THE GEOMETRY OF RELATIVE ARBITRAGE

SOUMIK PAL AND TING-KAM LEONARD WONG

Abstract. Consider an equity market with \( n \) stocks. The vector of proportions of the total market capitalization that belongs to each stock is called the market weights. Consider two portfolios, one is a passive buy-and-hold portfolio representing the entire market, and the other assigns a portfolio vector for each possible value of the market weights and requires trading to maintain this assignment. The evolution of stocks is taken to be any jump process in discrete time, or a path of any continuous semimartingale in continuous time. We do not make any stochastic modeling assumptions on the evolutions. We provide necessary and sufficient conditions on the assignment on portfolios to guarantee that, for all such evolutions in a diverse and sufficiently volatile market, the actively traded portfolio outperforms the buy-and-hold portfolio in the long run. This class of ‘relative arbitrage’ portfolios were discovered by Fernholz \cite{Fer99} and are called functionally generated portfolios. We prove that, in an appropriate sense, a slight generalization of these are the only possible ones. Remarkably, such portfolios correspond to solutions of Monge-Kantorovich optimal transport problems on the unit simplex with a cost function that can be described as log of the partition function. We also show how the presence of microstructure noise in stock price data leads to statistical arbitrage in high-frequency trading. Our primary tool is a property of multidimensional functions that we call multiplicative cyclical monotonicity.

1. Introduction

Consider investing in an equity market. At each point in time the investor allocates some proportions of his entire wealth into each of the stocks, forming a portfolio. In this article a portfolio is assumed to be a self-financing trading strategy, initially starting with one dollar, that is fully invested in the stock market, is long-only, and is never allowed to borrow from the money market. Mathematically, recall the unit simplex in dimension \( n \) to be the set

\[
\Delta^n = \left\{ p = (p_1, \ldots, p_n) : p_i \geq 0 \text{ for all } i, \text{ and } \sum_{i=1}^{n} p_i = 1 \right\}.
\]

A portfolio at any point of time is represented by a vector in the unit simplex in dimension given by the number of stocks. The individual coordinates of this vector are called portfolio weights. Over time this leads to a process with the unit simplex as the state space.

For example, a buy-and-hold portfolio is one where, initially, one buys a certain number of shares of each stock and holds them for all future time. The value of
this portfolio (i.e., total money generated as a process in time) can be thought of as reflecting the growth of the entire market. This is the case of the so-called market portfolio where the portfolio vector at any point of time is proportional to the capitalizations (called cap-weighted) of the individual firms. The portfolio weights of the market portfolio are called market weights and the value of this portfolio is called a market index. For example, in the US equity market, a standard benchmark is the S&P500 index which can be thought of (approximately) as the value process of a buy-and-hold portfolio.

The skill of an active investor is often judged by comparing the value of the portfolio with that of a market index. Informally, the portfolio is said to beat the market if its value tends to be higher than that of the index, normalized to have started with one dollar. Of course, both value processes are one-dimensional random processes and one has to make this concept precise. Leaving this formalism for later we ask the million-dollar question: when and how can we beat the market? That this is indeed sometimes possible has been widely observed; see the recent popular article in *The Economist* [Eco13]. Intuitively it is understood that the performance of an actively traded portfolio can be boosted by a careful buy-low-sell-high strategy when the market is volatile. A major step towards making this mathematically precise is the development of stochastic portfolio theory outlined in the book by Fernholz [Fer02]. We refer the reader to the survey by Fernholz and Karatzas [FK09] for a more recent exposition. This theory identifies two fundamental sufficient conditions. Namely, if the market is diverse, meaning that none of the stock is ever allowed to dominate the entire market, and sufficiently volatile, one can construct explicit portfolios whose value will be greater than that of the market portfolio after a finite (but large) time, with probability one. These portfolios are called relative arbitrages with respect to the market. A remarkable fact is that these portfolios can be chosen to depend only on the current market weights of the stocks. In other words, as long as the two conditions hold, forecasts of expected returns and the covariance matrix are not needed to beat the market (in the long run). Moreover these portfolios, being easily replicable and typically more diversified than the market portfolio, may serve as enhanced indexing strategies and alternative benchmarks; see Fernholz et al. [FGH98] and the more recent survey by Chow et al. [CHKL11] for other alternative indexing strategies.

For a precise statement of this result and the definitions of the italicized terms please consult [FK09 Section 7 and Section 11]. To appreciate our current results, we provide an informal description of such functionally generated portfolios in the set-up of [Fer02]. In that set-up, individual stock prices are modeled as a general continuous semimartingale. Let \( \mu(t) = (\mu_1(t), \ldots, \mu_n(t)) \) denote the market portfolio at time \( t \) where \( n \) is the number of stocks. Let us recall that \( \mu(t) \) is a vector in the unit simplex in dimension \( n \). By assumption, it is a continuous semimartingale. Suppose we are given a twice continuously differentiable positive function \( G \) defined on an open neighborhood of \( \Delta^n \) in \( \mathbb{R}^n \). Let \( D_i \) represent the partial derivative with respect to the \( i \)th coordinate in \( \mathbb{R}^n \). Assume that the maps \( x \mapsto x_i D_i \log G(x) \) are bounded for all \( i \). The portfolio generated by \( G \) is given by the weights

\[
\pi_i(t) = \left( D_i \log G(\mu(t)) + 1 - \sum_{j=1}^{n} \mu_j(t) D_j \log G(\mu(t)) \right) \mu_i(t), \quad i = 1, 2, \ldots, n.
\]
These portfolios were introduced by Fernholz in [Fer99] who showed by elementary but enigmatic stochastic calculus that, when $G$ is concave, (some of) these portfolios are relative arbitrages. We stress that relative arbitrages have to be model-free, i.e., not dependent on the probability law of the market but rather on weak structural assumptions such as diversity and sufficient volatility. The following question was posed [FK09, Remark 11.5, page 142]: are there portfolios $\pi$ which are not functionally generated by concave functions, and yet are relative arbitrages under similar assumptions? Pal and Wong [PW13] showed that the answer is yes by exhibiting a class of portfolios (called energy-entropy portfolios) that are relative arbitrages under suitable conditions. However, these portfolios depend on the entire history of the market weights. In this article we show that if our portfolios are restricted to be functions of the current market weights, then the answer to the above question is essentially no, irrespective of discrete or continuous time.

Our objectives in this current paper are: (i) to dispel the mystery of the formula (2) by providing intuition behind the construction, and (ii) to settle the posed question above. However these require expanding the definitions of relative arbitrage, diversity, and the concept of sufficient volatility. Below we describe these concepts informally. Details are left for Section 1.1.

As noted in the abstract, time is either discrete or continuous. When time is discrete we assume that the market process is a deterministic sequence of elements in $\Delta^n$. When time is continuous we say that the market follows a typical path of an arbitrary continuous semimartingale. This is a definition that means the following. A property of continuous paths in $\Delta^n$ is said to hold for a typical path of an arbitrary continuous semimartingale if it holds almost surely for any continuous semimartingale with state space given by the unit simplex.

Suppose $U$ is a subset of $\Delta^n$. Assume that the market weights stays within $U$ forever. Consider a portfolio $\pi$ given as a function of the market weights, and consider the ratio of the value of this portfolio to the value of the market portfolio. This will be called the relative value process. The precise definition of a pseudo-arbitrage opportunity is given in the following subsection. Informally, a portfolio $\pi$ will be called a pseudo-arbitrage opportunity on $U$ if the relative value process is bounded below by a positive constant uniformly over all paths (discrete or continuous) of the market weights and (i) if time is discrete, there are some paths along which this ratio tends to infinity with time, while (ii) if time is continuous, there are semimartingale models for which, with positive probability, this ratio tends to infinity with time. That is, the downside risk is uniformly bounded below regardless of the market movements, and there is a possibility of unbounded gain. Pseudo-arbitrages become relative arbitrages when $U$ is the simplex with a neighborhood of each corner vector $e(i)$ removed and certain stronger assumptions on the volatility matrix are assumed to guarantee that the ratio tends to infinity almost surely.

In Section 3 we prove that when the market weights are discrete jump processes the only possible pseudo-arbitrage strategies are given by a slight generalization of the functionally generated portfolios. This result remains true even if we assume that the size of the jumps is bounded above by an arbitrarily small constant. Since the only way to implement continuous trading is by (possibly high-frequency) discrete approximation, this provides an answer to the open question mentioned above. Our argument is convex-analytic and relies on a novel property we call multiplicative cyclical monotonicity, which is a variant of the classical cyclical monotonicity.
property. This property will be defined and studied in Section 2. Intuitively, this property requires that the portfolio will not underperform if the market goes over any discrete loop or excursion in the unit simplex. Remarkably, this implies the portfolio is functionally generated and is a pseudo-arbitrage, in both discrete and continuous time, without any modeling assumptions.

In practice it is rather unintuitive to create portfolios out of a generating function. The generating function has limited financial meaning and an investor might have several restrictions on the kind of portfolios that can be used. In Section 4 we take a different route to create pseudo-arbitrages. The map that takes a vector of market weights \( \mu \) to the corresponding vector of portfolio weights \( \pi \) can be thought of as a coupling or assignment of vectors on the unit simplex. We show that this assignment can be constructed by solving a Monge-Kantorovich optimal transport problem where the cost function can be informally described as the log of the partition function. In fact, we prove that all pseudo-arbitrages can be realized as solutions of this optimal transport problem.

Due to the presence of transaction costs and liquidity issues, stocks are not traded continuously. For example, the portfolio may be rebalanced - traded to achieve the target portfolio weights - every quarter. More generally, rebalancing actions can be triggered when the deviation of the current portfolio weights from the targeted ones exceeds a specified amount. We consider trigger-based rebalancing in the context of functionally generated portfolios and study how the discretely rebalanced portfolio resembles and differs from the continuously traded counterpart. We conclude with an analysis of the presence of statistical arbitrage opportunities in high frequency trading, based on the observations and ideas of Fernholz and Maguire, Jr. [FM07]. We prove the following result: Suppose high-frequency stock data is modeled as a hidden semimartingale model as in Zhang, Mykland, and Aït-Sahalia [ZMAS05], then, in a sense to be made precise, trading more frequently allows more gain from volatility to be captured.

Apart from the references mentioned above, the following articles are closely related to our current work. The roles of diversity and sufficient volatility in relative arbitrage have been studied by authors such as Fernholz and Karatzas [FK05] and Fernholz, Karatzas, and Kardaras [FKK05]. Functionally generated portfolio is used by Banner and Fernholz [BF08] to prove that sufficient volatility of the smallest stock implies the existence of short term relative arbitrages. Generalizations of functionally generated portfolios to stochastic generating functions and application to statistical arbitrage have been done in [Str12]. A recent trend in mathematical finance is to study models with model uncertainty and robust pricing and hedging of contingent claims. For example, Fernholz and Karatzas [FK11] characterizes optimal relative arbitrage when the covariance matrix is uncertain. The theory of optimal transport also arises in this context, see, for example, Beiglböck et al. [BHL13], although it is quite different from the one considered here.

1.1. Definitions and notations. The ambient space of our analysis is the unit simplex \( \Delta^n \). All topological and measure-theoretic aspects will be relative to the unit simplex. Here and everywhere following, for any two vectors \( p \) and \( q \) where \( q \) has positive components, the vector \( p/q \) refers to the vector of their coordinate-wise ratios. The vector \( p \cdot q \) will refer to their coordinate-wise product. The notation \( \langle x, y \rangle \) will refer to the Euclidean inner product of two vectors \( x \) and \( y \), and \( \| x \| \) is the...
Euclidean norm of \( x \). By an abuse of notation \( \langle X, Y \rangle \) will also refer to the quadratic covariation of continuous semimartingales \( X \) and \( Y \), and we write \( \langle X \rangle = \langle X, X \rangle \).

Let \( \mu(t) = (\mu_1(t), \ldots, \mu_n(t)) \) be the process of market weights with state space \( \Delta^n \). Here the time parameter set can be \( \{0, 1, 2, \ldots\} \) (discrete) or \( \mathbb{R}_+ \) (continuous). In discrete time \( \{\mu(t)\} \) can be any deterministic path. If time is continuous we assume that the process \( \{\mu(t)\} \) is a typical path of an arbitrary continuous semimartingale, as defined in the last section. Sometimes we will also use \( \mu \) to denote a generic point of \( \Delta^n \), and this should cause no confusion.

In this paper, we only consider portfolios which are functions of the current market weights. Hence we have the following restrictive definition.

**Definition 1** (Portfolio). Let \( U \) be a measurable subset of \( \Delta^n \). By a portfolio \( \pi \) on \( U \) we mean a Borel measurable map \( \pi : U \to \Delta^n \). We will also write \( \pi(t) = \pi(\mu(t)) \) as the portfolio weights process.

We will be interested in the performance of the portfolio relative to the market, and we will measure this by the ratio \( V \) of the value of this portfolio to the value of the market. Thus

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

and we will call this the relative value process of the portfolio. Here is the rigorous definition.

**Definition 2** (Relative value process). Let \( \pi \) be a portfolio. We will denote by \( V(\cdot) \) the relative value process when implementing the portfolio \( \pi \). That is, \( V \) is a positive process satisfying \( V(0) = 1 \) and

\[
\frac{V(t+1)}{V(t)} = \sum_{i=1}^{n} \frac{\pi_i(t)\mu_i(t+1)}{\mu_i(t)}, \quad \text{in discrete time, and}
\]

\[
\frac{dV(t)}{V(t)} = \sum_{i=1}^{n} \pi_i(t) \frac{d\mu_i(t)}{\mu_i(t)}, \quad \text{in continuous time.}
\]

A justification of (3) can be found in Lemma 2.1 of \([PW13]\). Here we make the simplifying assumptions that the number of firms is fixed, they do not merge or break up, and there are no transaction costs.

For later use, we note that for the discrete case we may write

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

We call \( w = \pi/\mu \) the vector of weight ratios. We also note that for any \( \pi \in \overline{\Delta^n} \) and \( p, q \in \Delta^n \),

\[
1 + \left\langle \frac{\pi}{p}, q - p \right\rangle = \sum_{i=1}^{n} \pi_i + \sum_{i=1}^{n} \frac{\pi_i(q_i - p_i)}{p_i} = \sum_{i=1}^{n} q_i \pi_i > 0.
\]

This verifies that \( V(t) \) is positive for all \( t \).

In our formulation, the value of the market portfolio serves as the numéraire, so we only need to specify the evolution of the market weights. We have used the same set up in \([PW13]\). This is similar to, but different from, the usual one
in stochastic portfolio theory, where the market capitalizations of stocks are the primary objects. Our approach treats the unit simplex intrinsically and emphasizes the fact that, if the relative value $V(t)$ is the main concern, the evolution of market weights contains all relevant information. For example, if all stocks experience the same returns, however large in magnitude, the market weights will not change, and so the relative value stays constant for any portfolio.

Next we introduce the concept of pseudo-arbitrage opportunity. As explained in the introduction, it is a portfolio $\pi$ such that $V(t)$ is required to be bounded below by a positive constant for every path of $\mu$ as long as it stays within a fixed set $U$, while, potentially, growing unbounded in the long-run.

**Definition 3** (Pseudo-arbitrage). Let $U$ be a measurable subset of $\Delta^n$. We say that a portfolio $\pi$ on $U$ is a pseudo-arbitrage on $U$ if there is a constant $C \geq 0$ such that

(i) (discrete time) for every sequence $\omega = \{\mu(t), t = 0, 1, 2, \ldots\} \subset U$, or

(ii) (continuous time) for every typical path $\omega = \{\mu(t), t \geq 0\}$ of an arbitrary continuous semimartingale with state space $U$,

the relative value of the portfolio with respect to the market allows the following decomposition

$$\log V(t) + C \geq A(t),$$

where $t \mapsto A(t) = A(\omega, t)$ is nonnegative, nondecreasing, starting at zero, and is not identically zero in $\omega$.

