ be a family of portfolio maps. We say that a portfolio $\pi\in\mathcal{S}$ is maximal in $\mathcal{S}$ if there is no portfolio map in $\mathcal{S}$, other than π itself, that dominates π on compacts, i.e., $\tau\in\mathcal{S}$ and $\tau\succeq\pi$ implies $\tau=\pi$.

We are interested in the case where π is functionally generated. To state the main result of this chapter we need a definition taken from [41, Definition 3.4.1].

Definition 7.1.6 (Measure of diversity)**.** We say that Φ is a measure of diversity if it is a C^2 (twice continuously differentiable) concave function on Δ_n which is symmetric, i.e.,

$$\Phi(p_1, \ldots, p_n) = \Phi(p_{\sigma(1)}, \ldots, p_{\sigma(n)})$$

for all $p \in \Delta_n$ and any permutation σ of $\{1, \ldots, n\}$.

Theorem 7.1.7. *Let π be a portfolio map generated by a measure of diversity Φ. Let $\overline{e} = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$ be the barycenter of the simplex Δ_n. If*

$$\int_0^1 \frac{1}{\Phi(te(1) + (1-t)\overline{e})^2} dt = \infty, \tag{7.1.5}$$

then π is maximal in the family of continouously differentiable portfolios maps.

This sufficient condition is satisfied by the equal and entropy weighted portfolios among many other portfolios. Note that by Theorem 7.1.4, if π is functionally generated and τ dominates π on compacts, then τ must be functionally generated. Thus we may rephrase Theorem 7.1.7 by saying that if (7.1.5) holds then π is maximal in the family of functionally generated portfolios with C^2 generating functions. A consequence of Theorem 7.1.7 is the following.

Corollary 7.1.8. *Under the setting of Theorem* 7.1.7*, suppose τ is a C^1 portfolio map not equal to π. Then there is a compact set $K \subset \Delta_n$ and a market path $\{\mu(t)\}_{t=0}^{\infty}$ taking values in K, such that the portfolio value of τ relative to π tends to zero as t tends to infinity.*

One can interpret Corollary 7.1.8 by saying that if π is maximal and $\tau \neq \pi$, it is possible to find a diverse and sufficiently volatile market in which π beats τ in the long run. In this sense, for a portfolio map π satisfying (7.1.5), it is impossible to find a portfolio map which is a relative arbitrage with respect to π in all diverse and sufficiently volatile markets.

Remark 7.1.9. The relation 'domination on compacts' refers to global properties of portfolio maps. Even if π is maximal, for a fixed subset $K \subset \Delta_n$ it may be possible to find a portfolio τ (depending on K) which beats π in the long run whenever $\{\mu(t)\}_{t=0}^{\infty} \subset K$. For example, when $n = 2$, it can be shown that the entropy-weighted portfolio beats the equal-weighted portfolio in the long run if $\{\mu(t)\}$ is sufficiently volatile and stays in a sufficiently small neighborhood of $\left(\frac{1}{2}, \frac{1}{2}\right)$. This, however, requires that K is known in advance. Maximality of π requires that there is no single τ which beats π on all compact sets $K \subset \Delta_n$.

7.2 Drift quadratic form

By Theorem 7.1.4, to prove Theorem 7.1.7 we need to study the relative concavities of generating functions, where concavity is measured by L-divergence. Since we restrict to generating functions that are twice continuously differentiable, the infinitesimal version of (7.1.3) leads to second order differential inequalities.

Definition 7.2.1. Let $\mathcal{FG}^2$ be the collection of portfolio maps generated by C^2 concave functions. We write $(\pi, \Phi) \in \mathcal{FG}^2$ to emphasize that π is generated by Φ.

Definition 7.2.2 (Drift quadratic form)**.** Let $(\pi, \Phi) \in \mathcal{FG}^2$. Its drift quadratic form, denoted by both H_π and H_Φ, is defined by

$$H_\pi(p)(v, v) := \frac{-1}{2\Phi(p)} \operatorname{Hess} \Phi(p)(v, v), \quad p \in \Delta_n, v \in T\Delta_n.$$

Here $\operatorname{Hess} \Phi$ is the Hessian of Φ regarded as a quadratic form. By definition, it is given by

$$\operatorname{Hess} \Phi(p)(v, v) = \frac{d^2}{dt^2} \Phi(p + tv)\bigg|_{t=0}. \tag{7.2.1}$$

Remark 7.2.3. In Chapter 10 we will interpret $2H_\pi(p)(\cdot, \cdot)$ as a Riemannian metric on Δ_n induced by the generating function.

Lemma 7.2.4. *The L-divergence and the drift quadratic form are concave in the portfolio weights in the following sense. Let* $(\pi^{(1)}, \Phi^{(1)}), (\pi^{(2)}, \Phi^{(2)}) \in \mathcal{FG}$*. For* $\lambda \in [0, 1]$*, let* $\pi = \lambda \pi^{(1)} + (1 - \lambda)\pi^{(2)}$ *and let* $\Phi = \left(\Phi^{(1)}\right)^\lambda \left(\Phi^{(2)}\right)^{1-\lambda}$ *be the generating function of* π*. Let* T*,* $T^{(1)}$ *and* $T^{(2)}$ *be the L-divergences of* (π, Φ)*,* $(\pi^{(1)}, \Phi^{(1)})$ *and* $(\pi^{(2)}, \Phi^{(2)})$ *respectively. Then*

$$T\left(q \mid p\right) \geq \lambda T^{(1)}\left(q \mid p\right) + (1 - \lambda) T^{(2)}\left(q \mid p\right), \quad p, q \in \Delta_n. \tag{7.2.2}$$

If $\Phi^{(1)}$ *and* $\Phi^{(2)}$ *are* C^2, *then* $H_\pi \geq \lambda H_{\pi^{(1)}} + (1-\lambda) H_{\pi^{(2)}}$ *in the sense that*

$$H_\pi(p)(v,v) \geq \lambda H_{\pi^{(1)}}(p)(v,v) + (1-\lambda) H_{\pi^{(2)}}(p)(v,v) \tag{7.2.3}$$

for all $p \in \Delta_n$ *and* $v \in T\Delta_n$.

Proof. To prove (7.2.2) we write the L-divergence $T(q \mid p)$ of a functionally generated portfolio (π, Φ) in the form

$$T(q \mid p) = \log\left(1 + \left\langle \frac{\pi(p)}{p}, q - p \right\rangle\right) - I_\pi(\gamma),$$

where $I_\pi(\gamma) = \log \Phi(q) - \log \Phi(p) = \int_\gamma \frac{\pi(p)}{p} dp$ is the line integral of the weight ratio along the line segment from p to q. Since the line integral is linear in π and the logarithm is concave, we see that $T(q \mid p)$ is concave in π. The statement for the drift quadratic form follows from the Taylor approximation (7.2.5). □

Lemma 7.2.5. *Let* $(\pi, \Phi), (\tau, \Psi) \in \mathcal{FG}^2$, *and let* T_π *and* T_τ *be their corresponding* L*-divergences. If* $\tau \succeq \pi$ *and therefore* $T_\tau(q \mid p) \geq T_\pi(q \mid p)$ *for all* $p, q \in \Delta_n$, *then* $H_\tau \geq H_\pi$ *in the sense that*

$$H_\tau(p)(v,v) \geq H_\pi(p)(v,v) \tag{7.2.4}$$

for all $p \in \Delta_n$ *and* $v \in T\Delta_n$.

Proof. The lemma follows immediately from the Taylor approximation

$$T_\pi(p + tv \mid p) = \frac{-t^2}{2\Phi(p)} \mathrm{Hess}\, \Phi(p)(v,v) + o\left(t^2\right). \tag{7.2.5}$$

where $p \in \Delta_n$, v is a tangent vector, and $t \in \mathbb{R}$ is small. □

As a consequence of Lemma 7.2.5, in order to show that a portfolio $\pi \in \mathcal{FG}^2$ is maximal in $\mathcal{FG}^2$, it is enough to show that its drift quadratic form H_π is not dominated (in the sense of (7.2.4)) by that of some other portfolio. This is the approach we use to prove Theorem 7.1.7. Simple examples show, however, that $H_\tau \geq H_\pi$ does not imply $T_\tau \geq T_\pi$.

Example 7.2.6 (Diversity-weighted portfolio). For $0 < r < 1$, the diversity-weighted portfolio π is generated by the function

$$\Phi(p) = \left(\sum_{j=1}^{n} p_j^r \right)^{\frac{1}{r}}.$$

It is easy to show that Φ is bounded below by 1. Let τ be the portfolio generated by $\Psi := \Phi - 1$. Then it can be shown that $\tau \succeq \pi$. To see this, write the L-divergence in the form

$$T_\pi \left(q \mid p \right) = \log \frac{\Phi(p) + D_{q-p}\Phi(p)}{\Phi(q)}, \quad p, q \in \Delta_n. \tag{7.2.6}$$

Then

$$T_\tau \left(q \mid p \right) = \log \frac{(\Phi(p) - 1) + D_{q-p}\Phi(p)}{\Phi(q) - 1} \geq T_\pi \left(q \mid p \right).$$

From (7.2.6), we can show that for a portfolio (π, Φ) to be maximal in $\mathcal{FG}^2$, it is necessary that the continuous extension of Φ to the closure $\overline{\Delta}_n$ (which exists by [87, Theorem 10.3]) vanishes at all the vertices $e(1), \ldots, e(n)$ (because otherwise we can subtract an affine function from Φ and make T larger). However this condition is not sufficient for π to be maximal in $\mathcal{FG}^2$.

7.3 Relative concavity

To illustrate the ideas of the proof of Theorem 7.1.7 we first give a proof of the maximality of the equal-weighted portfolio for $n = 2$.

7.3.1 Two-asset case

Proposition 7.3.1. *For $n = 2$, the equal-weighted portfolio $\pi \equiv \left(\frac{1}{2}, \frac{1}{2}\right)$ generated by the geometric mean $\Phi(p) = \sqrt{p_1 p_2}$ is maximal in $\mathcal{FG}^2$.*

Proof. Let $(\tau, \Psi) \in \mathcal{FG}^2$ be a portfolio which dominates (π, Φ) on compacts. Define $u(x) = \Phi(x, 1-x) = \sqrt{x(1-x)}$ and let $v(x) = \Psi(x, 1-x)$, $x \in (0, 1)$. Then u and v are positive C^2 concave functions on $(0, 1)$. By Theorem 7.1.4 and Lemma 7.2.5, the drift quadratic form of

τ dominates that of π. Using (7.2.1), we have the differential inequality

$$\frac{-v''(x)}{v(x)} \geq \frac{-u''(x)}{u(x)} = \frac{1}{4\left(x(1-x)\right)^2}, \quad x \in (0,1). \tag{7.3.1}$$

We claim that v also generates the equal-weighted portfolio, and so $\tau = \pi$.

We will use a transformation which amounts to a change of numéraire using $y = \log \frac{x}{1-x}$. See the binary tree model in [81, Section 4] for the motivation of this transformation and related results. Define a function $\tau_1 : (0,1) \to [0,1]$ by

$$\tau_1(x) = x + x(1-x)\frac{v'(x)}{v(x)} = x\left[1 + (1-x)(\log v)'(x)\right]. \tag{7.3.2}$$

By (4.3.3), this is the portfolio weight of stock 1 generated by v and τ_1 takes value in $[0,1]$. Let $y = \log \frac{x}{1-x}$, so $x = \frac{e^y}{1+e^y}$. Define $q : \mathbb{R} \to [0,1]$ by

$$q(y) = \tau_1(x) = \frac{e^y}{1+e^y} + \frac{e^y}{(1+e^y)^2}\frac{v'(x)}{v(x)}, \quad x = \frac{e^y}{1+e^y}, \quad y \in \mathbb{R}.$$

For the equal-weighted portfolio the corresponding portfolio weight function is identically $\frac{1}{2}$. It follows from a straightforward computation that

$$q(y)(1-q(y)) - q'(y) = \frac{-e^{2y}}{(1+e^y)^4}\frac{v''(x)}{v(x)}.$$

Now (7.3.1) can be rewritten in the form

$$q(y)(1-q(y)) - q'(y) \geq \frac{1}{4}, \quad y \in \mathbb{R}. \tag{7.3.3}$$

The proof is then completed by the following elementary lemma. □

Lemma 7.3.2. *Suppose $q : \mathbb{R} \to [0,1]$ is differentiable and $q(y)(1-q(y)) - q'(y) \geq 1/4$ on $\mathbb{R}$. Then $q(y) \equiv 1/2$.*

Proof. Since $0 \leq q(y) \leq 1$, we have

$$q'(y) \leq q(y)(1-q(y)) - \frac{1}{4} \leq \frac{1}{4} - \frac{1}{4} = 0,$$

so q is non-increasing. If $q(y_0) = q_0 < \frac{1}{2}$ for some y_0, then on $y \in [y_0, \infty]$, q must satisfy the differential inequality

$$q'(y) \leq q_0(1 - q_0) - \frac{1}{4} < 0,$$

which contradicts the fact that $q(y) \geq 0$. Similarly, if $q(y_0) = q_0 > \frac{1}{2}$ for some y_0, the same inequality is satisfied on $(-\infty, y_0]$, again a contradiction. Thus we get $q(y) \equiv \frac{1}{2}$ for all $y \in \mathbb{R}$. □

The main idea of the proof of Proposition 7.3.1 is that for a portfolio to dominate the equal-weighted portfolio π on compacts, it must be more aggressive than π everywhere on the simplex. This means buying more and more the underperforming stock at a sufficiently fast rate satisfying (7.3.3), but this is impossible to continue up to the boundary of the simplex. While there is a multi-dimensional analogue of the differential inequality (7.3.3) (see [83, Theorem 9]), we are unable to extend this proof to the multi-asset case since the market and portfolio weights can move in many directions. Instead, we will work with portfolio generating functions.

7.3.2 Relative concavity lemma

The main ingredient of the proof of Theorem 7.1.7 is the following ingenious observation taken from [25] and [26, Lemma 2] (it is called the relative convexity lemma in these references). It related to Sturm's comparison theorem for elliptic equations and can be proved by direct differentiation.

Lemma 7.3.3 (Relative concavity lemma). *[25] Let $-\infty < a < b \leq \infty$ and $c, C : [a, b) \to \mathbb{R}$ be continuous. Suppose $u, v : [a, b) \to (0, \infty)$ are C^2 and satisfy the differential equations*

$$u''(x) + c(x)u(x) = 0, \quad x \in [a, b),$$
$$v''(x) + C(x)v(x) = 0, \quad x \in [a, b).$$

Define $F : [a, b) \to [0, \infty)$ by

$$F(x) = \int_a^x \frac{1}{u(t)^2} dt, \quad x \in [a, b).$$

Let G be the inverse of F defined on $[0, \ell)$, where $\ell = \lim_{x \uparrow b} F(x)$. Then the function

$$w(y) := \frac{v(G(y))}{u(G(y))}$$

defined on $[0, \ell)$ satisfies the differential equation

$$w''(y) = -(C(x) - c(x))u(x)^4 w(y), \quad 0 \leq y < \ell, \quad x = G(y).$$

In particular, if $C(x) \geq c(x)$ on $[a, b)$, then w is concave on $[0, \ell)$.

7.4 Proof of Theorem 7.1.7

Let $\tau : \Delta_n \to \overline{\Delta}_n$ be a C^1 portfolio map which dominates π on compacts. We want to prove that $\tau = \pi$. By Theorem 7.1.4, τ is generated by a concave function $\Psi : \Delta_n \to (0, \infty)$. Since τ is C^1, by [83, Proposition 5(iii)] Ψ is C^2, so $\tau \in \mathcal{FG}^2$. Thus we may rephrase Theorem 7.1.7 by saying that π is maximal in $\mathcal{FG}^2$.

Let Ψ be a generating function of τ. Recall that $\overline{e}$ is the barycenter of Δ_n. By scaling, we may assume that $\Psi(\overline{e}) = \Phi(\overline{e})$. We will prove that Ψ equals Φ identically, so Ψ generates π and $\tau = \pi$. We divide the proof into the following steps.

Step 1 (Symmetrization). Let S_n be the set of permutations of $\{1, \ldots, n\}$. For $\sigma \in S_n$, define Ψ_σ by relabeling the coordinates, i.e.,

$$\Psi_\sigma(p) = \Psi(p_{\sigma(1)}, \ldots, p_{\sigma(n)}).$$

Since $\tau \succeq \pi$, by Lemma 7.2.5 (and relabeling the coordinates) we have $H_{\Psi_\sigma} \geq H_{\Phi_\sigma}$ for all $\sigma \in S_n$. But Φ is a measure of diversity, so $\Phi_\sigma = \Phi$ by symmetry and we have $H_{\Psi_\sigma} \geq H_\Phi$ for all $\sigma \in S_n$. Let

$$\widetilde{\Psi} = \prod_{\sigma \in S_n} (\Psi_\sigma)^{\frac{1}{n!}}$$

be the symmetrization of Ψ. By Proposition 4.3.3, $\widetilde{\Psi}$ generates the symmetrized portfolio

$$\widetilde{\tau}(p) = \frac{1}{n!} \sum_{\sigma \in S_n} \tau(p_{\sigma(1)}, \ldots, p_{\sigma(n)}), \quad p \in \Delta_n.$$

By Lemma 7.2.4, we have

$$H_{\widetilde{\Psi}} \geq \frac{1}{n!} \sum_{\sigma \in S_n} H_{\Psi_\sigma} \geq H_\Phi. \tag{7.4.1}$$

Thus $H_{\widetilde{\Psi}} \succeq H_\Phi$. Clearly $\widetilde{\Psi}$ is a measure of diversity and by symmetry it achieves its maximum at $\bar{e}$.

Step 2 ($\widetilde{\Psi} \leq \Phi$). We claim that $\widetilde{\Psi} \leq \Phi$ on Δ_n. Let $p \in \Delta_n$ and consider the one-dimensional concave functions

$$\begin{aligned} u(t) &= \Phi((1-t)\bar{e} + tp) \\ v(t) &= \widetilde{\Psi}((1-t)\bar{e} + tp) \end{aligned} \tag{7.4.2}$$

defined on $[0,1]$. We have $u(0) = v(0)$ and $u'(0) = v'(0) = 0$ since both Φ and $\widetilde{\Psi}$ achieve their maximums at $\bar{e}$. Since $H_{\widetilde{\Psi}} \geq H_\Phi$, we have

$$\frac{-v''(t)}{v(t)} \geq \frac{-u''(t)}{u(t)}, \quad t \in [0,1].$$

By the relative concavity lemma (Lemma 7.3.3),

$$w(y) = \frac{v(G(y))}{u(G(y))} \tag{7.4.3}$$

is a positive concave function on $[0, \ell]$, where $\ell = \int_0^1 \frac{1}{u(t)^2} dt$, with $w(0) = 1$ and $w'(0) = 0$ (by the quotient rule). Note that $\ell < \infty$ as Φ is continuous and positive on the line segment $[\bar{e}, p] \subset \Delta_n$. Also, it is straightforward to see that in this case the relative concavity lemma can be applied to $[0, \ell]$ instead of $[0, \ell)$. This implies that w is non-increasing and so $w(\ell) = \widetilde{\Psi}(p)/\Phi(p) \leq 1$.

Step 3 ($\widetilde{\Psi} \equiv \Phi$). Let $Z = \{p \in \Delta_n : \widetilde{\Psi}(p) = \Phi(p)\}$ and we claim that $Z = \Delta_n$. Here we follow an idea in the proof of [26, Theorem 3]. Define u and v on $[0, 1)$ by (7.4.2) with p replaced by $e(1)$. Then the function w defined as in (7.4.3) is positive and concave on $[0, \infty)$ since the integral in (7.1.5) (which defines $\ell = \int_0^1 \frac{1}{u(t)^2} dt$) diverges. Again w satisfies $w(0) = 1$ and $w'(0) = 0$. But since w is defined on an infinite interval, if $w'(y) < 0$ for some y, then

w must hit zero as w' is non-increasing by concavity. This contradicts the positivity of w, and so w is identically one on $[0,\infty)$. It follows that $\widetilde{\Psi} = \Phi$ on the line segment $[\overline{e}, e(1))$. By symmetry, Z contains the segments $[\overline{e}, e(i))$ for all i.

Next we show that the set Z is convex. Let $p, q \in Z$. Again we consider the pair of functions

$$\begin{aligned} u(t) &= \Phi((1-t)p + tq) \\ v(t) &= \widetilde{\Psi}((1-t)p + tq) \end{aligned} \tag{7.4.4}$$

on $[0,1]$. Let $\widetilde{w}(t) = \frac{v(t)}{u(t)}$, $t \in [0,1]$. By the relative concavity lemma again, we know that $\widetilde{w}$ is concave after a reparameterization. But $\widetilde{w}(t) \le 1$ by Step 2 and $\widetilde{w}$ equals one at the endpoints 0 and 1. By concavity, $\widetilde{w}$ is identically one on $[0,1]$. Hence if Z contains p and q, it also contains the line segment $[p,q]$. Now Z is a convex set containing $[\overline{e}, e(i))$ for all i. It is easy to see that Z is then the simplex Δ_n. Hence $\widetilde{\Psi}$ equals Φ identically.

Step 4 (Desymmetrization). We have shown that $\widetilde{\Psi} \equiv \Phi$, and so $H_{\widetilde{\Psi}} = H_\Phi$. By (7.4.1), we have

$$H_\Phi = H_{\widetilde{\Psi}} \ge \frac{1}{n!} \sum_{\sigma \in S_n} H_{\Psi_\sigma} \ge H_\Phi.$$

Since $H_{\Psi_\sigma} \ge H_\Phi$ for each $\sigma \in S_n$, we have $H_{\Psi_\sigma} = H_\Phi$ for all σ. In particular, taking σ to be the identity, we have $H_\Psi = H_\Phi$. It remains to show that Ψ equals Φ identically (recall that we assume $\Psi(\overline{e}) = \Phi(\overline{e})$).

Fix $i \in \{1, \ldots, n\}$ and consider

$$\begin{aligned} u(t) &= \Phi((1-t)\overline{e} + te(i)) \\ v(t) &= \Psi((1-t)\overline{e} + te(i)) \end{aligned}$$

for $t \in [0,1)$. By the argument in Step 3, if $\left(\frac{v}{u}\right)'(0) \le 0$, the integral condition (7.1.5) implies that v/u is identically one. So $\left(\frac{v}{u}\right)'(0) \le 0$ implies $\left(\frac{v}{u}\right)'(0) = 0$. For $\sigma \in S_n$ let

$$v_\sigma(t) = \Psi((1-t)\overline{e} + te(\sigma(i))).$$

Since $\widetilde{\Psi} = \Phi$, we have

$$\prod_{\sigma \in S_n} \left(\frac{v_\sigma(t)}{u(t)} \right)^{\frac{1}{n!}} = 1.$$

Taking logarithm on both sides and differentiating, we see that the average of the derivatives $\left(\frac{v}{u}\right)'(0)$ over i is 0 (recall that Φ is symmetric). Since all derivatives are non-negative by the above argument, in fact they are all 0, and so $\Psi = \Phi$ on $[\overline{e}, e(i))$ for all i.

Since the vectors $e(i) - \overline{e}$ span the plane parallel to Δ_n, the graphs of Ψ and Φ have the same tangent plane at $\overline{e}$. Since Φ achieves its maximum at $\overline{e}$, we see that Ψ achieves its maximum at $\overline{e}$ as well. Now we may apply the argument in Steps 2 and 3 to conclude that Ψ equals Φ identically on Δ_n. Thus $\tau = \pi$ and we have proved that π is maximal in $\mathcal{FG}^2$. This finishes the proof of Theorem 7.1.7.

Proof of Corollary 7.1.8. Let τ be a C^1 portfolio not equal to π. By the maximality of π, it is not the case that $\tau \succeq \pi$. By Theorem 7.1.7, τ does not satisfy MCM relative to π. Thus, there is a cycle $\{\mu(t)\}_{t=0}^{m+1}$ (with $\mu(0) = \mu(m+1)$) over which

$$\frac{V_\tau(m+1)}{V_\pi(m+1)} < 1. \tag{7.4.5}$$

Consider, as in the proof of Theorem 7.1.7, the market weight sequence which goes through this cycle again and again. Clearly $\{\mu(t)\}_{t=0}^{\infty}$ takes values in a finite set K which is compact. From (7.4.5), it is clear that $V_\tau(t)/V_\pi(t) \to 0$ as $t \to \infty$. □

It is interesting to remove the symmetry and differentiability conditions in Theorem 7.1.7.

Part II

OPTIMIZATION

Chapter 8
POINT ESTIMATION

In Part I we characterized functionally generated portfolios as the family of volatility harvesting portfolio maps (Theorem 4.2.5). A natural problem is to optimize over the family based on observed data and other information. Optimization of functionally generated portfolio is an open problem stated in [41, Problems 3.1.7–8]. To the best of our knowledge only limited progress has been made. In joint work with Soumik Pal [81] we made an attempt in the two asset case, and in Chapter 6 we studied an optimal transport problem. Also see [64] for a machine learning perspective. In this chapter we study an optimization problem analogous to nonparametric density estimation based on [104].

8.1 Nonparametric estimation of functionally generated portfolio

8.1.1 Relative value as an integral

We begin by introducing some notations which will be used in this and the next chapter. Let π be a functionally generated portfolio with generating function Φ. Given a market path $\{\mu(t)\}_{t=0}^{\infty} \subset \Delta_n$, the relative value of the portfolio π at time t is given by

$$\begin{aligned} \log V_\pi(t) &= \sum_{s=0}^{t-1} \log \left(\pi(\mu(s)) \cdot \frac{\mu(s+1)}{\mu(s)} \right) \\ &= \log \frac{\Phi(\mu(t))}{\Phi(\mu(0))} + \sum_{s=0}^{t-1} T\left(\mu(s+1) \mid \mu(s)\right). \end{aligned} \tag{8.1.1}$$

Recall that $T\left(\cdot \mid \cdot\right)$ is the L-divergence of π. Let $\ell_\pi : \Delta_n \times \Delta_n \to \mathbb{R}$ be defined by

$$\ell_\pi(p, q) = \log \left(\pi(p) \cdot \frac{q}{p} \right), \quad p, q \in \Delta_n. \tag{8.1.2}$$

In other words, $\ell_\pi(p, q)$ is the relative log return of the portfolio if the market weight jumps from p to q.