The process $A$ will be referred to as the cumulative energy. It depends on the portfolio and the market weight process and is closely related to volatility of market weights. In continuous time, not identically zero in $\omega$ means that there are continuous semimartingale models for which $\lim A(\omega, t)$ is strictly positive (possibly infinity) with positive probability.

The definition of pseudo-arbitrage looks rather weak, especially since we do not demand any bound on the growth of cumulative energy. However, as we will see, the definition puts a rather severe restriction on portfolios. In Theorem 7 we will characterize these as functionally generated portfolios whose generating functions are concave and not affine. The cumulative energy will go to infinity in time with probability one under natural assumptions on the volatility of market weights.

In Definition 3 we only require $A$ to be not identically zero. This provides conditions under which the portfolio can beat the market. The following simple observation shows it implies that $A$ tends to infinity for certain paths.

**Lemma 1.** Let $\pi$ be a pseudo-arbitrage opportunity. Then there exist paths $\omega = \{\mu(t)\}$ such that $A(\omega, \cdot)$ grows to infinity.

**Proof.** By assumption, there exists $t_0 > 0$ and paths $\omega = \{\mu(t)\}_{0 \leq t \leq t_0}$ such that $A(\omega, t_0) > 0$. Construct a path $\omega'$ which is the infinite concatenation of $\omega$ (if time is continuous, we insert a continuous path from $\omega(t_0)$ to $\omega(0)$ between each consecutive pairs). Since $A$ is non-decreasing by definition, it is easy to see that $\lim_{t \to \infty} A(\omega', t) = \infty$. \qed

We emphasize that the collection of ‘good paths’ (i.e., those such that $A(t)$ grows large) depends on the portfolio. As we shall see below, this is determined by the geometry of the portfolio map.
2. Convex analysis on the unit simplex

We begin by reviewing some notions of convex analysis, specializing to the unit simplex. A standard reference is the book [Roc97] by Rockafellar and we will refer to it frequently. Recall that a function $\Phi : \Delta^n \to \mathbb{R}$ is concave if for any $p, q \in \Delta^n$ and any $0 < \alpha < 1$, we have

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

Let $\Phi$ be a concave function on $\Delta^n$. The superdifferential (or supergradient) of $\Phi$ is a multi-valued function $\partial \Phi$ from $\Delta^n$ to $\mathbb{R}^n$. Let $p \in \Delta^n$. By definition, we say that a vector $\xi \in \mathbb{R}^n$ belongs to $\partial \Phi(p)$ if for all $q \in \Delta^n$ we have

$$\Phi(q) \leq \Phi(p) + \langle \xi, q - p \rangle. \tag{6}$$

Thus $\xi$ defines a supporting hyperplane of the hypograph of $\Phi$ at $(p, \Phi(p))$. It is important for our purpose that we allow $\xi \in \mathbb{R}^n$. Since $q - p$ is parallel to $\Delta^n$, $\partial \Phi(p)$ is a union of straight lines perpendicular to $\Delta^n$.

We will also consider differentiable functions on $\Delta^n$ and their derivatives. Implicitly, we regard $\Delta^n$ as an $(n-1)$-dimensional manifold embedded in $\mathbb{R}^n$, equipped with the global coordinate system $\varphi(p) = (p_1, \ldots, p_{n-1})$. For example, a function $\Phi$ on $\Delta^n$ is of class $C^k$ ($k$ times continuously differentiable) if the push forward $\Phi \circ \varphi^{-1}$ is of class $C^k$ on the open set $\varphi(\Delta^n)$ in $\mathbb{R}^{n-1}$.

For $i = 1, \ldots, n$, let $e(i) = (0, \ldots, 1, \ldots, 0)$ be the vertex of $\Delta^n$ in the $x_i$-direction. A tangent vector of $\Delta^n$ is a vector $v = (v_1, \ldots, v_n)$ satisfying $\sum_{i=1}^n v_i = 0$. Note that for any $p \in \Delta^n$ and $i = 1, \ldots, n$, the vector $e(i) - p$ is a tangent vector. Moreover, if $v$ is a tangent vector and $p \in \Delta^n$, we can write $v = \sum_{i=1}^n p_i(e(i) - p)$. If $\Phi$ is a function on $\Delta^n$, $p \in \Delta^n$ and $v$ is a tangent vector, we denote by $D_v \Phi(p)$ the one-sided directional derivative of $\Phi$ at $p$ in the direction $v$ whenever it exists. Explicitly, it is defined by

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

It is well known that if $\Phi$ is concave, then $D_v \Phi$ always exists. If $F$ is vector-valued and differentiable, $D.F(p)$ is the differential map (push forward of tangent vectors).

2.1. Multiplicative cyclical monotonicity. Let $\Phi : \Delta^n \to (0, \infty)$ be a concave function. Let $p, q \in \Delta^n$ and $\xi \in \partial \Phi(p)$. Dividing both sides of (6) by $\Phi(p)$, it follows that

$$1 + \frac{\langle \xi, q - p \rangle}{\Phi(p)} \geq \frac{\Phi(q)}{\Phi(p)}. \tag{7}$$

In particular, the left hand side of (7) is positive.

Taking log on both sides of (7) and using the inequality $\log(1 + x) \leq x$, for $x \geq -1$, we obtain

$$\frac{\langle \xi, q - p \rangle}{\Phi(p)} \geq \log \frac{\Phi(q)}{\Phi(p)}.$$ 

This shows that $\frac{1}{\Phi(p)} \partial \Phi(p)$ is a subset of $\partial \log \Phi(p)$, the superdifferential of the concave function $\log \Phi$ at $p$. If $\Phi$ is differentiable, these two sets are equal by the chain rule of calculus.

We now ask the following question. Let $U$ be a measurable subset of $\Delta^n$. Suppose $\pi : U \to \Delta^n$ is a map, possibly multivalued. Under what condition can we claim
the existence of a concave function \( \Phi : \Delta^n \to (0, \infty) \) such that \( \pi(p)/p \) is contained in \( \frac{\pi(p)}{p} \partial \Phi(p) \) for each \( p \)? If there is such a function \( \Phi \), we say that \( \pi \) is generated by \( \Phi \) or that \( \Phi \) is a generating function of \( \pi \).

**Definition 4** (Multiplicative cyclical monotonicity - MCM). Consider a (possibly multivalued) mapping \( \pi : U \subseteq \Delta^n \to \Delta^n \). It will be called multiplicative cyclically monotone (MCM) if, for any \( m \geq 1 \), and any set of points \( \{(p(i), \pi(i)) : i = 0, \ldots, m\} \subseteq U \times \Delta^n \) such that each \( \pi(i) \in \pi(p(i)) \), the following inequality holds:

\[
\left(1 + \frac{\pi(0)}{p(0)}, p(1) - p(0) \right) \left(1 + \frac{\pi(1)}{p(1)}, p(2) - p(1) \right) \ldots \left(1 + \frac{\pi(m)}{p(m)}, p(0) - p(m) \right) \geq 1.
\]

Equivalently, by taking log on both sides (which is legitimate by (5)), we get

\[
\sum_{i=0}^{m} \log \left(1 + \frac{\pi(i)}{p(i)}, p(i+1) - p(i) \right) \geq 0, \quad \text{where} \quad p(m+1) = p(0).
\]

The definition above is a multiplicative form of the classical cyclical monotonicity property in convex analysis [Roc97 Section 24]. A related definition is the following.

**Definition 5** (\( \delta \)-MCM). Let \( \pi : U \to \Delta^n \) be Borel measurable and \( \delta > 0 \). We say that \( \pi \) is \( \delta \)-MCM on \( U \) if the MCM inequality (8) holds for all cycles where the successive jump-sizes \( \|p(i+1) - p(i)\| \) are all less than \( \delta \).

**Lemma 2.** Let \( \pi : U \subseteq \Delta^n \to \Delta^n \) be a multivalued map.

(i) Suppose \( U \) is open. In order that a concave function \( \Phi : \Delta^n \to (0, \infty) \) exists such that \( \frac{\pi(p)}{p} \subseteq \frac{1}{\Phi(p)} \partial \Phi(p) \) for each \( p \in U \), it is necessary and sufficient that \( \pi \) is MCM.

(ii) Furthermore, if \( U \) is open and connected, the generating function \( \Phi \) is unique on \( U \) up to a positive multiplicative constant.

(iii) Suppose \( U \) is open and connected, and \( \pi : U \to \Delta^n \) satisfies \( \delta \)-MCM on \( U \). Then \( \pi \) satisfies the usual MCM property.

**Proof.** The proof is an adaptation of the proof of [Roc97 Theorem 24.8].

For the only if part in claim (i), suppose \( \Phi \) exists and \( \frac{\pi(p)}{p} \subseteq \frac{1}{\Phi(p)} \partial \Phi(p) \) for \( p \in U \). For any choice of points \( p(i), i = 0, 1, 2, \ldots, m \), in \( U \), and any choice of \( \xi(i) \in \pi(p(i)) \), we apply (7) to get

\[
\prod_{i=0}^{m} \left(1 + \left(\frac{\xi(i)}{p(i)}, p(i+1) - p(i) \right) \right) \geq \prod_{i=0}^{m} \frac{\Phi(p(i+1))}{\Phi(p(i))} = 1.
\]

The final equality holds since \( p(m+1) = p(0) \). Hence \( \pi \) is MCM.

For the if part, fix some \( p(0) \in U \). Define a function \( \Phi \) on \( \Delta^n \) by

\[
\Phi(p) = \inf \prod_{i=0}^{m} \left(1 + \left(\frac{\xi(i)}{p(i)}, p(i+1) - p(i) \right) \right),
\]

where \( p(m+1) = p \) and the infimum is taken over all \( m \geq 0 \) and all choices of points \( \{(p(i), \xi(i)) : i = 0, \ldots, m\} \) where \( p(i) \in U \) and \( \xi(i) \in \pi(p(i)) \).

Clearly \( \Phi \), being the pointwise infimum of a family of affine functions in \( p \), is a concave function on \( \Delta^n \). By (5), \( \Phi \) is non-negative, and the MCM property implies that \( \Phi(p(0)) = 1 \). Hence, it follows that \( \Phi \) must be everywhere positive.
We claim that \( \frac{\pi(p)}{p} \) is contained in \( \frac{1}{\Phi(p)} \partial \Phi(p) \). To see this, it suffices to show that for any two points \( p, q \in U \) and \( q \in \Delta^n \), and for any \( \alpha > \Phi(p) \), we have
\[
1 + \left\langle \frac{\xi}{p}, q - p \right\rangle \geq \frac{\Phi(q)}{\alpha},
\]
where \( \xi \in \pi(p) \).

Since \( \alpha > \Phi(p) \), there exists some \( m \geq 0 \) and some choice of \( \{p(1), \ldots, p(m)\} \subseteq U \), with \( p(m+1) = p \), such that
\[
\prod_{i=0}^{m} \left( 1 + \left\langle \frac{\xi(i)}{p(i)}, p(i+1) - p(i) \right\rangle \right) < \alpha, \quad \xi(i) \in \pi(p(i)).
\]
Setting \( p(m+2) = q \) and \( \xi = \xi(m+1) \in \pi(p) \), we have
\[
\Phi(q) \leq \left( 1 + \left\langle \frac{\xi}{p}, q - p \right\rangle \right) \alpha
\]
and claim (i) is proved.

To prove the uniqueness in claim (ii), suppose there are positive concave functions \( \Phi_1 \) and \( \Phi_2 \) such that
\[
w(p) := \frac{\pi(p)}{p} \in \frac{1}{\Phi_i(p)} \partial \Phi_i(p) \subseteq \partial \log \Phi_i(p), \quad i = 1, 2,
\]
for all \( p \in U \). Let \( p, q \in U \) be such that the line segment \( \ell \) from \( p \) to \( q \) lies in \( U \). Consider the restrictions of \( \log \Phi_i \) to \( \ell \), denoted by \( \log \Phi_i|_{\ell} \). They can be parametrized as one-dimensional concave functions. In particular, they are differentiable on \( \ell \) except at most for countably many points on \( \ell \). Since \( w \) defines parallel supporting hyperplanes for \( \log \Phi_1 \) and \( \log \Phi_2 \), the derivatives of \( \log \Phi_1|_{\ell} \) and \( \log \Phi_2|_{\ell} \) agree almost everywhere on \( \ell \). By the fundamental theorem of calculus for concave functions \cite[Corollary 24.2.1] {Roc77}, we have
\[
\log \Phi_1(q) - \log \Phi_1(p) = \log \Phi_2(q) - \log \Phi_2(p).
\]
Since any two points of an open connected \( U \) can be joined by a polygonal path contained in \( U \), there exists a constant \( c \) such that
\[
\log \Phi_2(p) = \log \Phi_1(p) + c
\]
for all \( p \in U \). This proves claim (ii).

The idea behind the proof of claim (iii) is to repeat the proof of part (i) with the additional restriction that the jumps have sizes less than \( \delta \).

Define a function \( \Phi \) on \( U \) by \( \Phi \) where the infimum is taken over all \( m \geq 0 \) and all choices of points \( \{p(1), \ldots, p(m)\} \) where \( p(i) \in U \) and \( \|p(i+1) - p(i)\| < \delta \). Since \( U \) is open and connected, there are always paths in \( U \) joining \( p(0) \) and \( p \) with jump size less than \( \delta \). As before, \( \Phi \) is a non-negative concave function on \( U \), with \( \Phi(p(0)) = 1 \), that can be extended to a positive concave function on \( \Delta^n \).

Now, we want to prove that \( \frac{\pi(p)}{p} \) is contained in \( \frac{1}{\Phi(p)} \partial \Phi(p) \), for each \( p \in U \). The weaker claim that \( \frac{\pi(p)}{p} \geq \frac{1}{\Phi(p)} \partial \Phi(p) \) holds for any two points \( p, q \in U \), with \( \|p - q\| < \delta \), and for any \( \alpha > \Phi(p) \) holds as in the proof of claim (i). Taking \( \alpha \downarrow \Phi(p) \) shows that
\[
1 + \left\langle \frac{\pi(p)}{p}, q - p \right\rangle \geq \frac{\Phi(q)}{\Phi(p)}, \quad \text{for any } q \text{ such that } \|q - p\| < \delta.
\]
Thus $\Phi(p)\pi(p)/p$ belongs to the superdifferential of the restricted concave function $\Phi|_V$ at $p$, where $V$ is a convex neighborhood of $p$ in $\Delta^n$.

However, by [Roc97] Theorem 23.2,
\[
\partial \Phi(p) = \{ \xi \in \mathbb{R}^n : D_v \Phi(p) \leq \langle \xi, v \rangle, \text{ for all tangent vector } v \}.
\]

Since the one-sided derivatives of $\Phi$ depends only on the values of $\Phi$ in a neighborhood of $p$, we observe that $\partial \Phi(p) = \partial (\Phi|_V)(p)$ where $V$ is any convex neighborhood of $p$. It follows from (11) that $\pi(p)/p \in \frac{1}{\pi(p)} \partial \Phi(p)$ for all $p \in U$, and so $\pi$ satisfies MCM on $U$. This completes the proof of the lemma. □

Now we give a general definition of functionally generated portfolios which extends the one in [Fer02, Theorem 3.1.5].