For each t, let

$$\mathbb{P}_t = \frac{1}{t}\sum_{s=0}^{t-1} \delta_{(\mu(s),\mu(s+1))} \tag{8.1.3}$$

be the empirical measure of the pair $(\mu(s), \mu(s+1))$ up to time t. It is a Borel probability measure on the product set $\Delta_n \times \Delta_n$. Now we may rewrite (8.1.1) in the form

$$\begin{aligned} \frac{1}{t}\log V_\pi(t) &= \int_{\Delta_n\times\Delta_n} \ell_\pi(p,q) d\mathbb{P}_t \\ &= \frac{1}{t}\log\frac{\Phi(\mu(t))}{\Phi(\mu(0))} + \int_{\Delta_n\times\Delta_n} T\left(q \mid p\right) d\mathbb{P}_t. \end{aligned} \tag{8.1.4}$$

For technical reasons, throughout this and the next chapter we will impose the following condition on the market path.

Assumption 8.1.1. There exists a constant $M > 0$ such that the market path $\{\mu(t)\}_{t=0}^\infty$ satisfies

$$\frac{1}{M} \leq \frac{\mu_i(t+1)}{\mu_i(t)} \leq M \tag{8.1.5}$$

for all $t \geq 0$ and $1 \leq i \leq n$. Let

$$\mathcal{S} = \left\{(p,q) \in \Delta_n \times \Delta_n : \frac{1}{M} \leq \frac{q_i}{p_i} \leq M \text{ for } 1 \leq i \leq n\right\}. \tag{8.1.6}$$

Then (8.1.5) states that $(\mu(t), \mu(t+1)) \in \mathcal{S}$ for all $t \geq 0$.

Assumption 8.1.1 states that the relative returns of the stocks are bounded and is common in the literature (see for example [28, 54, 29, 53]). We do not assume that the investor knows the value of M.

8.1.2 *Two optimization problems*

Let $\mathbb{P}$ be a Borel probability measure on $\mathcal{S}$ where $\mathcal{S}$ is defined by (8.1.6). We call $\mathbb{P}$ an intensity measure. It represents the intensity of jumps of the market path in the simplex.

Assumption 8.1.2. We assume that $\mathbb{P}$ is either discrete or is absolutely continuous with respect to the Lebesgue measure on $\mathcal{S} \times \mathcal{S}$.

Consider the following two optimization problems. The first problem maximizes the average logarithmic growth rate under $\mathbb{P}$:

$$\sup_{\pi \in \mathcal{FG}} \int_{\mathcal{S}} \ell_{\pi}(p, q) d\mathbb{P}. \tag{8.1.7}$$

The second problem maximizes the average L-divergence:

$$\sup_{\pi \in \mathcal{FG}} \int_{\mathcal{S}} T\left(q \mid p\right) d\mathbb{P}. \tag{8.1.8}$$

Remark 8.1.3. For any $\pi \in \mathcal{FG}$, its generating function Φ is differentiable almost everywhere. Thus π is almost everywhere given in terms of the unique gradient of $\varphi = \log \Phi$. This implies that the maps $\ell_{\pi}(\cdot, \cdot)$ and $T\left(\cdot \mid \cdot\right)$ are Lebesgue measurable. Assuming $\mathbb{P}$ to be either discrete or absolutely continuous, the integrals in (8.1.7) and (8.1.7) are well-defined. Note that the set of nondifferentiability depends on Φ; this is main source of technicality in this and the next chapters.

The second problem (8.1.8) is the one we studied in [103].[1] The intuitive idea is the following. The concavity of Φ allows the portfolio to harvest market volatility. If we have information about the future position of the market weight (quantified by the intensity measure $\mathbb{P}$), we should let Φ be most concave on the region where the market is most likely to fluctuate around. This leads naturally to the problem (8.1.8). Since we expect $\frac{1}{t} \log \Phi(\mu(t)) \to 0$ in typical markets, the two optimization problems should give similar results.

In this chapter we focus on the first problem (8.1.7) as it is more directly related to nonparametric density estimation.[2] Indeed, we treat the logarithmic growth rate $\frac{1}{t} \log V_{\pi}(t)$ as a quantity analogous to the likelihood function. Then a portfolio maximizing the growth rate can be regarded as the maximum likelihood estimator. Throughout the development it is helpful to keep in mind the analogy between (8.1.7) and maximum likelihood estimation

[1] In [103] we imposed the rather restrictive condition that $\mathbb{P}$ is supported on a compact subset of $\Delta_n \times \Delta_n$. This condition can be relaxed by imposing Assumption 8.1.5.

[2] The mathematical treatments of the two problems are essentially identical.

of a log-concave density. In that context, we are given a random sample $X_1, \ldots, X_N$ from a log-concave density f_0 on $\mathbb{R}^d$ (i.e., $\log f_0$ is concave). The log-concave maximum likelihood estimate (MLE) $\widehat{f}$ is the solution to

$$\max_f \sum_{j=1}^{N} \log f(X_j), \tag{8.1.9}$$

where f ranges over all log-concave densities on $\mathbb{R}^d$. It can be shown that the MLE exists almost surely (when $N \geq d+1$ and the support of f_0 has full dimension) and is unique; see [31] for precise statements of these results. We remark that (8.1.7) is conceptually more complicated than (8.1.9) because the portfolio weights correspond to selections of the superdifferential $\partial\varphi$, while (8.1.9) involves only the values of the density.

We end this section by giving some examples of intensity measures. The first example deals with infinite horizon while the second examples is concerned with a finite (but random) horizon. We will first study some theoretical properties of this abstract (unconstrained) optimization problem, and then focus on a discrete special case where numerical solutions are possible. In contrast to classical portfolio selection theory where the portfolio weights are optimized period by period, in (8.1.7) and (8.1.8) we optimize the portfolio weights over a region simultaneously.

Example 8.1.4. Suppose $\{(\mu(t-1), \mu(t))\}$ is an ergodic Markov chain on $\mathcal{S}$. Then we can take $\mathbb{P}$ to be its stationary distribution.

Example 8.1.5. We model $\{\mu(t)\}_{t=0}^{\infty}$ as a stochastic process. Let K be a subset of Δ_n containing $\mu(0)$. Let τ be the first exit time of K, i.e.,

$$\tau = \inf\{t \geq 0 : \mu(t) \notin K\}.$$

Consider the measure $\mathbb{G}$ on $K \times K$ defined by

$$\mathbb{G}(A) = \mathbb{E}\left[\sum_{t=1}^{\tau-1} 1_{\{(\mu(t-1),\mu(t)) \in A\}}\right],$$

for any Borel set $A \subset \mathcal{S}$. Suppose $\mathbb{G}(K \times K) = \mathbb{E}(\tau - 1) < \infty$, i.e., the exit time has finite expectation. Then

$$\mathbb{P}(A) = \frac{1}{\mathbb{G}(K \times K)} \mathbb{G}(A)$$

defines a Borel probability measure on $\mathcal{S}$.

8.2 Existence and uniqueness

We focus on the optimization problem (8.1.7) where $\mathbb{P}$ is a discrete or absolutely continuous intensity measure on $\mathcal{S}$ and $\mathcal{S}$ is defined by (8.1.6). By Proposition 4.3.3 the set $\mathcal{FG}$ is convex. Also, by the argument of Proposition 3.4.1, the map

$$\pi \mapsto \int_{\mathcal{S}} \ell_\pi(p, q) d\mathbb{P}$$

is concave in $\pi \in \mathcal{FG}$. Thus (8.1.7) is a convex optimization problem and we expect that it has good properties.

To formulate a uniqueness statement we need a technical condition. Given an intensity measure $\mathbb{P}$, it can be decomposed in the form

$$\mathbb{P}(dpdq) = \mathbb{P}_1(dp)\mathbb{P}(dq|p), \tag{8.2.1}$$

where $\mathbb{P}_1$ is the first marginal of $\mathbb{P}$ and $\mathbb{P}_2$ is the conditional distribution fo the second variable given p.

Assumption 8.2.1 (Support condition). Let $\mathbb{P}$ be an absolutely continuous measure on $\mathcal{S}$ with the decomposition (8.2.1). Write $\mathbb{P}_1(dp) = f(p)dp$ where f is the density of $\mathbb{P}_1$ with respect to the Lebesgue measure on Δ_n. We say that $\mathbb{P}$ satisfies the support condition if for almost all p for which $f(p) > 0$, for all $v \in T\Delta_n$ there exists $\lambda > 0$ such that $p + \lambda v$ belongs to the support of $\mathbb{P}_2(\cdot \mid p)$.

Theorem 8.2.2 (Existence and uniqueness)**.** *Consider the optimization problem* (8.1.7) *where $\mathbb{P}$ is a discrete or absolutely continuous intensity measure on $\mathcal{S}$.*

(i) The problem has an optimal solution $\widehat{\pi}$.

(ii) If $\widehat{\pi}^{(1)}$ *and* $\widehat{\pi}^{(2)}$ *are optimal solutions, then*

$$\frac{\widehat{\pi}^{(1)}(p)}{p} \cdot (q-p) = \frac{\widehat{\pi}^{(2)}(p)}{p} \cdot (q-p) \tag{8.2.2}$$

for $\mathbb{P}$*-almost all* (p,q)*. In particular, if* $\mathbb{P}(dpdq) = \mathbb{P}_1(dp)\mathbb{P}_2(dq|p)$ *is absolutely continuous with* $\mathbb{P}_1(dp) = f(p)dp$ *and satisfies the support condition, then* $\widehat{\pi}^{(1)} = \widehat{\pi}^{(2)}$ *almost everywhere on* $\{p \in \Delta_n : f(p) > 0\}$*.*

We first give some convex analytic lemmas that are useful in the proofs of Theorem 8.2.2 and other results.

Lemma 8.2.3. *Let* $p_0 \in \Delta_n$ *be fixed and let* $\mathcal{C}_0$ *be the collection of positive concave functions* Φ *on* Δ_n *satisfying* $\Phi(p_0) = 1$*. Then any sequence in* $\mathcal{C}_0$ *has a subsequence which converges locally uniformly on* Δ_n *to a function in* $\mathcal{C}_0$*.*

Proof. By [87, Theorem 10.9] it suffices to prove that $\mathcal{C}_0$ has a uniform upper bound. We first derive an upper bound in the one-dimensional case. Let f be a non-negative concave function on the real interval $[a,b]$. Let $x_0 \in (a,b)$ and suppose $f(x_0) = 1$. Let $x \in [a, x_0]$ and write $x_0 = \lambda x + (1-\lambda)b$ for some $\lambda \in [0,1]$. By concavity we have

$$1 = f(x_0) \geq \lambda f(x) + (1-\lambda) f(b) \geq \lambda f(x).$$

Thus

$$f(x) \leq \frac{1}{\lambda} = \frac{b-x}{b-x_0} \leq \frac{b-a}{b-x_0}, \quad x \in [a, x_0]$$

The case $x \in [x_0, b]$ can be handled similarly, and we get

$$f(x) \leq \frac{b-a}{\min\{|x_0 - a|, |x_0 - b|\}}, \quad x \in [a,b]. \tag{8.2.3}$$

Now let $\Phi \in \mathcal{C}_0$. Applying (8.2.3) to the restriction of Φ to line seqments in Δ_n containing p_0, we get

$$\Phi(p) \leq \frac{\mathrm{diam}(\Delta_n)}{\mathrm{dist}(p_0, \partial\Delta_n)}, \quad p \in \Delta_n,$$

where $\mathrm{diam}(\Delta_n)$ is the Euclidean diameter of Δ_n and $\mathrm{dist}(p_0, \partial\Delta_n)$ is the Euclidean distance from p_0 to the boundary of Δ_n. This completes the proof of the lemma. □

In view of Lemma 8.2.3 we give the following definition.

Definition 8.2.4. Let $\mathcal{C}_0$ be the set of all positive concave functions Φ on Δ_n satisfying the normalization $\Phi(\overline{e}) = 1$, where $\overline{e} = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$ is the barycenter of $\overline{e}$. We endow $\mathcal{C}_0$ with the topology of local uniform convergence. We define a metric on $\mathcal{C}_0$ as follows. For $m = 1, 2, \ldots$, let K_m be the compact set $\{p \in \Delta_n : p_i \geq \frac{1}{m}, 1 \leq i \leq n\}$. Then $\{K_m\}_{m=1}^{\infty}$ is a compact exhaustion of Δ_n. For $\Phi, \Psi \in \mathcal{C}_0$ we define

$$d(\Phi, \Psi) = \sum_{m=1}^{\infty} 2^{-m} \frac{\max_{p \in K_m} |\Phi(p) - \Psi(p)|}{1 + \max_{p \in K_m} |\Phi(p) - \Psi(p)|}.$$

With this metric $\mathcal{C}_0$ becomes a compact metric space.

Next we show that convergence of generating functions implies convergence of portfolio weights almost everywhere.

Lemma 8.2.5. *Let $\Phi_0 \in \mathcal{C}_0$ and $p_0 \in \Delta_n$. Let $K \subset \Delta_n$ be a compact set whose (relative) interior contains p_0. Then, for any $\epsilon > 0$, there exists $\delta > 0$ such that whenever $\Phi \in \mathcal{C}_0$, $\max_{p \in K} |\Phi(p) - \Phi_0(p)| < \delta$ and $|q - p_0| < \delta$, we have $\max_{p : |p - p_0| < \delta} |\pi(p) - \pi_0(p_0)| < \epsilon$.*

Proof. This is a uniform version of [87, Theorem 24.5]. We will proceed by contradiction. If the statement is false, there exists $\epsilon_0 > 0$ such that the following holds. For every $k \geq 1$, there exists $\Phi_k \in \mathcal{C}_0$ and $p_k \in \Delta_n$ such that

$$\max_{p \in K} |\Phi_k(p) - \Phi_0(p)| < \frac{1}{k}, \quad |p_k - p_0| < \frac{1}{k}$$

and

$$\partial \Phi_k(p_k) \subset \partial \log \Phi(p_0) + \epsilon_0 \overline{B}(0, 1),$$

where $\overline{B}(0, 1)$ is the closed unit ball. This contradicts [87, Theorem 24.5] and thus the lemma is proved. □

Using Lemma 8.2.5 and Proposition 4.3.1 we have the following corollary which is a refined version of [104, Lemma 11].

Lemma 8.2.6. *Let $\pi^{(0)}$ be a portfolio generated by $\Phi^{(0)}$. Let $p_0 \in \Delta_n$ be a point at which $\Phi^{(0)}$ is differentiable. For any $\epsilon > 0$ and any compact neighborhood K of p_0 in Δ_n, there exists $\delta > 0$ such that whenever π is generated by Φ and $\max_{p \in K} |\Phi(p) - \Phi^{(0)}(p)| < \delta$, we have $\max_{p : |p - p_0| < \delta} |\pi(p) - \pi^{(0)}(p_0)| < \delta$.*

In particular, suppose $\pi^{(k)}$ is generated by $\Phi^{(k)}$, π is generated by Φ, and $\Phi^{(k)}$ converges locally uniformly to Φ as $k \to \infty$. Then $\pi^{(k)}(p) \to \pi(p)$ almost everywhere on Δ_n.

Proof of Theorem 8.2.2. (i) The existence of an optimal solution will be proved by a compactness argument. Suppose $(\pi^{(k)}, \Phi^{(k)})$ is a maximizing sequence for the optimization problem (8.1.7). By scaling, we may assume that $\Phi^{(k)}(p_0) = 1$ where $p_0 \in \Delta_n$ is fixed. By Lemma 8.2.3, we may replace it by a subsequence such that $\Phi^{(k)}$ converges locally uniformly on Δ_n to a positive concave function Φ on Δ_n. Let π be any portfolio generated by Φ.

Case 1. $\mathbb{P}$ is discrete and has masses at $(p(j), q(j)) \in \mathcal{S}$. By a diagonal argument, we can extract a further subsequence (still denoted by $(\pi^{(k)}, \Phi^{(k)})$ such that $\lim_{k \to \infty} \pi^{(k)}(p(j))$ exists for all j. We claim that if we redefine π on $\{p(1), p(2), \ldots\}$ such that $\pi(p(j)) = \lim_{k \to \infty} \pi^{(k)}(p(j))$, then π is still generated by Φ. By definition of functionally generated portfolio (Definition 4.1.1), it suffices to show that

$$\lim_{k \to \infty} \pi(p(j)) \cdot \frac{q}{p(j)} \geq \frac{\Phi(q)}{\Phi(p(j))}$$

for all j and $q \in \Delta_n$. Since $\pi^{(k)}$ is generated by $\Phi^{(k)}$, we have

$$\lim_{k \to \infty} \pi^{(k)}(p(j)) \cdot \frac{q}{p(j)} \geq \frac{\Phi^{(k)}(q)}{\Phi^{(k)}(p(j))}.$$

Letting $k \to \infty$ proves the claim.

It follows that

$$\lim_{k \to \infty} \ell_{\pi^{(k)}}(p, q) = \lim_{k \to \infty} \log \left(\pi^{(k)}(p) \cdot \frac{q}{p} \right) = \ell_{\pi}(p, q)$$

$\mathbb{P}$-almost everywhere. By Assumption 8.1.1, we have $|\ell_{\pi^{(k)}}(p, q)| \leq \log M$ on $\mathcal{S}$. Thus, by the bounded convergence theorem, we have

$$\lim_{k \to \infty} \int_{\mathcal{S}} \ell_{\pi^{(k)}}(p, q) d\mathbb{P} = \int_{\mathcal{S}} \ell_{\pi}(p, q) d\mathbb{P}.$$

This shows that π is an optimal solution.

Case 2. $\mathbb{P}$ is absolutely continuous. By Lemma 8.2.6, $\pi^{(k)}$ converges almost everywhere to π. It follows that $\ell_{\pi^{(k)}}(p,q) \to \ell_\pi(p,q)$ almost everywhere on $\mathcal{S}$. Again by the bounded convergence theorem, we see that π is optimal.

(ii) Suppose $(\pi^{(1)}, \Phi^{(1)})$ and $(\pi^{(2)}, \Phi^{(2)})$ are optimal solutions. Define $\pi = \frac{1}{2}\pi^{(1)} + \frac{1}{2}\pi^{(2)}$ which is generated by the geometric mean $\Phi = \sqrt{\Phi^{(1)}\Phi^{(2)}}$. By concavity of log, we have

$$\log\left(\pi(p)\cdot\frac{q}{p}\right) \geq \frac{1}{2}\log\left(\pi^{(1)}(p)\cdot\frac{q}{p}\right) + \frac{1}{2}\log\left(\pi^{(2)}(p)\cdot\frac{q}{p}\right).$$

Hence π is optimal. But log is strictly concave, thus

$$\frac{\pi^{(1)}(p)}{p}\cdot(q-p) = \frac{\pi^{(2)}(p)}{p}\cdot(q-p)$$

for $\mathbb{P}$-almost all (p,q).

If $\mathbb{P}$ is absolutely continuous and satisfies sthe support condition, then for almost all p for which $f(p) > 0$, we have

$$\frac{\pi^{(1)}(p)}{p}\cdot v = \frac{\pi^{(2)}(p)}{p}\cdot v$$

for all tangent vectors of Δ_n. This implies that $\pi^{(1)}(p) = \pi^{(2)}(p)$ for these values of p. This completes the proof of the theorem. □

8.3 Consistency

Let $\mathbb{P}$ be an intensity measure on $\mathcal{S}$. Suppose $\{\mathbb{P}_N\}_{N\geq 1}$ is a sequence of discrete or absolutely continuous probability measures on $\mathcal{S}$ that converges weakly to $\mathbb{P}$. By definition, this means that

$$\lim_{N\to\infty}\int_{\mathcal{S}} f d\mathbb{P}_N = \int_{\mathcal{S}} f d\mathbb{P}$$

for all bounded continuous functions on $\mathcal{S}$. For example, one may sample i.i.d. observations $\{(p(j), q(j))\}_{j=1}^N$ from $\mathbb{P}$ and take $\mathbb{P}_N$ be the empirical distribution

$$\mathbb{P}_N = \frac{1}{N}\sum_{j=1}^N \delta_{(p(j),q(j))},$$

where $\delta_{(p(j),q(j))}$ is the point mass at $(p(j), q(j))$. From the perspective of statistical inference, the optimal portfolio $\widehat{\pi}^{(N)}$ which solves (8.1.7) for $\mathbb{P}^{(N)}$ can be regarded as a point estimate of the optimal portfolio $\widehat{\pi}$ for $\mathbb{P}$. The following result states that the estimator is consistent. See [30, Theorem 4] for an analogous statement in the context of log-concave density estimation.

Theorem 8.3.1 (Consistency)**.** *Let π be an optimal portfolio in* (8.1.7) *for $\mathbb{P}$ where $\mathbb{P}(dpdq) = \mathbb{P}_1(dp)\mathbb{P}_2(dq|p)$ is absolutely continuous with $\mathbb{P}_1(dp) = f(p)m(dp)$, supported on $\mathcal{S}$, and satisfies the support condition. Let $\{\mathbb{P}_N\}_{N=1}^{\infty}$ be a sequence of discrete or absolutely continuous probability measures on $\mathcal{S}$ converging to $\mathbb{P}$ weakly, and let $\widehat{\pi}^{(N)}$ be optimal for the measure $\mathbb{P}_N$, $N \geq 1$. Then $\widehat{\pi}^{(N)} \to \pi$ almost everywhere on $\{p : f(p) > 0\}$.*

Theorem 8.3.1 will be established by a series of lemmas that are also useful in Chapter 9. First we prove a 'strong law of large numbers' for individual elements of $\mathcal{FG}$.

Lemma 8.3.2. *Suppose $\mathbb{P}_N$ is a sequence of discrete or absolutely continuous probability measures on $\mathcal{S}$ that converges weakly to an absolutely continuous probability measure $\mathbb{P}$. Then for every $\pi \in \mathcal{FG}$ we have*

$$\lim_{N\to\infty} \int_{\mathcal{S}} \ell_\pi d\mathbb{P}_N = \int_{\mathcal{S}} \ell_\pi d\mathbb{P}. \tag{8.3.1}$$

Proof. Note that (8.3.1) does not follow directly from the definition of weak convergence because ℓ_π may have discontinuities. (But ℓ_π is bounded by Assumption 8.1.1.)

Let $\epsilon > 0$ be given. Let $\Phi \in \mathcal{C}_0$ be the generating function of π and consider the set

$$D = \{p \in \Delta_n : \Phi \text{ is differentiable at } p\}.$$

Then $\Delta_n \setminus D$ has Lebesgue measure 0. Given $\epsilon > 0$, there exists $\epsilon' > 0$ such that whenever $\pi^{(1)}, \pi^{(2)} \in \overline{\Delta}_n$ and $|\pi^{(1)} - \pi^{(2)}| < \epsilon'$, we have

$$\left|\log\left(\pi^{(1)} \cdot \frac{q}{p}\right) - \log\left(\pi^{(2)} \cdot \frac{q}{p}\right)\right| < \epsilon \tag{8.3.2}$$

for all $(p, q) \in \mathcal{S}$.

For each $p \in D$, by Lemma 8.2.6 there exists $\delta(p) > 0$ such that $B(p, \delta(p)) \subset \Delta_n$ and $|q - p| < \delta(p)$ implies

$$|\pi(p) - \pi(q)| < \epsilon'. \tag{8.3.3}$$

As a subspace of a separable metric space, D is separable. Hence, there exists a countable set $\{p_k\}_{k=1}^{\infty} \subset D$ such that

$$D \subset \bigcup_{k=1}^{\infty} B(p_k, \delta(p_k)).$$

Set $A_1 = B(p_1, \delta(p_1))$ and for $k \geq 2$ define

$$A_k = B(p_k, \delta(p_k)) \setminus \bigcup_{j=1}^{k-1} B(p_j, \delta(p_j)).$$

Then the sets $\{A_k\}$ are disjoint and

$$(D \times \Delta_n) \cap \mathcal{S} \subset \bigcup_{k=1}^{\infty} (A_k \times \Delta_n) \cap \mathcal{S}.$$

Since $\mathbb{P}((D \times \Delta_n) \cap \mathcal{S}) = 1$ by absolute continuity, by continuity of measure, there exists a positive integer k_0 such that

$$\mathbb{P}\left(\bigcup_{k=1}^{k_0} (A_k \times \Delta_n) \cap \mathcal{S}\right) > 1 - \epsilon.$$

Let

$$A_0 = \Delta_n \setminus \left(\bigcup_{k=1}^{k_0} A_k\right).$$

Then

$$\mathbb{P}((A_0 \times \Delta_n) \cap \mathcal{S}) \leq \epsilon. \tag{8.3.4}$$

Note that for $0 \leq k \leq k_0$, $(A_k \times \Delta_n) \cap \mathcal{S}$ is a $\mathbb{P}$-continuity set as it is formed by set-theoretic operations on $\mathcal{S}$ (which has piecewise smooth boundary), Δ_n and Euclidean balls. Also, by Assumption 8.1.1 $|\ell_\pi(\cdot, \cdot)|$ is bounded uniformly on $\mathcal{S}$ by $M' := \log M$. So, for each $1 \leq k \leq k_0$ the map

$$(p, q) \mapsto \ell_{\pi(p(k))}(p, q) := \log\left(\pi(p(k)) \cdot \frac{q}{p}\right)$$

is a bounded continuous function on $\mathcal{S}$

By weak convergence and Lemma 8.5.2 in Section 8.5, there exists a positive integer N_0 such that for $N \geq N_0$ we have

$$\mathbb{P}_N\left((A_0 \times \Delta_n) \cap \mathcal{S}\right) < 2\epsilon \tag{8.3.5}$$

and

$$\left|\int_{(A_k \times \Delta_n)\cap\mathcal{S}} \ell_{\pi(p(k))} d(\mathbb{P}_N - \mathbb{P})\right| < \frac{\epsilon}{k_0}. \tag{8.3.6}$$

(note that k_0 is fixed before t_0 is chosen).

Now we estimate the difference $\left|\int_{\mathcal{S}} \ell_\pi d\left(\mathbb{P}_N - \mathbb{P}\right)\right|$. We have

$$\left|\int_{\mathcal{S}} \ell_\pi d\left(\mathbb{P}_N - \mathbb{P}\right)\right| \leq \left|\sum_{k=1}^{k_0} \int_{(A_k\times\Delta_n)\cap\mathcal{S}} \ell_\pi d\left(\mathbb{P}_N - \mathbb{P}\right)\right| + \left|\int_{(A_0\times\Delta_n)\cap\mathcal{S}} \ell_\pi d\left(\mathbb{P}_N - \mathbb{P}\right)\right|. \tag{8.3.7}$$

Using the boundedness of ℓ_π, (8.3.4) and (8.3.5), the second term of (8.3.7) is bounded by $3M'\epsilon$. Now for each k, by (8.3.2), (8.3.3) and (8.3.6) we have

$$\begin{aligned}
&\left|\int_{(A_k\times\Delta_n)\cap\mathcal{S}} \ell_\pi d\left(\mathbb{P}_N - \mathbb{P}\right)\right| \\
&\leq \int_{(A_k\times\Delta_n)\cap\mathcal{S}} \left|\ell_\pi - \ell_{\pi(p_k)}\right| d\mathbb{P}_N + \int_{(A_k\times\Delta_n)\cap\mathcal{S}} \left|\ell_\pi - \ell_{\pi(p_k)}\right| d\mathbb{P} + \left|\int_{(A_k\times\Delta_n)\cap\mathcal{S}} \ell_{\pi(p_k)} d\left(\mathbb{P}_N - \mathbb{P}\right)\right| \\
&\leq \epsilon\mathbb{P}_N\left((A_k \times \Delta_n) \cap \mathcal{S}\right) + \epsilon\mathbb{P}\left((A_k \times \Delta_n) \cap \mathcal{S}\right) + \frac{\epsilon}{k_0}.
\end{aligned}$$

Summing the above inequality over k, we get

$$\left|\int_S \ell_\pi d\left(\mathbb{P}_N - \mathbb{P}\right)\right| \leq \epsilon + \epsilon + \epsilon + 3M'\epsilon, \quad t \geq t_0,$$

and the lemma is proved. □

Now we observe that the proof of Lemma 8.3.2 can be modified to yield a uniform version. Recall that $d(\Phi, \Psi)$ is the metric on $\mathcal{C}_0$ given in Definition 8.2.4.

Lemma 8.3.3. *Suppose $\mathbb{P}_N$ converges weakly to an absolutely continuous probability measure $\mathbb{P}$ on $\mathcal{S}$. Let $\pi^{(0)} \in \mathcal{FG}$ be generated by $\Phi^{(0)} \in \mathcal{C}_0$. For any $\epsilon > 0$, there exists $\delta > 0$ such that*

$$\limsup_{N\to\infty} \sup_{\pi\in\mathcal{FG}(\pi^{(0)},\delta)} \left|\int_{\mathcal{S}} \ell_\pi(p,q) d(\mathbb{P}_N - \mathbb{P})\right| < \epsilon, \tag{8.3.8}$$

where $\mathcal{FG}(\pi_0, \delta)$ is the set of all functionally generated portfolio π whose generating function $\Phi \in \mathcal{C}_0$ satisfies $d(\Phi, \Phi^{(0)}) < \delta$. In particular, we have the 'uniform strong law of large numbers'

$$\lim_{t \to \infty} \sup_{\pi \in \mathcal{FG}} \left| \int_{\mathcal{S}} \ell_\pi(p, q) d(\mathbb{P}_N - \mathbb{P}) \right| = 0. \tag{8.3.9}$$

Proof. Recall from Definition 8.2.4 that $K_m = \left\{p \in \Delta_n : p_i \geq \frac{1}{m}\right\}$. By continuity of measure, we can choose m so that

$$\mathbb{P}\left((K_m \times \Delta_n) \cap S\right) > 1 - \epsilon.$$

Since $(K_m \times \Delta_n) \cap \mathcal{S}$ is a $\mathbb{P}$-continuity set, for t sufficiently large we have

$$\left| \left(\int_{\mathcal{S}} - \int_{(K_m \times \Delta_n) \cap \mathcal{S}} \right) (\ell_\pi - \ell_{\pi_0})\, d\mathbb{P}_N \right| < 4M\epsilon,$$

where $M' = \log M$ is the upper bound of $|\ell_\pi|$ and $|\ell_{\pi_0}|$ on $\mathcal{S}$. This allows us to focus on the set $(K_m \cap \Delta_n) \cap \mathcal{S}$.

Fix $\epsilon' > 0$. By Lemma 8.2.6, for each p in the (relative) interior of K_m at which $\Phi^{(0)}$ is differentiable (call this set D_m), there exists $\delta'(p) > 0$ such that whenever $\max_{q \in K_m} |\Phi(q) - \Phi_0(q)| < \delta'(p)$ and $|q - p| < \delta'(p)$, we have $|\pi(q) - \pi^{(0)}(p)| < \epsilon'$.

As in the proof of Lemma 8.3.2, we may cover D_m by a disjoint countable union $\bigcup_{k=1}^{\infty} A_k$, where A_k is a $\mathbb{P}$-continuity set containing p_k and has diameter bounded by $\delta'(p_k)$.

Now choose a positive integer k_0 such that

$$\mathbb{P}\left(\left(\bigcup_{k=1}^{k_0} A_k \times \Delta_n \right) \cap \mathcal{S} \right) > 1 - 2\epsilon.$$

Also, choose $\delta > 0$ such that

$$d(\Phi, \Phi^{(0)}) < \delta \Rightarrow \max_{p \in K_m} |\Phi(p) - \Phi^{(0)}(p)| < \min_{1 \leq k \leq k_0} \delta'(p_k).$$

It follows that

$$\sup_{\pi \in \mathcal{FG}(\pi^{(0)}, \delta(p_k))} \sup_{p : |p - p_k| < \delta'(p_k)} \left|\pi(p) - \pi^{(0)}(p_k)\right| < \epsilon',$$

With this uniform local approximation, we may follow the same steps as the proof of Lemma 8.3.2 to prove that

$$\limsup_{N\to\infty} \sup_{\pi\in\mathcal{FG}(\pi^{(0)},\delta)} \left| \int_{\mathcal{S}} \ell_{\pi}(p,q) d\mathbb{P}_N - \int_{\mathcal{S}} \ell_{\pi^{(0)}}(p,q) d\mathbb{P}_N \right| < C\epsilon, \tag{8.3.10}$$

where $C > 0$ is a constant. Thus (8.3.8) follows by letting $\epsilon \to 0$.

Note that (8.3.8) implies that

$$\sup_{\pi\in\mathcal{FG}(\pi_0,\delta)} \left| \int_{\mathcal{S}} \ell_{\pi}(p,q) d\mathbb{P} - \int_{\mathcal{S}} \ell_{\pi^{(0)}}(p,q) d\mathbb{P} \right| \leq \epsilon.$$

Since $\mathcal{C}_0$ is compact, we may cover $\mathcal{FG}$ by finitely many sets of the form $\mathcal{FG}(\pi^{(0)}, \delta)$, and (8.3.9) follows. □

Proof of Theorem 8.3.1. Let $\Phi^{(N)} \in \mathcal{C}_0$ be the generating function of $\widehat{\pi}^{(N)}$. For any subsequence of $\{\Phi^{(N)}\}$, there exists a further subsequence $\{\Phi^{(N')}\}$ which converges locally uniformly on Δ_n. Let $\Phi^{(0)}$ be the limiting function, and let $\pi^{(0)}$ be a portfolio generated by $\Phi^{(0)}$.

By optimality of $\widehat{\pi}^{(N')}$, we have

$$\int_{\mathcal{S}} \ell_{\widehat{\pi}^{(N)}} d\mathbb{P}_N \geq \int_{\mathcal{S}} \ell_{\widehat{\pi}} d\mathbb{P}_{N'}.$$

Taking liminf on both sides as $N' \to \infty$, by Lemma 8.3.2 we have

$$\liminf_{N'\to\infty} \int_{\mathcal{S}} \ell_{\widehat{\pi}^{(N)}} d\mathbb{P}_{N'} \geq \liminf_{N'\to\infty} \int_{\mathcal{S}} \ell_{\widehat{\pi}} d\mathbb{P}_{N'} = \int_{\mathcal{S}} \ell_{\pi} d\mathbb{P}. \tag{8.3.11}$$

Since $\Phi^{(N')} \to \Phi^{(0)}$ locally uniformly, as a consequence of (8.3.10), we have

$$\lim_{N'\to\infty} \int_{\mathcal{S}} \ell_{\widehat{\pi}^{(N)}} d\mathbb{P}_{N'} = \int_{\mathcal{S}} \ell_{\pi^{(0)}} d\mathbb{P}. \tag{8.3.12}$$

By (8.3.11), it follows that $\pi^{(0)}$ is optimal for $\mathbb{P}$. By Theorem 8.2.2, we have $\pi^{(0)} = \widehat{\pi}$ almost everywhere on $\{p : f(p) > 0\}$. Since $\Phi^{(N')} \to \Phi^{(0)}$ locally uniformly, by Lemma 8.2.6 we have $\widehat{\pi}^{(N)} \to \widehat{\pi}$ almost everywhere on $\{p : f(p) > 0\}$. Since the subsequence is arbitrary, we have the statement of the theorem. □

8.4 Finite dimensional reduction

Without further constraints, the optimal portfolio weights of (8.1.7) may be highly irregular. Now we restrict to the special case where

$$\mathbb{P} = \frac{1}{N} \sum_{j=1}^{N} \delta_{(p(j),q(j))} \tag{8.4.1}$$

is a discrete measure and $(p(j), q(j)) \in \mathcal{S}$ for $j = 1, \ldots, N$. This presents no great loss of generality because in practice the market weights have finite precisions and we can choose the pairs $(p(j), q(j))$ to take values in a grid approximating $\mathcal{S}$. Moreover, from Theorem 8.3.1 we expect that when N is large the optimal solution approximates that of the continuous counter part. Consider the modified optimization problem

$$\begin{aligned} &\underset{\pi \in \mathcal{FG}}{\text{maximize}} \quad \int \ell_\pi(p, q) d\mathbb{P} \\ &\text{subject to} \quad (\pi(p(1)), \ldots, \pi(p(N))) \in C, \end{aligned} \tag{8.4.2}$$

where C is a given closed convex subset of $\left(\overline{\Delta}_n\right)^N$.[3] Some examples of C are given in Table 8.1, where each constraint is a cylinder set of the form $\{\pi(p(j)) \in C_j\}$ with C_j a closed convex set of $\overline{\Delta}_n$. It can be verified easily that the proof of Theorem 8.2.2 goes through without changes with these constraints, so (8.4.2) has an optimal solution. Moreover, if $\widehat{\pi}^{(1)}$ and $\widehat{\pi}^{(2)}$ are optimal solutions, then

$$\frac{\widehat{\pi}^{(1)}(p(j))}{p(j)} \cdot (q(j) - p(j)) = \frac{\widehat{\pi}^{(2)}(p(j))}{p(j)} \cdot (q(j) - p(j)), \quad j = 1, \ldots, N.$$

For maximum likelihood estimation of a log-concave density, it is shown in [31] that the logarithm of the MLE $\widehat{f}$ is polyhedral, i.e., $\log \widehat{f}$ is the pointwise minimum of finitely many affine functions (see [87, Section 19]). In particular, there exists a triangulation of the data points over which $\log \widehat{f}$ is piecewise affine, and this fact was used to derive an algorithm for computing $\widehat{f}$. We show that an analogous statement holds for (8.4.2). Let $D = \{p(j), q(j) : j = 1, \ldots, N\}$ be the set of data points.

[3] In (8.4.2) we may replace the function $\ell_\pi(p,)$ by $T(q \mid p)$.

Table 8.1: Examples of additional constraints imposed for $p \in \{p(1), \ldots, p(N)\}$. The parameters may be given functions of p.

Constraint	Interpretation
$a_i \leq \pi_i(p) \leq b_i$	Box constraints on portfolio weights
$m_i \leq \frac{\pi_i(p)}{p_i} \leq M_i$	Box constraints on weight ratios
$(\pi(p) - p)'\Sigma(\pi(p) - p) < \varepsilon$	Constraint on tracking error given a covariance matrix

Theorem 8.4.1. *Let* (π, Φ) *be an optimal portfolio for the problem* (8.4.2) *where* $\mathbb{P} = \frac{1}{N}\sum_{j=1}^{N} \delta_{(p(j),q(j))}$. *Let* $\overline{\Phi} : \Delta_n \to (0, \infty)$ *be the smallest positive concave function on* Δ_n *such that* $\overline{\Phi}(p) \geq \Phi(p)$ *for all* $x \in D$. *Then* $\overline{\Phi}$ *is a polyhedral positive concave function on* Δ_n *satisfying* $\overline{\Phi} \leq \Phi$ *and* $\overline{\Phi}(p) = \Phi(p)$ *for all* $p \in D$. *Moreover,* $\overline{\Phi}$ *generates a portfolio* $\overline{\pi}$ *such that* $\overline{\pi}(p(j)) = \pi(p(j))$ *for all* j. *In particular,* $(\overline{\pi}, \overline{\Phi})$ *is also optimal for the problem* (8.4.2).

Proof. It is a standard result in convex analysis that $\overline{\Phi}$ such defined is finitely generated ([87, Section 19]). By [87, Corollary 19.1.2], $\overline{\Phi}$ is a polyhedral concave function. By definition of $\overline{\Phi}$ and concavity of Φ, we have $\overline{\Phi}(p) = \Phi(p)$ for all $x \in D$ and $\overline{\Phi} \leq \Phi$. This implies that $\partial \log \Phi(p(j)) \subset \partial \log \overline{\Phi}(p(j))$ for all j. Thus $\overline{\Phi}$ generates a portfolio $\overline{\pi}$ which agrees with π on $\{p(1), \ldots, p(N)\}$. It follows that

$$\ell_\pi(p(j), q(j)) = \ell_{\overline{\pi}}(p(j), q(j))$$

for all j, and hence $(\overline{\pi}, \overline{\Phi})$ is optimal. □

Theorem 8.4.1 reduces (8.4.2) to a finite-dimensional problem which can in principle be solved numerically. We refer the reader to [83] for an empirical example for the problem (8.1.8). An interesting problem is to design efficient algorithms.

8.5 Appendix

The following lemmas are both standard results. Since we are unable to find suitable references, we will provide the proofs for completeness.

Lemma 8.5.1. *Let X be a topological space and Y be a subset of X equipped with the subspace topology. If $A \subset Y$, then*

$$\partial_X A \subset \partial_Y A \cup \partial_X Y.$$

Proof. We will argue by contradiction. Suppose $x \in \partial_X A$ and $x \notin \partial_Y A \cup \partial_X Y$.

By the definition of subspace topology and boundary, there exist neighborhoods U_1 and U_2 of x in X such that

$$(1)\ U_1 \cap Y \subset A \quad \text{or} \quad (2)\ U_1 \cap Y \subset Y \setminus A,$$

and

$$(\text{i})\ U_2 \subset Y \quad \text{or} \quad (\text{ii})\ U_2 \subset X \setminus Y.$$

We may replace U_1 and U_2 above by their intersection $U = U_1 \cap U_2$. Also, since $x \in \partial_X A$, U intersects both A and $X \setminus A$. We claim that the above statements are incompatible. We consider the following cases.

(1) and (i): Since $U \subset Y$ and $U \cap Y \subset A$, we have $U \subset A$. This contradicts the fact that U intersects $X \setminus A$.

(2) and (i): We have $U \subset Y \setminus A$. But $A \subset Y$, so U does not intersect A and we have a contradiction.

(ii): If $U \cap Y = \emptyset$, then U does not intersect A which is a contradiction. □

Lemma 8.5.2. *Suppose $\mathbb{P}_N$ converges weakly to $\mathbb{P}$. Let $f : \mathcal{S} \to \mathbb{R}$ be bounded continuous and let Y be a $\mathbb{P}$-continuity set in $\mathcal{S}$ with $\mathbb{P}(Y) > 0$. Then*

$$\lim_{N\to\infty} \int_Y f d\mathbb{P}_N = \int_Y f d\mathbb{P}.$$

Proof. Consider the measures conditioned on Y:

$$\widetilde{\mathbb{P}}_N(\cdot) = \frac{\mathbb{P}_N(\cdot \cap Y)}{\mathbb{P}_N(Y)}, \quad \widetilde{\mathbb{P}}(\cdot) = \frac{\mathbb{P}(\cdot \cap Y)}{\mathbb{P}(Y)}.$$

Since $\mathbb{P}_N(Y) \to \mathbb{P}(Y) > 0$ as A is a $\mathbb{P}$-continuity set, the measures $\widetilde{\mathbb{P}}_N$ are well defined for N sufficiently large.

We claim that $\widetilde{\mathbb{P}}_N$ converges weakly to $\widetilde{\mathbb{P}}$. This implies the statement because f is bounded continuous on Y and

$$\int_{\mathcal{S}} f d\widetilde{\mathbb{P}}_N = \frac{1}{\mathbb{P}_N(Y)} \int_Y f d\mathbb{P}_N \to \frac{1}{\mathbb{P}(Y)} \int_Y f d\mathbb{P} = \int_{\mathcal{S}} f d\widetilde{\mathbb{P}}.$$

To prove the claim, it suffices by the Portmanteau theorem to show that $\widetilde{\mathbb{P}}_N(A) \to \widetilde{\mathbb{P}}(A)$ for all $A \subset Y$ with $\widetilde{\mathbb{P}}(\partial_Y A) = \frac{1}{\mathbb{P}(Y)} \mathbb{P}\left(\partial_Y A \cap Y\right) = 0$. Note that $\partial_Y A \subset Y$, so $\mathbb{P}\left(\partial_Y A\right) = 0$. By Lemma 8.5.1, we have $\partial_{\mathcal{S}} A \subset \partial_Y A \cup \partial_{\mathcal{S}} Y$, and so $\mathbb{P}\left(\partial_{\mathcal{S}} A\right) = 0$ as Y is a $\mathbb{P}$-continuity set. Thus $A = A \cap Y$ is a $\mathbb{P}$-continuity set and we have $\mathbb{P}_N(A) \to \mathbb{P}(A)$. This completes the proof of the lemma. □

Chapter 9

UNIVERSAL PORTFOLIO AND LARGE DEVIATIONS

9.1 Introduction

9.1.1 Universal portfolio theory

Universal portfolio theory is a very active research field in mathematical finance and machine learning. Instead of giving an extensive review (for which we refer the reader to the recent survey [67] and the references therein) let us explain the main ideas of Cover's classic paper [28] which started the subject. Cover asked the following question: Without any knowledge of future stock prices, is it possible to invest in such a way that the resulting wealth is close to

$$V^*(t) = \max_{\pi \in \overline{\Delta}_n} V_\pi(t),$$

the performance of the best constant-weighted portfolio chosen with hindsight? While this seems an unrealistically ambitious goal, Cover constructed a non-anticipative sequence of portfolio weights $\widehat{\pi}(t)$ such that the resulting wealth $\widehat{V}(t)$ satisfies the universality property

$$\frac{1}{t} \log \frac{\widehat{V}(t)}{V^*(t)} \geq \frac{C}{t^{(n-1)/2}} \to 0, \tag{9.1.1}$$

where $C > 0$ is a constant, for arbitrary sequences of stock returns. Explicitly, Cover's universal portfolio is given by

$$\widehat{\pi}(t) = \frac{\int_{\overline{\Delta}_n} \pi V_\pi(t) d\pi}{\int_{\overline{\Delta}_n} V_\pi(t) d\pi}. \tag{9.1.2}$$

That is, $\widehat{\pi}(t)$ is the average of all constant-weighted portfolios weighted by their performances, and it can be shown that

$$\widehat{V}(t) = \frac{\int_{\overline{\Delta}_n} V_\pi(t) d\pi}{\int_{\overline{\Delta}_n} d\pi}.$$

This allows us to view Cover's portfolio as a buy-and-hold portfolio of all constant-weighted portfolios, where each portfolio receives the same infinitesimal wealth initially. Cover's result (9.1.1) states that the maximum and average of $V_\pi(t)$ over $\pi \in \overline{\Delta}_n$ have the same asymptotic growth rate, and can be viewed as a consequence of Laplace's method of integration and the fact that for constant-weighted portfolios the map $\pi \mapsto V_\pi(t)$ is essentially a multiple of a normal density. While numerous alternative portfolio selection algorithms have been proposed for constant-weighted and other families of portfolios, the idea of forming a wealth-weighted average underlies many of these generalizations.

9.1.2 Universal portfolio and stochastic portfolio theory

It is natural to ask if functionally generated portfolios and Cover's universal portfolio are connected in some way [48, Remark 11.7]. Recently, [18] showed that Cover's portfolio (9.1.2) is, in a generalized sense, functionally generated. With hindsight, this is not surprising since Cover's portfolio is a buy-and-hold portfolio of constant-weighted portfolios, and both buy-and-hold portfolios and constant-weighted portfolios are functionally generated (see Section 4.1). Instead, it is more interesting to think of Cover's portfolio as a market portfolio where each constituent asset is the value process of a portfolio in a family. The capital distribution then generalizes to the distribution of wealth over the portfolios. While the capital distribution of an equity market is typically stable and diverse (as mentioned in Chapter 1), this is not true for the distribution of wealth over a family of portfolios. Quite the contrary, wealth often concentrates exponentially around an optimal portfolio, and under suitable conditions this can be quantified by a pathwise large deviation principle (LDP). Moreover, we show that Cover's portfolio (9.1.1) can be generalized to the nonparametric family of functionally generated portfolios which contains the constant-weighted portfolios. Indeed, this chapter can be viewed as the Bayesian counterpart of Chapter 8.

9.1.3 Summary of main results

To formulate the main results we introduce some notations. We consider an equity market (as formulated in Chapter 2 satisfying Assumption 8.1.1. Consider a family $\{\pi_\theta\}_{\theta\in\Theta}$ of portfolio maps, where Θ is a topological index set and each π_θ is a map from Δ_n to $\overline{\Delta}_n$.

Imagine at time 0 we distribute wealth over the family according to a Borel probability measure ν_0 on Θ; we call ν_0 the *initial distribution*. The *wealth distribution* of the family $\{\pi_\theta\}_{\theta\in\Theta}$ at time t is the Borel probability measure ν_t on Θ defined by

$$\nu_t(B) = \frac{1}{\int_\Theta V_\theta(t) d\nu_0(\theta)} \int_B V_\theta(t) d\nu_0(\theta), \quad B \subset \Theta. \tag{9.1.3}$$

We will be interested in situations where the wealth distribution of the family $\{\pi_\theta\}_{\theta\in\Theta}$ concentrates exponentially around some optimal portfolio. A natural way to quantify this is to prove a *large deviation principle* (LDP) (see Figure 9.1). A standard reference of large deviation theory is [32].

Definition 9.1.1. Let $I : \Theta \to [0, \infty]$ be a lower-semicontinuous function, called the rate function. We say that the sequence $\{\nu_t\}_{t=0}^\infty$ satisfies the large deviation principle on Θ with rate I if the following statements hold.

(i) (Upper bound) For every closed set $F \subset \Theta$,

$$\limsup_{t\to\infty} \frac{1}{t} \log \nu_t(F) \leq -\inf_{\theta\in F} I(\theta).$$

(ii) (Lower bound) For every open set $G \subset \Theta$,

$$\liminf_{t\to\infty} \frac{1}{t} \log \nu_t(G) \geq -\inf_{\theta\in G} I(\theta).$$

A sufficient condition for existence of LDP is that the asymptotic growth rate

$$W(\theta) = \lim_{t\to\infty} \frac{1}{t} \log V_\theta(t) \tag{9.1.4}$$

exists for all $\theta \in \Theta$ and the map $\theta \mapsto V_\theta(t)$ is 'sufficiently regular'. As preparation, in Section 9.3 we study a simple situation where the family $\{\pi_\theta\}_{\theta\in\Theta}$, as maps from Δ_n to $\overline{\Delta}_n$, is totally bounded in the supremum metric.

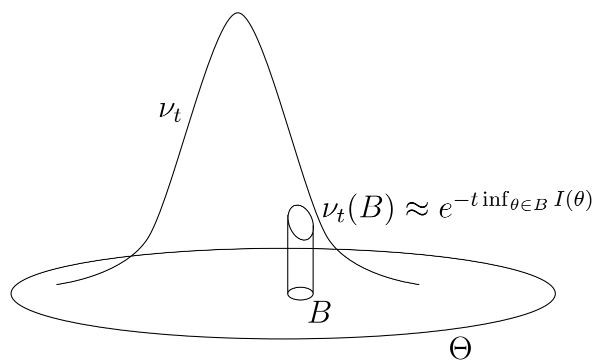

Figure 9.1: Wealth distribution of a family of portfolios.

Theorem 9.1.2. *Let $\{\pi_\theta\}_{\theta \in \Theta}$ be a totally bounded family of portfolio maps from Δ_n to $\overline{\Delta}_n$. Suppose the asymptotic growth rate $W(\theta) = \lim_{t \to \infty} \frac{1}{t} \log V_\theta(t)$ exists for all $\theta \in \Theta$ and the initial distribution ν_0 has full support on Θ. Then the sequence of wealth distribution satisfies LDP on Θ with rate function*

$$I(\theta) = W^* - W(\theta),$$

where $W^* = \sup_{\theta \in \Theta} W(\theta)$.

Consider the family $\mathcal{FG}$ of functionally generated portfolios. We endow $\mathcal{FG}$ with the topology of uniform convergence. Recall the notation

$$\mathbb{P}_t = \frac{1}{t} \sum_{s=0}^{t-1} \delta_{(\mu(s), \mu(s+1))}$$

for the empirical measure of the pair $(\mu(s), \mu(s+1))$ up to time t. Here is the main result of this chapter. Here we impose an asymptotic condition on $\{\mathbb{P}_t\}_{t=0}^{\infty}$ in the spirit of [58].

Theorem 9.1.3. *Suppose $\mathbb{P}_t$ converges weakly to an absolutely continuous Borel probability measure $\mathbb{P}$ on $\Delta_n \times \Delta_n$.*

(i) (Glivenko-Cantelli property) The asymptotic growth rate $W(\pi)$ defined by (9.1.4) exists for all $\pi \in \mathcal{FG}$. Furthermore, we have

$$\lim_{t \to \infty} \sup_{\pi \in \mathcal{FG}} \left| \frac{1}{t} \log V_\pi(t) - W(\pi) \right| = 0.$$

(ii) *(LDP) Let* ν_0 *be any initial distribution on* $\mathcal{FG}$*. Then the sequence* $\{\nu_t\}_{t=0}^{\infty}$ *of wealth distributions given by* (9.1.3) *satisfies LDP with rate*

$$I(\pi) = \begin{cases} W^* - W(\pi) & \text{if } \pi \in \operatorname{supp}(\nu_0), \\ \infty & \text{otherwise}, \end{cases}$$

where $W^* = \sup_{\pi \in \operatorname{supp}(\nu_0)} W(\pi)$.

(iii) *(Universality) There exists a probability distribution* ν_0 *on* $\mathcal{FG}$ *such that*

$$W^* = \sup_{\pi \in \operatorname{supp}(\nu_0)} W(\pi) = \sup_{\pi \in \mathcal{FG}} W(\pi)$$

for any absolutely continuous $\mathbb{P}$*. For this initial distribution, consider Cover's portfolio*

$$\widehat{\pi}(t) = \int_{\mathcal{FG}} \pi(\mu(t)) d\nu_t(\pi). \tag{9.1.5}$$

Let $\widehat{V}(t)$ *be the value of this portfolio and let* $V^*(t) = \sup_{\pi \in \mathcal{FG}} V_\pi(t)$*. Then*

$$\lim_{t\to\infty} \frac{1}{t} \log \widehat{V}(t) = \lim_{t\to\infty} \frac{1}{t} \log V^*(t) = W^*. \tag{9.1.6}$$

In particular, we have $\lim_{t\to\infty} \frac{1}{t} \log\left(\widehat{V}(t)/V^*(t)\right) = 0$.