**Definition 6** (Functionally generated portfolios). A portfolio $\pi$ on $U$ is said to be generated by a concave function $\Phi : \Delta^n \to (0, \infty)$ if $\frac{\pi(p)}{p} \in \frac{1}{\pi(p)} \partial \Phi(p)$ for each $p \in U$.

Thus, a portfolio is generated by a concave function if and only if it satisfies the MCM property. The next lemma gives some basic properties of portfolios generated by concave functions.

**Lemma 3.** Let $\pi$ be a portfolio on $\Delta^n$ generated by a concave function $\Phi : \Delta^n \to (0, \infty)$.

(i) For any $\mu \in \Delta^n$ and $i = 1, \ldots, n$,
\[
1 + D_{e(i)-\mu} \log \Phi(\mu) \leq \frac{\pi_i}{\mu_i} \leq 1 - D_{\mu-e(i)} \log \Phi(\mu).
\]

In particular, if $\Phi$ is differentiable, the portfolio is given by the formula
\[
(12) \quad \pi_i = \mu_i \left(1 + D_{e(i)-\mu} \log \Phi(\mu)\right), \quad i = 1, \ldots, n.
\]

(ii) If $\pi$ is continuous, then $\Phi$ is automatically continuously differentiable. More generally, if $\pi$ is of class $C^k$, then $\Phi$ is of class $C^{k+1}$.

(iii) Conversely, if $\Phi : \Delta^n \to (0, \infty)$ is concave and differentiable, and we define $\pi$ by (12), then $\pi$ is generated by $\Phi$. In particular, $\pi(\mu) \in \Delta^n$ for all $\mu$.

(iv) Moreover, whether $\Phi$ is differentiable or not, $\pi \in \Delta^n$ when $\mu$ belongs to the set
\[
U = \{ \mu \in \Delta^n : 1 + D_{e(i)-\mu} \log \Phi(\mu) > 0, \text{ for all } i = 1, \ldots, n \}.
\]

**Proof.** (i) By definition of $\pi$, for $h \in \mathbb{R}$ small enough such that $\mu + h(e(i) - \mu) \in \Delta^n$, the superdifferential inequality gives
\[
(13) \quad 1 + \left\langle \frac{\pi(\mu)}{\mu}, h(e(i) - \mu) \right\rangle \geq \frac{\Phi(\mu + h(e(i) - \mu))}{\Phi(\mu)}.
\]

Note that the inner product is given by
\[
\left\langle \frac{\pi(\mu)}{\mu}, h(e(i) - \mu) \right\rangle = h \left(\frac{\pi_i}{\mu_i} - 1\right).
\]

Taking log on both sides of (13), for $h$ sufficiently small, we have
\[
h \left(\frac{\pi_i}{\mu_i} - 1\right) \geq \log \Phi(\mu + h(e(i) - \mu)) - \log \Phi(\mu).
\]
Dividing by $h$ and taking the limits as $h \downarrow 0$ and $h \uparrow 0$, we obtain the desired inequalities. Formula (12) is proved by noting that if $\Phi$ (and hence $\log \Phi$) is differentiable, for every tangent vector $v$ we have $D_v \log \Phi = -D_{-v} \log \Phi$.

(ii) Suppose that $\pi$ is continuous. It follows that $p \mapsto \frac{\pi(p)}{p}$ is a continuous selection of the superdifferential of $\log \Phi$. By [Rai88, Proposition 4], we see that $\log \Phi$, and hence $\Phi$, is differentiable on $\Delta^n$. By [Roc97, Corollary 25.5.1], $\Phi$ is actually continuously differentiable.

If $\pi$ is of class $C^k$ where $k \geq 1$, we already know $\Phi$ is differentiable. In terms of the coordinate system $\varphi(\mu) = (\mu_1, ..., \mu_{n-1})$, for $i = 1, ..., n-1$ we have
\[
\frac{\partial}{\partial \mu_i} (\log \Phi \circ \varphi^{-1})(\mu_1, ..., \mu_{n-1}) = D_{e(i)-e(n)} \log \Phi(\mu) = \frac{\pi_i}{\mu_i} - \frac{\pi_n}{\mu_n}
\]
which is of class $C^k$. Hence $\log \Phi$, and hence $\Phi$, is of class $C^{k+1}$.

(iii) Suppose $\Phi$ is differentiable and define $\pi$ by (12). To show $\pi_i \geq 0$, fix $\mu$ and $i$, and consider the restriction of $\Phi$ to the segment $(\mu, e(i))$. The inequality $\pi_i \geq 0$ is equivalent to $D_{e(i)-\mu} \Phi(\mu) \geq -\Phi(\mu)$, and follows from concavity and positivity of $\Phi$. To verify our claim that $\pi \in \overline{\Delta^n}$ we only need to show that $\sum_{i=1}^n \pi_i = 1$, or
\[
\sum_{i=1}^n \mu_i D_{e(i)-\mu} \log \Phi(\mu) = 0.
\]
This follows from the identity $\sum_{i=1}^n \mu_i (e(i) - \mu) = 0$ and linearity of the directional derivative for differentiable functions.

Since $v = \sum_{i=1}^n v_i (e(i) - \mu)$, we must have
\[
D_v \log \Phi = \sum_{i=1}^n v_i D_{e(i)-\mu} \log \Phi = \sum_{i=1}^n v_i \left(1 + D_{e(i)-\mu} \log \Phi \right) = \left\langle v, \frac{\pi}{\mu} \right\rangle.
\]
Hence $\frac{\pi}{\mu} \in \partial \log \Phi(\mu)$ and $\pi$ is generated by $\Phi$. The last statement (iv) is immediate.

Our next lemma shows that under further conditions on $\Phi$, near the boundary of the simplex, $\pi$ will not put mass on coordinates where $\mu$ is vanishing. In particular, when the market approaches any of the corners $e(i)$, $\pi$ must also be solely concentrated on the $i$th stock.

**Lemma 4.** Let $\pi$ be a portfolio on $\Delta^n$ generated by a concave function $\Phi : \Delta^n \rightarrow (0, \infty)$.

(i) If $\log \Phi$ is Lipschitz on $\Delta^n$, then $\pi/\mu$ is bounded.

(ii) If $\pi$ has positive weights and $\log \Phi$ is bounded on $\Delta^n$, then $\mu/\pi$ is bounded.

**Proof.** To prove (i), note that since $\log \Phi$ is Lipschitz, there is a constant $C > 0$ such that $|D_v \log \Phi(\mu)| \leq C||v||$ for all $\mu$ and $v$. It follows from Lemma 3 that $\pi/\mu$ is bounded.

To prove (ii), we proceed by the method of contradiction. By [Roc97, Theorem 10.3], we know that $\log \Phi$ can be extended to a bounded concave function on $\overline{\Delta^n}$. Suppose that for some $\mu \in \Delta^n$ and some $1 \leq i \leq n$, we have
\[
\frac{\mu_i}{\pi_i} > M \gg 0.
\]
Then, by concavity of $\Phi$, we get

$$\log \left( \frac{\pi_i}{\mu_i} \right) = \log \left( 1 + \frac{\pi_i}{\mu} e(i) - \mu \right) \geq \log \Phi(\mu(i)) - \log \Phi(\mu).$$

Rearranging terms we get

(14) \( \log \Phi(\mu(i)) \leq \log \Phi(\mu) - \log M. \)

Now suppose there is a sequence \(\{ (\mu(k), \pi(k)) \}, k = 1, 2, \ldots\) such that \(\mu(k)/\pi(k)\) is unbounded. Then, there must exist some coordinate \(i\) such that

$$\lim_{k \to \infty} \frac{\mu_i(k)}{\pi_i(k)} = \infty.$$

The above observation, coupled with (14), contradicts the fact that \(\log \Phi\) is bounded. This completes the proof.

Note that expression (12) still defines a ‘portfolio’ \(\pi\) (which may have negative holdings) without assuming \(\Phi\) is concave. In this case we still say that \(\pi\) is generated by \(\Phi\). We close the current subsection by showing in the next lemma that our definition is consistent with Fernholz’s. It can be proved by direct computation. The second statement will be useful in Section 3.1.

**Lemma 5.** Let \(G\) be a twice continuously differentiable function on \(\Delta^n\) and suppose it is extended to a twice continuously differentiable function function on an open neighborhood of \(\Delta^n\) in \(\mathbb{R}^n\). Then for any \(i = 1, \ldots, n\),

$$D_{e(i)-\mu}G = D_iG - \sum_{j=1}^n \mu_j D_jG.$$

Moreover, for any tangent vectors \(u\) and \(v\) of \(\Delta^n\),

$$\text{Hess}(u, v) := \sum_{i,j=1}^n u_i v_j D_{ij}G = \sum_{i,j=1}^n u_i v_j D_{e(i)-\mu}D_{e(i)-\mu}G.$$

### 2.2. Energy functionals.

The most interesting financial properties of of a functionally generated portfolio are captured by the concavity of the generating function.

**Definition 7.** Suppose the portfolio \(\pi\) is generated by a concave function \(\Phi : U \to (0, \infty)\). The **discrete energy** of the pair \((\Phi, \pi)\) is defined by

(15) \( T(q | p) := \log \left( 1 + \frac{\pi(p)}{p} \left( q - p \right) \right) - \left( \log \Phi(q) - \log \Phi(p) \right), \quad p, q \in \Delta^n. \)

This resembles the definition of **Bergman divergence**, see [AC10]. It is clear from [7] that \(T(q | p) \geq 0\) and is strictly positive unless \(\Phi\) is affine on the straight line joining \(p\) and \(q\). When \(\Phi\) is strictly concave, \(T(q | p) = 0\) only if \(q = p\). In general, \(T\) is asymmetric and does not define a metric.

As an example, fix \(\pi \in \Delta^n\) and consider the function \(\Phi(p) = \prod_{i=1}^n p_i^{\pi_i}\). This is a concave function which generates the constant-weighted portfolio \(\pi\). We have

(16) \( T(q | p) = \log \left( 1 + \frac{\pi(p)}{p} \left( q - p \right) \right) - \sum_{i=1}^n \pi_i \log \left( \frac{q_i}{p_i} \right) =: \gamma(p, q | \pi). \)

We call \(\gamma(p, q | \pi)\) the **free energy**, see Definition 2.2 of [PW13].
The interest in $T$ lies in the following decomposition formula. It generalizes \cite[Theorem 3.1.5]{Fer02} which treats the case where time is continuous and $\Phi$ is twice continuously differentiable.

**Lemma 6.** Let $\pi$ be generated by a positive concave function $\Phi$. Let $T$ be the discrete energy of the pair $(\Phi, \pi)$, and $\{\mu(t)\}$ be any market weight process in discrete or continuous time. Then the relative value process $V(t)$ satisfies the decomposition

\begin{equation}
\log V(t) = \log \frac{\Phi(\mu(t))}{\Phi(\mu(0))} + A(t),
\end{equation}

where $A(t)$ is a non-decreasing process satisfying $A(0) = 0$. In discrete time $A$ is given by

\[ A(t) = \sum_{k=0}^{t-1} T(\mu(k+1) | \mu(k)). \]

Moreover, in both discrete and continuous time, $A$ is identically zero for all market weight processes if and only if $\Phi$ is affine.

**Proof.** We first consider the case of discrete time. By \cite{15} and \cite{4}, we have

\[ T(\mu(k+1) | \mu(k)) = \log \left( 1 + \left( \pi(\mu(k)) \mu(k) - \mu(k+1) - \mu(k) \right) \right) - \log \frac{\Phi(\mu(k+1))}{\Phi(\mu(k))} \]
\[ = \log \frac{V(k+1)}{V(k)} - \log \frac{\Phi(\mu(k+1))}{\Phi(\mu(k))}. \]

Summing over $k$, we have

\begin{equation}
\log V(t) = \log \frac{\Phi(\mu(t))}{\Phi(\mu(0))} + A(t),
\end{equation}

where $A(t)$ is the cumulative energy given by $A(0) = 0$ and

\[ A(t) = \sum_{k=0}^{t-1} T(\mu(k+1) | \mu(k)), \]

which is non-decreasing as $T \geq 0$. Since $T(\mu(k+1) | \mu(k)) > 0$ if and only if $\Phi$ is not affine on the line segment from $\mu(k)$ to $\mu(k+1)$, the proof is completed.

In continuous time, the existence of $A$ will be proved by using a generalization of Itô-Tanaka formula for multidimensional semimartingales as proved in Bouleau \cite{Bou84}. Since $\pi$ is Borel measurable, the function $p \mapsto \Phi(p) \pi(p)/p$ is a Borel measurable section of $\partial \Phi$, i.e., $\Phi(p) \pi(p)/p \in \partial \Phi(p)$ for all $p$. By \cite[Theorem 3]{Bou84}, we get

\[ d\Phi(\mu(t)) = \Phi(\mu(t)) \frac{\pi(\mu(t))}{\mu(t)} d\mu(t) - dC(t), \]

where $C$ is a non-decreasing process starting at zero.

Taking log on both sides and applying Itô’s formula, we get

\[ d\log \Phi(\mu(t)) = \frac{1}{\Phi(\mu(t))} d\Phi(\mu(t)) - \frac{1}{2\Phi(\mu(t))^2} d\langle \Phi(\mu(t)) \rangle \]
\[ = \frac{\pi(\mu(t))}{\mu(t)} d\mu(t) - \frac{1}{\Phi(\mu(t))} dC(t) - \frac{1}{2\Phi(\mu(t))^2} d\langle \Phi(\mu(t)) \rangle. \]
Thus,
\[
\frac{dV(t)}{V(t)} = \frac{\pi(\mu(t))}{\mu(t)} d\mu(t) = d\log \Phi(\mu(t)) + \frac{1}{\Phi(\mu(t))} dC(t) + \frac{1}{2\Phi(\mu(t))^2} d\langle \Phi(\mu(t)) \rangle.
\]

By Itô’s formula again
\[
d\log V(t) = \frac{dV(t)}{V(t)} V(t) - \frac{1}{2} V(t)^2 d\langle V(t) \rangle
\]
\[
= \frac{dV(t)}{V(t)} - \frac{1}{2} d\langle \log \Phi(\mu(t)) \rangle, \quad \text{by (19),}
\]
\[
= \frac{dV(t)}{V(t)} - \frac{1}{2\Phi(\mu(t))^2} d\langle \Phi(\mu(t)) \rangle
\]
\[
= d\log \Phi(\mu(t)) + \frac{1}{\Phi(\mu(t))} dC(t), \quad \text{again by (19).}
\]

Hence the decomposition (18) holds with \(dA(\cdot) = dC(\cdot)/\Phi(\mu(\cdot))\).