9.2 Wealth distributions of portfolios

In this section we give more details for the terms defined in the previous section. Let Θ be an index set and suppose each $\theta \in \Theta$ is associated with a portfolio map π_θ. The individual components of π_θ will be denoted by $(\pi_{\theta,1}, \ldots, \pi_{\theta,n})$. (Sometimes we will use $\pi_1, ..., \pi_k$ to refer to a sequence of portfolios; the meaning should be clear from the context.) We are interested in the properties of $V_\theta(t) := V_{\pi_\theta}(t)$ as a function of both t and θ. To this end, we will consider an imaginary market whose basic assets are the portfolios π_θ.

We assume that Θ is a topological space and we are given a Borel probability measure ν_0 on Θ. The measure ν_0 will be called the initial distribution. The support $\operatorname{supp}(\nu_0)$ of ν_0 is the smallest closed subset F of Θ satisfying $\nu_0(F) = 1$. We say that ν_0 has full support if supp $(\nu_0) = \Theta$. Intuitively, the imaginary market is defined by distributing unit wealth over the portfolios $\{\pi_\theta\}_{\theta\in\Theta}$ according to the initial distribution ν_0, and letting the portfolios evolve. At time 0, the portfolio π_θ receives wealth $\nu_0(d\theta)$ which grows to $V_\theta(t)\nu_0(d\theta)$ at time t. Thus

$$\widehat{V}(t) := \int_\Theta V_\theta(t) d\nu_0(\theta) \tag{9.2.1}$$

is the total relative value of the imaginary market at time t. In order that (9.2.1) and related quantities (such as (9.2.2)) are well defined, we assume that the map $(p, \theta) \mapsto \pi_\theta(p)$ on $\Delta_n \times \Theta$ is jointly measurable in (p, θ). This condition usually follow immediately from the definition of the family considered, and for now we take this as granted. By Assumption 8.1.1, we have $V_\pi(t+1)/V_\pi(t) \leq M$ for any portfolio, so $V^*(t) < \infty$ and the integral in (9.2.1) is finite.

Definition 9.2.1 (Wealth distribution)**.** Given a family of portfolios $\{\pi_\theta\}_{\theta\in\Theta}$ and an initial distribution ν_0, the wealth distribution is the sequence of Borel probability measures $\{\nu_t\}_{t=0}^\infty$ defined by

$$\nu_t(B) = \frac{1}{\widehat{V}(t)} \int_B V_\theta(t) d\nu_0(\theta), \tag{9.2.2}$$

where B ranges over the measurable subsets of Θ.

Note that $\frac{d\nu_t}{d\nu_0}(\theta) = \frac{1}{\widehat{V}(t)} V_\theta(t)$. The main interest in the quantity $\widehat{V}(t)$ is the following fact first exploited by Cover in [28], where $\{\pi_\theta\}_{\theta\in\Theta}$ is the family of constant-weighted portfolios. A proof can be found in [29, Lemma 3.1].

Lemma 9.2.2 (Cover's portfolio)**.** *For each t, define the portfolio weight vector*

$$\widehat{\pi}(t) := \int_\Theta \pi_\theta(\mu(t)) d\nu_t(\theta). \tag{9.2.3}$$

Then $V_{\widehat{\pi}}(t) \equiv \widehat{V}(t)$ for all t. We call $\widehat{\pi}$ Cover's portfolio.

Let

$$V^*(t) = \sup_{\theta\in\Theta} V_\theta(t) \tag{9.2.4}$$

be the performance of the best portfolio in the family over the time interval $[0, t]$. The original goal of Cover's portfolio (9.2.3) is to track $V^*(t)$ in the sense that

$$\frac{1}{t} \log \frac{\widehat{V}(t)}{V^*(t)} \to 0 \tag{9.2.5}$$

as $t \to \infty$. If (9.2.5) holds, the portfolio $\widehat{\pi}$ performs asymptotically as good as the best portfolio in the family. In Section 9.3.3 we give a simple example to show that (9.2.5) does not always hold. This question is naturally related to the concentration of the wealth distribution and motivates our study.

Remark 9.2.3. As pointed out by many authors (see for example [29]), the construction of Cover's portfolio (9.2.3) as a wealth-weighted average has a strong Bayesian flavor. Imagine the problem of finding the best portfolio in the family $\{\pi_\theta\}_{\theta \in \Theta}$. Little is known at time 0, but from historical data, experience and insider knowledge one may form a *prior distribution* ν_0 which describes the belief of the investor. At time t, having observed the returns of the portfolios up to time t, the investor updates the belief with the posterior distribution ν_t which satisfies

$$\frac{d\nu_t}{d\nu_0}(\theta) \propto \frac{V_\theta(t)}{V_\theta(0)} = V_\theta(t).$$

This corresponds to Bayes' rule where the relative return plays the role of the likelihood. Note that this procedure is time-consistent. Namely, for $t > s$, we have

$$\frac{d\nu_t}{d\nu_s}(\theta) \propto \frac{V_\theta(t)}{V_\theta(s)}.$$

Cover's portfolio (9.2.3) is then the posterior mean of $\pi_\theta(\mu(t))$.

9.3 LDP for totally bounded families

To gain intuition about how Cover's portfolio and the wealth distribution behave for a general (possibly nonparametric) family, and to prepare for the more technical treatment of functionally generated portfolio in Section 9.4, in this section we study large deviation properties of wealth distributions where the family of portfolios is totally bounded with respect to the uniform metric.

9.3.1 Finite state

To fix ideas we begin with an even simpler situation where the sequence $\{\mu(t)\}_{t=0}^{\infty}$ takes values in a *finite* set $E \subset \Delta_n$. The finite set E may be obtained by approximating the simplex by a finite grid. Let

$$\Theta = \left\{\pi : E \to \overline{\Delta}_n\right\} = \left(\overline{\Delta}_n\right)^E$$

be the set of all portfolio maps on E. (Note that the family is indexed by the symbol π itself.) We equip Θ with the topology of uniform convergence. Since E is finite, this is the same as the topology of pointwise convergence. Note that Θ is compact and convex. Recall the notations in Section 8.1.1.

Lemma 9.3.1. *Suppose $\mathbb{P}_t$ converges weakly to a probability measure $\mathbb{P}$ on $E \times E$. Then for each $\pi \in \Theta$, the asymptotic growth rate exists and we have*

$$W(\pi) = \lim_{t \to \infty} \frac{1}{t} \log V_\pi(t) = \int_{E \times E} \ell_\pi d\mathbb{P}.$$

Moreover, there is a portfolio $\pi^* \in \Theta$ satisfying

$$W(\pi^*) = W^* := \max_{\pi \in \Theta} W(\pi). \tag{9.3.1}$$

If we write $\mathbb{P}(p, q) = \mathbb{P}_1(p)\mathbb{P}_2(q \mid p)$, where $\mathbb{P}_1$ is the first marginal and $\mathbb{P}_2$ is the conditional distribution, then

$$\pi^*(p) = \arg\max_{x \in \overline{\Delta}_n} \int_E \log\left(x \cdot \frac{q}{p}\right) \mathbb{P}_2(dq \mid p) \tag{9.3.2}$$

for all p where $\mathbb{P}_1(p) > 0$.

A portfolio satisfying (9.3.1) may be called a log-optimal portfolio.

Proof. Since $E \times E$ is a finite set, by weak convergence we have

$$W(\pi) = \lim_{t \to \infty} \int_{E \times E} \ell_\pi d\mathbb{P}_t = \int_{E \times E} \ell_\pi d\mathbb{P}.$$

Thus the asymptotic growth rate exists for all $\pi \in \Theta$. Clearly $W(\cdot)$ is a continuous function on Θ. Since Θ is compact, it has a maximizer π^*. The last statement follows from the representation

$$W(\pi) = \int_E \left(\int_E \ell_\pi(p,q) \mathbb{P}_2 (dq \mid p) \right) \mathbb{P}(dp).$$

□

The following LDP is a special case of Theorem 9.1.2 which will be proved in the next subsection.

Theorem 9.3.2 (Finite state LDP)**.** *Suppose* $\{\mu(t)\}_{t=0}^\infty$ *takes values in a finite set* $E \subset \Delta_n$*. Let* $\Theta = \left(\overline{\Delta}_n\right)^E$ *and suppose that the initial distribution* ν_0 *has full support.*

(i) The portfolio $\widehat{\pi}$ *satisfies the universality property* (9.3.3).

$$\lim_{t\to\infty} \frac{1}{t} \log \frac{\widehat{V}(t)}{V^*(t)} = 0. \tag{9.3.3}$$

(ii) If $\mathbb{P}_t$ *converges weakly to a probability measure* $\mathbb{P}$ *on* $E \times E$*, the family* $\{\nu_t\}_{t=0}^\infty$ *satisfies the large deviation principle on* Θ *with the convex rate function*

$$I(\pi) = W^* - W(\pi).$$

Remark 9.3.3. In the setting of Theorem 9.3.2(i), it is not difficult to show (see [29, Theorem 3.1]) that $V^*(t)/\widehat{V}(t)$ is bounded above by a constant multiple of t^d, where $d = |E|(n-1)$ is the 'dimension' of Θ and $|E|$ is the cardinality of E.

9.3.2 LDP for totally bounded families

In this subsection we prove Theorem 9.1.2. Now $\{\mu(t)\}_{t=0}^\infty$ is any sequence in Δ_n satisfying Assumption 8.1.1.

Let Θ be a subset of $L^\infty\left(\Delta_n, \overline{\Delta}_n\right)$, the set of functions from Δ_n to $\overline{\Delta}_n$ equipped with the supremum metric $\|\cdot\|_\infty$ (defined in terms of the Euclidean norm $|\cdot|$ on $\overline{\Delta}_n$). We endow

Θ with the induced topology, i.e., the topology of uniform convergence. A consequence of Assumption 8.1.1 is that the function $\ell_\pi(\cdot,\cdot)$ defined by (8.1.2) is bounded on $\mathcal{S}$ between $\log \frac{1}{M}$ and $\log M$, for any $\pi \in \Theta$.

We say that Θ is *totally bounded* if for any $\epsilon > 0$, there exists $\pi_1, \ldots, \pi_N \in \Theta$ with the following property: for any $\pi \in \Theta$, there exists $1 \leq j \leq N$ such that $\|\pi - \pi_j\|_\infty < \varepsilon$. The smallest such N is called the ϵ*-covering number* of Θ. Thus Θ is totally bounded if and only if the covering number is finite for all $\epsilon > 0$.

First we prove a simple lemma which generalizes [29, Theorem 3.1] to nonparametric families. In this generality it seems that a quantitative bound like (9.1.1) is out of reach.

Lemma 9.3.4. *Suppose the market satisfies Assumption* 8.1.1*. Let* Θ *be a totally bounded subset of* $L^\infty(\Delta_n, \overline{\Delta}_n)$ *and let* ν_0 *be any initial distribution on* Θ *with full support. Then Cover's portfolio* $\widehat{\pi}$ *satisfies the universality property* (9.3.3)*.*

Proof. It suffices to show that $\liminf_{t\to\infty} \frac{1}{t} \log \frac{\widehat{V}(t)}{V^*(t)} \geq 0$. Let $\epsilon > 0$ be given. Then there exists $\epsilon' > 0$ and portfolios $\pi_1, \ldots \pi_N \in \Theta$ such that the set $\{\pi_j\}_{1\leq j\leq N}$ are ϵ'-dense in Θ, and whenever $\|\pi - \pi_j\|_\infty < \epsilon'$ we have $|\ell_\pi - \ell_{\pi_j}| < \epsilon$ on $\mathcal{S}$.

For every $t > 0$, there exists a portfolio $\pi^{[t]} \in \Theta$ such that

$$\frac{1}{t} \log V_{\pi^{[t]}}(t) > \frac{1}{t} \log V^*(t) - \epsilon,$$

and from the above construction there exists $1 \leq j^{[t]} \leq N$ such that $\pi^{[t]} \in B(\pi_{j^{[t]}}, \epsilon')$, the open ball in Θ with radius ϵ' centered at $\pi_{j^{[t]}}$. Thus

$$\left|\frac{1}{t} \log V_{\pi_{j^{[t]}}}(t) - \frac{1}{t} \log V^*(t)\right| < 2\epsilon. \tag{9.3.4}$$

Moreover, if $\pi \in B_{j^{[t]}} := B(\pi_{j^{[t]}}, \epsilon')$, then

$$\left|\frac{1}{t} \log V_\pi(t) - \frac{1}{t} \log V_{\pi_{j^{[t]}}}(t)\right| < \epsilon$$

for all t. Combining these inequalities, we have

$$\begin{aligned}\frac{1}{t}\log\widehat{V}(t) &\geq \frac{1}{t}\log\int_{B_{j[t]}} V_\pi(t)\nu_0(d\pi)\\ &\geq \frac{1}{t}\log\int_{B_{j[t]}} \exp\left(t\cdot\left(\frac{1}{t}\log V^*(t) - 3\epsilon\right)\right)\nu_0(d\pi)\\ &\geq \frac{1}{t}\log V^*(t) - 3\epsilon + \frac{1}{t}\log\nu_0(B_{j[t]}).\end{aligned} \tag{9.3.5}$$

Since ν_0 has full support, we have $\lim_{t\to\infty}\min_{1\leq j\leq N}\frac{1}{t}\log\nu_0(B_j) = 0$. Hence

$$\liminf_{t\to\infty}\frac{1}{t}\log\frac{\widehat{V}(t)}{V^*(t)} \geq -3\epsilon.$$

The proof is completed by letting $\epsilon \to 0$. □

Theorem 9.1.2 is a consequence Lemma 9.3.4 and the following 'uniform strong law of large numbers'. The proof is a standard bracketing argument similar to the proof of Lemma 9.3.4 and can be found, for example, in [96, Section 3.1].

Lemma 9.3.5. *Under the hypotheses of Theorem* 9.1.2*, we have*

$$\lim_{t\to\infty}\sup_{\pi\in\Theta}\left|\frac{1}{t}\log V_\pi(t) - W(\pi)\right| = 0.$$

Proof of Theorem 9.1.2. By assumption, $W(\pi) = \lim_{t\to\infty}\frac{1}{t}\log V_\pi(t)$ exists for all $\pi\in\Theta$. Using the argument of the proof of Lemma 9.3.4, it can be shown that

$$\lim_{t\to\infty}\frac{1}{t}\log\widehat{V}(t) = \lim_{t\to\infty}\frac{1}{t}\log V^*(t) = W^*(t). \tag{9.3.6}$$

Since

$$\begin{aligned}\frac{1}{t}\log\nu_t(B) &= \frac{1}{t}\log\left(\frac{1}{\widehat{V}(t)}\int_\Theta V_\pi(t)d\nu_0(\pi)\right)\\ &= \frac{1}{t}\log\left(\int_\Theta V_\pi(t)d\nu_0(\pi)\right) - \frac{1}{t}\log\widehat{V}(t)\end{aligned}$$

and thanks to (9.3.6), to prove the LDP it suffices to show that

$$\limsup_{t\to\infty}\frac{1}{t}\log\int_F V_\pi(t)d\nu_0(\pi) \leq \sup_{\pi\in F} W(\pi) \tag{9.3.7}$$

for all closed sets F and

$$\liminf_{t\to\infty} \frac{1}{t} \log \int_G V_\pi(t) d\nu_0(\pi) \geq \inf_{\pi\in G} W(\pi) \tag{9.3.8}$$

for all open sets G. Indeed, we will show that (9.3.7) holds for all measurable sets no matter it is closed or not.

By Lemma 9.3.5, the quantity

$$R(t) = \sup_{\pi\in\Theta} \left| \frac{1}{t} \log V_\pi(t) - W(\pi) \right|$$

converges to 0 as $t \to \infty$. To prove the upper bound, write

$$\begin{aligned} \frac{1}{t} \log \int_F V_\pi(t) d\nu_0(\pi) &\leq \frac{1}{t} \log \int_F \exp\left(t\left(W(\pi) + R(t)\right)\right) d\nu_0(\pi) \\ &\leq \sup_{\pi\in F} W(\pi) + R(t) + \frac{1}{t}\nu_0(F) \\ &\leq \sup_{\pi\in F} W(\pi) + R(t), \end{aligned}$$

since $\nu_0(F) \leq 1$ for all F. Letting $t \to \infty$ establishes the upper bound for all measurable sets. The lower bound for open sets can be proved in a similar manner using the fact that ν_0 has full support. □

Proof of Theorem 9.3.2. Since E is finite, Θ is a totally bounded family of functions from E to $\overline{\Delta}_n$. (It can be extended from E to Δ_n by setting $\pi(p) = \pi_0$ for $p \notin E$, where π_0 is a fixed element of $\overline{\Delta}_n$.) The first statement then follows from Lemma 9.3.4. Moreover, by Lemma 9.3.1 the limit $W(\pi) = \lim_{t\to\infty} \int_{E\times E} \ell_\pi d\mathbb{P}_t$ exists and equals $\int_{E\times E} \ell_\pi d\mathbb{P}$ for all $\pi \in \Theta$. Thus Theorem 9.1.2 applies. It is easy to see that $I(\pi)$ is convex in π. □

9.3.3 An example

Theorem 9.1.2 assumes that the family is totally bounded in the supremum metric and the asymptotic growth rates of all portfolios exist. Now we give a simple example to show what might go wrong. First, if the family is too large and the topology is not chosen appropriately, universality may fail. Second, the LDP may be trivial even if there is an optimal portfolio.

Consider a market with two stocks (so $n = 2$). Assume that the market weight takes values in the countable set

$$E = \{p = (p_1, p_2) \in \Delta_2 : p_1, p_2 \in \mathbb{Q}\}.$$

Let $\Theta = \left(\overline{\Delta}_2\right)^E$ be the set of portfolio maps on E and equip Θ with the topology of *pointwise convergence*. Let the initial distribution ν_0 be the infinite product of the uniform distribution on $\overline{\Delta}_2$. Then ν_0 has full support on Θ.

Let $\delta > 0$ be a rational number and consider the path $\{\mu(t)\}_{t \geq 0}$ in E defined recursively by

$$\mu(0) = \left(\frac{1}{2}, \frac{1}{2}\right), \quad \mu(t+1) = \left(\frac{\mu_1(t)}{1 + \delta\mu_2(t)}, \frac{(1+\delta)\mu_2(t)}{1 + \delta\mu_2(t)}\right). \tag{9.3.9}$$

Note that

$$\frac{\mu_2(t+1)}{\mu_2(t)} = (1+\delta)\frac{\mu_1(t+1)}{\mu_1(t)} \tag{9.3.10}$$

for all $t \geq 0$ and it can be verified directly that $\{\mu(t)\}_{t \geq 0}$ satisfies Assumption 8.1.1 with $M = 1 + \delta$.

From (9.3.10), it is clear that any optimal portfolio π up to time t satisfies $\pi(\mu(s)) = (0, 1)$ for all $0 \leq s \leq t - 1$. It follows that $V^*(t) = \max_{\pi \in \Theta} V_\pi(t) = \frac{\mu_2(t)}{\mu_2(0)}$ for all t.

Proposition 9.3.6. *For the market weight path given by* (9.3.9)*, Cover's portfolio* $\widehat{\pi}$ *satisfies*

$$\widehat{V}(t) = \frac{\mu_2(t)}{\mu_2(0)}\left(1 - \frac{1}{2}\frac{\delta}{1+\delta}\right)^t.$$

In particular, we have

$$\lim_{t \to \infty} \frac{1}{t} \log \frac{\widehat{V}(t)}{V^*(t)} = \log\left(1 - \frac{1}{2}\frac{\delta}{1+\delta}\right) < 0.$$

Thus Cover's portfolio does not satisfy the universality property (9.3.3) *for all market weight paths satisfying Assumption* 8.1.1*.*

Proof. Given a portfolio $\pi \in \Theta$, we have

$$V_\pi(t) = \prod_{s=0}^{t-1}\left(\pi_1(\mu(s))\frac{\mu_1(s+1)}{\mu_1(s)} + \pi_2(\mu(s))\frac{\mu_2(s+1)}{\mu_2(s)}\right).$$

By (9.3.10), we can write

$$\begin{aligned} V_\pi(t) &= \prod_{s=0}^{t-1} \left(\frac{\mu_2(s+1)}{\mu_2(s)} \left(\frac{1}{1+\delta} \pi_1(\mu(s)) + \pi_2(\mu(s)) \right) \right) \\ &= \frac{\mu_2(t)}{\mu_2(0)} \prod_{s=0}^{t-1} \left(1 - (1 - \pi_2(\mu(s))) \frac{\delta}{1+\delta} \right). \end{aligned}$$

The value of Cover's portfolio is

$$\widehat{V}(t) = \frac{\mu_2(t)}{\mu_2(0)} \int_\Theta \prod_{s=0}^{t-1} \left(1 - (1 - \pi_2(\mu(s))) \frac{\delta}{1+\delta} \right) d\nu_0(\pi).$$

Since ν_0 is an infinite product of uniform distributions, by independence we have

$$\widehat{V}(t) = \frac{\mu_2(t)}{\mu_2(0)} \left(1 - \frac{1}{2} \frac{\delta}{1+\delta} \right)^t.$$

□

Proposition 9.3.7. *For the market weight path given by* (9.3.9)*, the wealth distributions* $\{\nu_t\}_{t=0}^\infty$ *satisfies LDP on* Θ *with the trivial rate function* $I(\pi) \equiv 0$.

Proof. Let G be any open set of Θ. Then G contains a cylinder set of the form

$$C = \{(\pi(p_1), \ldots, \pi(p_\ell)) \in B\}, \tag{9.3.11}$$

where $p_1, \ldots, p_\ell \in E$ and B is an open subset of $\left(\overline{\Delta}_2\right)^\ell$. It follows that

$$\nu_t(G) \geq \frac{1}{\left(1 - \frac{1}{2}\frac{\delta}{1+\delta}\right)^t} \int_C \prod_{s=0}^{t-1} \left(1 - (1 - \pi_2(\mu(s))) \frac{\delta}{1+\delta} \right) d\nu_0(\pi).$$

Using the fact that C puts restrictions on only finitely many coordinates, we have

$$\lim_{t\to\infty} \frac{1}{t} \log \int_C \prod_{s=0}^{t-1} \left(1 - (1 - \pi_2(\mu(s))) \frac{\delta}{1+\delta} \right) d\nu_0(\pi) = \log \left(1 - \frac{1}{2} \frac{\delta}{1+\delta} \right).$$

So $\liminf_{t\to\infty} \frac{1}{t} \log \nu_t(G) \geq 0$ and $\lim_{t\to\infty} \frac{1}{t} \log \nu_t(G) = 0$. Since the upper bound holds trivially, the LDP is proved. □

9.4 LDP for functionally generated portfolios

This section is devoted to proving Theorem 9.1.3 for functionally generated portfolios. As in Section 9.3 we impose Assumption 8.1.1 on the market weight sequence $\{\mu(t)\}_{t=0}^{\infty}$. In fact, most of the hard work has been done in Chapter 8

Let $\mathcal{FG} \subset L^{\infty}\left(\Delta_n, \overline{\Delta}_n\right)$ be the family of all functionally generated portfolios $\pi : \Delta_n \rightarrow \overline{\Delta}_n$. It is known that $\mathcal{FG}$ is convex. Indeed, if π is generated by Φ and η is generated by Ψ, then for any $\lambda \in (0,1)$ the portfolio $\lambda\pi + (1-\lambda)\eta$ (a constant-weighted portfolio of π and η is generated by the geometric mean $\Phi^{\lambda}\Psi^{1-\lambda}$. We endow $\mathcal{FG}$ with the topology of uniform convergence. The following lemma shows that the current setting is not covered by Theorem 9.1.2.

Lemma 9.4.1. *$\mathcal{FG}$ is not totally bounded. In fact, $\mathcal{FG}$ is not separable.*

Proof. We give an example for $n = 2$ and similar considerations can be applied to all dimensions. For each $\theta \in (0,1)$, let $\pi_{\theta} : \Delta_2 \rightarrow \overline{\Delta}_2$ be the portfolio

$$\pi_{\theta}(p) = \begin{cases} (1,0) & \text{if } p_1 \leq \theta \\ (0,1) & \text{if } p_1 > \theta. \end{cases}$$

It is easy to verify that each π_{θ} is functionally generated and $\{\pi_{\theta}\}_{\theta \in (0,1)}$ forms an uncountable discrete set in $\mathcal{FG}$. Hence $\mathcal{FG}$ is not separable. $\square$

Although the portfolio maps $\pi : \Delta_n \rightarrow \overline{\Delta}_n$ are the primary objects, it is technically more convenient to work with their generating functions. Recall the family $\mathcal{C}_0$ introduced in Definition 8.2.4. By Proposition 4.3.2 the generating function of a functionally generated portfolio is unique up to a positive multiplicative constant. Thus by a normalization we may assume without loss of generality that $\mathcal{C}_0$ is the set of generating functions. Although $\mathcal{FG}$ is not totally bounded, by Lemma 8.2.3 it is 'almost the same' as $\mathcal{C}_0$ which is a compact metric space. This allows us to show under appropriate conditions that $V_{\pi}(t)$ behaves nicely as a function of π when t is large.

Remark 9.4.2. It is natural to ask why we do not use the compact set $\mathcal{C}_0$ as the index space. There are three reasons for this. First, the portfolio maps $\pi : \Delta_n \to \overline{\Delta}_n$ are the primary objects for portfolio analysis, and the generating functions are only derived entities. Second, even if π_1 and π_2 have the same generating function, over a finite horizon $V_{\pi_1}(t)$ and $V_{\pi_2}(t)$ may have quite different behaviors. Third, even though for each $\Phi \in \mathcal{C}_0$ we may choose a portfolio π_Φ generated by Φ, there is no canonical way of doing this so that the maps $\Phi \mapsto \pi_\Phi$ and $\Phi \mapsto V_{\pi_\Phi}(t)$ are measurable.

Now we prove Theorem 9.1.3. First we rephrase Lemma 8.3.2 as follows.

Lemma 9.4.3. *Suppose $\mathbb{P}_t$ converges weakly to an absolutely continuous probability measure $\mathbb{P}$ on $\mathcal{S}$. Then for every $\pi \in \mathcal{FG}$ the asymptotic growth rate $W(\pi) = \lim_{t\to\infty} \frac{1}{t} \log V_\pi(t)$ exists and is given by*

$$W(\pi) = \lim_{t\to\infty} \int_{\mathcal{S}} \ell_\pi d\mathbb{P}_t = \int_{\mathcal{S}} \ell_\pi d\mathbb{P}. \tag{9.4.1}$$

Recall that $d(\Phi, \Psi)$ is the metric on $\mathcal{C}_0$ given in Definition 8.2.4. The following lemma is a restatement of Lemma 8.3.3.

Lemma 9.4.4. *Suppose $\mathbb{P}_t$ converges weakly to an absolutely continuous probability measure $\mathbb{P}$ on $\mathcal{S}$. Let $\pi_0 \in \mathcal{FG}$ be generated by $\Phi_0 \in \mathcal{C}_0$. For any $\epsilon > 0$, there exists $\delta > 0$ such that*

$$\limsup_{t\to\infty} \sup_{\pi \in \mathcal{FG}(\pi_0,\delta)} \left| \frac{1}{t} \log V_\pi(t) - \frac{1}{t} \log V_{\pi_0}(t) \right| < \epsilon,$$

where $\mathcal{FG}(\pi_0, \delta)$ is the set of all functionally generated portfolio π whose generating function $\Phi \in \mathcal{C}_0$ satisfies $d(\Phi, \Phi_0) < \delta$. In particular, we have the 'uniform strong law of large numbers'

$$\lim_{t\to\infty} \sup_{\pi\in\mathcal{FG}} \left| \frac{1}{t} \log V_\pi(t) - W(\pi) \right| = 0.$$

Recall that $\widehat{V}(t) = \int_\Theta V_\pi(t) d\nu_0(\pi)$ and $V^*(t) = \sup_{\pi\in\Theta} V_\pi(t)$.

Lemma 9.4.5. *Suppose $\mathbb{P}_t$ converges weakly to an absolutely continuous probability measure $\mathbb{P}$ on $\mathcal{S}$. Let ν_0 be any initial distribution on $\mathcal{FG}$ and $W^* = \sup_{\pi\in\mathrm{supp}(\nu_0)} W(\pi)$. Then $\lim_{t\to\infty} \frac{1}{t} \log \widehat{V}(t) = W^*$.*

Proof. For $\pi \in \mathcal{FG}$ write

$$\frac{1}{t}\log V_\pi(t) = W(\pi) + R_\pi(t)$$

where $R_\pi(t)$ is the remainder. By Lemma 9.3.5 we have $\lim_{t\to\infty}\sup_{\pi\in\mathcal{FG}}|R_\pi(t)| = 0$. Write

$$\widehat{V}(t) = \int_{\text{supp}(\nu_0)} e^{t(W(\pi)+R_\pi(t))} d\nu_0(\pi).$$

It is clear that $\limsup_{t\to\infty}\frac{1}{t}\log\widehat{V}(t) \leq W^*$. To show the other inequality, note that $W(\pi)$ is continuous in $\pi \in \mathcal{FG}$. Thus for any $\pi \in \text{supp}\,(\nu_0)$ and $\epsilon > 0$, by restricting the integral to a neighborhood of π we have $\liminf_{t\to\infty}\frac{1}{t}\log\widehat{V}(t) \geq W(\pi) - \epsilon$. Taking supremum over $\pi \in \text{supp}\,(\nu_0)$ completes the proof. □

Proof of Theorem 9.1.3. (i) This has been proved in Lemma 9.4.4.

(ii) We argue as in the proof of Theorem 9.1.2. Write

$$\nu_t(B) = \frac{1}{\widehat{V}(t)}\int_{B\cap\text{supp}\,(\nu_0)} V_\pi(t)d\nu_0(\pi).$$

Using the uniform convergence property (i), we can show that

$$\limsup_{t\to\infty}\frac{1}{t}\log\int_F V_\pi(t)d\nu_0(\pi) \leq \sup_{\pi\in F\cap\text{supp}\,(\nu_0)} W(\pi) \tag{9.4.2}$$

for any set F with $F\cap\text{supp}\,(\nu_0) \neq \emptyset$, and

$$\liminf_{t\to\infty}\frac{1}{t}\log\int_G V_\pi(t)d\nu_0(\pi) \geq \inf_{\pi\in G\cap\text{supp}\,(\nu_0)} W(\pi) \tag{9.4.3}$$

for all open sets G such that $G\cap\text{supp}\,(\nu_0) \neq \emptyset$. These inequalities and Lemma 9.4.5 imply the LDP.

(iii) Let $\{\Phi_k\}_{k=1}^\infty$ be a countable dense set in the metric space $(\mathcal{C}_0, d)$. For each k, let π_k be a portfolio generated by Φ_k. Consider an initial distribution of the form

$$\nu_0 = \sum_{k=1}^\infty \lambda_k\delta_{\pi_k}, \tag{9.4.4}$$

where $\lambda_k > 0$ and $\sum_{k=1}^\infty \lambda_k = 1$.

To see that ν_0 works, let π be any functionally generated portfolio and $\Phi \in \mathcal{C}_0$ be its generating function. Then there is a sequence $\pi_{k'}$ whose generating functions $\Phi_{k'}$ converges locally uniformly to Φ. By Lemma 9.3.5, we have $W(\pi_{k'}) \to W(\pi)$. Thus $W^* = \sup_{\pi \in \text{supp}\,(\nu_0)} W(\pi) = \sup_{\pi \in \mathcal{FG}} W(\pi)$. By Lemma 9.4.5, to establish the asymptotic universality property (9.1.6) it remains to show that

$$\lim_{t \to \infty} \frac{1}{t} \log V^*(t) = W^*,$$

but this is a direct consequence of the uniform convergence property (i). □

Similar to [58], the results in this chapter are asymptotic in nature, and in this nonparametric setting we are unable to establish quantitative bounds that hold for all finite horizons. It is desirable to obtain quantitative bounds despite of the fact that they may be too conservative to be useful in practice. Even if the underlying market process is modeled correctly, the convergence $\frac{1}{t} \log V_\pi(t) \to W(\pi)$ may take a long time and the portfolio $\widehat{\pi}(t)$ may be dominated by noise. A possible remedy is to use a smaller family or to impose regularization via a suitable prior (initial distribution). Tackling this bias-variance trade-off in dynamic portfolio selection is an interesting problem of great practical importance.

Instead of using Cover's portfolio as a wealth-weighted average, we may use other portfolio selection algorithms to construct universal portfolios for functionally generated portfolios. Perhaps the follow-the-regularized-leader (FTRL) approach of [53] can be generalized to this nonparametric set up.

A classic result in asymptotic parametric statistics is the Bernstein von-Mises Theorem which states that the posterior distribution is asymptotically normal under appropriate scaling [97, Chapter 10]. Certain generalizations to nonparametric models are possible, see for example [21]. As noted in the Introduction, for constant-weighted portfolios the map $\pi \mapsto V_\pi(t)$ is essentially a multiple of a normal density (see [58] and [29]). Hence the wealth distribution, when suitably rescaled, is approximately normal if the initial distribution is sufficiently regular. Since the family of functionally generated portfolios is convex, it can be viewed as an infinite dimensional constant-weighted family of portfolios. It is interesting to formulate and prove a version of Bernstein von-Mises Theorem in the setting of Theorem 9.1.3.

Part III

INFORMATION GEOMETRY

Chapter 10

INFORMATION GEOMETRY OF L-DIVERGENCE

Information geometry is the geometric study of manifolds of probability distributions. In this chapter we study the unit simplex Δ_n regarded as the set of probability distributions on n atoms. We show that the L-divergence of any exponentially concave function induces a remarkable geometric structure on Δ_n which has deep connections with the optimal transport problem studied in Chapter 6. This chapter is based on joint work with Soumik Pal [82]. Throughout this chapter we use the notations of Section 6.4.

10.1 Introduction

10.1.1 Motivation: Optimal frequency of rebalancing

We motivate this topic by a question of great practical interest: the optimal rebalancing frequency of portfolios. Consider a portfolio π generated by an exponentially concave function φ on Δ_n. By Fernholz's decomposition, we have

$$\log V_\pi(t) = (\varphi(\mu(t)) - \varphi(\mu(0))) + \sum_{s=0}^{t-1} T\left(\mu(t+1) \mid \mu(t)\right).$$

In this formula it is assumed that the portfolio rebalances every period (say every week). In practice, due to transaction costs and other considerations, we may want to rebalance at other frequencies. To begin with a simple case, let $0 = t_0 < t_1 < t_2$ be three time points and consider two ways of implementing the portfolio π: (i) rebalance at times t_0 and t_1 (ii) rebalance at time t_0 only. The relative values of the two implementations at time t_2 are given by

$$\begin{aligned}
\log V_\pi^{(1)}(t_2) &= (\varphi(\mu(t_2)) - \varphi(\mu(t_0))) + T\left(\mu(t_1) \mid \mu(t_0)\right) + T\left(\mu(t_2) \mid \mu(t_1)\right), \\
\log V_\pi^{(2)}(t_2) &= (\varphi(\mu(t_2)) - \varphi(\mu(t_0))) + T\left(\mu(t_2) \mid \mu(t_0)\right).
\end{aligned}$$

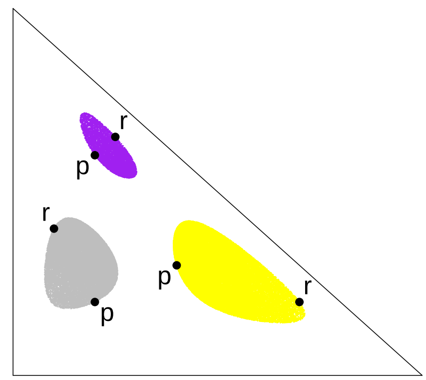

Figure 10.1: Plots of the region $\{q \in \Delta_n : T(q \mid p) + T(r \mid q) \leq T(r \mid p)\}$ for the equal-weighted portfolio $\pi = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$, for several values of p and r. Each point q on the boundary gives a 'right-angled geodesic triangle' in the sense of Theorem 10.1.1.

Letting $\mu(t_0) = p$, $\mu(t_1) = q$ and $\mu(t_2) = r$, the difference between the two values is

$$\log V_\pi^{(1)}(t_2) - \log V_\pi^{(2)}(t_2) = T(q \mid p) + T(r \mid q) - T(r \mid p).$$

This motivates the following question:

Given $p, q, r \in \Delta_n$, when is $T(q \mid p) + T(r \mid q) \leq T(r \mid p)$?

In Figure 10.1 we illustrate the idea using the equal-weighted portfolio of 3 stocks. In the figure, rebalancing at time t_1 creates extra profit if and only if q lies outside the region. This shows convincingly that rebalancing more frequently is not always better, even in the absence of transaction costs.

10.1.2 Generalized Pythagorean theorem

It turns out that the answer to the above question is given by a 'generalized Pythagorean theorem'. Let us describe the main ideas and leave the details for later. Consider the

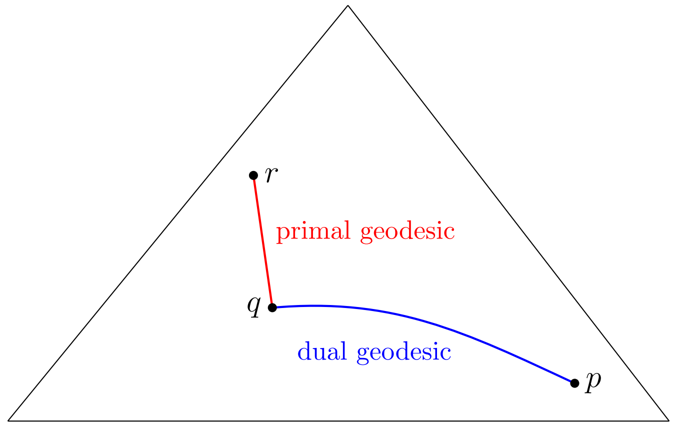

Figure 10.2: Generalized Pythagorean theorem for L-divergence.

L-divergence of the portfolio π. By Proposition 4.5.1, we have

$$T(q \mid p) = \log(1 + \nabla\varphi(p) \cdot (q - p)) - (\varphi(q) - \varphi(p))$$

for $p, q \in \Delta_n$. We only require that φ (the log generating function) is smooth and the Euclidean Hessian of e^{φ} is strictly positive definite everywhere (Assumption 6.4.3). The derivatives of this divergence defines a geometric structure on Δ_n consisting of a Riemannian metric g and a pair (∇, ∇^*) of dually coupled affine connections. These connections define via parallel transport to kinds of geodesics on Δ_n; we call them primal and dual geodesics.

With these concepts (to be made precise later) we can state the generalized Pythagorean theorem.

Theorem 10.1.1 (Generalized Pythagorean theorem)**.** *Given $p, q, r \in \Delta_n$, consider the dual geodesic joining q and p and the primal geodesic joining q and r. Then the difference*

$$T(q \mid p) + T(r \mid q) - T(r \mid p) \tag{10.1.1}$$

is positive, zero or negative depending on whether the Riemannian angle between the geodesics at q is less than, equal to, or greater than $\pi/2$ (see Figure 10.2).

The Pythagorean theorem is not the only application of this geometric structure. As we will see, this framework has deep connections with the optimal transport problem studied in Chapter 6. In particular, the difference (10.1.1) can also be given an optimal transport interpretation (Section 10.2.1). Moreover, the dual geodesics can be used to construct a 'displacement interpolation' for the transport problem.

10.1.3 Information geometry

Instead of the L-divergence, we may ask the same question for other divergences. In particular, we can consider the Bregman divergence (defined by (4.5.2)). Indeed, for Bregman divergences such questions have been studied thoroughly in information geometry. Among various divergences (such as α-divergence and f-divergence and others, see [93]), Bregman divergence plays a special role because it induces a dually flat geometry on the underlying space. First studied in the context of exponential families in statistical inference [77], it gave rise to information geometry – the geometric study of manifolds of probability distributions. Furthermore, Bregman divergence enjoys properties such as the generalized Pythagorean theorem and projection theorem which led to numerous applications. See [2, 3, 19, 62, 76] for introductions to this beautiful theory. The related concept of dual affine connection is also useful in affine differential geometry (see [34, 65, 92]).

10.1.4 Other results

We also prove other remarkable properties of the geodesics. (i) There exist explicit coordinate systems under which the primal and dual geodesics are time changes of Euclidean straight lines (Theorem 10.4.1). In other words, the new geometry is dually projectively flat. In particular, the primal geodesics are Euclidean straight lines up to time reparameterization. Moreover, the primal and dual connections have constant sectional curvature -1 with respect to the Riemannian metric, and thus satisfy an Einstein condition (Corollary 10.3.10). The primal and dual geodesics can also be constructed as time changes of Riemannian gradient flows for the functions $T(r \mid \cdot)$ and $T(\cdot \mid p)$ (Theorem 10.4.3). This is remarkable because while the geodesic equations depend only on the local properties of $T(\xi \mid \xi')$ near $\xi = \xi'$, the gradient flows are global as they involve the derivatives of $T(r \mid \cdot)$ and $T(\cdot \mid p)$. Indeed, this relation is known only for limited families of divergence including Bregman divergence and α-divergence [5].

(ii) We extend the static transport problem with cost $c(\theta, \phi) = \psi(\theta - \phi)$ to a time-dependent transport problem with a corresponding convex Lagrangian action. In Theorem 10.5.2 we show that the action minimizing curves are the (reparametrized) dual geodesics which, moreover, satisfy the intermediate time optimality condition. This allows for a consistent displacement interpolation formulation between probability measures on the unit simplex. Previously, such studies focused almost exclusively on the Wasserstein spaces corresponding to the cost functions $c(x, y) = d(x, y)^{\alpha}$ (here d is a metric on a Polish space with suitable properties and $\alpha \geq 1$). This is an immensely important topic in classical Wasserstein transport with fundamental implications in geometry, physics, probability and PDE. See [101, Chapter 7]. Our Lagrangian, although convex, is not superlinear, and, therefore, is not covered by the standard theory. However, we expect it to lead to many equally remarkable properties.

These results suggest plenty of problems for further research. Generalizing Theorem 10.1.1 to more than three points is of interest in stochastic portfolio theory. Displacement interpolation has become an extremely important topic in optimal transport theory. In particular, [68] defines Ricci curvatures on metric measure spaces in terms of displacement interpolation and displacement convexity. We expect that the displacement interpolation in this chapter will lead to a new Otto calculus ([101, Chapter 15]) and related PDEs (such as Hamilton-Jacobi equations). It appears that the Bregman divergence and L-divergence are only two of an entire family of divergences with special properties and corresponding Monge-Kantarovich optimal transport problems. In forthcoming papers we plan to study this general class. We also believe that this new information geometry will be useful in dynamic optimization problems where the cost function is multiplicative in time. Finally, it is naturally of interest to study exponential concavity on general convex domains.

10.1.5 Notation

Throughout this chapter, we assume we are given an exponentially concave function φ on Δ_n satisfying the regularity conditions in Assumption 6.4.3. We let π be the portfolio generated by φ and let $f = \varphi + \psi$ (in exponential coordinates). Recall that f is a c-concave function.

10.2 c-divergence

In this section we interpret the L-divergence from the point of view of optimal transport. Recall the notations and conventions in Section 6.4. In particular, recall the cost function

$$c(\theta, \phi) = \psi(\theta - \phi) = \log\left(1 + \sum_{i=1}^{n-1} e^{\theta_i - \phi_i}\right)$$

defined for $\theta, \phi \in \mathbb{R}^{n-1}$. By duality, we show that a pair of natural divergences on Δ_n can be defined for the c-concave functions f and f^*. Moreover, they coincide with L-divergence. It is clear that we may replace c by other cost functions. When c is the negative Euclidean inner product, we obtain the classical Bregman divergence. This covers both L-divergence and Bregman divergence under the same framework. To the best of our knowledge these definitions have not appeared in the literature. We will use the triple representation (p, θ, ϕ) for each point in Δ_n.

Definition 10.2.1 (c-divergence)**.** Consider the c-concave function f defined by (6.4.5) and its c-transform f^*.

(i) The c-divergence of f is defined by

$$D\left(p \mid p'\right) = c(\theta, \phi') - c(\theta', \phi') - (f(\theta) - f(\theta')), \quad p, p' \in \Delta_n. \tag{10.2.1}$$

(ii) The c-divergence of f^* is defined by

$$D^*\left(p \mid p'\right) = c(\theta', \phi) - c(\theta', \phi') - (f^*(\phi) - f^*(\phi')), \quad p, p' \in \Delta_n. \tag{10.2.2}$$

From Fenchel's identity (6.4.3) we see that D and D^* are non-degnereate, i.e., they vanish only on the diagonal of $\Delta_n \times \Delta_n$. The following is a generalization of the self-dual expression of Bregman divergence (see [2, Theorem 1.1]).

Proposition 10.2.2 (Self-dual expressions)**.** *We have*

$$D\left(p \mid p'\right) = c(\theta, \phi') - f(\theta) - f^*(\phi'), \tag{10.2.3}$$

$$D^*\left(p \mid p'\right) = c(\theta', \phi) - f^*(\phi) - f(\theta'). \tag{10.2.4}$$

In particular, for $p, p' \in \Delta_n$, we have $D\left(p \mid p'\right) = D^*\left(p' \mid p\right)$.

Proof. To prove (10.2.4), we use the Fenchel identity $f(\theta') + f^*(\phi') = c(\theta', \phi')$. It follows that

$$\begin{aligned} D\left(p \mid p'\right) &= c(\theta, \phi') - c(\theta', \phi') - (f(\theta) - f(\theta')) \\ &= c(\theta, \phi') - f(\theta) - f^*(\phi'). \end{aligned}$$

The proof of (10.2.4) is similar. □

Now we show that L-divergence is a special case of c-divergence when $c(\theta, \phi)$ is given by $\psi(\theta - \phi)$.

Proposition 10.2.3 (L-divergence as c-divergence)**.** *For $p, p' \in \Delta_n$ we have*

$$D\left(p \mid p'\right) = T\left(p \mid p'\right).$$

Proof. Using the primal-dual relation (6.4.6), we have

$$\psi(\theta - \phi') = \log\left(\sum_{i=1}^{n} e^{\theta_i - \theta_i' + \log \frac{\pi_i(\theta')}{\pi_n(\theta')}}\right) = \log\left(\pi(p') \cdot \frac{p}{p'}\right) - \log\left(\pi_n(p')\frac{p_n}{p_n'}\right).$$

Next, by Fenchel's identity, we have

$$f^*(\phi') = \psi(\theta' - \phi' - f(\theta') = \psi(\theta' - \phi') - \varphi(\theta') - \psi(\theta').$$

Using these identities, we compute

$$\begin{aligned} D\left(p \mid p'\right) &= \psi(\theta - \phi') - f(\theta) - f^*(\phi') \\ &= \log\left(\pi(p') \cdot \frac{p}{p'}\right) - \log\left(\pi_n(p')\frac{p_n}{p'_n}\right) \\ &\quad - (\varphi(\theta) + \psi(\theta)) - (\psi(\theta' - \phi') - \varphi(\theta') - \psi(\theta')) \\ &= \log\left(\pi(p') \cdot \frac{p}{p'}\right) - (\varphi(\theta) - \varphi(\theta')) \\ &= T\left(p \mid p'\right). \end{aligned}$$

□

For computations it is convenient to express $T\left(p \mid p'\right)$ solely in terms of either the primal or dual coordinates. We omit the details of the computations.

Lemma 10.2.4 (Coordinate representations)**.** *For $p, p' \in \Delta_n$ we have*

$$T\left(p \mid p'\right) = \log\left(\sum_{\ell=1}^{n} \pi_\ell(\theta') e^{\theta_\ell - \theta'_\ell}\right) - \left(f(\theta) - f(\theta')\right),$$

$$T\left(p \mid p'\right) = \log\left(\sum_{\ell=1}^{n} \pi_\ell(\phi) e^{\phi_\ell - \phi'_\ell}\right) - \left(f^*(\phi') - f^*(\phi)\right).$$

10.2.1 Transport interpretation of the generalized Pythagorean theorem

Using Proposition 10.2.3 we can give a transport interpretation of the expression (10.1.1) in the generalized Pythagorean theorem (Theorem 10.1.1). Let $p, q, r \in \Delta_n$ be given. Let $(\theta^{(j)}, \phi^{(j)})_{1 \leq j \leq 3}$ be the primal and dual coordinates of p, q and r respectively. Given the exponentially concave function φ, the coupling $(\theta, \phi = \nabla f^c(\theta))$ is c-cyclical monotone. Hence coupling $\theta^{(j)}$ with $\phi^{(j)}$ is optimal.

Consider two (suboptimal) perturbations of the optimal coupling:

(i) (Cyclical perturbation) Couple $\theta^{(1)}$ with $\phi^{(3)}$, $\theta^{(2)}$ with $\phi^{(1)}$, and $\theta^{(3)}$ with $\phi^{(2)}$. The associated cost is

$$c(\theta^{(1)}, \phi^{(3)}) + c(\theta^{(2)}, \phi^{(1)}) + c(\theta^{(3)}, \phi^{(2)}).$$

(ii) (Transposition) Couple $\theta^{(1)}$ with $\phi^{(3)}$, $\theta^{(3)}$ with $\phi^{(1)}$, and keep the coupling $(\theta^{(2)}, \phi^{(2)})$. The associated cost is

$$c(\theta^{(1)}, \phi^{(3)}) + c(\theta^{(3)}, \phi^{(1)}) + c(\theta^{(2)}, \phi^{(2)}).$$

Now we ask which perturbation has lower cost. The difference (i) $-$ (ii) is

$$c(\theta^{(2)}, \phi^{(1)}) + c(\theta^{(3)}, \phi^{(2)}) - c(\theta^{(3)}, \phi^{(1)}) - c(\theta^{(2)}, \phi^{(2)}).$$

By Proposition 10.2.3, this is nothing but the difference $T(q \mid p) + T(r \mid q) - T(r \mid p)$. Thus the generalized Pythagorean theorem gives an information geometric characterization of the relative costs of the two perturbations.

10.3 Geometry of L-divergence

In this section we derive the geometric structure induced by a given L-divergence $T(\cdot \mid \cdot)$. As always we impose the regularity conditions in Assumption 6.4.3. Using the primal and dual coordinate systems, we compute explicitly the Riemannian metric g, the primal connection ∇ and the dual connection ∇^*. We call (g, ∇, ∇^*) the induced geometric structure. An important fact in information geometry is that the Levi-Civita connection $\nabla^{(0)}$ is not necessarily the right one to use. Nevertheless, by duality we always have $\nabla^{(0)} = \frac{1}{2}(\nabla + \nabla^*)$.

10.3.1 Preliminaries

For differential geometric concepts such as Riemannian metric and affine connection we refer the reader to [2, Chapters 5] whose notations are consistent with ours. For computational convenience we define the geometric structure in terms of coordinate representations. It can be shown that the geometric structure does not depend on the choice of coordinates, and we refer the reader to [19, Chapter 11] for intrinsic definitions. The following definition (which holds for a general divergence on a manifold) is taken from [2, Section 6.2].

Definition 10.3.1 (Induced geometric structure)**.** Given a coordinate system $\xi = (\xi_1, \ldots, \xi_{n-1})$ of Δ_n, the coefficients of the geometric structure (g, ∇, ∇^*) are given as follows.

(i) The Riemannian metric is given by

$$g_{ij}(\xi) = -\frac{\partial}{\partial \xi_i}\frac{\partial}{\partial \xi'_j} T(\xi \mid \xi')\Big|_{\xi=\xi'}, \quad i,j = 1,\ldots,n-1. \tag{10.3.1}$$

By Assumption 6.4.3 the matrix $(g_{ij}(\xi))$ is strictly positive definite. The Riemannian inner product and length are denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ respectively.

(ii) The primal connection is given by

$$\Gamma_{ijk}(\xi) = -\frac{\partial}{\partial \xi_i}\frac{\partial}{\partial \xi_j}\frac{\partial}{\partial \xi'_k} T(\xi \mid \xi')\Big|_{\xi=\xi'}, \quad i,j,k = 1,\ldots,n-1. \tag{10.3.2}$$

(iii) The dual connection is given by

$$\Gamma^*_{ijk}(\xi) = -\frac{\partial}{\partial \xi_k}\frac{\partial}{\partial \xi'_i}\frac{\partial}{\partial \xi'_j} T(\xi \mid \xi')\Big|_{\xi=\xi'}, \quad i,j,k = 1,\ldots,n-1. \tag{10.3.3}$$

For a general divergence the above definitions were first introduced by Eguchi in [35, 36]. If we define the dual divergence by $T^*(p \mid p') = T(p' \mid p)$, the dual connection of T is the primal connection of T^*. The primal and dual connections are defined in such a way that they are dual to each other [2, Theorem 6.2]. While any divergence induces a geometric structure (g, ∇, ∇^*), it may not enjoy nice properties. For the geometric structure induced by any Bregman divergence, it can be shown that the Riemann-Christoffel curvatures of ∇ and ∇^* both vanish. Thus we say that the induced geometry is dually flat [2, Chapter 1]. We will show that L-divergence gives rise to a different geometry with many interesting properties.