Now we argue that if \(A(t) \equiv 0\) for all continuous semimartingale \(\mu\), then \(\Phi\) is affine. While the abstract definition of \(A\) is hard to work with, it suffices to prove that \(\Phi\) is affine on any line segment in \(\Delta^n\), where the usual Itô-Tanaka formula applies. Let \(v\) be a tangent vector and consider the line \(\ell = \{p_0 + sv : s \in [-\varepsilon, \varepsilon]\}\) lying in \(\Delta^n\). Let \(\Psi\) be the one-dimensional concave function on \([-\varepsilon, \varepsilon]\) defined by
\[
\Psi(s) = \Phi(p_0 + sv).
\]

Let \(\mu\) be a Brownian motion on \(\ell\) started at \(p_0\) and killed at the endpoints of \(\ell\), i.e., \(\mu(0) = p_0\) and
\[
d\mu(t) = vdB(t),
\]
where \(B\) is a one-dimensional Brownian motion killed at \(\pm \varepsilon\). Note that since \(\Psi\) is differentiable for all but countably many points \(p\) on \(\ell\), on those points of differentiability we have
\[
\left\langle \nabla p, v \right\rangle = D_v \log \Phi(p) = \frac{\Psi'(s)}{\Psi(s)}.
\]

By Itô-Tanaka formula \cite[Theorem VI.1.5]{RY99}, we have
\[
\Phi(\mu(t)) - \Phi(\mu(0)) = \Psi(B(t)) - \Psi(B(0))
\]
\[
= \int_0^t \Psi'(B(u)) dB(u) + \frac{1}{2} \int_{(-\varepsilon, \varepsilon)} L^\alpha_t \Psi''(da),
\]
where \(L^\alpha_t\) is the local time of \(B\) at \(a\), \(\Psi''\) is the second derivative of \(\Psi\) as a non-negative Radon measure \cite[Page 545]{RY99}, and in the last equality we also used the fact that \(B\) spends Lebesgue negligible amount of time on the set where \(\Psi\) is not differentiable.
By Itô’s formula,
\[
\log \frac{\Phi(\mu(t))}{\Phi(\mu(0))} = \log \frac{\Psi(B(t))}{\Psi(B(0))}
\]
\[
= \int_0^t \frac{1}{\Psi(B(u))} d\Psi(B(u)) - \frac{1}{2} \int_0^t \frac{1}{\Psi(B(u))^2} d\langle \Psi(B(u)) \rangle
\]
\[
= \int_0^t \frac{\Psi'(B(u))}{\Psi(B(u))} dB(u) + \frac{1}{2} \int_{-\epsilon}^\epsilon \left( \int_0^t \frac{dL^n_a(u)}{\Psi(B(u))} \right) \Psi''(da) - \frac{1}{2} (\log \Phi(\mu(t))).
\]

By \(20\) and \(21\), we have
\[
\int_0^t \frac{\Psi'(B(u))}{\Psi(B(u))} dB(u) = \int_0^t (\log \Psi)'(B(u)) dB(u) = \int_0^t \frac{\pi(\mu(u))}{\mu(u)} d\mu(u) = \int_0^t \frac{dV(u)}{V(u)},
\]
and by Itô’s formula, we have
\[
\log V(t) = \int_0^t \frac{dV(u)}{V(u)} - \frac{1}{2} (\log \Phi(\mu(t))).
\]

Combining, we get
\[
\log \frac{\Phi(\mu(t))}{\Phi(\mu(0))} = \log V(t) + \frac{1}{2} \int_{-\epsilon}^\epsilon \left( \int_0^t \frac{dL^n_a(u)}{\Psi(B(u))} \right) \Psi''(da).
\]

Hence the cumulative energy is
\[
A(t) = -\frac{1}{2} \int_{-\epsilon}^\epsilon \left( \int_0^t \frac{dL^n_a(u)}{\Psi(B(u))} \right) \Psi''(da).
\]

Suppose \(A(t) \equiv 0\) almost surely. The integrand in parentheses in \(22\) is a random non-negative continuous function in \(a\) (see \[RY99, Theorem VI.1.7\]) which is positive on a neighborhood of any fixed point in \((-\epsilon, \epsilon)\) with positive probability. From \(22\) we see that \(\Psi'' = 0\). Hence \(\Psi\) is affine on \(\ell\). Since \(\ell\) is arbitrary, \(\Phi\) is affine on the entire simplex. \(\square\)

3. Functionally generated portfolios

**Theorem 7.** Let \(U\) be a measurable subset of \(\Delta^n\) and let \(\pi\) be a portfolio on \(U\), where \(\pi : U \to \Delta^n\) is Borel measurable.

(i) Suppose \(\pi\) satisfies the MCM property, the generating function \(\Phi\) is not affine and \(\log \Phi\) is bounded on \(U\). Then it is a pseudo-arbitrage opportunity on \(U\) in both discrete and continuous time.

(ii) Suppose \(U\) is open and connected and \(\pi\) fails the MCM property. Then, given any \(\delta > 0\), there is a deterministic sequence of market weights in \(U\) of successive jump size less than \(\delta\), such that the relative value of the portfolio goes to zero as time tends to infinity. Thus \(\pi\) cannot be a pseudo-arbitrage strategy.

(iii) In fact, if \(U\) is open and convex, and \(\pi\) is not MCM, then, for any \(\delta > 0\), one can find an \(p \in U\) such that the deterministic sequence mentioned in (ii) can be chosen to lie entirely within the Euclidean ball of radius \(\delta\) around \(p\).

In this sense the only relative arbitrage opportunities are functionally generated.

**Proof.** Part (i) follows immediately from Lemma 6 and the definition of pseudo-arbitrage opportunity. Here we note that since \(\Phi\) is positive and continuous, \(\log \Phi\) is bounded on \(U\) whenever \(U\) is precompact in \(\Delta^n\).
To prove (ii), suppose that multiplicative cyclical monotonicity breaks down for some sequence \( \{p(0), p(1), \ldots, p(m)\} \) in \( U \). That is
\[
\eta := \prod_{i=0}^{m} \left( 1 + \left\langle \frac{\pi(i)}{p(i)}, p(i+1) - p(i) \right\rangle \right) < 1,
\]
where \( \pi(i) = \pi(p(i)) \) and \( p(m+1) = p(0) \). Consider a market process \( \{\mu(t), t = 0, 1, 2, \ldots\} \) such that
\[
\mu(t) = p(i), \quad \text{if } t = i \mod (m+1).
\]
In other words, the market cycles through the sequence \( \{p(0), p(1), \ldots, p(m)\} \). By (4), the relative value process satisfies
\[
\frac{V(t+1)}{V(t)} = 1 + \left\langle \frac{\pi(t)}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle.
\]
By (23), we must have
\[
\frac{V(k(m+1))}{V(0)} < \eta^k,
\]
which tends to zero as \( k \) tends to infinity. Hence this portfolio assignment cannot be a pseudo-arbitrage. The fact that the above sequence can be taken to have jump sizes less than \( \delta \) follows from Lemma 2(iii).

It remains to prove claim (iii). Suppose the MCM property fails on \( U \), assumed to be open and convex. We will show that there is a point \( p \in U \) such that the MCM property fails inside the Euclidean ball of radius \( \delta \) around \( p \). Then we may repeat the proof of claim (ii) in this ball.

This will be achieved by a method of contradiction using the following claim.

Claim. Suppose there is a \( \delta > 0 \) such that, for any \( p \in U \), the MCM property holds over any choice of points selected within a ball of radius \( \delta \) around \( p \). Then the MCM property must hold over \( U \).

To prove the above claim recall the notions of line integrals and conservative vector fields. Let \( \gamma \) be a piecewise smooth curve in \( \Delta^n \) indexed by some closed interval, say \([0, 1]\). The curve will be called a loop if \( \gamma(0) = \gamma(1) \). The line integral of the vector field \( \pi/\mu \) over any \( \gamma \) will be denoted by
\[
I_\gamma(\pi/\mu) := \int_\gamma \frac{\pi(\mu)}{\mu} d\mu = \int_0^1 \left\langle \frac{\pi(\mu(t))}{\mu(t)}, \mu'(t) \right\rangle dt.
\]
Recall that line integrals do not depend on parametrization, except for the orientation. In fact, if the direction of the line is flipped, the sign of the integral reverses. We will restrict ourselves to piecewise linear curves and loops. By a slight abuse of notation, the line from any \( a \) to any \( b \) in \( \Delta^n \), irrespective of parametrization, will be denoted by \([a, b]\).

Choose any point \( p \in U \) and let \( B_\delta(p) \) denote all points \( q \in \Delta^n \) such that the Euclidean distance between \( p \) and \( q \) is less than \( \delta \). Consider any loop \( \gamma \) whose range is contained in \( B_\delta(p) \). Then, we claim that
\[
I_\gamma(\pi/\mu) = 0.
\]
In other words, the vector field \( \pi/\mu \) is locally conservative restricted to every \( B_\delta(p) \).

To see (24), we use the fact that \( \pi \) satisfies MCM over \( B_\delta(p) \). Therefore, by Lemma 2, there is a positive concave function \( \Phi \) on \( \Delta^n \) such that \( \pi/\mu \) is an element
in the superdifferential of the concave function $\log \Phi$ at $\mu$. Consider any line $\ell$ contained in $B_\delta(p)$. Then $\log \Phi$ restricted to this line is concave, and, hence, $(\pi/\mu, \mu')$ over this line is non-increasing ([Roc97, Theorem 24.1]). If $p_1$ and $p_2$ are the two end points on this line, it follows that $I_\ell(\pi/\mu) = \log \Phi(p_2) - \log \Phi(p_1)$ ([Roc97, Theorem 24.2]). Thus, if $\gamma$ is a piecewise linear curve in $B_\delta(p)$, (24) follows.

We now show that any locally bounded and conservative vector field over $U$ must be **globally** conservative. While this statement is well known, we note that unlike the usual setting, we only assume that $\pi/\mu$ is measurable and locally bounded, and the resulting potential $\log \Phi$ is not necessarily differentiable. Let $w(\mu) := \pi(\mu)/\mu$ be locally conservative in the sense of (24). Fix two points $p$ and $q$ in $U$ and consider two piecewise linear curves $\gamma_1$ and $\gamma_2$ both of which start at $p$ and end at $q$. We will be done once we show

\[
\int_{\gamma_1} w(\mu)d\mu = \int_{\gamma_2} w(\mu)d\mu.
\]

Since $\gamma_1$ is arbitrary, it suffices to show above for $\gamma_2(t) = (1 - t)p + tq$.

In fact, we can assume that $\gamma_1$ has exactly three corners $p, r, q$ and is a concatenation of linear pieces $[p, r]$ and $[r, q]$ (we call such curves triangular). This is because once we establish (25) for such triangular curves, we can inductively eliminate corners in any other $\gamma_1$ and establish (25) in general.

For the rest of the argument we assume that $\gamma_1$ is triangular and $\gamma_2$ is $[p, q]$. Assume both $\gamma_1$ and $\gamma_2$ are indexed by $[0, 1]$. Now, two cases arise:

**Case 1.** $\sup_{0 \leq t \leq 1} \| \gamma_1(t) - \gamma_2(t) \| < \delta/2$. In this case, choose points $u_0 = 0 < u_1 < u_2, \ldots$ in $[0, 1]$ such that their images on $\gamma_2$ 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 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 $\gamma_1$ and $-\gamma_2$. Therefore this integral is zero, proving (25).

Notice that, in this argument we have not made use of the special shape of $\gamma_1$.

**Case 2.** We now argue that any other case can be reduced to Case 1 above. So, consider a triangle with vertices $p, r, q$. Not being Case 1 implies that the distance of
Figure 2. Transforming $\gamma_1$ (in thick lines) to $\tilde{\gamma}_1$ (in regular lines).

$r$ from $[p,q]$ is at least $\delta/2$. Consider now the ball of radius $\delta/4$ around $r$. Consider the two points $r_1$ and $r_2$ where the lines $[p,r]$ and $[r,q]$ uniquely intersects this ball. Consider the plane containing the three points $p, r, q$. This trace of the ball $B_{\delta/4}(r)$ on this plane is a disk of radius $\delta/4$. Consider the two points $r_1$ and $r_2$ where the lines $[p, r]$ and $[r, q]$ uniquely intersects this ball.

Consider the plane containing the three points $p, r, q$. This trace of the ball $B_{\delta/4}(r)$ on this plane is a disk of radius $\delta/4$. Consider the arc of this circle lying on the same side as $[p, q]$ and let $r_3$ be the intersection of this arc and the line which drops orthogonally from $r$ to $[p, q]$. Without loss of generality, we assume a planar picture that contains all the relevant lines and points. This is displayed in Figure 2.

Now consider a loop consisting of the oriented lines $[r_1, r], [r, r_2], [r_2, r_3]$ and $[r_3, r_1]$. This quadrilateral loop lies entirely in $B_{\delta}(r)$, and hence the integral of $w$ over this loop is zero. Therefore, the integral of $w$ over $\gamma_1$ is equal to its integral over the concatenation of two triangular curves, $[p, r_1] \cup [r_1, r_3]$ and $[r_3, r_2] \cup [r_2, q]$.

Consider each triangular piece separately. First consider $[p, r_1] \cup [r_1, r_3]$. We will show that the integral of $w$ over this path is the same as the integral over $[p, r_3]$. This follows from Case 1 once we show that if we index both paths by $[0, 1]$, then the supremum distance between these two paths is smaller than $\delta/2$.

Since $[p, r_1] \cup [r_1, r_3]$ has exactly one corner, it is clear that the supremum distance is the distance between $r_1$ and the line $[p, r_3]$. However this distance is smaller than $\|r_3 - r_1\|$, which, in turn, is smaller than $\delta/2$. Therefore, by Case I, the integral of $w$ over the curve $[p, r_1] \cup [r_1, r_3]$ is identical to the integral of $w$ over the line $[p, r_3]$. Similarly, the integral of $w$ over $[r_3, r_2] \cup [r_2, q]$ is identical to that over $[r_3, q]$.

Hence, we have shown that the integral of $w$ over the triangular curve $\gamma_1 = [p, r] \cup [r, q]$ is identical to that over the triangular curve $\tilde{\gamma}_1 = [p, r_3] \cup [r_3, q]$. However, now the distance between $\gamma_1$ and $\gamma_2$ is reduced by $\delta/4$. Therefore, we can repeat this same procedure finitely many times until our curve is within $\delta/2$ distance of $\gamma_2$, at which points Case 1 holds, and we are done.

Now that we have shown that $w$ is globally conservative, we can unambiguously define a function $\Phi$ on $U$ by fixing some $p_0 \in U$ and defining

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

where the integral is over any piecewise linear curve from $p_0$ to $p$. Over any $B_\delta(p)$, the function $\Phi$ must coincide (up to a constant) with the concave function resulting from the local MCM property of the vector field $\pi$. Thus, $\Phi$ is locally concave on a convex $U$, and hence it is concave (see [Hör07 page 58]) and generates $\pi$. However, this shows that $\pi$ is MCM over $U$ and this completes the proof of our claim. □
Multiplicative cyclical monotonicity implies a local curvature condition. To see this, suppose $U$ is open, $\pi : U \to \Delta^n$ is differentiable and satisfies MCM. Let $p \in U$ and consider a tangent vector $v$. Let $p_0 = p$ and $p_1 = p_0 + \varepsilon v$ for a small enough $\varepsilon > 0$ such that $p_1 \in U$. Then, by the MCM property with $m = 1$ we get

$$\left(1 + \frac{\langle \pi(p), \varepsilon v \rangle}{p + \varepsilon v} \right) \left(1 - \frac{\langle \pi(p + \varepsilon v), \varepsilon v \rangle}{p + \varepsilon v} \right) - 1 \geq 0.$$  

Let $w(p) = \pi(p)/p$. Expanding the left hand side of the above expression, we get

$$-\varepsilon \langle w(p + \varepsilon v) - w(p), v \rangle - \varepsilon^2 \langle w(p), v \rangle \langle w(p + \varepsilon v), v \rangle \geq 0.$$  

Dividing by $\varepsilon^2$ and taking the limit as $\varepsilon$ tends to zero, we get

$$\langle v, D_v w(p) \rangle \leq -\langle w(p), v \rangle^2.$$  

Hence we obtain the following lemma.

**Lemma 8.** For a differentiable portfolio function $\pi = \pi(\mu)$ to be a pseudo-arbitrage on an open set $U \subset \Delta^n$, it is necessary that the weight ratio $w(\mu) = \frac{\pi(\mu)}{\mu}$ satisfies the inequality

$$\langle v, D_v w(\mu) \rangle \leq -\langle w(\mu), v \rangle^2$$  

for any $\mu \in U$ and any tangent vector $v$ of $\Delta^n$.