10.3.2 Notations

We begin by clarifying the notations. Following our convention, we write $T(p \mid p') = T(\theta \mid \theta') = T(\phi \mid \phi')$ depending on the coordinate system used. The primal and dual coordinate representations have been computed in Lemma 10.2.4.

The Riemannian metric will be computed using both the primal and dual coordinate systems. To be explicit about the coordinate system we use $g_{ij}(\theta)$ to denote its coefficients in primal coordinates, and $g^*_{ij}(\phi)$ for its coefficients in dual coordinates:

$$g_{ij}(\theta) := -\left.\frac{\partial^2}{\partial\theta_i\partial\theta'_j} T\left(\theta \mid \theta'\right)\right|_{\theta=\theta'}, \quad g^*_{ij}(\phi) := -\left.\frac{\partial^2}{\partial\phi_i\partial\phi'_j} T\left(\phi \mid \phi'\right)\right|_{\phi=\phi'}.$$

The inverses of the matrices $(g_{ij}(\theta))$ and $\left(g^*_{ij}(\phi)\right)$ are denoted by $(g^{ij}(\theta))$ and $(g^{*ij}(\phi))$ respectively.

The primal connection ∇ will be computed using the primal coordinate system:

$$\Gamma_{ijk}(\theta) := -\left.\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta'_k} T\left(\theta \mid \theta'\right)\right|_{\theta=\theta'}, \quad \Gamma^k_{ij}(\theta) := \sum_{m=1}^{n-1} \Gamma_{ijm}(\theta) g^{mk}(\theta).$$

The dual connection ∇^* will be computed using the dual coordinate system:

$$\Gamma^*_{ijk}(\phi) := -\left.\frac{\partial^3}{\partial\phi_k\partial\phi'_i\partial\phi'_j} T\left(\phi \mid \phi'\right)\right|_{\phi=\phi'}, \quad \Gamma^{*k}_{ij}(\phi) := \sum_{m=1}^{n-1} \Gamma^*_{ijm}(\phi) g^{*mk}(\phi).$$

The following notations are useful. For $1 \leq i \leq n$ we define

$$\Pi_i(\theta, \theta') := \frac{\pi_i(\theta') e^{\theta_i - \theta'_i}}{\sum_{\ell=1}^n \pi_\ell(\theta') e^{\theta_\ell - \theta'_\ell}}, \quad \Pi^*_i(\phi, \phi') := \frac{\pi_i(\phi) e^{\phi_i - \phi'_i}}{\sum_{\ell=1}^n \pi_\ell(\phi) e^{\phi_\ell - \phi'_\ell}}. \tag{10.3.4}$$

As always we adopt the convention $\theta_n = \theta'_n = \phi_n = \phi'_n = 0$. Note that $\Pi_i(\theta, \theta')$ involves the portfolio at θ' (the second variable) while $\Pi^*_i(\phi, \phi')$ involves the portfolio at ϕ (the first variable). The partial derivatives of Π_i and Π^*_i are given in the next lemma and can be verified by direct differentiation. We let δ_{ij} be the Kronecker delta and $\delta_{ijk} = \delta_{ij}\delta_{jk}$.

Lemma 10.3.2 (Derivatives of Π_i and Π^*_i)**.**

(i) For $1 \leq i \leq n$ and $1 \leq j \leq n-1$, we have

$$\begin{aligned}
\frac{\partial \Pi_i(\theta, \theta')}{\partial \theta_j} &= \Pi_i(\theta, \theta')\left(\delta_{ij} - \Pi_j(\theta, \theta')\right), \\
\frac{\partial \Pi_i(\theta, \theta')}{\partial \theta'_j} &= -\Pi_i(\theta, \theta')\left(\delta_{ij} - \Pi_j(\theta, \theta')\right) \\
&\quad + \Pi_i(\theta, \theta')\left(\frac{1}{\pi_i(\theta')}\frac{\partial \pi_i}{\partial \theta'_j}(\theta') - \sum_{\ell=1}^n \Pi_\ell(\theta, \theta')\frac{1}{\pi_\ell(\theta')}\frac{\partial \pi_\ell}{\partial \theta'_j}(\theta')\right).
\end{aligned}$$

(ii) For $1 \le i \le n$ and $1 \le j \le n-1$, we have

$$\begin{aligned}\frac{\partial \Pi_i^*(\phi,\phi')}{\partial \phi_j} &= \Pi_i(\phi,\phi')(\delta_{ij} - \Pi_j(\phi,\phi')) \\ &\quad + \Pi_i(\phi,\phi')\left(\frac{1}{\pi_i(\phi)}\frac{\partial \pi_i}{\partial \phi_j}(\phi) - \sum_{\ell=1}^{n} \Pi_\ell(\phi,\phi')\frac{1}{\pi_\ell(\phi)}\frac{\partial \pi_\ell}{\partial \phi_j}(\phi)\right), \\ \frac{\partial \Pi_i(\phi,\phi')}{\partial \phi_j'} &= -\Pi_i(\phi,\phi')(\delta_{ij} - \Pi_j(\phi,\phi')).\end{aligned}$$

Lemma 10.3.3 (Derviatives of $\pi(\cdot)$). *For $1 \le i \le n$ and $1 \le j \le n-1$, we have*

$$\begin{aligned}\frac{\partial \pi_i}{\partial \theta_j}(\theta) &= \pi_i(\theta)(\delta_{ij} - \pi_j(\theta)) - \pi_i(\theta)\left(\frac{\partial \phi_i}{\partial \theta_j}(\theta) - \sum_{\ell=1}^{n-1} \pi_\ell(\theta)\frac{\partial \phi_\ell}{\partial \theta_j}(\theta)\right), \\ \frac{\partial \pi_i}{\partial \phi_j}(\phi) &= -\pi_i(\phi)(\delta_{ij} - \pi_j(\phi)) + \pi_i(\phi)\left(\frac{\partial \theta_i}{\partial \phi_j}(\phi) - \sum_{\ell=1}^{n-1} \pi_\ell(\phi)\frac{\partial \theta_\ell}{\partial \phi_j}(\phi)\right).\end{aligned} \tag{10.3.5}$$

Proof. We prove the second formula and the proof of the first is similar. Using (6.2.5), we write

$$\pi_i(\phi) = \frac{e^{\theta_i - \phi_i}}{\sum_{\ell=1}^n e^{\theta_\ell - \phi_\ell}}$$

and regard θ is a function of ϕ (recall that $\theta_n = \phi_n = 0$). Then

$$\begin{aligned}\frac{\partial \pi_i}{\partial \phi_j}(\phi) &= \frac{e^{\theta_i - \phi_i}\left(\frac{\partial \theta_i}{\partial \phi_j}(\phi) - \delta_{ij}\right)}{\sum_{\ell=1}^n e^{\theta_\ell - \phi_\ell}} - \frac{e^{\theta_i - \phi_i}}{\left(\sum_{\ell=1}^n e^{\theta_\ell - \phi_\ell}\right)^2}\sum_{\ell=1}^n e^{\theta_\ell - \phi_\ell}\left(\frac{\partial \theta_\ell}{\partial \phi_j}(\phi) - \delta_{\ell j}\right) \\ &= -\pi_i(\phi)(\delta_{ij} - \pi_j(\phi)) + \pi_i(\phi)\left(\frac{\partial \theta_i}{\partial \phi_j}(\phi) - \sum_{\ell=1}^{n-1} \pi_\ell(\phi)\frac{\partial \theta_\ell}{\partial \phi_j}(\phi)\right).\end{aligned}$$

Note that the nth term of the sum is omitted because $\theta_n = 0$. □

Thanks to these formulas, computations in the primal and dual coordinates are very similar except for a change of sign. In the following we will often give details for one coordinate system and leave the other one to the reader.

Last but not least, let $\frac{\partial \phi}{\partial \theta}(\theta) = \left(\frac{\partial \phi_i}{\partial \theta_j}(\theta)\right)$ be the Jacobian of the change of coordinate map $\theta \mapsto \phi$. Similarly, we let $\frac{\partial \theta}{\partial \phi}(\phi) = \left(\frac{\partial \theta_i}{\partial \phi_j}(\phi)\right)$ be the Jacobian of the inverse map $\phi \mapsto \theta$. The two Jacobians are inverses of each other, i.e.,

$$\frac{\partial \phi}{\partial \theta}(\theta)\frac{\partial \theta}{\partial \phi}(\phi) = I. \tag{10.3.6}$$

10.3.3 Riemannian metric

For intuition, we first compute the Riemannian inner product using Euclidean coordinates. We let $T_p\Delta_n$ be the tangent space at p.

Proposition 10.3.4. *Let $u, v \in T_p\Delta_n$ be represented in Euclidean coordinates, i.e., $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$ and $u_1 + \cdots + u_n = 0$, and similarly for v. Then*

$$\begin{aligned} \langle u, v\rangle &= u^T \left(-\mathrm{Hess}\,\varphi(p) - \nabla\varphi(p)\nabla\varphi(p)^T\right) v \\ &= u^T \left(\frac{-1}{\Phi(p)}\mathrm{Hess}\,\Phi(p)\right) v. \end{aligned} \tag{10.3.7}$$

Proof. By [19, Proposition 11.3.1] we have $\|v\|^2 = \frac{d^2}{dt^2} T\left(p + tv \mid p\right)\Big|_{t=0}$, where

$$T\left(p + tv \mid p\right) = \log\left(1 + t\nabla\varphi(p) \cdot v\right) - \left(\varphi(p + tv) - \varphi(p)\right).$$

Differentiating two times and setting $t = 0$ give the first equality in (10.3.7) when $u = v$,[1] and polarizing gives the general case. The second equality follows from the chain rule. □

Theorem 10.3.5 (Riemannian metric)**.**

(i) Under the primal coordinate system, the Riemannian metric is given by

$$g_{ij}(\theta) = \pi_i(\theta)(\delta_{ij} - \pi_j(\theta)) - \frac{\partial \pi_i}{\partial \theta_j}(\theta). \tag{10.3.8}$$

Its inverse is given by

$$g^{ij}(\theta) = \frac{1}{\pi_j(\theta)}\frac{\partial \theta_i}{\partial \phi_j}(\phi) + \frac{1}{\pi_n(\theta)}\sum_{\ell=1}^{n-1}\frac{\partial \theta_i}{\partial \phi_\ell}(\phi). \tag{10.3.9}$$

(ii) Under the dual coordinate system, the Riemannian metric is given by

$$g^*_{ij}(\phi) = \pi_i(\phi)(\delta_{ij} - \pi_j(\phi)) + \frac{\partial \pi_i}{\partial \phi_j}(\phi). \tag{10.3.10}$$

Its inverse is given by

$$g^{*ij}(\phi) = \frac{1}{\pi_j(\phi)}\frac{\partial \phi_i}{\partial \theta_j}(\phi) + \frac{1}{\pi_n(\phi)}\sum_{\ell=1}^{n-1}\frac{\partial \phi_i}{\partial \theta_\ell}(\phi). \tag{10.3.11}$$

[1]Note that this is two times the drift quadratic form (see Definition 7.2.2).

Proof. (i) By Lemma 10.2.4 and Lemma 10.3.2, we compute

$$\frac{\partial}{\partial\theta_i}T\left(\theta\mid\theta'\right) = \Pi_i(\theta,\theta') - \pi_i(\theta), \tag{10.3.12}$$

$$\frac{\partial}{\partial\theta_i'}T\left(\theta\mid\theta'\right) = -\Pi_i(\theta,\theta') + \pi_i(\theta) + \sum_{\ell=1}^{n}\Pi_\ell(\theta,\theta')\frac{1}{\pi_\ell(\theta')}\frac{\partial\pi_\ell}{\partial\theta_i'}(\theta'). \tag{10.3.13}$$

Differentiating (10.3.12) with respect to θ_j', we have

$$\begin{aligned}\frac{\partial^2}{\partial\theta_i\partial\theta_j'}T\left(\theta\mid\theta'\right) &= -\Pi_i(\theta,\theta')\left(\delta_{ij} - \Pi_j(\theta,\theta')\right)\\ &\quad + \Pi_i(\theta,\theta')\left(\frac{1}{\pi_i(\theta')}\frac{\partial\pi_i}{\partial\theta_j'}(\theta') - \sum_{\ell=1}^{n}\Pi_\ell(\theta,\theta')\frac{1}{\pi_\ell(\theta')}\frac{\partial\pi_\ell}{\partial\theta_j'}(\theta')\right).\end{aligned} \tag{10.3.14}$$

Setting $\theta = \theta'$, we get $g_{ij}(\theta) = \pi_i(\theta)(\delta_{ij} - \pi_j(\theta)) - \frac{\partial\pi_i}{\partial\theta_j}(\theta)$.

By Lemma 10.3.3, we have the alternative expression

$$g_{ij}(\theta) = \pi_i(\theta)\left(\frac{\partial\phi_i}{\partial\theta_j}(\theta) - \sum_{\ell=1}^{n-1}\pi_\ell(\theta)\frac{\partial\phi_\ell}{\partial\theta_j}(\theta)\right). \tag{10.3.15}$$

Expressing (10.3.15) in matrix form, we have

$$(g_{ij}(\theta)) = \operatorname{diag}\left(\pi(\theta)\right)(I - \mathbf{1}\pi'(\theta))\frac{\partial\phi}{\partial\theta}(\theta), \tag{10.3.16}$$

where $\pi(\theta) = (\pi_1(\theta),\ldots,\pi_{n-1}(\theta))'$, $\mathbf{1} = \mathbf{1}_{n-1} = (1,\ldots,1)'$ and $I = I_{n-1}$ is the identity matrix.

To invert (10.3.16) we use the fact that

$$(I - \mathbf{1}\pi'(\theta))^{-1} = I + \frac{\mathbf{1}\pi'(\theta)}{\pi_n(\theta)}.$$

This can be verified directly or seen as a special case of the Sherman-Morrison formula. Thus

$$\left(g^{ij}(\theta)\right) = \frac{\partial\theta}{\partial\phi}(\phi)\left(I + \frac{\mathbf{1}\pi'(\theta)}{\pi_n(\theta)}\right)\operatorname{diag}\left(\frac{1}{\pi(\theta)}\right).$$

Now (10.3.9) follows by expanding the matrix product.

(ii) The proofs of (10.3.10) and (10.3.11) follow the same lines. For later use we record the following formulas:

$$\frac{\partial}{\partial \phi_i} T(\phi \mid \phi') = \Pi_i^*(\phi, \phi') - \pi_i(\phi) + \sum_{\ell=1}^n \frac{1}{\pi_\ell(\phi)} \frac{\partial \pi_\ell}{\partial \phi_i}(\phi) \Pi_\ell^*(\phi, \phi'), \tag{10.3.17}$$

$$\frac{\partial}{\partial \phi_i'} T(\phi \mid \phi') = -\Pi_i^*(\phi, \phi') + \pi_i(\phi'), \tag{10.3.18}$$

$$\begin{aligned} \frac{\partial^2}{\partial \phi_i \partial \phi_j'} T(\phi \mid \phi') &= -\Pi_j^*(\phi, \phi')(\delta_{ij} - \Pi_i^*(\phi, \phi')) \\ &\quad - \Pi_j(\phi, \phi') \left(\frac{1}{\pi_j(\phi)} \frac{\partial \pi_j}{\partial \phi_i}(\phi) - \sum_{\ell=1}^n \frac{1}{\pi_\ell(\phi)} \frac{\partial \pi_\ell}{\partial \phi_i}(\phi) \Pi_\ell(\phi, \phi') \right). \end{aligned} \tag{10.3.19}$$

□

Remark 10.3.6. By Lemma 6.4.7 we have

$$\frac{\partial \pi_i}{\partial \theta_j}(\theta) = \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(\theta) = \frac{\partial \pi_j}{\partial \theta_i}(\theta). \tag{10.3.20}$$

Thus the right hand side of (10.3.8) is symmetric in i and j. Similarly, we have $\frac{\partial \pi_i}{\partial \phi_j} = \frac{\partial \pi_j}{\partial \phi_i}$.

10.3.4 Primal and dual connections

Theorem 10.3.7 (Primal and dual connections)**.**

(i) Under the primal coordinate system, the coefficients of the primal connection ∇ *is given by*

$$\Gamma_{ijk}(\theta) = \delta_{ij} g_{ik}(\theta) - \pi_i(\theta) g_{jk}(\theta) - \pi_j(\theta) g_{ik}(\theta), \tag{10.3.21}$$

$$\Gamma_{ij}^k(\theta) = \delta_{ijk} - \delta_{ik}\pi_j(\theta) - \delta_{jk}\pi_i(\theta). \tag{10.3.22}$$

(ii) Under the dual coordinate system, the coefficients of the dual connection ∇^* *is given by*

$$\Gamma_{ijk}^*(\phi) = -\delta_{ij} g_{ik}^*(\phi) + \pi_i(\phi) g_{jk}^*(\phi) + \pi_j(\phi) g_{ik}^*(\phi), \tag{10.3.23}$$

$$\Gamma_{ij}^{*k}(\phi) = -\delta_{ijk} + \delta_{ik}\pi_j(\phi) + \delta_{jk}\pi_i(\phi). \tag{10.3.24}$$

Proof. We prove (ii) and leave (i) to the reader. By (10.3.19), we have

$$\frac{\partial^2}{\partial \phi_k \partial \phi_i'} T(\phi \mid \phi') = -\Pi_i^*(\phi, \phi')(\delta_{ik} - \Pi_k^*(\phi, \phi'))$$
$$- \Pi_i^*(\phi, \phi') \left(\frac{1}{\pi_i(\phi)} \frac{\partial \pi_i}{\partial \phi_k}(\phi) - \sum_{\ell=1}^{n} \frac{1}{\pi_\ell(\phi)} \frac{\partial \pi_\ell}{\partial \phi_k}(\phi) \Pi_\ell(\phi, \phi') \right).$$

For notational convenience we momentarily suppress ϕ and ϕ' in the computation (later we will do so without comment). Differentiating one more time, we have

$$\begin{aligned}
&\frac{\partial^3}{\partial \phi_k \partial \phi_i' \partial \phi_j'} T(\phi \mid \phi') \\
&= -\delta_{ik} \frac{\partial \Pi_i^*}{\partial \phi_j'} + \Pi_i^* \frac{\partial \Pi_k^*}{\partial \phi_j'} + \Pi_k^* \frac{\partial \Pi_i^*}{\partial \phi_j'} - \frac{\partial \Pi_i^*}{\partial \phi_j'} \left(\frac{1}{\pi_i} \frac{\partial \pi_i}{\partial \phi_k} - \sum_{\ell=1}^{n} \frac{1}{\pi_\ell} \frac{\partial \pi_\ell}{\partial \phi_k} \Pi_\ell^* \right) + \Pi_i \sum_{\ell=1}^{n} \frac{1}{\pi_\ell} \frac{\partial \pi_\ell}{\partial \phi_k} \frac{\partial \Pi_\ell^*}{\partial \phi_j'} \\
&= \delta_{ik} \Pi_i^* (\delta_{ij} - \Pi_j^*) - \Pi_i^* \Pi_k^* (\delta_{jk} - \Pi_j^*) - \Pi_k^* \Pi_i^* (\delta_{ij} - \Pi_j^*) \\
&\quad + \Pi_i^* (\delta_{ij} - \Pi_j^*) \left(\frac{1}{\pi_i} \frac{\partial \pi_i}{\partial \phi_k} - \sum_{\ell=1}^{n} \frac{1}{\pi_\ell} \frac{\partial \pi_\ell}{\partial \phi_k} \Pi_\ell^* \right) - \Pi_i^* \sum_{\ell=1}^{n} \frac{1}{\pi_\ell} \frac{\partial \pi_\ell}{\partial \phi_k} \Pi_\ell^* (\delta_{\ell j} - \Pi_j^*).
\end{aligned}$$

Evaluating at $\phi = \phi'$ and simplifying, we get

$$\begin{aligned}
\Gamma_{ijk}^*(\phi) &= -\delta_{ijk} \pi_i - 2\pi_i \pi_j \pi_k + \delta_{ij} \pi_i \pi_k + \delta_{jk} \pi_j \pi_i + \delta_{ki} \pi_k \pi_j \\
&\quad - \delta_{ij} \frac{\partial \pi_i}{\partial \phi_k} + \pi_j \frac{\partial \pi_i}{\partial \phi_k} + \pi_i \frac{\partial \pi_j}{\partial \phi_k}.
\end{aligned} \tag{10.3.25}$$

By (10.3.10), we have $\frac{\partial \pi_i}{\partial \phi_j} = g_{ij}^* - \pi_i(\delta_{ij} - \pi_j)$. Plugging this into (10.3.25) and simplifying, we have $\Gamma_{ijk}^*(\phi) = -\delta_{ij} g_{ik}^*(\phi) + \pi_i(\phi) g_{jk}^*(\phi) + \pi_j(\phi) g_{ik}^*(\phi)$. Finally,

$$\begin{aligned}
\Gamma_{ij}^{*k}(\phi) &= \sum_{m=0}^{n-1} \left(-\delta_{ij} g_{im}^*(\phi) + \pi_i(\phi) g_{jm}^*(\phi) + \pi_j(\phi) g_{im}^*(\phi) \right) g^{mk}(\phi) \\
&= -\delta_{ijk} + \delta_{ik} \pi_j(\phi) + \delta_{jk} \pi_i(\phi).
\end{aligned}$$

□

Remark 10.3.8. It is interesting to note that although the connections are defined in terms of third order derivative of $T(\cdot \mid \cdot)$, the coefficients $\Gamma_{ij}^k(\theta)$ and $\Gamma_{ij}^{*k}(\phi)$ are given in terms of the portfolio π which is a normalized gradient of φ.

10.3.5 Curvatures

It is well known that the induced geometry of any Bregman divergence is dually flat. This is not the case for the geometry of L-divergence whenever $n \geq 3$ (when $n = 2$ the simplex Δ_2 is one-dimensional). To verify this we compute the Riemann-Christoffel curvature tensors of the primal and dual connections. In this (and only this) subsection we adopt the Einstein summation notation (see [2, p.20]).

The Riemann-Christoffel (RC) curvature tensor of a connection ∇ is defined for smooth vector fields X, Y and Z by

$$R(X, Y)Z = \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X,Y]}Z,$$

where $[X, Y]$ is the Lie bracket. Its coordinate representation is defined in terms of the coefficients R_{ijk}^{ℓ} by $R\left(\frac{\partial}{\partial\theta_i}, \frac{\partial}{\partial\theta_j}\right)\frac{\partial}{\partial\theta_k} = R_{ijk}^{\ell}\frac{\partial}{\partial\theta_\ell}$. By [2, (5.66)], we have

$$R_{ijk}^{\ell} = \frac{\partial}{\partial\theta_i}\Gamma_{jk}^{\ell} - \frac{\partial}{\partial\theta_j}\Gamma_{ik}^{\ell} + \Gamma_{im}^{\ell}\Gamma_{jk}^{m} - \Gamma_{jm}^{\ell}\Gamma_{ik}^{m}.$$

Theorem 10.3.9 (Primal and dual Riemann-Christoffel curvatures)**.** *Let R and R^* be the RC curvature tensors of the primal and dual connections respectively.*

(i) In primal coordinates, the coefficients of R are given by

$$R_{ijk}^{\ell}(\theta) = \delta_{\ell j}g_{ik}(\theta) - \delta_{\ell i}g_{jk}(\theta). \tag{10.3.26}$$

(ii) In dual coordinates, the coefficients of R^* are given by

$$R_{ijk}^{*\ell}(\phi) = \delta_{\ell j}g_{ik}^{*}(\phi) - \delta_{\ell i}g_{jk}^{*}(\phi). \tag{10.3.27}$$

In particular, for $n \geq 3$ both R and R^* do not vanish anywhere on Δ_n.

Proof. We prove the statements for R. Using (10.3.21) and suppressing the argument, we have

$$\frac{\partial}{\partial\theta_i}\Gamma_{jk}^{\ell} = -\delta_{\ell j}\frac{\partial\pi_k}{\partial\theta_i} - \delta_{\ell k}\frac{\partial\pi_j}{\partial\theta_i}, \quad \frac{\partial}{\partial\theta_j}\Gamma_{ik}^{\ell} = -\delta_{\ell i}\frac{\partial\pi_k}{\partial\theta_j} - \delta_{\ell k}\frac{\partial\pi_i}{\partial\theta_j}.$$

From (10.3.20) it follows that

$$\frac{\partial}{\partial \theta_i}\Gamma_{jk}^{\ell} - \frac{\partial}{\partial \theta_j}\Gamma_{ik}^{\ell} = -\delta_{\ell j}\frac{\partial \pi_k}{\partial \theta_i} + \delta_{\ell i}\frac{\partial \pi_k}{\partial \theta_j}. \tag{10.3.28}$$

Next we compute (with some work)

$$\Gamma_{im}^{\ell}\Gamma_{jk}^{m} - \Gamma_{jm}^{\ell}\Gamma_{ik}^{m} = -\delta_{\ell i}\delta_{jk}\pi_j + \delta_{\ell j}\delta_{ik}\pi_i + \delta_{\ell i}\pi_j\pi_k - \delta_{\ell j}\pi_i\pi_k. \tag{10.3.29}$$

Combining (10.3.28) and (10.3.29), we have

$$\begin{aligned} R_{ijk}^{\ell}(\theta) &= -\delta_{\ell j}\frac{\partial \pi_k}{\partial \theta_i} + \delta_{\ell i}\frac{\partial \pi_k}{\partial \theta_j} - \delta_{\ell i}\delta_{jk}\pi_j + \delta_{\ell j}\delta_{ik}\pi_i + \delta_{\ell i}\pi_j\pi_k - \delta_{\ell j}\pi_i\pi_k \\ &= \delta_{\ell j}\left(\delta_{ik}\pi_i - \pi_i\pi_k - \frac{\partial \pi_k}{\partial \theta_i}\right) - \delta_{\ell i}\left(\delta_{jk}\pi_j - \pi_j\pi_k - \frac{\partial \pi_k}{\partial \theta_j}\right) \\ &= \delta_{\ell j}g_{ik} - \delta_{\ell i}g_{jk}. \end{aligned}$$

To see that R does not vanish for $n \geq 3$, suppose on the contrary that $R(\theta) = 0$. Then $R_{ijk}^{\ell}(\theta) = \delta_{\ell j}g_{ik}(\theta) - \delta_{\ell i}g_{jk}(\theta) = 0$ for all values of i, j, k, ℓ. Fix i and k. Letting $\ell = j$, we have $g_{ik}(\theta) = \delta_{ij}g_{jk}(\theta)$. Next let $j \neq i$ (here we need $\dim \Delta_n = n - 1 \geq 2$). Then we get $g_{ik}(\theta) = 0$. Since i and k are arbitrary, we have $g(\theta) = 0$ which is a contradiction. □

We end this section by showing that the primal and dual connections have constant sectional curvature -1. For the definitions of sectional and Ricci curvatures we refer the reader to [66, Chapter 7] (these are compatible with the notations in [2]). Note that the sectional and Ricci curvatures can be defined with respect to any given affine connection and Riemannian metric.