In the next subsection we will relate this with the cumulative energy of the relative value process as well as the Fisher information metric in information geometry.

### 3.1. Energy-information inequality

In this subsection time is continuous. Suppose for each $p \in \Delta^n$, $b(p)(\cdot, \cdot)$ is a bilinear or quadratic form of tangent vectors of $\Delta^n$ given by

$$b(p)(v, v) = \sum_{i,j=1}^{n} b_{ij}(p)v_iv_j,$$

where $b_{ij}(\cdot)$ are functions on $\Delta^n$. The $b$-variation of the market weight process $\mu$ is defined by

$$\int b(d\mu, d\mu) := \sum_{i,j=1}^{n} \int b_{ij}(\mu)d(\mu_i, \mu_j).$$  

More generally, the coefficients $b_{ij}$ can themselves be stochastic processes. If $b$ is positive definite, i.e., $b(v, v) \geq 0$, the $b$-variation is non-decreasing. See [EM89] for more details.

Let $\pi$ be the portfolio (which may have negative holdings) generated via (12) by a twice continuously differentiable function $\Phi : \Delta^n \to (0, \infty)$ which is not necessarily concave. Fernholz [Fer02, Theorem 3.1.5] showed that the relative value process satisfies the decomposition

$$d \log V(t) = d \log \Phi(\mu(t)) + d\Theta(t),$$  

where $\Theta$ is a finite variation process called the drift process. It is given by

$$\Theta(t) = \int_0^t \frac{-1}{2\Phi(\mu(s))} \text{Hess}\Phi(\mu(s))(d\mu(s), d\mu(s)), \quad s \geq 0,$$

where $\text{Hess}\Phi$ is the Hessian quadratic form of $\Phi$.  

The expression of the drift process (29) can be motivated by discrete approximation. Given $\Phi$, we can define the associated discrete energy $T(q \mid p)$ by (15), and the discrete time decomposition formula (28) still holds. Note that since $\Phi$ is not necessarily concave, $T(q \mid p)$ may be negative. By Taylor series approximation, one can show that when $q - p$ is small, then

$$T(q \mid p) \approx -\frac{1}{2\Phi(p)}\text{Hess}\Phi(p)(q - p, q - p).$$

(30)

It follows that $\Theta(t)$ is the continuous time limit of accumulated discrete energy. This viewpoint will be discussed in detail when we study trigger-based rebalancing in Section 5.1.

Let $U \subset \Delta^n$ be open. Suppose $\Phi$ is uniformly concave in the sense that for some $\varepsilon > 0$, we have

$$-\text{Hess}\Phi(\mu)(v, v) \geq \varepsilon \|v\|^2$$

for all $\mu \in U$ and all tangent vectors $v$. If $\Phi(\mu) \geq \alpha$ on $U$ and the market weight process $\mu$ stays within $U$, it is easy to show that

$$\Phi(t) \geq \frac{\varepsilon}{2\alpha} \sum_{i=1}^{n} \langle \mu_i \rangle(t).$$

In general, the drift process is not monotone and may be decreasing. Our next theorem provides characterizations of those which are non-decreasing for all market weight processes. To formulate it we need the Fisher information metric in information geometry [ANH07]. Here we include a factor of $1/2$.

**Definition 8 (Fisher information metric).** The Fisher information metric on $\Delta^n$ defines an inner product of tangent vectors for each $p \in \Delta^n$. If $u = (u_1, ..., u_n)$ and $v = (v_1, ..., v_n)$ are tangent vectors of $\Delta^n$, and $p \in \Delta^n$, the inner product is defined by

$$\langle\langle u, v \rangle\rangle_p = \frac{1}{2} \sum_{i,j=1}^{n} \frac{1}{p_i} u_i v_j.$$ 

Since we are not doing differential geometry in this paper we will not distinguish the base point of the tangent vectors.

**Definition 9 (Free energy).** Let $\pi$ be a portfolio. The free energy quadratic form $\Gamma_\pi$ (at $\mu \in \Delta^n$) is defined by

$$\Gamma_\pi(v, v) = \frac{1}{2} \sum_{i,j=1}^{n} \frac{\pi_i(\delta_{ij} - \pi_j)}{\mu_i \mu_j} v_i v_j,$$

where $v$ is any tangent vectors of $\Delta^n$.

It can be verified that for any fixed $\pi \in \Delta^n$, $\Gamma_\pi(q - p, q - p)$ is the second order approximation of the discrete free energy $\gamma(q, p|\pi)$. Moreover, when $\pi = \mu$, then $\Gamma_\mu(\cdot, \cdot) = \langle\langle \cdot, \cdot \rangle\rangle_\mu$. The free energy corresponds to the ‘excess growth rate’ in [Fer02, page 8].

**Theorem 9.** Let $\pi$ be the portfolio generated by a twice continuously differentiable function $\Phi : \Delta^n \rightarrow (0, \infty)$. The following statements are equivalent.

(i) The drift process is non-decreasing for all market weight processes.

(ii) The generating function $\Phi$ is concave.
(iii) The weight ratio $w(\mu) = \frac{Z}{\mu}$ satisfies the inequality in Lemma 8:
$$\langle v, D_v w(\mu) \rangle \leq -\langle w(\mu), v \rangle^2,$$
for any $\mu \in \Delta^n$ and any tangent vector $v$ of $\Delta^n$.

(iv) The following energy-information inequality is satisfied: for any $\mu \in \Delta^n$ and any tangent vector $v$ of $\Delta^n$,
$$\Gamma_\pi(v, v) - \langle \langle D_v \pi(\mu), v \rangle \rangle_\mu \geq 0.$$

Recall that $D_v w$ and $D_v \pi$ mean the push forwards of tangent vectors (images of the differential map). Intuitively, the inner product $\langle \langle D_v \pi(\mu), v \rangle \rangle_\mu$ measures how much the portfolio weights move in the direction of the increment of market weights. The inequality (32) says that for $\pi$ to be a pseudo-arbitrage, this cannot be more than the free energy of the portfolio. This gives the meaning of concavity at the portfolio map level.

**Proof.** Without loss of generality, we may assume that $\Phi$ is twice continuously differentiable on an open neighborhood of $\Delta^n$ in $\mathbb{R}^n$. We have already shown that (ii) implies (i); this is also proved in [Fer02, Proposition 3.4.2]. To see that (i) implies (ii), suppose there exists $p_0 \in \Delta^n$ and a tangent vector $v$ of $\Delta^n$ such that $\text{Hess} \Phi(p_0)(v, v) > 0$.

Now we construct a continuous semimartingale $\mu$ in $\Delta^n$ such that
$$d\mu = vdM,$$
where $M$ is a (real-valued) continuous semimartingale with strictly increasing quadratic variation, and $\mu$ takes values on a line segment in $\Delta^n$ containing $p_0$. Then the drift process satisfies
$$d\Theta = -\frac{1}{2\Phi(\mu)}\text{Hess} \Phi(\mu)(v, v)d\langle M \rangle.$$

By construction $\Theta$ is decreasing whenever $\mu$ is near $p_0$.

To show that (ii) and (iii) are equivalent, we express the drift process in terms of $w$. Here all functions and derivatives are evaluated at $\mu$. Since $w_i = \frac{Z}{\mu_i} = 1 + D_{e_i - \mu} \log \Phi$, we have
$$D_{e_j - \mu} w_i = D_{e_j - \mu} D_{e_i - \mu} \log \Phi, \quad i, j = 1, ..., n.$$

Fix a tangent vector $v$. Using both statements of Lemma 5
$$\langle v, D_v w \rangle = \sum_{i,j=1}^n v_i v_j D_j w_i = \sum_{i,j=1}^n v_i v_j D_{e_j - \mu} w_i$$
$$= \sum_{i,j=1}^n v_i v_j D_{e_j - \mu} D_{e_i - \mu} \log \Phi = \sum_{i,j=1}^n v_i v_j D_{ij} \log \Phi.$$
Using the identity \( D_{ij} \log \Phi = \frac{D_i \Phi}{\Phi} - D_j \log \Phi \), we get
\[
\langle v, D_v w \rangle = \frac{1}{\Phi} \sum_{i,j=1}^{n} v_i v_j D_{ij} \Phi - \left( \sum_{i=1}^{n} v_i D_i \log \Phi \right)^2
\]
\[
= \frac{1}{\Phi} \text{Hess}\Phi(v, v) - (D_v \log \Phi)^2
\]
\[
= \frac{1}{\Phi} \text{Hess}\Phi(v, v) - \langle v, w \rangle^2.
\]
Thus
\[
- \frac{1}{2\Phi} \text{Hess}\Phi(v, v) = - \frac{1}{2} \left( \langle v, D_v w \rangle + \langle w, v \rangle^2 \right).
\]
From (33), it is clear that (ii) and (iii) are equivalent.

Finally we show that (iii) and (iv) are equivalent. Using the product rule,
\[
\frac{1}{2} \langle v, D_v w \rangle = \frac{1}{2} \sum_{i,j=1}^{n} v_i v_j \frac{\partial}{\partial \mu_j} \frac{\pi_i}{\mu_i}
\]
\[
= \frac{1}{2} \sum_{i,j=1}^{n} v_i v_j \left( \frac{1}{\mu_i} \frac{\partial \pi_i}{\partial \mu_j} - \pi_i \delta_{ij} \frac{1}{\mu_j^2} \right)
\]
\[
= \langle \langle v, D_v \pi \rangle \rangle_{\mu} - \frac{1}{2} \sum_{i=1}^{n} \frac{\pi_i}{\mu_i^2} v_i^2.
\]
It follows that
\[
\frac{1}{2} \langle v, D_v w \rangle + \frac{1}{2} \langle w, v \rangle^2 = \langle \langle v, D_v \pi \rangle \rangle_{\mu} - \frac{1}{2} \sum_{i,j=1}^{n} \frac{\pi_i (\delta_{ij} - \pi_i)}{\mu_i \mu_j} v_i v_j
\]
\[
= \langle \langle v, D_v \pi \rangle \rangle_{\mu} - \Gamma_{\pi}(v, v).
\]
This completes the proof of the theorem. \( \square \)

As a corollary, we have the following alternative expression of Fernholz’s decomposition.

**Corollary 10.** Let \( \pi \) be the portfolio generated by \( \Phi \). Using the notation in (27), the drift process \( \Theta \) satisfies
\[
d\Theta = \Gamma_{\pi}(d\mu, d\mu) - \langle \langle d\mu, D_d \pi \rangle \rangle_{\mu}.
\]

**Example 1** (Diversity-weighted portfolio). For \( 0 < p < 1 \), the diversity-weighted portfolio [Fer02, Example 3.4.4] is defined by
\[
\pi_i = \frac{\mu_i^p}{\sum_{k=1}^{n} \mu_k^p}, \quad i = 1, ..., n.
\]
It is generated by \( (\sum_{k=1}^{n} \mu_k^p)^{1/p} \). It is easy to check that for any tangent vector \( v \),
\[
\langle \langle v, D_v \pi(\mu) \rangle \rangle_{\mu} = p \Gamma_{\pi}(v, v)
\]
is a constant proportion of free energy. This allows us to state quantitatively, for example, that the portfolio with \( p = 0.7 \) follows 70% of market movement.
Remark 1. In [PW13, Section 4], we showed that when $n = 2$, a functionally generated portfolio $\pi = (q(Y), 1 - q(Y))$, where $Y = \log \frac{\mu_1}{\mu_2}$ and $q : \mathbb{R} \to (0, 1)$, has a non-decreasing drift function if and only if

$$q'(y) \leq q(y)(1 - q(y)), \quad y \in \mathbb{R}. \quad (34)$$

Using the identity $d\langle Y \rangle = \frac{1}{\mu_1 \mu_2} d\langle \mu_1 \rangle$, which follows from Itô’s formula, we can show that

$$\Gamma_\pi(d\mu, d\mu) = \frac{1}{2} q(1 - q)d\langle Y \rangle$$

and

$$\langle d\mu, D_d\pi(\mu) \rangle = \frac{1}{2} q' d\langle Y \rangle.$$ 

We get

$$d\Theta = \Gamma_\pi(d\mu, d\mu) - \langle d\mu, D_d\pi(\mu) \rangle = \frac{1}{2} (q(1 - q) - q') d\langle Y \rangle.$$ 

Thus Theorem 9 generalizes (34) and provides a geometric interpretation.

Remark 2. By Theorem 9, a continuously differentiable portfolio $\pi$ satisfies MCM if and only if the vector field $w = \pi/\mu$ is conservative and satisfies the inequality (26) of Lemma 8. When $n = 2$, all continuously differentiable portfolios are functionally generated as shown in [PW13, Lemma 4.6], and thus (26) implies the MCM property. In general, a portfolio may satisfy (26) without being generated by a concave function. The following example is inspired by [Roc97, Section 24].

The construction works for any $n \geq 3$, but for concreteness we let $n = 3$. Consider the matrices

$$A = \begin{pmatrix} -1 & -1 & -1 \\ 0 & -1 & -1 \\ 0 & 0 & -1 \end{pmatrix}, \quad B = \frac{1}{2} (A' + A),$$

where $A'$ is the transpose of $A$. Let $\lambda > 0$ be a parameter to be chosen, and define $A_\lambda := \lambda A$, $B_\lambda := \lambda B$. Note that $A$ is non-symmetric and $B$ is (strictly) negative definite. Here we use matrix notation whenever convenient. The matrix $A_\lambda$ defines a portfolio $\pi$ via the weight ratio, given by

$$w(\mu) = \frac{\pi}{\mu} = A_\lambda \mu + \alpha_\lambda(\mu) 1$$

$$= \lambda \begin{pmatrix} -1 \\ -(\mu_2 + \mu_3) \\ -\mu_3 \end{pmatrix} + \alpha_\lambda(\mu) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

where $\alpha_\lambda(\cdot) : \Delta^n \to \mathbb{R}$ is some smooth function. Using the identity $\sum_i \pi_i = \sum_i \mu_i w_i = 1$, we have

$$\alpha_\lambda(\mu) = 1 + \lambda \left[ \mu_1 + \mu_2(\mu_2 + \mu_3) + \mu_3^2 \right] > 1. \quad (36)$$

The portfolio is then given by

$$\pi(\mu) = \mu \cdot A_\lambda \mu + \alpha(\mu) \mu.$$ 

When $\lambda < 1$, the entries of $A_\lambda \mu$ are greater than $-1$. It follows from (30) that $w > 0$, so the portfolio has positive weights.

If this portfolio is generated by a concave function, Lemma 3(ii) implies that the generating function is smooth. By [Fer02, Proposition 3.1.11], there exists a continuously differentiable function $F$ on a neighborhood of $\Delta^n$ in $\mathbb{R}^n$ such that
\[ \sum (w_i + F) d\mu_i \text{ is an exact differential 1-form. It follows that } D_j (w_i + F) = D_i (w_j + F) \text{ for any } i \text{ and } j. \]

Letting \( \widetilde{F} = \alpha + F \), we see that 
\[ D_j \widetilde{F} = D_i \widetilde{F} = D_k \widetilde{F}, \]
while \( \lambda + D_3 \widetilde{F} = D_2 \widetilde{F} \), which is clearly a contradiction. Thus the portfolio is not functionally generated.

It remains to check that (26) holds. From (35), we see that 
\[ \langle v, D_v w \rangle + \langle v, w \rangle^2 = v' B v + (v' A \mu)^2 \]
(37)

For each \( \mu \), \( v \mapsto (v' A \mu)^2 \) is a non-negative definite quadratic form in \( v \). By continuity, there exists a constant \( C \) such that 
\[ (v' A \mu)^2 \leq C \| v \|^2 \text{ for all } \mu \in \Delta^n. \]
Since \( B \) is negative definite, by choosing \( \lambda > 0 \) sufficiently small, we can make the sum non-positive definite. Hence (26) holds on \( \Delta^n \).

4. Arbitrages constructed by optimal transport

In applications it is unintuitive to generate a portfolio from a given concave function. Rather the reverse question is more natural: suppose we are given a class of portfolios and a class of possible market weights. How can we build a map from the set of market weights to the set of given portfolios that is a pseudo-arbitrage opportunity? We show now that optimal transport provides a natural way of building such maps.

Let \( \mathcal{P} \) be a probability measure on \( \Delta^n \), considered as a Polish space under the standard Euclidean metric. The other Polish space we need is \([0, \infty)^n \) with the standard Euclidean metric. As will become clear below, a vector \( f \in [0, \infty)^n \setminus \{0\} \), where 0 is the zero vector, should be thought of as the unnormalized weight ratios of a portfolio. Let \( \mathcal{Q} \) be a probability measure on \([0, \infty)^n \). These measures can be atomic or have a density.

We now define a cost function. This is a continuous function \( c : \Delta^n \times [0, \infty)^n \to [-\infty, \infty) \) defined by 
\[ c(p, f) = \log \langle f, p \rangle, \quad p \in \Delta^n, f \in [0, \infty)^n. \]
(38)

Let \( \Pi(\mathcal{P}, \mathcal{Q}) \) be the set of probability measures on \( \Delta^n \times [0, \infty)^n \) such that the marginal distributions under any \( \mathcal{R} \in \Pi(\mathcal{P}, \mathcal{Q}) \) are given respectively by \( \mathcal{P} \) and \( \mathcal{Q} \). In other words, it defines a coupling \( (p, f) \) such that \( (p, f) \sim \mathcal{R}, p \sim \mathcal{P} \) and \( f \sim \mathcal{Q} \). We use \( \mathcal{R} [\cdot] \) to denote the expectation under the probability \( \mathcal{R} \). Consider the Monge-Kantorovich optimal transport problem [Vil03 page 2] that minimizes the expected cost function:
\[ \inf_{\mathcal{R} \in \Pi(\mathcal{P}, \mathcal{Q})} \mathcal{R} [\log \langle f, p \rangle]. \]
(39)

If the infimum is finite, it will be called the value of the optimal transport problem.

We show in this section that if the value of (39) is finite, the support of an optimal \( \mathcal{R} \) generates a portfolio assignment satisfying the MCM property. If the generating function is not affine one obtains a pseudo-arbitrage strategy.

We first consider a discrete case that will illuminate the connection with the MCM property. Our analysis will follow the line of argument described in Section 2.3 in [Vil03]. The major differences will be highlighted.
Lemma 11. Consider vectors $p(1), p(2), \ldots, p(N) \in \Delta^n$ and $f(1), f(2), \ldots, f(N) \in [0, \infty)^n \setminus \{0\}$ which may not be all distinct. Let $\mathcal{P}$ be the empirical law $\frac{1}{N} \sum_{j=1}^{N} \delta_{p(j)}$ and let $\mathcal{Q}$ be $\frac{1}{N} \sum_{j=1}^{N} \delta_{f(j)}$. Also, let $\mathcal{R}$ be the law $\frac{1}{N} \sum_{j=1}^{N} \delta_{(p(j), f(j))}$. Then $\mathcal{R}$ is the solution of the Monge-Kantorovich problem (39) if and only if the following condition holds. For all $1 \leq m \leq N$ and all $1 \leq k_1, k_2, \ldots, k_m \leq N$ we have

$$\sum_{j=1}^{m} \log \langle f(k_j), p(k_j) \rangle \leq \sum_{j=1}^{m} \log \langle f(k_j), p(k_{j+1}) \rangle, \quad m + 1 \equiv 1. \quad (40)$$

Define a portfolio $\pi$ on $\{p(j)\}_{j=1}^{N}$ by

$$\pi_i(p(j)) = \frac{f_i(j) p_i(j)}{\langle f(j), p(j) \rangle}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq N. \quad (41)$$

This is well-defined since $f(j) \neq 0$ implies $\langle f(j), p(j) \rangle > 0$.

Moreover, (41) holds if and only if the following MCM property holds:

$$\prod_{j=1}^{m} \left(1 + \left\langle \frac{\pi(p(k_j))}{p(k_j)}, p(k_{j+1}) - p(k_j) \right\rangle \right) \geq 1, \quad m + 1 \equiv 1. \quad (42)$$

Proof. The proof of (40) follows exactly as in Exercise 2.21 in [V103] page 79, parts (i) and (ii). The idea is that, by Birkhoff’s theorem, any transference plan can be realized as a convex combination of permutations. Thus, it suffices to optimize over permutations. Since every permutation is a product of disjoint cycles, we can restrict ourselves to optimal cyclic transference.

We now show the equivalence between (40) and (42). Choose some $m$ and some $k_1, \ldots, k_m$. For simplicity of notation, relabel such that $k_1 = 1, k_2 = 2, \ldots, k_m = m$. Notice that by (5), we have

$$\left(1 + \left\langle \frac{\pi(p(j))}{p(j)}, p(j+1) - p(j) \right\rangle \right) = \sum_{i=1}^{n} \pi_i(p(j)) p_i(j+1) - p_i(j) = \frac{1}{\langle f(j), p(j) \rangle} \sum_{i=1}^{n} f_i(j) p_i(j+1) = \frac{\langle f(j), p(j+1) \rangle}{\langle f(j), p(j) \rangle}. \quad (43)$$

Therefore the MCM property is equivalent to

$$\prod_{j=1}^{m} \langle f(k_j), p(k_j) \rangle \leq \prod_{j=1}^{n} \langle f(k_j), p(k_{j+1}) \rangle. \quad (44)$$

The equivalence now follows by taking logarithm on both sides. \hfill \square

The following is our main result in this section.

Theorem 12. Let $\mathcal{P}$ and $\mathcal{Q}$ be as above. Assume that $\inf \log \langle f, p \rangle > -\infty$, where the infimum runs over $p$ over the support of $\mathcal{P}$ and $f$ over the support of $\mathcal{Q}$. Then an optimal transference plan $\mathcal{R}$ exists and achieves the infimum in the Monge-Kantorovich problem (39). Define a multivalued map $\pi$ from the support of $\mathcal{P}$ to $\Delta^{\mathcal{P}}$ by

$$p \mapsto \left\{ \frac{1}{\langle f, p \rangle} f \cdot p : \quad (p, f) \in \text{supp} \mathcal{R} \right\}. \quad (45)$$

Then $\pi$ satisfies multiplicative cyclical monotonicity. In particular, $\pi$ is generated by a positive concave function.
Proof. Let \(-M, M \geq 0\), be the infimum \( \log (f,p) \) that is assumed to be finite. Under such a condition the existence of an optimal plan \( R \) is standard. The rest of the proof follows from [GM96] Theorem 2.3, page 128.

To put our problem in their framework, consider the cost function (also called \( c \)) in [GM96] to be \( M + \log (f,p) \). This is then a nonnegative cost function on the product of the supports of \( P \) and \( Q \). Theorem 2.3 in [GM96] then shows that the support of any optimal transference plan must be \( c \)-cyclically monotone. The notion of \( c \)-cyclically monotone results in a statement similar to (40), except now we are free to choose any \( m \geq 1 \) elements \((f(k_j), p(k_j))\) from the support. That this is equivalent to the MCM property of the derived portfolio has already been demonstrated in the proof of Lemma 3.4. \( \square \)

Now we give some explicit examples whose optimality can be verified directly. For \( \sigma = (\sigma_1, \ldots, \sigma_n) \in S_n \), the set of permutations of \( n \) labels, and \( x \in \mathbb{R}^n \), let \( \sigma \cdot x \) denote the vector \((x_{\sigma_1}, \ldots, x_{\sigma_n})\). Let \( E_\sigma \) be the set defined by

\[
E_\sigma = \{ x \in \mathbb{R}^n : \ x_{\sigma_1} > x_{\sigma_2} > \ldots > x_{\sigma_n} \}.
\]

In other words, the rank of the \( \sigma_i \)-th coordinate of a point in \( E_\sigma \) is \( i \). Note that \( E_\sigma = \sigma \cdot E_1 \), by extending our notion.

**Example 2.** Let \( U \) be a precompact subset of \( \Delta^n \). Suppose that \( U \) is defined only by the ranked coordinates. That is, \( U \cap E_\sigma = \sigma \cdot (U \cap E_1) \). Let \( \mathcal{P} \) be a probability measure on \( U \) which is exchangeable. Assume that \( \mathcal{P} \) does not charge the boundaries of \( E_\sigma \)'s.

Let \( V = \{ e(i), 1 \leq i \leq n \} \) and let \( Q \) be the uniform distribution on this finite set. With every \( p \in U \cap E_1 \), couple \( f = e(n) \). Extend by symmetry: with every \( p \in U \cap E_\sigma \), couple \( f = e(\sigma_n) \). Thus, this portfolio invests all in the smallest stock.

Similar examples can be worked out for \( V = \{ e(i) + e(j), 1 \leq i < j \leq n \} \). In this case the optimal portfolio will invest according to the rescaled market weights in the lowest two stocks. If \( V \) is the set of sums of \( k \) many distinct \( e(i)'s \), the resulting portfolio invests according to the market in the lowest \( k \) stocks. This is an example of portfolio selected by rank. See [Fer02] Example 4.3.2.

**Example 3.** Let \( p, q \in \Delta^n \). It follows from Jensen’s inequality that

\[
- \sum_{i=1}^{n} p_i \log p_i \leq - \sum_{i=1}^{n} p_i \log q_i.
\]

Let \( \mathcal{P} \) be any probability measure supported on \( \Delta^n \), and let \( \mathcal{Q} \) be the law of \(- \log p\), where \( p \sim \mathcal{P} \). Then, it follows trivially from (44) that the optimal coupling between \( \mathcal{P} \) and \( \mathcal{Q} \) is given by \( (p, -\log p) \). This produces the portfolio given by

\[
\pi_i = \frac{-p_i \log p_i}{\sum_{j=1}^{n} -p_j \log p_j}.
\]

This is the entropy-weighted portfolio described in [Fer02] Example 3.12.

It is natural to ask whether all pseudo-arbitrages can be realized as solutions of the optimal transport problem (39). We show that this is the case under general conditions.

**Theorem 13.** Let \( \pi \) be a portfolio generated by a concave function \( \Phi : \Delta^n \to (0, \infty) \). Let \( \mathcal{P} \) be any probability measure on \( \Delta^n \), and let \( \mathcal{Q} \) be the pushforward of \( \mathcal{P} \)
under the map \( w(\mu) = \pi(\mu) / \mu \). If \( \inf \log \langle f, p \rangle > -\infty \) where \( p \) and \( f \) range over the supports of \( P \) and \( Q \) respectively, then the coupling \( (\mu, w(\mu)) \) where \( \mu \sim P \) solves the optimal transport problem \( 39 \).

Proof. This is an immediate consequence of the general Kantorovich duality stated in [Vil08, Theorem 5.10]. Restricting to the product of the supports of \( P \) and \( Q \), the cost function is real-valued, continuous and bounded below. Let \( R \) be the law of the coupling \( (\mu, w(\mu)) \); it is clearly an element of \( \Pi(P, Q) \). By [Vil08, Theorem 5.10(ii)], \( R \) is optimal if and only if it is \( c \)-cyclically monotone. We have seen that this is equivalent to the MCM property of \( \pi \). Since \( \pi \) is generated by a concave function, it is MCM and the proof is completed. □

4.1. Relative entropy as a cost function. There is, in fact, another optimal transport problem whose solutions produce pseudo-arbitrages. Recall the relative entropy in information theory:

\[
H(q | p) = \sum_{i=1}^{n} q_i \log \frac{q_i}{p_i}, \quad (q, p) \in \Delta^n \times \Delta^n.
\]

Consider the relative entropy as a cost function \( H(q | p) \) on \( \Delta^n \times \Delta^n \). Now let \( \Pi(Q, P) \) be the set of probability measures concentrated on \( \Delta^n \times \Delta^n \) such that the marginal distributions under any \( R \in \Pi(Q, P) \) are \( Q \) and \( P \) respectively. Consider the following (negative) Monge-Kantorovich optimal transport problem:

\[
(45) \sup_{R \in \Pi(Q, P)} R[H(q | p)].
\]

Here \( q \sim Q \) and \( p \sim P \). Notice that the value of \( 45 \) is infinite whenever there is a coordinate \( i \) and some sequence of \( \mu \)'s in the support of \( P \) such that along this sequence this coordinate approaches zero, while the corresponding coordinate is bounded away from zero on the support of \( Q \).

**Theorem 14.** Let \( P \) be supported on \( U \) and let \( Q \) be supported on \( V \). Suppose that

\[
\gamma := \sup_{\pi \in V, \mu \in U} H(\pi | \mu) < \infty.
\]

Then the value of the problem \( 45 \) is finite. Moreover,

(i) There exists a concave function \( \Phi : \Delta^n \to (0, \infty) \) such that any optimal coupling \( R \) is concentrated on the set

\[
\left\{ (\pi, \mu) \in U \times \Delta^n : \pi/\mu \in \frac{1}{\Phi(\mu)} \partial \Phi(\mu) \right\}.
\]

(ii) For any two points \( p, q \) in \( U \) we have \( \Phi(p) / \Phi(q) \leq \exp(\gamma) \). Therefore, \( \log \Phi \) is bounded on \( U \).

The proof of the above theorem depends on the following lemma.

**Lemma 15.** Let \( \varphi \) be a proper convex function on \( (0, \infty)^n \). Consider \(-\log \Delta^n \) as a subset of \( (0, \infty)^n \). Assume that, for every point \( p \in \Delta^n \), there is a \( \pi \in \Delta^n \) such that the following subdifferential inequality holds

\[
\varphi(-\log q) \geq \varphi(-\log p) + \langle \pi, \log p - \log q \rangle, \quad \text{for all } q \in \Delta^n.
\]

That is, \( \pi \) is a subdifferential of \( \varphi \) at \(-\log \phi \). Then \( \Phi(p) = \exp[-\varphi(-\log p)] \) is a concave function on \( \Delta^n \). Moreover, \( \pi/p \in \frac{1}{\Phi(p)} \partial \Phi(p) \) for all \( p \in \Delta^n \).
**Proof.** Consider any two points \( p, q \in \Delta^n \). By (47) we get,
\[
\phi(-\log p) - \phi(-\log q) \leq \langle \pi, \log q - \log p \rangle = \sum_{i=1}^{n} \pi_i \log \left( \frac{q_i}{p_i} \right)
\]
\[
\leq \log \left( \sum_{i=1}^{n} \pi_i \frac{q_i}{p_i} \right), \quad \text{by Jensen's inequality,}
\]
\[
= \log \left( 1 + \left( \frac{\pi}{p}, q - p \right) \right).
\]
Exponentiating both sides of the above inequality, we get
\[
\left( 1 + \left( \frac{\pi}{p}, q - p \right) \right) \geq \exp \left[ \phi(-\log p) - \phi(-\log q) \right] = \frac{\Phi(q)}{\Phi(p)}.
\]
This shows that \( \Phi \) has a supergradient at \( p \) given by \( \Phi(p) \pi/p \). Since \( \Delta^n \) is a convex domain, \( \Phi \) is concave by [BSS13, Theorem 3.2.6]. This completes the proof. \( \square \)

**Proof of Theorem 14.** It is clear that the value of the problem is finite.

For (i) we will use a known fundamental result on optimal transport for quadratic cost. See Theorem 2.12 in [Vil03]. To use this result we will reformulate problem (45). Let \( \tilde{\mathcal{P}} \) be the law of \( \zeta = -\log \mu \) when \( \mu \sim \mathcal{P} \). Then both \( \tilde{\mathcal{P}} \) and \( \mathcal{Q} \) are probability measures in \( \mathbb{R}^n \), and the optimization problem (45) becomes
\[
\sup_{R \in \Pi(Q, \tilde{\mathcal{P}})} R \left( \langle \pi, \zeta - H(\pi) \rangle \right),
\]
where \( H(\pi) \) is the Shannon entropy of \( \pi \). Since \( R[H(\pi)] = Q[H(\pi)] \), it can be dropped. By completing the squares we get that this problem is equivalent to
\[
\inf_{R \in \Pi(Q, \tilde{\mathcal{P}})} \mathcal{R} \left( \| \pi - \zeta \|^2 \right),
\]
which is the usual optimal transport problem for the quadratic cost.