Corollary 10.3.10 (Primal and dual sectional curvatures)**.** *The primal and dual connections have constant sectional curvature -1 with respect to g. In particular, the primal and dual Ricci curvatures satisfy the Einstein condition*

$$\mathrm{Ric} = \mathrm{Ric}^* = -(n-2)g.$$

Proof. Recall that ∇ has constant sectional curvature k with respect to g if

$$R(X, Y)Z \equiv k(\langle Y, Z\rangle X - \langle X, Z\rangle Y)$$

for all X, Y and Z. For the primal Riemann-Christoffel curvature tensor we have

$$R\left(\frac{\partial}{\partial\theta_i}, \frac{\partial}{\partial\theta_j}\right)\frac{\partial}{\partial\theta_k} = R_{ijk}^{\ell}\frac{\partial}{\partial\theta_\ell} = -\left(\left\langle \frac{\partial}{\partial\theta_j}, \frac{\partial}{\partial\theta_k}\right\rangle \frac{\partial}{\partial\theta_i} - \left\langle \frac{\partial}{\partial\theta_i}, \frac{\partial}{\partial\theta_k}\right\rangle \frac{\partial}{\partial\theta_j}\right),$$

which implies that the sectional curvature is $k = -1$. The claim for Ricci curvature follows immediately by taking trace (see for example [94, (4.31)]). The proof for the dual curvatures is the same. □

10.4 Geodesics and generalized Pythagorean theorem

Armed with the primal and dual connections we can formulate the primal and dual geodesic equations. Their solutions are the primal and dual geodesics which will be studied in this section. The highlight of this section is the generalized Pythagorean theorem (Theorem 10.1.1). Along the way we will prove some remarkable properties of the geometric structure (g, ∇, ∇^*).

10.4.1 Primal and dual geodesics

Note that in Figure 10.2 the primal geodesic is drawn as a straight line in Δ_n. We now prove that this is indeed the case. The same is true for the dual geodesic in dual Euclidean coordinates.

Let $\gamma : [0, 1] \to \Delta_n$ be a smooth curve. We denote time derivatives by $\dot{\gamma}(t)$. Let $\theta(t)$ and $\phi(t)$ be the primal and dual coordinate representations of γ. We say that γ is a primal geodesic if its satisfies the primal geodesic equation

$$\ddot{\theta}_k(t) + \sum_{i,j=1}^{n-1} \Gamma_{ij}(\theta(t))\dot{\theta}_i(t)\dot{\theta}_j(t) = 0, \quad k = 1, \ldots, n-1.$$

It is a dual geodesic if its satisfies the dual geodesic equation

$$\ddot{\phi}_k(t) + \sum_{i,j=1}^{n-1} \Gamma_{ij}^*(\phi(t))\dot{\phi}_i(t)\dot{\phi}_j(t) = 0, \quad k = 1, \ldots, n-1.$$

By Theorem 10.3.7, the primal geodesic equation in primal coordinates is

$$\ddot{\theta}_k(t) + 2\dot{\theta}_k(t) \sum_{\ell=1}^{n-1} \pi_\ell(\theta(t))\dot{\theta}_\ell(t) = 0, \quad k = 1, \ldots, n-1. \tag{10.4.1}$$

The dual geodesic equation in dual coordinates is

$$\ddot{\phi}_k(t) - 2\dot{\phi}_k(t) \sum_{\ell=1}^{n-1} \pi_\ell(\phi(t))\dot{\phi}_\ell(t) = 0, \quad k = 1, \ldots, n-1. \tag{10.4.2}$$

Theorem 10.4.1 (Primal and dual geodesics)**.**

(i) Let $\gamma : [0,1] \to \Delta_n$ be a primal geodesic. Then the trace of γ in Δ_n is the Euclidean straight line in Δ_n joining $\gamma(0)$ and $\gamma(1)$.

(ii) Let $\gamma^* : [0,1] \to \Delta_n$ be a dual geodesic. For each t, let $p^*(t)$ be the dual Euclidean coordinate of $\gamma(t)$. Then the trace of p^* in Δ_n is the Euclidean straight line in Δ_n joining $p^*(0)$ and $p^*(1)$.

Proof. (i) Let $q, r \in \Delta_n$ be fixed. Let their primal coordinates be θ^q and θ^r respectively. Consider the curve $\gamma : [0,1] \to \Delta_n$ defined in terms of the primal coordinate system by

$$\theta_k(t) = \log\left((1 - h(t))e^{\theta_k^q} + h(t)e^{\theta_k^r}\right), \quad k = 1, \ldots, n-1, \tag{10.4.3}$$

where $h(t)$ is a time parameterization to be chosen. Suppose that h is a solution to the one-dimensional dfferential equation

$$h''(t) - 2(h'(t))^2 \sum_{\ell=1}^{n-1} \pi_\ell(\theta(t)) \frac{\theta_\ell^r - \theta_\ell^q}{(1 - h(t))e^{\theta_\ell^q} + h(t)e^{\theta_\ell^r}} = 0, \tag{10.4.4}$$

where $\theta(t)$ is given by (10.4.3). Plugging (10.4.3) into the primal geodesic equation (10.4.1), it can be shown after some computation that γ is a primal geodesic.

It remains to show that there exists a solution h such that $h(0) = 0$, $h'(t) > 0$ for all $t \in [0,1]$ and $h(1) = 1$. With this choice of h the curve (10.4.3) is a primal geodesic from q to r.

First we note that if $h(t)$ is a solution, then $h(ct)$ is a solution for any $c > 0$. Also, if $h'(t_0) = 0$ then $h(t) = h(t_0)$ for all $t \geq t_0$. Let $h_0(t)$ be a maximal solution to (10.4.4) defined on an interval $[0, t_{\max})$ with $h_0(0) = 0$ and $h_0'(0) > 0$. By the previous remark, h_0 is strictly increasing on $[0, t_{\max})$. If $h_0(t)$ hits 1 at some $t = t_0 < t_{\max}$, the function $h(t) = h_0(t/t_0)$ is a solution with the desired properties. In fact, we claim that

$$\lim_{t \uparrow t_{\max}} h_0(t) = \sup_{t < t_{\max}} h_0(t) = M := \min_{1 \leq \ell \leq n-1} \left(\frac{e^{\theta_\ell^q}}{e^{\theta_\ell^q} - e^{\theta_\ell^r}} 1_{\{\theta_\ell^q > \theta_\ell^r\}} + \infty \cdot 1_{\{\theta_\ell^q \leq \theta_\ell^r\}} \right) > 1.$$

Suppose on the contrary that

$$M' := \sup_{t < t_{\max}} h(t) = \lim_{t \uparrow t_{\max}} h(t) < M.$$

Let $h_1(t)$, $t \in (-\epsilon, \epsilon)$ be a solution to (10.4.4) satisfying $h_1(0) = M'$ and $h_1'(0) > 0$. Note that h_1 exists because by construction the fractions in (10.4.4) are well-defined near M'. Then the range of h_1 contains an open interval containing M'. Thus there exists $t_0 < t_{\max}$, $c > 0$ and $t_1 < 0$ such that $ct_1 > -\epsilon$, $h_0(t_0) = h_1(ct_1)$ and $h_0'(t_0) = \frac{d}{dt} h_1(ct)\big|_{t=t_1}$. This allows us to extend the range of h_0 beyond M' which contradicts the maximality of M'. Thus there is a primal geodesic $\gamma : [0, 1] \to \Delta_n$ from q to r. By the uniqueness of the solution of the primal geodesic equation with given initial position and velocity (note that $\gamma(t)$ is a solution if and only if $\gamma(ct)$ is a solution where $c > 0$), we see that γ is the unique primal geodesic beginning at q at time 0 and reaching r at time 1.

To see that the trace of γ is a Euclidean straight line in Δ_n, consider its Euclidean representation $p(t) = (p_1(t), \ldots, p_n(t))$. By (10.4.3) we have

$$e^{\psi(\theta(t))} = \frac{1}{p_n(t)} = (1 - h(t)) e^{\psi(\theta^q)} + h(t) e^{\psi(\theta^r)}.$$

Solving for $\psi(\theta(t))$ gives

$$h(t) = \frac{e^{\psi(\theta(t))} - e^{\psi(\theta^q)}}{e^{\psi(\theta^r)} - e^{\psi(\theta^q)}}. \tag{10.4.5}$$

Expressing (10.4.3) in Euclidean coordinates and using (10.4.5), we get after some algebra that

$$p_k(t) = e^{\theta_k^q} p_n(t) + \left(1 - e^{\psi(\theta^q)} p_n(t)\right) \frac{e^{\theta_k^r} - e^{\theta_k^q}}{e^{\psi(\theta^r)} - e^{\psi(\theta^q)}}, \quad k = 1, \ldots, n-1.$$

Hence there exists a_k, b_k such that

$$p_k(t) = a_k + b_k p_n(t), \quad k = 1, \ldots, n-1. \tag{10.4.6}$$

Together with the identity $p_1(t) + \cdots + p_n(t) \equiv 1$, (10.4.6) shows that γ is the time change of the Euclidean straight line from q to r.

Using the dual coordinate system ϕ and dual Euclidean coordinate system p^* (ii) can be proved in a similar way by considering the curve defined by

$$\phi_k(t) = \log\left(\frac{1}{(1-h(t))e^{-\phi_k^q} + h(t)e^{-\phi_k^p}}\right), \quad k = 1, \ldots, n-1. \tag{10.4.7}$$

□

A manifold is said to be projectively flat (with respect to a given connection) if there is a coordinate system under which the geodesics are straight lines up to time reparameterization. In view of Theorem 10.4.1 we have the following corollary.

Corollary 10.4.2. *The manifold Δ_n equipped with the geometric structure (g, ∇, ∇^*) is dually projectively flat, but is not flat for $n \geq 3$.*

10.4.2 Gradient flows and inverse exponential maps

Motivated by the recent paper [5] we relate the primal and dual geodesics with gradient flows under the L-divergence. Fix $p, q, r \in \Delta_n$. Consider the following gradient flows starting at q:

$$\begin{cases} \dot{\gamma}(t) = -\mathrm{grad}\ T(r \mid \cdot)(\gamma(t)) \\ \gamma(0) = q \end{cases} \quad \text{(primal flow)} \tag{10.4.8}$$

and

$$\begin{cases} \dot{\gamma}^*(t) = -\mathrm{grad}\ T(\cdot \mid p)(\gamma^*(t)) \\ \gamma^*(0) = q \end{cases} \quad \text{(dual flow)} \tag{10.4.9}$$

Here grad denotes the Riemannian gradient with respect to the metric g. We call (10.4.8) the primal flow and (10.4.9) the dual flow.

It can be verified easily that

$$\frac{d}{dt}T(r \mid \gamma(t)) = -\|\dot{\gamma}(t)\|^2 \quad \text{and} \quad \frac{d}{dt}T(\gamma^*(t) \mid p) = -\|\dot{\gamma}^*(t)\|^2.$$

Since $T(q \mid p) = 0$ if and only if $p = q$, by standard ODE theory it can be shown that the solutions $\gamma(t)$ and $\gamma^*(t)$ are defined for $t \in [0, \infty)$ and

$$\lim_{t\to\infty} \gamma(t) = r, \quad \lim_{t\to\infty} \gamma^*(t) = p.$$

In other words, both gradient flows converge to the unique minimizers.

Theorem 10.4.3 (Gradient flows)**.**

(i) The primal flow $\gamma(t)$ is a time change of the primal geodesic from q to r.

(ii) The the dual flow $\gamma^*(t)$ is a time change of the dual geodesic from q to p.

Recall the concept of exponential map. For $q \in \Delta_n$ and $v \in T_q\Delta_n$, consider the primal geodesic γ starting at q with initial velocity v. If γ is defined up to time 1, we define $\exp_q(v) = \gamma(1)$. The dual exponential map $\exp^*$ is defined analogously. As a corollary of Theorems 10.4.1 and 10.4.3 we have the following characterization of the primal and dual inverse exponential maps.

Corollary 10.4.4 (Inverse exponential maps)**.** *Let* $\exp$ *and* $\exp^*$ *be the exponential maps with respect to the primal and dual connections respectively. For $p, q \in \Delta_n$ we have*

(i) $\exp_q^{-1}(p) \propto -\text{grad } T(p \mid \cdot)(q)$.

(ii) $\left(\exp_q^*\right)^{-1}(p) \propto -\text{grad } T(\cdot \mid p)(q)$.

To prove Theorem 10.4.3 we begin by computing the Riemannian gradients of $T(r \mid \cdot)$ and $T(\cdot \mid p)$.

Lemma 10.4.5 (Riemannian gradients)**.** *Let $p, q, r \in \Delta_n$.*

(i) Under the primal coordinate system, we have

$$\begin{aligned}\operatorname{grad}\ T\left(r \mid \cdot\right)(q) &= \sum_{i=1}^{n-1}\left(-\frac{\Pi_i(\theta^r,\theta^q)}{\pi_i(\theta^q)} + \frac{\Pi_n(\theta^r,\theta^q)}{\pi_n(\theta^q)}\right)\frac{\partial}{\partial\theta_i^q} \\ &= \frac{1}{\sum_{\ell=1}^n \pi_\ell(\theta^q)e^{\theta_\ell^r-\theta_\ell^q}}\sum_{i=1}^{n-1}\left(-e^{\theta_i^r-\theta_i^q}+1\right)\frac{\partial}{\partial\theta_i^q}.\end{aligned} \tag{10.4.10}$$

(ii) Under the dual coordinate system, we have

$$\begin{aligned}\operatorname{grad}\ T\left(\cdot \mid p\right)(q) &= \sum_{i=1}^{n-1}\left(\frac{\Pi_i^*(\phi^q,\phi^p)}{\pi_i(\phi^q)} - \frac{\Pi_n^*(\phi^q,\phi^p)}{\pi_n(\phi^q)}\right)\frac{\partial}{\partial\phi_i^q} \\ &= \frac{1}{\sum_{\ell=1}^n \pi_\ell(\phi^q)e^{\phi_\ell^q-\phi_\ell^p}}\sum_{i=1}^{n-1}\left(e^{\phi_i^q-\phi_i^p}-1\right)\frac{\partial}{\partial\phi_i^q}.\end{aligned} \tag{10.4.11}$$

Proof. (i) To prove the first formula in (10.4.10), we compute, using (10.3.9) and (10.3.12),

$$\begin{aligned}\left(\operatorname{grad}\ T\left(\cdot \mid \theta^p\right)(\theta^q)\right)_i &= \sum_{j=1}^{n-1} g^{ij}(\theta^q)\frac{\partial}{\partial\theta_j^q}T\left(\cdot \mid \theta^p\right)(\theta^q) \\ &= \sum_{j=1}^{n-1}\left(\frac{1}{\pi_j(\theta^q)}\frac{\partial\theta_i}{\partial\phi_j}(\theta^q) + \frac{1}{\pi_n(\theta^q)}\sum_{k=1}^{n-1}\frac{\partial\theta_i}{\partial\phi_k}(\theta^q)\right)\left(\Pi_j(\theta^q,\theta^p)-\pi_j(\theta^q)\right) \\ &= \sum_{j=1}^{n-1}\left(\frac{\Pi_j(\theta^q,\theta^p)}{\pi_j(\theta^q)} - \frac{\Pi_n(\theta^q,\theta^p)}{\pi_n(\theta^q)}\right)\frac{\partial\theta_i}{\partial\phi_j}(\theta^q).\end{aligned}$$

For the second formula, we first prove a

Claim. We have

$$\frac{\partial}{\partial\theta_j^q}T\left(\theta^r \mid \theta^q\right) = \sum_{\ell=1}^{n-1}\frac{\partial\phi_\ell}{\partial\theta_j^q}(\theta^q)(\pi_\ell(\theta^q)-\Pi_\ell(\theta^r,\theta^q)). \tag{10.4.12}$$

To see this, we use (10.3.13), (10.3.5) and compute as follows:

$$\begin{aligned}&\frac{\partial}{\partial\theta_j^q}T\left(\theta^r \mid \theta^q\right) \\ &= -\Pi_j(\theta^r,\theta^q)+\pi_j(\theta^q)+\sum_{\ell=1}^{n}\frac{1}{\pi_\ell(\theta^q)}\frac{\partial\pi_\ell}{\partial\theta_j^q}(\theta^q)\Pi_\ell(\theta^r,\theta^q) \\ &= -\Pi_j(\theta^r,\theta^q)+\pi_j(\theta^q)+\sum_{\ell=1}^{n}\left(\delta_{\ell j}-\pi_j(\theta^q)-\left(\frac{\partial\phi_\ell}{\partial\theta_j^q}(\theta^q)-\sum_{m=1}^{n-1}\pi_m(\theta^q)\frac{\partial\phi_m}{\partial\theta_j^q}(\theta^q)\right)\Pi_\ell(\theta^r,\theta^q)\right) \\ &= \sum_{\ell=1}^{n-1}\frac{\partial\phi_\ell}{\partial\theta_j^q}(\theta^q)(\pi_\ell(\theta^q)-\Pi_\ell(\theta^r,\theta^q)).\end{aligned}$$

Now we compute, using Theorem 10.3.5,

$$
\begin{aligned}
&(\text{grad } T\,(r\mid\cdot)\,(q))_i \\
&= \sum_{j=1}^{n-1} g^{ij}(\theta^q)\frac{\partial}{\partial\theta^q_j}T\,(\theta^r\mid\cdot)\,(\theta^q) \\
&= \sum_{j=1}^{n-1}\left(\frac{1}{\pi_i(\theta^q)}\frac{\partial\theta_j}{\partial\phi_i}(\phi^q)+\frac{1}{\pi_n(\theta^q)}\sum_{k=1}^{n-1}\frac{\partial\theta_j}{\partial\phi_k}(\phi^q)\right)\cdot\sum_{\ell=1}^{n-1}\frac{\partial\phi_\ell}{\partial\theta^q_j}(\theta^q)(\pi_\ell(\theta^q)-\Pi_\ell(\theta^r,\theta^q)) \\
&= \sum_{\ell=1}^{n-1}(\pi_\ell(\theta^q)-\Pi_\ell(\theta^r,\theta^q))\cdot\sum_{j=1}^{n-1}\left(\frac{1}{\pi_i(\theta^q)}\frac{\partial\phi_\ell}{\partial\theta^q_j}(\phi^q)\frac{\partial\theta_j}{\partial\phi_i}(\phi^q)+\frac{1}{\pi_n(\theta^q)}\sum_{k=1}^{n-1}\frac{\partial\phi_\ell}{\partial\theta^q_j}(\phi^q)\frac{\partial\theta_j}{\partial\phi_k}(\phi^q)\right) \\
&= \sum_{\ell=1}^{n-1}(\pi_\ell(\theta^q)-\Pi_\ell(\theta^r,\theta^q))\left(\frac{1}{\pi_i(\theta^q)}\delta_{\ell i}+\frac{1}{\pi_n(\theta^q)}\right) \\
&= -\frac{\Pi_i(\theta^r,\theta^q)}{\pi_i(\theta^q)}+\frac{\Pi_n(\theta^r,\theta^q)}{\pi_n(\theta^q)}.
\end{aligned}
$$

In the second last equality we used (10.3.6). The proof of (ii) is similar. □

Proof of Theorem 10.4.3. We prove (i) and leave (ii) to the reader. Let $\theta(t)$ be the primal representation of the primal flow starting at q. By Lemma 10.4.5, at any time t we have

$$\dot{\theta}_k(t) \propto e^{\theta^r_k-\theta_k(t)}-1=\frac{1}{e^{\theta_k(t)}}\left(e^{\theta^r_k}-e^{\theta_k(t)}\right),$$

where the constant of proportionality depends on $\theta(t)$ but is independent of k. It follows that

$$\frac{d}{dt}e^{\theta_k(t)} \propto e^{\theta^r_k}-e^{\theta_k(t)},\quad k=1,\ldots,n-1. \tag{10.4.13}$$

Comparing (10.4.13) and (10.4.3) we see that the primal flow is a time change of the primal geodesic. □

10.4.3 Generalized Pythagorean theorem

Having characterized the primal and dual geodesics, we are ready to prove the generalized Pythagorean theorem. Our proof makes use of the Riemannian gradients given in Lemma 10.4.5. The reason is that these gradients appear to have the correct scaling which is easier to handle, as can be seen in the proof (see (10.4.16)).

Proof of Theorem 10.1.1. Given $p, q, r \in \Delta_n$, consider the primal geodesic from q to r and the dual geodesic from q to p. Let

$$u = -\text{grad } T(\cdot \mid p)(q) \quad \text{and} \quad v = -\text{grad } T(r \mid \cdot)(q). \tag{10.4.14}$$

By Theorem 10.4.3 u and v are proportional to the initial velocities of the two geodesics. Thus, it suffices to prove that the sign of (10.1.1) is the same as that of $\langle u, v\rangle$. This claim will be established by the following two lemmas.

Lemma 10.4.6. *The sign of* $T(q \mid p) + T(r \mid q) - T(r \mid p)$ *is the same as that of*

$$1 - \sum_{k=1}^{n} \frac{\Pi_k(q,p)\Pi_k(r,q)}{\pi_k(q)}. \tag{10.4.15}$$

Proof. By Lemma 10.2.4, the sign of $T(q \mid p) + T(r \mid q) - T(r \mid p)$ is the same as that of

$$\left(\sum_{i=1}^{n} \pi_i(\theta^p) e^{\theta_i^q - \theta_i^p}\right)\left(\sum_{j=1}^{n} \pi_j(\theta^q) e^{\theta_j^r - \theta_j^q}\right) - \sum_{i=1}^{n} \pi_i(\theta^p) e^{\theta_i^r - \theta_i^p}.$$

Rearranging, we have

$$-\sum_{i,j=1}^{n} \pi_i(\theta^p)(\delta_{ij} - \pi_j(\theta^q)) e^{\theta_j^r - \theta_j^q} e^{\theta_i^q - \theta_i^p}.$$

Since scaling does not change sign, we may consider instead the quantity

$$-\sum_{i,j=1}^{n} \pi_i(\theta^p)(\delta_{ij} - \pi_j(\theta^q)) \frac{\Pi_j(\theta^r, \theta^q)}{\pi_j(\theta^q)} \frac{\Pi_i(\theta^q, \theta^p)}{\pi_i(\theta^p)}.$$

We get (10.4.15) by expanding. □

Lemma 10.4.7. *Consider the tangent vectors* u *and* v *defined by* (10.4.14)*. Then*

$$\langle u, v\rangle = 1 - \sum_{k=1}^{n} \frac{\Pi_k(\theta^q, \theta^p)\Pi_k(\theta^r, \theta^q)}{\pi_k(\theta^q)}. \tag{10.4.16}$$

Proof. For this computation we use the primal coordinate system. We have

$$u = -\text{grad } T(\cdot \mid p)(q) = -\sum_{i,k=1}^{n-1} g^{ik}(q) \frac{\partial}{\partial \theta_k^q} T(\cdot \mid p)(\theta^q) \frac{\partial}{\partial \theta_i^q}$$

$$v = -\text{grad } T(r \mid \cdot)(q) = -\sum_{j,\ell=1}^{n-1} g^{j\ell}(q) \frac{\partial}{\partial \theta_\ell^q} T(r \mid \cdot)(\theta^q) \frac{\partial}{\partial \theta_j^q}.$$

Using the definition of the Riemannian inner product, we compute

$$
\begin{aligned}
\langle u, v\rangle &= \sum_{i,j=1}^{n-1} g_{ij}(q) \sum_{k,\ell=1}^{n-1} g^{ik}(q) g^{j\ell}(q) \frac{\partial}{\partial \theta_k^q} T(\cdot \mid p)(\theta^q) \frac{\partial}{\partial \theta_\ell^q} T(r \mid \cdot)(\theta^q) \\
&= \sum_{k,\ell=1}^{n-1} g^{k\ell}(\theta^q) \frac{\partial}{\partial \theta_k^q} T(\cdot \mid p)(\theta^q) \frac{\partial}{\partial \theta_\ell^q} T(r \mid \cdot)(\theta^q).
\end{aligned}
$$

By (10.3.9), (10.3.18) and (10.4.12), we have

$$
\begin{aligned}
g^{k\ell}(q) &= \frac{1}{\pi_k(q)} \frac{\partial \theta_\ell}{\partial \phi_k}(\phi^q) + \frac{1}{\pi_n(\theta^q)} \sum_{\alpha=1}^{n-1} \frac{\partial \theta_\ell}{\partial \phi_\alpha}(\phi^q), \\
\frac{\partial}{\partial \theta_k^q} T(\cdot \mid p)(\theta^q) &= \Pi_k(\theta^q, \theta^p) - \pi_k(\theta^q), \\
\frac{\partial}{\partial \theta_\ell^q} T(r \mid \cdot)(\theta^q) &= \sum_{\beta=1}^{n-1} \frac{\partial \phi_\beta}{\partial \theta_\ell^q}(\theta^q) \left(\pi_\beta(\theta^q) - \Pi_\beta(\theta^r, \theta^q)\right).
\end{aligned}
$$

Claim. We have

$$
\langle u, v\rangle = \sum_{k,\ell=1}^{n-1} (\Pi_k(\theta^q, \theta^p) - \pi_k(\theta^q))(\pi_\ell(\theta^q) - \Pi_\ell(\theta^r, \theta^q)) \left(\frac{\delta_{k\ell}}{\pi_k(\theta^q)} + \frac{1}{\pi_n(\theta^q)} \right). \tag{10.4.17}
$$

To see this, write

$$
\begin{aligned}
\langle u, v\rangle = \sum_{k,\beta=1}^{n-1} &(\Pi_k(\theta^q, \theta^p) - \pi_k(\theta^q))(\pi_\beta(\theta^q) - \Pi_\beta(\theta^r, \theta^q)) \\
&\cdot \sum_{\ell=1}^{n-1} \left(\frac{1}{\pi_k(q)} \frac{\partial \phi_\beta}{\partial \theta_\ell^q}(\theta^q) \frac{\partial \theta_\ell}{\partial \theta_k}(\phi^q) + \frac{1}{\pi_n(q)} \sum_{\alpha=1}^{n-1} \frac{\partial \phi_\beta}{\partial \theta_\ell^q}(\theta^q) \frac{\partial \theta_\ell}{\partial \phi_\alpha}(\phi^q) \right).
\end{aligned}
$$

The last expression can be simplified using the identities

$$
\frac{\partial \phi_\beta}{\partial \theta_\ell^q}(\theta^q) \frac{\partial \theta_\ell}{\partial \theta_k}(\phi^q) = \delta_{\beta k}, \quad \frac{\partial \phi_\beta}{\partial \theta_\ell^q}(\theta^q) \frac{\partial \theta_\ell}{\partial \phi_\alpha}(\phi^q) = \delta_{\alpha\beta},
$$

and this gives the claim.