By (i) there is at least one solution, say \( \tilde{\mathcal{R}} \). The Knott-Smith optimality criterion [Vil03, Theorem 2.12] states that there is a lower semi-continuous convex function \( \varphi \) on \( \mathbb{R}^n \) such that the support of the optimal coupling is contained in the graph of \( \partial \varphi \). That is, for \( \tilde{\mathcal{R}} \) almost surely \( (\pi, \zeta) \), we have \( \pi \in \partial \varphi(\zeta) \).

Consider the map \( \mu = \exp(-\zeta) \mapsto \pi \). Let \( \varsigma(\mu) = -\varphi(-\log \mu) = -\varphi(\zeta) \). We now use Lemma 15 to claim that \( \Phi = \exp(\varsigma(\mu)) \) is a concave function in \( \mu \) and that \( \Phi(\mu) \pi/\mu \in \partial \Phi(\mu) \).

For (ii), let \( (q, \pi) \) be in the support of \( \tilde{\mathcal{R}} \) as above. We use the convexity of \( \varphi \) and the fact that \( \pi \) is a supergradient at \( -\log q \) to claim that
\[
\varphi(-\log p) - \varphi(-\log q) \geq \langle \pi, -\log p + \log p \rangle = H(\pi | p) - H(\pi | q) \geq -\gamma.
\]
Multiplying the above inequalities by the negative sign and then exponentiating proves our claim. \( \square \)

We now give some examples constructed via the previous result. Recall the sets \( E_{\sigma}, \sigma \in \mathcal{S}_n \), from (43).
Example 4. Let $U$ and $\mathcal{P}$ be as in Example 2. Let $V = \{e_i, \ 1 \leq i \leq n\}$ and let $\mathcal{Q}$ be the uniform distribution on this finite set. It is then clear that the following coupling maximizes the cross-entropy. With every $\mu \in U \cap E_1$, couple $e_n$. Extend by symmetry. With every $\mu \in U \cap E_\sigma$, couple $e(\sigma_n)$. Hence, the optimal coupling is the portfolio that puts its entire holding on the minimum stock. This is the same portfolio we get in Example 2.

When we take $V = \{(e(i) + e(j))/2, \ i < j\}$ or the set of sums of $k$ many distinct $e(i)$’s, the optimal portfolio will invest equally among the lowest $k$ stocks. This is in contrast to Example 2 where the portfolio invests according to the market in the lowest $k$ stocks.

The problem with the optimal transport approach with relative entropy cost over the entire simplex is that the value can become infinite. For example, one can couple some $\mu$ with a $\pi$ that is singular with respect to $\mu$ and make the relative entropy infinite. In general, these couplings produce portfolios that are more risky (or, contrarian, as known in some literature) than those produced by Theorem 12.

5. Approximations by Trigger-based Rebalancing

Implementation of functionally generated portfolios can be tricky. Portfolios generated by a strictly concave generating function $\Phi$ requires trading at every time point. If time is taken to be continuous, this is plainly impossible. Even when time is discrete, transactions costs can be forbiddingly high. Therefore, a portfolio manager typically tries to approximate a target portfolio with a sequence of buy-and-hold portfolios.

One such approximation scheme is called trigger-based rebalancing. Suppose $\pi$ is the target portfolio. We start at a point in the market $\mu(0)$ with a portfolio $\pi(\mu(0))$. We buy and hold this portfolio. As the market weights change, our portfolio also changes in the absence of trading. We keep on holding until the first time $\tau_1$ when our portfolio weights differ significantly from the target weights. At this point we perform the necessary trades to bring our portfolio weights to $\pi(\mu(\tau_1))$. The procedure is then repeated.

For completeness, we include the definition of buy-and-hold portfolio.

Definition 10 (buy-and-hold portfolio). A portfolio $\pi : \Delta^n \rightarrow \Delta^n$ is a buy-and-hold portfolio if

$$\pi_i = \frac{c_i \mu_i}{\sum_{j=1}^{n} c_j \mu_j},$$

where $c_1, ..., c_n$ are non-negative constants which are not all zero. It is generated by any positive multiple of the linear function $\Phi(p) = \sum_{j=1}^{n} c_j \mu_j$.

Since the generating function is linear, the discrete energy of a buy-and-hold portfolio is identically zero.

5.1. Local approximation of a portfolio by a buy-and-hold portfolio. Let us now formalize the above procedure. Suppose $\pi$ is generated by a concave function $\Phi : \Delta^n \rightarrow (0, \infty)$. Let $T$ be the discrete energy of the pair $(\Phi, \pi)$. Let $p_0 \in \Delta^n$. The buy-and-hold portfolio started at the weights $\pi(p_0)$ is generated by the following
as a telescoping sum, we get
\[ \hat{\Phi}_{p_0}(p) = \Phi(p_0) \left( 1 + \left\langle \frac{\pi(p_0)}{p_0}, p - p_0 \right\rangle \right) = \Phi(p_0) \sum_{i=1}^{n} \frac{\pi_i(p_0)}{p_{0i}} p_i, \quad p \in \Delta^n. \]

Note that \( \hat{\Phi}_{p_0}(p_0) = \Phi(p_0) \) and, from (4), that
\[ (50) \quad \hat{\Phi}_{p_0}(p) \geq \Phi(p), \quad p \in \Delta^n. \]

In fact, \( \hat{\Phi}_{p_0}(p) \) defines a supporting hyperplane of the hypograph of \( \Phi \) at \( (p_0, \Phi(p_0)) \).

For simplicity of notation we work in discrete time; the case of continuous time is similar. Consider an arbitrary strictly increasing sequence of stopping times \( \{\tau_k\}_{k=0}^{\infty} \) with \( \tau_0 = 0 \). These are the times when the portfolio rebalances. Consider the market weights \( \mu(0) \) at time 0. Let \( \hat{\Phi}_0 \) be the function \( \hat{\Phi}_{\mu(0)}(\cdot) \). The trigger-based rebalancing procedure will follow the buy-and-hold portfolio generated by \( \hat{\Phi}_0 \) until \( \tau_1 \). At this point the portfolio weights jump to those generated by \( \hat{\Phi}_1 = \hat{\Phi}_{\mu(\tau_1)}(\cdot) \). The procedure continues by successively defining the generating functions \( \hat{\Phi}_i \). Let \( \hat{\pi} \) be the resulting sequence of portfolio weights; it is an adapted process with values in \( \Delta^n \).

We analyze the relative value process \( V \) for this procedure. Recall that \( V(0) = 1 \) and the identity (4) which obviously holds in this context:
\[ \log V(t+1) - \log V(t) = \log \left( 1 + \left\langle \frac{\hat{\pi}(t)}{\mu(t)}, \mu(t+1) - \mu(t) \right\rangle \right). \]

Two cases arise. Either \( \tau_i \leq t < \tau_{i+1} \), in which case
\[ \log V(t+1) - \log V(t) = \log \hat{\Phi}_i(\mu(t+1)) - \log \hat{\Phi}_i(\mu(t)). \]

There is no energy collected during this period since \( \hat{\Phi}_i \) is linear. Adding these up as a telescoping sum, we get
\[ \log V(\tau_{i+1}) - \log V(\tau_i) = \log \hat{\Phi}_i(\mu(\tau_{i+1})) - \log \hat{\Phi}_i(\mu(\tau_i)). \]

Fix a \( k \geq 1 \), and sum the above for \( 0 \leq i \leq k \) to get
\[ \log V(\tau_{k+1}) = \sum_{i=0}^{k} \left[ \log \hat{\Phi}_i(\mu(\tau_{i+1})) - \log \hat{\Phi}_i(\mu(\tau_i)) \right] \]
\[ = \log \hat{\Phi}_k(\mu(\tau_{k+1})) - \log \hat{\Phi}_0(\mu(\tau_0)) + \sum_{i=1}^{k} \left[ \log \hat{\Phi}_{i-1}(\mu(\tau_i)) - \log \hat{\Phi}_i(\mu(\tau_i)) \right]. \]

Note that \( \hat{\Phi}_i(\mu(\tau_i)) = \Phi(\mu(\tau_i)) \), and, thus, from (50),
\[ (51) \quad \log \hat{\Phi}_{i-1}(\mu(\tau_i)) - \log \hat{\Phi}_i(\mu(\tau_i)) = T(\mu(\tau_i) | \mu(\tau_{i-1})) \geq 0. \]

Hence
\[ (52) \quad \log V(\tau_{k+1}) = \log \hat{\Phi}_k(\mu(\tau_{k+1})) - \log \Phi(\mu(\tau_0)) + \hat{A}(\tau_{k+1}), \]
where
\[ \hat{A}(\tau_{k+1}) = \sum_{i=1}^{k} T(\mu(\tau_i) | \mu(\tau_{i-1})) \]
is an increasing nonnegative process starting at zero.
Extend the above process to all nonnegative discrete time points by defining $\hat{A}(t) = \hat{A}(\tau_k)$ for all $t \in [\tau_k, \tau_{k+1} - 1]$. Hence, for a general time point $t$ such that $t \in [\tau_k, \tau_{k+1} - 1]$, we have

\begin{equation}
\log V(t) = \log \hat{\Phi}_k(\mu(t)) - \log \Phi(\mu(0)) + \hat{A}(t).
\end{equation}

Thus, if $\Phi$ is strictly concave and the difference of the first two terms on the right side of the above equation is bounded below, the trigger-based rebalancing procedure, although not functionally generated, behaves as a pseudo-arbitrage strategy.

The boundedness condition can be easily seen to hold. Let $U$ be an open subset of the unit simplex. If $\log \Phi$ is bounded on $U$ then for all paths of $\mu$ contained in $U$,

$$
\log \hat{\Phi}_k(\mu(t)) - \log \Phi(\mu(0)) \geq \log \Phi(\mu(t)) - \log \Phi(\mu(0))
$$

is clearly bounded below.

The above argument does not change significantly in continuous time, except for the fact that the definitions of the stopping times are now in continuous time.

Notice the following conclusion. Consider the market weight process sampled at the stopping times: $\{\mu(\tau_k), k = 0, 1, 2, \ldots\}$. By a change of time, we define a new process $\nu(k) = \mu(\tau_k), k = 0, 1, 2, \ldots$. Then the relative value process of this new market weight process, when the portfolio is generated by $\Phi$, coincides with the trigger-based rebalancing procedure. In this sense portfolio analysis for discrete market weight process can be thought as trigger-based rebalancing applied to continuous market weight process.

5.2. **Statistical arbitrage in a hidden semimartingale model.** Trigger-based rebalancing need not be a very good approximation. For example, the continuously traded portfolio can potentially collect a lot of energy in between $\tau_i$ and $\tau_{i+1}$, while the buy-and-hold approximation will collect none. The fact that a piecewise buy-and-hold portfolio can outperform a continuously traded portfolio has been already exhibited in [PW13, Section 5]. See, for example, the case of bang-bang process with a large drift $\gamma > 1$. This brings us to the interesting question: what is the optimal frequency at which we should trade? The examples in [PW13] work because we make assumptions on where the process will be in the future. In the absence of such predictability it is not obvious which is better. Fernholz and Maguire, Jr., [FM07] analyze data in the case of high-frequency trading that maintains an equal-weighted portfolio at regular intervals. They observe that empirically the free energy collected by trading at longer intervals is smaller than that collected by trading in shorter intervals. This explains how the difference can be translated to a statistical arbitrage by building a long-short portfolio.

In this subsection we attempt to prove the observation in [FM07] under a hidden semimartingale model of market microstructure similar in spirit to ones described in [ZMAS05, JLM+09]. We refer the readers to the introductory discussions in the above papers for the references in the econometric literature regarding the history of these models and their validity. In short, this is a model where the underlying stock prices follow a semimartingale model which may or may not be observed. The true prices are observed when sampled at somewhat sparse intervals (typically once in a few minutes). When prices are sampled at high frequency (say several times per second) one observes the true price plus a noise of significant size. What is observed cannot be effectively approximated by a semimartingale.
We mention, in passing, that although the model suggested in [FM07, Appendix B] shows the presence of statistical arbitrage “in expectation”, it fails to produce a statement that holds with high probability. Straightforward calculations suggest that any version of high-probability statistical arbitrage is impossible in semimartingale models.

We model the ‘true’ stock prices as a continuous semimartingale:

\[ d \log X_i(t) = \gamma_i(t)dt + dM_i(t), \quad i = 1, 2, \ldots, n. \]

Here \( M_1, \ldots, M_n \) are continuous martingales with \( \langle M_i, M_j \rangle(t) = a_{ij}(t)dt \). The drift parameters \( \gamma(t) \) and the volatility matrix \( a(t) \) are taken to be arbitrary progressively measurable processes such that \( X = (X_1, \ldots, X_n) \) exists. We also write \( \log X = (\log X_1, \ldots, \log X_n) \).

Let \( \delta > 0 \) and partition the positive half-line into subintervals of size \( \delta \). Suppose on the even intervals we observe \( \log X((2k + 1)\delta) + \sqrt{\delta} \epsilon(k) \), where \( \epsilon(1), \epsilon(2), \ldots \) is a sequence of i.i.d. \( \mathbb{R}^n \)-valued random variables with mean zero and covariance matrix \( \Sigma \) which are jointly independent of the process \( X \).

Let \( \{Y(k) : k \geq 0\} \) be the observed log prices. For \( k \geq 0 \), we have

\[ Y(2k) = \log X(2k\delta), \]
\[ Y(2k + 1) = \log X((2k + 1)\delta) + \sqrt{\delta} \epsilon(k). \]

Let \( \{\bar{\mu}(k) : k \geq 0\} \) denote the market weight process corresponding to the observed prices \( Y \). We analyze the performance of two constant-weighted portfolios \( \pi = (\pi_1, \ldots, \pi_n) \), one rebalances at intervals of size \( \delta \) and the other at interval of size \( 2\delta \). The two portfolios differ in the accumulated free energies generated by the processes \( \{\bar{\mu}(k), k \geq 0\} \) and \( \{\bar{\mu}(2k), k \geq 0\} \), where the free energy is

\[ T(q | p) = \gamma(p, q | \pi) = \log \left( \sum_{i=1}^{n} \pi_i e^{\log(q_i/p_i)} \right) - \sum_{i=1}^{n} \pi_i \log(q_i/p_i), \quad p, q \in \Delta^n. \]

In a practical setting one can have several sample points of the finer sampling scheme between successive samples from the coarser scheme. Moreover, one can assume that all the fine samples will be observed with noise. It is not hard to see that the proof of the following theorem can be easily extended to that case.

**Theorem 16.** Suppose that \( E(\|\epsilon(1)\|^3) < \infty \). Let \( \gamma \) be the positive constant given by

\[ \gamma := \sum_{i,j=1}^{n} \pi_i (\delta_{ij} - \pi_j) \Sigma_{ij}. \]

Let \( t_0 > 0 \) be some fixed time. Let \( T_1 \) be the accumulated free energy from the finely process \( \bar{\mu}(\cdot) \) during the time interval \([0, T]\), and let \( T_2 \) be the accumulated free energy from the sparsely sampled process \( \bar{\mu}(2\cdot) \), i.e.,

\[ T_1 = \sum_{k: (k+1)\delta \leq t_0} T(\bar{\mu}(k + 1) | \bar{\mu}(k)), \]
\[ T_2 = \sum_{k: 2(k+1)\delta \leq t_0} T(\bar{\mu}(2(k + 1)) | \bar{\mu}(2k)). \]

Then \( T_1 - T_2 \) converges to \( \gamma t_0/2 \) in probability as \( \delta \) tends to zero.
Proof. We start with a Taylor approximation of the discrete energy \( T(q \mid p) \). Write \( \xi = \log (q/p) \in \mathbb{R} \). Then we have the approximation

\[
T(q \mid p) = \frac{1}{2} \sum_{i,j=1}^{n} \pi_i (\delta_{ij} - \pi_j) \xi_i \xi_j + R(\xi),
\]

where \( R(\xi) \) is the remainder. This is equivalent to the free energy quadratic form in (31) except for the logarithmic change of variables. We also note that, by the numéraire invariance property of free energy proved in [PW13, Lemma 2.3], if there are positive vectors \( x \) and \( y \) such that

\[
p = \sum_{i=1}^{n} x_i, \quad q = \sum_{i=1}^{n} y_i,
\]

then we have

\[
T(q \mid p) = \log \left( \sum_{i=1}^{n} \pi_i e^{\log(y_i/x_i)} \right) - \sum_{i=1}^{n} \pi_i \log(y_i/x_i).
\]

Thus we may also take \( \xi = y/x \).

Without loss of generality, we can take \( t_0 = 1 \). There are \( N = \lfloor 1/\delta \rfloor \) time points in \([0, 1]\). Define the sequence of increments

\[
\nu(k) = \log \left( X((k+1)\delta) - X(k\delta) \right).
\]

The accumulated free energy for the sparsely sampled process is

\[
T_2 = \sum_{k=0}^{\lfloor N/2 \rfloor - 1} T(\mu(2k + 2) \mid \mu(2k))
\]

\[
= \frac{1}{2} \sum_{k=0}^{\lfloor N/2 \rfloor - 1} \sum_{i,j=1}^{n} \pi_i (\delta_{ij} - \pi_j) \left( \nu_i(2k) + \nu_i(2k + 1) \right) \left( \nu_j(2k) + \nu_j(2k + 1) \right)
\]

\[
+ \sum_{k=0}^{\lfloor N/2 \rfloor - 1} R(\nu(2k) + \nu(2k + 1)).
\]

On the other hand, the accumulated free-energy for the finely sampled process is given by

\[
T_1 = \frac{1}{2} \sum_{k=0}^{\lfloor N/2 \rfloor - 1} \sum_{i,j=1}^{n} \pi_i (\delta_{ij} - \pi_j) \left[ \left( \nu_i(2k) + \sqrt{\delta} \epsilon_i(k) \right) \left( \nu_j(2k) + \sqrt{\delta} \epsilon_j(k) \right) \right]
\]

\[
+ \frac{1}{2} \sum_{k=0}^{\lfloor N/2 \rfloor - 1} \sum_{i,j=1}^{n} \pi_i (\delta_{ij} - \pi_j) \left[ \left( \nu_i(2k + 1) - \sqrt{\delta} \epsilon_i(k) \right) \left( \nu_j(2k + 1) - \sqrt{\delta} \epsilon_j(k) \right) \right]
\]

\[
+ \sum_{k=0}^{\lfloor N/2 \rfloor - 1} \left[ R(\nu(2k) + \sqrt{\delta} \epsilon(k)) + R(\nu(2k + 1) - \sqrt{\delta} \epsilon(k)) \right].
\]
Consider the first two quadratic terms on the right side of (56). Their sum can be additionally decomposed as the sum of four terms:

\[ H_1 = \frac{1}{2} \sum_{k=0}^{N} \sum_{i,j=1}^{n} \pi_i (\delta_{ij} - \pi_j) \nu_i(k) \nu_j(k), \]

\[ H_2 = \delta \sum_{k=0}^{[N/2]-1} \sum_{i,j=1}^{n} \pi_i (\delta_{ij} - \pi_j) \epsilon_i(k) \epsilon_j(k), \]

\[ H_3 = \sqrt{\delta} \sum_{k=0}^{[N/2]-1} \sum_{i,j=1}^{n} \pi_i (\delta_{ij} - \pi_j) \nu_i(2k) \epsilon_j(k), \]

\[ H_4 = -\sqrt{\delta} \sum_{k=0}^{[N/2]-1} \sum_{i,j=1}^{n} \pi_i (\delta_{ij} - \pi_j) \nu_i(2k+1) \epsilon_j(k). \]

Therefore, we can write

(57)

\[ T_1 = H_1 + H_2 + H_3 + H_4 + \sum_{k=0}^{[N/2]-1} \left[ R(\nu(2k) + \sqrt{\delta} \epsilon(k)) + R(\nu(2k+1) - \sqrt{\delta} \epsilon(k)) \right]. \]

A similar decomposition for \( T_2 \) will be

(58)

\[ T_2 = \overline{H}_1 + \sum_{k=0}^{[N/2]-1} R(\nu(2k) + \nu(2k+1)), \]

where \( \overline{H}_1 \) is the sole quadratic term in the Taylor expansion.

Since \( X \) is a continuous semimartingale there is a set \( \Omega_1 \) of paths of \( X \) of probability one such that for every \( \omega \in \Omega_1 \) we have

\[ \lim_{\delta \downarrow 0} \overline{H}_1 = \lim_{\delta \downarrow 0} H_1 = \int_{0}^{1} \Gamma_{\pi}(d\mu, d\mu), \]

\[ \lim_{\delta \downarrow 0} \sum_{k=0}^{[N/2]-1} R(\nu(2k) + \nu(2k+1)) = 0, \]

\[ \lim_{\delta \downarrow 0} \sum_{k=0}^{[N/2]-1} [R(\nu(2k)) + R(\nu(2k+1))] = 0, \]

where \( \mu(\cdot) \) is the continuous time market weight process corresponding to the ‘true’ stock price \( X \), and \( \int_{0}^{1} \Gamma_{\pi}(d\mu, d\mu) \) is the accumulated free energy for \( \mu \).

Now, consider \( H_2 \). It follows immediately that \( E(H_2) = \gamma \delta \lfloor N/2 \rfloor \), and its variance is \( O(\delta) \). Thus \( H_2 \) converges in probability to \( \gamma/2 \) in probability as \( \delta \) tends to zero. Consider \( H_3 \) and \( H_4 \). They both have zero mean and variance \( O(\delta) \), where the constant in the big-O is a quadratic function of the increments of \( X \). Thus, both \( H_3 \) and \( H_4 \) converge to zero in probability.
Finally, consider
\[ R_1 := \sum_{k=0}^{\lfloor N/2 \rfloor - 1} \left[ R(\nu(2k)) + \sqrt{\delta \epsilon(k)} \right] + R(\nu(2k + 1) - \sqrt{\delta \epsilon(k)}) \],
\[ R_2 := \sum_{k=0}^{\lfloor N/2 \rfloor - 1} \left[ R(\nu(2k)) + R(\nu(2k + 1)) \right]. \]

We claim that the difference \( R_1 - R_2 \) goes to zero in probability. Then, appealing to the last limit in (59) we will be done.

Consider again the expression for \( R \) from (30):
\[ R(\xi) = \log \left( \sum_{i=1}^{n} \pi_i e^{\xi_i} \right) - \langle \pi, \xi \rangle - \frac{1}{2} \sum_{i,j=1}^{n} \pi_i (\delta_{ij} - \pi_j) \xi_i \xi_j. \]

We now compute the first two derivatives of \( R \).
\[ D_i R = \frac{\pi_i e^{\xi_i}}{\sum_{j=1}^{n} \pi_j e^{\xi_j}} - \pi_i - \pi_i \xi_i + \sum_{j=1}^{n} \pi_j \xi_j, \]
\[ D_{ij} R = -\frac{\pi_i \pi_j e^{\xi_i + \xi_j}}{(\sum_{k=1}^{n} \pi_k e^{\xi_k})^2} + \pi_i \pi_j, \quad i \neq j, \]
\[ D_{ii} R = \frac{\pi_i e^{\xi_i}}{\sum_{k=1}^{n} \pi_k e^{\xi_k}} - \frac{\pi_i^2 e^{2\xi_i}}{(\sum_{k=1}^{n} \pi_k e^{\xi_k})^2}. \]

It is obvious that the Hessian matrix of \( R \) is uniformly bounded. In fact, similar computation shows that every element of the third derivative of \( R \) is uniformly bounded in absolute value.

Hence, there is a constant \( C_1 > 0 \) such that \( C_1 \delta^{3/2} \| \epsilon(k) \|^3 \) is an upper bound on the difference
\[ \left| R(\nu(2k)) + \sqrt{\delta \epsilon(k)} \right| - \left| R(\nu(2k)) \right| - \left| \langle \epsilon, \nabla R(\nu(2k)) \rangle \right| - \frac{\delta}{2} \| \epsilon(k) \| \left\langle \nabla^2 R(\nu(2k)) \right\rangle \epsilon(k). \]

In particular, by our moment assumption on \( \epsilon(\cdot) \)'s, the expectation of the above, summed over \( k \), converges to zero as \( \delta \) tends to zero. In particular, it converges to zero in probability.

However, repeating arguments as before, we see that
\[ \sum_{k=0}^{\lfloor N/2 \rfloor} \left[ \sqrt{\delta} \langle \epsilon(k), \nabla R(\nu(2k)) \rangle + \frac{\delta}{2} \langle \epsilon(k), \nabla^2 R(\nu(2k)) \rangle \epsilon(k) \right] \]
also converges to zero in probability. Therefore,
\[ \sum_{k=0}^{\lfloor N/2 \rfloor - 1} R(\nu(2k)) + \sqrt{\delta \epsilon(k)} - R(\nu(2k)) \]
converges to zero in probability as well.

Repeating the above argument for the odd numbered samples, we have shown \( R_1 - R_2 \) (and hence \( R_1 \)) converges to zero in probability.

Combining all the previous steps we have shown that \( T_1 \) in (56) converges in probability to \( \gamma/2 + \int_{0}^{1} \Gamma_\pi(d\mu, d\mu) \) and \( T_2 \) converges in probability to \( \int_{0}^{1} \Gamma_\pi(d\mu, d\mu) \).
Therefore, $T_1 - T_2$ converges to $\gamma / 2$ is probability and we have proved the intended statement for $t_0 = 1$. The case of a general $t_0$ follows by scaling. □

References

[AC10] S.-I. Amari and A. Cichocki, Information geometry of divergence functions, Bulletin of the Polish Academy of Sciences: Technical Sciences 58 (2010), no. 1, 183–195.

[ANH07] S. Amari, H. Nagaoka, and D. Harada, Methods of information geometry, Translations of mathematical monographs, American Mathematical Society, 2007.

[BF08] Adrian D. Banner and Daniel Fernholz, Short-term relative arbitrage in volatility-stabilized markets, Annals of Finance 4 (2008), no. 4, 445–454.

[BHLPL3] Mathias Beiglböck, Pierre Henry-Labordère, and Friedrich Penkner, Model-independent bounds for option prices: a mass transport approach, Finance and Stochastics (2013), 1–25.

[Bou84] N. Bouleau, Formules de changement de variables, Annales de l'I.H.P., section B 20 (1984), no. 2, 133–145.

[BSS13] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, Nonlinear programming: Theory and algorithms, Wiley, 2013.

[CHKL11] Tzee-man Chow, Jason Hsu, Vitali Kalesnik, and Bryce Little, A survey of alternative equity index strategies, Financial Analysts Journal 67 (2011), no. 5, 37–57.

[Eco13] The Economist, Fund management: The rise of smart beta, Print edition July 6 2013, Available at http://www.economist.com/news/finance-and-economics/21580518-terrible-name-interesting-trend-rise-smart-

[EM89] Michel Emery and P. A. Meyer, Stochastic calculus in manifolds, Berlin, 1989.

[Fer99] Robert Fernholz, Portfolio generating functions, Quantitative Analysis in Financial Markets, River Edge, NJ. World Scientific (1999).

[Fer02] E. R. Fernholz, Stochastic portfolio theory, Applications of Mathematics, Springer, 2002.

[FGH98] Robert Fernholz, Robert Garvy, and John Hannon, Diversity-weighted indexing, The Journal of Portfolio Management 24 (1998), no. 2, 74–82.

[FK05] E. R. Fernholz and I. Karatzas, Relative arbitrage in volatility-stabilized markets, Annals of Finance 1 (2005), no. 2, 149–177.

[FK09] E. R. Fernholz and I. Karatzas, Stochastic portfolio theory: an overview, Handbook of Numerical Analysis (P. G. Ciarlet, ed.), Handbook of Numerical Analysis, vol. 15, Elsevier, 2009, pp. 89 – 167.

[FK11] Daniel Fernholz and Ioannis Karatzas, Optimal arbitrage under model uncertainty, The Annals of Applied Probability 21 (2011), no. 6, 2191–2225.

[FKK05] E. R. Fernholz, I. Karatzas, and C. Kardaras, Diversity and relative arbitrage in equity markets, Finance and Stochastics 9 (2005), no. 1, 1–27.

[FM07] R. Fernholz and Carey Maguire, Jr., The statistics of statistical arbitrage, Financial Analysts Journal 63 (2007), no. 5, 46–52.

[GM96] Wilfrid Gangbo and Robert J McCann, The geometry of optimal transportation, Acta Math. 177 (1996), 113–161.

[Hör07] L. Hörmander, Notions of convexity, Modern Birkhäuser classics, Springer London, Limited, 2007.

[JLM+09] J. Jacob, Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter, Microstructure noise in the continuous case: The pre-averaging approach, Stochastic processes and their applications 119 (2009), 2249–2276.

[PW13] S. Pal and T.-K. L. Wong, Energy, entropy, and arbitrage, ArXiv e-prints (2013), no. 1308.5376.

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

[Roc97] R. T. Rockafellar, Convex analysis, Convex Analysis, Princeton University Press, 1997.

[RY99] D. Revuz and M. Yor, Continuous martingales and brownian motion, Grundlehren der Mathematischen Wissenschaften, Springer, 1999.

[Str12] Winslow Strong, Generalizations of functionally generated portfolios with applications to statistical arbitrage, Arxiv e-prints (2012), no. 1212.1877.
[Vil03] Cédric Villani, Topics in optimal transportation, Graduate studies in mathematics, American Mathematical Society, 2003.

[Vil08] Cédric Villani, Optimal transport: old and new, vol. 338, Springer, 2008.

[ZMAS05] L. Zhang, P. A. Mykland, and Y. Ait-Sahalia, A tale of two time scales: Determining integrated volatility with noisy high-frequency data, Journal of the American Statistical Association 100 (2005), 1394–1411.

Department of Mathematics, University of Washington, Seattle, WA 98195

E-mail address: soumikpal@gmail.com, wongting@uw.edu