Finally, expanding and simplifying (10.4.17), we obtain the desired identity(10.4.16). □

□

10.5 Displacement interpolation

In this section we consider displacement interpolation for the optimal transport problem. We refer the reader to [100, Chapter 5] and [101, Chapter 7] for introductions to displacement interpolation.

10.5.1 Time dependent transport problem

Let $P^{(0)}$ and $P^{(1)}$ be Borel probability measures on $\mathbb{R}^{n-1}$. Consider the transport problem with cost $c(\theta, \phi) = \psi(\theta - \phi)$. Suppose the transport problem is solved in terms of the exponentially concave function φ on Δ_n. Letting $f = \varphi + \psi$, the optimal transport map is given by the c-supergradient of f. In particular, $P^{(1)}$ is the pushforward of $P^{(0)}$ under $F := \nabla^c f$:

$$P^{(1)} = F_{\#} P^{(0)}.$$

The idea of displacement interpolation is to introduce an additional time structure. We want to define an 'action' $\mathcal{A}(\cdot)$ on curves such that the cost function is given by

$$c(\theta, \phi) = \min_{\gamma} \mathcal{A}(\gamma),$$

where the minimum is taken over smooth curves $\gamma : [0, 1] \to \mathbb{R}^{n-1}$ satisfying $\gamma(0) = \theta$ and $\gamma(1) = \phi$. For each pair (θ, ϕ), a minimizing curve γ gives a time-dependent map transporting θ to ϕ along a continuous path. Let $F^{(t)} : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ be defined by $F^{(t)}(\theta) = \gamma(t)$, where γ is the minimizing curve for the pair $(\theta, F(\theta))$. We want to define $\mathcal{A}$ in such a way that $F^{(t)}$ is an optimal transport map for the probability measures $P^{(0)}$ and $P^{(t)}$ where

$$P^{(t)} = \left(F^{(t)}\right)_{\#} P_0.$$

For the classical Euclidean case with cost $|x - y|^2$ the action is $\mathcal{A}(\gamma) = \int_0^1 |\dot{\gamma}(t)|^2 dt$ and the optimal transport map has the form $F(x) = x - \nabla h(x)$ where h is an ordinary concave function. The displacement interpolations are linear interpolations:

$$F^{(t)} = (1 - t)\mathrm{Id} + tF$$

(See [100, Theorem 5.5, Theorem 5.6].) In particular, the individual trajectories (minimizing curves) are Euclidean straight lines which can be regarded as the geodesics of a flat geometry. In this section we formulate and prove an analogous statement for our transport problem.

10.5.2 Lagrangian action and portfolio interpolation

We begin by defining an appropriate action. Let $\gamma : [0,1] \to \mathbb{R}^{n-1}$ be a smooth curve with $\gamma(0) = \theta$. For each t, define $q(t) \in \Delta_n$ such that its exponential coordinate is $\theta - \gamma(t)$, i.e.,

$$\frac{q_i(t)}{q_n(t)} = e^{\theta_i - \gamma_i(t)}, \quad 1 \leq i \leq n-1. \tag{10.5.1}$$

Equivalently, we have $q_i(t) = e^{\theta_i - \gamma_i(t) - \psi(\theta - \gamma(t))}$, for $1 \leq i \leq n-1$. Intuitively, we think of $q(t)$ as the portfolio at time t (in the sense of interpolation). Note that $q(0) = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$.

We define the Lagrangian action by

$$\mathcal{A}(\gamma) = \int_0^1 -\log\left(\frac{1}{n} + \dot{q}_n(t)\right) dt. \tag{10.5.2}$$

We take $-\log(\cdot) = \infty$ if the argument is not in $(0, \infty)$. An alternative representation of the action is

$$\mathcal{A}(\gamma) = \int_0^1 -\log\left(\frac{1}{n} + \frac{d}{dt} e^{-\psi(\gamma(0) - \gamma(t))}\right) dt. \tag{10.5.3}$$

Lemma 10.5.1. *For any $\theta, \phi \in \mathbb{R}^{n-1}$ we have*

$$c(\theta, \phi) = \psi(\theta - \phi) = \min\left\{\mathcal{A}(\gamma) : \gamma(0) = \theta, \gamma(1) = \phi\right\}. \tag{10.5.4}$$

The action is minimized by the curve

$$\gamma_i(t) = \theta_i - \log\frac{(1-t)\frac{1}{n} + tq_i(1)}{(1-t)\frac{1}{n} + tq_n(1)}, \quad 1 \leq i \leq n-1. \tag{10.5.5}$$

In particular, for this minimizing curve we have

$$q(t) = (1-t)\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) + tq(1). \tag{10.5.6}$$

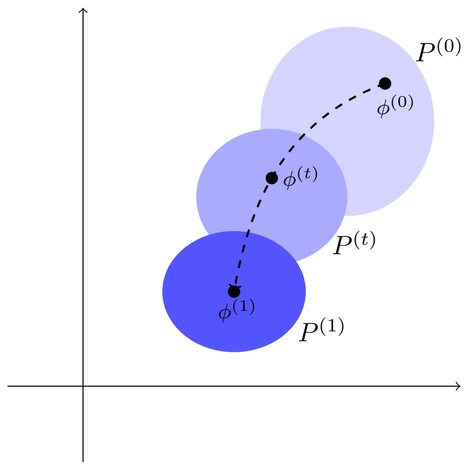

Figure 10.3: Displacement interpolation

Proof. Fix a smooth curve $\gamma : [0,1] \to \mathbb{R}^{n-1}$ from θ to ϕ. Since $-\log$ is convex, by Jensen's inequality we have

$$\begin{aligned} \int_0^1 -\log\left(\frac{1}{n} + \dot{q}_n(t)\right) dt &\geq -\log\left(\frac{1}{n} + \int_0^1 \dot{q}_n(t) dt\right) \\ &= -\log\left(\frac{1}{n} + q_n(1) - q_n(0)\right) = -\log q_n(1) = \psi(\theta - \phi). \end{aligned} \tag{10.5.7}$$

For the curve defined by (10.5.5), $\dot{q}(t) = q(1) - \frac{1}{n}$ is constant and so equality holds in (1.3.3). Finally (10.5.6) follows by a direct calculation. □

10.5.3 Displacement interpolation

We work under the following setting. Let $P^{(0)}$ and $P^{(1)}$ be Borel probability measures on $\mathbb{R}^{n-1}$. Let $\varphi : \Delta_n \to \mathbb{R}$ be an exponentially concave function, satisfying Assumption 6.4.3, such that $F^{(1)} := \nabla^c f$ is an optimal transport map (here f is the c-concave function $\varphi + \psi$). Let $\pi^{(1)} : \Delta_n \to \Delta_n$ be the portfolio map generated by $\varphi^{(1)} = \varphi$.

Consider the flow $(t, \theta) \mapsto \phi^{(t)}(\theta)$ defined by the minimizing curves (10.5.5), i.e.,

$$\phi_i^{(t)}(\theta) = \theta_i - \log \frac{\pi_i^{(t)}(\theta)}{\pi_n^{(t)}(\theta)}, \quad 1 \leq i \leq n-1, \quad t \in [0,1], \tag{10.5.8}$$

where each $\pi^{(t)} : \Delta_n \to \Delta_n$ is the portfolio map defined by

$$\pi^{(t)} = (1-t)\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) + t\pi^{(1)}, \quad t \in [0,1]. \tag{10.5.9}$$

See Figure 10.3 for an illustration.

The following is the main result of this section. It is interesting to note that the displacement interpolation can be interpreted naturally as the linear interpolation between the equal-weighted portfolio and the terminal portfolio.

Theorem 10.5.2 (Displacement interpolation)**.** *Consider the setting of Section* 10.5.3*.*

(i) For each $t \in [0,1]$, the portfolio map $\pi^{(t)}$ is generated by the exponentially concave function $\varphi^{(t)}$ on Δ_n defined by

$$\varphi^{(t)}(p) = (1-t)\sum_{i=1}^{n} \frac{1}{n} \log p_i + t\varphi(p), \quad p \in \Delta_n. \tag{10.5.10}$$

(ii) For each $t \in [0,1]$, let $f^{(t)} = \varphi^{(t)} + \psi$ and let $F^{(t)} = \nabla^c f^{(t)}$. If θ is distributed according to $P^{(0)}$, then $\theta^{(t)}$ is distributed according to $P^{(t)}$ where

$$P^{(t)} = \left(F^{(t)}\right)_{\#} P^{(0)}.$$

Moreover, $F^{(t)}$ is an optimal transport map for the transport problem for $(P^{(0)}, P^{(t)})$.

(iii) Endow Δ_n with the geometric structure induced by the L-divergence of φ. We further assume that the c-gradient $F^{(1)} = \nabla^c f : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ is surjective. For each $\theta \in \mathbb{R}^{n-1}$ fixed, consider the curve $t \mapsto \varphi^{(t)}(\theta)$ in dual coordinates. Then the trace of the curve is the dual geodesic joining θ and $\varphi^{(1)}(\theta)$.

Proof. (i) Follows directly from Proposition 4.3.3.

(ii) It is clear that $(\theta, F^{(t)}(\theta))$ is a coupling of $(P^{(0)}, P^{(t)})$. By Theorem 6.3.1, the graph of the map $F^{(t)}$ is c-cyclical monotone. This proves that $F^{(t)}$ is an optimal transport map.

(iii) We write (10.5.8) in the form

$$\begin{aligned} e^{-\phi_i^{(t)}} &= e^{-\theta_i} \frac{(1-t)\frac{1}{n} + t\pi_i(\theta)}{(1-t)\frac{1}{n} + t\pi_n(\theta)} \\ &= \frac{(1-t)\frac{1}{n}}{(1-t)\frac{1}{n} + t\pi_n(\theta)} e^{-\theta_i} + \frac{t\pi_n(\theta)}{(1-t)\frac{1}{n} + t\pi_n(\theta)} e^{-\theta_i - \log \pi_n(\theta)} \\ &=: (1-h(t))e^{-\theta_i} + h(t)e^{-\phi_i^{(1)}(\theta)}. \end{aligned}$$

By (10.4.7) we see that $t \mapsto \phi^{(t)}$ is a time change of a dual geodesic. The surjectivity assumption guarantees that the curve lies within $\mathcal{Y}'$, the range of the dual coordinate system.

□

10.5.4 Another interpolation

From the financial perspective there is another natural interpolation, namely the linear interpolation between the market portfolio μ and the portfolio π:

$$\pi^{(t)} = (1-t)\pi + t\mu. \tag{10.5.11}$$

The corresponding log generating function is $\varphi^{(t)} = (1-t)\varphi$. From the transport perspective, the market portfolio corresponds to the trivial transport map $F(\theta) \equiv 0$ (recall in Theorem 6.3.1(iii) that the portfolio has exponential coordinate given by $\theta - F(\theta)$). By the argument of Theorem 10.5.2 we have the following result.

Proposition 10.5.3. *Consider the geometric structure induced by φ and assume that the range of the dual coordinate system is $\mathbb{R}^{n-1}$. Consider the flow $(t, \theta) \mapsto \phi^{(t)}(\theta)$ in (10.5.8) where $\pi^{(t)}$ is given by the interpolation (10.5.11). Then for each θ, in dual coordinates, the trace of the curve $t \mapsto \phi^{(t)}(\theta)$ is a time change of the dual geodesic from $\phi^{(0)}(\theta)$ to 0.*

BIBLIOGRAPHY

[1] Paul H Algoet and Thomas M Cover. Asymptotic optimality and asymptotic equipartition properties of log-optimum investment. *The Annals of Probability*, pages 876–898, 1988.

[2] Shun-ichi Amari. *Information Geometry and Its Applications*. Springer, 2016.

[3] Shun-ichi Amari and Hiroshi Nagaoka. *Methods of information geometry*. American Mathematical Soc., 2007.

[4] Luigi Ambrosio and Nicola Gigli. A users guide to optimal transport. In *Modelling and optimisation of flows on networks*, pages 1–155. Springer, 2013.

[5] Nihat Ay and Shun-ichi Amari. A novel approach to canonical divergences within information geometry. *Entropy*, 17(12):8111–8129, 2015.

[6] Kerry Back. *Asset pricing and portfolio choice theory*. Oxford University Press, 2010.

[7] Adrian D. Banner and Daniel Fernholz. Short-term relative arbitrage in volatility-stabilized markets. *Annals of Finance*, 4(4):445–454, 2008.

[8] Adrian D Banner, Robert Fernholz, and Ioannis Karatzas. Atlas models of equity markets. *The Annals of Applied Probability*, 15(4):2296–2330, 2005.

[9] Erhan Bayraktar, Yu-Jui Huang, and Qingshuo Song. Outperforming the market portfolio with a given probability. *The Annals of Applied Probability*, pages 1465–1494, 2012.

[10] Jean-David Benamou, Brittany D. Froese, and Adam M. Oberman. Numerical solution of the optimal transportation problem using the Monge–Ampère equation. *Journal of Computational Physics*, 260:107–126, 2014.

[11] Peter L Bernstein. *Capital ideas: the improbable origins of modern Wall Street*. Simon and Schuster, 1993.

[12] Peter L Bernstein. *Capital ideas evolving*. John Wiley & Sons, 2011.

[13] Paul Bouchey, Vassilii Nemtchinov, Alex Paulsen, and David M Stein. Volatility harvesting: Why does diversifying and rebalancing create portfolio growth? *The Journal of Wealth Management*, 15(2):26, 2012.

[14] Paul Bouchey, Vassilii Nemtchinov, and Ting-Kam Leonard Wong. Volatility harvesting in theory and practice. *The Journal of Wealth Management*, 18(3):89–100, 2015.

[15] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[16] Lev M Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217, 1967.

[17] L. Breiman. Optimal gambling systems for favorable games. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, pages 65–78. Univ. California Press, Berkeley, Calif., 1961.

[18] Marcel R. Brod. Generating the Universal Portfolio. Master's thesis, ETH, Switzerland, 2014.

[19] Ovidiu Calin and Constantin Udrişte. *Geometric modeling in probability and statistics*. Springer, 2014.

[20] Álvaro Cartea, Sebastian Jaimungal, and José Penalva. *Algorithmic and High-Frequency Trading*. Cambridge University Press, 2015.

[21] I. Castillo and R. Nickl. Nonparametric bernstein–von mises theorems in gaussian white noise. *The Annals of Statistics*, 41(4):1999–2028, 2013.

[22] D. R. Chambers and J. S. Zdanowicz. The limitations of diversification return. *Journal of Portfolio Management*, 40(4):65–76, 2014.

[23] Sourav Chatterjee and Soumik Pal. A phase transition behavior for brownian motions interacting through their ranks. *Probability theory and related fields*, 147(1-2):123–159, 2010.

[24] L. B. Chincarini and D. Kim. *Quantitative Equity Portfolio Management: An Active Approach to Portfolio Construction and Management*. McGraw-Hill Library of Investment and Finance. McGraw-Hill, 2006.

[25] Martin Chuaqui, Peter Duren, and Brad Osgood. Schwarzian derivative criteria for valence of analytic and harmonic mappings. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 143, pages 473–486. Cambridge Univ Press, 2007.

[26] Martin Chuaqui, Peter Duren, Brad Osgood, and Dennis Stowe. Oscillation of solutions of linear differential equations. *Bulletin of the Australian Mathematical Society*, 79(01):161–169, 2009.

[27] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 2006.

[28] Tomas M. Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, 1991.

[29] Jason E. Cross and Andrew R. Barron. Efficient universal portfolios for past-dependent target classes. *Mathematical Finance*, 13(2):245–276, 2003.

[30] Madeleine Cule and Richard Samworth. Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electronic Journal of Statistics*, 4:254–270, 2010.

[31] Madeleine Cule, Richard Samworth, and Michael Stewart. Maximum likelihood estimation of a multi-dimensional log-concave density. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(5):545–607, 2010.

[32] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*. Springer, 1998.

[33] Michael AH Dempster, Igor V Evstigneev, and Klaus R Schenk-Hoppé. Volatility-induced financial growth. *Quantitative Finance*, 7(2):151–160, 2007.

[34] Franki Dillen, Katsumi Nomizu, and Luc Vranken. Conjugate connections and radon's theorem in affine differential geometry. *Monatshefte für Mathematik*, 109(3):221–235, 1990.

[35] Shinto Eguchi. Second order efficiency of minimum contrast estimators in a curved exponential family. *The Annals of Statistics*, 11(3):793–803, 1983.

[36] Shinto Eguchi. Geometry of minimum contrast. *Hiroshima Math. J*, 22(3):631–647, 1992.

[37] Claude B Erb and Campbell R Harvey. The strategic and tactical value of commodity futures. *Financial Analysts Journal*, 62(2):69–97, 2006.

[38] Eugene F Fama and Kenneth R French. The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427–465, 1992.

[39] Daniel Fernholz and Ioannis Karatzas. On optimal arbitrage. *The Annals of Applied Probability*, 20(4):1179–1204, 2010.

[40] Daniel Fernholz and Ioannis Karatzas. Optimal arbitrage under model uncertainty. *The Annals of Applied Probability*, 21(6):2191–2225, 2011.

[41] E. Robert Fernholz. *Stochastic Portfolio Theory.* Springer, 2002.

[42] R. Fernholz and B. Shay. Stochastic portfolio theory and stock market equilibrium. *The Journal of Finance*, 37(2):615–624, 1982.

[43] Robert Fernholz. Portfolio generating functions. *Quantitative Analysis in Financial Markets, River Edge, NJ. World Scientific*, 1999.

[44] Robert Fernholz. Equity portfolios generated by functions of ranked market weights. *Finance and Stochastics*, 5(4):469–486, 2001.

[45] Robert Fernholz, Robert Garvy, and John Hannon. Diversity-weighted indexing. *The Journal of Portfolio Management*, 24(2):74–82, 1998.

[46] Robert Fernholz, Tomoyuki Ichiba, and Ioannis Karatzas. A second-order stock market model. *Annals of Finance*, 9(3):439–454, 2013.

[47] Robert Fernholz and Ioannis Karatzas. Relative arbitrage in volatility-stabilized markets. *Annals of Finance*, 1(2):149–177, 2005.

[48] Robert Fernholz and Ioannis Karatzas. Stochastic portfolio theory: an overview. In P. G. Ciarlet, editor, *Handbook of Numerical Analysis*, volume 15 of *Handbook of Numerical Analysis*, pages 89 – 167. Elsevier, 2009.

[49] Robert Fernholz, Ioannis Karatzas, and Constantinos Kardaras. Diversity and relative arbitrage in equity markets. *Finance and Stochastics*, 9(1):1–27, 2005.

[50] Booth David G. and Fama Eugene F. Diversification returns and asset contributions. *Financial Analysts Journal*, 48(3), 1992.

[51] W. G. Hallerbach. Disentangling rebalancing return. *Journal of Asset Management*, 15(5):301–316, 2014.

[52] Joel Hasbrouck. *Empirical market microstructure: The institutions, economics, and econometrics of securities trading.* Oxford University Press, 2006.

[53] E. Hazan and S. Kale. An online portfolio selection algorithm with regret logarithmic in price variation. *Mathematical Finance*, 25(2):288–310, 2015.

[54] D. P. Helmbold, R. E. Schapire, Y. Singer, and M. K. Warmuth. On-line portfolio selection using multiplicative updates. *Mathematical Finance*, 8(4):325–347, 1998.

[55] Lars Hörmander. *Notions of convexity.* Springer, 2007.

[56] Tomoyuki Ichiba, Soumik Pal, and Mykhaylo Shkolnikov. Convergence rates for rank-based models with applications to portfolio theory. *Probability Theory and Related Fields*, 156(1-2):415–448, 2013.

[57] Tomoyuki Ichiba, Vassilios Papathanakos, Adrian Banner, Ioannis Karatzas, Robert Fernholz, et al. Hybrid atlas models. *The Annals of Applied Probability*, 21(2):609–644, 2011.

[58] Farshid Jamshidian. Asymptotically optimal portfolios. *Mathematical Finance*, 2(2):131–150, 1992.

[59] Chloé Jimenez and Filippo Santambrogio. Optimal transportation for a quadratic cost with convex constraints and applications. *Journal de mathématiques pures et appliquées*, 98(1):103–113, 2012.

[60] Ioannis Karatzas and Johannes Ruf. Trading strategies generated by lyapunov functions. ArXiv e-prints 1603.08245, 2016.

[61] Constantinos Kardaras, Dörte Kreher, and Ashkan Nikeghbali. Strict local martingales and bubbles. *The Annals of Applied Probability*, 25(4):1827–1867, 2015.

[62] Robert E Kass and Paul W Vos. *Geometrical foundations of asymptotic inference.* John Wiley & Sons, 2011.

[63] J. Kelly. A new interpretation of information rate. *Bell Systems Technical Journal*, 35:917–926, 1956.

[64] Yves-Laurent Kom Samo and Alexander Vervuurt. Stochastic portfolio theory: A machine learning perspective. ArXiv e-prints 1605.02654, 2016.

[65] Takashi Kurose. Dual connections and affine geometry. *Mathematische Zeitschrift*, 203(1):115–121, 1990.

[66] John M Lee. *Riemannian manifolds: an introduction to curvature*. Springer, 2006.

[67] Bin Li and Steven CH Hoi. Online portfolio selection: A survey. *ACM Computing Surveys (CSUR)*, 46(3):35, 2014.

[68] John Lott and Cédric Villani. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics*, 169(3):903–991, 2009.

[69] David G. Luenberger. *Investment science*. Oxford University Press, 1998.

[70] Leonard C MacLean, Edward O Thorp, and William T Ziemba. *The Kelly capital growth investment criterion: Theory and practice*, volume 3. world scientific, 2011.

[71] Burton G Malkiel. Returns from investing in equity mutual funds 1971 to 1991. *The Journal of finance*, 50(2):549–572, 1995.

[72] Burton G Malkiel. The efficient market hypothesis and its critics. *The Journal of Economic Perspectives*, 17(1):59–82, 2003.

[73] Burton G Malkiel and Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417, 1970.

[74] Burton Gordon Malkiel. *A random walk down Wall Street: including a life-cycle guide to personal investing*. WW Norton & Company, 1999.

[75] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[76] Michael K Murray and John W Rice. *Differential geometry and statistics*, volume 48. CRC Press, 1993.

[77] Hiroshi Nagaoka and Shun-ichi Amari. Differential geometry of smooth families of probability distributions. *Univ. Tokyo, Tokyo, Japan, METR*, pages 82–7, 1982.

[78] Rishi K. Narang. *Inside the Black Box: A Simple Guide to Quantitative and High Frequency Trading*, volume 883. John Wiley & Sons, 2013.

[79] Soumik Pal. Exponentially concave functions and high dimensional stochastic portfolio theory. ArXiv e-prints 1603.01865, 2016.

[80] Soumik Pal and Philip Protter. Analysis of continuous strict local martingales via h-transforms. *Stochastic Processes and their Applications*, 120(8):1424–1443, 2010.

[81] Soumik Pal and Ting-Kam Leonard Wong. Energy, entropy, and arbitrage. ArXiv e-prints 1308.5376, 2013.

[82] Soumik Pal and Ting-Kam Leonard Wong. Exponentially concave functions and a new information geometry. ArXiv e-prints 1605.05819, 2016.

[83] Soumik Pal and Ting-Kam Leonard Wong. The geometry of relative arbitrage. *Mathematics and Financial Economics*, 10:263–293, 2016.

[84] Y. Plyakha, R. Uppal, and G. Vilkov. Why does an equal-weighted portfolio outperform value-and price-weighted portfolios? *Available at SSRN 1787045*, 2012.

[85] William Poundstone. *Fortune's Formula: The untold story of the scientific betting system that beat the casinos and wall street*. Macmillan, 2010.

[86] John Rainwater. Yet more on the differentiability of convex functions. *Proceedings of the American Mathematical Society*, 103(3):773–778, 1988.

[87] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1997.

[88] Mark Rubinstein. *A history of the theory of investments: My annotated bibliography*. John Wiley & Sons, 2011.

[89] Johannes Ruf. Hedging under arbitrage. *Mathematical Finance*, 23(2):297–317, 2013.

[90] Paul A Samuelson. The fallacy of maximizing the geometric mean in long sequences of investing or gambling. *Proceedings of the National Academy of Sciences*, 68(10):2493–2496, 1971.

[91] William F Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964.

[92] Hirohiko Shima. *The Geometry of Hessian Structures*. World Scientific, 2007.

[93] G. Shub, B. Beardsley, H. Donnadieu, K. Kramer, M. Kumar, A. Maguire, P. Morel, and T. Tang. Global asset management 2013: Capitalizing on the recovery. *Boston Consulting Group*, 2013.

[94] Shlomo Sternberg. *Curvature in mathematics and physics*. Dover, 2012.

[95] Winslow Strong. Generalizations of functionally generated portfolios with applications to statistical arbitrage. *SIAM Journal on Financial Mathematics*, 5(1):472–492, 2014.

[96] Sara A van de Geer. *Empirical Processes in M-estimation.* Cambridge University Press, 2000.

[97] Aad W. Van der Vaart. *Asymptotic statistics*. Cambridge University Press, 2000.

[98] Alexander Vervuurt. Topics in stochastic portfolio theory. ArXiv e-prints 1504.02988, 2015.

[99] Alexander Vervuurt and Ioannis Karatzas. Diversity-weighted portfolios with negative parameter. *Annals of Finance*, 11(3-4):411–432, 2015.

[100] Cédric Villani. *Topics in optimal transportation*. American Mathematical Soc., 2003.

[101] Cédric Villani. *Optimal transport: old and new*. Springer, 2008.

[102] Scott Willenbrock. Diversification return, portfolio rebalancing, and the commodity return puzzle. *Financial Analysts Journal*, 67(4):42–49, 2011.

[103] Ting-Kam Leonard Wong. Optimization of relative arbitrage. *Annals of Finance*, 11(3):345–382, 2015.

[104] Ting-Kam Leonard Wong. Universal portfolios in stochastic portfolio theory. ArXiv e-prints 1510.02808, 2015.